IJNS

International Journal of Network Security

ISSN 1816-353X (Print) Vol. 18, No. 1 (Jan. 2016)
ISSN 1816-3548 (Online)

INTERNATIONAL JOURNAL OF NETWORK SECURITY

Editor-in-Chief

Prof. Min-Shiang Hwang
Department of Computer Science & Information Engineering, Asia University, Taiwan

Co-Editor-in-Chief:

Prof. Chin-Chen Chang (IEEE Fellow)
Department of Information Engineering and Computer Science, Feng Chia University, Taiwan

Publishing Editors

Shu-Fen Chiou, Chia-Chun Wu, Cheng-Yi Yang

PUBLISHING OFFICE

Min-Shiang Hwang
Department of Computer Science & Information Engineering, Asia University, Taichung 41354, Taiwan, R.O.C.
Email: mshwang@asia.edu.tw

International Journal of Network Security is published both in traditional paper form (ISSN 1816-353X) and in Internet (ISSN 1816-3548) at http://ijns.jalaxy.com.tw

PUBLISHER: Candy C. H. Lin

International Journal of Network Security

Vol. 18, No. 1 (Jan. 1, 2016)

Design and Analysis of Lightweight Trust Mechanism for Secret Data using Lightweight Cryptographic Primitives in MANETs

Adarsh Kumar[1], Krishna Gopal[1], and Alok Aggarwal[2]
(Corresponding author: Adarsh Kumar)

Computer Science Engineering and Information Technology Department, Jaypee Institute of Information Technology[1]
A-10, Sector-62, Noida, India
(Email: adarsh.kumar@jiit.ac.in)
JP Institute of Engineering and Technology, Meerut[2]
Mawana Road, P.O. RAJPURA, Rajpura Meerut, Uttar Pradesh, India

(Received May 17, 2013; revised and accepted Apr. 20 & Nov. 6, 2014)

Abstract

Lightweight trust mechanism with lightweight cryptography primitives and post-quantum cryptosystems are having important concerns in resource constraint wireless sensor based Mobile Ad Hoc Networks (MANETs). In post-quantum cryptosystems, error correcting codes (ECC) help in code based cryptography for lightweight identification, authentication, distance bounding and tag with ownership transfer protocols to provide security. In this work, a novel approach is designed to secure the RFID-Sensor based MANET that uses ECC for assigning identification to resource constrained mobile nodes. This assignment helps to create centralized environment with subgroups, groups and hierarchies. Group or subgroups boundaries are limited through distance bounding protocols. Trust management plays the role of maintaining the relationship between nodes for long endeavor. Probability analysis of distance bounding protocol shows that the proposed approach is protected from mafia fraud, distance fraud, terrorist fraud, and distance hijacking attacks. The success of these attacks on the proposed mechanism dependence on trust score: lesser trust score (≤ 50) increases the chances of these attacks whereas higher trust score protects the network from these attacks and improves the network performance as well. In performance analysis, it is observed that the Zone Routing Protocol (ZRP) outperforms the other MANET routing protocols in terms of network performance and security for the proposed scheme. However, the probabilistic analysis proves that it is still possible to control outliers in the network despite the new inserted defenses with trust management and limited resources.

Keywords: MANET, RFID, zone routing protocol

1 Introduction

Radio frequency identification (RFID) devices are the low cost computing devices for automatic identification, locating and tracking objects using radio frequency (RF). RFID networks are having many applications like: access rights, object tracking, inventory management, library management etc. RFID devices are classified into three major components: tag, reader and back-end system. Tag includes the identification mark and a small memory unit to store information about product, object or environment. Reader helps to write and/or read information to tag. The read information is delivered to backend system for storage, migration etc. Wireless sensor networks (WSNs) and RFIDs are the two complementary technologies. WSNs consist of small sensing devices with wireless communication medium. In compliment to RFID, WSNs consist of multi-hop, smart sensing, tracking and reprogrammable devices. However, integration of WSNs and RFIDs provides sensors to read tags, intelligence, sensing, ad-hoc and wireless communication facilities. These facilities result in many advantages which include: network-resource-data expandability, network-information scalability, portable readers extendability for speeding the on spot and random data collection, reducing hardware cost etc. [45, 74]. Requirements to integrate RFID-sensor network include accurate and reliable communication, energy efficiency and network maintenance [19, 74]. Various proposals are given to integrate RFID and sensor networks. In [72, 74], three types of integration mechanisms are proposed. In first integration mechanism, RFID tags are integrated with sensor devices. In this mechanism, two approaches are suggested to integrate RFID tags and sensors. In first approach, tags are integrated with sensor devices and communicate only with readers. Second approach suggest to integrate tag with sensor devices and they communicate with each other to construct an ad hoc network. In second integration mechanism, reader are integrated with sensor devices [24, 74]. In this mechanism, readers attached with sensors collect data from RFID tags. Readers-sensor attachment communicates to route the information and construct an ad hoc network. In [26], a commercial solution to integrate RFID and mobile devices is proposed. This solution helps to construct MANET. In third integration mechanism, a mixed architecture is proposed. In this architecture, tags and sensor nodes are kept independent but coexist in same network. Mixed architecture consist of smart stations, RFID tags and sensor nodes. Smart stations are composed of RFID reader, a microprocessor and a network interface. Both RFID and Sensor networks are pervasive networks and require more attention on all aspects of its security. Security aspects in these networks include access rights, identification, authentication, authorization, ownership transfer, hardware cryptographic implementation, message delivery guarantee, security threats, tampering, forging etc. [5, 40]. Among WSNs, security and privacy issues include physical attacks, jamming, tampering at physical layer, packet disruption and collision at data link layer, spoofing, sybil, altering, replaying, wormhole and sinkhole attacks at network layer, flooding at transport layer, cloning, incorrect location reference, data aggregation, time synchronization and masquerading attacks in service and application layer. Among RFIDs, security and privacy issues include spoofing, cloning, tampering, tracking, denial of service, etc. [62]. Solutions to these security and privacy issues are achievable through cryptography or detection and prevention mechanisms [62]. Cryptography is an art of writing or solving the codes which is classified into symmetric and asymmetric cryptosystem.

Asymmetric cryptosystem is considered to be more secure than symmetric cryptosystem. In asymmetric cryptosystem, key can be easily shared between two parties without the need to pre-establish any key. But algorithms of asymmetric cryptosystem can be easily broken using quantum computers [58]. Thus, Elliptic Curve Cryptosystem (ECCr), ElGamal Cryptosystem, RSA, etc. are not considered to be secure against quantum computers [14]. Hence demand of designing secure system increases and it results to post quantum cryptosystem [51]. Post quantum cryptosystem can be classified as: Hash based, Lattice based, Coding based, Multivariate-quadratic and Secret key cryptosystem [11]. These systems are considered to be secure against quantum computers. Both RFID and sensor based Mobile Ad Hoc Networks (MANETs) are resource constraint devices and thus require lightweight cryptographic primitives. These lightweight cryptographic aspects should be accommodated within one third of the total hardware available. This space may increase three to four times at lesser cost in future [76]. Lightweight hierarchical error correcting codes are an efficient approach for node interconnection in resource constraint devices [10]. Such hierarchical systems decrease the losses, errors, noises, implementation overhead and improve performance, throughput, goodput, etc. In order to achieve complete security, lightweight cryptographic primitives can be integrated with hierarchical distribution for achieving the necessary performance and security.

For achieving complete system security, a three dimensional McCumber Cubes model suggests various cryptographic primitives: transmission, storage, processing, confidentiality, integrity, availability, human factor, policy with practices and technology [47]. During these phases various aspects are taken into consideration like: user rights and roles, usage policies, trust policies, password policy, authentication policy, security policies, educating security policy, training policies, privacy rights, etc. Trust management is an important aspect of consideration. Trust is a behavior assessment and it is defined in many ways [4, 22, 23, 33, 46, 48, 64]. Trust can be measured based on various aspects like: integrity, ability and benevolence, key generation, identification, information secrecy, simulator aspects, etc. [32, 69]. In this work trust is used to establish and maintain relationships between nodes.

The current study proceeds as follows. Section 2 provides background on lightweight cryptographic primitives, protocols and trust management. Section 3 introduces the assumption and premises used in this work. In section 4, proposed method for integrating lightweight identification, lightweight authentication, lightweight distance bounding, lightweight tag and ownership transferred is presented using lightweight trust management mechanism. Section 5 describes the probability based attack analysis in distance bounding protocols. Simulation and protocol policy analysis of proposed hierarchical network is also presented in section 5. Finally, section 6 concludes the work.

2 Background

Lightweight cryptography is classified as: lightweight primitives and lightweight protocols [2]. Two major classes of lightweight primitives are: symmetric and asymmetric primitives. Symmetric primitives include block cipher, stream cipher, hash function, pseudo random number generation and asymmetric primitives include number based system, discrete logarithmic construction and curve based cryptosystem. Lightweight Protocols can be classified as: identification, authentication, distance bounding, yoking, tag ownership protocols, etc. In resource constraint devices, upto 30% of gate equivalents (GEs) are available for lightweight cryptographic primitives and protocols [34, 53]. These GEs can increase with advancement of technology [49].

On radio frequency signal, authenticity and validity of users and messages is achieved through cryptographic primitives, ultra-lightweight operations, EPC-global Class1 Generation2 protocols, physical primitives, etc. [2]. Unique serial number generation [35, 41, 44, 65] and plausibility check [44, 52] are the authentication mechanisms without using tags. These protocols are application dependent solutions for authentication with proper justification. A leak in justification enhances the chance of un-authenticated users become part of network. Authentication solutions through cryptography avoid cloning. For example: encryption/decryption, hash-lock, hash based synchronous secret, Hopper and Blum (HB), pseudo random number based protocols, zero knowledge device authentication, etc. are cryptography mechanisms for providing authentication [44]. In another solution [50], physical properties of product stores the unique and cryptography based data for avoiding counterfeiting and un-authorized access. Apart solutions from cryptography, specific security model based requirements for authentication is considered to be a valid choice [13]. Among these protocols, traceability, de-synchronization, man-in-middle, cracking codes using basic binary operation, etc. are commonly found to be the attacks [6, 15, 61]. Cryptography based authentication solutions are costlier also. For example, although hash based solution are found to be perfect in security but the hardware cost for implementing a hash based solution proposed is almost infeasible solution [60]. Hash based solutions like: RIP, RAP, O-RAP, O-RAKE, etc. easily avoids the traceability attacks. Cryptography based stored information containing unique identification, anonymity and anti-cloning mechanism provides maximum security through hashing only [12]. In [3], it is found that computational workload and scalability are the major challenges in hash based schemes. However, solutions has been proposed to increased the scalability and security of authentication protocols through hashing. For example, Avoine mutual authentication protocol is a two phase hash based mechanism and it is designed to increase the scalability and security. Here, scalability is limited with distance bounding and removal of distance based frauds.In lightweight cryptography, various solutions for lightweight authentication protocols are proposed. For example, Lightweight Mutual Authentication protocol (LMAP) [67]. LMAP provides security against replay, forgery, anonymity, etc. However, this protocol is not secure against traceability attack. Protocol for Lightweight Authentication of IDentity (PLAID) provides authentication and enhances the privacy through confidentiality and integrity [9]. This solution is designed for contactless smart card systems. Efficiency and reduction of costs are the real advantages of this protocol. It also provides fast and strong security between smart card and terminal devices. Strong security is achieved by not leaking the identity information.

Trust Management involves trust measurement, trust propagation, trust accumulation, trust prediction and trust application [20, 28, 29]. Trust measurement is a subjective calculation that one node has to establish on another. Trust measurement among various nodes of a resource constraint network is another challenge. CuboidTrust is a positive or negative signal based global trust computational method [18]. This method also helps to determine quality and contribution of nodes in a network. EigenTrust is satisfactory or unsatisfactory transaction based method with malicious node identification [36].

Health of resource constraint mobile nodes plays an important role in measuring the trust score. In this work, health is measured with the help of three components: lightweight energy measurements, lightweight route acting algorithms and lightweight vibration signals. Lightweight energy conservation and measurement algorithms in lightweight mobile sensor networks with ability of full coverage play an important role in trust computation. Energy in ad hoc networks is consumed through three modes: transmitting, receiving or simply ”on” [25]. Saving energy increases the lifetime and utilization of ad hoc nodes. Transmitting data is major source of energy consumption among three components [25]. Receiving or collecting information is divided into four major components: discovery, data transfer, routing and motion control [27]. Discovery information can be collected from either of the two methods: Mobility independent protocols or knowledge based protocols. Mobility independent protocols are further classified into three schemes: scheduled rendezvous, on-demand and asynchronous [27]. Schedules based protocols classification involve time slot, frequency based and spread spectrum codes [75]. In these types of networks, slots are fixed for every node thus no chance of collision or overhead, easy to implement and energy efficient but assigning numbers to nodes for specific slot can prolonged delay. For example, Chakrabarti et. al. proposed a wake up mechanism on time schedule [17]. Zhang et. al. proposed ZebraNet based on global positioning system (GPS) and derivation of schedule mechanism [73]. Other examples of scheduling based protocols developed for sensor nodes are: TRAMA [56], FLAMA [55], SMACS [59], SRSA [68], R-MAC [71], DW-MAC [62, 75], etc. On-demand protocols are based on wakeup calls. Whenever some event signals to channel, it intimates to the sensor node and that node power up the data radio and start transmission. In this type of protocols, two types of signals are required to complete the process: one for wakeup call and second for data transmission. Various mechanisms are used to complete this functionality. Wakeup call could be performed using low frequency and data transmission through high frequency [57], wakeup call and data transmission call are performed using separate messages [70].

3 Proposed Scheme

Table 1 shows the symbols used in this work.

3.1 Lightweight Identification

In order to reduce the computation cost, Reed-Muller codes is used for identifying the tags. $BC_{2^n}^{M^{(a,b)}_{(c,d)}}$

Table 1: Symbols

Symbol	Quantity
$M_{(c,d)}^{(a,b)}$	c^{th} mobile node in d_{th} subgroup at a^{th} layer with b^{th} network. Here, $a, b, c, d\epsilon\{1, 2...\infty\}$.
$BC_{2^n}^{M_{(c,d)}^{(a,b)}}$	binary code selected for $M_{(c,d)}^{(a,b)}$.
$SM_{(e,d)}^{HL_a}$	e^{th} subgroup member in d^{th} subgroup at a^{th} layer.
$CW_{BC_{2^n}^{M_{(c,d)}^{(a,b)}}}$	codeword generated with length L and distance D.
$SG_d^{HL_a}$	d^{th} subgroup at a^{th} layer. Selection of $SG_d^{HL_a}$ is based on HEALTH, i.e. HEH^{MN_a}.
HEH^{MN_a}	HEALTH,$HEH^{MN_a}\epsilon f\{ES^{M_a}, RAS^{M_a}, VIB_+^{SM_{(e,d)}^{HL_a}}\}$.
ES^{MN_a}	energy state
RAS^{MN_a}	router acting strength moment
$VIB_+^{SM_{(e,d)}^{HL_a}}/VIB_-^{SM_{(e,d)}^{HL_a}}$	positive/negative vibration signals send from subgroup member
PS^{MN_a}	a^{th} mobile node in its full energy and without being attacked
$SG_{SC_d^{HL_a}}$	subgroup controller of d^{th} subgroup at hierarchical layer HL_a

is an ary code with elements $(CW_{BC_{2^n}^{M_{(c,d)}^{(a,b)}}}, L, D)$. A new binary code for next node is generated as $BC(m)_{2^n}^{M_{(c,d)}^{(a,b)}} = BC(m_1)_{2^n}^{M_{(c,d)}^{(a,b)}} * BC(m_2)_{2^n}^{M_{(c,d)}^{(a,b)}} =\{(X, X + Y), X\epsilon BC(m_1)_{2^n}^{M_{(c,d)}^{(a,b)}}$ and $Y\epsilon BC(m_2)_{2^n}^{M_{(c,d)}^{(a,b)}}\}$. Major strengths of this coding technique are: (i) with the help of small key size it provides strong security, (ii) it reduces the probability of cheating some node to a great extent and (iii) computational complexity is very less. Weakness of this coding technique is that it is prone to structural attack.

3.2 Lightweight Grouping

Trust management plays an important role for forming secure local subgroups for information exchange. It is also necessary to integrate additional trust security layer to resource constraint sensor nodes since cryptographic primitives do not provide complete security and any extra computation is not feasible on these nodes [37]. In order to compute trust, following steps are followed: (a) gather node information, (b) propagate information, (c) map to trust model and make trust decision [37].

3.2.1 Gather Node Information

Taarget node's reliability for information transfer can easily be calculated through neighboring nodes. Neighbor node can send $VIB_+^{SM_{(e,d)}^{HL_a}}$ or $VIB_-^{SM_{(e,d)}^{HL_a}}$ signal towards $SM_{(e,d)}^{HL_a}$. Strength of signal can be calculated through different ways such as: forwarded packets, intentionally dropped packets, intentionally forward packet through some legitimate intermediate node, impersonation or masquerading of data to bogus data, probability of some event, etc. Probability of finding an anomaly in attending or reporting in a regular event is helpful for providing neighboring node information [43]. Now, probability of following a path from source $(SR^{(x_1,y_1)})$ to destination $(DT^{(x_n,y_n)})$ is identified using Markov chain. $P(SR_1^{(x_1,y_1)}, SR_2^{(x_2,y_2)}, SR_3^{(x_3,y_3)}.DT_n^{(x_n,y_n)}) = P(SR_1^{(x_1,y_1)} = SR_1^{(x_1,y_1)} * p_{x_1x_2} * p_{x_2x_3} ..... * p_{x_{n-1}x_n} = P_S$, i.e. when probability reaches zero then that particular region is called an event region. When a node follows a particular path, Frisbee model [16] is used to construct subgroups. This model in resource constraint network reduces losses. Figure 1 shows the construction of Frisbees with fixed number of nodes. In the process of creating single-hop Frisbees, node communicates with other node through lightweight and energy efficient authentication mechanism.

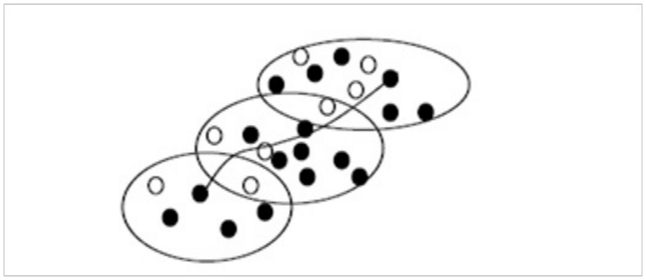


Figure 1: Frisbee construction with mobility of node

3.2.2 Propagate Information

Once subgroups are constructed then these subgroups are merged to form hierarchy. Each $SG_{M_{(c,d)}^{(a,b)}}^{HL_i}$ at every hierarchical layer will contain a subgroup controller. Figure 2 shows the construction of hierarchy with movement of $M_{(c,d)}^{(a,b)}$ that may take the form of $SG_{SC_d}^{HL_i}$. As shown in Figure 2, $M_{(c,d)}^{(a,b)}$ will act as producer (P_i) or consumer

($C_i^{'}$ or $C_i^{''}$). These producer and consumer will perform multiple tasks like: (i) distribution of $BC(m_1)_{2^n}^{M_{(c,d)}^{(a,b)}}$, (ii) with the help of $BC(m)_{2^n}^{M_{(c,d)}^{(a,b)}}, SG_{SC_d}^{HL_i}$ generate keys and distribute to consumers and (iii) nodes exchange messages using lightweight encryption mechanism.

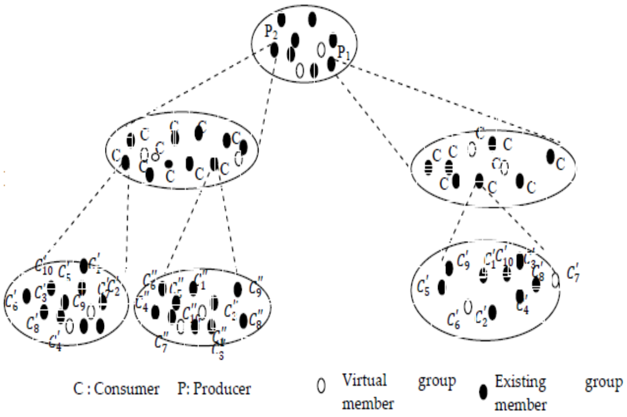


Figure 2: Hierarchical formation using real and virtual nodes

- During distribution of $BC(m)_{2^n}^{M_{(c,d)}^{(a,b)}}, P_i$ will fetch the reed-muller binary code from the database and distribute to $C_i^{'}$ or $C_i^{''}$. The producer-consumer module to exchange reed-muller code using interface, port and channel is shown in Figure 3. Here, n-consumer modules are connected to single producer and each producer/consumer module is associated with an interface. These are writing and reading interfaces at producer and consumer ends respectively. Since producers generate and consumers accept reed-muller codes thus port associated with producer is output and consumer is input.

- With the help of $BC(m)_{2^n}^{M_{(c,d)}^{(a,b)}}, SG_{SC_d}^{HL_i}$ generate keys and distribute to consumers. In [42], efficient hierarchical threshold based symmetric group key management protocol is proposed. It is found that inclusion of virtual nodes reduces the energy losses and joining/leaving expenses of nodes. Extension to Teo and Tan's group key management protocol is integrated to generate and distribute a group symmetric key 'K'' [42, 66]. Major strengths of this process are: (i) protected from forward and backward secrecy, (ii) strong authentication mechanism and (iii) efficient in terms of small subgroup formation in close vicinity.

- With help of symmetric key 'K'', messages are exchanged using protocol1 between smart nodes. Here, smart node is integration of RFID reader with mobile sensor node. Reader reads the information from nearby tags and communicates to other sensor nodes through radio frequency. A microcontroller is used to make the RFID reader data compatible for sensor node in a smart node.

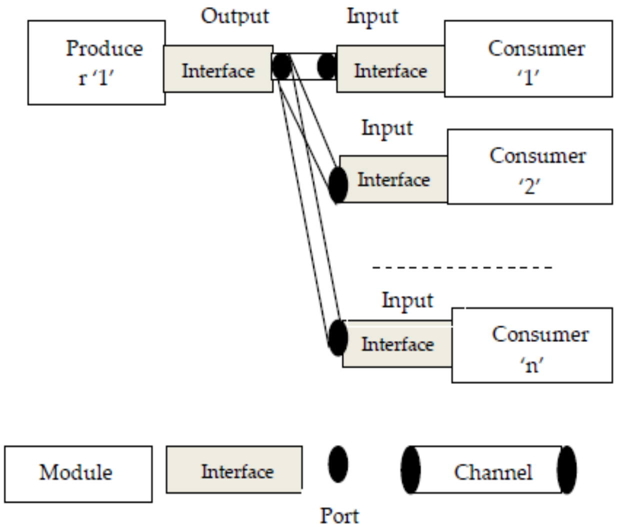


Figure 3: Exchange of $BC(m)_{2^n}^{M_{(c,d)}^{(a,b)}}, P_i$ using producer-consumer

Protocol 1: Messages exchange using lightweight encryption/decryption mechanisms.

Premise: Let E_K, D_K and H represents the lightweight encryption, decryption and hashing functions respectively.

1) $SG_{SM_j}^{HL_i} \to SG_{SM_r}^{HL_o} : \{E_K\{Message\}, H(Message)\}$.

2) $SG_{SM_r}^{HL_o}$ verifies the message digest by regenerating it using $H(D_K(E_K\{Message\}))$. If $H(D_K(E_K\{Message\})) = H(Message)$ then message is accepted otherwise rejected.

3) if message is accepted then $SG_{SM_r}^{HL_o} \to SG_{SM_j}^{HL_i}$: $\{E_K\{Acknowledgement\}, H(Acknowledgement)\}$ and if message is rejected then $SG_{SM_r}^{HL_o} \to SG_{SM_j}^{HL_i}$: $\{E_K\{Negative\_Acknowledgement\}, H(Negative\_Acknowledgement)\}$.

4) $SG_{SM_j}^{HL_i}$ verifies the receipt of message through acknowledgement as: $H(D_K(E_K\{Acknowledgement\}))$. If $H(D_K(E_K\{Acknowledgement\})) = H(Acknowledgement)$ then message is accepted otherwise retransmission start with timer.

These steps of message exchange ensures: (i) confidentiality of message exchange through encryption/decryption, (ii) message integrity through lightweight hashing hashing, (iii) pre-image resistant and collision resistant properties of messages through lightweight hashing, (iv) compression of message through hashing and (v) retransmission of messages in case of message loss or corruption.

3.2.3 Map to Trust Model

As discussed, trust management includes trust generation, trust propagation, trust accumulation, trust prediction and trust application [29]. Once subgroup is constructed using protocol1 then it can be protected from various attacks and maintains the relationships using trust management. Trust mechanism assumes every member of constructed hierarchy as PS^{MN_a} and passes through following phases for maintaining relationships.

Trust Generation: Trust on a mobile node is calculated from its HEH^{MN_a} score. Trust is directly proportional to HEH^{MN_a} score. Initially, all nodes are considered to be PS^{MN_a} and vibrate $VIB_{+}^{SM_{(e,d)}^{HL_a}}$ signal only. Here, health is calculated from three factors i.e. $HEH^{MN_a} \epsilon f\{ES^{M_a}, RAS^{M_a}, VIB_{+}^{SM_{(e,d)}^{HL_a}}\}$. Three component's values are rated on grading scheme in order to calculate the trust value of any node and this grading process is explained as follows:

- RAS^{M_a} ensures reliability and quality of service. Since all nodes are considered to be PS^{MN_a} thus reliability and quality of nodes is assumed to be very high. Reliability of node is dependent upon delivery ratio, goodput, coverage, fairness, jitter and routing cost [54]. Quality of service is calculated from number and type of interactions, which is calculated as probability score value (PSV) and it is calculated as number of times the $P(SR_1^{(x_1,y_1)}, SR_2^{(x_2,y_2)}, SR_3^{(x_3,y_3)} \cdots DT_n^{(x_n,y_n)})$ of any $M_{(c,d)}^{(a,b)}$ reaches zero in some region 'R'. Interactions in this region may transmit very good, good, average, poor or very poor quality of transmissions.

- ES^{M_a} is measured in form of bursts and sleep time. These bursts are scaled based on traffic. Low traffic consumes less energy and heavy traffic consumes high energy. In order to rate energy levels, bursts are divided into four major categories: zero, low, medium and high. Zero bursts do not consume energy and in this state, nodes are assumed to be in sleep state. Low bursts are the minimum consumption states. Medium bursts are the frequent consumption states but do not increase its value with time as compared to high bursts which are more frequent. Energy consumption increases with time if high bursts are contineously observed. Section 5 describes the energy consumption analysis.

- $VIB_{+}^{SM_{(e,d)}^{HL_a}}$ are the positive vibration signals and present experiences of neighboring nodes. A node can send positive or negative vibration signals. Positive signals are used to indicate trust and negative for un-trust. In this work, counts on positive signals are made to measure the trust. This count value ranges from 1(Low) to 10 (High). Rating is the number of trust response coming from neighboring nodes. If number of neighboring nodes exceed ten then it is considered to be highly trusted but if number of neighboring nodes response is less than ten then 10 minus total response will give negative vibration score. Subgroup signal value is also calculated from the average score of it's node's trust vibration scores. Subgroup controller can debar any subgroup from hierarchy because of its malicious operations. Which is calculated from its subgroup members health score.

Table 2: Lightweight automatic trust propagation-intruder analysis (time in msec)

Percentage age of SCORE ($HEH_{neighbor}^{MN_a}$)	Intruder Assertions	Proposed Trusted Strategy	
		Time (Steps)	**Result**
More than 90	1/5/10	20/21/26 (120/226/351)	Proved
90 to 75	1/5/10	35/42/61 (222/350/595)	Proved
75 to 60	1/5/10	41/61/74 (332/530/650)	Proved
60 to 45	1/5/10	52/74/85 (436/626/751)	Proved
Less than 45	1/5/10	62/84/95 (546/726/881)	Proved

Trust Propagation: Once trust of node is calculated then its value is propagated to other nodes. This propagation is made through selective algorithm [63]. Range of $SCORE(HEH_{neighbor}^{MN_a})$ selected for selective algorithm is analyzed using Alloy [30, 31]. Alloy is a lightweight, powerful, simple design, automatic and animation analysis tool. Table 2 shows that there are five ranges of health score: more than 90, 90 to 75, 75 to 60, 60 to 45 and less than 45. There are three variations of intruders: 1, 5 and 10 to analyze the proposed mechanism. This analysis shows that with change in every score range, there is an increase in minimum of 10 msec and 100 steps to detect intruders. However, intruders are detectable and results are proved in this tool. According to selective algorithm, single high health score neighbor is selected if $SCORE(HEH_{neighbor}^{MN_a}) \geq$ 90%, two high score neighbor are selected if 90% $\gnsim$ $SCORE(HEH_{neighbor}^{MN_a}) \geq$ 75% ,three high score neighbor are selected if 75% $\gnsim$ $SCORE(HEH_{neighbor}^{MN_a}) \geq$ 60%, four high score neighbor are selected if 60% $\gnsim$ $SCORE(HEH_{neighbor}^{MN_a}) \geq$ 45%, transmit to all neighboring nodes if 45%$SCORE(HEH_{neighbor}^{MN_a})$. Multiple entities of trust are re-evaluated in trust prediction phase through identification marks since each communication contains its identification, i.e. $BC(m)_{2^n}^{M_{(c,d)}^{(a,b)}}) \parallel SCORE(HEH_{neighbor}^{MN_a})$. This mechanism of trust propagation through health score help in protecting the network from various attacks.

Table 3: Lightweight automatic trust accumulation-intruder analysis (time in msec)

Percentage age of SCORE (HEH_{AVG})	Intruder Assertions	Proposed Trusted Strategy	
		Time (Steps)	Result
More than 90	1/5/10	31/41/44 (131/233/362)	Proved
90 to 80	1/5/10	42/52/71 (222/362/493)	Proved
80 to 70	1/5/10	54/61/88 (341/466/645)	Proved
70 to 60	1/5/10	64/81/101 (531/771/823)	Proved
60 to 50	1/5/10	74/93/118 (666/902/1120)	Proved
Less than 50	1/5/10	92/104/165 (786/966/1481)	Proved

Trust Accumulation: At destination, trust values are accumulated and evaluated. Since, trust value passes through multiple paths hence source‘s trust value is predicted from health of the path followed. Health of each routed node is accumulated along with its trust value. Average of health is calculated using: $HEH_{AVG} = (HEH^{MN_1} + HEH^{MN_2} + .. + HEH^{MN_n})/N$. Based on $score(HEH_{AVG})$ value, path is selected and rated. Table 3 shows that there are six range of score(HEH_{AVG}). With decrease in $score(HEH_{AVG})$ of 10% there is an increase in minimum of 10 msec and 100 steps to detect intruders. However, intruders are detectable and results are proved on alloy tool. This measurement is taken to rate the path followed for trust accumulation. If $score(HEH_{AVG}) \geq 90\%$, then path is considered as excellent, very good if $90\% \gneq score(HEH_{AVG}) \geq 80\%$, good if $80\% \gneq score(HEH_{AVG}) \geq 70\%$, average if $70\% \gneq score(HEH_{AVG}) \geq 60\%$, below average if $60\% \gneq score(HEH_{AVG}) \geq 50\%$, poor if $50\% \gneq score(HEH_{AVG})$.

Trust Prediction: Now, after transmitting the trust score in the form of health, healthiness of route is determined. If route health is below average then trust is recomputed at destination using lightweight trust computation based on prejudice, experience and hearsay. It is calculated as: $T^i = C * Exp. + (1 - C) * Her.$, where T, C, Exp. and Her. are respectively the trust, self confidence level, experience and hearsay values. Experience is the average value of current observation and immediate observation. Hearsay is calculated as: $H(MN^j) = (\Sigma_{i=1}^{n} T^i)/N$. Here, N is the number of neighboring connected nodes to MN_a and T^i is the i^{th} response of trust.

Trust Application: Once basic trust relationship is established then application specific trust depends upon user operations. Secure and safe transmission of information is necessary and confirmed through authentication procedures. Applications that are required to be operated in basic trusted environment should have to produce application trust value (T_a). This trust value is compared with basic trust value (T_i). If $T_a \lneq T_i$ then access to application fails. Failure or success of the application for operation is broadcasted to other subgroup members using broadcasting mechanism. Protocol 2 describes this mechanism.

Protocol 2: Application trust broadcasting for access rights.
Goal: To compare trust value with required application trust value. After this comparison, if application trust value is less then access to application is not allowed and this information is broadcasted to all subgroup members.

1) $SG^{HL_i}_{SM_j} \rightarrow SG^{HL_i}_{SM_k}$: "ALLOW" ∥ "DENY".

2) $SG^{HL_i}_{SC_k} \rightarrow SG^{HL_{i-1}}_{SC_k}$: "ALLOW" ∥ "DENY". This step is repeated until top subgroup controller receives the message.

3) $SG^{HL_o}_{SC_k}$ initiated the process of collecting information about applications whose access rights are managed through trust comparison.

Here, ALLOW and DENY are single bit messages. These messages help to debar the applications that can maliciously harm the network. If 'h' is the height of hierarchy constructed and 'n' is the total number of subgroup constructed then total number of messages required to broadcast this information are '$h * n * 10$'. In this work, a set of two node based trust applications are integrated for distance bounding. This trust application is explained in next sections.

3.3 Lightweight Trust Based Distance Bounding and Authentication

In this section, distance bounding and authentication protocols are integrated to hierarchical model for limiting the distance between two nodes and to authenticate each other. Distance bounding and authentication are two set of protocols but an integrated form of these protocols is used to reduce the hardware cost. In this work, modified form of Avoine mutual authenticated KA2 (MA-KA2) protocol is integrated with lightweight parameters [7]. The modified form of this mechanism is explained in Protocol 3. There are two phases of protocol: slow and fast. In slow phase, nonce values are exchanged and in fast phase, authentication is performed using challenge-verify process.

Protocol 3: Modified MA-KA2 and Distance Bounding Protocol.

Premises: Let $R^{M^{(a,b)}_{(c,d)}}$ be the random number selected by $M^{(a,b)}_{(c,d)}$.$N_{SG^{HL_a}_{SC_d}}$ represents the nonce generated by d^{th} subgroup with its subgroup controller. Here, every subgroup member act as a prover or a verifier. When direction bit $DIR^i_{M^{(a,b)}_{(c,d)}}$ of some mobile node is zero then $M^{(a,b)}_{(c,d)}$ sends a random challenge $CHA^i_{M^{(a,b)}_{(c,d)}} \epsilon \{0,1\}$ towards another mobile node $M^{(a,b)}_{(f,d)}$. Now, this mobile node replies with verification process ($VER^{CHA^i}_{M^{(a,b)}_{(f,d)}}$). When $DIR^i_{M^{(a,b)}_{(c,d)}}$ is one then $M^{(a,b)}_{(f,d)}$ will send $CHA^i_{M^{(a,b)}_{(f,d)}} \epsilon \{0,1\}$ and $M^{(a,b)}_{(c,d)}$ will verify. If the random number generated is not verified, i.e. $R^{M^{(a,b)}_{(c,d)}} \neq VER^{CHA^i}_{M^{(a,b)}_{(a,d)}}$ then communication is put in protected mode. This protected mode behaves differently than regular rounds. In this mode, nodes have to regularly produce and verify the challenges. Let b and r are the number of bits used in direction bit and number of rounds in two phases of distance bound mutual authentication protocol. $T_{M^{(a,b)}_{(c,d)}}$ represents the timer from $M^{(a,b)}_{(c,d)}, T_{MAX}$ is the maximum time elapsed for checking distance bounding and H is a pseudorandom number function.

Goal: Limit the distance between two subgroup controllers or members and authenticate each other.

Step 1: Slow Phase

1) Every subgroup member from both subgroups will select a random number, i.e. $R^{M^{(a,b)}_{(1,d)}}, R^{M^{(a,b)}_{(2,d)}} ... R^{M^{(a,b)}_{(9,d)}}$ and $R^{M^{(a,b)}_{(1,e)}}, R^{M^{(a,b)}_{(2,e)}} ... R^{M^{(a,b)}_{(9,e)}}$.

2) Since a symmetric key K is already shared between subgroup members thus nonce are generated using: $N_{SG^{HL_a}_{SC_d}} = H(K, R^{M^{(a,b)}_{(1,d)}} \parallel R^{M^{(a,b)}_{(2,d)}} \parallel ... \parallel R^{M^{(a,b)}_{(9,d)}})$ and $N_{SG^{HL_a}_{SC_e}} = H(K, R^{M^{(a,b)}_{(1,e)}} \parallel R^{M^{(a,b)}_{(2,e)}} \parallel ... \parallel R^{M^{(a,b)}_{(9,e)}})$. Here, H is a lightweight cryptographic hash function.

3) Two subgroup controller exchanges these nonce values as: $SG^{HL_a}_{SC_d} \rightarrow SG^{HL_a}_{SC_e} : N_{SG^{HL_a}_{SC_d}}, SG^{HL_a}_{SC_e} \rightarrow SG^{HL_a}_{SC_d} : N_{SG^{HL_a}_{SC_e}}, \{DIR^i_{SG^{HL_a}_{SC_d}} \parallel DIR^i_{SG^{HL_a}_{SC_e}} \parallel VER^{CHA^0} \parallel VER^{CHA^1} \parallel VER^{CHA^2}\} = h(K, N_{SG^{HL_a}_{SC_d}}, N_{SG^{HL_a}_{SC_e}}$ Number of bits $(DIR^i_{SG^{HL_a}_{SC_d}})$ = Number of bits$(DIR^i_{SG^{HL_a}_{SC_e}}) = r$, Number of bits (VER^{CHA^0}) $=Numberofbits(VER^{CHA^1}) = 2(b-r)-1$, Number of bits $(VER^{CHA^2}) = 2b$.

Step 2: Fast bit exchange phase

1) $SG^{HL_a}_{SC_d}$ computes $COM^1_{SG^{HL_a}_{SC_d}} = DIR^1_{SG^{HL_a}_{SC_d}}$ and start timer $T_{SG^{HL_a}_{SC_d}}$. During this time, it sends $COM^1_{SG^{HL_a}_{SC_d}}$ towards $SG^{HL_a}_{SC_e}$.

2) $SG^{HL_a}_{SC_e}$ checks if $COM^1_{SG^{HL_a}_{SC_d}} = DIR^1_{SG^{HL_a}_{SC_d}}$ then computes $COM^1_{SG^{HL_a}_{SC_e}} = VER_1^{CHA^2}$. With start of $T_{SG^{HL_a}_{SC_e}}$, $SG^{HL_a}_{SC_e}$ sends $COM^1_{SG^{HL_a}_{SC_e}}$ to $SG^{HL_a}_{SC_d}$. But if $COM^1_{SG^{HL_a}_{SC_d}} \neq DIR^1_{SG^{HL_a}_{SC_d}}$ then error is detected and instead of sending random answers until end of the protocol it check value of $HEH^{MN_a}_{SG^{HL_a}_{SC_e}}$ and HEH_{AVG}. if any value is below satisfactory level then it adds the communication in protected mode.

3) $SG^{HL_a}_{SC_d}$ stops $T_{SG^{HL_a}_{SC_d}}$ and compute $DOM^{b-1}_{SG^{HL_a}_{SC_e}} = COM^{b-1}_{SG^{HL_a}_{SC_d}} \oplus VER^{CHA^2}_{2b-3}$.if $DOM^{b-1}_{SG^{HL_a}_{SC_e}} = DIR^{b-1}_{SG^{HL_a}_{SC_e}}$ then $COM^b_{SG^{HL_a}_{SC_d}} = VER^{CHA^2}_{2b-2} \oplus DIR^b_{SG^{HL_a}_{SC_d}}$. Further, if $DOM^{b-1}_{SG^{HL_a}_{SC_e}} \neq DIR^{b-1}_{SG^{HL_a}_{SC_d}}$ then again it check for $HEH^{MN_a}_{SG^{HL_a}_{SC_d}}$ and HEH_{AVG}. If any of these values are unsatisfactory then it adds the communication to protected mode. Also, $SG^{HL_a}_{SC_d}$ sends $COM^b_{SG^{HL_a}_{SC_d}}$ to $SG^{HL_a}_{SC_e}$ and start $T_{SG^{HL_a}_{SC_d}}$.

4) $SG^{HL_a}_{SC_e}$ stops $T_{SG^{HL_a}_{SC_e}}$ and compute $DOM^b_{SG^{HL_a}_{SC_d}} = COM^b_{SG^{HL_a}_{SC_d}} \oplus VER^{CHA^2}_{2b-2}$. If $DOM^b_{SG^{HL_a}_{SC_d}} \neq DIR^b_{SG^{HL_a}_{SC_d}}$ then $HEH^{MN_a}_{SG^{HL_a}_{SC_d}}$ and HEH_{AVG} are checked before sending unsatisfactory report for protected mode. Also, $SG^{HL_a}_{SC_e}$ start $T_{SG^{HL_a}_{SC_e}}$ and send $DOM^b_{SG^{HL_a}_{SC_d}}$ to $SG^{HL_a}_{SC_d}$.

5) $SG^{HL_a}_{SC_d}$ stops $T_{SG^{HL_a}_{SC_d}}$ and compute $DOM^b_{SG^{HL_a}_{SC_e}} = DOM^b_{SG^{HL_a}_{SC_d}} \oplus VER^{CHA^2}_{2b-1}$. If $DOM^b_{SG^{HL_a}_{SC_e}} = DIR^b_{SG^{HL_a}_{SC_e}}$ then compute $COM^{b+1}_{SG^{HL_a}_{SC_d}} = VER^{CHA^2}_{2b} \oplus R^{M^{(a,b)}_{(1,d)}}$. Further, if $DOM^b_{SG^{HL_a}_{SC_e}} \neq DIR^b_{SG^{HL_a}_{SC_e}}$ then $HEH^{MN_a}_{SG^{HL_a}_{SC_d}}$ and HEH_{AVG} values are checked before sending unsatisfactory report for protected mode. $SG^{HL_a}_{SC_d}$ sends $COM^{b+1}_{SG^{HL_a}_{SC_d}}$ to $SG^{HL_a}_{SC_e}$ and start $T_{SG^{HL_a}_{SC_d}}$

6) $SG^{HL_a}_{SC_e}$ stops $T_{SG^{HL_a}_{SC_e}}$ and computes $R^{M^{(a,b)}_{(1,d)}} = COM^{b+1}_{SG^{HL_a}_{SC_d}} \oplus VER^{CHA^2}_{2b-2}$. If $R^{M^{(a,b)}_{(1,d)}} = 0$ then $COM^{b+1}_{SG^{HL_a}_{SC_e}} = VER^{CHA^0} \oplus R^{M^{(a,b)}_{(1,e)}}$ else if $R^{M^{(a,b)}_{(1,d)}} =$

1 then $COM^{b+1}_{SG^{HL_a}_{SC_e}} = VER^{CHA^1} \oplus R^{M^{(a,b)}_{(1,e)}}$. Also, $SG^{HL_a}_{SC_e}$ starts $T_{SG^{HL_a}_{SC_e}}$ and sends $COM^{b+1}_{SG^{HL_a}_{SC_e}}$ to $SG^{HL_a}_{SC_d}$.

7) $SG^{HL_a}_{SC_d}$ stops $T_{SG^{HL_a}_{SC_d}}$ and computes $R^{M^{(a,b)}_{(r-b-1,e)}} = DOM^{b-1}_{SG^{HL_a}_{SC_d}} \oplus VER^{CHA^{r-b-1}}_{2r-2b-2}$. Now, if $R^{M^{(a,b)}_{(r-1,e)}} = 0$ then $DOM^{b}_{SG^{HL_a}_{SC_d}} = VER^{CHA^0_{2r-2b-1}} \oplus R^{M^{(a,b)}_{(r-b,d)}}$ else if $R^{M^{(a,b)}_{(r-1,e)}} = 1$ then $COM^{b}_{SG^{HL_a}_{SC_d}} = VER^{CHA^1}_{2r-2b-1} \oplus R^{M^{(a,b)}_{(r-b,d)}}$. Also, $SG^{HL_a}_{SC_d}$ starts $T_{SG^{HL_a}_{SC_d}}$ and sends $COM^{b}_{SG^{HL_a}_{SC_d}}$ to $SG^{HL_a}_{SC_e}$.

8) $SG^{HL_a}_{SC_e}$ stops $T_{SG^{HL_a}_{SC_e}}$ and computes $R^{M^{(a,b)}_{(r-b-1,d)}} = DOM^{b}_{SG^{HL_a}_{SC_d}} \oplus VER^{CHA^{r-b-1}}_{2r-2b-2}$. Now, if $R^{M^{(a,b)}_{(r-b,d)}} = 0$ then $COM^{b}_{SG^{HL_a}_{SC_e}} = VER^{CHA^0}_{2r-2b-1} \oplus R^{M^{(a,b)}_{(r-b,e)}}$ else if $R^{M^{(a,b)}_{(r-b,d)}} = 1$ then $COM^{b}_{SG^{HL_a}_{SC_e}} = VER^{CHA^1_{2r-2b-1}} \oplus R^{M^{(a,b)}_{(r-b,e)}}$. Also, $SG^{HL_a}_{SC_e}$ sends $COM^{b}_{SG^{HL_a}_{SC_e}}$ to $SG^{HL_a}_{SC_d}$.

9) $SG^{HL_a}_{SC_d}$ stops $T_{SG^{HL_a}_{SC_d}}$.

Step 3: End of fast bit exchange phase and start check for processing delay.

1) $SG^{HL_a}_{SC_d}$ checks for $H(K, R^{M^{(a,b)}_{(1,d)}} \parallel R^{M^{(a,b)}_{(2,d)}} \parallel .... \parallel R^{M^{(a,b)}_{(9,d)}}) = N_{SG^{HL_a}_{SC_d}}$ and $SG^{HL_a}_{SC_e}$ checks for $H(K, R^{M^{(a,b)}_{(1,e)}} \parallel R^{M^{(a,b)}_{(2,e)}} \parallel ... \parallel R^{M^{(a,b)}_{(9,e)}} = N_{SG^{HL_a}_{SC_e}}$. If both are true and time elapsed is less than T_{MAX} then communication is successful.

Major strengths of this protocol are: (i) one subgroup controller or member can put distance limit to another subgroup controller or member, (ii) unilateral authentication is provided to protect against dismantling attack, (iii) distance bounding protocols protects from location based attacks using cryptographic characteristics integrated with physical attributes of the nodes and (iv) attack analysis in section 5 shows that the modified protocol is efficient, secure and having lowest False Acceptance Rate (FAR). The FAR is the rate of possibility of acceptance of nodes when there are chances of attack.

4 Result Analysis

4.1 Attack Analysis

4.1.1 Distance Bounding Protocol Attack Analysis

In this section, probability of success of mafia fraud, distance fraud, terrorist fraud and distance hijacking attacks are analyzed on distance bounding protocols. The analysis is explained as follows:

Attack: Mafia Fraud Attack

Description: In this attack, a malicious subgroup controller ($MSG^{HL_a}_{M^{(a,b)}_{(c,d)}}$) and a malicious group member ($MM^{(a,b)}_{(c,d)}$) are inserted in subgroups. These malicious entities communicate with original subgroup controller and members and convince them to reveal secret information [59, 68, 71]. $MSG^{HL_a}_{M^{(a,b)}_{(c,d)}}$ and $MM^{(a,b)}_{(c,d)}$ start man-in-middle attack by sending $MSG^{HL_a}_{SC_d} \rightarrow SG^{HL_a}_{SC_e} : N_{MSG^{HL_a}_{SC_d}}$ and $MSG^{HL_a}_{SC_e} \rightarrow SG^{HL_a}_{SC_d} : N_{MSG^{HL_a}_{SC_e}}$. This effects the rounds of fast bit exchange. Now, success probability of this attack is determined by defining the following events:

- $AND^{i}_{SG^{HL_a}_{SC_d}}$ attack is not detected at i^{th} round by $SG^{HL_a}_{SC_d}$.
- $AD^{i}_{SG^{HL_a}_{SC_d}}$ attack is detected at i_{th} round by $SG^{HL_a}_{SC_d}$.
- $HEH\_AND^{MN_a}_{SG^{HL_a}_{SC_d}}$ health score of $SG^{HL_a}_{SC_d}$ at time when attack is not detected at i^{th} round by $SG^{HL_a}_{SC_d}$.
- $UAND^{i}_{SG^{HL_a}_{SC_d}}$ attack is not detected at until the i^{th} round by $SG^{HL_a}_{SC_d}$.
- $AND^{i}_{SG^{HL_a}_{SC_e}}$ attack is not detected at i^{th} round by $SG^{HL_a}_{SC_e}$.
- $AD^{i}_{SG^{HL_a}_{SC_e}}$ attack is detected at i_{th} round by $SG^{HL_a}_{SC_e}$.
- $HEH\_AND^{MN_a}_{SG^{HL_a}_{SC_e}}$ health score of $SG^{HL_a}_{SC_e}$ at time when attack is not detected at i^{th} round by $SG^{HL_a}_{SC_e}$.
- $UAND^{i}_{SG^{HL_a}_{SC_e}}$ attack is not detected at until the i^{th} round by $SG^{HL_a}_{SC_e}$.
- $COL^{i}_{SG^{HL_a}_{SC_d}}$ is an event when collision occurs at $SG^{HL_a}_{SC_d}$ side in i^{th} round.
- $COL^{i}_{SG^{HL_a}_{SC_e}}$ is an event when collision occurs at $SG^{HL_a}_{SC_e}$ side in i^{th} round.

Now, success probability of Mafia fraud attack can be calculates as follows:

$$\begin{aligned} & P[FAR] \\ = \; & P[UAND^i_{SG^{HL_a}_{SC_d}}/UAND^i_{SG^{HL_a}_{SC_e}}]P[UAND^i_{SG^{HL_a}_{SC_e}}] \\ & + \sum_{i=1}^{n} P[UAND^i_{SG^{HL_a}_{SC_e}}/AD^i_{SG^{HL_a}_{SC_d}}]P[AD^i_{SG^{HL_a}_{SC_d}}] \\ & + \sum_{i=1}^{n} P[UAND^i_{SG^{HL_a}_{SC_d}}/AD^i_{SG^{HL_a}_{SC_e}}]P[AD^i_{SG^{HL_a}_{SC_e}}] \end{aligned} \quad (1)$$

$$\begin{aligned} & P[UAND^i_{SG^{HL_a}_{SC_e}}/AD^i_{SG^{HL_a}_{SC_d}}]P[AD^i_{SG^{HL_a}_{SC_d}}] \\ = \; & \Pi_{j=1}^{i-1} P[\frac{UAND^i_{SG^{HL_a}_{SC_e}}}{AND^i_{SG^{HL_a}_{SC_d}}}]_{HEH\_AND^{MN_a}_{SG^{HL_a}_{SC_d}}=satisfactory} \\ & \Pi_{j=1}^{i-1} P[\frac{UAND^i_{SG^{HL_a}_{SC_e}}}{AD^i_{SG^{HL_a}_{SC_d}}}]_{HEH\_AND^{MN_a}_{SG^{HL_a}_{SC_d}}=satisfactory} \end{aligned} \quad (2)$$

Now, there are five case when $HEH\_AND^{MN_a}_{SG^{HL_a}_{SC_d}} = satisfactory$. Let $\frac{1}{p_{90}}, \frac{1}{p_{80}}, \frac{1}{p_{70}}, \frac{1}{p_{60}}$ and $\frac{1}{p_{50}}$ are the five case probabilities when $HEH\_AND^{MN_a}_{SG^{HL_a}_{SC_d}} \geq 90\%$, $HEH\_AND^{MN_a}_{SG^{HL_a}_{SC_d}} \geq 80\%$, $HEH\_AND^{MN_a}_{SG^{HL_a}_{SC_d}} \geq 70\%$, $HEH\_AND^{MN_a}_{SG^{HL_a}_{SC_d}} \geq 60\%$,and $HEH\_AND^{MN_a}_{SG^{HL_a}_{SC_d}} \geq 50\%$ respectively. If $\frac{1}{p_{i-1}}$ be the probability that collision is not detected until $(i-1)^{th}$ round and $\frac{1}{p_{protected}}$ is the probability of moving to protected mode then:

$$\begin{aligned} & \Pi_{j=1}^{i-1} P[\frac{UAND^i_{SG^{HL_a}_{SC_e}}}{AND^i_{SG^{HL_a}_{SC_d}}}]_{HEH\_AND^{MN_a}_{SG^{HL_a}_{SC_d}}=satisfactory} \\ = \; & (\frac{1}{p_{j-1}})^{j-1}(\frac{1}{p_{protected}})^{j-1} \\ & +(\frac{1}{p_{90}}+\frac{1}{p_{80}}+\frac{1}{p_{70}}+\frac{1}{p_{60}}+\frac{1}{p_{50}})^{j-1}. \end{aligned}$$

Thus Equation (2) can be written as:

$$\begin{aligned} & P[UAND^i_{SG^{HL_a}_{SC_e}}/AD^i_{SG^{HL_a}_{SC_d}}]P[AND^i_{SG^{HL_a}_{SC_d}}] \\ = \; & (\frac{1}{p_{i-1}})^{i-2}(\frac{1}{p_{protected}})^{i-2} \\ & +(\frac{1}{p_{90}}+\frac{1}{p_{80}}+\frac{1}{p_{70}}+\frac{1}{p_{60}}+\frac{1}{p_{50}})^{i}. \end{aligned} \quad (3)$$

Similarly,

$$\begin{aligned} & \sum_{i=1}^{n} P[UAND^i_{SG^{HL_a}_{SC_d}}/AD^i_{SG^{HL_a}_{SC_e}}]P[AD^i_{SG^{HL_a}_{SC_e}}] \\ = \; & (\frac{1}{p_{i-1}})^{i-2}(\frac{1}{p_{protected}})^{i-2} \\ & +(\frac{1}{p_{90}}+\frac{1}{p_{80}}+\frac{1}{p_{70}}+\frac{1}{p_{60}}+\frac{1}{p_{50}})^{i}. \end{aligned} \quad (4)$$

From Equations (3) and (4), one of the equation is used to find error thus it reduces the probability of finding a collision to be $\frac{1}{2}$. After putting values of Equations (3) and (4) in (2), probability of false acceptance rate can be calculated as:

$$\begin{aligned} P[FAR_n] \; = \; & (\frac{1}{p_{i-1}})^{n}(\frac{1}{p_{protected}})^{n} \\ & +(\frac{1}{p_{90}}+\frac{1}{p_{80}}+\frac{1}{p_{70}}+\frac{1}{p_{60}}+\frac{1}{p_{50}})^{n} \\ & +\sum_{i=1}^{n}((\frac{1}{p_{i-1}})^{n-i-2}(\frac{1}{p_{protected}})^{n-i-2} \\ & +(\frac{1}{p_{90}}+\frac{1}{p_{80}}+\frac{1}{p_{70}}+\frac{1}{p_{60}}+\frac{1}{p_{50}})^{n-i-2}). \end{aligned} \quad (5)$$

Equation (5) gives the false acceptance probability. Higher value of this probability give less protection against intruders at earlier stage. However, progression of relationship through trust decreases the probability and increases the security of network for finding an attack. If health score does not permit to accept any subgroup controller or member then collision can stop the process of communication at early stage.

Attack: Distance Fraud Attack

Description: A malicious node can come closer to subgroup and make false claim to be the nearest node. [7, 38, 39]. Let $EVENT^i_{SG^{HL_a}_{SC_e}}$ and $EVENT^i_{SG^{HL_a}_{SC_d}}$ are the events when $SG^{HL_a}_{SC_e}$ and $SG^{HL_a}_{SC_d}$ find collision. A collision can occur when some bits are not verified. Now, success probability of distance fraud attack can be calculated as:

$$\begin{aligned} & P[EVENT^i_{SG^{HL_a}_{SC_e}} \cap EVENT^i_{SG^{HL_a}_{SC_d}}] \\ = \; & (P[EVENT^1_{SG^{HL_a}_{SC_e}}]P[\frac{EVENT^2_{SG^{HL_a}_{SC_e}}}{EVENT^1_{SG^{HL_a}_{SC_e}}}] \\ & \cdots P[\frac{EVENT^n_{SG^{HL_a}_{SC_e}}}{\Pi_{i=1}^{n-1} EVENT^i_{SG^{HL_a}_{SC_e}}}]_{HEH=satisfactory} \\ & +(P[EVENT^1_{SG^{HL_a}_{SC_d}}]P[\frac{EVENT^2_{SG^{HL_a}_{SC_d}}}{EVENT^1_{SG^{HL_a}_{SC_d}}}] \\ & \cdots P[\frac{EVENT^n_{SG^{HL_a}_{SC_d}}}{\Pi_{i=1}^{n-1} EVENT^i_{SG^{HL_a}_{SC_d}}}]_{HEH=satisfactory}. \end{aligned}$$

Now, when $DIR^1_{SG^{HL_a}_{SC_d}}$ or $DIR^1_{SG^{HL_a}_{SC_e}}$ is zero then:

$$\begin{aligned} & P[EVENT^i_{SG^{HL_a}_{SC_e}} \cap DIR^i_{SG^{HL_a}_{SC_e}} \\ & \quad \cap HEH\_AND^{MN_a}_{SG^{HL_a}_{SC_d}} = satisfactory] \\ = \; & \frac{1}{2}(\frac{1}{p_{90}}+\frac{1}{p_{80}}+\frac{1}{p_{70}}+\frac{1}{p_{60}}+\frac{1}{p_{50}}) \\ = \; & P[EVENT^i_{SG^{HL_a}_{SC_d}} \cap DIR^i_{SG^{HL_a}_{SC_d}} \\ & \quad \cap HEH\_AND^{MN_a}_{SG^{HL_a}_{SC_d}} = satisfactory]. \end{aligned} \quad (6)$$

When $DIR^1_{SG^{HL_a}_{SC_d}}$ or $DIR^1_{SG^{HL_a}_{SC_e}}$ is one then:

$$\begin{aligned}
& P[EVENT^i_{SG^{HL_a}_{SC_e}} \cap DIR^1_{SG^{HL_a}_{SC_e}} \\
& \qquad \cap HEH\_AND^{MN_a}_{SG^{HL_a}_{SC_e}} = satisfactory] \\
= \; & P[EVENT^i_{SG^{HL_a}_{SC_e}} \cap DIR^i_{SG^{HL_a}_{SC_e}}] \\
& \qquad P[HEH\_AND^{MN_a}_{SG^{HL_a}_{SC_e}} = satisfactory] \\
= \; & P[(EVENT^i_{SG^{HL_a}_{SC_e}} \\
& \cap VER^{CHA^1} = h[K, N_{SG^{HL_a}_{SC_d}}, N_{SG^{HL_a}_{SC_e}})] \\
& P[DIR^i_{SG^{HL_a}_{SC_e}}]P[HEH\_AND^{MN_a}_{SG^{HL_a}_{SC_e}} = satisfactory] \\
& +P[(EVENT^i_{SG^{HL_a}_{SC_e}} \\
& \cap VER^{CHA^1} \neq h[K, N_{SG^{HL_a}_{SC_d}}, N_{SG^{HL_a}_{SC_e}})] \\
& P[DIR^i_{SG^{HL_a}_{SC_e}}]P[HEH\_AND^{MN_a}_{SG^{HL_a}_{SC_e}} = satisfactory] \\
= \; & (\frac{3}{4})^i + (\sum_{i=1}^{n}(\frac{1}{p_{i-1}})^{n-i-2} * (\frac{1}{p_{protected}})^{n-i-2} \\
& +(\frac{1}{p_{90}} + \frac{1}{p_{80}} + \frac{1}{p_{70}} + \frac{1}{p_{60}} + \frac{1}{p_{50}})^{n-i-2}). \qquad (7)
\end{aligned}$$

Since collision is found in one of the two sides thus in this case also probability is considered to be $\frac{1}{2}$. Equation (7) gives the value of acceptance rate of attack. Higher value of trust reduces the chances of this attack to a great extent.

Attack: Terrorist Fraud Attack.

Description: In this attack, existing $M^{(a,b)}_{(c,d)}$ act as malicious entity. $M^{(a,b)}_{(c,d)}$ collaborate with $MM^{(a,b)}_{(c,d)}$ and tries to convince $MSG^{HL_a}_{M^{(a,b)}_{(c,d)}}$ that he is nearby when he is not [7, 39, 38]. This attack can be protected using secret sharing scheme [8]. P[success of terrorist fraud attack] $\geq P$ [success of mafia fraud attack]. Let $P[M^{(a,b)}_{(c,d)} \rightarrow MM^{(a,b)}_{(c,d)} : Cert(MN^{HL_i}_{SM_{j+1}}), N_M, SKAL_M] = \frac{1}{pterrorist}$. $P[MM^{(a,b)}_{(c,d)} \rightarrow M^{(a,b)}_{(c,d)} : Cert(VN^{(a,b)}_{(c,d)})]_{HEH\_AND^{MN_a}_{M^{(a,b)}_{(c,d)}}} = \frac{1}{p_{terrorist}} + \frac{1}{p_{90}} + \frac{1}{p_{80}} + \frac{1}{p_{70}} + \frac{1}{p_{60}} + \frac{1}{p_{50}}$. Since symmetric key K is known to all thus $P[M^{(a,b)}_{(c+1,d)} \rightarrow MM^{(a,b)}_{(c,d)} : E_{PK_{VN^{(a,b)}_{(c,d)}}}\{SK_{M^{(a,b)}_{(c+1,d)}}\}]_{HEH\_AND^{MN_a}_{M^{(a,b)}_{(c,d)}}} = (\frac{1}{p_{terrorist}} + \frac{1}{p_{90}} + \frac{1}{p_{80}} + \frac{1}{p_{70}} + \frac{1}{p_{60}} + \frac{1}{p_{50}})^2$. and it is easy to mislead any communication by existing members. With increase in such communication chances of terrorist fraud detection increases because trust score decreases. If probability of $M^{(a,b)}_{(c,d)}$ for self answered question is marked as $\frac{1}{p_{self\_answered}}$ then P[success of terrorist fraud attack]$=(\frac{1}{p_{self\_answered}})^q * (\frac{1}{p_{terrorist}} + \frac{1}{p_{90}} + \frac{1}{p_{80}} + \frac{1}{p_{70}} + \frac{1}{p_{60}} + \frac{1}{p_{50}} + (t-1)/t + p_{self\_answered})^q$. where '$t$' is the total number of queries exchanged between $M^{(a,b)}_{(c,d)}$ and $MM^{(a,b)}_{(c,d)}$ and collision does not found in q rounds.

Attack: Distance Hijacking Attack

Description: This attack is different from distance fraud and terrorist fraud attack. In distance fraud, a dishonest prover and verifier are involved. In terrorist fraud, dishonest prover involves with other attacker but in the distance hijacking attack, dishonest prover interacts with honest prover and involves them for false distance [21]. In distance hijacking attack, minimum single dishonest prover is involved with the other honest parties. If other parties behave like dishonest prover or verifiers then this attack become distance fraud attack. Now, P[Success of distance hijacking attack] $\leq P$[Success of distance fraud attack] [38]. P[Success of distance hijacking attack] $=$ P [honest nodes reveal secret information without being dishonest]. Any dishonest node can behave as honest through masquerading, impersonation, taking false ownership, etc. This dishonest behavior in tags can be checked through birthday paradox and trust score. Now according to birthday paradox, probability of matching two numbers when number of nodes are 10 in each subgroup is less than $\frac{1}{8}$. Further, trust score reduces the probability of this attack to $(\frac{1}{p_{90}} + \frac{1}{p_{80}} + \frac{1}{p_{70}} + \frac{1}{p_{60}} + \frac{1}{p_{50}})$. This probability of success of distance hijacking attack due to trust score is much less than $\frac{1}{8}$.

4.2 Performance Analysis

In this section, network performance is analyzed using various QoS parameters: delivery ratio, goodput, coverage, energy consumption and jitter. This analysis is performed using 150-nodes scenarios on ns-3 simulator. In order to construct MANET, a smart node is formed by integrating RFID reader with mobile sensor node. These mobile smart nodes constitute a hierarchical Ad-hoc network as shown in Figure 2. Reader collects the data from its local network and transmits to other nodes through radio frequency antenna of sensor nodes. Performance analysis of QoS parameters is as follows.

Delivery Ratio. It is the ratio of number of sent packets to number of delivered packets toward sink. Figure 4 shows the delivery ratios of 150 nodes over five MANETs routing protocols: Ad-hoc On Demand Distance Vector (AODV), Destination Sequenced Distance Vector (DSDV), Dynamic Source Routing (DSR), Temporarily Ordered Routing Algorithm (TORA) and Zone Routing Protocol (ZRP). From both scenarios, it is observed that ZRP protocol outperforms the other routing protocols. In 150 nodes scenarios, delivery ratio decreases with increase in time for every protocol because the number of available nodes for data transmission decreases and more number of nodes are occupied for routing.

Goodput. Another non-overlapping term with delivery ratio is goodput. It is the total number of successfully delivered packets to sink [54]. With addition of more number of packets and delay parameters, value of goodput can be increased. Figure 5 and Figure 6

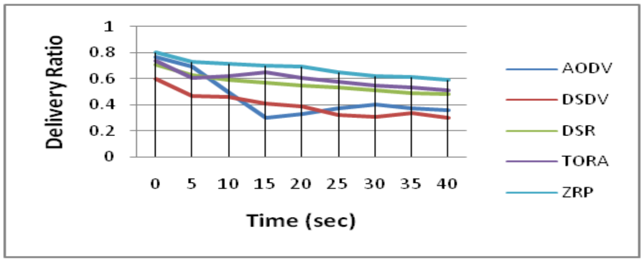


Figure 4: Delivery ratio for 150 nodes over MANETs routing protocols

Table 4: Lightweight node-packet delivery analysis using alloy (time in msec)

Percentage of routed or delivered packets	Intruder Assertions	Proposed Trusted Strategy	
		Time (Steps)	Result
More than 75	1/5/10	10/21/32 (80/113/131)	Proved
More than 65	1/5/10	60/94/145 (150/173/224)	Proved
More than 55	1/5/10	113/146/211 (170/210/563)	Proved

show the goodput for 150 nodes at offer load of 1 packet/second and 5 packets/second respectively. In 150-nodes scenarios, ZRP protocol outperforms than any other protocol. Performance of ZRP protocol is average and it is increasing exponentially with time at lesser rate, i.e. 1 pkt/sec.. In 5pkt/sec. for 150 nodes, ZRP is having improved performance as compared to 1 pkt/sec. In these scenarios, other protocols also show increase in performance but this increase is lesser as compared to ZRP protocol. It is also observed that in 150 nodes scenarios, growth of throughput for ZRP is linear than linear but for other protocol, it is linear or less.

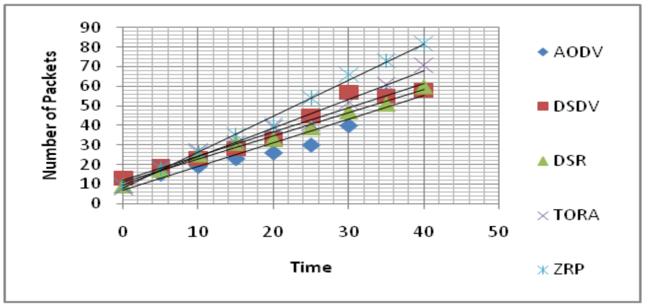


Figure 5: Goodput for 150 nodes at 1 packet/second

Coverage. It is defined as number of nodes used per unit time for successful transmission of packets. In Table 4, three scenarios are taken into consideration to find the coverage range for proposed scheme. Results shows that if a node deliver more than 75% of packets then intrusion detection take 50 msec and 70 steps which is lesser than delivery percentage of 65. It takes a minimum difference of 100 msec. and 90 steps when compared with 55% of delivery. Hence, a node is considered to be covered if it successfully delivers 75% of packets it receive and loss 25% only for performance analysis. Figure 7 and Figure 8 show the coverage of 150 nodes at 1pkt/sec and 5 pkts/sec respectively. In 1pkt/sec and 5 pkts/sec. scenarios, DSR and TORA are having worst performance variance. In both such scenarios, ZRP outperforms the other protocols because of its hybrid routing nature. This protocol, internally divides the nodes into zone and these zones with energy saving Frisbee formation save nodes energy for communication. Most of the nodes are silent during simulation initialization and this property is common among all scenarios. High coverage is observed during peak hours which varies from protocol to protocol.

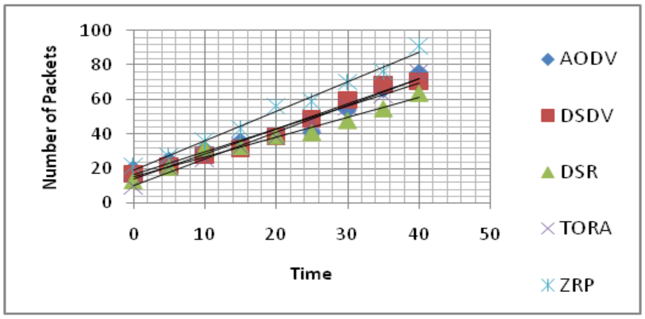


Figure 6: Goodput for 150 nodes at 5 packets/second

Energy Consumption. The evaluation of energy consumption in simulation environment is observed through throughput. Whenever radio of any node is on and a byte is transferred then energy of node is considered to be consumed. As discussed in section 4, this energy is calculated from RSSI and it is a function of distance. More is the distance parameter more will be the energy consumption. Figure 9 shows the average energy consumption for 150 nodes scenario. If bursts of any protocol are closer to the outer ring then average energy consumption for that protocol is higher and it is called as high burst (0.04-0.05 Joules). Low bursts are the minimum consumption values that are close to origin (0.01 Joules). Whereas, medium bursts are the intermediate values between high and low bursts (0.02-0.03 Joules). As shown in Figure 9, ZRP and TORA protocol are having higher average energy consumption than AODV, DSDV and DSR for o.1 pkt/sec, 1 pkt/sec. and 5 pkts/sec. In DSR and DSDV protocol, energy consumption shows

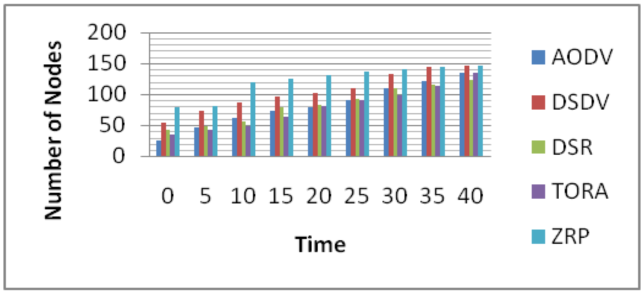


Figure 7: Coverage for 150 nodes at 1 packets/second

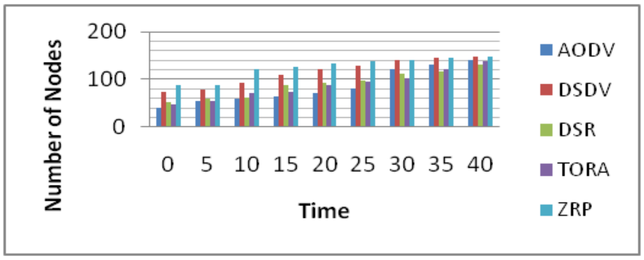


Figure 8: Coverage for 150 nodes at 5 packets/second

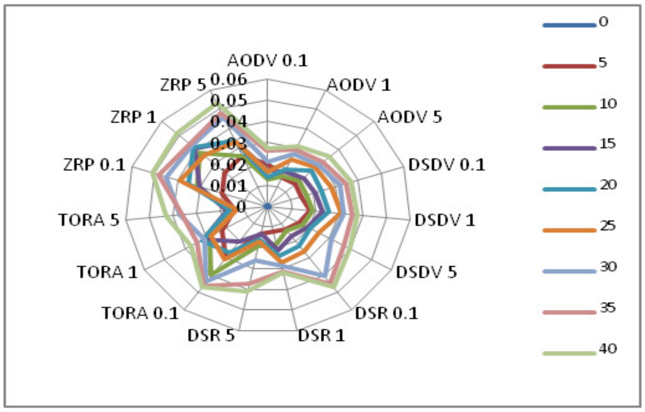


Figure 9: Energy consumption for 150 nodes during simulation time

variations with increases in packet/second. because of dynamic nature of routing protocol. Whenever there is need to transmit packets, then only nodes are activated and energy consumption starts.

Jitter. It is an average value of root mean square delay. Figure 10 shows the jitter values at different packet delivery rates, i.e. 1 pkt/sec. and 5 pkts/sec. Jitter values of TORA and AODV are worst as compared to other protocols. Since ZRP provides higher throughput but at minimum jitter thus it is considered to be the best protocol. Jitter value decreases with increase in number of nodes because more nodes are available to route the packets thus delay decreases. But this delay does not affect much on the performance because the packet delivery rate also increases. Performance improvement because of increased number of nodes is compensated by increase in packet delivery ratio. Also, with increase in packet delivery ratio the jitter decreases because once routes are established then it does not affect much on the performance.

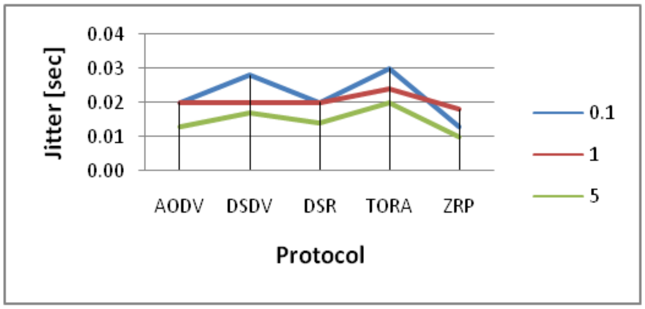


Figure 10: Jitter for 150 nodes at different delivery rate

4.3 Lightweight Analysis

4.3.1 Lightweight Primitive Analysis

Confidentiality as well as authentication mechanisms are integrated with protocol 1 and protocol 3 whereas only authentication mechanism is integrated with protocol 2. Table 5 shows the comparative analysis of substitution permutation network (SPN) based lightweight primitives for Protocols 1, 2 and 3. Two lightweight primitives are taken for analysis: LED and PHOTON. Result of lightweight primitives are compared with classical mechanism, i.e. Advanced Encryption Standard (AES). All three are based on confusion and diffusion layer principle in SPNs. LED and AES are used to achieve confidentiality and PHOTON is used for authentication. Alloy analysis shows that the number of variable generated, clauses formed and computational time in Protocol 1 and Protocol 3 for LED and PHOTON are much lesser than AES. Both confusion and diffusion layers are showing similar results. Multiple challenges and verifications in Protocol 2 increases the resource consumption and time required to complete the operations. Comparison of lightweight primitives with classical primitive shows that integration of LED and PHOTON in proposed mechanism enhances the performance of protocols as compared to AES based classical confidentiality mechanism.

4.3.2 Lightweight Policy Analysis

Figure 11 shows the proposed trust policy for subgroup member in proposed scheme. Trust based proposed mechanism is having: subgroup controller, subgroup member, virtual subgroup member and virtual subgroup controller. Each entity in hierarchical model acts as either producer or consumer. While acting as producer or consumer, there will be change of permissions. A subgroup controller will be having READ, WRITE, ACCESS, USE, MODIFY permissions for trust management. Whereas, a subgroup

Table 5: Simple vs. lightweight primitive analysis for proposed scheme

Protocol	Primitives	Layer	Variables	Clauses	Time(msec)
Protocol1	LED	Confusion	22025	15174	1463
		Diffusion	20451	13012	1231
	PHOTON	Confusion	42314	44101	2112
		Diffusion	36110	23603	1642
	AES	Confusion	80178	25545	3463
		Diffusion	60145	160234	2654
Protocol2	PHOTON	Confusion	44114	46045	2414
		Diffusion	37111	26032	2001
Protocol3	LED	Confusion	22544	160112	1513
		Diffusion	20653	13009	1213
	PHOTON	Confusion	41015	44023	2104
		Diffusion	36009	23112	1672
	AES	Confusion	81534	26123	3413
		Diffusion	62435	16144	2611

```
(MemberAssigned = (Interested s a r):- (Assigned s r) (AssignID a) (SubGroup r))
(MemberConflict = (Interested s a r):- (Conflicted s r) (RetrieveID a) (SubGroup r))
(MemberTrust = (TrustGeneration s a r):- (Assigned s r) (AssignID a) (SubGroup r))
(MemberTrust = (TrustPropagation s a r):- (Assigned s r) (AssignID a) (SubGroup r))
(MemberTrustConflict = (TrustAccumulation s a r):- (Conflicted s r) (SubGroup r))
(MemberTrust = (TrustPrediction s a r):- (Assigned s r) (AssignID a) (SubGroup r))
(MemberTrustConflict = (TrustEvaluate s a r):- (Conflicted s r) (SubGroup r))
(MemberTrustConflict = (TrustApplication s a r):- (Conflicted s r) (SubGroup r))
```

Figure 11: Margrave policy for access control in proposed scheme

member will be having READ, ACCESS, USE permissions only. So, each member will have its own policy in the network. Figure 11 shows the subgroup member policy for TrustGeneration, TrustPropagation, TrustAccumulation, TrustPrediction, TrustEvaluation and TrustApplication. A subgroup member can act as producer to assign new identification to new node or retrieve its identification. Trust generation, propagation and prediction are permissible for subgroup member. Trust accumulation and application comparison are not allowed for member but these are considered to be the functions of subgroup controller. After designing and analyzing the policies of every member in proposed scheme, it is analyzed through Margrave that there is no conflict in any policy [1].

5 Conclusions

The current study examines RFID-Sensor based MANETs using ECCr in code based cryptography. MANETs are constructed by extending the trust management approach in resource constraint environment with Teo and Tan protocol for key exchange using hierarchical model [66] and Avoine MA-KA2 protocol for distance bounding and mutual authentication [7]. These approaches are perceived as efficient lightweight approaches with strong protection against distance bounding attacks. QoS parameters taken for network performance analysis are: delivery ratio, goodput, coverage, energy consumption and jitter. In conclusion, 150 nodes scenario shows that ZRP protocol outperforms any other protocol for proposed security system using trust management. Maximum goodput that is achievable through best routing protocol is approximately 80 packets per second to minimum delay of 0.03 msec. Probability attack analysis is performed for mafia fraud attack, distance fraud attack, terrorist fraud attack and distance hijacking attack in distance bounding protocol. In this analysis, fault acceptance rate of system is checked and in result it is found that system is strong enough against all these attacks. Lightweight primitives and policies for subgroup members are also analyzed. It is found that integration of lightweight primitives reduce computation and time complexity. Lightweight policy analysis shows that there is no conflict in access domains of any subgroup member.

References

[1] *The Margrave Policy Analyzer*, Jan. 19, 2015. (http://www.margrave-tool.org)

[2] M. R. S. Abyaneh, *Security Analysis of Lightweight Schemes for RFID Systems*, PhD thesis, University of Bergen, Norway, 2012.

[3] B. Alomair, A. Clark, J. Cuellar, and R. Poovendran, "Scalable RFID systems: A privacy-preversing proto with constant-time identification," *IEEE Transactions on Parallel Distribution Systems*, vol. 23, no. 8, pp. 1536–1550, 2012.

[4] R. J. Anderson, *Security Engineering: A guide to Building Dependable Distributed Systems*, New York, USA, John Wiley & Sons, 2001.

[5] S. A. Anson and M. Ilyas, *RFID handbook: Application, technology, security and privacy*, Boca Raton, Florida, USA, CRC, 2008.

[6] P. D'Arco and A. De Santis, "On ultralightweight RFID authentication protocols," *IEEE Transactions on Dependable and Secure Computing*, vol. 8, no. 4, pp. 548–563, 2011.

[7] G. Avoine and C. H. Kimh, "Improving program analyses by structure untupling," *IEEE Transactions on Mobile Computing*, vol. 12, no. 5, pp. 830–839, 2013.

[8] G. Avoine, C. Lauradoux, and B. Martin, "How secret-sharing can defeat terrorist fraud," in *Proceedings of the 4th ACM Conference on Wireless Network Security*, pp. 145–155, Hamburg, Germany, June 15-17, 2011.

[9] N. Bagheri and M. Safkhani, "Secret disclosure attack on kazahaya, a yoking-proof for low-cost RFID tags," Technical Report Cryptology ePrint Archive: Report 2013/453, July 2013.

[10] J. D. Bakos, D. M. Chiarulli, and S. P. Levitan, "Lightweight error correction coding for system-level interconnects," *IEEE Transactions on Computing*, vol. 56, no. 3, pp. 289–304, 2007.

[11] D. J. Bernstein, J. Buchmann, and E. Dahmen, *Post-Quantum Cryptography*, Springer-Verlag Berlin Heidelberg, New York, USA, Springer, 2009.

[12] M. Burmester, T. V. Le, and B. D. Medeirosn, "Universally composable RFID identification and authentication protocols," *ACM Transaction on Information and Systems Security*, vol. 12, no. 4, pp. 21:1–21:33, 2012.

[13] M. Burmester and J. Munilla, "Lightweight RFID authentication with forward and backward security," *ACM Transactions on Information and System Security*, vol. 14, no. 1, pp. 11:1–11:26, 2011.

[14] A. Canteaut and F. Chabaud, "Improvement of the attacks on cryptosystems based on error-correcting codes," Research Report: LIENS-95-21, École Normale Supérieure, Paris, July 1995.

[15] T. Cao, E. Bertino, and H. Lei, "Security analysis of the sasi protocol," *IEEE Transactions on Dependable and Secure Computing*, vol. 6, no. 1, pp. 73–77, 2009.

[16] A. Cerpa, J. Elson, D. Estrin, L. Girod, M. Hamilton, and J. Zhao, "Habitat monitoring application driver for wireless communication technology," in *Proceedings of the ACM SIGCOMM Workshop on Data Communication in Latin America and the Caribean*, pp. 20–41, San Jose, Costa Rica, Apr. 2001.

[17] A. Chakrabarti, A. Sabharwal, and B. Aazhang, "Using predictable observer mobility for power efficient design of sensor networks," in *Proceedings of the 2nd International Workshop on Information Processing in Sensor Networks (IPSN-03)*, pp. 129–145, Palo Alto, CA, USA, Apr. 2003.

[18] R. Chen, X. Chao, L. Tang, J. Hu, and Z. Chen, "A global reputation-based trust model in peer-to-peer networks," in *4th International Conference Automatic and Trusted Computing (ATC 2007)*, pp. 203–215, Hong Kong, China, 2007.

[19] J. Cho, Y. Shim, T. Kwon, and Y. Choi, "Sarif: A novel framework for integrating wireless sensor and RFID networks," *IEEE Wireless Communications*, vol. 14, no. 6, pp. 50–56, Dec. 2007.

[20] M. Conrad, T. French, and W. Huang, "A lightweight model of trust propagation in a multiclient network environment. to what extent does experience matter?," in *International Conference on Avaiability, Reliability and Security (ARES'06)*, pp. 482–487, Vienna University of Technology, Austria, Apr. 20-22, 2006.

[21] C. Cremers, K. B. Rasmussen, B. Schmidt, and S. Capkun, "Distance hijacking attacks on distance bounding protocols," in *IEEE Symposium on Security and Privacy (SP'12)*, pp. 113 – 127, San Francisco, CA, USA, 20-23 May 2012.

[22] D. Denning, "A new paradigm for trusted systems," in *Proceedings on the 1992-1993 Workshop on New Security Paradigms*, pp. 36–41, New York, NY, USA, 1993.

[23] M. Deutch, "Cooperation and trust: Some theoretical notes," in *Nebraska Symposium on Motivation*, pp. 275–319, University of Nebraska Press, Lincoln NE, USA, 1962.

[24] C. Englund and H. Wallin, "RFID in wireless sensor network," Master Thesis, Communication Systems Group, Department of Signals and Systems, Chalmers University of Technology, Goteborg, Sweden.

[25] A. Ephremides, "Energy concerns in wireless networks," *IEEE Transactions on Wireless Communication*, vol. 9, no. 4, pp. 48–59, 2002.

[26] R. B. Ferguson, "Gentag patent adds RFID sensor network feature to mobile devices," Dec. 2006. (`http://www.eweek.com/c/a/Mobile-and-Wireless/Gentag-Patent-Adds-RFID-Sensor-Network\-Feature-to-Mobile-Devices\`)

[27] M. D. Francesco, S. K. Das, and G. Anastasi, "Data collection in wireless sensor networks with mobile elements: A survey," *ACM Transaction on Sensor Networks*, vol. 8, no. 1, pp. 7:1–7:31, 2011.

[28] D. Gambetta, "Can we trust?," in *Trust: Making and Breaking Cooperative Relations*, vol. 13, pp. 213–237, Department of Sociology, University of Oxford, England, 2000.

[29] K. Govindan and P. Mohapatra, "Trust computations and trust dynamics in mobile adhoc networks: A survey," *IEEE Communications Surveys and Tutorials*, vol. 14, no. 2, pp. 279–298, 2012.

[30] D. Jackson, "Alloy: a lightweight object modelling notation," *ACM Transactions on Software Engineering and Methodology*, vol. 11, no. 2, pp. 256–290, 2002.

[31] D. Jackson, "Micromodels of software: Lightweight modelling and analysis with alloy," Technical Report MIT Lab Manual, Feb. 2002.

[32] S. Jarvenpaa, K. Knoll, and E. L. Dorothy, "Is anybody out there?: antecedents of trust in global virtual teams," *Journal of Management*, vol. 14, no. 4, pp. 29–64, 1998.

[33] A. Josang, "The right type of trust for distributed systems," in *Proceedings of the ACM New Security Paradigm Workshop*, pp. 119–131, Lake Arrowhead, CA, USA, 1996.

[34] A. Juel and S. Weis, "Authenticating pervasive devices with human protocols," in *Advances in cryptology (Crypto'05)*, pp. 293–298, Santa Barbara, California, USA, 2005.

[35] A. Juels, "RFID security and privacy: A research survey," *IEEE Journal on Selected Areas in Communication*, vol. 24, no. 2, pp. 381–394, 2005.

[36] S. Kamvar, M. Schlosser, and H. Garcia-Molina, "The eigentrust algorithm for reputation management in p2p networks," in *Proceedings of the International World Wide Web Conference (WWW'03)*, pp. 640–651, Budapest, Hungary, 2003.

[37] O. Khalid, U. S. Khan, S. A. Madani, et al., "Comparative study of trust and reputation systems for wireless sensor networks," *International Journal on Security and Communication networks*, vol. 6, no. 6, pp. 669–688, 2013.

[38] C. H. Kim, "Security analysis of ykhl distance bounding protocol with adjustable false acceptance ratio," *IEEE Communications Letters*, vol. 15, no. 10, pp. 1078–1080, 2011.

[39] C. H. Kim and G. Avoine, "RFID distance bounding protocols with mixed challenges," *IEEE Transactions on Wireless Communications*, vol. 10, no. 5, pp. 1618–1626, 2011.

[40] P. Kitsos and Y. Zhang, *RFID security, techniques, protocols and system-on-chip design*, New York, USA, Springer, 2008.

[41] R. Koh, E. Schuster, I. Chackrabarti, and A. Bellman, "Securing the pharmaceutical supply chain," White Paper, 2003.

[42] A. Kumar and A. Aggarwal, "Efficient hierarchical threshold symmetric group key management protocol for mobile ad hoc networks," in *International Conference on Contemporary Computing (IC3'12)*, pp. 335–346, Noida, India, 2012.

[43] A. Kumar, K. Gopal, and A. Aggarwal, "Outlier detection and treatment for lightweight mobile ad hoc networks," in *Qshine'13*, pp. 750–763, Greater Noida, India, 2013.

[44] M. Lehtonem, T. Staake, F. Michahelles, and E. Fleisch, "From identification to authentication- a review of RFID product authentication techniues," in *Networked RFID Systems and Lightweight Cryptography*, pp. 169–187, USA, 2007.

[45] A. Mason, A. Shaw, A. I. Al-Shamma'a, and T. Welsby, "RFID and wireless sensor integration for intelligent tracking systems," in *Proceedings of 2nd GERI Annual Research Symposium (GARS'06)*, Liverpool, U.K., 2006.

[46] R. C. Mayer, J. H. Davis, and F. D. Schoorman, "An integrative model of organizational trust," *Academy of Management Executive*, vol. 20, no. 3, pp. 709–734, 1995.

[47] A. McCumber, *Assessing and Managing Security Risk in IT Systems: A Structured Methodology*, Boca Raton, Florida, USA, Auerbach Publications, 2005.

[48] D. H. McKnight and N. L. Chervany, "Trust and distrust definitions: One bite at a time," in *Deception, Fraud, and Trust in Agent Societies*, pp. 27–54, Barcelona, Spain, 2000.

[49] G. E. Moore, "Cramming more components onto integrated circuits," *Electronics Magazine*, vol. 38, no. 8, pp. 114–117, 1965.

[50] Z. Nochta, T. Staake, and E. Fleisch, "Product specific security features based on RFID technology," in *International Symposium on Applications and the Internet Workshops (SAINTW'06)*, pp. 72–75, Phoenix, AZ, USA, 2006.

[51] D. M. R. Overbeck, *Public Key Cryptography based on Coding Theory*, Ph.D. Thesis, Technische Universitat Darmstadt, 64277 Darmstadt, 2007.

[52] J. Pearson, "Securing the pharmaceutical supply chain with RFID and public key infrastructure (PKI) technologies," White Paper, June 2005.

[53] P. Peris-Lopez, J. C. Hernandez-Castro, J. M. Esteveze-Tapiador, and A. Ribagorda, "RFID systems: A survey on security threats and proposed solutions," in *International Conference on Personal Wireless Communication (PWCA'06)*, pp. 159–170, Albacete, Spain, 2006.

[54] D. Puccinelli and M. Haenggi, "Reliable data delivery in large scale low-power sensor networks," *IEEE Transactions on Mobile Computing*, vol. 6, no. 4, pp. 28:1–28:41, 2010.

[55] V. Rajendran, J. J. Garcia-Luna-Aceves, and K. Obraczka, "Energy-efficient, application-aware medium access for wireless sensor networks," in *Proceedings of the 2005 International Conference on Mobile Ad Hoc and Sensor Systems Conference (MASS'05)*, pp. 623–630, Washington, DC, USA, 2005.

[56] V. Rajendran, K. Obraczka, and J. J. Garcia-Luna-Aceves, "Energy-efficient, collision free medium ac-

cess control for wireless sensor networks," *Wireless Networks*, vol. 12, no. 1, pp. 63–78, 2006.

[57] C. Schurgers, V. Tsiatsis, S. Ganeriwal, and M. B. Srivastava, "Optimizing sensor networks in the energy-latency-density design space," *IEEE Transactions on Mobile Computing*, vol. 1, no. 1, pp. 70–80, 2002.

[58] P. W. Shor, "Algorithm for quantum computation: Discrete logarithms and factoring," in *35th Annual Symposium on Foundations of Computer Science*, pp. 124–134, Santa Fe, New Mexico, USA, 1994.

[59] K. Sohrabi, J. Gao, V. Ailawadhi, and G. J. Pottie, "Protocols for self-organization of a wireless sensor networks," *ACM Computer Communication Review*, vol. 7, no. 5, pp. 16–27, 2000.

[60] A. W. Stephen, S. E. Sarma, R. L. Rivest, and D. W. Engels, "Security and privacy aspects of low-cost radio frequency identification systems," in *First International Conference on Security in Pervasive Computing*, pp. 201–212, Boppard, Germany, 2003.

[61] H. M. Sun, W. C. Ting, and K. H. Wang, "On the security of chien's ultralightweight RFID authentication protocol," *IEEE Transactions on Dependable and Secure Computing*, vol. 8, no. 2, pp. 315–317, 2011.

[62] Y. Sun, S. Du, O. Gurewitz, and D. B. Johnson, "Dw-mac: A low latency energy efficient demand wakeup mac protocol for wireless sensor networks," in *proceedings of the 9th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc'08)*, pp. 53–62, New York, USA, 2008.

[63] Y. Sun, W. Yu, Z. Han, and K. J. Ray Liu, "Information theoretic framework of trust modeling and evaluation for ad hoc networks," *IEEE Journal of Select Area on Communications*, vol. 24, no. 2, pp. 305–317, 2006.

[64] P. Sztompka, *Trust: A sociological theory*, Cambridge, United Kingdom: Cambridge University Press, 1999.

[65] K. Takaragi, M. Usami, R. Imura, R. Itsuki, and T. Satoh, "An ultra small individual recognition security chip," *IEEE Micro*, vol. 21, no. 6, pp. 43–49, 2001.

[66] J. C. M. Teo and C. H. Tan, "Energy-efficient and scalable group key agreement for large ad hoc networks," in *ACM International Workshop on Performance Evaluation of Wireless Ad Hoc, Sensor, and Ubiquitous Networks (PE-WASUN'05)*, pp. 114–121, Montreal, Qc. Canada, 2005.

[67] M. Najam ul islam U. Mujahid and J. Ahmed, "Ultralightweight cryptography for passive RFID systems," Technical Report Cryptology ePrint Archive: Report 2013/847, Dec. 2013.

[68] T. Wu and S. Biswas, "A self-reorganizing slot allocation protocol for multi-cluster sensor networks," in *proceedings of the 4th International Symposium on Information Processing in Sensor Networks*, pp. 309–316, Los Angeles, California, USA, 2005.

[69] R. Yahalom, B. Klein, and Th. Beth, "Trust relationships in secure systems- a distributed authentication perspective," in *Proceedings 1993 IEEE Symposium on Research in Security and Privacy*, pp. 150–164, Oakland, CA, USA, 1993.

[70] X. Yang and N. Vaidya, "A wakeup scheme for sensor networks: Achieving balance between energy saving and end-to-end delay," in *Proceedings of the 10th IEEE Real-Time and Embedded Technology and Applications Symposium (RTAS'04)*, pp. 19–26, King Edward, Toronto, Canada, 2004.

[71] S. Yessad, F. Nait-Abdesselam, T. Taleb, and B. Bensaou, "R-mac: Reservation medimum access control protocol for wireless sensor networks," in *Proceedings of the 32nd IEEE conference on Local computer networks*, pp. 719–724, Dublin, Ireland, 2007.

[72] L. Zhang and Z. Wang, "Integration of RFID into wireless sensor networks: Architectures, opportunities and challenging problems," in *Proceedings of the 5th International Conference on Grid and Cooperative Computing Workshops (GCCW'06)*, pp. 463–469, Changsha, China, 2006.

[73] P. Zhang, C. M. Sadler, S. A. Lyon, and M. Martonosi, "Hardware design experiences in zebranet," in *Proceedings of the 2nd ACM Conference on Embedded Networked Sensor Systems (SenSys'04)*, pp. 227–238, Baltimore, Maryland, 2004.

[74] Y. Zhang, L. T. Yang, and J. Chen, *RFID and Sensor Networks: Architectures, Protocols, Security and Integrations*, Boca Raton, London, New York: CRC, 2009.

[75] Y. Z. Zhao, C. Miao, M. Ma, J. B. Zhang, and C. Leungi, "A survey and projection on medium access control protocols for wireless sensor networks," *ACM Computing Surveys*, vol. 45, no. 1, pp. 7:1–7:37, 2012.

[76] L. Zhou and Z. J. Haas, "Securing ad hoc networks," *IEEE Network*, vol. 13, no. 6, pp. 24–30, 1999.

Adarsh Kumar received his ME degree in software engineering from Thapar University, Patiala, Punjab, India, in 2003. Since 2003, he has been with the Department of Computer Science Engineering and Information Technology, Jaypee Institute of Information Technology, Noida, Uttar Pardesh, India, where he is now an assistant professor. His main research interests are cryptography, network security, and adhoc networks.

Krishna Gopal biography. received his BTECH degree in electrical engineering from the Department of Electrical Engineering, IIT, Madras, India, in 1966 and his MS and PhD degrees in engineering from the REC Kurukshetra, Kurukshetra, India, in 1972 and 1979, respectively. Since 2011, he has been working asa dean (Academic & Research) with Jaypee Institute of Information Technology, Noida, India. He has forty-five years of teaching and research experience. He is a member of various professional bodies, such as the Life Member System Society of India, the Indian Society for

Technical Education, and the IEEE. .

Alok Aggarwal received his BTECH and MTECH degrees in computer science engineering from the Department of Computer science, Kurukshetra University, India, in 1995 and 2001, respectively and his PhD degree in engineering from IIT, Roorkee, India, in 2010. From 2009 to 2012, he worked for the Jaypee Institute of Information Technology, Noida, India. Since 2012, he has been with the JP Institute of Engineering and Technology, Meerut, India, where he is now a professor and director. His main research interests are wired/wireless networks, security, and coding theory.

Asynchronous Invariant Digital Image Watermarking in Radon Field for Resistant Encrypted Watermark

Dhekra Essaidani, Hassene Seddik, and Ezzedine Ben Braiek
(Corresponding author: Dhekra Essaidani)

Product Research Center, CEREP Research Laboratory, ESSTT
5 Av. Taha Hussein, 1008, Tunis
(Email: dhekraessaidani89@gmail.com)

(Received Aug. 19, 2013; revised and accepted Jan 10 & Mar. 15, 2014)

Abstract

With the rapid evolution of processing multimedia technologies (text, audio, image and video) and its internet application (wide and easy transmission of digital multimedia contents), the copyright protection has been receiving an increasing attention. Among the existing strategies to protect multimedia online, digital watermarking provides a promising way of protecting data online from illegal manipulation and duplication. In this paper, a new image of watermarking scheme is presented. It performs imperceptible watermarking of color image in the radon domain. The proposed algorithm can resist to geometrical attacks. Experimental results show that the proposed watermarking approach not only can meet the demand on invisibility and robustness of the watermark, but also presents a good performance compared to other proposed methods considered in the comparative study. A mathematical study is developed to demonstrate how and why this approach is robust against geometric transforms.

Keywords: Asynchronous attacks, circular integration transform (CIT), color image watermarking, discrete radon transform, radial integration transform (RIT), robustness

1 Introduction

The rapid development of processing data technologies and internet applications has improved the ease of access to information online. It also increases the problem of illegal copying and redistribution of digital media. Encryption and Stenography are the two techniques introduced to solve data on line. In 1992, the research suggested to use the watermarking technique in data protection.

Nowadays, image watermarking is a protection technology that has attracted a lot of attention. The basic idea of watermarking involves integrating a message into a digital content. This last covers the information to be transmitted in a holder in a way to be invisible and correctly reversible (an algorithm allows the exact extraction of the embedded watermark). Its algorithm requires equilibrium between three constraints: imperceptibility, robustness and embedding capacity [2].

Image watermarking schemes have to keep the image quality and to be robust against general image processing and geometric transformation (scaling, rotation and translation) [15]. There are many watermarking algorithms which have been presented in recent literature to protect data against geometric attacks. They can be divided in three main categories. The first one includes watermarking approach which is a watermark detection performing in an invariant domain to geometric attacks. The second category includes methods that detect and correct the geometric attack of the watermarked image in order to perform the detection process. However, another approach for resisting geometric attacks is based on synchronizing, in terms of position, orientation and scaling, use image features to embed and extract the correlating watermark [13].

Various watermarking schemes are proposed for the digital multimedia protection. Most of the schemes perform on the spatial domain where the watermarking techniques directly modify the intensities of selected pixels [4, 5, 6, 8, 9]. Also, several schemes perform on the transformation domain (Fourier-Mellin Transform, Discrete Cosine Transform (DCT), Discrete Fourier Transform (DFT), Discrete Wavelet Transform and the Complex Wavelet Transform (CWT)) where the watermarking algorithm modifies the selected transformed coefficients [10, 16, 18]. In [18], authors used the properties of Fourier transform to develop a watermarking scheme resisting the unavoidable noise and cropping. This algorithm presents a robust watermark strategy for quantum images. The watermark is embedded into the Fourier coefficients of the quantum carrier image. Authors in [15] proposed a state-coding based on blind watermarking algorithm to embed color image watermark to color host image. This approach used the Integer Wavelet Transform (IWT) and the rules of state coding of the components, R, G and B, of color image watermark and the components, Y, Cr and Cb, of color host image. In the extraction process, authors used also the rules of state coding to recover the original watermark or original host image. In [11], authors proposed an invariant image watermarking scheme by introducing the Polar Harmonic Transform (PHT). This algorithm proposed to resist geometric transformation. Furthermore, Xiao, Ma and Cui have been used for invariance watermarking scheme against global geometric that transforms the Radon field and pseudo-Fourier-Mellin transforms. This combination is named Radon and pseudo-Fourier-Mellin invariants (RPFMI) [17].

In order to resist geometric attacks, we propose a new watermarking algorithm for RGB color image. The proposed approach belongs to the second of the categories that were described above. Imperceptible watermark embedding and detection are performed in the non-conventional radon domain. Our approach selects specific coefficients based on their energy in Radon field to embed watermark. The simulation results proved by mathematical study proved the high efficiency and robustness of the proposed approach. This paper is organized as follows: Section 2 presents an overview of Radon Transformation (RT). Section 3 details our watermarking method. In Section 4, we study the robustness of this technique against different STIRMARK attacks, and we test the ability to detect the embedded watermark in the host image. A study of the watermarked image distortions before and after different attacks is also presented. In Section 5, a mathematical study is developed to explain the resistance of the proposed method and prove the results found. A comparative study with recent published techniques is also presented.

2 Mathematical Recall of Radon Transform

2.1 Generalized Radon Transform

In 1917, Radon, Austrian mathematician, defined the theory of Radon Transform. He proved the possibility to reconstruct a function of a space from knowledge of its integration along the hyper-plans in the same space. This theory establishes the reversibility of the Radon transform and the transition between the native function space and the Radon space, or the space of projections [3]. In image processing, the Radon Transform represents a collection of projections along various directions [14]. The generalized Radon transformation of a 2D continuous function is defined in [1] by the following equation:

$$R(\rho,\theta) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\delta(x\cos\theta + y\sin\theta - \rho)dxdy,$$

where ρ represents the perpendicular distance of a straight line from the origin, and represents the angle between the distance vector and the x-axis.

The literature proposed two categories of one-dimensional Radon transformation; the first is based on Radial Integration Transform (RIT) and the second is based on the Circular Integration Transform (CIT).

2.1.1 One-Dimensional Radial Integration Transform

The RIT of a function $f(x,y)$ is defined as the integral of $f(x,y)$ along a straight line that begins from the origin $(x_0,y_0)$and has angle θ with respect to the horizontal axis (see Figure 1). It is given by the following equation [13]:

$$R_f(\theta) = \int_0^{+\infty} f(x_0 + u\cos\theta, y_0 + u\sin\theta)du,$$

where u is the distance from the origin $(x_0,y_0)$and $f(x,y)$ is presented by the integral along a straight line that begins from the origin (x_0,y_0) and has an angle with respect to the horizontal axis.

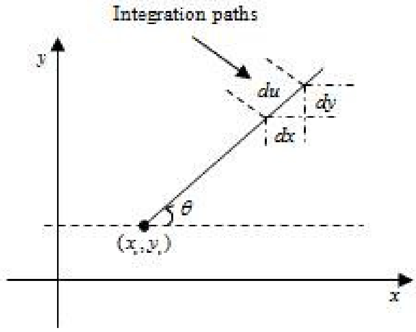


Figure 1: Representation of the radial integration transform

2.1.2 One-Dimensional Circular Integration Transform

The CIT of a function $f(x,y)$ is defined as the integral of $f(x,y)$ along a circle curve with center x_0, y_0 and radius ρ(see Figure 2). It is given at [13] by the following equation:

$$C_f(\rho) = \int_0^{2\pi} f(x_0 + \rho\cos\theta, y_0 + \rho\sin\theta)\rho d\theta,$$

where $d\theta$ is the corresponding elementary angle and $f(x,y)$ represents the circle integrated function around the center $(x_0,y_0)$and by the radius ρ.

2.2 Discrete Radon Transform (DRT)

The discreet Radon transformation 'DRT' of an image $I(x,y)$ can be defined by the following equations [13]:

$$R(t\Delta\theta) = \tfrac{1}{J}\textstyle\sum_{j=1}^{J} I(x_0 + j\Delta s\cos(t\Delta\theta), \\ y_0 + j\Delta s\sin(t\Delta\theta)).$$

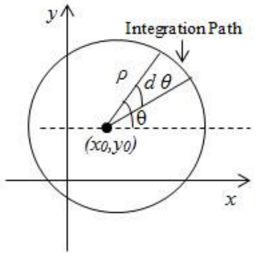


Figure 2: Representation of the circular integration transform

$$C(k\Delta\rho) = \frac{1}{T}\sum_{t=1}^{T} I(x_0 + k\Delta\rho\cos(t\Delta\theta), \\ y_0 + k\Delta\rho\sin(t\Delta\theta)),$$

where $d\theta$ represents the angular variation step, Δs is the scaling step, $k\Delta p$ represents the radius of the smallest circle that encircles the image, J represents the number of samples on the radius with orientation θ, $t = 1, ..., \frac{360}{\Delta\theta}$ and $k = 1, ..., \frac{360}{\Delta\theta}$.

The Radon transformation of the image $I(x, y)$ of size $[MN]$ generate a matrix $R(\rho, \theta)$ of a size equal to N_ρ, N_e with real coefficient representing the radon filed with:

$$\begin{cases} N_\rho &= \sqrt{N^2 + M^2} + 1 \\ M_\rho &= \frac{\theta_{\max}}{\Delta\theta}. \end{cases} \quad (1)$$

3 The Proposed Watermarking Approach

The literature suggested that the Radon Transform properties are much recommended in watermarking applications in which resistance to geometric attacks. A watermark embedding and detection scheme using these properties are described in the proposed watermarking approach. Also, due to the expansion of the projected image matrix from its size $[M, N]$ to $[M_\rho, N_e]$ this filed allows a higher amount of embedded data. In fact the DRT increases the size of the transformed image (see Equation (1)). The proposed method consists in embedding the watermark in selected coefficients in the radon field. These coefficients are chosen from the area of maximal energy. They represent maxims in the Radon coefficients and must respect the following three essential characteristics:

- These coefficients are set on the integral line of projection, so they will be well recovered from the inverse radon transform.
- Secondly, they contain the most important details of the original images. Consequently, they are the most adapted to a code in a watermark with better imperceptibility.
- Their high values enable us to hide the binary coefficients of the watermark without any perceptual degradation.

3.1 Details of the Proposed Algorithm

In the following Sections, we will note the original watermark as W_o, the encrypted embedded watermark as W, the recovered encrypted watermark as $W^{'}$,the recovered decrypted watermark as $W_o^{'}$, the Original RGB color image (support or host image) as I, the original blue matrix of the host image as I_b,the transformed channel blue (matrix blue) of the host image in Radon field as R_b, the watermarked channel blue (matrix blue) of the host image in Radon field as R_{bw}, the watermarked blue matrix in spatial field as I_{bw}, I_w represents the watermarked spatial image and the transformed channel blue (matrix blue) of the watermarked host image in Radon field as $R^{'}_{bw}$. Likewise, (x, y) represents the spatial coordinates of the original image, (ρ, θ) represents the coordinates of the color image in the radon field and k and l are the coded bits representing the watermark, $[M_w\, N_w]$ represents the size of the original image, $[M\, N]$represents the size of original watermark, $[N_\rho\, N_\theta]$ represents the size of the image in radon domain and G is the embedding strength.

3.1.1 Watermark Embedding Process

The main concept of the watermark embedding process is shown in Figure 3.

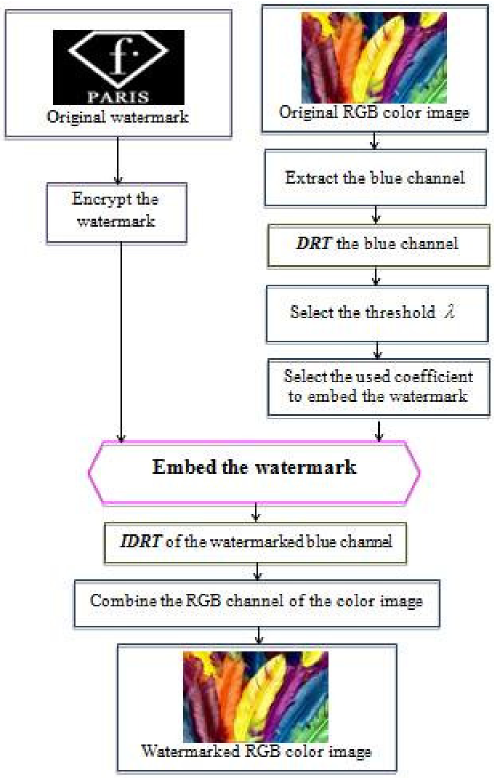

PARIS
Original watermark
Original RGB color image
Extract the blue channel
Encrypt the watermark
DRT the blue channel
Select the threshold λ
Select the used coefficient to embed the watermark
Embed the watermark
IDRT of the watermarked blue channel
Combine the RGB channel of the color image
Watermarked RGB color image

Figure 3: Watermarking algorithm

For the RGB color image, the red, green and blue channels are candidates for watermark embedding as human eyes are not sensitive on the modification of blue channel than the green and red channels. Besides, Watermarking in the blue channel allows good invisibility and higher embedding capacity. Therefore, we propose to embed the watermark in the blue channel of the selected color image [7]. The proposed algorithm is described in the following steps:

Step 1: Encrypt the watermark.

In order to encrypt the watermark, we use the following steps:

1) Transform the original watermark W_o into a one dimensional vector V_{water} by the following equation:

$$W_o(x, y) \rightarrow V_{water}(l).$$

2) Decompose the watermark in N equal blocks B_i, where

$$V_{water}(N) = \{B_1, B_2, ..., B_N.$$

3) Generate a key $key0$ witch its length is equal to the length of the block B_i.

4) Encrypt the first block B_1 by using the following equation:

$$Bc_1 = B_1 \oplus key0.$$

5) Encrypt the second block B_2 by using the following equation:

$$Bc_2 = B_2 \oplus Bc_1.$$

6) Generally, after each encrypting iterates of each block B_i , the resulting encrypted blocks Bc_i is used to encrypt the next block Bc_{i+1}, where $i = 3, 4, 5, ..., N$.

$$Bc_i = B_i \oplus Bc_{i-1}.$$

7) After encrypting the watermark by using the function "XOR", we applied S_{max} iteratively permutations on the encrypted watermark vector in order to improve the encryption system. The first permutated iteration is defined by the following equation:

$$P_{s=1}(B_1, ..., B_N) = \\ W_o : B_{N/2}, ..., B_1, B_N, ..., B_{((N/2)+1)}.$$

We continue the permutation process by applying the defined function as follows:

$$P_{s=\alpha}(Bc_1, ..., Bc_N) = \\ P_{s=\alpha-1} Bc_N, ..., Bc_{(N/2)}, Bc_1, ..., Bc_{((N/2)+1)},$$

where $\alpha = 2 \rightarrow S_{max}$ and S_{max} represents the number of the permutation iteration. The encrypted watermark vector V_c is obtained after S_{max} permutations and it is defined as follows:

$$V_c(l) = P_{s=20}(l).$$

To obtain the encrypted watermark, we transformed the vector V_c to matrix with size equal to $[M_w \, N_w]$ defined as follows:

$$V_c(l) \rightarrow W(x, y).$$

Step 2: Select the radon coefficients to embed the watermark.

In this step, a discrete radon transform is applied only on the blue channel I_b of the color image I. A selection of a set of coefficients having the higher energy from this transformed matrix called R_b is done. The number of the selected coefficient is equal to $M_w \times N_w$ which represents the length of the encrypted watermark. For this reason, we transform the image matrix into a vector V by using the following equation:

$$Rb(\rho, \theta) \rightarrow V(k).$$

Then, the vector V is organized in downward order. It is defined as fellows:

$$V(1) > V(2) > V(3) > ... > V(k-1) > V(k),$$

where $K = M_\theta * N_\theta$ Next, we select the used coefficients to embed the watermark. They represent the $M_w \times N_w$ first highest coefficients in the matrix R_b. For this process, we use the following steps:

1) Define the threshold λ_{opt} : It represents the coefficient number $M_w \times N_w$ in the vector V:

$$\lambda_{opt} = V(M_w * N_w).$$

2) Select the coefficients to be used for watermark coding:

$$R_E(\rho, \theta) = R_b(\rho, \theta) \qquad where R_b(\rho, \theta) \geq \lambda_{opt}.$$

So, the selected coefficient to encode the watermark represent the $M_w \times N_w$ first coefficients in the vector V:

$$R_E(\rho, \theta) = V(l).$$

Step 3: Embedding process.

The embedding process is described in Figure 4.

Each selected coefficient coded one bits of the encrypted watermark vector V_w. To embed watermark in the selected coefficient of the matrix R_b, we used the following equation:

$$\begin{cases} R_{bw}(\rho, \theta) &= R_b(\rho, \theta) + G & if \;\; V_c(l) = 1 \\ R_{bw}(\rho, \theta) &= R_b(\rho, \theta) - G & if \;\; V_c(l) = 0. \end{cases}$$

Also, we use the vectors Er, E_ρ and E_θ to save the amplitude, the position ρ and the position θ of each selected coefficient $R_b(\rho, \theta)$. These vectors have the length $M_w \times N_w$. They are used later to recover the embedding watermark. These vectors are filled by using the following equations:

$$Er(c) = R_b(\rho, \theta).$$

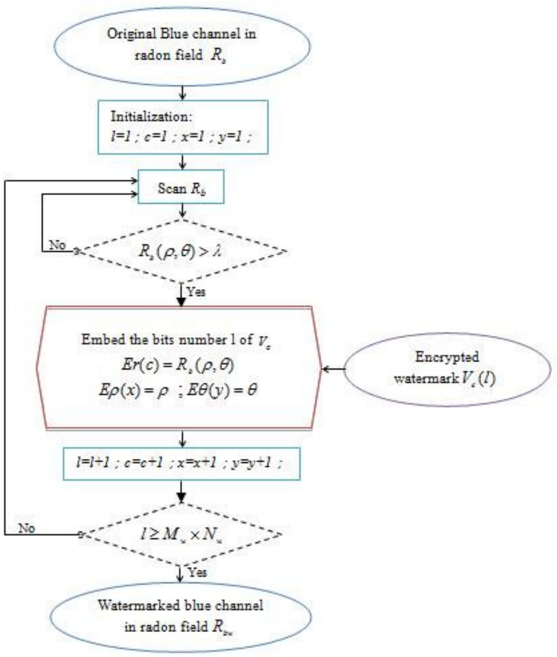

Original Blue channel in radon field R_b
Initialization: $l=1$; $c=1$; $x=1$; $y=1$;
Scan R_b
No
$R_b(\rho,\theta) > \lambda$
Yes
Embed the bits number l of V_c
$Er(c) = R_b(\rho,\theta)$
$E\rho(x) = \rho$; $E\theta(y) = \theta$
Encrypted watermark $V_c(l)$
$l=l+1$; $c=c+1$; $x=x+1$; $y=y+1$;
No
$l \geq M_w \times N_w$
Yes
Watermarked blue channel in radon field R_{bw}

Figure 4: Embedding process

$$\begin{cases} E_\theta(y) = \theta \\ E_\rho(x) = \rho. \end{cases}$$

Step 4: Watermarking image in spatial field. We transform the watermarking blue matrix R_{bw} by the inverse DRT (IDRT) and we combine the RGB channels of the image to create the watermarked color image in spatial domain I_w.

3.1.2 Watermark Recovering Process

This algorithm represents the second principal algorithm in every watermarking approach. It serves to recover the embedded information with minimal loss. For this reason, it is necessary to respect the used parameter of the processing field and to use the details of the embedding program in this approach. This process is detailed by the illustrated in Figure 5.

In this algorithm, we extract the blue channel I_{bw}.Then, we apply a radon transform with the same parameter used in the embedding process (step2 and step 3 in paragraph 3.1.1) for the extracted channel. This transformation gives the watermarked blue channel in radon field R'_{bw}. Next, we use the two saved vectors E_ρ and E_θ detect the position of the used coefficient to embed the encrypted watermark. The amplitudes of the detecting coefficients are saved in the vector E'_r by using the following equation:

$$Er'(cl) = R'_{bw}(E_\rho(cl), E_\theta(cl)),$$

where $cl = 1, 2, 3, ..., M_w \times N_w$ This vector contains the used coefficients in embedding the watermark. In order to decide if the embedding bits is equal to 1 or 0. We

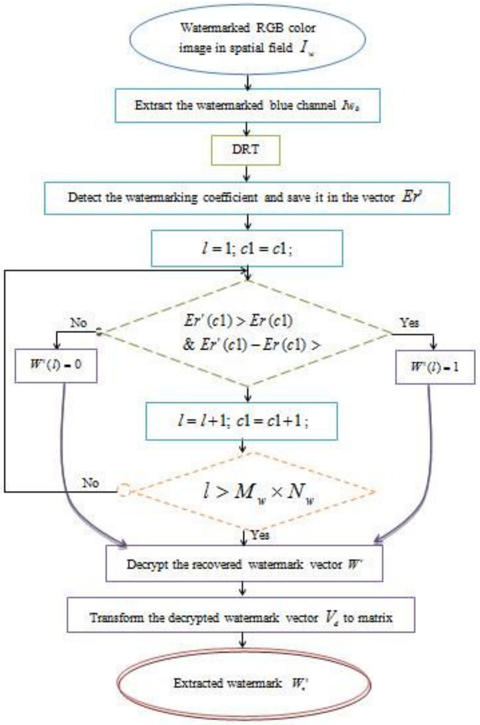

Watermarked RGB color image in spatial field I_w
Extract the watermarked blue channel Iw_b
DRT
Detect the watermarking coefficient and save it in the vector Er'
$l=1$; $c1=c1$;
No
$Er'(c1) > Er(c1)$
$\&\ Er'(c1) - Er(c1) >$
Yes
$W'(l)=0$
$W'(l)=1$
$l=l+1$; $c1=c1+1$;
No
$l > M_w \times N_w$
Yes
Decrypt the recovered watermark vector W'
Transform the decrypted watermark vector V_d to matrix
Extracted watermark W'_r

Figure 5: Watermark extraction process

compare the two saved vectors E'_r and E_r bit by bits and its difference with the selected threshold T. In this step, we use the comparison test defined as follows:

$$\begin{cases} W'(l) = 1\ if Er(l) \geq Er'(l)\ and\ Er'(l) - Er(l) > T \\ W'(l) = 0\ else. \end{cases}$$

The recovered watermark W' is a one-dimensional vector which represents the encrypted watermark. To decrypt it, we use the same algorithm used to encrypt the original watermark with the same parameters and steps. It is defined as follows:

1) Decompose the recovered encrypted watermark vector in N equal blocks B'_i with length equal to the length of the blocks B_i where $i = 1, 2, ..., N$:

$$V'_{water}(N) = B'_1, B'_2, ..., B'_N.$$

2) In order to recover the original watermark, we apply S_{max} iteratively inverses permutations to the recover encrypted watermark vector.The first inverse permutation is defined as follows:

$$P^{-1}_{S'=1}(B'_1, B'_2, ..., B'_N) \\ = V'_{water}(B'_{((N/2)+1}, ..., B'_N, B'_{(N/2)}, ..., B'_1).$$

The next of the inverse permutation process defined by the following equation:

$$P^{-1}_{S'=\alpha}(B'_1, B'_2, ..., B'_N) \\ = P^{-1}_{S'=\alpha-1}(B'_{((N/2)+1}, ..., B'_N, B'_{(N/2)}, ..., B'_1),$$

where $\alpha = 2 \rightarrow S_{max}$ and S_{max} represents the number of the permutation iteration used in the embedding watermark process. The decrypted watermark vector after S_{max} inverse permutation V_d it is defined as follows:

$$V_d(l) = P^{-1}_{S=S_{max}}(l).$$

3) Devise V_d into blocks $B_i^{'}$ with equal length equal to the length of the blocks B_i where $i = 1, 2, ..., N$. Use the same key sing in the embedding process $key0$ to decrypt the first block Br_1 by using the following equation:

$$Br_1 = B_1^{'} \oplus key0.$$

4) Decrypt the second block $B_2^{'}$ by using the following equation:

$$Br_2 = B_2^{'} \oplus B_1^{'}.$$

5) Generally, after each decrypting iterates of each block Br_i, the resulting decrypted blocks is used to decrypt the next block Br_{i+1} where $i = 3, 4, ..., N-1, N$.

$$Br_i = B_i^{'} \oplus B_{i-1}^{'}.$$

6) The decrypted watermark after application of XOR function is defined as follows:

$$V_d^{'}(N) = Br_1, Br_2, ..., Br_N.$$

Finally, to obtain the decrypted watermark, we transformed the vector V_d to matrix with size equal to $[M_w N_w]$ defined as follows:

$$V_d^{'}(N) \rightarrow W_O^{'}(x, y).$$

$W_o^{'}$ is the recover watermark.

3.2 Experimental Results

To evaluate the performance of the proposed watermarking scheme, we use a data base composed with 100 logical watermarks coded on 0 and 1 binary ($d = 2$) and 50 host cover images. Different tests give results close to the present results in this paper. In this work, we present the results of the standard Lena image RGB color (see Fig.18) with size (256×256) and a watermark with a size (80×80) (see Figure 20). We apply a discrete radon transform to the original image with an integration angle path "in degree $\Delta\theta = 1$ for $\theta \in [02\pi]$ and an integration scale path $\Delta\rho = 1 pixels$ for $\rho \in [0\sqrt{M^2 + N^2} + 1]$.

The similitude rate between the extracted watermark and the original watermark is continuously computed to test the robustness of this approach. This is done by the normalized Cross-correlation presented in the following equation:

$$NC = \frac{\sum_{i=1}^{M_q} \sum_{i=1}^{N_q} W_o W_o^{'}}{\sqrt{\sum_1^{M_q} [\sum_1^{N_q} W_o^2] \sum_1^{M_q} [\sum_1^{N_q} W_o^{'2}]}}.$$

On the other hand, the imperceptibility of the embedded watermark is a constraint that must be respected. A measure of similarity rate based on the PSNR described by the following equation is computed after each watermarking process with respect to the gain factor used. A threshold of $37\,dB$ is fixed to verify if some distortions begin to appear on the watermarked image in addition to a psycho-visual decision.

$$PSNR = 10 \log(\frac{d^2}{MSE}).$$

Where d represents the maximal image intensities ($d = 256$ for the host cover image and $d = 2$ for the used logical watermark in our case) and MSE is calculated in the following equation:

$$MSE = \frac{1}{M_w \times N_w} \sum_{i=1}^{M_w \times N_w} (W_o - W_o^{'})^2.$$

The first step of this simulation study consists to select the threshold λ uses to select the radon coefficients used to embed the watermark.

3.2.1 Selection of the Threshold λ

The threshold λ which is chosen for selecting the radon coefficients will be used to embed the watermark in order to improve the robustness of the proposed watermarking scheme against different attacks categories. For that purpose, it is provided in the following five different values of λ, (λ_0, λ_{min}, $\lambda_{means-inf}$,$\lambda_{means-max}$ and λ_{max}), used to select the encoding coefficients. A comparative study is performed to select the optimum threshold used to select the embedding coefficients.

1) Watermarking in coefficients equal to 0.
To select the encoded coefficients $R_E(\rho, \theta)$, we used a threshold $\lambda_0 = 0$. The number of the selected coefficient is equal to $M_w * N_w$ where:

$$if\ \mathrm{R}_b(\rho, \theta) = \lambda_0 \qquad then \quad R_E(\rho, \theta) = R_b(\rho, \theta).$$

The simulations tests show that we cannot recover the embedding watermark in the coefficients equal to 0. So, the comparative parameters give the following results: $C = NaN$, $PSNR = NaN$ and $BER = 6400bits = 100\%$.

2) Watermarking in minimal coefficients different to 0.
In this algorithm, the coefficients whose values are minimal and different from zero are selected. For this reason we used a counter "count" to compute the number of zero in the matrix $R_b(\rho, \theta)$:

$$\begin{cases} count = 0; \\ if\ \ R_b(\rho, \theta) = 0\ \ then\ \ count = count + 1; \\ end. \end{cases}$$

The threshold λ_{min} is selected as follows:

$$\lambda_{min} = V(((len - count) - (M_w \times N_w)) + 1),$$

where len is the length V of the vector representing the coefficients of the host radon image organized in downward order.
The selected coefficient to embed watermark are:

$$R_E(\rho,\theta) = R_b(\rho,\theta) \quad where\, R_b(\rho,\theta) \leq \lambda_{min} \\ and \;\; R_b(\rho,\theta) \neq 0.$$

The simulations tests give a threshold $\lambda_{min} = 5073$ and they show that we can recover the embedding watermark by the following parameter quality results: $C = 0.3023$, $PSNR = 52.3670dB$, $BER = 2413bits = 37.70$ %. The visual results given in Figures 6 and 7.

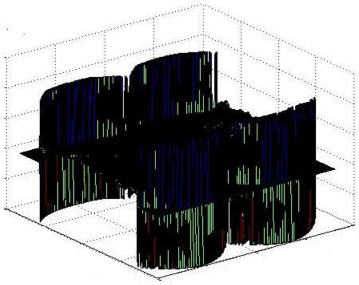


Figure 6: Watermarking image in radon field

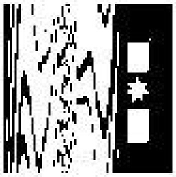


Figure 7: Recovered watermark

3) Watermarking in maximal coefficients.
The used process to define λ_{max} is same of the defined process in step 2 of the paragraph 3.1.1 (watermark embedding process) where $\lambda_{max} = \lambda_{opt}$.

The simulations tests give a threshold $\lambda_{max} = 32208$ and they show that we can recover the embedding watermark by the following parameters quality results: $C = 1$, $PSNR = Inf$ and $BER = 0bits = 0$ %. The visual results given in Figures 8 and 9.

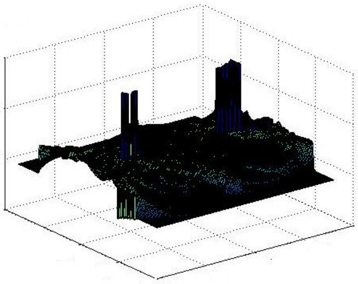


Figure 8: Watermarking image in radon field

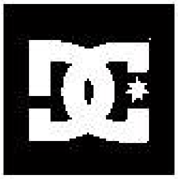


Figure 9: Recovered watermark

4) Watermarking in the higher coefficients to λ_{means} and different to the maximal coefficients.
In this algorithm the selected coefficient to embed watermark are:

$$R_E(\rho,\theta) = R_b(\rho,\theta) \\ where\; R_b(\rho,\theta) > \lambda_{means}\, and\, R_b(\rho,\theta) < \lambda_{max}.$$

The used threshold in this algorithm is noted by $\lambda_{means-sup}$.The simulations results give $\lambda_{means} = \lambda_{means-sup} = 18842$ and $\lambda_{max} = 32208$. The simulation tests show that we can recover the embedding watermark by the following parameter quality results: $C = 0.9149$, $PSNR = 62.5942$ and $BER = 229bits = 3.57\%$. The visual results given in Figures 10 and 11.

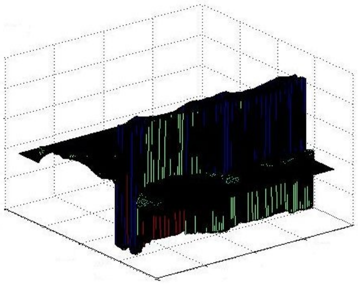


Figure 10: Watermarking image in radon field

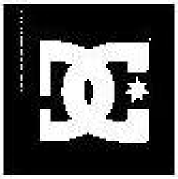


Figure 11: Recovered watermark

5) Watermarking in the less coefficients to λ_{means} and different to the minimal coefficients.
In this algorithm the selected coefficient to embed watermark are:

$$R_E(\rho,\theta) = R_b(\rho,\theta) \qquad where \quad R_b(\rho,\theta) < \lambda_{means} \\ and \;\; R_b(\rho,\theta) > \lambda_{min}.$$

The used threshold in this algorithm is noted by $\lambda_{means-inf}$. The simulations results give $\lambda_{means} = \lambda_{means-inf} = 18842$ and $\lambda_{min} = 5073$. The simulation tests show that we can recover the embedding watermark by the following parameter quality results: $C = 0.7957$, $PSNR = 58.8606$ and $BER = 541bits = 8.45\%$. The visual results given Figures 12 and 13.

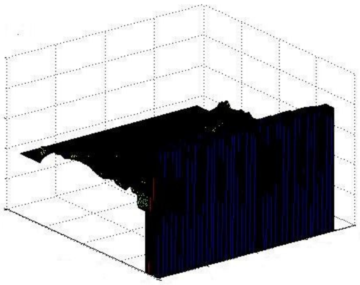


Figure 12: Watermarking image in radon field

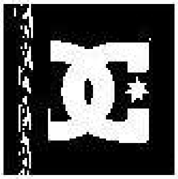


Figure 13: Recovered watermark

6) Comparative study to select λ.

Figures 14, 15, and 16 show the variation of the correlation, PSNR and BER of the recovered watermark for different values of threshold λ.

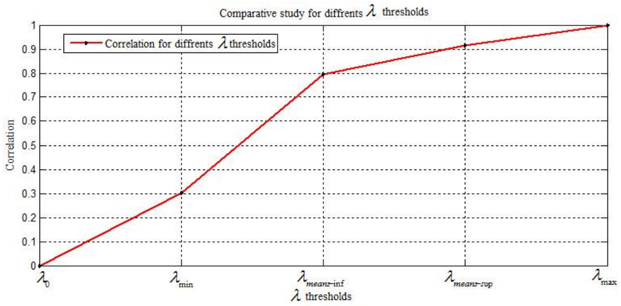


Figure 14: Correlation variation of the recovered watermark for different values of threshold λ

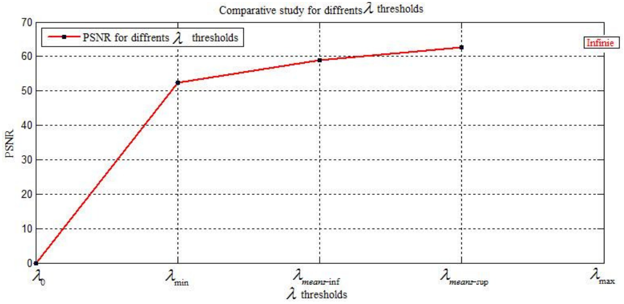


Figure 15: PSNR variation of the recovered watermark for different values of threshold λ

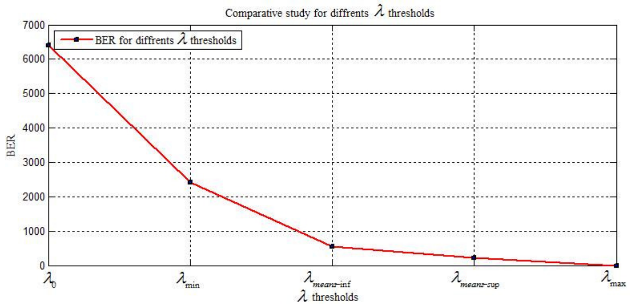


Figure 16: BER variation of the recovered watermark for different values of threshold λ

Figures 14, 15, and 16 show that the more threshold λ increases the better the correlation of the recovered watermark becomes. So, the optimum threshold λ is $\lambda_{opt} = \lambda_{max}$.Generally, the robustness of the proposed method against synchronous and asynchronous attacks for the higher coefficient depends on the highest energy of the radon region. The important peaks of radon domain are located at the points corresponding to the projection parameter. Besides, the coefficients of the radon region in which the highest energy is located and set on the line of projection contain the significant information of the original image. So, they allow a good recovery of the transformed information with inverse radon transform.

3.2.2 Robustness of the Proposed Watermarking Approach

An example of an original and a watermarked image is illustrated in Figure 17 and 18. The set of Figures 19, 20 and 21 illustrate respectively the original watermark, the encrypted one and the decrypted recovered watermark from the radon field.

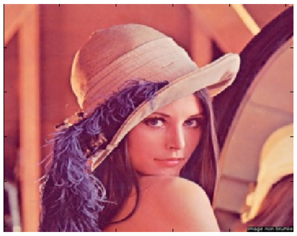


Figure 17: Original image

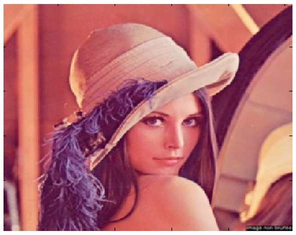


Figure 18: watermarked image

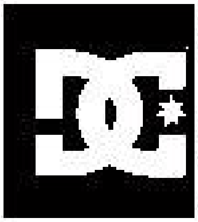


Figure 19: Original watermark

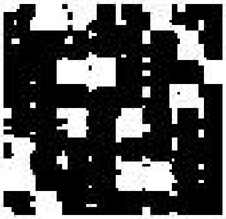


Figure 20: Encripted watermark

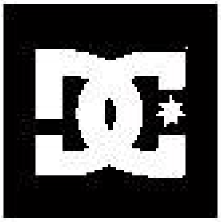


Figure 21: Recovered watermark

In the simulation test, we use a factor gain $G = 1000$ to embed the watermark in the selected coefficient in radon field. These coefficients are higher to $\lambda_{opt} = 32208$. This application gives a PSNR value between the original and the watermarked host image equal to $PSNR = 30.89dB$. The normalized cross-correlation between original and correlating watermark is $NC = 1$. So, no visible differences are detected between the original and the recovered watermark.

In order to improve the correction of the recovered watermark, we insist in this version that the embedded watermark is binary coded on 0 and 1. So, it allows just two different intensities scale. Consequently, the used factor d to compute the PSNR between original and recovered watermark is equal to 2 and gives the present results in Tables 1 and 2.

Tables 1 and 2 show the effectiveness of the proposed watermarking approach to resist the different STIRMARK attacks. We note that the resistivity of the proposed approach against geometric attacks is very high. This efficiency is related to the properties of the Discreet Radon Transform. Also, the effectiveness of the proposed method to resist common image processing attacks is related to an accurate selection of the coefficients which are selected from the image in Radon field. These coefficients presenting the highest energy in the Radon field contain

Table 1: Resistance of the proposed method against the common image processing attacks

ATTACKS	NC	PSNR	BER	
			bits	percent
$Conv_2$	1	inf	0	0
$Median_3$	0.9878	71.0075	33	0.51
$Median_5$	0.9793	68.7107	56	0.875
$Median_7$	0.9841	69.8579	43	0.67
$PSNR_0$	0.9149	62.5942	229	3.578
$PSNR_50$	1	inf	0	0
$Noise_20$	0.9863	70.5106	37	0.578
$Noise_40$	0.9826	69.4716	47	0.73
$JPEG_50$	0.9413	63.9654	167	2.6
$JPEG_70$	0.9974	77.7416	7	0.109
$JPEG_80$	1	inf	0	0

the significant information of the original image, which allow the withstanding against the common image processing attacks. These coefficients are located and set on the line of projection, so they allow a good recuperation of transformed information with inverse radon transform. This robustness is mathematically proved in the following Section. The following images illustrate zoomed views of the selected coefficients in the Radon field "Figure 22" and the coded watermark in these coefficients illustrated in the figure "Figure 23".

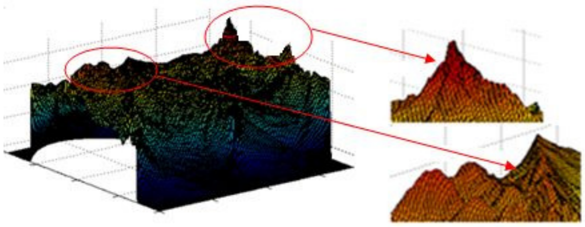


Figure 22: The used sets of coefficient to embed watermark in Radon field

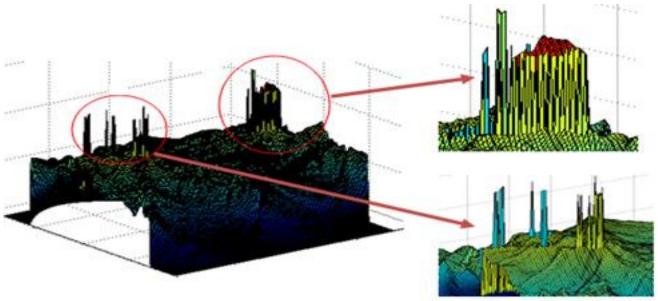


Figure 23: The sets of watermarked coefficient the watermark presence in the Radon field

Table 2: Resistance of the proposed method against the geometric attacks

ATTACKS	NC	PSNR	BER	
			bits	percent
$RESC_90$	0.9993	83.1828	2	0.03
$RESC_75$	0.9989	81.4214	3	0.046
$RNDDIST_1$	1	inf	0	0
$LARDIST$	1	inf	0	0
RML_10	1	inf	0	0
RML_40	0.9985	80.1720	4	0.06
$AFFINE_1$	1	inf	0	0
$AFFINE_8$	0.9996	86.1926	1	0.015
ROT_51	1	inf	0	0
$ROT_0.75$	0.9952	75.0532	13	0.2
ROT_1	1	inf	0	0
ROT_90	1	inf	0	0
$ROTCROP_0.5$	1	inf	0	0
$ROTCROP_1$	0.9996	86.1926	1	0.015
$ROTCROP_10$	1	inf	0	0
$ROTSCALE_0.5$	1	inf	0	0
$ROTSCALE_1$	1	inf	0	0
$SCALING_0.9$	0.9989	81.4214	3	0.046
$SCALING_0.7$	0.9896	71.7210	28	0.437

In another hand, the watermarked image is illustrated in Figure 18. It represents an imperceptible watermarking scheme. We will prove in the following mathematical study why the watermarking system in radon field represents an imperceptible watermarking approach. In fact, embedding information in radon coefficient especially in the higher radon coefficients under the perceptibility threshold is defined by the Weber law. We note the data loss "error" between the recovered images after inverse radon transform of the original image $I'(x,y)$ and the watermarked image $I(x,y)$ by ε. The following equations prove that error has no visual effect on the watermarked image.

$$I^{'}(x,y) - I_w(x,y) = \varepsilon(x,y)$$

or the Weber law imposes that:

$$\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{I^{'}(i,j) - I(i,j)}{I^{'}(i,j)} \leq \tau \qquad \text{where } \tau \cong (2\% - 3\%).$$

Since the simulation results proved that $\varepsilon < \tau$. So, perceptually we can say that no visible changes are engendered by the watermarking approach in radon domain, then:

$$I^{'} = I \qquad \text{and } \varepsilon \rightarrow 0.$$

Figures 24, 25, and 26 illustrate an original and Radon transformed image followed by the positions (In red) of the imperceptible distortions introduced by applying the radon transform and recovering the image by the inverse Radon transform without any watermark embedding.

Figure 24: Original image

Figure 25: IDRT watermarked image

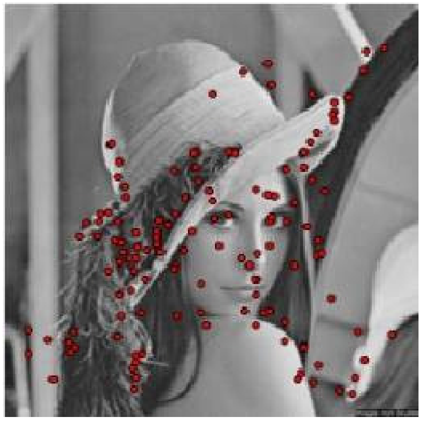


Figure 26: Difference between IDRT original image and IDRT watermarked image

In these Sections, we improve that the watermarking image in radon field presenting a more robustness against asynchronous attacks presented in table I is proved. In addition, when dealing with image, this transform doesn't engender any perceptual degradation.

4 Justifying Robustness Against Asynchronous Attacks

In this Section, we will detail why the proposed approach resists against the geometric attacks. This is done through the mathematical characteristics of the Radon transform.

The robustness of the proposed approach and its resistance against asynchronous attacks can be justified only if we prove mathematically that a change of the Radon matrix coefficients presented by a watermark insertion is invariant against geometric transforms and has no mathematical or visual impact on the image in its spatial representation. In addition, this transform has to be entirely reversible and conservative. In the case it is almost conservative, we have a loss of data in the inverse process when applying the inverse DRT. The data loss does not affect in any way the perceptibility of the watermarked image in the spatial domain based on the Weber law. All these constraints must be tested and proved in the following Section. Given an image $I(x,y)$ defined in $\Re_2^+$, $R(\rho,\theta)$ represents its radon projection in $\Re_2^+$. (x,y) and (ρ,θ) represent respectively the coordinates of the image in spatial and Radon domain.$[MN]$ and $[N_\rho N_\theta]$ represent the size of the image in spatial and Radon field. The invariance faces to the following geometric transforms have to be proved.

4.1 Linearity

Given $g(x,y)=\beta I(x,y)$ where β is a constant. The DRT of $g(x,y)$ gives the following relation:

$$DRT[g(x,y)] = R_1(\rho,\theta) \qquad \text{where } \frac{R_1(\rho,\theta)}{R(\rho,\theta)} = \beta',$$

or $\beta' = \beta + \Delta\beta$ through different tests we find that:

$$\begin{array}{rcl} \Delta\beta <<<<<<< \beta & & \text{so } \beta' \cong \beta. \\ Do:\ R_1(\rho,\theta) & = & \beta R(\rho,\theta). \\ So:\ DRT[g(x,y)] & = & \beta DRT[I(x,y)]. \end{array}$$

Similarly, we define the following relationship:

$$If:\ g(x,y) = \beta_1 I_1(x,y) + \beta_2 J(x,y).$$

The Radon Transformation of J gives the following results:

$$DRT[g(x,y)] = \beta_1 DRT[I_1(x,y)] + \beta_2 DRT[J(x,y)],$$

where $[I1(x,y)]$ and $[J(x,y)]$ are two image defined in spatial field and $\beta1$ and $\beta2$ are two constants. This relation proves that the DRT is linearly invariant.

4.2 Image Scaling

A scaling on the $\overrightarrow{X}$ and $\overrightarrow{Y}$ axis of the image is applied as presented in Figure 27 and the following equation:

$$g(x,y) = I(x-x_0, y-y_0).$$

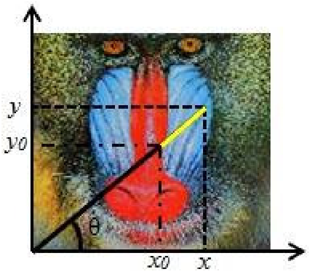


Figure 27: Scaling image on X and Y axis

The Radon transform of $J(x,y)$ is as follows:

$$DRT[g(x,y)] = R_1(\rho_1,\theta) = DRT[I(x-x_0,y-y_0)],$$

where

$$\begin{aligned}\rho_1 &= (x - x_0)\cos\theta + (y - y_0)\sin\theta\\ &= \quad x\cos\theta - x_0\cos\theta + y\sin\theta - y_0\sin\theta\\ &= \underbrace{x\cos\theta + y\sin\theta}_{\rho} - x_0\cos\theta - y_0\sin\theta\\ &= \rho - x_0\cos\theta - y_0)\sin\theta.\end{aligned}$$

So:

$$DRT[g(x,y)] = R_1(\rho_1,\theta) = R(\rho - x_0\cos\theta - y_0\sin\theta, \theta).$$

Also, the error defined by $\Delta\rho = -x_0\cos\theta - y_0\sin\theta, \theta$ is lower than the value of ρ. Since the variation $\Delta\rho$ cannot allow the projection of a pixel neighbor defined by its coordinates (ρ,θ) due to its size $\Delta\rho <<<< \rho$ then, the radon transformation depends only on the value of ρ.

4.3 Image Rotation

Supposing that $K(\rho,\theta)$ represents the polar coordinate of $I(x,y)$ and $g(\rho,\theta) = K(\rho,\theta-\varphi)$ with φ represents the angle of circular shifting. The results of rotating a spatial image in the radon field is studied and presented in Figure 28 and the following equations:

$$\begin{aligned}DRT[g(\rho,\theta)] &= DRT[(\rho,\theta-\varphi)] = R_1(\rho_1,\theta_1)\\ &= I(x\cos\theta\cos\varphi + y\sin\theta\sin\varphi,\\ &\qquad -y\cos\theta\sin\varphi + y\sin\theta\cos\varphi).\\ R_1(\rho_1,\theta_1) &= I(x\cos(\theta-\varphi), y\sin(\theta-\varphi)).\\ &= R(\rho,\theta-\varphi).\\ \rho_1 &= \rho\\ \theta_1 &= \theta-\varphi.\end{aligned}$$

So,

$$DRT[g(\rho,\theta)] = R_1(\rho_1,\theta_1) = R(\rho,\theta-\varphi).$$

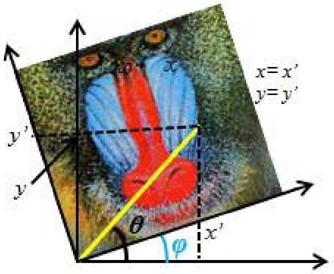


Figure 28: Circular shifted image by angle

This relation shows that the radon transform depends only on the value of $\theta - \varphi$ and its magnitude is constant. Consequently the projected pixel will change its location in the Radon field according to the angular rotation applied. This proves that angular rotations are conserved to generate the correspondent Radon coefficients.

4.4 Cropping Rotation and Scaling

In this section, we test the invariance of the DRT if different asynchronous attacks are combined simultaneously such as rotation and scaling or cropping without changing the axis projection.

Case 1. Scaling on Y axis and circular shifting by an angle (See Figure 29).

$$\begin{aligned}DRT[g(\rho,\theta)] &= DRT[K(\rho,\theta-\varphi)] = R_1(\rho_1,\theta_1)\\ &= I(x\cos\theta\cos\varphi + x\sin\theta\sin\varphi,\\ &\qquad -(y-y_0)cos\theta\sin\varphi\\ &\qquad +(y-y_0)\sin\theta\cos\varphi.\\ R_1(\rho_1,\theta_1) &= I(x\cos(\theta-\varphi),(y-y_0)\sin(\theta-\varphi)).\\ &= R(\rho - y_0\sin(\theta-\varphi),\theta-\varphi).\\ \rho_1 &= \rho - y_0\sin(\theta-\varphi)\\ \theta_1 &= \theta-\varphi.\end{aligned}$$

So, in this case:

$$DRT[K(\rho,\theta-\varphi)] = R(\rho - y_0\sin(\theta-\varphi),\theta-\varphi).$$

The error is defined by:

$$\Delta\rho = -y_0\sin(\theta-\varphi) <<<< \rho.$$

Experimental test proved that the error found is very small compared with ρ ($\Delta\rho <<<< \rho$).

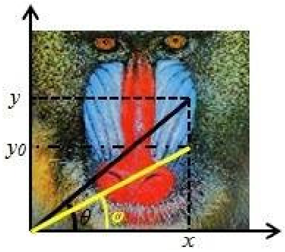


Figure 29: Scaling on Y axis and circular shifting by an angle

Case 2. Scaling on Xaxis and circular shifted of crop in image (See Figure 30).

$$\begin{aligned}DRT[g(\rho,\theta)] &= DRT[K(\rho,\theta-\varphi)] = R_1(\rho_1,\theta_1)\\ &= I((x-x_0)\cos\theta\cos\varphi\\ &\qquad +(x-x_0)\sin\theta\sin\varphi,\\ &\qquad -y\cos\theta\sin\varphi + y\sin\theta\cos\varphi.\\ R_1(\rho_1,\theta_1) &= I((x-x_0)\cos(\theta-\varphi), ysin(\theta-\varphi))\\ &= R(\rho - x_0\cos(\theta-\varphi),\theta-\varphi)\\ \rho_1 &= \rho - x_0\cos(\theta-\varphi)\\ \theta_1 &= \theta-\varphi.\end{aligned}$$

So,

$$DRT[K(\rho,\theta-\varphi)] = R(\rho - x_0\cos(\theta-\varphi),\theta-\varphi).$$

The error is

$$\Delta\rho = -x_0\cos(\theta-\varphi) <<<< \rho.$$

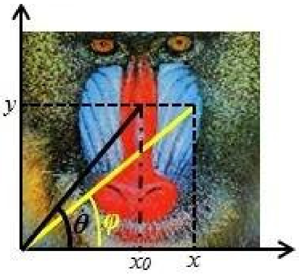


Figure 30: Scaling on Xaxis and circular shifted of crop in image

Case 3. Scaling on Xaxis and circular shifted of crop in image (See Figure 31).

$$\begin{aligned}
DRT[g(\rho,\theta)] &= DRT[K(\rho,\theta-\varphi)] = R_1(\rho_1,\theta_1)\\
&= I((x-x_0)\cos\theta\cos\varphi\\
&\quad +(x-x_0)\sin\theta\sin\varphi,\\
&\quad -(Y-Y_0)\cos\theta\sin\varphi\\
&\quad +(Y-Y_0)\sin\theta\cos\varphi).\\
R_1(\rho_1,\theta_1) &= I((x-x_0)\cos(\theta-\varphi),\\
&\qquad (Y-Y_0)\sin(\theta-\varphi)).\\
&= R(\rho-x_0\cos(\theta-\varphi)\\
&\quad -y_0\sin(\theta-\varphi),\theta-\varphi).\\
\rho_1 &= \rho-x_0\cos(\theta-\varphi)-y_0\sin(\theta-\varphi)\\
\theta_1 &= \theta-\varphi.
\end{aligned}$$

So, in this case:

$$\begin{aligned}
&DRT[K(\rho,\theta-\varphi)]\\
=\;& R(\rho-x_0\cos(\theta-\varphi),-y_0\sin(\theta-\varphi),\theta-\varphi).
\end{aligned}$$

The error is defined by:

$$\Delta\rho = x_0\cos(\theta-\varphi)y_0\sin(\theta-\varphi).$$

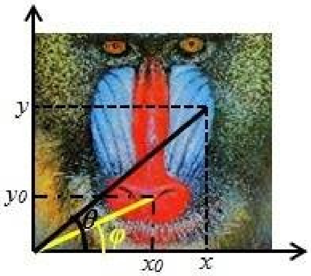


Figure 31: Scaling on X and T axis and circular shifted of crop in image

Due to its feeble value, the error $\Delta\rho$ cannot change the original integer quantized coefficient in Radon field. So, the used coefficient to code the watermark does not change and the watermark is correctly recovered. As proved above, face to singular or composed geometric transform, the Radon transform offers invariance to the transformed image. Consequently, the distortions and geometric transforms applied on the spatial watermarked image have generally no effect on the embedded watermark in the radon field since these variations are conserved over the Radon coefficients where the watermark is coded.

5 Comparative Study

In order to prove the efficiency and high robustness of the proposed method, a comparison study illustrated in Figures 32 and 33 is conducted with recent proposed approach in the literature exploiting the Radon domain [12] and [17].

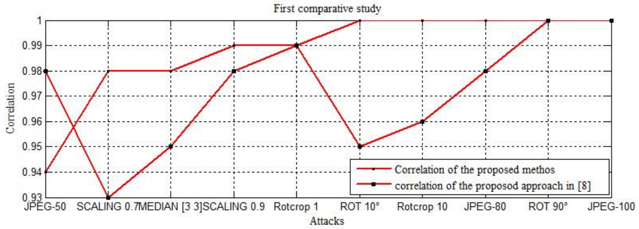


Figure 32: Comparative study between our proposed method and the developed technique in [12]

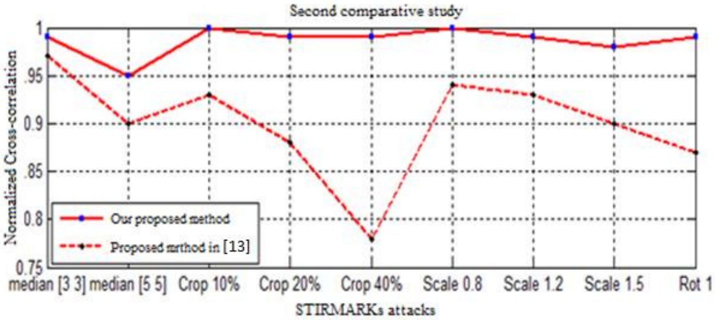


Figure 33: Comparative study between our proposed method and the Wang method in [17]

Compared to the proposed methods in [12] and [17], Figure 32 and Figure 33 show that the robustness of the proposed watermarking scheme in the radon field is more effective than the previous schemes. This efficiency is due to the propriety of radon field and its performance in detecting the important peaks by sing the inverse radon transform. The proposed watermarking scheme does not degrade the visual perception of the watermarked image and it is robust against the two categories of attacks.

6 Conclusions

In this paper a watermarking approach based on the Radon transform is presented. The watermark is coded in selected coefficients with respect to specific mathematical characteristics and the energy is characterized by higher robustness against various attacks types. The proposed scheme presents high robustness especially against asynchronous attacks. This resistance against these geometric attacks is studied and proved mathematically. Accordingly, the equilibrium between watermarking constraints is achieved. Robustness and imperceptibility are respected and the embedding capacity is increased.

Acknowledgments

This study was supported by CEREP Research laboratory in College of Sciences and techniques in Tunisia. The authors gratefully acknowledge the anonymous reviewers for their valuable comments.

References

[1] A. Averbuch, I. Sedelnikov, and Y. Shkolnisky, "CT reconstruction from parallel and Fan-Beam projections by a 2-D discrete radon transform," *IEEE Transactions on Image Processing*, vol. 21, no. 2, pp. 733–741, 2012.

[2] L. Baisa and R. R. Manthalkar, "An overview of transform domain robust digital image watermarking algorithms," *Journal of Emerging Trends in Computing and Information Sciences*, vol. 2, no. 1, pp. 2079–8407, 2010.

[3] G. Beylkin, "Discrete radon transform," *IEEE Transactions on acoustics speech and signal processing*, vol. 35, no. 2, pp 162–172, 1987.

[4] C. C. Chang, K. F. Hwang, and M. S. Hwang, "A digital watermarking scheme using human visual effects," *Informatica: An International Journal of Computing and Informatics*, vol. 24, no. 4, pp. 505–511, 2000.

[5] C. C. Chang, K. F. Hwang, and M. S. Hwang, "A Feature-Oriented copyright owner proving technique for still images," *International Journal of Software Engineering and Knowledge Engineering*, vol. 12, no. 3, pp. 317–330, 2002.

[6] RSR. Channapragada, AS. Mantha, and Munaga V. N. K. Prasad, "Study of contemporary digital watermarking techniques," *IJCSI International Journal of Computer Science Issues*, vol. 9, no. 1, pp. 1694–0814, 2012.

[7] Y. Fu, "Robust oblivious image watermarking scheme based on coefficient relation," *International Journal for Light and Electron Optics*, vol. 124, no. 6, pp. 517–521, 2013.

[8] M. S. Hwang, C. C. Chang, and K. F. Hwang, "A watermarking technique based on One-way hash functions," *IEEE Transactions on Consumer Electronics*, vol. 45, no. 2, pp. 286–294, 1999.

[9] M. S. Hwang, C. C. Chang, and K. F. Hwang, "Digital watermarking of images using neural networks," *Journal of Electronic Imaging*, vol. 9, no. 4, pp. 548–555, 2000.

[10] G. Dayalin Leena, S. Selva Dhayanithy, and M. S. Hwang, "Robust image watermarking in frequency domain," *International Journal of Innovation and Applied Studies*, vol. 2, no. 4, pp. 582–587, 2013.

[11] L. Li, S. Li, A. Abraham, and J. S. Pan, "Geometrically invariant image watermarking using polar harmonic transforms," *Information Sciences*, vol. 199, pp. 1–19, 2012.

[12] I. Nasir, F. Khelifi, J. Jiang, and S. Ipson, "Robust image watermarking via geometrically invariant feature points and image normalisation," in *Proceeding of IET Image*, vol. 6, pp. 354, 2012.

[13] D. Simitopoulos, D. E. Koutsonanos, and M. G. Strintzis, "Robust image watermarking based on generalized radon transformations," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 8, pp. 732–745, 2003.

[14] Q. Su, Y. Niu, X. Liu, and T. Yao, "A novel blind digital watermarking algorithm for embedding color image into color image," *International Journal for Light and Electron Optics*, vol. 124, no. 18, pp. 3254–3259, 2013.

[15] Q. Su, Y. Niu, X. Liu, and Y. Zhu, "A blind dual color image watermarking based on IWT and state coding," *Optics Communications*, vol. 285, no. 7, pp. 1717–1724, 2012.

[16] H. H. Tsaia, Y. J. Jhuanga, and Y. S. Lai, "An SVD-based image watermarking in wavelet domain using SVR and PSO," *Applied Soft Computing*, vol. 12, no. 8, pp. 2442–2453, 2012.

[17] E. Vahedi, R. A. Zoroofi, and M. Shiva, "Toward a new wavelet-based watermarking approach for color images using bio-inspired optimization principles," *Digital Signal Processing*, vol. 22, no. 1, pp. 153–162, 2012.

[18] W. W. Zhang, F. Gao, B. Liu, Q. Y. Wen, and H. Chen, "A watermark strategy for quantum images based on quantum fourier transform," *Quantum Information Processing*, vol. 12, no. 2, pp. 793–803, 2013.

Dhekra Essaidani was born in Bizerte, Tunis on 10 January 1989. In 2010, she received the Fundamental license in Electrical Engineering: Automatic and industrial computer from the ESTI (Tunis College of Technology and Computer science in Tunis) in Tunisia and Master degree in Electrical engineering (Automatic and industrial computer) from the ESSTT, in 2012.Since 2012; she enrolled in a doctoral thesis in the Department of Electrical Engineering in College of Sciences and techniques: CEREP Research laboratory. Here domain of interest is: Audio-image and video processing. Here main research interests are the multimedia security.

Hassene Seddik is born in 15 October 1970 in Tunisia, he has obtained the electromechanical engineer degree in 1995 and followed by the master degree in signal processing: speaker recognition and the thesis degree in image processing watermarking using non conventional transformations. He has over 14 international journals papers and 65 conference papers. His domain of interest is: Audio-image and video processing applied in filtering, encryption and watermarking. He belongs to the CEREP research unit and supervises actually five thesis and 08 masters in the field.

Ezzedine Ben Braiek obtained his HDR on 2008 in Electrical engineering from ENSET Tunisia. He is, presently, full professor in the department of electrical engineering at the National High School of Engineering of Tunis (ENSIT) and manager of the research group on vision and image processing at the CEREP. His fields of interest include automatics, electronics, control, computer vision, image processing and its application in handwritten data recognition.

Provably Secure Group Key Exchange Protocol in the Presence of Dishonest Insiders

Ziba Eslami, Mahnaz Noroozi, and Saideh Kabiri Rad
(Corresponding author: Ziba Eslami)

Department of Computer Science, Shahid Beheshti University
G. C., Tehran, Iran
(Email: z_eslami@sbu.ac.ir)

(Received July 16, 2013; revised and accepted Oct. 10 & Dec. 20, 2013)

Abstract

The most important security concern in group key exchange protocols is the semantic security of the produced shared key which dictates that outsiders should not be able to learn anything about the key. It is also challenging for these protocols to retain their security even in the presence of dishonest insiders who do not follow the protocol specifications. In this paper, we propose an identity-based group key exchange protocol which addresses these security concerns. We prove that our scheme achieves semantic security in a well-known adversarial model. We then show that the success probability of recognizing dishonest insiders in the proposed scheme is almost one. We further provide a comparison between our protocol and some other schemes in terms of computation and communication cost, as well as security properties.

Keywords: Bilinear pairing, dishonest insiders, elliptic curve, group key exchange, provable security

1 Introduction

In an Internet conference the participants communicate with each other over an insecure network. In order to prevent the conference contents to be revealed, their communications must be encrypted. Therefore, the participants should agree upon a common key and use it for encrypting the messages. One solution to establish such a common key is using group key exchange protocols in which the group members compute a common key via an insecure public channel cooperatively [16, 20, 21, 36].

The first key exchange protocol was proposed in 1976 by Diffie and Hellman [15]. The scheme enabled two participants to establish a common key and its security was based on the discrete logarithm problem. But it was not suitable for groups of users. In 1982, Ingemaresson et al. [23] proposed the first group key exchange protocol. Both of these schemes were vulnerable against the man-in-the-middle attack, i.e., an adversary could impersonate the participants without being detected. So the authentication property was added to the key exchange protocols [25] and therefore, one could assure that he is establishing the key with legitimate participants [19, 30]. In other words, while security in key exchange protocols is considered against a passive adversary who only eavesdrops, in authenticated key exchange protocols, a stronger class of adversaries is involved who is capable of controlling all communication in the network.

In 1984, Shamir [32] introduced the concept of identity-based cryptosystems. In this setting, a user's private key is generated by a trusted key generator center (KGC), enabling any party to derive the user's public key from his identity and thereby removing the need for public-key certificates. An authenticated key exchange protocol is called identity-based if the users use an identity-based asymmetric key pair instead of a traditional public/private key pair for authentication and determination of the established key. Since identity-based systems simplify the process of key management (compared to the traditional public key systems), they have been considered extensively for designing key exchange protocols [11, 33, 35].

In real world, it is not reasonable to assume that all participants are honest and in fact an important issue in key exchange protocols is the presence of malicious insiders. Two types of malicious behavior are considered in the literature. One is impersonation in which the malicious participant impersonates another entity in the group who is not present [13]. Another type of malicious act which we consider in the paper as of this point, pertains to participants who prohibit the group from computing the same shared key by broadcasting fake information, i.e., values which are not produced according to the protocol specifications [18]. The word "dishonest" refers to this kind of users throughout this paper. To deal with dishonest participants, the fault-tolerance property was introduced for group key exchange protocols in 2002 [34]. This property ensures that honest participants are able to acquire a conference-key; no matter how many dishonest participants exist. Since then, some other fault-tolerant group key exchange protocols have been proposed [22, 37].

So far and to the best of our knowledge, no formal security proof exists for the fault-tolerance property of these schemes. It is one of the goals of this paper to introduce a formal proof for security against dishonest insiders (see Section 4.1). In fact, until recently, 'informal' definitions and proofs were widely used for group key exchange protocols. Most of such definitions were originally stated for two-party protocols and then adapted to a group setting. The informal definitions serve as foundations for the subsequent formal security models for group key exchange protocols. Examples are notions like key privacy, known-key security, key freshness, forward and backward secrecy, entity authentication, unknown key-share resilience, key confirmation, and key control. For more details, the interested reader is referred to [31].

In the paradigm of provable security for key exchange protocols, a 'formal model' must be defined. In this model, the capabilities of the adversary as well as the players should be captured. It has to be clearly stated what it means for the scheme to be secure and provide a proof of its security. The security proof aims to show that the scheme actually achieves the claimed security goals under computational assumptions. The proof usually works via reduction to an underlying hard problem.

Bellare and Rogaway [3] in 1993 proposed the first computational security model for authenticated two-party key exchange protocols. Since then some other security models were proposed [1, 10]. In 2001, Bresson et al. [8] proposed the first computational (game-based) security model for group key exchange protocols (referred to as the BCPQ model) which builds on prior work from the 2-party setting [2, 3, 4]. This model has been widely used to analyze group key exchange protocols [6, 7].

In 1994, Burmester and Desmedt [9] proposed an efficient group key exchange protocol. A variant of this protocol was later proposed in 2008 by Dutta and Barua [17]. The variant has a lower communication cost and can handle the joining and leaving processes of the participants. However, this scheme has some weaknesses: (1) The authors claimed that their scheme could detect the presence of dishonest participants. However, Eslami and Kabiri in [18] designed an attack to prove that this claim was not true and showed how two dishonest insiders could prohibit legitimate participants from obtaining the same shared key; (2) It is unable to identify dishonest participants.

In this paper, we employ elliptic curves and bilinear pairings to propose a group key exchange protocol which does not suffer from these weaknesses. The advantage of elliptic curve-based cryptosystems is their short key size, high processing throughput, and low bandwidth. We prove that our proposed scheme is secure in BCPQ model. Since this model does not consider dishonest participants, we design an experiment to formally prove the security of our protocol in the presence of dishonest participants.

The rest of this paper is organized as follows: Section 2 briefly explains preliminary concepts. Our proposed protocol is described in Section 3. In Section 4, we discuss the security concepts of the proposed protocol. Section 5 is devoted to performance analysis and comparisons. Finally, a conclusion is drawn in Section 6.

2 Preliminaries

We briefly describe the preliminaries needed in the paper. First, the definition of bilinear pairings is given. Then we introduce elliptic curves. The computational problems used for security analysis are listed afterwards. Finally, we describe the security model in which the security of our group key exchange protocol is proven.

2.1 Bilinear Pairings

Let P denote a generator of G_1 where G_1 is an additive group of large order q and let G_2 be a multiplicative group. A pairing is a map $e : G_1 \times G_1 \rightarrow G_2$ which has the following properties:

Bilinearity. For every $S, Q, R \in G_1$ we have:

$$e(S+Q,R) = e(S,R) \cdot e(Q,R).$$

$$e(S,Q+R) = e(S,Q) \cdot e(S,R).$$

Non-degeneracy. There exist $R, Q \in G_1$ such that $e(R,Q) \neq 1_{G_2}$, where 1_{G_2} is the identity element of G_2.

Computability. There exists an efficient algorithm to compute $e(R,Q) \in G_2$ for any $R, Q \in G_1$.

Note that non-degeneracy means that if $e(R,Q)$ is the identity element of G_2, then either R is the identity of G_1 or Q is the identity of G_1. (See [12], pp. 29)

2.2 Elliptic Curves

An elliptic curve defined over $GF(q)$ is given by the equation: $E : y^2 = x^3 + ax + b$, where $a, b \in GF(q)$ and $4a^3 + 27b^2 \neq 0$. The points of E (plus an infinite point O) together with a special operator "+", form a finite Abelian group.

2.3 Cryptographic Assumptions

Definition 1. *(Elliptic Curve Discrete Logarithm (ECDL) Problem) Given kA where A is a point on elliptic curve E, find k.*

The advantage of a distinguisher $\mathcal{A}$ against the $ECDL$ problem is defined as

$$adv_{\mathcal{A},E}^{ECDL} = Pr[\mathcal{A}(kA) = k].$$

Definition 2. *(Elliptic Curve Discrete Logarithm (ECDL) Assumption) Given $kA \in$ elliptic curve E, $adv_{\mathcal{A},E}^{ECDL}$ of a distinguisher $\mathcal{A}$ whose goal is to solve the $ECDL$ problem is negligible.*

Note that a negligible function f has the property that for every polynomial $p(.)$ there exists an N such that for all integers $n > N$ it holds that $f(n) < \frac{1}{p(n)}$.

Definition 3. *(Computational Diffie-Hellman (CDH) Problem) Given (P, aP, bP) where P is the generator of additive group G from order q and $a, b \in Z_q^*$, find abP.*

The advantage of a distinguisher $\mathcal{A}$ against the CDH problem is defined as

$$adv_{\mathcal{A},G}^{CDH} = Pr[\mathcal{A}(P, aP, bP) = abP].$$

Definition 4. *(Computational Diffie-Hellman (CDH) Assumption) Given (P, aP, bP), $adv_{\mathcal{A},G}^{CDH}$ of a distinguisher $\mathcal{A}$ whose goal is to solve the CDH problem is negligible.*

2.4 Security Model

We describe below the adversarial model following Bresson et al.'s definition [8], denoted by $BCPQ$. We then adopt it for the security analysis of our protocol. In this model, the process controlled by a player running on some machine is modeled as an instance of the player. The various types of attacks are modeled by queries to these instances and the security of the session key is modeled through semantic security. In the model, it is assumed that players are honest. Therefore, the security in the presence of dishonest participants is studied separately in Section 4.1. Note that we describe the model in the case of a passive adversary.

Notations: Throughout this section, we use the following notations:

n:	The number of participants;
U_i:	The ith participant;
Π_i^s:	Instance s of U_i;
LL_i:	The long-lived key of player U_i;
SK_i^s:	The session key related to instance s of U_i.

Description. The model consists of protocol participants (n players that should agree on a common key in different sessions through protocol P) and an adversary $\mathcal{A}$ (which is not a player in the formalization, and is given enormous capabilities). A player U_i can have many instances called oracles, involved in distinct concurrent executions of P. As defined in notations, we denote instance s of player U_i as Π_i^s with $s \in \mathbb{N}$. Each player U_i holds a long-lived key LL_i which is a pair of matching public/private keys. The adversary $\mathcal{A}$ controls all communications in a game $Game^{ke}(\mathcal{A}, P)$. This is a game between $\mathcal{A}$ and the oracles Π_i^s involved in the executions of P. During the game, $\mathcal{A}$ can ask a set of queries $(Execute, Reveal, Corrupt, Test)$ defined below with the restriction that the $Test$-query $Test(\Pi_i^s)$ must be asked only once. Moreover, it is only available if Π_i^s is fresh which means that Π_i^s or its partners involved in the execution of P, has not been asked for a *Reveal*-query, and none of them has been asked for a *Corrupt*-query. We now describe the queries that $\mathcal{A}$ can ask in $Game^{ke}(\mathcal{A}, P)$:

- $Execute(U_{l_1}, s_1, \cdots, U_{l_m}, s_m)$: This query models adversary $\mathcal{A}$ initiating an execution of protocol P. This executes the protocol between the (unused) instances $\{\Pi_{l_i}^{s_i}\}_{1 \le i \le m \le n}$, and outputs the transcript of the execution.
- $Reveal(\Pi_i^s)$: This query models the attacks resulting in the session key being revealed. The *Reveal*-query unconditionally forces Π_i^s to release session key SK_i^s.
- $Corrupt(U_i)$: This query models the attacks resulting in the player U_i's LL-key being revealed. Adversary $\mathcal{A}$ gets back LL_i but does not get the internal data of any instances of U_i executing P.
- $Test(\Pi_i^s)$: This query models the semantic security of the session key SK_i^s. After flipping a coin b, $\mathcal{A}$ is given SK_i^s if $b = 1$ or a random string if $b = 0$.

At the end of the game, adversary $\mathcal{A}$ outputs a bit b' and wins the game if $b = b'$.

The security is now defined as follows:

Definition 5. *The key exchange protocol P is $\mathcal{A}$-secure if adversary $\mathcal{A}$ succeeds in $Game^{ke}(\mathcal{A}, P)$ with probability that is at most negligibly greater than $\frac{1}{2}$.*

For more details, the interested reader is referred to [8].

3 The Proposed Scheme

In this section, we propose an identity-based group key exchange protocol based on elliptic curves. Suppose a set of n users $U = \{U_1, U_2, \cdots, U_n\}$ wish to establish a common session key among themselves. We assume that ID_i is the identity of U_i. Our protocol involves four phases:

1) The **initialization** phase in which the Key Generation Center (KGC) outputs the public parameters and generates the private key of each user.
2) The **information exchange** phase which consists of two rounds.
3) The **dishonest user elimination** phase which eliminates the dishonest user if such a person exists.
4) The **key computation** phase in which each user computes the common key.

Now, we explain these phases in details.

The initialization phase. In this phase which is executed once, the KGC performs the following steps:

1) Chooses an elliptic curve group G_1, a multiplicative group G_2 (both of prime order q), a generator $P \in G_1$, a bilinear pairing $e : G_1 \times G_1 \rightarrow G_2$, and a hash function $H : \{0, 1\}^* \rightarrow Z_q^*$.

2) Chooses randomly $s \in Z_q^*$ (KGC's private key).

3) Computes $P_{pub} = sP$ (KGC's public key).

4) Outputs the public parameters $params = \{G_1, G_2, q, P, H, P_{pub}\}$.

5) Computes U_i'private key $S_i = \frac{1}{s+H(ID_i)}P$ and sends it to him securely.

The identity-based key pair produced in this phase will be used in Section 4 to authenticate the exchanged information. This is done by applying the compiler proposed in [27].

The information exchange phase (round 1). Each user $U_i(i = 1, \cdots, n)$ performs the following steps:

1) Chooses randomly $k_i \in Z_q^*$.

2) Computes P_i as follows:

$$P_i = k_i P.$$

3) Sends P_i to all other users.

The information exchange phase (round 2). Each user $U_i(i = 1, \cdots, n)$ performs the following steps:

1) Computes Y_i using P_{i-1} and P_{i+1}.

$$Y_i = k_i(P_{i+1} - P_{i-1}).$$

2) Sends Y_i to all other users.

The dishonest user elimination phase. Each user $U_i(i = 1, \cdots, n)$ verifies the following equation:

$$e(Y_j, P) = e(P_{j+1} - P_{j-1}, P_j), (j = 1, \cdots, i-1, i+1, \cdots, n). \tag{1}$$

If the equation does not hold for some j, then U_j will be considered dishonest. After eliminating dishonest participants from the session, the honest participants restart the protocol.

The key computation phase. Each user $U_i(i = 1, \cdots, n)$ obtains the common key using the following equation:

$$\mathrm{K_i = nk_iP_{i-1} + (n-1)\,Y_i + (n-2)\,Y_{i+1} + \ldots + Y_{i+n-2}}.$$

It may be easily verified that all users compute the same key: $(k_1k_2 + k_2k_3 + \cdots + k_nk_1)P$.

4 Security Analysis

In this section, we analyze security of the proposed scheme, denoted throughout this section by Π. The proof is presented in two parts. In Section 4.1, we prove that Π is capable of identifying dishonest participants, i.e., participants who prohibit the group from computing a common key by broadcasting fake values which are not produced according to the protocol specifications. Recall that the $BCPQ$ model does not support such participants. Therefore, in this part of the proof, we design an experiment which helps us formulate a suitable definition of security for this purpose.

The second part of the proof is devoted to showing that Π achieves security following the $BCPQ$ model, i.e., in the sense of Definition (5). This part is essentially adopted from [27]. Recall from Section 1, that security in key exchange protocols is considered against a passive adversary while in authenticated key exchange protocols, we must consider an active one. In [27], Katz and Yung presented an efficient compiler that transforms any group key exchange protocol secure against a passive eavesdropper to an authenticated protocol which is secure against an active adversary. This is achieved by adding a signature scheme which is strongly unforgeable under adaptive chosen message attack. Therefore, we show in Section 4.2, that our protocol is a secure key exchange protocol following Definition 5. Then applying Katz and Yung's compiler with values produced in initialization phase, we conclude that our scheme is a provably secure protocol for authenticated group key exchange.

4.1 Identification of Dishonest Participants

The aim of this section is to formally prove that in our scheme it is impossible for a participant to prohibit group members from obtaining a common key and remain unnoticed. Although, there exist protocols having formal proof for detecting dishonest participants [26], to the best of our knowledge, the present paper is the first to demonstrate a formal proof of security for the purpose of identifying dishonest participants.

In existing research papers where only the detection of dishonest participants is concerned, the approach is to show that the success probability of an insider in disrupting establishment of a key among honest participants is negligible. However, since we are after identifying dishonest insiders, we prefer to adopt the following definition: *Identification of dishonest insiders is achieved if the success probability of an honest participant in recognizing dishonest insiders is negligibly less than 1.*

We now propose the following experiment which is a game played between a user U_i and an imaginary tester who wishes to see if U_i succeeds in distinguishing protocol values from fake ones. Therefore, a protocol has security against dishonest insiders if the success probability of any honest participant playing this experiment is at most negligibly less than 1.

Formalizing the above notions, let Π be a group key exchange protocol containing n users $\{U_i\}_{i=1}^n$, R rounds of information exchange, and k as the security parameter. We define the following experiment:

Capability of identifying dishonest participants experiment $GKE_{U_i,\Pi}^{disP}(k)$:

- A matrix of $R\times(n-1)$ random bits $b_{r,j} \leftarrow \{0,1\}, r \in \{1,\cdots,R\}, j \in \{1,2,\cdots,i-1,i+1,\cdots,n\}$ is chosen by tester.
- The tester and U_i execute protocol Π where the tester essentially plays the role of all the users except U_i. This execution of the protocol results in a transcript $Trans$ containing all the messages exchanged among participants during different rounds of the information exchange phase of the protocol. Let $M_{r,j}^0$ denote the value produced honestly by U_j in round r (for every $r \in \{1,\cdots,R\}$, $j \in \{1,2,\cdots,i-1,i+1,\cdots,n\}$) and let $M_{r,j}^1$ be the corresponding fake value, i.e., $M_{r,j}^1$ is not computed following the protocol's steps. The tester computes $M_{r,j}^{b_{r,j}}$ as the value of round r produced by U_j.
- At the end, U_i outputs a matrix containing $b'_{r,j}, r \in \{1,\cdots,R\}, j \in \{1,2,\cdots,i-1,i+1,\cdots,n\}$.
- The output of the experiment is defined to be 1 if $b'_{r,j} = b_{r,j}, \forall r \in \{1,\cdots,R\}, \forall j \in \{1,2,\cdots,i-1,i+1,\cdots,n\}$ and 0 otherwise. If $GKE_{U_i,\Pi}^{disP}(k) = 1$, we say that U_i succeeds.

Definition 6. *A group key exchange scheme* Π *is capable of identifying dishonest insiders if for all probabilistic polynomial-time honest users* U_i*, there exists a negligible function negl such that:*

$$Pr[GKE_{U_i,\Pi}^{disP}(k) = 1] \geq 1 - negl(k).$$

We now show that the protocol is secure against dishonest participants with respect to this definition. The following two lemmas form the basis of the proof.

Lemma 1. *In round 1 of the information exchange phase of the proposed protocol, the following holds for an honest participant* U_i *and for all* $j(\neq i)$*:*

$$Pr[U_i(Trans, M_{1,j}^0) = 0] - Pr[U_i(Trans, M_{1,j}^1) = 0] = 1.$$

In other words, U_i *is able to differentiate between* $M_{1,j}^0$ *and* $M_{1,j}^1$ *in Trans with probability 1.*

Proof. Suppose that U_i wants to determine the value of $b_{1,j}$ by observing $M_{1,j}^{b_{1,j}}$ after executing round 1 of the information exchange phase. Note that $M_{1,j}^{b_{1,j}}$ is the value produced honestly in round 1 by U_j if and only if $M_{1,j}^{b_{1,j}}$ is in G_1. Therefore, U_i checks the membership of $M_{1,j}^{b_{1,j}}$ in G_1 and returns $b'_{1,j} = 0$ if it is a member of that group, and $b'_{1,j} = 1$ otherwise. So U_i can distinguish $M_{1,j}^0$ from $M_{1,j}^1$ with probability exactly 1. □

Lemma 2. *In round 2 of the information exchange phase, the following holds for an honest participant* U_i *and for all* $j(\neq i)$*:*

$$Pr[U_i(Trans, M_{2,j}^0) = 0] - Pr[U_i(Trans, M_{2,j}^1) = 0] = 1.$$

In other words, U_i *is able to differentiate between* $M_{2,j}^0$ *and* $M_{2,j}^1$ *with probability equal to 1.*

Proof. Suppose that U_i wants to determine the value of $b_{2,j}$ by observing $M_{2,j}^{b_{2,j}}$ after executing round 2 of the information exchange phase. We show that $M_{2,j}^{b_{2,j}}$ satisfies Equation (1) if and only if it is the value produced honestly in round 2 by U_j.

First assume that $M_{2,j}^{b_{2,j}}$ satisfies Equation (1). Then we have: $e(M_{2,j}^{b_{2,j}}, P) = e(P_{j+1} - P_{j-1}, P_j) = e(P_{j+1} - P_{j-1}, k_jP) = e(k_j(P_{j+1} - P_{j-1}), P)$. Therefore, we have two possibilities: either $M_{2,j}^{b_{2,j}} = k_j(P_{j+1} - P_{j-1}) = Y_j$ or there exists α such that $M_{2,j}^{b_{2,j}} = Y_j + \alpha$ where $e(\alpha, P) = 1$. But as mentioned in Section 2, $e(\alpha, P) = 1$ if and only if α is the identity element of G_1. So $M_{2,j}^{b_{2,j}} = Y_j + identity = Y_j$. It means that $M_{2,j}^{b_{2,j}}$ is the value produced honestly in round 2 by U_j.

Now, suppose that $M_{2,j}^{b_{2,j}}$ is the value produced honestly in round 2 by U_j, i.e., Y_j. Then we have: $M_{2,j}^{b_{2,j}} = Y_j = k_j(P_{j+1} - P_{j-1})$. So we have: $e(M_{2,j}^{b_{2,j}}, P) = e(k_j(P_{j+1} - P_{j-1}), P) = e(P_{j+1} - P_{j-1}, k_jP) = e(P_{j+1} - P_{j-1}, P_j)$, which means that $M_{2,j}^{b_{2,j}}$ satisfies Equation (1).

Therefore, it is enough for U_i to check if $M_{2,j}^{b_{2,j}}$ satisfies equation $e(M_{2,j}^{b_{2,j}}, P) = e(P_{j+1} - P_{j-1}, P_j)$ and if so returns $b'_{2,j} = 0$, otherwise $b'_{2,j} = 1$ is returned. So U_i can distinguish $M_{2,j}^0$ from $M_{2,j}^1$ with probability equal to 1. □

Theorem 1. *The proposed protocol (denoted by* Π*) is capable of identifying dishonest participants.*

Proof. Let U_i be a probabilistic polynomial-time honest user. Using the definition of experiment $GKE_{U_i,\Pi}^{disP}(k)$ we have:

$$\begin{aligned} &Pr[GKE_{U_i,\Pi}^{disP}(k) = 1] = \\ &Pr[(b_{1,j} = b'_{1,j} \forall j \neq i) \wedge (b_{2,j} = b'_{2,j} \forall j \neq i)]. \end{aligned}$$

So it's sufficient to prove that U_i's guesses are correct in round 1 and 2 of the information exchange phase. In other words, we should prove the following two claims:

Claim 1. *In round 1 of the information exchange phase, we have:* $\forall j \neq i : Pr[b_{1,j} = b'_{1,j}] = 1$.

Proof. Since $\forall j, r : Pr[b_{r,j} = 0] = Pr[b_{r,j} = 1]$, we have $\forall j \neq i$:

$$\begin{aligned} Pr[b_{1,j} = b'_{1,j}] &= \frac{1}{2}Pr[b'_{1,j} = 0 | b_{1,j} = 0] \\ &\quad +\frac{1}{2}Pr[b'_{1,j} = 1 | b_{1,j} = 1] \\ &= \frac{1}{2}Pr[U_i(Trans, M^0_{1,j}) = 0] \\ &\quad +\frac{1}{2}Pr[U_i(Trans, M^1_{1,j}) = 1] \\ &= \frac{1}{2}Pr[U_i(Trans, M^0_{1,j}) = 0] \\ &\quad +\frac{1}{2}(1 - Pr[U_i(Trans, M^1_{1,j}) = 0]) \\ &= \frac{1}{2} + \frac{1}{2}(Pr[U_i(Trans, M^0_{1,j}) = 0] \\ &\quad -Pr[U_i(Trans, M^1_{1,j}) = 0]) \\ &= \frac{1}{2} + \frac{1}{2}(1) \\ &= 1, \end{aligned}$$

where the last line follows from Lemma 1. □

Claim 2. *In round 2 of the information exchange phase, we have:* $\forall j \neq i : Pr[b_{2,j} = b'_{2,j}] = 1$.

Proof. Since $\forall j, r : Pr[b_{r,j} = 0] = Pr[b_{r,j} = 1]$, we have $\forall j \neq i$:

$$\begin{aligned} Pr[b_{2,j} = b'_{2,j}] &= \frac{1}{2}Pr[b'_{2,j} = 0 | b_{2,j} = 0] \\ &\quad +\frac{1}{2}Pr[b'_{2,j} = 1 | b_{2,j} = 1] \\ &= \frac{1}{2}Pr[U_i(Trans, M^0_{2,j}) = 0] \\ &\quad +\frac{1}{2}Pr[U_i(Trans, M^1_{2,j}) = 1] \\ &= \frac{1}{2}Pr[U_i(Trans, M^0_{2,j}) = 0] \\ &\quad +\frac{1}{2}(1 - Pr[U_i(Trans, M^1_{2,j}) = 0]) \\ &= \frac{1}{2} + \frac{1}{2}(Pr[U_i(Trans, M^0_{2,j}) = 0] \\ &\quad -Pr[U_i(Trans, M^1_{2,j}) = 0]) \\ &= \frac{1}{2} + \frac{1}{2}(1) \\ &= 1, \end{aligned}$$

where the last line follows from Lemma 2. □

Combining these two claims, we conclude that

$$Pr[GKE^{disP}_{U_i,\Pi}(k) = 1] = 1 \geq 1 - negl(k).$$

completing the proof. □

4.2 BCPQ-Security of the Proposed Scheme

In this section, we consider semantic security of the proposed protocol following BCPQ-model described in Section 2.4. We will show that the success probability of an attacker to learn anything about the shared key produced by the proposed scheme is negligible. Our proofs are inspired by [27].

Theorem 2. *The proposed protocol (denoted here by* Π*) is a secure group key exchange protocol in the sense of Definition 5.*

Proof. Let $\mathcal{A}$ be an adversary attacking the security of Π. We are going to show that his success probability is at most negligibly greater than $\frac{1}{2}$ in the game $Game^{ke}(\mathcal{A}, \Pi)$ defined in Section 2.4. So we claim that the value of the session key is indistinguishable for $\mathcal{A}$ from a random value in G_1.

First we assume that $\mathcal{A}$ eavesdrops on a single execution of the protocol. So he uses just one *Execute* query and he cannot use any *Reveal* query (because there is only one session key and it must be fresh for meaningful *Test* query). Since the participants have no LL-keys, there is no Corrupt query here. Thereupon, $\mathcal{A}$ can only use the knowledge of the transcript of this execution to output b'.

Consider the distribution $D = (Trans, sk)$, where $Trans = (\{P_i\}, \{Y_i\})$ is the transcript of an execution of the protocol and sk is the resulting session key. Let D' be another distribution in which (as in D) all the $\{P_i\}$ are uniformly distributed in G_1, but in which (in contrast to D) all the $\{Y_i\}$ are uniformly distributed in G_1 subject to the constraint $\sum_i Y_i = 0$. We have the following two claims:

Claim 3. *No efficient adversary can distinguish between the distributions* D *and* D'.

Proof. In order to show that the distributions D and D' are computationally indistinguishable, hybrid argument can be used. We do this by defining a sequence of n hybrid distributions in which one Y_i at a time is replaced with a random group element (subject to the above-mentioned constraint). Since adjacent distributions in this definition differ by only one Y_i, it is computationally easy to conclude their indistinguishability. Then since computational indistinguishability is transitive across a polynomial number of distributions, we conclude that D and D' are computationally indistinguishable. □

Claim 4. *In distribution* D'*, the value of the session key is uniformly distributed in* G_1*, independent of the value of the transcript.*

Proof. Let $c_{i,i+1} := k_i k_{i+1}$ for $1 \leq i \leq n$. Given $Trans = (\{P_i\}, \{Y_i\})$, the values $c_{1,2}, \cdots, c_{n,1}$ are constrained by the following n equations (only $n-1$ of which are linearly

Table 1: Computation cost of our protocol and the protocols proposed in [17, 22], and [37]

	Dutta and Barua [17]	**Huang et al. [22]**	**Zhao et al. [37]**	**Ours**
T_M	$2n-2$	$2n-2$	-	-
T_D	1	-	-	-
T_{exp}	3	$2n-1$	$3n-2$	-
T_A	-	-	-	$2n-1$
T_{SM}	-	-	-	$n+1$
T_{BP}	-	-	-	$2n-2$
T_H	-	-	n	-
T_{Sign}	2	1	-	2
T_{Vrfy}	$n+1$	$n-1$	-	$2n-2$
Total	$(2n-2)T_M+1T_D+3T_{exp}+2T_{Sign}+(n+1)T_{Vrfy}$	$(2n-2)T_M+(2n-1)T_{exp}+1T_{Sign}+(n-1)T_{Vrfy}$	$(3n-2)T_{exp}+nT_H$	$(2n-1)T_A+(n+1)T_{SM}+(2n-2)T_{BP}+2T_{Sign}+(2n-2)T_{Vrfy}$

independent):

$$\frac{1}{P}Y_1 = c_{1,2} - c_{n,1}$$

$$\vdots$$

$$\frac{1}{P}Y_n = c_{n,1} - c_{n-1,n}$$

Furthermore, $sk = (c_{1,2} + c_{2,3} + \cdots + c_{n,1})P$; equivalently, we have

$$\frac{1}{P}sk = c_{1,2} + c_{2,3} + \cdots + c_{n,1}.$$

Since this final equation is linearly independent from the set of equations above, sk is independent of $(\{P_i\}, \{Y_i\})$. This implies that even for a computationally-unbounded adversary $\mathcal{A}$, we have

$$Pr[(\{P_i\}, \{Y_i\}, sk_0) \leftarrow D^{'}; sk_1 \leftarrow G_1; b \leftarrow \{0,1\} : A(\{P_i\}, \{Y_i\}, sk_b) = b] = \frac{1}{2}.$$

□

The above two claims, prove the security of the protocol for an adversary making only a single Execute query. The case of multiple Execute queries can be dealt with using a straight forward hybrid argument. □

5 Performance Analysis and Comparison

In this section, we first consider the communication and computation cost of the proposed protocol. The analysis of the communication cost is done in terms of the number of messages sent by any single user. Note that we consider the communication cost in the point-to-point model in which each message sent to a different party is counted separately. Computation complexity analysis is also done in terms of the basic time-consuming operations. We then provide a comparison between our protocol and some other schemes in terms of computation cost, communication cost, and security properties. In order to evaluate the computation cost, we use the following notations:

T_M: Execution time for one multiplication operation in multiplicative groups;

T_D: Execution time for one division operation in multiplicative groups;

T_{exp}: Execution time for one exponentiation operation in multiplicative groups;

T_A: Execution time for one addition operation in elliptic curve groups;

T_{SM}: Execution time for one scalar multiplication operation in elliptic curve groups;

T_{BP}: Execution time for one bilinear pairing operation of two elements over an elliptic curve;

T_H: Computation time of a hash function;

T_{Sign}: Generation time of one signature;

T_{Vrfy}: Execution time for one signature verification.

Here, we first analyze the communication and computation cost of the proposed protocol with n participants. In round 1 and 2 of the information exchange phase, each user should send P_i and Y_i to $n-1$ other users. Therefore, $2n-2$ messages are communicated in total by each participant. Moreover, in round 1, one T_{SM} is required for computing P_i and in round 2, each participant requires $1T_A+1T_{SM}$ to compute Y_i. In the dishonest user elimination phase $(n-1)(2T_{BP}+1T_A)$ is needed as well. Finally, in the key computation phase, we require $(n-1)T_{SM}+(n-1)T_A$ to compute the common key. As a result, $(n+1)T_{SM}+(2n-2)T_{BP}+(2n-1)T_A$ is required for each participant in our protocol. Note that by applying the compiler of Katz and Yung to this protocol, an extra message has to be sent by each user to the others which can be sent along with the first broadcasted

Table 2: Comparison between our protocol and the protocols proposed in [17, 22], and [37]

	Dutta and Barua [17]	**Huang et al. [22]**	**Zhao et al. [37]**	**Ours**
Communication cost	$n+1$	$2n-2$	$2n-2$	$2n-2$
Computation cost	$(2n+728)T_M +$ $2T_{Sign}+(n+1)T_{Vrfy}$	$(482n-242)T_M +$ $1T_{Sign}+(n-1)T_{Vrfy}$	$(720.4n-480)T_M$	$(189.1n-125.9)T_M +$ $2T_{Sign}+(2n-2)T_{Vrfy}$
	$(352.9n+1618.7)T_M$	$(832.9n-323)T_M$	$(720.4n-480)T_M$	$(890.9n-287.9)T_M$
Formal proof of semantic security	Yes	No	No	Yes
Formal proof of security against dishonest participants	No	No	No	Yes
Detecting the presence of dishonest participants	No	Yes	Yes	Yes
Identifying dishonest participants	No	Yes	Yes	Yes

message. Moreover, one signature generation and $n-1$ signature verifications (for each of the two messages sent in round 1 and 2) are added to the computation complexity of the proposed protocol.

Now, we compare our protocol with those of Dutta and Barua [17], Huang et al. [22] and Zhao et al. [37]. The computation cost of these protocols is shown in Table 1. Table 2 lists the comparison between our protocol and group key exchange protocols [17, 22, 37] in terms of performance and security properties.

In order to make comparison more clear, we have used the relationship between the execution times of operations as in [5, 24, 28, 29]. We assume that $T_{exp} \cong 8.24T_{SM}$, $T_{exp} \cong 240T_M$, $T_{exp} \cong 600T_H$, $T_{exp} \cong 3.2T_{BP}$, $T_A \cong 5T_M$, and $T_D \cong 10T_M$. Besides, we have set the execution times of the signature and verification algorithms used in [14] as T_{Sign} and T_{Vrfy} respectively to simplify cost relations. The results are summarized in Table 2. According to Table 2, the scheme of Dutta and Barua outperforms the others in communication costs; however the detection and identification of dishonest participants are not achieved. The schemes of [22] and [37] both detect and identify dishonest participants but no formal proofs are provided. In our scheme, capabilities of the method against dishonest participants are formally proved. Moreover, detection and identification of dishonest participants is achieved at reasonable cost.

6 Conclusions

In this paper, elliptic curves are employed to propose an identity-based group key exchange protocol. It is proved the proposed protocol achieves security following the adversarial model of Bresson et al. The security of the protocol in the presence of dishonest participants is proved formally as well. The performance of the scheme in terms of computation cost, communication cost, and security properties is also considered.

Acknowledgments

The authors gratefully acknowledge the anonymous reviewers for their valuable comments.

References

[1] M. Bellare, R. Canetti, and H. Krawczyk, "A modular approach to the design and analysis of authentication and key exchange protocols," in *Proceedings of the 30th Annual ACM Symposium on the Theory of Computing*, pp. 419–428, New York, NY, USA, 1998.

[2] M. Bellare, D. Pointcheval, and P. Rogaway, "Authenticated key exchange secure against dictionary attacks," in *Proceedings of the 19th International Conference on Theory and Application of Cryptographic Techniques*, vol. 1807, pp. 139–155, Berlin Heidelberg, 2000.

[3] M. Bellare and P. Rogaway, "Entity authentication and key distribution," in *Proceedings of the 13th Annual International Cryptology Conference on Advances in Cryptology (Crypto'93)*, LNCS 773, pp. 232–249, Springer, 1993.

[4] M. Bellare and P. Rogaway, "Provably secure session key distribution - the three party case," in *Proceedings of the 27th ACM Symposium on the Theory of Computing*, pp. 57–66, New York, NY, USA, 1995.

[5] G. M. Bertoni, L. Breveglieri, L. Chen, P. Fragneto, K. A. Harrison, and G. Pelosi, "A pairing SW implementation for smart-cards," *Journal of Systems and Software*, vol. 81, no. 7, pp. 1240–1247, 2008.

[6] C. Boyd and J. M. G. Nieto, "Round-optimal contributory conference key agreement," in *Proceedings of the 6th International Workshop on Theory and Practice in Public Key Cryptography*, LNCS 2567, pp. 161–174, Springer, 2003.

[7] E. Bresson, O. Chevassut, and D. Pointcheval, "Dynamic group diffie-hellman key exchange under standard assumptions," in *Proceedings of the International Conference on the Theory and Applications of Cryptographic Techniques*, LNCS 2332, pp. 321–336, Springer, 2002.

[8] E. Bresson, O. Chevassut, D. Pointcheval, and J. J. Quisquater, "Provably authenticated group Diffie-Hellman key exchange," in *Proceedings of the 8th ACM Conference on Computer and Communications Security*, pp. 255–264, 2001.

[9] M. Burmester and Y. Desmedt, "A secure and efficient conference key distribution system," in *Proceedings of the Workshop on the Theory and Application of Cryptographic Techniques*, LNCS 950, pp. 275–286, Springer, 1995.

[10] R. Canetti and H. Krawczyk, "Analysis of key-exchange protocols and their use for building secure channels," in *Proceedings of the International Conference on the Theory and Application of Cryptographic Techniques*, LNCS 2045, pp. 453–474, Springer, 2001.

[11] X. Cao, W. Kou, and X. Du, "A pairing-free identity-based authenticated key agreement protocol with minimal message exchanges," *Information Sciences*, vol. 180, no. 15, pp. 2895–2903, 2010.

[12] S. Chatterjee and P. Sarkar, *Identity-Based Encryption*, Springer, 2011.

[13] K. Y. Choi, J. Y. Hwang, and D. H. Lee, "Id-based authenticated group key agreement secure against insider attacks," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E91-A, no. 7, pp. 1828–1830, 2008.

[14] S. Cui, P. Duan, C. W. Chan, and X. Cheng, "An efficient Identity-based signature scheme and its applications," *International Journal of Network Security*, vol. 5, no. 1, pp. 89–98, 2007.

[15] W. Diffie and M. Hellman, "New directions in cryptography," *IEEE Transactions on Information Theory*, vol. 22, no. 6, pp. 644–654, 1976.

[16] R. Dutta and R. Barua, "Password-Based encrypted group key agreement," *International Journal of Network Security*, vol. 3, no. 1, pp. 23–34, 2006.

[17] R. Dutta and R. Barua, "Provably secure constant round contributory group key agreement in dynamic setting," *IEEE Transactions on Information Theory*, vol. 54, no. 5, pp. 2007–2025, 2008.

[18] Z. Eslami and S. Kabiri Rad, "Another security weakness in an authenticated group key agreement," *Journal of Internet Technology*, vol. 11, no. 4, pp. 573–576, 2010.

[19] A. A. Hafez, A. Miri, and L. O. Barbosa, "Authenticated group key agreement protocols for Ad-hoc wireless networks," *International Journal of Network Security*, vol. 4, no. 1, pp. 90–98, 2007.

[20] M. Hietalahti, "A clustering-based group key agreement protocol for Ad-hoc networks," *Electronic Notes in Theoretical Computer Science*, vol. 192, no. 2, pp. 43–53, 2008.

[21] S. Hong, "Queue-based group key agreement protocol," *International Journal of Network Security*, vol. 9, no. 2, pp. 135–142, 2009.

[22] K. H. Huang, Y. F. Chung, H. H. Lee, F. Lai, and T. S. Chen, "A conference key agreement protocol with fault-tolerant capability," *Computer Standards & Interfaces*, vol. 31, no. 2, pp. 401–405, 2009.

[23] I. Ingemaresson, D. T. Tang, and C. K. Wong, "A conference key distribution system," *IEEE Transaction on Information Theory*, vol. 28, no. 5, pp. 714–720, 1982.

[24] W.-S. Juang, "Ro-cash: An efficient and practical recoverable pre-paid offline e-cash scheme using bilinear pairings," *The Journal of Systems and Software*, vol. 83, no. 4, pp. 638–645, 2010.

[25] M. Just and S. Vaudenay, "Authenticated multiparty key agreement," in *Proceedings of The International Conference on the Theory and Applications of Cryptology and Information Security*, pp. 36–49, London, UK, 1996.

[26] J. Katz and J. S. Shin, "Modeling insider attacks on group key-exchange protocols," in *Proceedings of the 12th ACM conference on Computer and communications security*, pp. 180–189, New York, NY, USA, 2005.

[27] J. Katz and M. Yung, "Scalable protocols for authenticated group key exchange," *Journal of Cryptology*, vol. 20, no. 1, pp. 85–113, 2007.

[28] K. Lauter, "The advantages of elliptic curve cryptography for wireless security," *IEEE Wireless Communications*, vol. 11, no. 1, pp. 62–67, 2004.

[29] A. Lenstra, E. Tromer, A. Shamir, W. Kortsmit, B. Dodson, J. Hughes, and P. Leyland, "Factoring estimates for a 1024-bit RSA modulus," in *Proceedings of the 9th International Conference on the Theory and Application of Cryptology and Information Security*, LNCS 2894, pp. 55–74, Springer, 2003.

[30] J. P. Lin and J. M. Fu, "Authenticated key agreement scheme with Privacy-Protection in the Three-party setting," *International Journal of Network Security*, vol. 15, no. 3, pp. 179–189, 2013.

[31] M. Manulis, *Survey on Security Requirements and Models for Group Key Exchange*, Technical Report TR-HGI-2006-002, Ruhr-Universität Bochum, Jan. 5, 2008.

[32] A. Shamir, "Identity-based cryptosystems and signature schemes," in *Proceedings of Advances in Cryptology (CRYPTO'84)*, LNCS 196, pp. 47–53, Springer, 1985.

[33] Z. Tan, "Efficient identity-based authenticated multiple key exchange protocol," *Computers & Electrical Engineering*, vol. 37, no. 2, pp. 191–198, 2011.

[34] W. G. Tzeng, "A secure fault-tolerant conference key agreement protocol," *IEEE Transactions on Computers*, vol. 80, no. 4, pp. 373–379, 2002.

[35] S. Wang, Z. Cao, and F. Cao, "Efficient Identity-based authenticated key agreement protocol with PKG forward secrecy," *International Journal of Network Security*, vol. 7, no. 2, pp. 181–186, 2008.

[36] Z. You and X. Xie, "A novel group key agreement protocol for wireless mesh network," *Computers & Electrical Engineering*, vol. 37, no. 2, pp. 218–239, 2011.

[37] J. Zhao, D. Gu, and Y. Li, "An efficient fault-tolerant group key agreement protocol," *Computer Communications*, vol. 33, no. 7, pp. 890–895, 2010.

Ziba Eslami received her B.S., M.S., and Ph.D. in Applied Mathematics from Tehran University in Iran. She received her Ph.D. in 2000. From 1991 to 2000, she was a resident researcher in the Institute for Studies in Theoretical Physics and Mathematics (IPM), Iran. During the academic years 2000-2003, she was a Post Doctoral Fellow in IPM. She served as a non-resident researcher at IPM during 2003-2005. Currently, she is associate professor in the Department of Computer Sciences at Shahid Beheshti University in Iran. Her research interests include design theory, combinatorial algorithms, cryptographic protocols, and steganography.

Mahnaz Noroozi received her B.S. degree in Computer Sciences in 2010 from Sharif University of Technology, Tehran, Iran. In 2012, she received her M.S. degree in Computer Sciences from Shahid Beheshti University, Tehran, Iran. She is currently doing research on cryptographic protocols and their security.

Saideh Kabiri Rad received the B.S. degree from Shahid Bahonar University, in Kerman and the M.S. degree from Shahid Beheshti University, in Tehran, both in Computer Science, Iran. She received the B.S. in 2007 and M.S. in 2010. She is currently doing research on information security and cryptography.

A Quantitative and Qualitative Analysis-based Security Risk Assessment for Multimedia Social Networks

Zhiyong Zhang[1], Lijun Yang[1], Hanman Li[1], and Fei Xiang[2]
(Corresponding author: Zhiyong Zhang)

Department of Computer Science, Information Engineering College, Henan University of Science and Technology[1]
Luoyang 471023, P. R. of China
(Email: z.zhang@ieee.org)
Department of Electronic Technology, Electrical Engineering College, Henan University of Science and Technology[2]

(Received Apr. 29, 2013; revised Dec. 13, 2013; accepted Jan 29, 2014)

Abstract

The emerging Multimedia Social Network (MSN) provides much more conveniences for the transmission and sharing of multimedia digital contents. However, the scenario on the distribution and spreading of copyrighted digital contents between users at will brings about a burning problem of Digital Rights Management (DRM). In addition, the open Internet and MSN platform are facing the security risks of digital contents copyrights infringements. The paper proposed a quantitative and qualitative-based risk analysis and assessment method, considering potential paths existence in MSN. Several risk impact factors was introduced, such as trust risk and user demands. Specifically, Value at Risk, a risk calculation method widely used in the financial field, as a quantitative analysis, was employed here. While an expert scoring sheet, as a qualitative approach, is used to evaluate non-quantifiable factors. Finally, the effectiveness of the security risk assessment method and related algorithm was verified by a well-designed experiment. We defined the size of the community followed by the "Rule of 150," and construct a random non-overlapped multimedia social network by using YouTube dataset. The experiment indicates that the relationships of risk loss with average rate of risk occurrence and risk preference of content providers are revealed.

Keywords: Digital rights management, multimedia social network, qualitative and quantitative analysis, risk assessment

1 Introduction

With the rapid development of network socialization, a large quantity of multimedia social network services and tools emerge, aiming at providing network tools, services and applications for transmission and sharing of the digital multimedia contents (such as digital images, audio and video, Java mobile applications, etc.) for MSN users. At present, the popular multimedia social networks throughout the world include Youtube, SongTaste, Youku, etc. These networks that are organized by users' social relationships are mainly used for using, sharing, and disseminating digital media content. They show obvious advantages in directly, quickly and flexibly transmitting digital contents. But, it also brings some risks for insecure transmission and uncontrollable sharing of the copyrighted digital contents. The unauthorized distribution, transmission and misuse of the digital contents make the problems of digital rights management increasingly prominent [14, 16, 17]. In the conditions of ubiquitous security flaws and malicious attacks, security risk management belongs to an effective way to ensure contents security and mitigate valuable digital asset risk. Therefore, the problem as to how to evaluate the transmission risk of digital rights under the multimedia social networks becomes our focus of concern. Further, DRM of multimedia social network will be better solved.

In recent years, the secure information spread and sharing in social networks, including multimedia social network and the special-purpose social network [7], has attracted much attention gradually. The research of DRM has two major technical paths: the preventive DRM technology and the reactive DRM technology. The preventive DRM technology is mainly based on the theory of cryptography and the usage control technology. The reactive one employs digital watermarking to protect digital contents and the corresponding copyrights. Whatever, for MSN scenario, the sharing strengthening of digital contents increases their exposure and leads to greater threat. Recently, some researchers focus on the information spread risk of social network, and the main research objective is to evaluate the risks of the unauthorized information access among users based on traditional access control policies. Different from the existing methods, this paper's main contribution is to integrate quantitative with qualitative approaches to evaluating the risks of potential digital rights distribution in multimedia social network scenario.

The remaining parts are arranged as below: firstly, Section 2 introduces the relevant researches on social network DRM and risk assessment. Section 3 describes background and theoretical knowledge involved in this article. Section 4 analyses the transmission risk of digital rights, and puts forward the risk assessment method by combining the qualitative and quantitative approaches. Section 5 provides Risk assessment algorithm. Section 6 makes an experiment by YouTube dataset and analyses the experimental results. Finally, Section 7 summarizes this research and shows the subsequent work.

2 Related Works

In recent years, some researchers have made extensive researches on social network platform and DRM. With regard to access control of digital contents for multimedia social networks, Barbara et al. [1] pointed out that the enhanced social network access control system is the first step to solve the existing security and privacy issues in online social networks. In order to resolve the current limitations, they proposed an expandable, fine-grain access control model based on semantic, web-online social networking. Sachan et al. [9] come up with an effective fine-grained access control model based on bit-vector transform. This model is able to convert certificates related to the digital contents into an effective structure. Security, storage and execution efficiency of this scheme are verified through mathematical and simulation experiment. Villegas [10] represented a personal data access control (PDAC) scheme. PDAC computes a "trusted distance" measure between users that is composed of the hop distance on the social network and an affine distance derived from experiential data. For online social network, Park et al. [8] proposed a user-activity-centric framework for access control, which determines four key control behaviors: attribute, policy, relation and session controls. This frame not only supports access control based on user relations, but also applies to common attribute-based access control.

In order to improve media content copyright protection and to diminish the illegal spread of media content in social networks, Lian et al. [6] proposed a content distribution and copyright authentication system based on the media index and watermarking technology. The results of the experiments confirmed that the system had strong robustness and stability. In addition, Chung et al. [3] proposed a novel video matching algorithm, as well as developed an intelligent copyright protection system based on this algorithm. Confirmed by experiments, the proposed algorithm can effectively conduct video matching; and the proposed system was suitable for copyright protection for video sharing networks. With the intention of solving the problem of content security in online social networks, Yeh et al. [13] proposed a security model based on multi-party authentication and key agreement. This proposed model can achieve user authentication between communities with a strong non-repudiation and flexibility. We proposed a MSN trust model based on small-world theory [18]. This model can effectively evaluate and dynamically update the value of trust between users, as well as identify malicious share users.

Unfortunately, with the development of MSN, DRM-enabling digital contents are suffering from huge risks owing to increasingly serious copyright infringement and misuse. More attention should be paid to security risk assessment for the social networks. A probability-based method to evaluate unauthorized access risk (UAR) was proposed in [2]. This scheme is capable of accurately computing the probability of information transmission in all connected paths between two users, and the method practicability has been proved by tests. In addition, Wang et al. [11] presented a statistical risk assessment method to quantify the threats in the networks. The information flow between two users in the social information network scenarios is thus evaluated. For a generic security risk assessment, Huang et al. [5] proposed a novel approach to addressing some implementation issues involved in employing such an information security risk assessment standard of ISO/IEC 27005:2011(E), and use the chlorine processing system in a water treatment plant as an example to well indicate the effectiveness of the proposed method. We have ever tried to highlight a multidisciplinary method for all-around examinations on risks to digital assets in the contents sharing scenario [15]. The method is a qualitative and quantitative fuzzy risk assessment, which is used for estimating a novel concept called Risk-Controlled Utility (RCU) in DRM. Then, we emphasize on an application case of the emerging trusted computing policy, and analyze the influences of different content sharing modes.

Summarily, there is a lack of risk assessment method for potential digital rights distribution in multimedia social network scenario, further better solved DRM issue.

3 Background Knowledge

Based on relations between users, the existing researches have mined potential transmission paths and credible potential paths for the multimedia social networks. Since the judgment of the credible potential paths is closely related to user-defined trust threshold, and the setting of the trust threshold has certain risk, the copyrighted digital contents transmitted and shared through the credible potential paths are still risky. For this reason, this article mainly carries out researches on risk assessment of digital right transmission via the credible potential paths. Thus, the relevant concepts of MSN potential paths are

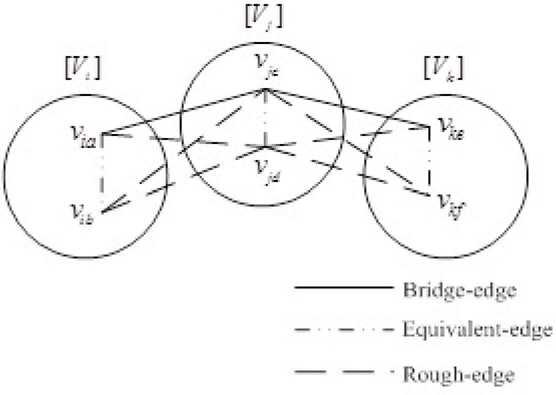


Figure 1: The potential paths in MSN

introduced in the following.

3.1 The Potential Paths in MSN

A social network consists of several locally dense "communities." In social networks, each community represents an actual social organization formed on the basis of social relationship or interest. That is, the node-node connection within community is relatively dense, but between which connections are very loose. A weak relation tends to transfer non-recurring information between different communities. Therefore, more alternation and information spread are performed between communities through weak relations, hence making it become an "information bridge" [12].

As shown in Figure 1, the MSN are divided into three communities, and the relation between users in the same community is very strong, whereas in different communities are relatively weak. Nodes in the same community have an equivalence relation. The connected edge of any nodes in the same community is called equivalent edge. The weak connection edge for connecting different communities is called the bridge edge. Through the weak relation between the communities, the rough relation will be produced, such as $< v_{ia}, v_{jd} >$, $< v_{jd}, v_{kf} >$ (as shown in the dotted line), what is called the rough edge. A path composed by continuous rough edges refers to a potential path (PP). So, in Figure 1, $< v_{ia}, v_{jd} >$, $< v_{jd}, v_{kf} >$ is a potential path from v_{ia} to v_{kf}. $< v_{ib}, v_{jd} >$, $< v_{jd}, v_{ke} >$ is the potential path in S from v_{ib} to v_{ke}; and $< v_{ib}, v_{jd} >$, $< v_{jd}, v_{kf} >$ is the potential path in S from v_{ib} to v_{kf}.

3.2 Trust Measurements of (Credible) Potential Paths

There is the direct trust between two users connected equivalent-edge and bridge-edge. So, the equivalent-edge trust calculation between communities is same to bridge-edge trust calculation, which can adopt a trust model for MSN in the reference [18]. Further, rough-edge trust (RT) can be obtained by integrating equivalent-edge trust (EDT) and bridge-edge trust (BDT). The EDT in a community is the direct trust between two nodes of connected equivalent edges in that community. Bridge-edge direct trust (BDT) is the direct trust between two users with a weak connected edge. In Figure 2, the trust value of the potential paths from W to U is shown as Equation (1).

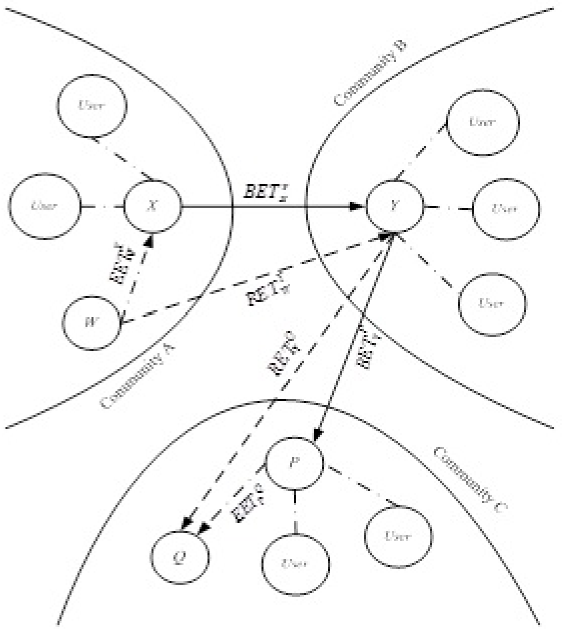


Figure 2: Potential path trust calculation

$$RT_w^u = EDT_w^v \cdot BDT_v^u. \quad (1)$$

The definition of potential paths indicates that all edges on the potential paths are rough-edges. The trust value of the potential paths (expressed as T_{pp}) is a product of all rough-edge trust values in the path, as shown in Equation (2):

$$T_{pp}(v_1, v_2, \cdots, v_n) = \Pi_{i=1}^{n} RT_{v_i}^{v_{i+1}} (i = 1, 2, \cdots, n), \quad (2)$$

where the relation between v_i and v_{i+1} $(i = 1, 2, \cdots, n)$ is the rough relation.

Figure 2 also shows the definition of potential path, in which a potential path exists from W to $T < W, U >< U, T >$. Moreover, Equation (2) indicates that the trust value of this potential path $T_{pp}(W, U, T) = RT_w^u \cdot RT_U^t$, where RT_w^u and RT_u^t are rough-edge trust values. The trust value of potential paths is described in Equation (1).

Based on the trust value of potential paths, we try to find the credible potential paths. The method is described as follows: First, trust value T_{pp} of the potential path is calculated. Second, the user defines a trust threshold, which is denoted by $T_{threshold}$. Finally, there is a comparison between T_{pp} and $T_{threshold}$.

Definition 1 (Credible Potential Path, CPP). *If $T_{pp_{v_1 \to v_n}} \geq T_{threshold}$, the potential path v_1 to v_n is called as a credible potential path $CPP_{v_1 \to v_n}$.*

4 Security Risk Assessment on Potential Digital Rights Distribution

In order to effectively control the risk of digital rights transmission on the credible potential paths, the main problems focus on identification, quantification and evaluation of transmission risk. Through effectively analyzing and calculating risk of digital content transmission on the potential paths of multimedia social networks, and evaluating possible loss brought by the risks, the rights are flexibly and safely shared and transmitted between users. Security of digital content transmission is enhanced for the multimedia social networks.

4.1 Quantitative and Qualitative Analytic Approach

In order to evaluate transmission risks of copyrighted digital contents in multimedia social network, this article adopts both quantitative and qualitative risk assessment approaches. In risk management, Annualized Loss Expectancy (ALE) is a common quantitative analysis tool used for computing an expected loss for an annual unit, and in general it includes the following elements:

- Asset Value (AV) denotes a tangible or intangible worth of digital assets by using monetary or other styles, and it is determined by the potential impact caused by the loss of assets.
- Annual Rate of Risk Occurrence (ARRO) is a prediction of how often a specific risk event is likely to happen each year.
- Exposure Factor (EF) indicates the impact of risks on a target system.

According to the transmission feature of digital rights among MSN users, we introduce Risk of Trust (RT), which refers to the transaction process by the trust relationship reflects the risk of the interactive event. With regard to such a risk severity factor as user demands for contents in DRM ecosystem, we introduce User Demand (UD). So, ALE is defined as Equation (3).

$$ALE = AV \cdot ARRO \cdot EF \cdot UD \cdot RT, \tag{3}$$

where, for main parameters of ALE, Asset Value is easily acquired and depicted by the monetary value of digital contents, ARRO is calculated by Poisson Distribution of the annual risk occurrence, EF and UD are yielded through the fuzzy assessments, and RT is quantified through the relation between trust and risk.

4.1.1 VaR-based Calculation on Maximum ARRO

Value at Risk (VaR) [4] denotes the maximum possible loss of certain assets value in normal fluctuations of market. The basic idea is to take advantage of the historical volatility information to infer future situation, and this inference refers to a probability distribution. VaR is an essential calculation method for estimating the maximum risk values based on a confidence degree $(1-\alpha)$ in a given time period, and it was defined as

$$Prob(L \leq VaR) = 1 - \alpha,$$

where L is an expected risk loss, VaR is the maximum loss, and α is determined by Content Providers' opinions on risks to a specific DRM ecosystem, that is,

$$\begin{cases} 0 \leq \alpha < 0.5 & \text{Risk-averse} \\ \alpha = 0.5 & \text{Risk-neutral} \\ 0.5 < \alpha \leq 1 & \text{Risk-seeking} \end{cases} \tag{4}$$

Taking it into consideration that the Poisson Distribution is a common probability function depicting the likelihood of random events occurrence, we attempted to employ the Poisson Distribution and VaR to calculate the $ARRO$, that is an estimation on the maximum occurrence rate of a random copyrights infringement/illicit usage event of digital contents. Thus, the maximum occurrence rate is in line with Poisson Distribution with the parameter λ. And then, by using Equation (5), the maximum value of $ARRO$ can be calculated.

$$Prob(x \leq MAX_{ARRO}) = 1 - \alpha. \tag{5}$$

In this case, a Multimedia Social Network has the specific annual occurrence rate of copyrights infringements threat, as is compliant to Poison Distribution with λ that denotes the average $ARRO$ of random risky events. According to Equations (4) and (5), when λ respectively is equal to 1.8, 5, 9, MAX_{ARRO} can be gained for three different Providers' opinions, that is,
When $\lambda = 1.8$,

$$MAX_{ARRO} = \begin{cases} 5 & \alpha = 0.03 & \text{Risk-averse} \\ 2 & \alpha = 0.5 & \text{Risk-neutral} \\ 1 & \alpha = 0.54 & \text{Risk-seeking} \end{cases} \tag{6}$$

When $\lambda = 5$,

$$MAX_{ARRO} = \begin{cases} 9 & \alpha = 0.03 & \text{Risk-averse} \\ 4 & \alpha = 0.5 & \text{Risk-neutral} \\ 2 & \alpha = 0.88 & \text{Risk-seeking} \end{cases} \tag{7}$$

When $\lambda = 9$,

$$MAX_{ARRO} = \begin{cases} 16 & \alpha = 0.01 & \text{Risk-averse} \\ 8 & \alpha = 0.5 & \text{Risk-neutral} \\ 3 & \alpha = 0.98 & \text{Risk-seeking} \end{cases} \tag{8}$$

Obviously, the maximum of $ARRO$ decreases with the increase of α.

4.1.2 Fuzzy Assessments on EF and UD by Using Triangular Fuzzy Number

With respect to two fundamental parameters EF and UD, we adopt the triangular fuzzy number-based subjection function to estimate risk factors influencing digital rights transmission. Besides, about fuzzy assessments of UD and EF, there were six judges participating in the risk assessments. In this case, we firstly presented the assessment scale and corresponding semantics of UD and EF, which is shown by Table 1. And, a group of assessment values for parameters UD and EF were given in Table 2.

Table 1: Five-level scale descriptions of UD and EF factors

Level	Scale	UD Description	EF Description
1	90	Strong	High
2	70	Medium to Strong	Medium to High
3	50	Medium	Medium
4	30	Weak to Medium	Low to Medium
5	10	Weak	Low

According to the fuzzy assessment method, UD and EF were calculated as follows:

- As UD is a single-factor assessment participated by six reviewers, $zeta_{s_i}(x) = (\sum_{t=1}^{6} \zeta_{s_i}(x^t))/6$, $s_i \in \{90, 70, 50, 30, 10\}$. According Table 2, the subjection degree vector of UD is $SD_{UD} = (0.333, 0.425, 0.167, 0.1, 0)$. In terms of the principle for the maximum subjection degree, the optimal SD is 0.425, and UD is 70.

- EF is a multi-factor fuzzy assessment procedure, and the final value of SD should consider two factors (Credible potential path length and the trust value of credible potential paths) and its weights, which are shown in Table 2. So we gained $SD_{EF} = (0.382, 0.286, 0.21, 0.123, 0)$. The optimal EF is 0.382, and EF is 90.

In the calculations of ALE, UD and EF was normalized. That is, UD is 0.7, and EF is 90.

4.2 Trust Risks of Credible Potential Paths

Digital content sharing between users under multimedia social networks is based on certain trust relation, which will directly affect sharing and transmission of digital contents. For multimedia social networks, trust is closely associated with the risk, i.e. the higher the trust value between two users, the smaller the risk in sharing content information. So trust values of two interacting parties can reflects the risks in interaction events. Other conditions being equal, the higher the trust, the lower the risk would be; otherwise, the higher the risk. So we can suppose that trust value plus value-at-risk is approximately equivalent to 1. Based on the relation between trust and risk, trust risk value RT of the credible potential paths is given by Equation (9):

$$RT = 1 - T_{pp} (0 \leq T_{pp} \leq 1). \tag{9}$$

5 Algorithm Design

Risk assessment process of digital right transmission for multimedia social network divides into the following steps: first, all potential paths between two user-nodes in different communities are identified, and then the trust values of the potential paths are calculated. The credible potential paths in the range of user-defined trust threshold are found. Finally, the quantitative and qualitative approaches are proposed in this article to evaluate the risks in the credible potential paths. The process of the algorithm is described as the follows Algorithm 1.

Algorithm 1 Mining and Risk Assessment of Credible Potential Paths between any Two Nodes in MSN

```
1:  Input: All the potential paths $PP_{i \to j}$ from $i$ to $j$; the number of share cycle $ShareNum$; the trust calculation window size $WindowSize$; feedback weight factor $Rate$; the trust threshold $T_{threshold}$; Asset Value $AV$; the value of Annual Rate of Risk Occurrence $ARRO$; the value of Exposure Factor $EF$; the value of User Demand $UD$.
2:  Output: The trust values $T_{PP_{i \to j}}$ of $PP_{i \to j}$, all the credible paths $CPP_{i \to j}$, and Annualized Loss Expectancy ($ALE$).
3:  Begin
4:  To calculate direct trust value between the nodes from the same equivalence community class based on $ShareNum$, $WindowSize$ and $Rate$;
5:  Based on the potential paths trust calculation method, trust value $T_{PP_{i \to j}}$ of all potential paths between $i$ and $j$ is computed;
6:  The obtained trust value $T_{PP_{i \to j}}$ from Step 4 is compared with the trust threshold $T_{threshold}$. Then, according Definition 1, $CPP_{i \to j}$ will be obtained;
7:  Return trust value $T_{PP_{i \to j}}$ and the credible potential paths $CPP_{i \to j}$.
8:  To calculate $ALE$ of the credible potential paths $CPP_{i \to j}$ according to Equation (3), here, $RT_{PP_{i \to j}}$ is equal to $(1 - T_{PP_{i \to j}})$;
9:  Return $ALE$ value of the credible potential paths $CPP_{i \to j}$;
10: End
```

6 Experiment and Analysis

In order to verify the effectiveness of the quantitative-and-qualitative-combined method for risk assessment, simulation experiment is made. The hardware of the simulation experiment is listed below: AMD Athlon(tm) X2 240 Processor 2.8G, 2G, and Microsoft Windows 7 ultimate. We made an experiment based on a representative real-world MSN YouTube dataset (http://socialnetworks.mpi-sws.org/data-imc2007.html), and further found a random multimedia social network with non-overlapped communities, as shown in Figure 3. Three sharing commu-

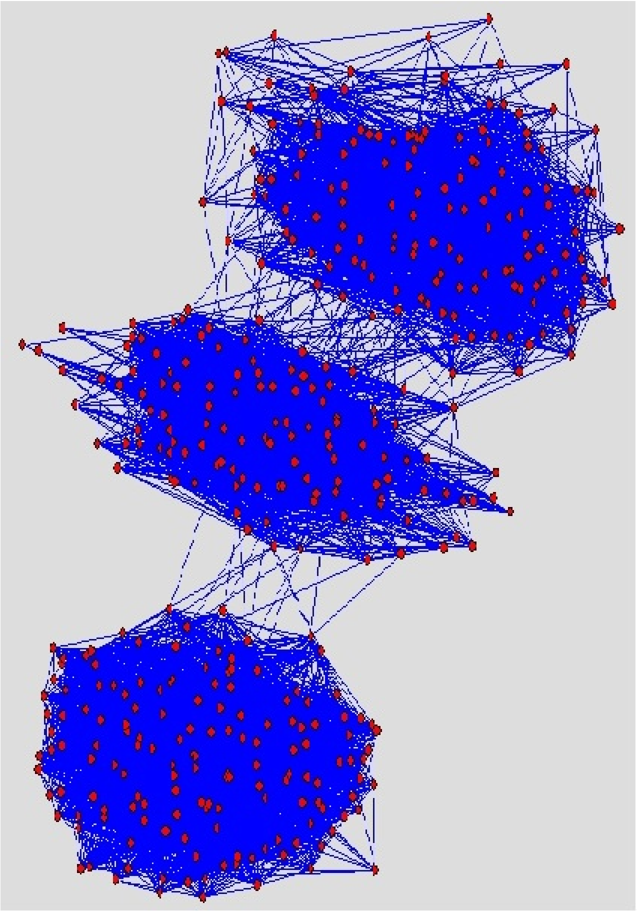


Figure 3: Random non-overlapped multimedia social network found by using YouTube

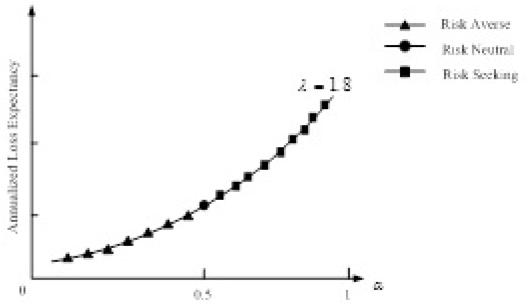


Figure 4: Effects of providers' risks opinions on ALE ($\lambda = 1.8$)

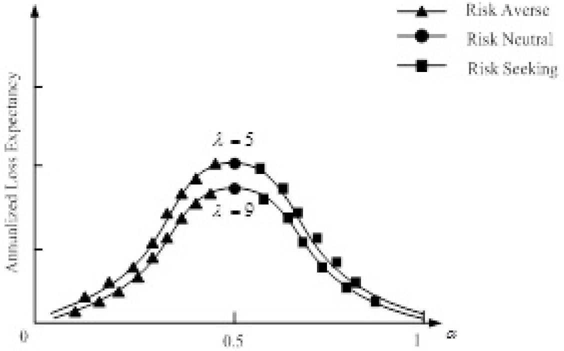


Figure 5: Effects of providers' risks opinions on ALE ($\lambda = 5$, $\lambda = 9$)

nities is involved with the random MSN, and they are written by C_1, C_2, and C_3, which are connected by some extra-community bridge edges called as weak ties compared with inner-community. We defined the size of the community followed by the "Rule of 150," which indicates that the node number of each sharing community is 150 or so for any user.

In Figure 3 networks, we accomplished the mining of the credible potential paths and their risk assessment. The experiment realizes that when two random nodes belong to different communities are input into MSN, all credible potential paths between these two nodes can be found, and risk loss expectation on the credible potential paths is evaluated.

Three groups of different data obtained from Equations (6), (7), and (8) are experimented. Tables 3, 4, and 5 show all credible potential paths from the starting point 123 to the end point 256 along with ALE experimental results. In the experiment, ALE is computed under VA=5000, EF=0.9 and UD=0.7. (Note: "-" in the table refers to it is not a credible potential path).

Table 3 show annual loss expectation ALE corresponding to three different risk preferences when $\lambda = 1.8$. The results show that in this conditions ALE will increases with α increasing.

Tables 4 and 5 show annual loss expectation ALE corresponding to three different risk preferences when $\lambda = 5$ and $\lambda = 9$, respectively. The results show that in these two conditions ALE will first increases and then decrease with α increasing. Based on the results listed by the tables above mentioned, we gained that ALE changing tendency as the following Figures 4 and 5, when λ is constant.

According to the above tables and figures, the experimental results are observed to be different. In the first case, when $\lambda = 1.8$, ALE will increases with α increasing. If the content providers are risk-averse, the loss is the lowest; if they are risk-neutral, the loss is larger; if they are risk-seeking, the loss is the maximum. In the second case, when $\lambda = 5$ and $\lambda = 9$, if the content providers are risk-averse, the loss is the lowest; if they are risk-seeking, the loss is larger; if they are risk-neutral, the loss is the maximum. In Poisson distribution, λ is average occurrence of random events within a limited time period. When the value of λ is higher, its probability distribution tends to show a standard normal distribution.

Thus, the relationships between risk loss and average risk occurrence, risk preference of content providers can be revealed: when the average risk occurrence is smaller, if the content providers are risk-averse, the loss is the

Table 2: Qualitative assessments

Target/Factor(s)	Assessment Scores						Weights
	J1	J2	J3	J4	J5	J6	
UD	82	58	73	38	95	75	-
Credible potential path length	88	39	79	50	69	85	0.6
The trust value of credible potential paths	77	84	92	30	81	52	0.4

Table 3: The credible potential paths and its ALE ($\lambda = 1.8$)

	The credible potential paths from point 123 to 256	ALE of credible potential paths ($\lambda = 1.8$)		
		$\alpha = 0.03$	$\alpha = 0.5$	$\alpha = 0.054$
1	$< 123, 341 >$, $< 341, 256 >$	47.2452	425.207	472.452
2	$< 123, 420 >$, $< 420, 256 >$	47.1301	424.171	471.301
3	$< 123, 381 >$, $< 381, 256 >$	47.0412	423.371	470.412
4	$< 123, 387 >$, $< 387, 256 >$	47.037	423.333	470.37
$\vdots$	$\dots$	$\dots$	$\dots$	$\dots$
129	$< 123, 383 >$, $< 383, 256 >$	35.0675	315.608	350.675
130	$< 123, 353 >$, $< 353, 256 >$	34.6403	311.763	346.403
131	$< 123, 377 >$, $< 377, 256 >$	32.1209	289.088	321.209
132	$< 123, 256 >$	16.5939	149.345	165.939

Table 4: The credible potential paths and its ALE ($\lambda = 5$)

	The credible potential paths from point 123 to 256	ALE of credible potential paths ($\lambda = 5$)		
		$\alpha = 0.03$	$\alpha = 0.5$	$\alpha = 0.88$
1	$< 123, 341 >$, $< 341, 256 >$	62.9936	283.471	125.987
2	$< 123, 420 >$, $< 420, 256 >$	62.8401	282.781	125.68
3	$< 123, 381 >$, $< 381, 256 >$	62.7216	282.247	125.443
4	$< 123, 387 >$, $< 387, 256 >$	62.716	282.222	125.432
$\vdots$	$\dots$	$\dots$	$\dots$	$\dots$
129	$< 123, 383 >$, $< 383, 256 >$	46.7567	210.405	93.5134
130	$< 123, 353 >$, $< 353, 256 >$	46.1871	207.842	92.3743
131	$< 123, 377 >$, $< 377, 256 >$	42.8279	192.726	85.6558
132	$< 123, 256 >$	22.1252	99.5633	44.2504

Table 5: The credible potential paths and its ALE ($\lambda = 9$)

	The credible potential paths from point 123 to 256	ALE of credible potential paths ($\lambda = 9$)		
		$\alpha = 0.01$	$\alpha = 0.5$	$\alpha = 0.98$
1	$< 123, 341 >$, $< 341, 256 >$	15.7484	204.729	31.4968
2	$< 123, 420 >$, $< 420, 256 >$	15.71	204.23	31.4201
3	$< 123, 381 >$, $< 381, 256 >$	15.6804	203.845	31.3608
4	$< 123, 387 >$, $< 387, 256 >$	15.679	203.827	31.358
$\vdots$	$\dots$	$\dots$	$\dots$	$\dots$
129	$< 123, 383 >$, $< 383, 256 >$	11.6892	151.959	23.3784
130	$< 123, 353 >$, $< 353, 256 >$	11.5468	150.108	23.0936
131	$< 123, 377 >$, $< 377, 256 >$	10.707	139.191	21.414
132	$< 123, 256 >$	5.53129	71.9068	11.0626

lowest; if they are risk-neutral, the loss is larger; if they are risk-seeking, the loss is the maximum. However, under a higher average risk occurrence, if the content providers are risk-averse, the loss is the lowest; if they are risk-seeking, the loss is larger; if they are risk-neutral, the loss is the maximum.

7 Conclusion

As transmission and sharing of digital content information has some potential threats due to openness and dynamic characteristics of the multimedia social networks, risk assessment for digital right transmission is an effective way to address security problems. This article mainly evaluates security risks in the credible potential paths for the multimedia social networks, and then proposes a risk assessment method based on combination of quantitative and qualitative approaches. Next, an algorithm is designed and later used to evaluate the risks in the credible potential paths through simulation experiment. The experimental results show that the average risk occurrence and risk preference of content providers jointly influence the risk-associated loss. In the following work, we will provide the specific security risk control model based on the risk assessment results, so as to reduce piracy and misuse risks of the digital contents protected by copyrights in the multimedia social networks.

Acknowledgments

The work was sponsored by National Natural Science Foundation of China (Grant No. 61370220, No.61003234), Program for Innovative Research Team (in Science and Technology) in University of Henan Province Grant No.15IRTSTHN010, Plan for Scientific Innovation Talent of Henan Province (Grant No. 134100510006), Key Program for Basic Research of The Education Department of Henan Province (Grant No.13A520240, No.14A520048). We would like to thank the reviewers for their valuable comments, questions, and suggestions.

References

[1] C. Barbara, F. Elena, H. Raymond, and K. Murat, "Semantic web-based social network access control", *Computers and Security*, vol. 30, no. 2, pp. 108–115, 2011.

[2] C. Barbara, F. Elena, M. Sandrol and T. Davide, "A probability-based approach to modeling the risk of unauthorized propagation of information in on-line social networks", in *Proceedings of the 1st ACM Conference on Data and Application Security and Privacy*, pp. 51–62, San Antonio, United States, Feb. 2011.

[3] M. B. Chung and I. J. Ko, "Intelligent copyright protection system using a matching video retrieval algorithm", *Multimedia Tools and Applications*, vol. 59, no. 1, pp. 383–401, 2012.

[4] M. A. H. Dempster, *Risk Management: Value at Risk and Beyond*, Cambridge: Cambridge University Press, 2002.

[5] C. C. Huang, K. J. Farn, and F. Y. S. Lin, "A study on implementations of information security risk assessment: Application to chlorine processing systems of water treatment plants", *International Journal of Network Security*, vol. 16, no. 4, pp. 241–248, 2014.

[6] S. Q. Lian, X. Chen, and J. W. Wang, "Content distribution and copyright authentication based on combined indexing and watermarking", *Multimedia Tools and Applications*, vol. 57, no. 1, pp. 49–66, 2012.

[7] M. J. H. Lim, M. Negnevitsky, and J. Hartnett, "Personality trait based simulation model of the e-mail system", *International Journal of Network Security*, vol. 3, no. 2, pp. 172–190, 2006.

[8] J. Park, R. Sandhu, and Y. Cheng, "A User-Activity-centric framework for access control in online social networks", *IEEE Internet Computing*, vol. 15, no. 5, pp. 62–65, 2011.

[9] A. Sachan, S. Emmanuel, and M. Kankanhalli, "An efficient access control method for multimedia social networks", in *Proceedings of the 2nd ACM SIGMM Workshop on Social Media*, pp. 33–38, Firenze, Italy, Oct. 2010.

[10] W. Villegas, *A Trust-Based Access Control Scheme for Social Networks*, Montreal: McGill University, 2008.

[11] T. Wang, M. Srivatsa, D. Agrawal, and L. Liu, "Modeling data flow in socio-information networks: A risk estimation approach", in *Proceedings of the 16th ACM Symposium on Access Control Models and Technologie*, pp. 113–122, Innsbruck, Austria, June 2011.

[12] L. J. Yang, Z. Y. Zhang, and J. X. Pu, "Rough set and trust assessment-based potential paths analysis and mining for multimedia social networks", *International Journal of Digital Content Technology & Its Applications*, vol. 6, no. 22, pp. 640–647, 2012.

[13] L. Y. Yeh, Y. L. Huang, A. D. Joseph, S. W. Shieh, and W. Tsaur, "A batch-authenticated and key agreement framework for P2P-Based online social networks", *IEEE Transactions on Vehicular Technology*, vol. 61, no. 4, pp. 1907–1924, 2012.

[14] Z. Y. Zhang, Q. Q. Pei, L. Yang and J. F. Ma, "Security and trust in digital rights management: A survey", *International Journal of Network Security*, vol. 9, no. 3, pp. 247–263, 2009.

[15] Z. Y. Zhang, S. G. Lian, Q. Q. Pei, and J. X. Pu, "Fuzzy risk assessments on security policies for digital rights management", *Neural Network World*, vol. 20, no. 3, pp. 265–284, 2010.

[16] Z. Y. Zhang, "Digital rights management ecosystem and its usage control: A survey", *International Journal of Digital Content Technology & Its Applications*, vol. 5, no. 3, pp. 255–272, 2011.

[17] Z. Y. Zhang, Security, *Trust and Risk in Digital Rights Management Ecosystem*, Beijing: Science Press, 2012.

[18] Z. Y. Zhang and K. L. Wang, "A trust model for multimedia social networks", *Social Networks Analysis and Mining*, vol. 3, no. 4, pp. 969–979, 2013.

Zhiyong Zhang born in 1975, earned his Master, Ph.D. degrees in Computer Science from Dalian University of Technology and Xidian University, China, respectively. He has ever been post-doctoral fellowship of School of Management at Xi'an Jiaotong University, China. He is currently full-professor with Department of Computer Science, Henan University of Science & Technology, and his research interests include digital rights management and multimedia social networks, trusted computing and access control, as well as security risk management and soft computing. Recent years, he has published over 80 scientific papers on the above research fields, held 5 authorized patents and drafted 2 China National Standards. Prof. Zhang is ACM/IEEE Senior Member, IEEE Systems, Man, Cybermetics Society Technical Committee on Soft Computing, World Federation on Soft Computing Young Researchers Committee. Besides, he is responsible for Topic Editor-in-Chief, Associate Editor and Guest Editor for several international journals, and Chair/Co-Chair and TPC Member for numerous international workshops/sessions.

Lijun Yang born in 1988, is a postgraduate majoring in Computer Science, at Department of Computer Science, Henan University of Science & Technology. Her research interests include digital rights management and rough set and soft computing, multimedia social networks and security risk assessment.

Hanman Li born in 1990, is a postgraduate majoring in Computer Science, at Department of Computer Science, Henan University of Science & Technology. Her research interests include multimedia social networks analysis, mining and simulation.

Fei Xiang born in 1980, received her Master degree in Theory and New Technology of Electronic Engineering from Huaqiao University, and earned her PhD. degree in Circuits and Systems at the Electronic and Information Engineering College, South China University of Technology. She is nowadays an associate professor at Electrical Engineering College, Henan University of Science & Technology. Her research interests include Chaos cryptography and its applications.

Reversible Data Hiding Based on Geometric Structure of Pixel Groups

Zhi-Hui Wang[1], Xu Zhuang[2], Chin-Chen Chang[3,4], Chuan Qin[4] and Yan Zhu[5]
(Corresponding author: Chin-Chen Chang)

School of Software, Dalian University of Technology[1]
Economy and Technology Development Area, Dalian 116620, China
Department of Computer Science and Technology, Southwest JiaoTong University[2]
Jinniu District, Chengdu, China, 610031
Department of Information Engineering and Computer Science, Feng Chia University[3]
Taichung, Taiwan, 40724
(Email: alan3c@gmail.com)
Department of Computer Science and Information Engineering, Asia University[4]
Taichung, Taiwan, 41354
Department of Computer Science and Technology, Southwest JiaoTong University[5]
Jinniu District, Chengdu 610031, China

(Received Sept. 23, 2013; revised and accepted Feb. 10 & May 10, 2014)

Abstract

Many reversible data hiding schemes have been developed in recent decades. Traditional schemes typically must deal with the problem of overflow and underflow, or transmission of hiding parameters. The new reversible data hiding scheme proposed in this paper is based on a combination of pixel groups' geometric structure and secret sharing mechanism. Experimental results confirm that the proposed scheme not only achieves the goal of hiding information without memorization of location map or any other parameter, but also generates very high quality stego-images with very large capacity of secret information embedded.

Keywords: Data hiding, geometric structure, reversible data embedding, watermarking

1 Introduction

Reversible data hiding conceals information (also called payload) into a cover image. Once an authentic user receives a stego-image containing the hidden information, the user can extract the hidden information exactly and recover the cover image without loss using a predefined procedure. Fridrich et al. [4] classified data hiding applications into two groups depending on the relationship between the hidden information and the cover image. The first group is the set of applications for which there is no relationship between the cover image and the hidden information. Both the coder and decoder are interested only in the hidden information. In the second group, the information has a close relationship to the cover image. For example, in digital watermarking, the hidden information usually is embedded into the cover image as a supplement to the cover image. In such cases, especially in medical, art, and military applications, two basic requirements must be met: (1) the receiver can extract the hidden information correctly and (2) the cover image can be recovered without distortion. The following three factors are generally used to evaluate a reversible data hiding scheme:

1) Payload capacity. Payload capacity is the maximum amount of information that can be embedded. The payload capacity is usually determined by the embedding algorithm and the size and content of the cover image. Generally, researchers use bits per pixel (bpp) to evaluate the payload capacity.

2) Imperceptibility. Imperceptibility means that people cannot perceive the existence of the hidden information with their eyes directly. For example, a meaningless image has no imperceptibility since it may readily attract a thief. In addition, if a coder adds information to the cover image directly, it is not imperceptible because of its expanded size. Thus, imperceptibility requires high similarity between the stego-image and the cover image, which means the stego-image has low distortion compared with the cover image. Generally, researchers use the peak signal to noise ratio (PSNR) to evaluate the degree of distortion of the stego-image versus the cover image.

3) Complexity of the algorithm. To indicate which pixels or pixel groups are used to embed information, many proposed schemes [5, 6, 7, 12] need a location map. Typically, these schemes use a compression technique to decrease the overhead of storing the location map. However, both the use of a location map and the compression technique increase the complexity of the algorithm.

Many data hiding schemes have been proposed in recent decades. Fridrich et al. [4] proposed the RS method for uncompressed image formats. The RS method is based on embedding messages in the status (regular or singular) of groups of pixels. However, as reported in [6], there are two drawbacks to the RS method. First, since the optimal size of a group is four pixels, the maximum bpp is only 0.25. In addition, due to the overhead of embedding the comprised RS-vector, the actual bpp is always smaller than the maximum bpp. Second, the RS method does not use information of neighboring groups so that it may lose useful information, which may increase its performance. Apart from these two drawbacks, the RS method also needs a lossless data compression technique, which increases the complexity of the algorithm. In 2003, Tian [12] proposed a difference expansion (DE) based algorithm. The DE method is based on the Haar wavelet transform, and it uses differences between two grayscale values in a pixel pair to embed information. Based on the experimental results shown in [12], Tian's algorithm can achieve very large hiding capacity. Kamstra et al. [6] developed an LSB predication embedding technique that has the same main idea described in [4]. Both the LSB predication embedding technique and the method described in [4] find a subset that is losslessly compressed from the cover image and embed the payload into the excess space after compression. Since it is difficult to construct a predication function, although [6] used another function to estimate the correctness of LSB predications, this method achieves only a very low embedding capacity. Kamstra et al. [6] also proposed an improved scheme based on Tian's method [12] in which two issues were considered: one is the capacity control problem and the other is the overhead costs caused by the location map. Due to Tian's excellent work [12], there are many variants based on the DE transform [1, 2, 3, 5, 7]. However, all these schemes must deal with the capacity problem and the overhead caused by the location map. In 2006, Ni et al. [9] proposed a reversible data hiding scheme with high PSNR (greater than 48dB) and considerable pure payload. Their algorithm is also light since there is not any transform operation such as DCT and DWT. Based on this work, Tai et al. [11] and Li et al. [8] presented improved methods that can achieve larger hiding capacity but keep embedding distortion low.

In this paper, we present a reversible data hiding method for digital images based on a combination of pixel groups' geometric structure and secret sharing mechanism [10]. The proposed method has four main merits: (1) it has meaningful stego-images, (2) the meaningful stego-images have high image quality, which means low distortion compared with the original cover image, (3) the original cover image can be recovered without loss, and (4) it can hide secrets with very large capacity without the use of a location map or any other extra information.

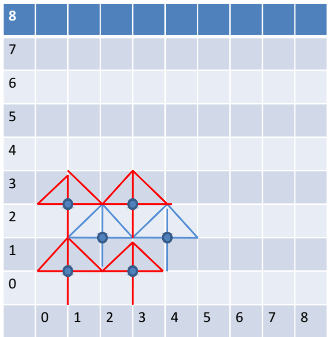


Figure 1: An example for illustrating data embedding rules

The rest of the paper is organized as follows. The proposed algorithm is presented in Section 2. Experimental results and further discussions are given in Section 3. Finally, conclusions are described in Section 4.

2 Proposed Algorithm

In this section, we provide an overview of our hiding algorithm, followed by a detailed presentation of the data embedding process and data extraction process.

2.1 Overview

Our algorithm uses a cover image to produce two stego-images (also called shadow images in secret sharing schemes) based on the to-be-embedded information and the cover image itself. We also call the to-be-embedded information a secret message in this paper. Any authentic user who holds these two shadow images at the same time can extract the secret message and recover the cover image without data loss. Next, we introduce the main idea of our method.

First, we create a two-dimensional pixel coordinate, as shown in Figure 1, and both the ordinate and abscissa of the pixel coordinate represent the grayscale value. Then we draw all the umbrellas in the pixel coordinate using a predefined method. Each umbrella has a center point called the embeddable point since just these points have the ability to embed a secret message without causing any confusion at the receiver end. We then create two copies of the cover image as the two shadows and pair all the pixels in these three images by a same pattern. Thus, each pair of the cover image can be located in the pixel coordinate as a point. In the data embedding process, if a pair of the cover image is located on an embeddable point, we change the corresponding pair of the shadows, referring to the embedding rules to embed a secret value. The data extraction is the inverse process of the data embedding. Next, we will describe the method for creating the pixel coordinate and drawing the umbrellas.

For convenience, we use a grayscale image with each pixel consisting of 3 bits to explain the procedure of building umbrellas. In Figure 1, we have drawn four red umbrellas and two blue umbrellas in the pixel coordinate, and each umbrella has a center point. These center points are (1, 1), (3, 1), (2, 2), (4, 2), (1, 3), and (3, 3). In total, there are 32 umbrellas in the pixel coordinate, and all of them can be drawn using the following steps:

1) Assuming the point that is being scanned is $(a,\ b)$, if $(a,\ b)$ is an edge point in the pixel coordinate, it does not have all four neighboring points, so it can't be the center point of an umbrella. Therefore, in the scanning process, we can ignore all the edge points and scan row by row, from point (1, 1) to point (6, 6) in this example.

2) If $(a,\ b)$ is not an edge point and not a point of any other umbrella, one can draw an umbrella by making $(a,\ b)$ the center point. Otherwise, one goes to step 3 without doing anything.

3) If there is a next point, get it and perform step 2 for it. Otherwise, end the process.

In our proposed scheme, every center point is an embeddable point that can be used to embed a secret message. We can use the bellowing method to decide whether a point is an embeddable point or not. For point $(i,\ j)$, each pixel consists of n bits:

1) If $i = 0$ or $j = 0$, $(i,\ j)$ is not an embeddable point.

2) If $i = 2^{n-1}$ or $j = 2^{n-1}$, $(i,\ j)$ is not an embeddable point.

3) If $(i,\ j)$ is not an edge point and i, j are odd numbers, $(i,\ j)$ is an embeddable point.

4) If $(i,\ j)$ is not an edge point and i, j are even numbers, $(i,\ j)$ is an embeddable point.

2.2 Data Embedding

In the data embedding process, we use all the embeddable points to embed the secret message. To improve the hiding capacity, we translate the to-be-embedded information to a seventeen-ary message, which means each digit value belongs to [0-16]. However, for convenience, we always use the hexadecimal system for computations in this paper.

An embeddable point $(a,\ b)$ is a center point of an umbrella, shown in Figure 2, and there are four neighboring points (point $(a\text{-}1,\ b)$, $(a,\ b\text{-}1)$, $(a\text{+}1,\ b)$, and $(a,\ b\text{+}1)$)

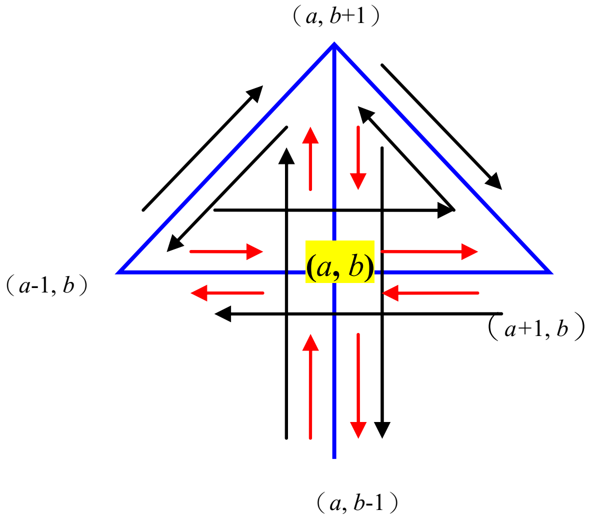


Figure 2: The proposed scheme

Table 1: Data embedding rules

Secret	EP	PPFS	PPSS
0	$(a,\ b)$	$(a,\ b)$	$(a,\ b)$
1	$(a,\ b)$	$(a\text{-}1,\ b)$	$(a,\ b\text{+}1)$
2	$(a,\ b)$	$(a,\ b\text{+}1)$	$(a\text{-}1,\ b)$
3	$(a,\ b)$	$(a\text{-}1,\ b)$	$(a,\ b)$
4	$(a,\ b)$	$(a,\ b)$	$(a\text{-}1,\ b)$
5	$(a,\ b)$	$(a,\ b\text{+}1)$	$(a,\ b)$
6	$(a,\ b)$	$(a,\ b)$	$(a,\ b\text{+}1)$
7	$(a,\ b)$	$(a,\ b\text{+}1)$	$(a,\ b\text{-}1)$
8	$(a,\ b)$	$(a,\ b\text{-}1)$	$(a,\ b\text{+}1)$
9	$(a,\ b)$	$(a,\ b\text{+}1)$	$(a\text{+}1,\ b)$
10	$(a,\ b)$	$(a\text{+}1,\ b)$	$(a,\ b\text{+}1)$
11	$(a,\ b)$	$(a\text{+}1,\ b)$	$(a,\ b)$
12	$(a,\ b)$	$(a,\ b)$	$(a\text{+}1,\ b)$
13	$(a,\ b)$	$(a\text{+}1,\ b)$	$(a\text{-}1,\ b)$
14	$(a,\ b)$	$(a\text{-}1,\ b)$	$(a\text{+}1,\ b)$
15	$(a,\ b)$	$(a,\ b\text{-}1)$	$(a,\ b)$
16	$(a,\ b)$	$(a,\ b)$	$(a,\ b\text{-}1)$

EP:Embeddable point
PFS:Pixel pair of first shadow
PSS:Pixel pair of second shadow

surrounding it. Thus, we can draw all arrows shown in Figure 2. That means there should be two different direction arrows between any two points linked by a blue line, e.g. (a-1, b) → (a, b), (a, b) → (a-1, b), (a, b+1) → (a, b-1), (a, b-1) → (a, b+1), and so on. We call the pixel pairs that determine the arrows as arrow pairs, e.g., ((a-1, b), (a, b)) is an arrow pair and ((a, b), (a-1, b)) is another arrow pair. Obviously, there are 16 arrow pairs in Figure 2 and we use these pairs to embed the secret message.

For an 8-bit grayscale cover image L and its two shadows $L1$ and $L2$, we first pair all the pixels in the cover image. For convenience, we can pair the two neighboring points row by row or column by column directly. If the size of the cover image is M×N, there are (M×N)/2 pixel pairs after pairing. Note that in this process, since the two shadows are copies of the cover image, all their pixels should be paired in the same way as the cover image. We use the following notations to represent the resultant sets of pairs for the cover image and its two shadows:

For cover image, PS={P_1, P_2,..., P_n}.

For the first shadow, PS_1={ P'_1, P'_2,..., P'_n}.

For the second shadow, PS_2={ P''_1, P''_2, ..., P''_n}.

Next, we scan PS, PS_1, and PS_2 simultaneously using the same scanning pattern, e.g., from the first element to the last sequentially. If a pixel pair in PS is an embeddable point, we embed a secret value into it by changing the corresponding pixel pairs of the two shadows in PS_1 and PS_2 according to the embedding rules. Since we have translated the binary secret message to a seventeen-ary message, we can use an embeddable point to embed a seventeen-ary value (0-16). Without losing generality, we use Table 1 to illustrate the embedding rules for the embeddable point (a, b). That means the pixel pairs of a certain arrow pair will be used to replace the corresponding original pixel pairs of these two shadows. In addition, the secret value 0 is mapped to the center point.

Now, we give an example of the data embedding process. We select the embeddable point (2, 2) in the pixel coordinate shown in Figure 1 to explain how to embed the secret message by using these umbrellas. Assuming the point (a, b) shown in Figure 2 is (2, 2), we can compute the four points surrounding (2, 2). They are (1, 2), (2, 1), (3, 2), and (2, 3). Then we use the embedding rules described in Table 1 to map a secret value to an arrow pair based on the value (0-16) of the secret itself. The embedding results are:

1) If secret = 0, (a1, b1) = (2, 2) and (a2, b2) = (2, 2);

2) If secret = 1, (a1, b1) = (1, 2) and (a2, b2) = (2, 3);

3) If secret = 2, (a1, b1) = (2, 3) and (a2, b2) = (1, 2);

4) If secret = 3, (a1, b1) = (1, 2) and (a2, b2) = (2, 2);

5) If secret = 4, (a1, b1) = (2, 2) and (a2, b2) = (1, 2);

6) If secret = 5, (a1, b1) = (2, 3) and (a2, b2) = (2, 2);

7) If secret = 6, (a1, b1) = (2, 2) and (a2, b2) = (2, 3);

8) If secret = 7, (a1, b1) = (2, 3) and (a2, b2) = (2, 1);

9) If secret = 8, (a1, b1) = (2, 1) and (a2, b2) = (2, 3);

10) If secret = 9, (a1, b1) = (2, 3) and (a2, b2) = (3, 2);

11) If secret = 10, (a1, b1) = (3, 2) and (a2, b2) = (2, 3);

12) If secret = 11, (a1, b1) = (3, 2) and (a2, b2) = (2, 2);

13) If secret = 12, (a1, b1) = (2, 2) and (a2, b2) = (3, 2);

14) If secret = 13, (a1, b1) = (3, 2) and (a2, b2) = (1, 2);

15) If secret = 14, (a1, b1) = (1, 2) and (a2, b2) = (3, 2);

16) If secret = 15, (a1, b1) = (2, 1) and (a2, b2) = (2, 2);

17) If secret = 16, (a1, b1) = (2, 2) and (a2, b2) = (2, 1).

By this way, we map a secret to a certain arrow pair based on the value of the secret itself. After we use all the embeddable points of a cover image to embed the secret messages, two shadows that have high similarity with the cover image are created. It is easy to extend this method to the grayscale image with each pixel consisting of 8 bits. That means the pixel coordinate in Figure 1 has the size of 255×255.

Table 2: Data extraction rules

(a1,b1)-(a2,b2)	Secret	Recovering
(0,0)	0	((a1,b1)=((a1,b1)
(-1,-1)	1	((a1,b1)=((a1+1,b1)
(1,1)	2	((a1,b1)=((a1,b1-1)
(-1,0) and (a2,b2)*	3	((a1,b1)=((a2,b2)
(1,0) and (a1,b1)*	4	((a1,b1)=((a1,b1)
(0,1) and (a2,b2)*	5	((a1,b1)=((a2,b2)
(0,-1) and (a1,b1)*	6	((a1,b1)=((a1,b1)
(0,2)	7	((a1,b1)=((a1,b1-1)
(0,-2)	8	((a1,b1)=((a1,b1+1)
(-1,1)	9	((a1,b1)=((a1,b1-1)
(1,-1)	10	((a1,b1)=((a1-1,b1)
(1,0) and (a2,b2)*	11	((a1,b1)=((a2,b2)
(-1,0) and (a1,b1)*	12	((a1,b1)=((a1,b1)
(2,0)	13	((a1,b1)=((a1-1,b1)
(-2,0)	14	((a1,b1)=((a1+1,b1)
(0,-1) and (a2,b2)*	15	((a1,b1)=((a2,b2)
(0,1) and (a1,b1)*	16	((a1,b1)=((a1,b1)

The notation (a,b)* represents that (a,b) is an embeddable point

2.3 Data Extraction

Once an authentic user receives two shadows, he/she can extract the embedded secret message and recover the cover image without loss using the following steps:

1) Pair all the pixels in these two shadows $L1$ and $L2$, respectively, according to the same pairing rule used in the data embedding process. This step produces two sets of pairs PS_1 (for $L1$) and PS_2 (for $L2$).

Lena Lena's shadow1 Lena's shadow2

Figure 4: Cover image Lena and its two shadows

2) ScanPS_1 and PS_2 in the same order. To recover the cover image, we need to recover one of these two shadows. For convenience, we always use the first shadow for the recovery. We use the extraction rules shown in Table 2 to extract the secret value and perform a recovering operation for the first shadow.

After the above two steps, an authentic user can extract all the embedded information correctly and reconstruct the cover image exactly.

We use an example to clarify these steps. Assuming that the recevier is scanning the first pixel pairs in the two stego-images (denoted as $L1$ and $L2$), he gets ($a1$,$b1$) for $L1$ and ($a2$,$b2$) for $L2$. Then he performs a difference operation on these two pairs and gets the result ($a1$ - $a2$, $b1$- $b2$). Next he extracts the secret value and recovers the original image using rules shown in Table 2. For exampe, ($a1$,$b1$) = (1,2) and ($a2$,$b2$) = (2,2), the recevier gets ($a1$ - $a2$, $b1$- $b2$)=(-1, 0). From Table 2, he finds that he must know whether ($a1$,$b1$) is an embeddable point or ($a2$,$b2$) is an an embeddable point to decide the secret value is 3 or 12. Based on the discussion in Section 2.1, ($a2$,$b2$) is an an embeddable point so that the embedded secret is 3. The correspoding pixel pair of cover image is ($a2$,$b2$). Because recevier uses the first shadow image to recover the original image, he performs ($a1$,$b1$) = ($a2$,$b2$). Then, he scans next pairs in the same way as he done for the first pairs. Finally, he extracs all the secret values and the first stego image is the cover image after performing all recovery operations.

3 Experimental Results and Discussions

We used 12 standard cover images sized 512×512 pixels for the experiments. In our experiments, all pixels pairs are paired derectly row by row. Table 3 shows the experimental results and Figure 3 shows all the cover images; Figure 4 shows Lena and its shadows. From Table 3, we can see that using the method proposed in the paper, all of these cover images have very large hiding capacity but with low distortion shadows. All PSNR of shadows shown in Table 3 are greater than 51dB and the smallest payload size is also greater than 256kb. The comparison with some other schemes using Lena is shown in Table 4. In Table 4, we compare the payload size of some others methods when they achieve their maximum PSNR. We use this method to compare our result because our scheme always reach a very high PSNR so that we don't have any data to compare with other schemes with relative low PSNR. The detailed analyses of the experimental results are presented below.

Table 3: Experimental results

Images	PSNR 1(db)	PSNR 2(db)	Payload(bits)
Lena	52.05	52.07	524288
Baboon	51.48	51.50	524216
Airplane	54.82	54.81	262172
Goldhill	51.94	51.96	524288
Boat	55.12	55.12	261760
Girl	51.41	51.42	348120
Woman	55.63	55.62	262964
Crowd	55.05	55.05	300240
Lake	55.12	55.11	262192
Barbara	55.56	55.55	263312
Bridge	53.32	53.33	449368
Couple	52.62	52.63	523752

PSNR 1: PSNR for fisrt shadow
PSNR 2: PSNR for second shadow

Table 4: Comparison with other schemes of cover image Lena with size 512×512×8. In the comparison, the payload of our method is divided by two because two stego-images are used in our method.

Schemes	Payload(bits)	PSNR
Tian's scheme [12]	39,566	44.20
Ni et al.' scheme [5]	5460	48.2
Tai et al.'s scheme [11]	24,377	48.35
Li et al.'s scheme APD$_1$ [8]	32,995	50.82
Li et al.'s scheme APD$_2$ [8]	60,785	48.40
Our scheme	262,144	52.05

3.1 Imperceptibility

Since we use two shadows having high similarity with the cover image to embed the secret message, it is difficult for anyone to perceive the embedded information in these two shadows with his/her eyes directly. As shown in Table 3, the lowest PSNR is 51.41dB among all the shadows, and it is already very high compared with other methods [1-4, 10].

Consider an extreme situation where all the pixel pairs of the cover image are embeddable points. In the worst case that each grayscale value of all pixels of the two shadows will be incremented or decremented by 1, the most mean square error (MMSE) is equal to 1. Therefore, the PSNR of each shadow versus the original cover image is:

$$PSNR = 10 \times \log_{10}(\frac{255 \times 255}{2}) = 48.13db. \quad (1)$$

The low limitation of PSNR of our method is already very high, and the experimental results show that the

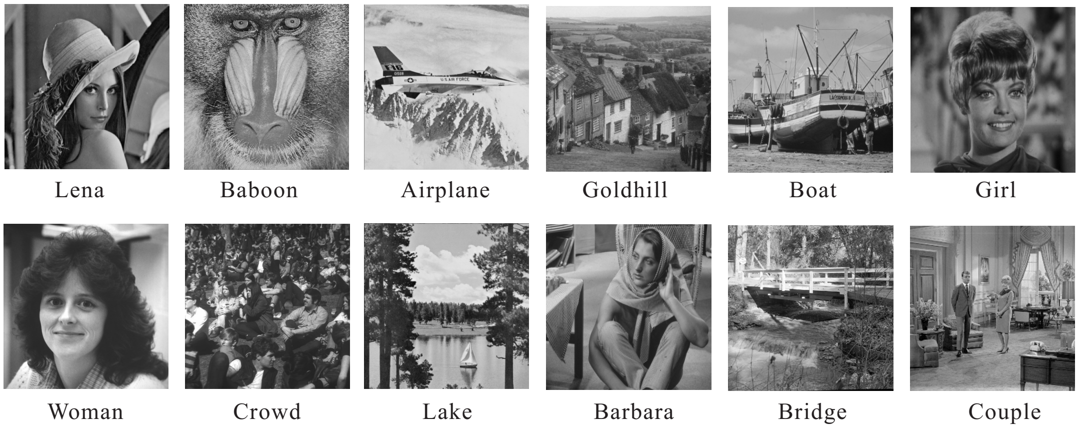


Figure 3: All cover images

actual PSNR is usually greater than the low limitation for 3dB-7dB. Thus, the two shadows produced by our algorithm have high performance in imperceptibility.

3.2 Capacity

For convenience, we think of the binary secret message as transformed to hexadecimal instead of seventeen-ary. Thus, each pixel pair can embed four bits of information. The hiding capacity for each stego-image in the above extreme situation is equal to:

$$payload = 4 \times (\frac{512 \times 512}{2 \times 2}) = 262,144 \quad \text{(bits)}. \tag{2}$$

As you can see, in an ideal situation, our algorithm can achieve very high performance with both large capacity and high PSNR. The experimental results show that all the pixel pairs of the cover images Lena and Goldhill are embeddable points so that these two images can achieve the maximum hiding capacity in our method.

3.3 Complexity

Assuming the size of the cover image is M×N, recall that in the data embedding process we use two copies (shadows) of the cover image and then pair all the pixels in these three images using the same pairing rule. Actually, we can pair the cover image, assuming the resultant pair set is $P = \{P1,\ P2,\ ...,\ P\text{n}\}$, and then produce two copies $P1$ and $P2$ of the pair set P. In this way, we achieve the same purpose but the time complexity is reduced to $O(\text{MN})$.

For data extraction, we need to first pair the shadows and then scan the resultant pair sets simultaneously. The time complexity of pairing is $O(2\text{MN})$ and of scanning is also $O(2\text{MN})$.

Our proposed algorithm is very light since we don't use any complex transform operation (such as DWT and DCT) or any compression technique.

3.4 Capacity Control

We have assumed that all the embeddable points will be used to embed the secret value. Obviously, this is not practical and we provide a simple answer to solve this problem.

We can use header information to indicate how many embeddable points are used to embed the secret message. For a grayscale image sized M×N, the maximum number of embeddable points is (M×N)/2. If we use the first n embeddable points as the header information, for easy in calculation, we also use hexadecimal here; since each embeddable point can embed four bits, we can choose n for the inequality:

$$2^{4n} - 1 \geq (M \times N)/2. \tag{3}$$

For example, considering the grayscale image sized 512×512, the maximum number of embeddable points in this image is (512×512)/2 = 131072. Because $2^{4\times5} - 1 \geq 131072$, we select 5 as the value of n. That means if the size of the cover image is 512×512, the first five embeddable points in the cover image are always used to indicate the number of embeddable points that are used to embed information in the embedding process. Thus, for the data extraction process, the decoder needs to first extract the header information and decide how many embeddable points should be extracted.

3.5 Comparisons

Table 4 shows a simple comparison with the other schemes since these four papers present their experimental results for Lena clearly. Tian's DE method [12] can achieve high embedding capacity when the capacity is larger than 260,000 bits, but PSNR of the stego-image is lower than 30dB. In such case, the imperceptibility of the stego-image decreases. As shown Table 4, when Tian's method achieves a high PSNR, its payload suddenly drops. On the other hand, because of the requirement for a location map, it is not easy for Tians method [12] to deal with the capacity control problem, so this method is not capable of embedding small payloads with low distortions as described in [6]. The RS method [4] cannot achieve higher performance in both capacity and imperceptibility than our method since its maximum bpp is 0.25 due to the optimal size of a group is four. The variants of the DE method, such as [5, 6, 7], also need to face up to the capacity control problem. Although these schemes propose new methods to decrease the size of the location map and control the embedding capacity, they are complex compared with our method since they usually need a lossless data compression technique. Ni et al.'s method [9] and Li et al.'s [8] can achieve very high PSNR compared with many other schemes, but both the PSNR and the hiding capacity are lower than these in our proposed method. Based on the discussions and the experimental results shown in Table 4, our method can achieve higher performance in imperceptibility, embedding capacity, and algorithm complexity than all the other methods [1, 2, 3, 4, 5, 6, 7, 9, 11, 12] mentioned in this paper.

4 Conclusions

In this paper, we proposed a reversible data hiding algorithm based on a combination of pixel groups' geometric structure and secret sharing mechanism. Since our method does not need a location map to indicate which pixel pairs are used to embed the secret message and each pixel can be changed at most one value, our algorithm has high performance in both capacity and imperceptibility. As the embeddable points are not edge points in the pixel coordinate, there would not be any overflow or underflow in our scheme. The proposed scheme is also light since we do not use any image compression technique. Our experimental results confirm that the proposed reversible data hiding algorithm outperforms all the other algorithms mentioned in the paper.

Acknowledgments

This work was supported by the National Nature Science Foundation of China under Grant No. 61272374.

References

[1] A. M. Alattar, "Reversible watermark using difference expansion of triplets," in *Proceedings of International Conference on Image Processing*, vol. 1, pp. 501–504, 2003.

[2] A. M. Alattar, "Reversible watermark using difference expansion of a generalized integer transform," *IEEE Transactions on Image Process*, vol. 13, pp. 1147–1156, 2004.

[3] A. M. Alattar, "Reversible watermark using difference expansion of quads," in *Proceedings of IEEE International Conference on Acoustics*, vol. 3, pp. 337–380, 2004.

[4] J. Fridrich, M. Goljan, and R. Du, "Lossless data embeddingnew paradigm in digital watermarking," *Journal of Applied Signal Processing*, vol. 1, pp. 185–196, 2002.

[5] Y. J. Hu, H. K. Lee, and J. W. Li, "De-based reversible data hiding with improved overflow location map," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, pp. 250–260, 2009.

[6] L. Kamstra and H. J. A. M. Heijmans, "Reversible data embedding into images using wavelet techniques and sorting," *IEEE Transactions on Image Process*, vol. 14, pp. 2082–2090, 2005.

[7] H. J. Kim, Y. Q. Shi, J. Nam, and H. G. Choo, "A novel difference expansion transform for reversible data embedding," *IEEE Transactions on Information Forensics and Security*, vol. 3, pp. 456–465, 2008.

[8] Y. C. Li, C. M. Yeh, and C. C. Chang, "Data hiding based on the similarity between neighboring pixels with reversibility," *Digital Signal Processing*, vol. 20, pp. 1116–1128, 2010.

[9] Z. Ni, Y. Q. Shi, N. Ansari, and W. Su, "Reversible data hiding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, pp. 354–362, 2006.

[10] A. Shamir, "How to share a secret," *Communication of the ACM*, vol. 11, pp. 612–613, 1979.

[11] W. L. Tai, C. M. Yeh, and C. C. Chang, "Reversible data hiding based on histogram modification of pixel differences," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, pp. 906–910, 2009.

[12] J. Tian, "Reversible data embedding using a difference expansion," *IEEE Transactions on Circuits System Video Technology*, vol. 13, pp. 890–896, 2012.

Zhi-Hui Wang received the BS degree in software engineering in 2004 from the North Eastern University, Shenyang, China. She received her MS degree in software engineering in 2007 and the PhD degree in software and theory of computer in 2010, both from the Dalian University of Technology, Dalian, China. Since November 2011, she has been a visiting scholar of University of Washington. Her current research interests include information hiding and image compression.

Xu Zhuang received his B.S. degree in Computer Science from Southwest Jiaotong University (SWJTU), Chengdu, China, in 2006, and is currently pursuing the Ph.D. degree in the Software Engineering Laboratory in Southwest Jiaotong University (SWJTU), Chengdu, China. His research interests include data mining, data hiding and information security.

Chin-Chen Chang received his Ph.D. degree in computer engineering from National Chiao Tung University. His first degree is Bachelor of Science in Applied

Mathematics and master degree is Master of Science in computer and decision sciences. Both were awarded in National Tsing Hua University. Dr. Chang served in National Chung Cheng University from 1989 to 2005. His title is Chair Professor in Department of Information Engineering and Computer Science, Feng Chia University, from Feb. 2005. He is a Fellow of IEEE and a Fellow of IEE, UK. His research interests include database design, computer cryptography, image compression and data structures.

Chuan Qin received the B.S. and M.S. degrees in electronic engineering from Hefei University of Technology, Anhui, China, in 2002 and 2005, respectively, and the Ph.D. degree in signal and information processing from Shanghai University, Shanghai, China, in 2008. Since 2008, he has been with the faculty of the School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, where he is currently a Lecturer. He also has been with Feng Chia University at Taiwan as a Postdoctoral Researcher from July 2010 to June 2012. His research interests include image processing and multimedia security.

Yan Zhu received her B.S. and M.S. degrees in Computer Science from Southwest Jiaotong University (SWJTU), Chengdu, China, in 1986 and 1989, respectively. She received her Ph.D. degree in Computer Science from Darmstadt University of Technology, Germany in 2004. Yan Zhu is currently a professor of the School of Information Science and Technology, SWJTU and the director of the Laboratory of Software Engineering. Her research interests include data mining, Web information security, and Web spam detection.

Construction of Extended Multivariate Public Key Cryptosystems

Shuaiting Qiao, Wenbao Han, Yifa Li, and Luyao Jiao
(Corresponding author: Shuaiting Qiao)

Zhengzhou Information Science and Technology Institute
Zhengzhou, Henan Province, 450001, China
(Email: qsting2012@163.com)

(Received Sep. 03, 2013; revised and accepted Jan. 20 & Mar. 23, 2014)

Abstract

Based on the ideas: "invertible cycle", "tame transformation" and "special oil and vinegar", three different nonlinear invertible transformations were constructed separately. Then making use of the idea of the extended multivariate public key cryptosystem, and combining the nonlinear invertible transformations above with Matsumoto-Imai (MI) scheme, three methods of designing extended multivariate public key cryptosystem were proposed. Next, the corresponding encryption and signature algorithms were given. Analysis results demonstrate that the new extended cryptosystems inherit the merit of MI scheme, i.e., efficient computation. Meanwhile, the new extended cryptosystems can also resist the linearization attack, differential attack and algebraic attack.

Keywords: Extended multivariate public key cryptosystem, invertible cycle, matsumoto-imai scheme, special oil and vinegar, tame transformation

1 Introduction

The 21st century is the era of information. With the rapid development of electronic information science and technology, information security has become so important. After electronic information science and technology, quantum and other new information science are building up and developing [9]. But the development of quantum computers will pose a threat to the widely-used public key cryptosystems, which are based on discrete logarithm problem and large integer factorization problem [10, 16]. Therefore, great attention has been paid to the post-quantum public key cryptography [2], and multivariate public key cryptosystems (MPKCs) develop rapidly in this background. MPKC is considered to be a candidate of secure cryptosystems in post-quantum era for its higher efficiency, better security and easy access to the hardware implementation, etc. During the last twenty years, MPKCs have received more and more attention.

The security of MPKCs depends on the difficulty of solving a set of nonlinear multivariate quadratic equations over a finite field [7] and the isomorphism of polynomials problem [17]. Its research began in the 1990s. According to different central maps, MPKCs have been divided into five schemes, which are Matsumoto-Imai (MI) scheme, Hidden Field Equation (HFE) scheme, Unbalanced Oil and Vinegar (UOV) scheme, Stepwise Triangular Systems (STS) scheme and Medium Field Equation (MFE) scheme [7]. Especially, in the past few years, many cryptosystems have emerged in sequence, such as the CyclicRainbow cryptosystem [14], the Double-Layer square cryptosystem [3], the Enhanced STS cryptosystem [19], etc, which make MPKCs develop and complete. Meanwhile, researchers have also applied MPKCs to identification [15], special signatures [22, 24]and other fields. So far, MPKCs have been a hot topic in cryptography.

In 1988, Matsumoto and Imai proposed MI scheme with high efficiency, which was seen as the first scheme of MPKCs [11]. In 1995, Patarin et al present linearization attack aimed at MI scheme [12]. To resist linearization attack, Jacques Patarin et al came up with the Flash cryptosystem in 2000 [13], and Ding Jin-tai et al put forward the PMI cryptosystem in 2004 [4], but both of them were vulnerable to differential attack [5, 8]. In 2011, by incorporating the hash authentication technique and traditional multivariate public key cryptography algorithm, Wang Hou-zhen et al proposed Extended Multivariate Public Key Cryptosystem (EMC), which can resist both linearization attack and differential attack [20]. The emerging of EMC pointed out a new idea to construct novel multivariate public key cryptosystems.

In this work, on the base of the ideas: invertible cycle [6], tame transformation [21] and special oil and vinegar [23], three different nonlinear invertible transformations are built separately. Then by combining these nonlinear invertible transformations above and MI scheme, three different methods to construct novel EMCs are proposed. Finally, the performance analysis and security analysis will be given.

The rest of this paper is organized as follows. In Section 2, we give a brief overview of general structure of MPKCs, EMC and MI scheme. Section 3 presents 3 the design of the novel EMCs. Section 4 gives the operation efficiency and security analysis of our proposed cryptosystems. Finally, Section 5 concludes this paper.

2 Preliminaries

2.1 General Structure of MPKCs

The trapdoor function of MPKCs is a set of nonlinear multivariate quadratic polynomials over a finite field, i.e.,$P =: \mathbb{F}^n \to \mathbb{F}^m$,

$$P = (p_1(x_1, \cdots, x_n), \cdots, p_n(x_1, \cdots, x_n)).$$

For $1 \leq j \leq k \leq n$, $1 \leq i \leq m$, each $p_i(x_1, \cdots, x_n)$ is organized as follows

$$y_i = p_i(x_1, \cdots, x_n) = \sum_{1\leq j\leq k\leq n} a_{ijk}x_jx_k + \sum_{j=1}^{n} b_{ij}x_j + c_i,$$

where $x_i \in \mathbb{F}_q, 1 \leq i \leq n$, and coefficients $a_{ijk}, b_{ij}, c_i \in \mathbb{F}_q$.

The construction of MPKCs is mainly based on the hardness of Multivariate Quadratic (MQ) problem and Isomorphism of Polynomials (IP) problem.

The trapdoor of MPKCs $P = (p_1(x_1, \cdots, x_n), \cdots, p_n(x_1, \cdots, x_n))$ is constructed as follows:

$$\begin{tikzcd} u=(u_1 , \cdots , u_n) \arrow[d, "S"] \\ x=(x_1 , \cdots , x_n) \arrow[d, "Q"] \\ y=(y_1 , \cdots , y_m) \arrow[d, "T"] \\ v=(v_1 , \cdots , v_m). \end{tikzcd}$$

The public key P consists of three maps, i.e.,$P = T \circ Q \circ S$, where $S : u \to x = M_S u + c_S$ and $T : y \to v = M_T y + c_T$ are random invertible affine maps in $\mathbb{F}^n$ and $\mathbb{F}^m$ respectively. They mask the structure of the central map together, and are important parts of the secret key.

2.2 EMC

In 2011, by combining the hash authentication technique with traditional multivariate public key algorithm, Wang Hou-zhen et al proposed Extended Multivariate Public Key Cryptosystem (EMC). It can be used as a signature scheme and an encryption scheme simultaneously. It was an essential expansion of traditional multivariate public key cryptography, and it improved security of the traditional MPKCs [20].

Tame transformation is used to construct EMC. Define tame transformation first.

Definition 1 (Tame Transformation)**.** *Tame transformation is a special mapping $G : GF(q)^n \to GF(q)^n$*

$$\begin{pmatrix} t_1 \\ t_2 \\ \vdots \\ t_{n-1} \\ t_n \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 + g_2(x_1) \\ \vdots \\ x_{n-1} + g_{n-2}(x_1, \cdots, x_{n-2}) \\ x_n + g_{n-1}(x_1, \cdots, x_{n-1}) \end{pmatrix}$$

where g_i are arbitrary quadratic polynomials. The mapping G is so special that it can be easily inverted.

Definition 2 (Hash-based Transformation)**.** *A Hash-based Transformation (HT) $L : \mathbb{F}_q^n \to \mathbb{F}_q^n$*

$$\left\{\begin{array}{l} \begin{pmatrix} y_1 \\ \vdots \\ y_{n-\delta} \end{pmatrix} = A \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_{n-\delta} \end{pmatrix} + \alpha_1 \\ \begin{pmatrix} y_{n-\delta+1} \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_{n-\delta+1} \\ \vdots \\ x_n \end{pmatrix} + D \cdot \begin{pmatrix} x_{n+1} \\ \vdots \\ x_{n+\delta} \end{pmatrix} + \\ \quad B \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_{n-\delta} \end{pmatrix} + \alpha_2 \end{array}\right.$$

where α_1, α_2 are $(n-\delta)$-dimension vector and δ-dimension vector respectively, A is a $(n-\delta) \times (n-\delta)$ matrix, and D must be a diagonal and full-rank $\delta \times \delta$ matrix; B is a random $\delta \times (n-\delta)$ matrix; all the coefficients are chosen over $\mathbb{F}_q$. The extended variables $x_{n+i}(1 \leq i \leq \delta)$ are defined by $x_{n+i} = H_k(x_1||x_2||\cdots||x_{n-\delta+i-1})$.

Known from the definition above, L can be seen as a compression mapping from $\mathbb{F}_q^{n+\delta}$ to $\mathbb{F}_q^n$, i.e.,$(y_1, \cdots, y_n) = L(x_1, \cdots, x_n, x_{n+1}, \cdots, x_{n+\delta})$.

The public key of the extended MQ cryptosystem is designed as follows:

$$P' = P \circ L = T \circ F \circ U \circ L = (p_1', \cdots, p_n').$$

The public key is a set of multivariate quadratic polynomials, which is a mapping from $\mathbb{F}_q^{n+\delta}$ to $\mathbb{F}_q^n$, and the corresponding secret key consists of L^{-1},U^{-1},T^{-1}, F^{-1}.

2.3 MI Scheme

In 1988, Matsumoto and Imai proposed the first multivariate public key cryptosystem, i.e., MI scheme [11].

Let $k = \mathbb{F}_q$ be a finite field of characteristic 2, where $q = 2^m$, and K be an extension field of degree n of K. Then define a standard K-linear isomorphism map $\Phi : K \to k^n$,$\Phi(a_0 + a_1x + \cdots + a_{n-1}x^{n-1}) = (a_0, a_1, \cdots, a_{n-1})$.Define $F : K \to K, F(X) = X^{1+q^\theta}$, where θ is an integer such that $1 \leq \theta \leq n$ and $gcd(1 + q^\theta, q^n - 1) = 1$. F is an invertible map and its inverse is given by $F^{-1}(X) = X^t$, where $t(1 + q^\theta) \equiv 1 mod(q^n - 1)$. Let $\overline{F} : k^n \to k^n$ be a central map:

$$\begin{aligned} \bar{F}(x_1 , \cdots , x_n) &= \phi \circ F \circ \phi^{-1}(x_1 , \cdots , x_n) \\ &= (\bar{F}_1 (x_1 , \cdots , x_n), \cdots , \bar{F}_n(x_1 , \cdots , x_n)) \end{aligned}$$

where $\overline{F}_i(x_1, \cdots, x_n)$ are quadratic polynomials of n variables.

Finally, let L_1 and L_2 be two randomly chosen invertible affine linear maps over k^n.

$\hat{F}(x_1, \cdots, x_n) = L_1 \circ \overline{F} \circ L_2(x_1, \cdots, x_n) = (\hat{F_1}(x_1, \cdots, x_n), \cdots, \hat{F_n}(x_1, \cdots, x_n))$ is the ciphertext suggested by MI scheme. The public key is $\hat{F}(x_1, \cdots, x_n)$, and the secret key is $(L_1^{-1}, L_2^{-1}, \theta)$.

3 Design of the Novel EMCs

Nowadays most algorithms of MPKCs cannot be a signature scheme and an encryption scheme simultaneously and most of them are under attack. How to construct a secure and efficient MPKC enabling both signature and encryption remain a hot topic and an open problem.

The key of our proposed cryptosystems is to build a nonlinear and invertible transformation L. By making use of the idea of EMC and incorporating MI scheme with nonlinear invertible transformations L, the novel EMCs are produced:

$$\begin{aligned} \tilde{F}(x_1\ ,\ \cdots\ ,\ x_n) &= \hat{F} \circ L(x_1\ ,\ \cdots\ ,\ x_n) \\ &= L_1 \circ \bar{F} \circ L_2 \circ L(x_1\ ,\ \cdots\ ,\ x_n) \quad (1) \end{aligned}$$

3.1 Construction of L

Constructing nonlinear and invertible transformations L is the key to design the novel EMCs. Three kinds of nonlinear and invertible transformations will be introduced based on different ideas below.

3.1.1 Construction of Invertible Transformation Based on "Invertible Cycle"

Assume the invertible transformation is L_3, in order to facilitate the inverse, L_3 is defined in cases. Suppose the order is n, to express properly the successor of $\{1, \cdots, n\}$, define

$$\mu : \{1, \cdots, n\} \to \{1, \cdots, n\} : \mu(i) = \begin{cases} 1 & \text{for } i = n \\ i+1 & \text{otherwise} \end{cases}$$

Lemma 1. *For a fixed integer $n \geq 2$, define a nonlinear transformation $L_3 : (x_1, \cdots, x_n) \to (t_1, \cdots, t_n)$ as follows:*

$$\begin{cases} t_1 := \begin{cases} c_1 x_1 x_2 & for\ n\ odd \\ c_1 {x_1}^q x_2 & for\ n\ even \end{cases} , \\ t_i := c_i x_i x_{\mu(i)} \quad for \quad 2 \leqslant i \leqslant n \end{cases}$$

Then the inverse image of $(t_1, \cdots, t_n)$,where $t_i \in \mathbb{F}_q^* :=$ $\mathbb{F}_q\backslash\{0\}$, $\forall i$ is given by

$$\begin{cases} x_1 := \begin{cases} \sqrt{\dfrac{\prod_{i=0}^{(n-1)/2} t_{2i+1}}{\prod_{i=1}^{(n-1)/2} t_{2i}}} \cdot \sqrt{\dfrac{\prod_{i=1}^{(n-1)/2} c_{2i}}{\prod_{i=0}^{(n-1)/2} c_{2i+1}}} & for\ n\ odd \\ \sqrt[q-1]{\dfrac{\prod_{i=0}^{n/2-1} t_{2i+1}}{\prod_{i=1}^{n/2} t_{2i}}} \cdot \sqrt[q-1]{\dfrac{\prod_{i=1}^{n/2} c_{2i}}{\prod_{i=0}^{n/2-1} c_{2i+1}}} & for\ n\ even \end{cases} \\ x_i := \dfrac{t_i}{c_i x_{\mu(i)}} \qquad for \quad i = n\ ,\ \cdots\ ,\ 2. \end{cases}$$

Remark 1. *However, for some $x_i = 0$ in the set $\{x_1, \cdots, x_n\}$, the definition of L_3 is the same as that of Lemma 1, so the conditions $t_i = 0$ and $t_{i+1} = 0$ are set up. Take odd n for example, the inverse of L_3, i.e., $(x_1, \cdots, x_n) = L_3^{-1}(t_1, \cdots, t_n)$ is organized as follows:*

Either $x_i = 0$ or $x_{\mu(i)} = 0$ is set up, where $t_i = 0$.

1) For $x_i = 0$, choose $x_{\mu(i)} = a \in \mathbb{F}_q\backslash\{0\}$ randomly, and other x_i can be worked out in sequence:

$$\begin{cases} x_{\mu(i)+1} = \dfrac{t_{\mu(i)}}{c_{\mu(i)}\ a}\ ,\ \cdots\ ,\ x_n = \dfrac{t_{n-1}}{c_{n-1}\ x_{n-1}} \quad for \quad i = 1, \\ x_{\mu(i)+1} = \dfrac{t_{\mu(i)}}{c_{\mu(i)}\ a}\ ,\ \cdots\ ,\ x_n = \dfrac{t_{n-1}}{c_{n-1}\ x_{n-1}}\ ,\ x_1 = \dfrac{t_n}{c_n\ x_n} \\ \qquad for \quad i = 2, \\ x_{\mu(i)+1} = \dfrac{t_{\mu(i)}}{c_{\mu(i)}\ a}\ ,\ \cdots\ ,\ x_n = \dfrac{t_{n-1}}{c_{n-1}\ x_{n-1}}\ ,\ x_1 = \dfrac{t_n}{c_n\ x_n} \\ \qquad \cdots\ ,\ x_{i-1} = \dfrac{t_{i-2}}{c_{i-2}\ x_{i-2}} \quad for\ i = 3\ ,\ \cdots\ ,\ n. \end{cases}$$

2) For $x_{\mu(i)} = 0$, choose $x_i = b \in \mathbb{F}_q\backslash\{0\}$ randomly, and other x_i can be calculated in sequence:

$$\begin{cases} x_{i-1} = \dfrac{t_{i-1}}{c_{i-1}\ b}\ ,\ \cdots\ ,\ x_1 = \dfrac{t_1}{c_1\ x_2}\ ,\ \cdots\ ,\ x_{\mu(i)+1} = \\ \qquad \dfrac{t_{\mu(i)+1}}{c_{\mu(i)+1}\ x_{\mu(i)+2}} \quad for \quad i = n\ ,\ n-1\ ,\ \cdots\ ,\ 2, \\ x_n = \dfrac{t_n}{c_n\ b}\ ,\ x_{n-1} = \dfrac{t_{n-1}}{c_{n-1}\ x_n} \quad \cdots\ ,\ x_{\mu(i)+1} = \\ \qquad \dfrac{t_{\mu(i)+1}}{c_{\mu(i)+1}\ x_{\mu(i)+2}} \quad for \quad i = 1. \end{cases}$$

3) For $x_i = 0$ and $x_{\mu(i)} = 0$, choose $x_{\mu(i)+1} = c \in \mathbb{F}_q\backslash\{0\}$ randomly, and do the following work:

$$\begin{cases} x_{\mu(i)+2} = \dfrac{t_{\mu(i)+1}}{c_{\mu(i)+1}\ c}\ ,\ \cdots\ ,\ x_n = \dfrac{t_{n-1}}{c_{n-1}\ x_{n-1}} \quad for \quad i = 1, \\ x_{\mu(i)+2} = \dfrac{t_{\mu(i)+1}}{c_{\mu(i)+1}\ c}\ ,\ \cdots\ ,\ x_n = \dfrac{t_{n-1}}{c_{n-1}\ x_{n-1}}\ ,\ x_1 = \\ \qquad \dfrac{t_n}{c_n\ x_n} \quad for \quad i = 2, \\ x_{\mu(i)+2} = \dfrac{t_{\mu(i)+1}}{c_{\mu(i)+1}\ c}\ ,\ \cdots\ ,\ x_n = \dfrac{t_{n-1}}{c_{n-1}\ x_{n-1}}\ ,\ x_1 = \\ \dfrac{t_n}{c_n\ x_n}\ ,\ \cdots\ ,\ x_{i-1} = \dfrac{t_{i-2}}{c_{i-2}\ x_{i-2}} \quad for \quad i = 3\ ,\ \cdots\ ,\ n. \end{cases}$$

From the discussions above, it can be seen that if there exists a singularity, i.e.,$\{x_1, \cdots, x_n\}$ such that

$\{x_1, \cdots, x_n | x_1 = 0 \vee \cdots \vee x_n = 0\}$, there must be some $t_i = 0$. In this situation, the inverse image of $\{t_1, \cdots, t_n\}$ is multiple, and the checksum need to be estimated. As the number of singularities is $q^n - (q-1)^n$, the probability of the existence of a singularity is $p_1 = 1 - \frac{(q-1)^n}{q^n}$, where $q = 2^m$. Proper parameters can guarantee that the probability is small enough and improve the decryption efficiency. Under the parameters $m = 12, n = 28$, the probability is $p_1 = 0.007$; $p_2 = 0.002$, for $m = 14, n = 28$; $p_1 = 0.0005$, for $m = 16, n = 27$, so the parameters $m = 16, n = 27$ are recommended.

3.1.2 Construction of Invertible Transformation based on Tame Transformation

Lemma 2. *Suppose the invertible transformation to construct is L_4. Choose positive integers n, d such that $n > 2d$, and define the invertible transformation based on tame transformation $L_4 : (x_1, \cdots, x_n) \to (t_1, \cdots, t_n)$*

$$\begin{cases} t_1 & = & x_1 & + & x_{d+1}x_n \\ t_2 & = & x_2 & + & x_{d+2}x_{n-1} \\ \vdots & & & & \\ t_d & = & x_d & + & x_{2d}x_{n-d+1} \\ t_{d+1} & = & x_{d+1} & & \\ \vdots & & & & \\ t_n & = & x_n. & & \end{cases}$$

Then the inverse image of $(t_1, \cdots, t_n)$, i.e., $L_4^{-1}(t_1, \cdots, t_n) = (x_1, \cdots, x_n)$ is given by

$$\begin{cases} x_1 & = & t_1 & + & t_{d+1}t_n \\ x_2 & = & t_2 & + & t_{d+2}t_{n-1} \\ \vdots & & & & \\ x_d & = & t_d & + & t_{2d}t_{n-d+1} \\ x_{d+1} & = & t_{d+1} & & \\ \vdots & & & & \\ x_n & = & t_n. & & \end{cases}$$

3.1.3 Construction of Invertible Transformation based on "Special Oil and Vinegar"

Suppose the invertible transformation to construct is L_5, and choose positive integers o, v and n such that $o > v$ and $n = o + v$. Divide the variables $\{x_1, \cdots, x_n\}$ into two parts:$\{x_1, \cdots, x_v, \cdots, x_o\}$ and $\{x_{o+1}, \cdots, x_n\}$.

Lemma 3. *Randomly choose $r_i \in \mathbb{F}_q$, $i = 1, \cdots, n$ and define $L_5(x_1, \cdots, x_n) = (t_1, \cdots, t_n)$ as follows:*

$$\begin{cases} t_1 & = & x_1 \\ \vdots & & \\ t_v & = & x_v \\ \vdots & & \\ t_o & = & x_o \\ t_{o+1} & = & \frac{(x_1+r_1)}{x_{o+1}} \\ \vdots & & \\ t_n & = & \frac{(x_v+r_v)}{x_n} \end{cases}$$

where variables $(t_1, \cdots, t_o)$ containing the first degree parts can be seen as "oil variables"; and variables $(t_{o+1}, \cdots, t_n)$ containing the quadratic parts can be seen as "vinegar variables".

Then the inverse image of $(t_1, \cdots, t_n)$, i.e., $L_5^{-1}(t_1, \cdots, t_n) = (x_1, \cdots, x_n)$, where $t_i \in \mathbb{F}_q^* := \mathbb{F}_q \backslash \{0\}$, $i = o+1, \cdots, n$ is given by

$$\begin{cases} x_1 & = & t_1 \\ \vdots & & \\ x_v & = & t_v \\ \vdots & & \\ x_o & = & t_o \\ x_{o+1} & = & \frac{t_{o+1}}{(t_1+r_1)} \\ \vdots & & \\ x_n & = & \frac{t_n}{(t_v+r_v)}. \end{cases}$$

Remark 2. *For some $t_i = 0, i = o+1, \cdots, n$, the inverse image of $(t_1, \cdots, t_n)$, i.e.,$L_5^{-1}(t_1, \cdots, t_n) = (x_1, \cdots, x_n)$ is obtained as follows.*

For some $t_i = 0$, either $x_i = 0$ or $x_{i-o} + r_{i-o} = 0$ is set up. If $x_{i-o} + r_{i-o} = 0$, choose $x_i \in \mathbb{F}_q$, and solve $(x_1, \cdots, x_n)$ from $(t_1, \cdots, t_n)$ directly, otherwise, utilize Lemma 3.1.1.

In conclusion, the existence of $t_i = 0$ makes the solution $(x_1, \cdots, x_n)$ not unique, the checksum need to be calculated. Similar to Section 3.1.1, the probability of the existence of a singularity is $p_2 = 1 - \frac{(q-1)^n}{q^n}$, so it can lower the probability, and improve the decryption efficiency by choosing proper parameters. For $m = 16, n = 27$, the probability is $p_2 = 0.0005$, so the parameters $m = 16, n = 27$ are proper.

3.2 Construction of Three Kinds of EMCs

By using the idea "function composition", and combining MI scheme with those nonlinear invertible transformations L_i in Section 3.1, the public key polynomials of the novel EMCs are deduced as follows:

$$\begin{aligned} \tilde{F}_i(x_1, \cdots, x_n) &= L_1 \circ \bar{F} \circ L_2 \circ L_i(x_1, \cdots, x_n) \\ &= (\tilde{F}_{i1}(x_1, \cdots, x_n), \cdots, \tilde{F}_{in}(x_1, \cdots, x_n)), \\ & \qquad i = 3, 4, 5 \qquad . \end{aligned}$$

Conversely, the secret keys consist of $(L_1^{-1}, L_2^{-1}, L_i^{-1}, \theta), i = 3, 4, 5$.

3.3 Encryption Algorithms

It can be seen that the secret keys of encryption algorithms are $D = (L_1^{-1}, L_2^{-1}, L_i^{-1}, \theta), i = 3, 4, 5$ from the construction process of the novel EMCs.

According to the construction of nonlinear transformation L in Section 3.1, when there exists a singularity $\{t_1, \cdots, t_n | t_1 = 0 \vee \cdots \vee t_n = 0\}$, the inverse images of $(t_1, \cdots, t_n)$: $L_3^{-1}(t_1, \cdots, t_n)$ and $L_5^{-1}(t_1, \cdots, t_n)$ can be multiple. Therefore, the encryption process and the decryption process will be discussed in two cases.

1) When there does not exist a singularity, the solution of $L_i^{-1}(t_1, \cdots, t_n)$ is unique.

   The encryption process. Given the plaintext $(x_1, \cdots, x_n) \in \mathbb{F}_q^n$, use the public key $\tilde{F}_i$ to calculate the ciphertext $(y_1, \cdots, y_n) = \tilde{F}_i(x_1, \cdots, x_n)$.

   The decryption process. Received the ciphertext $(y_1, \cdots, y_n) \in \mathbb{F}_q^n$, calculate the corresponding plaintext $(x_1, \cdots, x_n)$ as follows:

   a. Compute $(y_1^{'}, \cdots, y_n^{'}) = L_1^{-1}(y_1, \cdots, y_n)$;

   b. Use the secret key θ to get the inverse transformation of the central map $\bar{F}$, i.e., $(x_1^{'}, \cdots, x_n^{'}) = \bar{F}^{-1}(y_1^{'}, \cdots, y_n^{'})$;

   c. Compute $(t_1, \cdots, t_n) = L_2^{-1}(x_1^{'}, \cdots, x_n^{'})$;

   d. Finally, compute the corresponding plaintext $(x_1, \cdots, x_n) = L_i^{-1}(t_1, \cdots, t_n)$.

2) When there exists a singularity, i.e., the solution of $L_i^{-1}(t_1, \cdots, t_n)$ isn't unique.

   The encryption process. Given the plaintext $(x_1, \cdots, x_n) \in \mathbb{F}_q^n$, use the public key $\tilde{F}_i$ to calculate the corresponding ciphertext $(y_1, \cdots, y_n) = \tilde{F}_i(x_1, \cdots, x_n)$. Meanwhile, utilize the public hash function $Hash_1$ to calculate the checksum of plaintext $Hash_1(x_1, \cdots, x_n) = v$.

   The decryption process. Received the ciphertext $(y_1, \cdots, y_n) \in \mathbb{F}_q^n$ and the checksum $Hash_1(x_1, \cdots, x_n) = v$, the plaintext can be obtained as follows:

   a. Compute $(y_1^{'}, \cdots, y_n^{'}) = L_1^{-1}(y_1, \cdots, y_n)$;

   b. Use the secret key θ to get the inverse transformation of the central map $\bar{F}$, i.e., $(x_1^{'}, \cdots, x_n^{'}) = \bar{F}^{-1}(y_1^{'}, \cdots, y_n^{'})$;

   c. Compute $(t_1, \cdots, t_n) = L_2^{-1}(x_1^{'}, \cdots, x_n^{'})$;

   d. Use the secret key L_i^{-1} to get $(\bar{x}_1, \cdots, \bar{x}_n) = L_i^{-1}(t_1, \cdots, t_n)$,and compute $Hash_1(\bar{x}_1, \cdots, \bar{x}_n) = v^{'}$. If $v = v^{'}$, the corresponding solution $(\bar{x}_1, \cdots, \bar{x}_n)$ is the right plaintext, otherwise, discard the solution $(\bar{x}_1, \cdots, \bar{x}_n)$.

3.4 Signature Algorithms

The signature process. Suppose that the message M is the document to sign, and compute $(y_1, \cdots, y_n) = Hash_2(M)$. The secret keys of signature algorithms are the same as those of encryption algorithms, so are the process of calculating the signature $(x_1, \cdots, x_n)$ and the encryption process in Section 3.3. The difference is that whether there exists a singularity. When the signature isn't unique, choose one of the signatures randomly.

The verification process. Received the message M and signature $(x_1, \cdots, x_n)$, do the verification as follows:

1) Use another public Hash function $Hash_2$ to compute $Hash_2(M) = (y_1, \cdots, y_n)$;

2) Compute $\tilde{F}(x_1, \cdots, x_n) = (y_1^{'}, \cdots, y_n^{'})$, then determine whether the condition $(y_1^{'}, \cdots, y_n^{'}) = (y_1, \cdots, y_n)$ is true, otherwise, discard the invalid signature.

4 Operation Efficiency and Security Analysis

The operation efficiency and the security analysis of three novel EMCs will be given in the next installment.

4.1 Operation Efficiency

Encryption (verification) efficiency. Compared to the encryption (verification) efficiency of the MI scheme, the novel EMCs need simply do another operation L_i, $i = 1, 2, 3$. It can be seen that their efficiencies are high, and barely affect the whole efficiency of our proposed cryptosystem from the construction of L_i in Section 3.1.

Decryption (signature) efficiency. During the decryption process, when there exists a singularity, the solution isn't unique, and the verification need to be done many times. But the existence of a singularity can be avoided by choosing the proper parameters. During the signature process, just choose one of the solutions.

Above all, under the proper parameters, the three novel EMCs inherit high efficiency of MI, and the whole operation efficiency keeps high.

4.2 Security Analysis

Generally, attacks aimed at MPKCs are divided into two groups: structure-based attack and direct attack. Structure-based attack aims at the special structure of MPKCs, and it mainly includes linearization attack and differential attack. Direct attack starts with the public key polynomials of MPKCs. The common tools are comprised of the *Gröbner* base algorithm and the XL algorithm. Next, the security analysis of three EMCs will be performed. To keep it simple, take the EMC based on invertible cycle for example.

4.2.1 Linearization Attack

In 1995, Patarin present a linearization attack to the MI scheme, which simplified linear equations and posed threat to MI [12]. Next, it will be demonstrated that our proposed cryptosystem can be resistant against linearization attack.

Definition 3. *Let $P = (p_1, \cdots, p_m)$ be polynomials with n variables over F_q,with regard to P, a linearization equation is organized as*

$$\sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}x_iy_j + \sum_{i=1}^{n} b_ix_i + \sum_{i=1}^{m} c_iy_i + d$$
$$\in \mathbb{F}_q[x_1 , \cdots , x_n , y_1 , \cdots , y_m]$$

s.t. when plugging p_i into y_i, a zero polynomial about $(x_1, \cdots, x_n)$ the variable are obtained:

$$\sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}x_ip_j + \sum_{i=1}^{n} b_ix_i + \sum_{i=1}^{m} c_ip_i + d = 0.$$

According to the central map of MI $F : X \mapsto X^{q^\theta+1}$, the following special algebraic relation:$Y^{q^\theta-1} = X^{q^{2\theta}-1}$. Multiply XY on both sides of the relation to acquire the relation: $XY^{q^\theta} = YX^{q^{2\theta}}$.

Further, it can be easy to obtain n multivariate quadratic equations over F_q by the isomorphic mapping ϕ. Each equation is organized as follows:

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_iy_j + \sum_{i=1}^{n} b_ix_i + \sum_{i=1}^{n} c_iy_i + d = 0 \qquad (2)$$

Given $O((n + 1)^2)$ plaintext-ciphertext pairs $(x_1, \cdots, x_n, y_1, \cdots, y_n)$, it is feasible to work out the coefficients of the equation above. Once worked out all the coefficients and given the ciphertext $y = (y_1, \cdots, y_n)$, n linear equations about the plaintext $x = (x_1, \cdots, x_n)$ can be obtained.

Theorem 1. *The EMC based on invertible cycle puts forward the nonlinear invertible transformation L_3, s.t. the structure of public key polynomial will be changed to resist linearization attack.*

Proof. Similar to MI scheme, n multivariate quadratic equations of our proposed cryptosystem can be obtained, and each equation is organized as follows:

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}t_iy_j + \sum_{i=1}^{n} b_it_i + \sum_{i=1}^{n} c_iy_i + d = 0 \qquad (3)$$

If given $O((n + 1)^2)$ plaintext-ciphertext pairs $(t_1, \cdots, t_n, y_1, \cdots, y_n)$, the coefficients of the equations above can be calculated. However, since the ciphertext $(y_1, \cdots, y_n)$ is known, and the intermediate variables $(t_1, \cdots, t_n)$ remain unknown, the coefficients cannot be calculated when plugging the ciphertext $(y_1, \cdots, y_n)$ into the equation. In the worst case that all the coefficients are calculated, linear equations about t_i can be obtained by plugging $(y_1, \cdots, y_n)$ to Equation (3).

After plugging the expression of t_i, multivariate quadratic equations about $(x_1, \cdots, x_n)$ will be derived. Solving this kind of equation is still a NP problem, so it can be concluded that: our proposed cryptosystem is resistant against linearization attack.

4.2.2 Differential Attack

Differential attack aims at the type such as MI scheme. Initially, it was used to attack PMI [5], and it was also used to attack the SFLASH cryptosystem, i.e.,C^*- scheme later [8]. Next, it will be proved that the novel cryptosystem can resist differential attack.

Definition 4. *For any function $F(x)$, its differential at point F_{q^θ} is defined by $DF(a, x)$:*

$$DF(a, x) = F(x + a) - F(x) - F(a) + F(0).$$

When F is a quadratic function, if regard $DF(a, x)$ as a function of variables x and a, $DF(a, x)$ is a symmetric bilinear function about x and a.

In MI scheme, the inner function is $\tilde{F}(x) = x^{1+q^\theta}$, so $D\tilde{F}(a, x) = xa^{q^\theta} + ax^{q^\theta}$. Obviously, is symmetric bilinear., the differential function has a very specific multiplicative property:

$$D\bar{F}(a , \xi \cdot x) + D\bar{F}(\xi \cdot a , x) = (\xi + \xi^{q^\theta})D\bar{F}(a , x) \quad (4)$$

Similarly, the differential function of public key $P = L_1 \circ \bar{F} \circ L_2$ is $DP(a, x) = T \circ DF(U(a), U(x))$, which satisfies the following relation:

$$\begin{aligned} & DP(\xi a , x) + DP(a , \xi x) \\ = \; & L_1 \circ D\bar{F}(\xi \cdot L_2(a) , L_2(x)) \\ & \quad + \; L_1 \circ D\bar{F}(L_2(a) , \xi \cdot L_2(x)) \\ = \; & L_1 \circ (\xi + \xi^{q^\theta}) \circ L_1{}^{-1} \circ DP(a , x) \qquad (5) \end{aligned}$$

Let $P_\Pi = T_\Pi \circ \bar{F} \circ L_2$ be the public key of the C^{*-} scheme. It is entirely feasible to find the non-trivial map N_ξ such that

$$P^{'}{}_{\Pi} = P_\Pi \circ N_\xi = T_\Pi \circ M_\xi \circ \bar{F} \circ L_2 \qquad (6)$$

where N_ξ and M_ξ denote two linear maps with regard to ξ.

Therefore, a new MI public key can be obtained by comprising r equations randomly chosen from $P^{'}{}_{\Pi}$ with $(n - r)$ equations of the public key, and the probability of success is $1 - 1/q$. Then make use of linearization attack above to forge the signature.

Theorem 2. *The EMC based on invertible cycle utilizes the idea function composition and adds the nonlinear transformation L_3 to the MI scheme, therefore, it can break the special multiplicative property of MI and avoid differential attack.*

Proof. Relative to MI scheme, the public key of the novel EMC is transformed from P to $P^{'}$,

$$P^{'} = L_1 \circ \bar{F} \circ L_2 \circ L_3 \xlongequal{L^{'}{}_2 = L_2 \circ L_3} L_1 \circ \bar{F} \circ L^{'}{}_2 \qquad (7)$$

Since L_3 is a nonlinear transformation, $L^{'}{}_2$ in Relation (7) is also a nonlinear transformation.$\forall x\ ,\ \xi \in GF(q^n)$, there obviously exists the following relation:

$$\xi \circ L^{'}{}_2(x) \neq L^{'}{}_2(\xi x) \qquad (8)$$

To the novel EMC, the differential function of public key $P^{'} = L_1 \circ \bar{F} \circ L^{'}{}_2$ is

$$\begin{aligned} & DP^{'}(\xi a\ ,\ x) + DP^{'}(a\ ,\ \xi x) \\ = \ & L_1 \circ D\bar{F}(\xi \cdot L^{'}{}_2(a)\ ,\ L^{'}{}_2(x)) + \\ & \quad L_1 \circ D\bar{F}(L^{'}{}_2(a)\ ,\ \xi \cdot L^{'}{}_2(x)) \qquad (9) \\ \neq \ & L_1 \circ (\xi + \xi^{q^{\theta}}) \circ L_1{}^{-1} \circ (DP^{'}(a\ ,\ x)\ ) \end{aligned}$$

Expression (9) shows that the introduction of the transformation L_3 breaks the special multiplicative property of MI scheme.

In conclusion, our proposed cryptosystem can resist differential attack.

4.2.3 Algebraic Attack

The common tools of algebraic attack consist of the *Gröbner* base algorithm and the XL algorithm. So far, the most efficient methods to computer *Gröbner* bases are F4 and F5 algorithms.

According to the relations of the number of equations m and the number of variables n, and algebraic attack in three cases are discussed: $m > n$, $m < n$, and m=n. The equations satisfying the relation $m > n$ are called overdetermined equations [1], when $m < n$, underdetermined equations [18], and when m=n, permutation equations [21]. In our cryptosystem, the public key polynomial $P(x_1\ ,\ \cdots\ ,\ x_n) = (y_1\ ,\ \cdots\ ,\ y_n)$ satisfies the relation m=n. Therefore, the cases $m > n$ and $m < n$ are described no longer.

To the best of our knowledge, when $K = GF(q)$ $(q \neq 2)$ is big, and m=n, the complexity to solve the permutation equations is proved to be $O(2^{3m})$ [21].

In the novel cryptosystem, the corresponding equations are expressed as follows:

$$\begin{cases} p_1(x_1\ ,\ \cdots\ ,\ x_n) = y_1 \\ \qquad \vdots \\ p_n(x_1\ ,\ \cdots\ ,\ x_n) = y_n \end{cases} \qquad (10)$$

where the number of equations is equal to the number of variables. According to Section 4.1.1, the public key $p_i(x_1\ ,\ \cdots\ ,\ x_n)$ is multivariate quartic polynomial. The complexity to solve Equation (10) is much greater than the corresponding quadratic equations. Under the recommended parameters n=27 and $q = 2^{16}$, the complexity to solve multivariate quadratic equations is about $O(2^{81})$, therefore, the complexity to solve the public key polynomials of the novel EMC is more than $O(2^{81})$, that is, our proposed cryptosystem can be resistant against algebraic attack.

All in all, from Sections 4.2.1, 4.2.2, and 4.2.3, it can be concluded that the EMC based on invertible cycle can resist linearization attack, differential attack and algebraic attack. Similarly, the EMC based on tame transformation and the EMC based on special oil and vinegar are also secure, and detailed proofs are not given here.

5 Conclusions

In this paper, three different nonlinear invertible transformations are put forward. Incorporated with MI scheme, three novel EMCs are recommended. Next, the corresponding encryption and signature algorithms are provided. Finally, the operation analyses and security analyses of three novel cryptosystems are implemented. It can be demonstrated that our proposed cryptosystems can resist linearization attack, differential attack, and algebraic attack. Whether there is a new attack to our novel EMCs and the selection and optimized implementation of concrete parameters need further research.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (61003291).

References

[1] M. R. Albrecht, C. Cid, J. C. Faugere, and et al, "On the relation between the MXL family of algorithms and groebner basis algorithms," *Journal of Symbolic Computation*, vol. 47, no. 8, pp. 926–941, 2012.

[2] D. J. Bernstein, J. Buchmann, and E. Dahmen, *Post-quantum Cryptography*, Berlin: Springer Heidelberg, 2009.

[3] C. L. Clough and J. Ding, "Secure variables of the square encryption scheme," in *Post-quantum cryptography*, pp. 153–164, Springer Berlin Heidelberg, 2010.

[4] J. Ding, "A new variant of the Matsumoto-Imai cryptosystem through perturbation," in *Public Key Cryptography (PKC'04)*, pp. 305–318, Springer Berlin Heidelberg, 2004.

[5] J. Ding and J. E. Gower, "Inoculating multivariate schemes against differential attacks," in *Public Key Cryptography (PKC'06)*, pp. 290–301, Springer Berlin Heidelberg, 2006.

[6] J. Ding, C. Wolf, and B. Yang, "Invertible cycles for multivariate quadratic (MQ) public key cryptography," in *PKC'07*, pp. 266–281, Beijing, China, 2007.

[7] J. T. Ding and B. Y. Yang, *Multivariate Public Key Cryptography*, Berlin: Springer Heidelberg, 2009.

[8] V. Dubois, P. A. Fouque, and J. Stern, "Cryptanalysis of sflash with slightly modified parameters," in *Advances in Cryptology (Eurocrypt'07)*, pp. 264–275, Springer Berlin Heidelberg, 2007.

[9] D. S. A. Elminaam, H. M. Abdual-Kader, and M. M. Hadhoud, "Evaluating the performance of symmetric encryption algorithms," *International Journal of Network Security*, vol. 10, no. 3, pp. 216–222, 2010.

[10] X. Q. Fu, W. S. Bao, and C. Zhou, "Speeding up implementation for shor factorization quantum," *Chinese Sci Bull*, vol. 55, no. 4-5, pp. 322–327, 2010.

[11] T. Matsumoto and H. Imai, "Public quadratic polynomial-tuples for efficient signature verification and message-encryption," in *Advances in Cryptology (Eurocrypt'88)*, pp. 419–453, Springer Berlin Heidelberg, 1988.

[12] J. Patarin, "Cryptanalysis of the matsumoto and imai public key scheme of eurocrypt '88," in *Advances in Cryptology (Crypto'95)*, pp. 248–261, Springer Berlin Heidelberg, 1995.

[13] J. Patarin, N. Courtois, and L. Goubin, "Flash, a fast multivariate signature algorithm," in *Topics in Cryptology (T-RSA'01)*, pp. 298–307, Springer Berlin Heidelberg, 2001.

[14] A. Petzoldt, S. Bulygin, and J. Buchmann, "Cyclicrainbow - a multivariate signature scheme with a partially cyclic public key," in *Progress in Cryptology (Indocrypt'10)*, pp. 33–48, Hyderabad, India, 2010.

[15] K. Sakumoto, "Public-Key identification schemes based on multivariate cubic polynomials," in *Public Key Cryptography*, pp. 172–189, Springer Berlin Heidelberg, 2012.

[16] P. Shor, "Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer," *SIAM Review*, vol. 41, no. 2, pp. 303–332, 1999.

[17] S. Tang and L. Xu, "Proxy signature scheme based on isomorphisms of polynomials," in *Network and System Security*, pp. 113–125, Springer Berlin Heidelberg, 2012.

[18] E. Thomae and C. Wolf, "Solving underdetermined systems of multivariate quadratic equations revisited," in *Public Key Cryptography*, pp. 156–171, Springer Berlin Heidelberg, 2012.

[19] S. Tsujii, M. Gotaishi, K. Tadaki, and et al, "Proposal of a signature scheme based on STS trapdoor," in *Post-quantum Cryptography*, pp. 201–217, Springer Berlin Heidelberg, 2010.

[20] H. Wang and H. Zhang, "Extended multivariate public key cryptosystems with secure encryption function," *Science China Information Sciences*, vol. 54, no. 6, pp. 1161– 1171, 2011.

[21] H. Wang, H. Zhang, and H. Guan, "Multivariate algebra theory and its application in cryptography," *Journal of Beijing University Technology*, vol. 36, no. 5, pp. 9–17, 2010.

[22] S. Wang, R. Ma, Y. Zhang, and et al, "Ring signature scheme based on multivariate public key cryptosystems," *Computers and Mathematics with Applications*, vol. 62, no. 10, pp. 3973–3979, 2012.

[23] C. Wolf and B. Preneel, *Taxonomy of Public Key Schemes Based on the Problem of Multivariate Quadratic Equations*, Technical Report, Cryptology ePrint Archive, Report 2005/077, Dec. 2005.

[24] G. Yang, S. Tang, and L. Yang, "A novel group signature scheme based on MPKC," in *Information Security Practice and Experience*, pp. 181–195, Springer Berlin Heidelberg, 2011.

Shuaiting Qiao received his B.S. degree in applied mathematics from the Henan university, Kaifeng,China, in 2011. He is currently pursuing his M.S degree in department of Information Research, Zhengzhou Information Science and Technology Institute, Zhengzhou, China. His research fields include multivariate public key cryptography and information security.

Wenbao Han received his Ph.D. degree in mathematics from Sichuan University. He is currently a professor in the Department of Information Research, Zhengzhou Information Science and Technology Institute, Zhengzhou, China. His research field is information security.

Yifa Li received his Ph.D. degree in applied mathematics from the Zhengzhou Information Science and Technology Institute, China.He is a associate professor in the Department of Information Research, Zhengzhou Information Science and Technology Institute, Zhengzhou, China. His research field is information security.

Luyao Jiao received his B.S. degree in applied mathematics from the Henan university,Kaifeng, China, in 2010. His research field is multivariate public key cryptography.

A Key Based Secure Threshold Cryptography for Secret Image

Prabir Kr. Naskar[1], Hari Narayan Khan[2], Atal Chaudhuri[3]
(Corresponding author: Prabir Kr. Naskar)

Department of Computer Science & Engineering, MCKV Institute of Engineering[1]
Liluah, Howrah-711204, West Bengal, India.
Department of Computer Science & Engineering, Regent Education and Research Foundation[2]
Barrackpore, Kolkata-700121, West Bengal, India.
Department of Computer Science & Engineering, Jadavpur University[3]
Jadavpur, Kolkata-700032, West Bengal, India.
(Email: cse.prabir@gmail.com, manik1984@gmail.com, atalc23@gmail.com)

(Received June 12, 2013; revised and accepted Jan, 22 & Apr. 29, 2014)

Abstract

This paper presents a key based secured (k, n) threshold cryptography where key is used to encrypt the secret and then the secret as well as key is shared among set of n participants. In sharing phase, each secret byte is selected randomly from secret fields depending upon the key. That provides additional protection of the secret data. Also, each share has some bytes missing and these missing bytes can be recovered from a set of exactly k shares. Thus a given byte position can be confirmed for any k shares, but not less than k. Hence k shares are required to give back the secret. As a result, the generated shares are compressed and if k is closer to n then the compression ratio is increased. That provides strong protection of secret data. At the reconstruction phase only when a qualified set of legitimate shares comes together then reconstruction is possible. The proposed scheme is described in detail along with its security analysis, such as key sensitivity analysis and statistical analysis. This scheme has been tested using different images to prove that the scheme has great potential and has a good ability to achieve the confidential security.

Keywords: Compression, image Sharing, key-based threshold cryptography, perfect secret sharing (PSS), random selection of secret bytes, statistical analysis

1 Introduction

Protection of sensitive data is an important issue, during transmission over internet. Many cryptographic techniques are there to protect secret data. However, a common weakness of these technique is that an entire secret data is kept in a single medium. The secret data cannot be revealed if the medium or key is lost or corrupted. This is termed as a single point failure. To overcome this drawback secret sharing becomes more popular. In the secret sharing scheme, there is one dealer and n participants. The dealer gives a secret to the participants, but only when specific conditions are fulfilled. The dealer accomplishes this by giving each participant a share in such a way that any group of k or more participants (i.e., qualified participant) reconstruct the secret but no group of less than k players can. Such a system is called a (k, n)-threshold based secret sharing scheme. Threshold Secret sharing scheme thus says that: A secret is some data S. Our goal is to divide S into n shares V_1, V_2. $\cdots$, V_n in such a way that:

1) Knowledge of any k or more V_i shares makes S easily computable, where $1 \leq i \leq n$ and $2 < k \leq n$.

2) Knowledge of any $k-1$ or fewer V_i shares leaves S completely undetermined (in the sense that all its possible values are equally likely).

If $k = n$, then all the shares are required in the (n, n)-threshold scheme to recover the secret. However, the lost of any of the share produced using the (n, n)-threshold scheme results in inaccessible secret messages. Figure 1 shows the conceptual view of a (k, n)-threshold sharing scheme.

Well known secret sharing schemes (SSS) in the literature include Shamir's SSS [16] based on polynomial interpolation, Blakley's SSS [2] based on hyper plane geometry, Asmuth-Bloom's SSS [1] based on Chinese Remainder theorem. Karnin et al. [11] suggested the concept of perfect secret sharing (PSS) where zero information of the secret is revealed for an unqualified group of $(k-1)$ or fewer members. Specifically, Karnin et al. [11] used a term referred as information entropy (a measurement of the uncertainty of the secret), denoted as $H(s)$ where s is

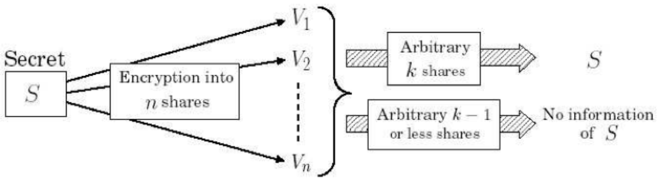


Figure 1: Concept of shared cryptography

a secret shared among n participants. The claim of PSS (Perfect Secret Sharing) schemes must satisfy the following:

1) A qualified coalition of n or more participants, C can reconstruct the secret (s), s: $H(s|C) = 0$, $\forall |C| \geq k$;

2) An unqualified coalition of $(n - 1)$ or few participants, C has no information about the secret (s), s: $H(s|C) = H(s)$, $\forall |C| < k$.

For these requirements in PSS schemes, a secret has zero uncertainty if the secret can be discovered by n or more participants. On the contrary, the secret, in PSS schemes, remains the same uncertainty for $(k-1)$ or fewer members. Therefore, there is no information exposed to the $(k - 1)$ or fewer members.

A shortcoming of above secret sharing schemes is the need to reveal the secret shares during the reconstruction phase. The system would be more secure if the subject function can be computed without revealing the secret shares or reconstructing the secret back. This is known as function sharing problem where the function computation is distributed according to underlying SSS such that distributed parts of computation are carried out by individual user and then the partial results can be combined to yield the final result without disclosing the individual secrets. Various function sharing protocols are been proposed [4, 5, 6, 7, 10, 15, 17] mostly based on Shamir secret sharing as the underlying scheme. A better image secret sharing approach was also proposed by Thien & Lin [18]. With some cryptographic computation, they cleverly used Shamir SSS to share a secret image. Chao et al. [3] proposed a method to extend (n, n) scheme to (k, n) scheme by using shadows-assignment matrix. Dong and Ku [8] proposed a new (n, n) secret image sharing scheme with no pixel expansion. In their scheme reconstruction is based on addition which has low computational complexity. Dong et al. [9] proposed a $(2, n)$ secret sharing scheme based on Boolean operation. The reconstructed image is totally the same with the original secret image and the scheme has no pixel expansion and contrast value was ideal. Apart from above secret sharing schemes, we propose a secret sharing scheme, where each share contains partial secret information. As a result, each generated shares are compressed. That provides strong protection of the secret data. In our scheme, each share contains secret data and header data as shared form. A header structure is constructed by the key, k, n and total number of bytes in secret and individual share number. At the reconstruction phase, only when k numbers of shares come together, then original header is reconstructed that is used to reconstruct the original secret. In our scheme secret reconstruction is not possible for less than threshold (k) number of shares; so it is Perfect Secret Sharing scheme.

2 Background and Related Work

2.1 Shamir's Secret Sharing Scheme

Shamir's secret sharing scheme [16] is based on (k, n)-threshold based secret sharing technique ($k \leq n$). The technique allows any k out of n shares to construct a given secret, a $(k - 1)$ degree polynomial is necessary. This polynomial function of order $(k - 1)$ is constructed as,

$$F(x) = a_0 + a_1 x + a_2 x^2 + ... + a_{k-1} x^{k-1} \bmod p.$$

Now we can easily generate n number of shares by using above equation. Where a0 is the secret, p is a prime number and all other coefficients are random elements from the secret. Each of the n shares is a pair (x_i, y_i) of numbers satisfying $f(x_i) = y_i$ and $x_i > 0$, $1 \leq i \leq$n and $0 < x_1 < x_2 < ... < x_k \leq p - 1$. Given any k shares, the polynomial is uniquely determined and hence the secret a0 can be computed via Lagrange interpolation. However, given $k - 1$ or fewer shares, the secret can be any element in the field.

The polynomial function f(x) is destroyed after each shareholder possesses a pair of values (x_i, y_i) so that no single shareholder knows the secret value a_0. In fact, no groups of k-1 or fewer shares can discover the secret a_0. On the other hand, when k or more secret shares are available, then we may set at least k linear equations $y_i = f(x_i)$ for the unknown a_i. The unique solution to these equations shows that the secret value a_0 can be easily obtained by using Lagrange interpolation.

2.2 Blakley's Secret Sharing Scheme

Blakley's [2] scheme is less space-efficient than Shamir's, while Shamir's shares are individually as large as the original secret. This scheme uses hyperplane geometry to solve the secret sharing problem. The secret is a point in a k-dimensional space and n shares are affine hyperplanes that pass through this point. An affine hyperplane in a k-dimensional space with coordinates in a field can be described by a linear equation of the following form:

$$a_1 x_1 + a_2 x_2 + a_3 x_3 + ... + a_k x_k = b.$$

The intersection point is obtained by finding the intersection of any k of these hyperplanes. The secret can be any of the coordinates of the intersection point or any function of the coordinates. We take the secret to be the first coordinate of the point of intersection.

2.3 Asmuth-Bloom's Secret Sharing Scheme

A fundamentally different Secret sharing scheme is Asmuth-Bloom's secret sharing scheme [1] which shares a secret among the parties using modular arithmetic and reconstructs it by Chinese Remainder Theorem (CRT). In Asmuth-Bloom's Secret Sharing Scheme, the sharing and Reconstruction of the secret can be done as follows.

Sharing Phase: To share the secret d among a group of n users, the dealer does the following:

1) A set of relatively prime integers $m_0 < m_1 < m_2 < ... < m_n$ where $m_0 > d$ is a prime, are chosen such that $\prod_{i=1}^{k} m_i > m_0 \ \ \prod_{i=1}^{k-1} m_{n-i+1}$;

2) Let M denote $\prod_{i=1}^{k} m_i$, the dealer computes,

$$y = d + am_0,$$

where a is a positive integer generated randomly subjected to the condition that $0 \leq y < M$.

3) The share of the i^{th} user $1 \leq i \leq n$, $y_i = y$ mod m_i.

Reconstruction Phase: Assume S is a coalition of k users to reconstruct the secret, let M_s denote$\prod_{i \in S^{m_i}}$.

1) Given the system $y = y_i$ mod m_i. For i belongs to S, solve y in Z_{MS} using the Chinese Remainder Theorem.

2) Compute the secret as: $d = y$ mod m_0.

According to Chinese Remainder Theorem, y can be determined uniquely in Z_{MS}. Since $y < M \leq M_S$, the solution is also unique in Z_M.

2.4 Thien and Lin's Secret Sharing Scheme

Thien and Lin proposed a (k, n)-threshold-based image secret sharing scheme [18] by cleverly using Shamir's SSS [16] to generate shares. The essential idea is to use a polynomial function of order $(k-1)$ to construct n image shares from a $L \times L$ pixel secret image (denoted as I) as

$$\begin{aligned} S_x &= I(i_{k+1}, j) + I(i_{k+2}, j)x + I(i_{k+3}, j)x^2 \\ & \quad +... + I(i_{k+k}, j)x^{k-1} \bmod p, \end{aligned} \tag{1}$$

where $0 \leq i \leq L/k$ and $1 \leq j \leq L$. This method reduces the size of image shares to become 1/k of the size of the size of the secret image. Any k image shares are able to reconstruct every pixel value in the secret image.

In above secret sharing schemes, each share contains the complete secret information in encrypted or ciphered form. Apart from above schemes, the idea behind our proposed scheme is that every share has some bytes missing and these missing bytes can be recovered from a set of exactly k shares. Thus a given byte position can be confirmed for any k shares, but not less than k. Hence k shares are required to give back the secret. Here we use image file as a secret data, but it is equally applicable for any digital data. In our first work [12], we proposed the basic concept of our scheme. Where we have shown, a secret can be shared among a set of participants by information sharing that means all the shares contain partial information about the secret. Then we have applied this scheme [14] for audio file by applying intermediate encryption using the digest of a given key and share the header information by applying simple ANDing with individual mask. But in example (Section 3.3), we have shown that if we apply simple ANDing with individual mask, some header information will be opened for attacker. Then [13] we have used this scheme for a digital image by sharing header with the concept linear geometry, where coefficient values are selected from generated shares. Therefore, if and only if k numbers of legitimate shares come together, then header reconstruction is possible, as a result lossless secret data will be reconstructed. Here we use the previous scheme in modified form where secret byte selection is randomly depending upon the key from the secret field which provide additional protection of the secret data and we discuss the strength of our scheme by analyzing the scheme in Section 5 with comparing existing schemes. Some additional experimental results are shown in Section 7 and the strength of our scheme is tested using statistical analysis (e.g. histogram analysis and correlation value, etc.) in Section 8, which shows that our scheme is completely prefect and secure secret sharing scheme. To establish the strength of our scheme, we have shown mask generation algorithm and our proposed secret sharing scheme with suitable example.

3 Mask Generation Algorithm

The proposed work is based upon masking which employs ANDing for share generation and ORing the predefined minimal number of shares to reconstruction the original.

3.1 Concept

For better understanding let us consider any secret as a binary bit file (i.e. bit is the smallest unit to work upon, in actual implementation one can consider a byte or group of bytes or group of pixels as the working unit). The secret could be an image, an audio or text etc. We shall decompose the bit file of any size onto n shares in such a way that the original bit file can be reconstructed only ORing any k number of shares where $k \leq$n$\geq$ 2 but in practice we should consider 2 $\leq$k<n$\geq$ 3.

Our basic idea is based on the fact that every share should have some bits missing and those missing bits will be replenished by exactly $(k-1)$ other shares but not less than that. So every individual bit will be missed from exactly $(k-1)$ shares and must be present in all remaining (n-k+1) shares, thus the bit under consideration is available in any set of k shares but not guaranteed in less than k shares. Now for a group of bits, for a particular bit position, $(k-1)$ number of shares should have the bit missed and $(n-k+1)$ number of shares should have the bit present and similarly for different positions there should be different combinations of $(k-1)$ shares having the bits missed and $(n-k+1)$ number of shares having the bits present. Clearly for every bit position there should be C^n_{k-1} such combinations and in our scheme thus forms the mask of size C^n_{k-1}, which will be repeatedly ANDed over the secret in any regular order. Different masks will produce different shares from the secret. Thus 0 on the mask will eliminate the bit from the secret and 1 in the mask will retain the bit forming one share. Different masks having different 1 and 0 distributions will thus generate different shares.

Next just ORing any k number of shares we get the secret back but individual share having random numbers of $1'$s and $0'$s reflect no idea about the secret.

A possible set of masks for 5 shares with threshold of 3 shares is shown below:

Share-1:	1	0	1	0	1	0	1	0	1	1
Share-2:	1	0	1	1	1	1	0	1	0	0
Share-3:	1	1	0	0	0	1	1	1	1	0
Share-4:	0	1	1	1	0	0	1	1	0	1
Share-5:	0	1	0	1	1	1	0	0	1	1

One can easily check that ORing any three or more shares we get all $1'$s but with less than three shares some positions still have $0'$s, i.e. remain missing.

Algorithm 1 Pseudo Code for mask generation

```
Input: $n$, $k$
Output: $mask[n][]$
Integer $mask\_generator(n, k, mask[n][])$
{
$bin\_arr[][n]$: Integer array;
$mask\_pattern\_len = 0$;
$max\_val = 2^n - 1$;
// calculate decimal value of $n$ numbers of 1s.
for $i = max\_val$ to 1 do
    $Decimal\_to\_Binary(i, bin[][])$;
    // calculate binary equivalent of decimal $i$ and store in $bin\_arr[][]$ array.
    if $(Zero\_Check(bin[mask\_pattern\_len[n], k)))$ then
        $mask\_pattern\_len = mask\_pattern\_len + 1$;
        // check whether $(k-1)$ nos. of zero exist or not, if exist then increment $mask\_pattern\_len$ by 1.
    end if
end for
$Rearrange\_Array(bin)$;
// rearrange the row of $bin[][n]$ array.
$Transpose(mask, bin)$;
//take transpose matrix of $bin[][n]$ and store in $mask[n][]$.
Return $mask\_pattern\_len$;
}
End
```

3.2 Algorithm

Here we are presenting the algorithm for designing the masks for n shares with threshold k.

Step 1. List all row vectors of size n having the combination of $(k-1)$ numbers of $0'$s and $(n-k+1)$ numbers of $1'$s and arranged them in some predefined order in terms of their decimal equivalent and finally organized them in the form of a matrix. Obvious dimension of the matrix will be $C^n_{k-1} \times n$.

Step 2. Transpose the matrix generated in Step-1. Obvious dimension of the transposed matrix will be $n \times C^n_{k-1}$. Each row of this matrix will be the individual mask for n different shares. The size of each mask is C^n_{k-1} bits, i.e. the size of the mask varies with the value of n and k.

Let us consider the previous example where $n = 5$ and $k = 3$.

Step 1. List of row vectors of size 5 bits with 2 numbers of $0'$s and 3 numbers of $1'$s, arranged in predefined manner as agreed during sharing phase in order to get masks identical to those used in share generation phase. (Here the arrangement is the highest followed by lowest then next highest followed by next lowest and so on, which appears here as 28, 7, 26, 11, 25, 13, 22, 14, 21, 19).

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 \end{bmatrix} C^5_2 \times 5 = 10 \times 5$$

Step 2. Take the transpose of the above matrix and we get the desired masks for five shares as listed above in the form of matrix of dimension $5 \times C^5_2$, i.e. 5×10. There are five masks each of size 10 bits.

3.3 Example

Consider a secret message (M) is WBPRACSE27 and the size of M is 10 bytes. Now by applying logical ANDing with individual mask following shares are generated.

S_i	$Mask$	$Shared\ Message$
$Share-1:$	1010101011	$W0P0A0S027$
$Share-2:$	1011110100	$W0PRAC0E00$
$Share-3:$	1100011110	$WB000CSE20$
$Share-4:$	0111001101	$0BPR00CS07$
$Share-5:$	0101110011	$0B0RAC0027$

From above shares we can easily notice that each share contains partial secret information. That is a secret byte corresponding to one in the mask is kept as it is and the secret byte corresponding to zero in the mask is kept as zero. So every share has some bytes missing and these missing bytes can be recovered from a set of exactly k shares. In mask generation algorithm for n shares and k threshold, size of each mask is C^n_{k-1}, where the number of zeros and ones are C^{n-1}_{k-2} and C^{n-1}_{n-k} respectively. Total size of all shares is 50 bytes ($\sum_{i=1}^{n} Sizeof(S_i)$).

Therefore, each share contains C^{n-1}_{n-k} numbers of partial secret bytes for a set of C^n_{k-1} numbers of secret bytes. Now all zero bytes corresponding to zero bit in the mask are discarded, that introduced a unique compression technique [Section 6]. Therefore, above shared message becomes compressed message.

S_i	$Compressed\ Message$
$Share-1:$	$WPAS27$
$Share-2:$	$WPRACE$
$Share-3:$	$WBCSE2$
$Share-4:$	$BPRCS7$
$Share-5:$	$BRAC27$

Now the total size of all shares is 30 bytes (< 50). Here all secret data are partially open to the participants. To overcome this problem encrypts individual share using a key. After that, the key itself is shared and concatenated with individual share to generate complete shares. Details of this scheme are discussed in the following section.

4 Secret Image Sharing Protocol

Our proposed scheme shares both secret data and header structure including key. Therefore every share has two parts secret share and header share.

4.1 Concept

Our proposed scheme is key based secure threshold cryptography. Initially a 16-byte digest string is generated from user given variable length key (UK_y) using MD5. This 16 bytes digest string is used as encryption key (K_y). The concept is variable length key becomes fixed length key. The length of UK_y should be greater than or equal to 16 bytes. Consider UK_y is "testkey@encry185" then generated fixed length K_y is shown in hex form as "CC7B269F18A6DDB8255EAF4799982131" and the length of K_y is 16 bytes. Here we use MD5 but one can use any strong hash function or random number generator. An advantage of using key based encryption is that it provides authentication as long as the key stays secret. It allows encryption and decryption using same key that is symmetric encryption. This scheme is free to save the key, because the key is also shared among the set of participants. So it reduces the chance of compromising. Also, depending upon k and n, n number of masks are generated and each mask is used to generate individual share and a secret byte (SB) corresponding to one in the mask is kept by encrypting using Equation (3). Also the secret byte (SB) corresponding to zero in the mask is simply discarded. Therefore, every share is compressed. Each share has some bytes missing, all missing bytes can be recovered from a set of exactly k shares. Here each SB is selected randomly from secret field using the following equation

$$f(x) = (Z^2 + c) \bmod L. \quad (2)$$

Where L is total number of secret bytes, Z is random value from K_y and c is $f(x-1)$. Therefore, it provides additional protection of the secret data. Here a complete header structure is to be generated using n, k, K_y and the total number of secret bytes in secret. After that header information is shared using Equation (4) and then each shared header (with individual share number) is appended with secret share to generate a complete share. In reconstruction phase, first collect k number of shares and then reconstruct the complete header information. Now from reconstructed header, the value of n and k are used to generate same mask as sharing phase using same mask generation algorithm. Then apply Equation (5) to decrypt shared bytes, which are selected from shares. This is applicable for nonzero (one) value of that mask with same index position. Thus the missing byte is recovered by inserting zero corresponding to zero in the same index position of that mask. Now apply ORing of k numbers of reconstructed bytes to generate the original secret byte. Therefore the missing bytes can be recovered from a set of exactly k shares. After that, each reconstructed secret bytes are placed in proper position to reconstruct the secret as lossless manner. Now stepwise sharing and reconstruction phases are discussed in following section. The conceptual view of our scheme is described in Figure 2.

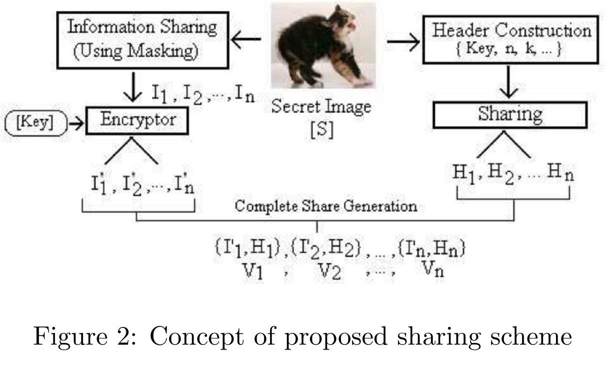


Figure 2: Concept of proposed sharing scheme

4.2 Sharing Phase

The sharing phase of the proposed scheme is stated in the following steps.

Input: threshold (k), total number of share (n), user given variable length key (UK_y), secret image (SI).

Step 1. Initially generate 16 bytes digest from user given variable length key (UK_y). This 16 bytes digest string is used as encryption key (K_y). Here we use MD5 hash function to generate 16 bytes digest string. So variable length key becomes fixed length key.

Step 2. Now construct Header Structure (h) of five fields and put share number (S_n) in 1^{st} field, total number (n) of shares in 2^{nd} field, threshold number (k) in 3^{rd} field, key (K_y) in 4^{th} field, and the total secret bytes (for image, only consider width) (W) in 5^{th} field (See Table 1).

The size of this header structure is 23-bytes. This structure or size of individual field may vary according to our requirements.

Step 3. Generate n masks for n individual shares using the proposed mask generation algorithm [Section 3.2]. Consider the mask pattern length is ML. Therefore, $ML = mask\_generator(n, k, mask[n][])$.

Step 4. Calculate total number of secret bytes (L) present in SI using height and width (W). For other digital files such as text, audio the total number of secret bytes L will be equal to W (i.e. actual size of secret). Also consider an array of N location say Index[L] and initialize all the location by zero, i.e. Index[L] = {0}, where zero indicates unread secret byte at i^{th} position and one indicates read i^{th} secret byte.

$$if \quad \begin{cases} index[i] = 0, & \text{unread} \\ index[i] = 1, & \text{read} \end{cases}$$

Step 5. Now select specific secret byte (SB) from a random position (PS) using Algorithm 2.

The secret byte (SB) corresponding to zero in the mask is simply discarded. Therefore, each share contains partial secret information and for each retained secret bytes apply Step 6 to generate confused secret bytes.

Step 6. Then the i^{th} retained byte (P_i) is ciphered by the j^{th} byte (K_{y_j}) by the following operation:

$$R_i = P_{i-1} \oplus (P_i \times K_{y_j}) \bmod 251, \qquad (3)$$

where $i = 0, 1, 2, \cdots, (L-1)$ and $j = \bmod(i, 16)$ and 251 is largest prime number in 8 bits.

0^{th} retained byte is ciphered by the 0^{th} byte of K_y. $R_0 = P_{-1} \oplus (P_0 \times K_{y_0}) \bmod 251$, where P_{-1} is zero.

Algorithm 2 Secret byte (SB) selection and distribution

```
$PS = 0$;
For $j = 0$ to $(L-1)$
$PS = K_y[j\%16] \times K_y[(j+1)\%16] + PS$;
$PS = PS\%L$;
while $(Index[PS]! = 0)$ do
    $PS = (PS + 1)\%L$;
end while
Read secret byte (SB) from $PS^{th}$ position;
$Index[PS] = 1$; // 0 for unread and 1 for read
For $i = 1$ to $n$
if $(mask[i][j\%ML] = 1)$ then
    Apply Step-6 for ciphering the SB;
    Write the ciphered byte in $i^{th}$ share;
end if
End For;
EndFor;
End
```

1^{st} retained byte is ciphered by the 1^{st} byte of K_y. $R_1 = P_0 \oplus (P_1 \times K_{y_1}) \bmod 251$.

2^{nd} retained byte is ciphered by the 2^{nd} byte of K_y. $R_2 = P_1 \oplus (P_2 \times K_{y_2}) \bmod 251$,

...

Same as t^{th} retained byte is ciphered by $(t \bmod 16)_{th}$ byte of K_y. $R_t = P_{t-1} \oplus (P_t \times K_{y_{(t \bmod 16)}}) \bmod 251$.

Step 7. Now from each of n ciphered shares collect k numbers of nonzero sample bytes from prefixed locations and thus matrix (A) of dimension $(n \times k)$ is formed.

$$\begin{bmatrix} a_{[0,0]} & a_{[0,1]} & \cdots & a_{[0,k-2]} & a_{[0,k-1]} \\ a_{[1,0]} & a_{[1,1]} & \cdots & a_{[1,k-2]} & a_{[1,k-1]} \\ . & . & \cdots & . & . \\ . & . & \cdots & . & . \\ . & . & \cdots & . & . \\ a_{[n-2,0]} & a_{[n-2,1]} & \cdots & a_{[n-2,k-2]} & a_{[n-2,k-1]} \\ a_{[n-1,0]} & a_{[n-1,1]} & \cdots & a_{[n-1,k-2]} & a_{[n-1,k-1]} \end{bmatrix}_{\text{n}\times\text{k}}$$

Step 8. The header (Table 1) excluding the leftmost field is also shared by applying following operation:-

$$V_i = \sum_{j=0,1,\cdots,k-1}^{i=0,1,\cdots,n-1} (a[i,j] \times h[j]). \qquad (4)$$

Step 9. Next each header share is appended with the share number (S_n) in the first field and concatenated with the corresponding secret share, which forms one complete share for transmission. (For shared image we have to add extra single height to add the shared header, i.e. if height of SI is h, then shared image height will be $(h+1)$).

Table 1: Header Structure (h)

1-Byte	1-Byte	1-Byte	16-Bytes	4-Bytes
Share number	Total number of Shares	Threshold	Encryption Key	Size of Secret
$[S_n]$	$[n]$	$[k]$	$[K_y]$	$[W]$

4.3 Reconstruction Phase

The reconstruction phase of the proposed scheme is stated in the following steps.

Input: k number of shares.

Step 1. Collect k-numbers of share and extract confused header information. Also generates $(k \times k)$ matrix (A).

Step 2. Now applying any conventional linear equation solving technique to reconstruct the original Header information.

Step 3. Once the original Header is reconstructed, we extract the 16 bytes digest string as well as the encryption key (K_y).

Step 4. Now using n and k, extracted from reconstructed header structure, generate n masks using same mask generation algorithm used in share generation phase (same set of mask in same order is reconstructed at the receiving end, used for expanding the compressed shares).

Step 5. According to the share number of the share holder appropriate mask is used to expand the secret share part by inserting zero bytes corresponding to zero in the corresponding mask.

Step 6. Calculate L using shared image height (h) and extracted W from header structure, i.e. $L = (h - 1) \times W$.

Step 7. Ciphered bytes (R_i) corresponding to 1 position in the mask, have generated by the Equation (3). So, apply following operation to get original byte (P_i).

$$P_i = P_{i-1} \oplus (R_i \times M_j^{-1}) \bmod 251. \qquad (5)$$

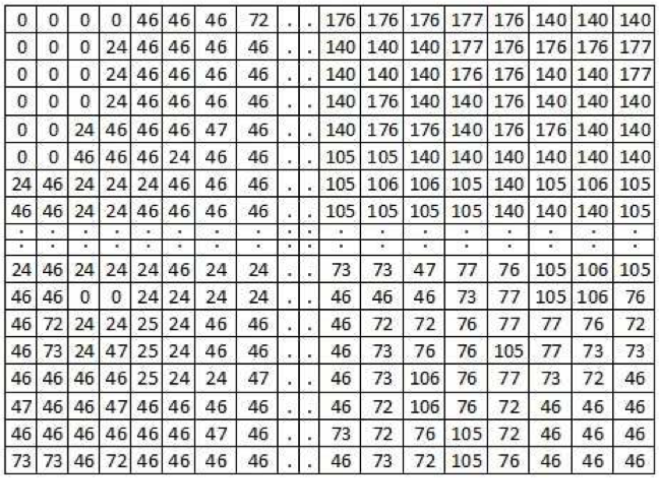
Where M_j^{-1} is the multiplicative inverse of K_{yj}.

Step 8. Now k numbers of P_i are ORed to generate a single secret byte (SB) and placed in PS position using Algorithm 3 (PS generation will be same as to the sharing phase).

Step 9. Secret image is reconstructed in a lossless manner by reconstruction of L numbers of secret bytes.

Algorithm 3 Reconstructed secret byte (SB) writing

```
$PS = K_y[i\%16] \times K_y[(i+1)\%16] + PS$;
$PS = PS\%L$;
while $(Index[PS]! = 0)$ do
    $PS = (PS + 1)\%L$;
end while
Write the secret byte (SB) in $PS^{th}$ position;
$Index[PS] = 1$; // 0 for unread and 1 for read
End
```



0	0	0	0	46	46	46	72	.	.	176	176	176	177	176	140	140	140
0	0	0	24	46	46	46	46	.	.	140	140	140	177	176	176	176	177
0	0	0	24	46	46	46	46	.	.	140	140	140	176	176	140	140	177
0	0	0	24	46	46	46	46	.	.	140	176	140	140	176	140	140	140
0	0	24	46	46	46	47	46	.	.	140	176	176	140	176	176	140	140
0	0	46	46	46	24	46	46	.	.	105	105	140	140	140	140	140	140
24	46	24	24	24	46	46	46	.	.	105	106	106	105	140	105	106	105
46	46	24	24	46	46	46	46	.	.	105	105	105	105	140	140	140	105
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
24	46	24	24	24	46	24	24	.	.	73	73	47	77	76	105	106	105
46	46	0	0	24	24	24	24	.	.	46	46	46	73	77	105	106	76
46	72	24	24	25	24	46	46	.	.	46	72	72	76	77	77	76	72
46	73	24	47	25	24	46	46	.	.	46	73	76	76	105	77	73	73
46	46	46	46	25	24	24	47	.	.	46	73	106	76	77	73	72	46
47	46	46	47	46	46	46	46	.	.	46	72	106	76	72	46	46	46
46	46	46	46	46	46	47	46	.	.	73	72	76	105	72	46	46	46
73	73	46	72	46	46	46	46	.	.	46	73	72	105	76	46	46	46

Figure 3: Secret Image: Size (61 × 52)

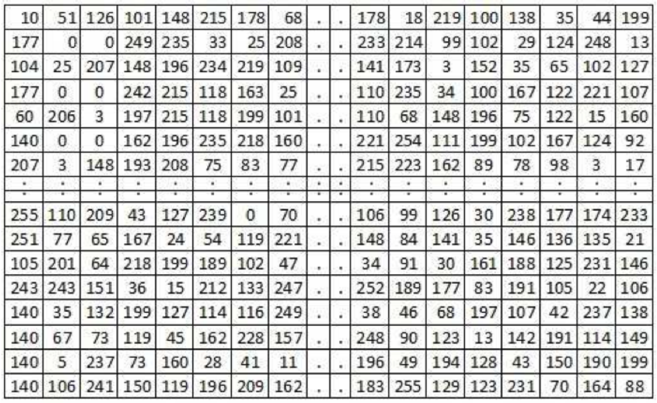
4.4 Example

Here we discuss the sharing phase of our proposed scheme using an example. Consider an image of height and width are 52 and 61 respectively (See Figure 3).

Here we use (3, 5)-threshold sharing scheme with the key $K_y = \{42, 74, 98, 119, 50, 68, 47, 180, 137, 245, 201, 168, 67, 254, 105, 254\}$.

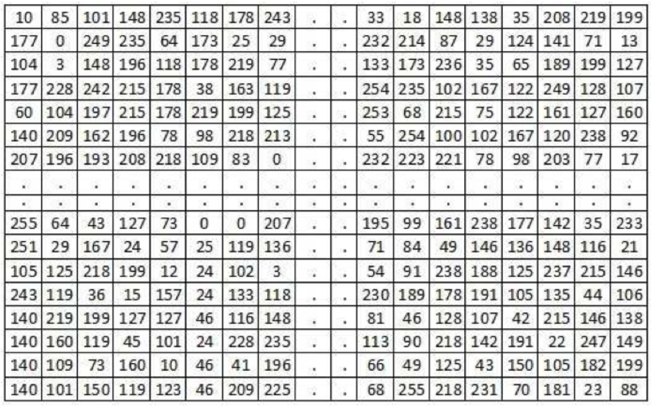
Now select a specific secret byte from a random field among 3172 bytes (61×52), which will be selected depending upon the K_y using Equation (2). Figures 4, 5, and 6 show different share after applying intermediate operation. These figures show each share contains partial confused secret information. That provides additional protection of the secret data. Now generate matrix (A), by taking first k number of nonzero fixed sample values from n shares. Here we consider first non-zero bytes. One can take any non-zero sample bytes from encrypted shares, but it should be same for both sharing and reconstruction phases.

$$A = \begin{array}{c} I_1^{'} \\ I_2^{'} \\ I_3^{'} \\ I_4^{'} \\ I_5^{'} \end{array} \left[\begin{array}{ccc} 10 & 51 & 126 \\ . & . & . \\ 10 & 85 & 101 \\ . & . & . \\ 126 & 85 & 148 \end{array} \right]$$



10	51	126	101	148	215	178	68	.	.	178	18	219	100	138	35	44	199
177	0	0	249	235	33	25	208	.	.	233	214	99	102	29	124	248	13
104	25	207	148	196	234	219	109	.	.	141	173	3	152	35	65	102	127
177	0	0	242	215	118	163	25	.	.	110	235	34	100	167	122	221	107
60	206	3	197	215	118	199	101	.	.	110	68	148	196	75	122	15	160
140	0	0	162	196	235	218	160	.	.	221	254	111	199	102	167	124	92
207	3	148	193	208	75	83	77	.	.	215	223	162	89	78	98	3	17
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
255	110	209	43	127	239	0	70	.	.	106	99	126	30	238	177	174	233
251	77	65	167	24	54	119	221	.	.	148	84	141	35	146	136	135	21
105	201	64	218	199	189	102	47	.	.	34	91	30	161	188	125	231	146
243	243	151	36	15	212	133	247	.	.	252	189	177	83	191	105	22	106
140	35	132	199	127	114	116	249	.	.	38	46	68	197	107	42	237	138
140	67	73	119	45	162	228	157	.	.	248	90	123	13	142	191	114	149
140	5	237	73	160	28	41	11	.	.	196	49	194	128	43	150	190	199
140	106	241	150	119	196	209	162	.	.	183	255	129	123	231	70	164	88

Figure 4: $I_1^{'}$: Size (37×52)



10	85	101	148	235	118	178	243	.	.	33	18	148	138	35	208	219	199
177	0	249	235	64	173	25	29	.	.	232	214	87	29	124	141	71	13
104	3	148	196	118	178	219	77	.	.	133	173	236	35	65	189	199	127
177	228	242	215	178	38	163	119	.	.	254	235	102	167	122	249	128	107
60	104	197	215	178	219	199	125	.	.	253	68	215	75	122	161	127	160
140	209	162	196	78	98	218	213	.	.	55	254	100	102	167	120	238	92
207	196	193	208	218	109	83	0	.	.	232	223	221	78	98	203	77	17
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
255	64	43	127	73	0	0	207	.	.	195	99	161	238	177	142	35	233
251	29	167	24	57	25	119	136	.	.	71	84	49	146	136	148	116	21
105	125	218	199	12	24	102	3	.	.	54	91	238	188	125	237	215	146
243	119	36	15	157	24	133	118	.	.	230	189	178	191	105	135	44	106
140	219	199	127	127	46	116	148	.	.	81	46	128	107	42	215	146	138
140	160	119	45	101	24	228	235	.	.	113	90	218	142	191	22	247	149
140	109	73	160	10	46	41	196	.	.	66	49	125	43	150	105	182	199
140	101	150	119	123	46	209	225	.	.	68	255	218	231	70	181	23	88

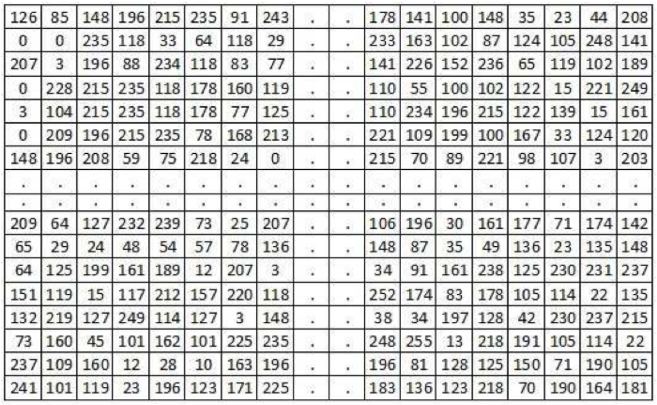
Figure 5: $I_3^{'}$: Size (37×52)

Now generate a header structure as Table 1, here key part contains our encryption key.

$$A = \begin{bmatrix} 10 & 51 & 126 \\ ... & ... & ... \\ 10 & 85 & 101 \\ ... & ... & ... \\ 126 & 85 & 148 \end{bmatrix} \times \begin{bmatrix} 5 \\ 3 \\ 42 \end{bmatrix}$$

$$= \begin{bmatrix} (5495) \\ ... \\ (4547) \\ ... \\ (7101) \end{bmatrix} \Rightarrow \begin{bmatrix} (0 & 21 & 119) \\ & ... & \\ (0 & 17 & 195) \\ & ... & \\ (0 & 27 & 189) \end{bmatrix}$$

Therefore shared header information is as Table 2.



126	85	148	196	215	235	91	243	.	.	178	141	100	148	35	23	44	208
0	0	235	118	33	64	118	29	.	.	233	163	102	87	124	105	248	141
207	3	196	88	234	118	83	77	.	.	141	226	152	236	65	119	102	189
0	228	215	235	118	178	160	119	.	.	110	55	100	102	122	15	221	249
3	104	215	235	118	178	77	125	.	.	110	234	196	215	122	139	15	161
0	209	196	215	235	78	168	213	.	.	221	109	199	100	167	33	124	120
148	196	208	59	75	218	24	0	.	.	215	70	89	221	98	107	3	203
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
209	64	127	232	239	73	25	207	.	.	106	196	30	161	177	71	174	142
65	29	24	48	54	57	78	136	.	.	148	87	35	49	136	23	135	148
64	125	199	161	189	12	207	3	.	.	34	91	161	238	125	230	231	237
151	119	15	117	212	157	220	118	.	.	252	174	83	178	105	114	22	135
132	219	127	249	114	127	3	148	.	.	38	34	197	128	42	230	237	215
73	160	45	101	162	101	225	235	.	.	248	255	13	218	191	105	114	22
237	109	160	12	28	10	163	196	.	.	196	81	128	125	150	71	190	105
241	101	119	23	196	123	171	225	.	.	183	136	123	218	70	190	164	181

Figure 6: $I_5^{'}$: Size (37×52)

Table 2: The shared header information

	$[S_n]$	[Shared Header]
H_1	1	(0 21 119) - - -
	- - -	- - -
H_3	3	(0 17 195) - - -
	- - -	- - -
H_5	5	(0 27 189) - - -

Now, actual shares after concatenation of I_i', V_i are as follows.

$$\begin{aligned} V_1 &= I_1', H_1. \quad \text{Size}(37 \times 53) \\ V_2 &= I_2', H_2. \quad \text{Size}(37 \times 53) \\ &\vdots \quad \vdots \\ V_5 &= I_5', H_5. \quad \text{Size}(37 \times 53). \end{aligned}$$

V_1, V_2, $\cdots$, V_5 are the complete shares for transmission. It shows, if and only if k number of shares are came together, then reconstruction is possible, otherwise reconstructed data will be completely different from original.

5 Analysis of the Protocol

In our algorithm for n shares with threshold k size of each mask is C_{k-1}^{n} where we have C_{k-2}^{n-1} zero and C_{n-k}^{n-1} ones. Then each share contains C_{n-k}^{n-1} number of bytes for C_{k-1}^{n} number of bytes of secret image. So percentage of information contain in each share is $(C_{n-k}^{n-1}/C_{k-1}^{n}) \times 100$. This clearly indicates that higher the number of shares and higher the threshold value, i.e. nearer to the number of shares lesser the content of information in each share.

Table 3 shows percentage of information in each share of proposed scheme and other schemes. Next during logical ANDing of the mask with secret image, the bytes of the image corresponding to the one bits in the mask are retained and the zero bytes corresponding to zero bit in the mask will be collapsed, which finally produces a set

Table 3: Percentage of information in each share for different k, n

n	k	C_{n-k}^{n-1}	C_{k-1}^{n}	Our Scheme	[16]	[2]	[1]
8	8	1	8	12.5	100	100	100
8	7	7	28	25	100	100	100
8	6	21	56	37.5	100	100	100
8	5	35	70	50	100	100	100
6	5	5	15	33.3	100	100	100
6	4	10	20	50	100	100	100
5	3	6	10	60	100	100	100

of scrambled bytes; effectively the retained information in compressed and thus being further ciphered.

Next the scrambled bytes are further ciphered using modulo multiplication techniques with the MD5 digest of the key given at the time of transmission (key). It may be noted that instead of using the key directly we have used the digest of the key for encryption, then the size of the key is immaterial which provides additional strength against cryptanalysis for the key.

Finally here the key may be the session key, i.e. the key will be varied with every transmission and our algorithm is free from key distribution hazard as the key itself is further shared in the secret shares.

Thus from individual shares there is hardly any leakage of information vis-a-vis unless minimal number of untampered share which is not tampered is collected nothing is revealed. Only when minimal number of valid share is collected one can form the key and get the information about total number of shares created. After knowing number of shares and the threshold (which is known) one can form the mask and from the key one can get the digest. From the digest using multiplicative inverse we get the compressed shares and the actual shares using the masks. Finally ORing the shares we get the original secret.

6 Analysis of Compression

All masking pattern has equal number of zeros with different distribution only. In every share we collapse all zero bytes corresponding to zero bit in the corresponding mask. It may be noted that as k is closer to n, more is compression, i.e. maximum for $k = n$.

Next for lossless expanding, knowing n and k we can redesign all n masks using our original mask generation algorithm. According to the share number of the share holder appropriate mask is used to expand the secret share by inserting zero bytes corresponding to zero bit in the corresponding mask.

In our example of (3, 5) the mask size is of 10 bits and every mask has 4 zeroes, thus every secret can be compressed by approximately 40 percent, obviously the compression varies with (k, n). In case of an example of (5, 6) the mask size is 15 bits and every mask has 10 zeroes, thus compression will be 66.6 % (See Table 4 and Figures 7 and 8).

Table 4: Compression rate for different k and n

Threshold (k)	Length of Masking Pattern	Number of zero in masking pattern	Approximate Compression Rate (percent)
2	5	1	20
3	10	4	40
4	10	6	60
5	5	4	80

Total number of Shares (n) = 5.

Threshold (k)	Length of Masking Pattern	Number of zero in masking pattern	Approximate Compression Rate (percent)
2	6	1	16
3	15	5	33
4	20	10	50
5	15	10	66
6	6	5	83

Total number of Shares (n) = 6.

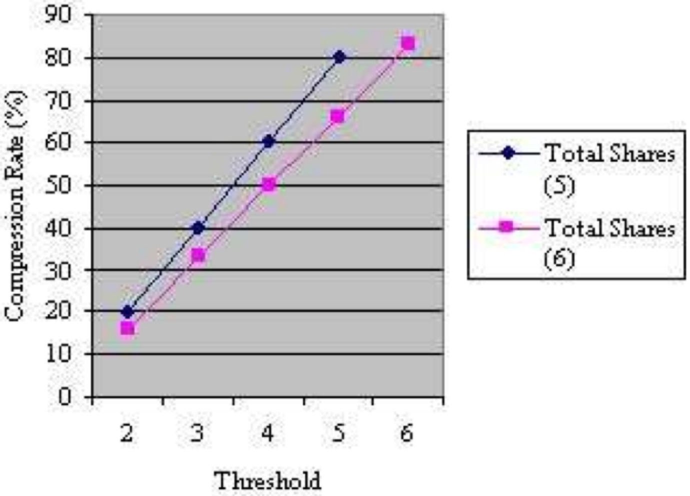
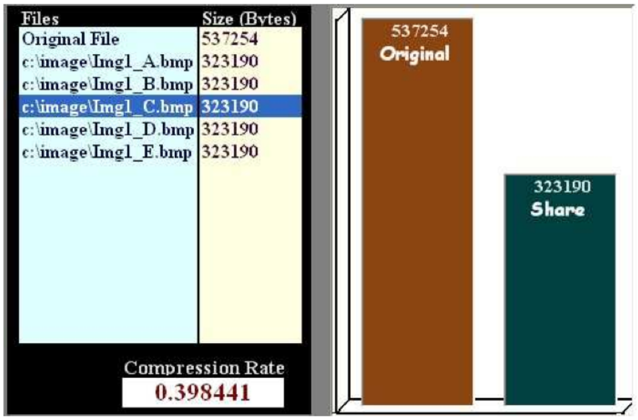


Figure 7: Threshold vs. compression ratio

7 Experimental Result

7.1 Experimental Result for 24-bit bmp Image

Figure 9 shows a secret image (Figure 9(a)) is shared among 5 participants by the user given key (UK_y) "2936451090872310". Here threshold value is 3. At the reconstruction phase, if we collect only 3 or more shares



Files	Size (Bytes)
Original File	537254
c:\image\Img1_A.bmp	323190
c:\image\Img1_B.bmp	323190
c:\image\Img1_C.bmp	323190
c:\image\Img1_D.bmp	323190
c:\image\Img1_E.bmp	323190

Compression Rate
0.398441

Figure 8: Compression rate

then reconstructed secret is lossless, but less than 3 shares are not sufficient to reconstruct the secret data. In the time of reconstruction no need to remember the key, because key as well as secret data is shared among set of participants.

7.2 Experimental Result for Gray Image

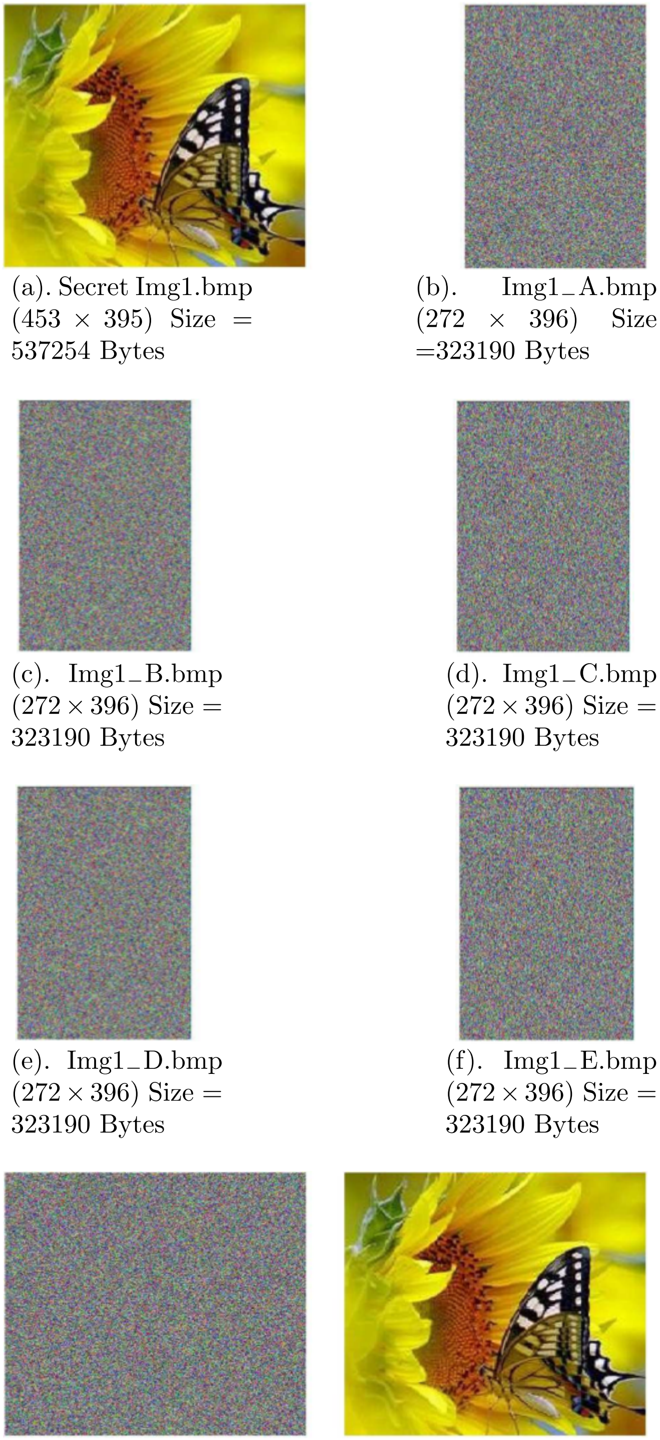
Figure 10 shows a secret image (Figure 10(a)) of size 181200 bytes is shared among 5 participants by the user given key "2936451290874310" and the value of k is 3. Here generated shares size are less than secret image and which hold the partial secret information. That provides an additional protection of secret image.

8 Strength and Security Analysis

A secure shared cryptography algorithm should be robust against all types of attacks such as cryptanalytic, statistical. Here we discuss the security analysis of the proposed algorithm by addressing key sensitivity analysis and different statistical analysis. The resistance against different types of attack is useful measure of the performance of a cryptosystem. Therefore some security analysis results are incorporated in the following section to prove the validity of our proposed scheme.

8.1 Key Sensitivity Analysis

In our scheme, shared data is highly sensitive to the secret key. Here user given variable length key (UK_y) is converted as a fixed length key (16 bytes) using MD5 hash function. Here we use MD5, but one can use any hash function or random number generator. Now this fixed length key is used as encryption key (K_y). Generated shares are varied for a single bit/byte changes in the key, because secret bytes are selected randomly from secret field depending upon the key and secret bytes are also encrypted using the key. The merits of our scheme is that key based shared cryptography that introduce the



(a). Secret Img1.bmp (453×395) Size = 537254 Bytes

(b). Img1_A.bmp (272×396) Size =323190 Bytes

(c). Img1_B.bmp (272×396) Size = 323190 Bytes

(d). Img1_C.bmp (272×396) Size = 323190 Bytes

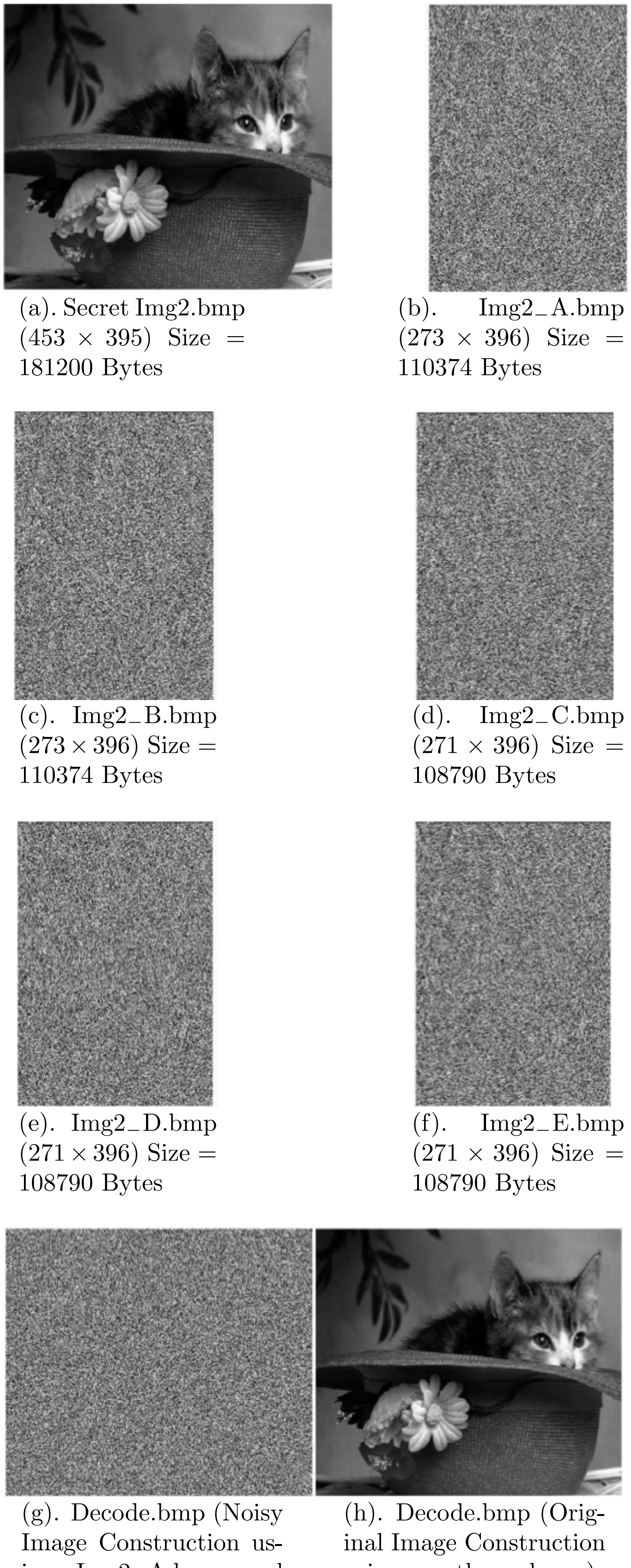
(e). Img1_D.bmp (272×396) Size = 323190 Bytes

(f). Img1_E.bmp (272×396) Size = 323190 Bytes

(g). Decode.bmp (Noisy Image Construction using Img1_B.bmp and Img1_E.bmp)

(h). Decode.bmp (Original Image Construction using any three shares)

Figure 9: (3, 5)-Sharing and Reconstruction for 24-bit image



(a). Secret Img2.bmp (453 × 395) Size = 181200 Bytes

(b). Img2_A.bmp (273 × 396) Size = 110374 Bytes

(c). Img2_B.bmp (273 × 396) Size = 110374 Bytes

(d). Img2_C.bmp (271 × 396) Size = 108790 Bytes

(e). Img2_D.bmp (271 × 396) Size = 108790 Bytes

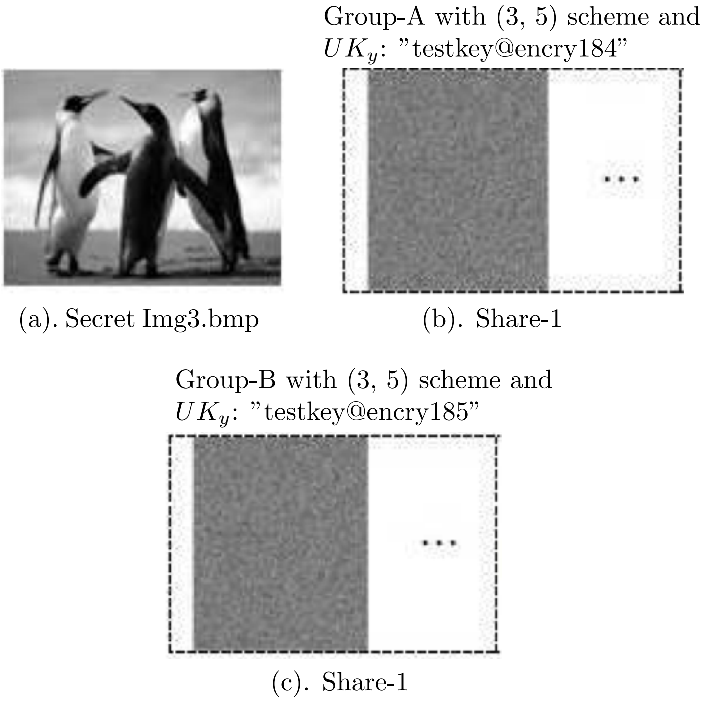
(f). Img2_E.bmp (271 × 396) Size = 108790 Bytes

(g). Decode.bmp (Noisy Image Construction using Img2_A.bmp and Img2_E.bmp)

(h). Decode.bmp (Original Image Construction using any three shares)

Figure 10: (3, 5)-Sharing and Reconstruction for gray scale image

concept of avalanche effect and no need to remember the key that overcomes the concept of single point failure, which provides additional protection to the secret data.



Group-A with (3, 5) scheme and UK_y: ”testkey@encry184”

(a). Secret Img3.bmp

(b). Share-1

Group-B with (3, 5) scheme and UK_y: ”testkey@encry185”

(c). Share-1

Figure 11: Two groups with same secret and different keys

Figure 11 shows a secret image is shared between two groups with different keys using (3, 5) scheme. Among five shares, only first share is shown for each group. Last character indicates the difference between two keys. Here two keys have single byte difference. Correlation value between Figure 11(b) (first share from group-A) and Figure 11(c) (first share from group-B) is 0.0018. This value (closer to zero) indicates two shared images are completely different and there is no such statistical relation. This part tuned an additional protection of secret data. Only when qualified a set of legitimate shares comes together, then reconstruction is possible. Figure 12 shows that a secret data is shared ((3, 5) scheme) among two groups by two different keys. In reconstruction phase, 3 selected shares are A.1, A.3 and A.5 (shares belong to same group) then S is reconstructed whereas if collected shares are 3 but taking from two groups, i.e. A.2, B.3 and A.5, at this situation reconstruction is not possible.

Figure 13 shows collision free random index positions for selecting secret bytes from secret field of size 20 bytes with two users given keys ’testkey@encry184’ and ’testkey@encry185’.

8.2 Statistical Analysis

Statistical analysis is crucial importance for a cryptosystem. An ideal cryptosystem should be resistive against

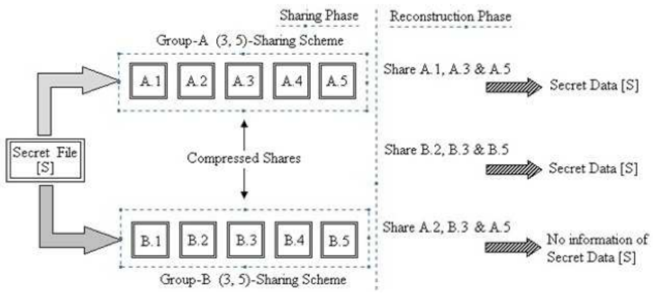


Figure 12: Reconstruction of a Secret file using qualified set of legitimate shares

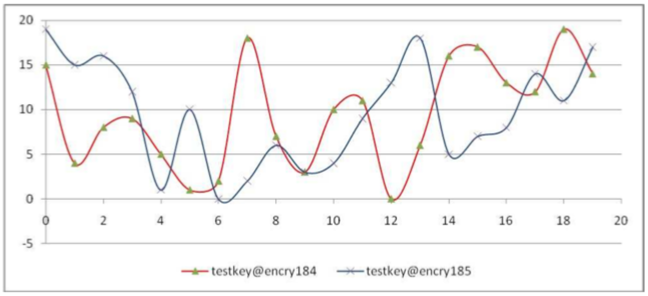


Figure 13: Random index positions for two keys

any statistical attack. To prove the robustness of the proposed algorithm, we have performed the following statistical test such as histogram analysis, correlation analysis, etc.

8.2.1 Histogram Analysis

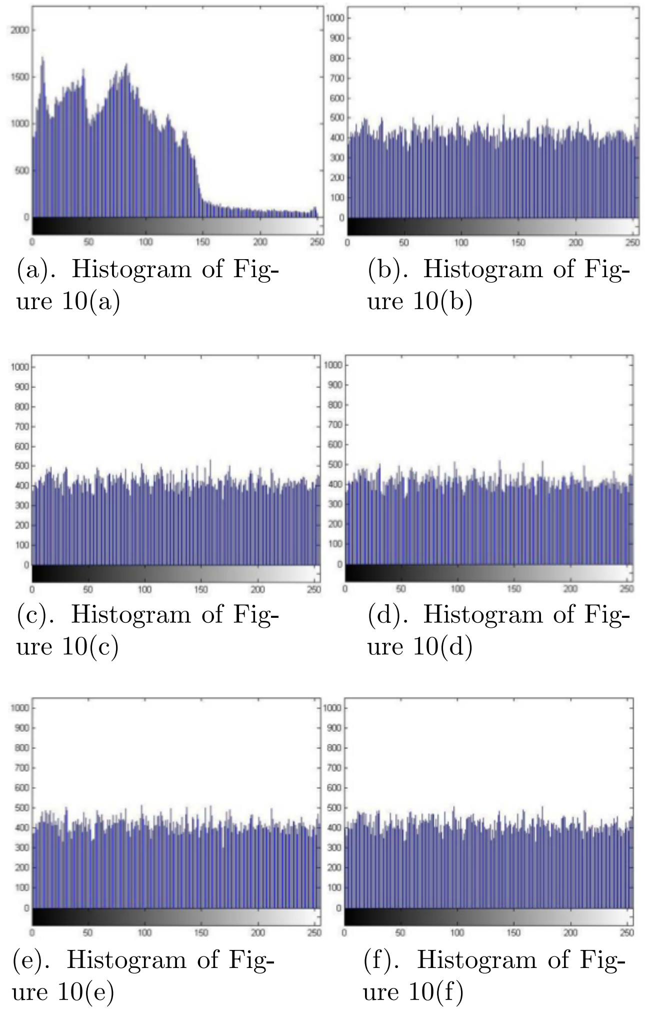
The histogram analysis clarifies how pixels in an image are distributed by plotting the number of pixels at each intensity level. Histogram analysis of gray scale secret image (SecretImg2.bmp) with respect to shared images is shown in Figure 14. The histogram of shared images has uniform distribution which is significantly different from the original image and has no statistical similarity in appearance (See Figure 14).

8.2.2 Correlation Value

A secret shared cryptography scheme must generate shared images independent of the original secret image. Therefore, they must have a very low correlation coefficient value. Here, we have calculated the correlation between the shares. Also the correlation between original and reconstructed image is shown in Table 5. The correlation value is calculated using Equation (6).

$$r = \frac{\sum_m \sum_n (A_{mn} - A')(B_{mn} - B')}{\sqrt{(\sum_m \sum_n (A_{mn} - A')^2 \sum_m \sum_n (B_{mn} - B')^2)}}, \quad (6)$$

where $A^{'}$ and $B^{'}$ are mean of A and B respectively. A low value of correlation coefficient shows that there is no straight relation between the original and encrypted images. Here generated shared images are compressed. So it is impossible to show the correlation value between secret image and shared images. Following table (Table 5) shows the correlation value among shared images and secret image and reconstructed images.



(a). Histogram of Figure 10(a)

(b). Histogram of Figure 10(b)

(c). Histogram of Figure 10(c)

(d). Histogram of Figure 10(d)

(e). Histogram of Figure 10(e)

(f). Histogram of Figure 10(f)

Figure 14: Histogram of Secret image (Figure 10(a)) and Shared images

Table 5: Correlation value

SL No.	Images		Correlation value
1	Figure 10(b) and (c)	:	0.5125
2	Figure 10(d) and (e)	:	0.5057
3	Figure 10(d) and (f)	:	0.1764
4	Figure 10(e) and (f)	:	0.5058
5	Figure 10(a) and (g)	:	-0.0030
6	Figure 10(a) and (h)	:	1

In the above table the 6^{th} entry shows the correlation value between secret image (Figure 10(a)) and reconstructed image (Figure 10(h)) using 3 shares and the value is one, so the reconstruction is lossless. The 5^{th} entry shows a value close to zero for the correlation between secret image (Figure 10(a)) and reconstructed image (Figure 10(g)) using 2 shares.

8.2.3 MSE and PSNR Measure

The Mean Square Error (MSE) and Peak Signal to Noise Ratio (PSNR) for the proposed technique have been computed for different images. The high value of MSE and low value of PSNR cause the resulting encrypted image more randomness. MSE is calculated using the formula

$$MSE = (\Sigma_{i-1}^{N}\Sigma_{j=1}^{M}[C(i,j) - C^{'}(i,j)]^2)/MN,$$

where, $c(I,j)$ and $c^{'}(I,j)$ are the i^{th} row and j^{th} column pixel of two images C and $C^{'}$, respectively. M and N are the number of rows and columns of an image. PSNR can be computed by

$$PSNR = 10 \times log_{10}[R^2/MSE],$$

where R is 255 as grey image has been used in this experiment. Calculated results of MSE and PSNR are tabulated in Table 6.

Table 6: MSE and PSNR value

SL No.	Images		MSE	PSNR
1	Figure 10(b) and (c)	:	115.6822	27.4981
2	Figure 10(d) and (e)	:	117.2003	27.4415
3	Figure 10(d) and (f)	:	194.8199	25.2345
4	Figure 10(e) and (f)	:	117.2881	27.4383
5	Figure 10(a) and (g)	:	233.7130	24.4437
6	Figure 10(a) and (h)	:	0	Infinity

High value MSE and low value PSNR indicate that two images are completely different. On the other hand, the high value of PSNR indicates the high quality image.

9 Conclusions

This paper shows a secured key based secret sharing scheme where key as well as secret data is shared among set of participants. Here image is selected as a secret data, although proposed scheme is strongly applicable for other digital data, such as text, audio, etc. In the sharing phase all secret bytes are selected randomly from secret field depending upon the key and each generated share holds partial secret information in scrambled and encrypted form. That provides additional protection of the secret image and also reduces the bandwidth required for transmission.

In our scheme if and only if numbers of collecting shares are equal to k or more and none of the share is tampered, then only the original secret image is reconstructed; otherwise reconstructed image will be completely ciphered, because fewer shares cannot reconstruct the original header, thus we cannot have either right key (K_y) or the information to construct the correct masking pattern. So our proposed scheme can claim to be a Perfect Secret Sharing (PSS) Scheme. Not only that, if a legitimate group of threshold number of shares comes together (i.e. shares from different groups cannot mix as keys are different), then only the original secret is reconstructed. Moreover, key sensitivity analysis and statistical analysis prove the high acceptability of the proposed algorithm.

Acknowledgments

We are thankful to the department of Computer Science & Engineering of MCKV Institute of Engineering, Regent Education & Research Foundation and Jadavpur University, Kolkata for giving us the platform for planning and developing this work in departmental laboratories.

References

[1] C. Asmuth and J. Bloom, "A modular approach to key safeguarding," *IEEE Transaction on Information Theory*, vol. 29, no. 2, pp. 208–210, 1983.

[2] G. R. Blakley, "Safeguarding cryptographic keys," in *Proceedings of AFIPS International Workshop on Managing Requirements Knowledge*, pp. 313, 1979.

[3] K. Y. Chao and J. C. Lin, "Secret image sharing: a boolean-operations based approach combining benefits of polynomial-based and fast approaches," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 23, no. 2, pp. 263–285, 2009.

[4] Y. Desmedt, "Some recent research aspects of threshold cryptography," in *Proceedings of 1st International Information Security Workshop (ISW'97)*, pp. 158–173, Ishikawa, Japan, 1997.

[5] Y. Desmedt and Y. Frankel, "Shared generation of authenticators and signatures," in *Advances in Cryptolog (RYPTO'91)*, LNCS 576, pp. 457–469, Springer Verlag, 1992.

[6] Y. Desmedt and Y. Frankel, "Threshold cryptosystems," in *Proceedings on Advances in Cryptology (CRYPTO'89)*, LNCS 435, pp. 307–315, 1990.

[7] Y. Desmedt and Y. Frankel, "Homomorphic zero knowledge threshold schemes over any finite abelian group," *SIAM Journal on Discrete Mathematics*, vol. 7, no. 4, pp. 667–675, 1994.

[8] L. Dong and M. Ku, "Novel (n,n) secret image sharing scheme based on addition," in *Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP'10)*, pp. 583–586, 2010.

[9] L. Dong, D. Wang, M. Ku, and Y. Dai, "$(2,n)$ secret image sharing scheme with ideal contrast," in *International Conference on Computational Intelligence and Security (CIS'10)*, pp. 421–424, 2010.

[10] H. F. Huang and C. C. Chang, "A novel efficient (t, n) threshold proxy signature scheme," *Information Sciences*, vol. 176, no. 10, pp. 1338–1349, 2006.

[11] E. D. Karnin, J. W. Greene, and M. E. Hellman, "On secret sharing systems," *IEEE Transactions on Information Theory*, vol. IT-29, no. 1, pp. 35–41, 1983.

[12] P. K. Naskar, A. Chaudhuri, D. Basu, and A. Chaudhuri, "A novel image secret sharing scheme," in *Second International Conference on Emerging Applications of Information Technology (EAIT'11)*, pp. 177–180, 2011.

[13] P. K. Naskar, H. N. Khan, U. Roy, A. Chaudhuri, and A. Chaudhuri, "Secret image sharing with embedded session key," in *Computer Information Systems Analysis and Technologies (CISIM'11)*, Communications in Computer and Information Science, vol. 245, pp 286-294, 2011.

[14] P. K. Naskar, H. N. Khan, U. Roy, A. Chaudhuri, and A. Chaudhuri, "Shared cryptography with embedded session key for secret audio," *International Journal of Computer Applications*, vol. 26, no. 8, pp. 5–9, 2011.

[15] A. De Santis, Y. Desmedt, Y. Frankel, and M. Yung, "How to share a function securely?," in *Proceedings of the Twenty-Sixth Annual ACM Symposium on Theory of Computing (STOC'94)*, pp. 522–533, 1994.

[16] A. Shamir, "How to share a secret?," *Communications of the ACM*, vol. 22, no. 11, pp. 612–613, 1979.

[17] V. Shoup, "Practical threshold signatures," in *Proceedings of Eurocrypt'00*, LNCS 1807, pp. 207–220, Springer-Verlag, 2000.

[18] C. C. Thien and J. C. Lin, "Secret image sharing," *Computers and Graphics*, vol. 26, no. 5, pp. 765–770, 2002.

Prabir Kr. Naskar, B.Tech from Govt. College of Engineering & Leather Technology (WBUT), West Bengal, India and M.Tech from Jadavpur University, West Bengal, India, is presently working as Assistant Professor in the Department of Computer Science & Engineering, MCKV Institute of Engineering, West Bengal, India. Currently he is doing his research work at Jadavpur University, West Bengal, India. His current research interests include: cryptography, information sharing, steganography, watermarking and image processing.

Hari Narayan Khan is an Assistant Professor in the Department of Computer Science & Engineering, Regent Education & Research Foundation, West Bengal University of Technology, Kolkata, India and presently pursuing his research work at Jadavpur University, Kolkata, West Bengal, India. He completed his M.Tech in Computer Technology at Jadavpur University, Kolkata, India. He completed his B.Tech in Electronics & Communication Engineering from Institute of Technology & Marine Engineering under West Bengal University of Technology, Kolkata, India. His research interest includes cryptography, network security, information sharing, steganography and watermarking.

Prof. Atal Chaudhuri, B.E., M.E. & PhD from Jadavpur University, West Bengal, India, is working in the Department of Computer Science & Engineering, Jadavpur University, West Bengal, India for last 29 years. His current research interests include: embedded system, cryptography, information sharing, steganography, watermarking and data mining.

An Improved RSA-based Certificateless Signature Scheme for Wireless Sensor Networks

Gaurav Sharma, Suman Bala, and Anil K. Verma
(Corresponding author: Gaurav Sharma)

Computer Science and Engineering Department, Thapar University, Patiala 147004, India
(Email: gaurav.sharma@thapar.edu)

(Received Aug. 15, 2013; revised and accepted Jan. 7 & Apr. 28, 2014)

Abstract

The entire world is looking to fulfill the need of the hour in terms of security. Certificateless cryptography is an efficient approach studied widely due to two reasons: first, it eliminates the need of certificate authority in public key infrastructure and second, it can resolve key escrow problem of ID-based cryptography. Recently, Zhang et al. proposed a novel security scheme based on RSA, applicable to real life applications but could not cope up with the well defined attacks. This paper, presents an RSA-based CertificateLess Signature (RSA-CLS) scheme applicable to wireless sensor networks. The security of RSA-CLS is based on the hardness assumption of Strong RSA. The scheme is proven to be secure against Type I and Type II attack in random oracle model.

Keywords: Certificateless public key cryptography, digital signature, RSA

1 Introduction

Express progress in the field of technology made it feasible to foster wireless sensor networks technology [1, 7, 21]. Wireless Sensor Networks (WSN) are comprised of large number of tiny sensor nodes with constrained resources in terms of processing power, energy and storage. WSN can be used in various applications mainly environmental monitoring, medical, military, and agriculture [1]. Since, devices used in sensor networks is not tamper resistant, so adversary can gain its physical access easily. Hence, the main objective is to protect the data from unauthorized access, which can be done by using some security mechanisms [6, 9, 29]. The technology faces lots of security problems as it has a wireless mode of communication and access to such sensor devices is quite easy. There are two approaches to restrict the unauthorized access in to the network: symmetric cryptography [39] and asymmetric cryptography [30]. Initially, symmetric algorithms are preferable to asymmetric algorithms, as they are simple and less computational for 8-bit micro-controller. Symmetric algorithms need key pre-distribution. The problem with this approach is the number of keys stored in each sensor node and the network is not forward secure after the compromising of the key. Moreover, it causes greater configuration effort before deployment and generating ample traffic, thus consequently higher energy consumption [8]. As a result, researchers are redirecting their attention to asymmetric algorithms but asymmetric algorithms are very challenging for constrained resources in WSN.

In traditional Public Key Infrastructure (PKI), the user selects a public key but it needs to be validated by a trusted third party known as Certificate Authority (CA) [3]. The CA provides a digital certificate to tag the public key with the user's identity. PKI has a problem of high computation and storage. To avoid this, Shamir [24] introduced the concept of Identity-based Infrastructure. It allows the user to choose a public key of its own choice such as email-id, phone number, name, etc. and the private key is generated by trusted third party server, Private Key Generator (PKG) causing a key escrow problem. Then, Al-Riyami et al. [2] presented a novel approach to solve the key escrow problem familial to Identity-based Cryptography (IBC) and eliminated the use of certificates in traditional Public Key Cryptography (PKC) known as CertificateLess Public Key Cryptography (CL-PKC). In CL-PKC, the trusted third party server, Key Generation Centre (KGC) generates a partial private key for the user wherein user's secret key and partial private key are used to generate public key of the user. In other words, CL-PKC differs from IBC in terms of arbitrary public key and when a signature is transmitted, user's public key is attached with it but not certified by any of the trusted authority. Thus, the KGC does not come to know the secret key of the user.

Thereafter, lots of CertificateLess Signature (CLS) schemes based on Discrete Logarithm Problem (DLP) have been presented and cryptanalyzed [12, 15, 17, 36]. Later, CLS schemes based on Elliptic Curve Discrete Logarithm Problem (ECDLP) has been presented and cryptanalyzed such as [10, 27, 28, 35, 40, 41]. Xu et al. [33, 34] presented two CLS schemes for emergency mobile wireless cyber-physical systems and mobile wireless cyber-physical systems respectively. But Zhang et al. [37] proved it insecure against public key replacement attack. Another, authenticated scheme for WSN was presented by Li et al. [19].

Since, the pairing operation is the most expensive operation among all, so there was a need to find the solution. In 2009, Wang et al. [31] presented a scheme which need not to compute the pairing at the sign stage, rather it precomputes and publishes as the system parameters. But, this is not the solution for the removal of pairing operation. In 2011, He et al. [13] developed an efficient short CLS scheme without pairing. After a while, few schemes have been presented and cryptanalyzed based on ECDLP without pairing [11, 26].

Another aspect of pairing free CLS scheme was presented by Zhang et al. [38] in 2012, based on Strong RSA assumption and proven to be secure against Super Type I adversary in random oracle model. But, proved insecure in [14, 25] independently against key replacement attack. Watro et al. [32] initiated the concept of RSA based cryptography in public-key based protocols for wireless sensor networks known as TinyPK. Bellare et al. [5] presented an Identity-Based Multi-Signature (IBMS) scheme based on RSA, which is secure under the one-wayness of RSA in the random oracle model.

This paper presents a new RSA-based Certificateless Signature (RSA-CLS) scheme for WSN based Strong RSA assumption and proven to be secure against Super Type I and Super Type II attacks in random oracle model.

1.1 Organization of the Paper

The rest of the paper is organized as follows: In the next section, we will discuss motivation and our contribution. Section 2 describes preliminaries, which includes complexity assumptions, formal model and security model of CLS scheme. Section 3, describes the proposed RSA-based CertificateLess Signature (RSA-CLS) scheme. Section 4 discusses about the analysis of the proposed RSA-CLS scheme, including security proofs against Type I and Type II adversary in the random oracle model and performance analysis w.r.t WSN followed by conclusion.

1.2 Motivation and Our Contribution

The real truth is far from imagination, i.e. there are many theories proposed for secure transmission. At present most of the theories are on paper but far from the real application. RSA has been implemented already in various applications like WSN, cloud computing etc. So, it would be preferable to upgrade the existing system rather implementing a new system. In present scenario, CL-PKC is the most convincing approach to provide secure communication. The main benefits of RSA based CertificateLess Signature (RSA-CLS) scheme is to avoid pairing operations which is the most expensive operation for resource-constrained WSN. Zhang et al. [38] proposed a scheme and claimed that their scheme based on Strong RSA assumption, is: (i) more practical as far as industry standard goes, (ii) secure in random oracle model, (iii) more efficient than existing schemes as no pairing operation is involved, (iv) secure against Super Type I (discuss in Section 2.3) adversary [16] (which implies the security against Strong and Normal Type both) and left an open problem of designing of CLS scheme secure in standard model. Sharma et al. [25] and He et al. [14] independently found that the scheme [38] is not secure against key replacement attack. Sharma et al. [25] proved that the [38] is insecure against Strong Type I attack. In Strong Type I attack, the adversary has a privilege to choose a private key, and query the challenger to replace the public key and breach the security of the scheme. We have avoid such kind of attack in scheme [38], by modifying the value of $R_1 = H_0(ID)^{r_1}$ to $R_1 = x_{ID}^e H_0(ID)^{r_1}$ and the corresponding value of u_1.

2 Preliminaries

In this section, we briefly review some fundamental concepts akin to CLs, which includes formal model and security notions. We further state the hardness assumptions required in the proposed RSA-CLS scheme.

2.1 Complexity Assumptions

In this section, we describe the complexity assumptions which are requisite for the security proof of the proposed scheme. The security of our proposed signature scheme will be attenuated to the hardness of the Strong RSA Assumption [22] in the group in which the signature is constructed. We briefly review the definition of the Strong RSA Assumption and Discrete Logarithm Problem (DLP) [20]:

Definition 1. *(Strong RSA Assumption). Let $n = pq$ be an RSA-like modulus and let G be a cyclic subgroup of Z_n^* of order $\#G, \lceil \log_2(\#G) \rceil = l_G$. Given (n, e) and $z \in G$, the strong RSA problem consists of finding $u \in Z_n$ satisfying $z = u^e$ mod n.*

Definition 2. *(Discrete Logarithm Problem). Let $n = pq$ be a RSA modular number which satisfying $p = 2p' + 1$, $q = 2q' + 1$, $g \in Z_n^*$ is a generator of order $p'q'$, for given elements g, y, n, its goal is to compute the exponent x such that $y = g^x \bmod n$.*

2.2 Formal Model of Certificateless Signature Scheme

This section describes the formal model of a certificateless signature scheme, which consists of seven polynomial-time algorithms. These are:

Setup. This algorithm is run by the KGC to initialize the system. It takes as input a security parameter 1^k and outputs a list of system parameters *params* and the master secret key d. The system parameters *params* is public to all where as the master secret key d is known to KGC only.

Partial-Private-Key-Extraction. This algorithm is run by the KGC, takes the system parameters *params*, master secret key d, and an identity $ID \in \{0,1\}^*$ as input, and outputs the partial private key d_{ID}, which is sent to the user via a secure channel.

Set-Secret-Value. This is a probabilistic algorithm, run by the user. It takes the system parameters *params* and the user's identity ID as input and outputs a secret value x_{ID}.

Set-Private-Key. This is a deterministic algorithm, run by the user. It takes the system parameters *params*, a partial private key d_{ID}, and a secret value x_{ID} as inputs and outputs a full private key SK_{ID} .

Set-Public-Key. This is a deterministic algorithm, run by the user. It takes the system parameters *params*, the user's identity d_{ID}, and the private key $SK_{ID} = (d_{ID}, x_{ID})$ as inputs and outputs a public key PK_{ID}.

CL-Sign. This algorithm is run by the user, takes the system parameters *params*, the user's identity ID, and the private key SK_{ID} and a message M as input and outputs a correct certificateless signature δ on message M.

CL-Verify. This algorithm is run by the user, takes the system parameters *params*, the user's identity ID, public key PK_{ID}, message M, and the signature δ as input and outputs *true* if the signature is correct, or else *false*.

2.3 Security Models

As for security model [16], a CLS scheme is different from an ordinary signature scheme. Certificateless signature scheme is vulnerable to two types (Type I and Type II) of adversaries. The adversary $\mathcal{A}_I$ in Type I represents a normal third party attacker against the CLS scheme. That is, $\mathcal{A}_I$ is not allowed to access to the master key but $\mathcal{A}_I$ may request public keys and replace public keys with values of its choice. The adversary $\mathcal{A}_{II}$ in Type II represents a malicious KGC who generates partial private keys of users. The adversary $\mathcal{A}_{II}$ is allowed to have access to the master-key but not replace a public key.

3 Proposed RSA-based Certificateless Signature Scheme (RSA-CLS)

In this section, we describe the proposed certificateless signature scheme based on Strong RSA assumption. The scheme works as follows.

Setup: Given a security parameter 1^k as input, a RSA group (n, p, q, e, d) is generated, where p' and q' are two large prime numbers which satisfy $p = 2p' + 1$ and $q = 2q' + 1$, $n = pq$ is a RSA modular number, $e < \phi(n)$ is the public key of Key Generation Center (KGC) and satisfies $gcd(e, \phi(n) = 1$ and $ed \equiv 1 \bmod \phi(n)$, where $\phi(n)$ denotes the Euler Totient function. Choose two cryptographic hash functions H and H_0 which satisfy $H_0 : \{0,1\}^* \to Z_n^*$ and $H : Z_n^4 \times \{0,1\}^* \to \{0,1\}^l$, where l is a security parameter. The master secret key is d and the public parameters of system is $params = \{n, e, H, H_0\}$.

Partial-Key-Extract: For a user with identity $ID \in \{0,1\}^*$, KGC computes partial private key by using the master secret key as $d_{ID} = H_0(ID)^d \bmod n$.

Set-Secret-Value: Given *params* and an identity ID, the user randomly chooses $x_{ID} \in Z_{2^{|n|/2-1}}$, where $|n|$ denotes the binary length of n.

Set-Private-Key: Given the partial private key d_{ID} and the secret value x_{ID} of a user with identity ID, output $SK_{ID} = (x_{ID}, d_{ID})$.

Set-Public-Key: Given the partial private key d_{ID} and the secret value x_{ID} of a user with identity ID, output $PK_{ID} = H_0(ID)^{x_{ID}} \bmod n$.

Sign: Given a message m and system parameters *params*, a user with identity ID computes the following steps by using its private key.

1) Randomly choose two numbers r_1, $r_2 \in Z_{2^{|n|/2-1}}$.
2) Compute $R_1 = x_{ID}^e H_0(ID)^{r_1} \bmod n$ and $R_2 = H_0(ID)^{r_2} \bmod n$.
3) Compute $h = H(R_1, R_2, ID, PK_{ID}, m)$.
4) Set $u_1 = x_{ID} d_{ID}^{r_1 - h} \bmod n$ and $u_2 = r_2 - x_{ID}h$.
5) Finally, the resultant certificateless signature on message m is $\delta = (u_1, u_2, h)$.

Verify: Given a certificateless signature $\delta = (u_1, u_2, h)$ on message m, a verifier executes as follows:

1) Compute $R_1' = u_1^e H_0(ID)^h \bmod n$ and $R_2' = H_0(ID)^{u_2} PK_{ID}^h \bmod n$.
2) Accept, if and only if the following equation holds $h = H(u_1^e H_0(ID)^h \bmod n, H_0(ID)^{u_2} PK_{ID}^h \bmod n, ID, PK_{ID}, m)$.

Correctness: In the following, we show that our scheme is correct and satisfies completeness.

$$\begin{aligned}
& H(u_1^e H_0(ID)^h, H_0(ID)^{u_2} PK_{ID}^h \bmod n, ID, PK_{ID}, m) \\
= \; & H((x_{ID} d_{ID}^{r_1 - h})^e H_0(ID)^h, \\
& \quad H_0(ID)^{r_2 - x_{ID}h} PK_{ID}^h \bmod n, ID, PK_{ID}, m) \\
= \; & H(x_{ID}^e H_0(ID)^{d(r_1 - h)e} H_0(ID)^h, \\
& \quad H_0(ID)^{r_2 - x_{ID}h + x_{ID}h} \bmod n, ID, PK_{ID}, m) \\
= \; & H(x_{ID}^e H_0(ID)^{r_1}, H_0(ID)^{r_2} \bmod n, ID, PK_{ID}, m) \\
= \; & h.
\end{aligned}$$

4 Analysis of RSA-CLS Scheme

4.1 Security Analysis

In this section, we prove that the proposed scheme is secure against Type I and Type II adversaries defined in Section 2.3 in the random oracle model H_0 and H. The following theorems are provided for the security.

Theorem 1. *If there exists a Type I adversary $\mathcal{A}_I$ who can ask at most q_{H_0} and q_H* ***Hash*** queries *to random oracles H_0 and H, q_s* ***Sign*** queries*, q_{ppk}* ***Partial-Private-Key-Extract*** queries*, and q_p* ***Private-Key-Extract*** queries*, and can break the proposed scheme in polynomial time τ with success probability ϵ, then there exists an algorithm β that solves the RSA problem with advantage, $\eta > \frac{(q_s+1)(q_s+q_{H_0})\epsilon}{2^l q_H \tau(q_{ppk}+q_p+q_s+1)}$.*

Proof. If there exists an adversary $\mathcal{A}_I$, who can break the proposed certificateless signature scheme. Then, we can construct another adversary β, known as RSA adversary, such that β can use $\mathcal{A}_I$ as a black-box and solve the RSA problem. The aim is to find $y^e = z$, where $y \in Z_n^*$ in n RSA modulus and (n, e, z) is an instance of RSA problem.

Setup: The adversary β selects two hash functions H_0 and H as random oracle. d is the master secret key, which satisfies $ed \equiv 1 \bmod \phi(n)$ and is unknown to β. The system parameters (e, n) is public to all. The adversary β maintains three lists H_0-*list*, *H-list* and *KeyList*, which are initially empty. The adversary β sends (e, n, z, H_0, H) as a final output to the adversary $\mathcal{A}_I$.

Queries: At any time, $\mathcal{A}_I$ is allowed to access the following oracles in a polynomial number of times. Then, β simulates the oracle queries of $\mathcal{A}_I$ as follows:

1) **H_0-Hash Queries:** $\mathcal{A}_I$ can query the random oracle H_0 at any time with an identity ID. In response to these queries, β flips a biased $coin \in \{0, 1\}$ at random such that $Pr[coin = 0] = \rho$. Then, β randomly chooses $t_{ID} \in Z_n$ and compute $h^0_{ID} = z_{coin} t^e_{ID}$ and send it to $\mathcal{A}_I$. β add $(ID, h_0, t_{ID}, coin)$ to the H_0-*list*.

2) ***H*-Hash Queries:** $\mathcal{A}_I$ can query the random oracle H at any time with $h = H(R_1, R_2, ID, PK_{ID}, m)$. For each query $(R_1, R_2, ID, PK_{ID}, m)$, β first checks the *H-list*:

   a. If $(R_1, R_2, ID, PK_{ID}, m, h)$ exists in the *H-list*, then β sets $H(R_1, R_2, ID, PK_{ID}, m) = h$ and returns h to $\mathcal{A}_I$.

   b. Else, β randomly chooses $h \in Z_n^*$, and add the record $(R_1, R_2, ID, PK_{ID}, m, h)$ to the *H-list*. β sends h to $\mathcal{A}_I$ as the corresponding response.

3) **Partial-Private-Key-Extract Queries:** At any time, $\mathcal{A}_I$ can query the oracle by giving an identity ID. β outputs a symbol $\perp$ if ID has not been created. Else, if ID has been created and $coin = 0$, then β returns t_{ID} to the adversary $\mathcal{A}_I$. Otherwise, β returns failure and terminates the simulation.

4) **Public-Key-Request Queries:** At any time, $\mathcal{A}_I$ can query the oracle by giving an identity ID. β randomly chooses $x_{ID} \in Z_{2^{|n|/2-1}}$ and searches the H_0-*list* for $(ID, h^0_{ID}, t_{ID}, coin)$. Then, β adds $(ID, PK_{ID} = h^0_{ID}{}^{x_{ID}}, x_{ID}, coin)$ to *KeyList* and send PK_{ID} to $\mathcal{A}_I$.

5) **Private-Key-Extract:** For a given identity ID chosen by $\mathcal{A}_I$, β searches $(ID,\ h^0_{ID},\ t_{ID},$ $coin)$ in the H_0-*list*. If $coin = 1$, then β aborts it, else, β searches $(ID,\ PK = h^0_{ID}{}^{x_{ID}},\ x_{ID},$ $coin)$ in the *KeyList*. β return $SK_{ID} = (x_{ID},$ $t_{ID})$ to $\mathcal{A}_I$ as a final output.

6) **Public-Key-Replace Queries:** $\mathcal{A}_I$ can request a query to replace public key PK_{ID} of an identity ID with a new public key PK'_{ID} chosen by $\mathcal{A}_I$ itself. As a result, β replaces the original public key PK_{ID} with PK'_{ID} if ID has been created in the H_0-*list*. Otherwise, output $\perp$.

7) **Sign Queries:** For each query on an input (m, ID), output $\perp$ if ID has not been queried before. For any input (m, ID) with ID which has already been queried, β searches H_0-*list* and *KeyList* for $(ID, h^0_{ID}, t_{ID}, coin)$ and $(ID, PK_{ID}, x_{ID}, coin)$. If $coin = 0$, then β produces a certificateless signature δ on message m by the returned private key (x_{ID}, t_{ID}). Otherwise, β computes as follows:

   a. β randomly chooses $u_1 \in Z_n^*$, $h \in \{0, 1\}^l$, and $u_2 \in Z_{2^{|n|/2-1}}$.

   b. β computes $R_1 = u_1^e H_0(ID)^h$ and $R_2 = H_0(ID)^{u_2} PK_{ID}{}^h$, where PK_{ID} may be a replaced public key.

   c. β searches whether $(R_1, R_2, ID, PK_{ID}, m)$ exists in the *H-list*. If it exists, then abort it. Else, β sets $H(R_1, R_2, ID, PK_{ID}, m) = h$ and adds $(R_1, R_2, ID, PK_{ID}, m, h)$ in the *H-list*.

   d. The resultant signature $\delta = (u_1, u_2, h)$ is returned to $\mathcal{A}_I$.

Output: After all the queries, $\mathcal{A}_I$ outputs a forgery $(ID^\star, PK^\star_{ID}, m^\star, \delta^\star = (u_1{}^\star, u_2{}^\star, h^\star))$ and win this game. It must satisfy the following conditions:

1) If $\delta^\star$ is a valid forgery, then $h^\star = H(R_1{}^\star, R_2{}^\star, PK_{ID^\star}, ID^\star, m)$, which is in the *H-list*, where $R_1{}^\star = u_1{}^{\star e} H_0(ID^\star)^{h^\star}$ and $R_2{}^\star = H_0(ID^\star)^{u_2{}^\star} PK^{h^\star}_{ID^\star})$.

2) $coin^\star = 1$ of the record $(ID^\star, h^0_{ID^\star}, t_{ID^\star}, coin^\star)$ in the H_0-*list*.

By applying Forking Lemma [22], after replaying $\mathcal{A}_I$ with the same random tape but different choices of oracle H, β can obtain another valid certificateless signature $(ID^\star, PK_{ID^\star}, m^\star, \delta'^\star = (u_1'^\star, u_2'^\star, h'^\star))$. Then, they should satisfy $R_1{}^\star = u_1{}^{\star e} H_0(ID^\star)^{h^\star}$ and $R_1{}^\star = u_1'{}^{\star e} H_0(ID^\star)^{h'^\star}$. Thus, we have the following relation

$$\begin{aligned}
u_1{}^{\star e} H_0(ID^\star)^{h^\star} &= u_1'{}^{\star e} H_0(ID^\star)^{h'^\star} \\
(\frac{u_1{}^\star}{u_1'{}^\star})^e &= H_0(ID^\star)^{h'^\star - h^\star} \\
(\frac{u_1{}^\star}{u_1'{}^\star})^e &= (zt_{ID^\star}{}^e)^{h'^\star - h^\star} \\
(\frac{u_1{}^\star}{t_{ID^\star}{}^{h'^\star - h^\star} u_1'{}^\star})^e &= (z)^{h'^\star - h^\star}.
\end{aligned}$$

Because e is a prime number, it means that $gcd(e, h'^\star - h^\star) = 1$, then there exists two numbers a, b satisfying $ae + b(h'^\star - h^\star) = 1$. Thus, we can obtain

$$\begin{aligned}
z &= z^{ae + b(h'^\star - h^\star)} \\
&= z^{ae} z^{b(h'^\star - h^\star)} \\
&= z^{ae} (\frac{u_1{}^\star}{t_{ID^\star}{}^{h'^\star - h^\star} u_1'{}^\star})^{eb} \\
&= (z^a (\frac{u_1{}^\star}{t_{ID^\star}{}^{h'^\star - h^\star} u_1'{}^\star})^{b})^{e}.
\end{aligned}$$

This shows that the RSA problem can be solved by β. Hence, it is in contradiction to the RSA problem.

Analysis: We show that β solves the given instance of the RSA problem with the probability η. We will observe that β does not abort during the whole simulation, $\mathcal{A}_I$ can forge the signature and the valid certificateless signature $(ID^\star, PK_{ID^\star}, m^\star, \delta'^\star = (u_1'^\star, u_2'^\star, h'^\star))$ satisfies $R_1{}^\star = u_1{}^{\star e} H_0(ID^\star)^{h^\star}$ and $R_1{}^\star = u_1'{}^{\star e} H_0(ID^\star)^{h'^\star}$. In **Partial-Private-Key-Extract** phase and **Private-Key-Extract** phase, the probability of β does not abort is at most $(1-\rho)^{q_{ppk}}$ and $(1-\rho)^{q_p}$, respectively. In **Signing phase**, the probability of no aborting is at most $(1-\rho)^{q_s}/q_H$. Thus, the probability of β does not abort in the simulation is at most $(1-\rho)^{q_{ppk}+q_p+q_s}(1-\rho)\cdot 1/q_H$ which is maximized at $\rho = 1 - 1/(q_{ppk}+q_p+q_s+1)$. That is to say, the probability of β does not abort is at most $1/\tau(q_{ppk} + q_p + q_s + 1)$, where τ denotes the base of the natural logarithm. Therefore, the probability of solving the RSA problem is $\eta > \frac{(q_s+1)(q_s+q_{H_0})\epsilon}{2^l q_H \tau (q_{ppk}+q_p+q_s+1)}$.

□

Theorem 2. *In the random oracle model, if there exists a type II adversary $\mathcal{A}_{II}$, who is allowed to request at most q_{H_0}, q_H* ***Hash*** queries *to random oracles H_0 and H, respectively, and q_s* ***Sign*** queries*, can break the proposed certificateless signature scheme with probability ϵ and within a time bound τ, then there exists another algorithm β who can make use of $\mathcal{A}_{II}$ to solve the discrete logarithm problem.*

Proof. Suppose there exists a Type II adversary $\mathcal{A}_{II}$ can break the proposed scheme. We are going to construct an adversary β that makes use of $\mathcal{A}_{II}$ to solve the discrete logarithm problem. Let us recall the discrete logarithm problem: for a given number $g \in Z_n^*$ and (n, p, q), y is a random number of Z_n, its goal is to compute x which satisfies $y = g^x \bmod n$. In order to solve this problem, β needs to simulate a challenge and the **Secret-Key-Extract** queries, **Hash** queries and **Sign** queries for $\mathcal{A}_{II}$. Thereby, β does in the following ways:

Setup: β maintains three lists H_0-*list*, *H-list* and *KeyList* which are initially empty. Let (e, n) be the system parameters. The master secret key is d and satisfies $ed \equiv 1 \bmod \phi(n)$, and the master secret key d and (p, q) are known for β, where $n = pq$. Choose two hash functions H_0 and H as random oracle. Let $PK_{ID^\star} = y$ be a challenged user $U^\star$'s public key and $ID^\star$ be the identity of the challenged user $U^\star$. Finally, β sends public parameters (e, d, n, g, H_0, H) to the adversary $\mathcal{A}_{II}$.

Queries: At any time, $\mathcal{A}_{II}$ is allowed to access the following oracles in a polynomial number of times. Then, β simulates the oracle queries of $\mathcal{A}_{II}$ as follows:

1) **H_0-Hash Queries:** $\mathcal{A}_{II}$ can query this oracle by given an identity ID. β randomly chooses $t_{ID} \in \phi(n)$ to set $H_0(ID) = g^{t_{ID}}$ and returns it to $\mathcal{A}_{II}$, where $\phi(n)$ is the Euler totient function and can be obtained by p, q. Finally, add $(ID, H_0(ID), t_{ID})$ to the H_0-*list*.

2) **H-Hash Queries:** In this process, $\mathcal{A}_{II}$ can request at most q_H Hash queries. For each query $(R_1,\ R_2,\ ID,\ PK_{ID},\ m)$, β randomly chooses $k_{ID} \in \{0,1\}^l$ and sets $H(R_1,\ R_2,\ ID,\ PK_{ID},\ m) = k_{ID}$. Finally, return k_{ID} to $\mathcal{A}_{II}$ and add $(R_1,\ R_2,\ ID,\ PK_{ID},\ m,\ k_{ID})$ to the *H-list*.

3) **Public-Key-Request Queries:** At any time, $\mathcal{A}_{II}$ can query the oracle by given an identity ID. If $ID \neq ID^\star$, β randomly chooses $x_{ID} \in \phi(n)$ to compute $PK_{ID} = H_0(ID)^{x_{ID}}$, then add $(ID, PK_{ID} = H_0(ID)^{x_{ID}}, x_{ID})$ to *KeyList*. Otherwise, β searches the H_0-*list* for $(ID^\star, H_0(ID^\star), t_{ID^\star})$ and computes $PK_{ID^\star} = y^{t_{ID^\star}}$. And add the record $(ID^\star, PK_{ID^\star}, \perp)$ to *KeyList*. Finally, send PK_{ID} to $\mathcal{A}_{II}$.

4) **Private-Key-Extract Queries:** When $\mathcal{A}_{II}$ makes this query with ID, if $ID \neq ID^\star$, β searches (ID, PK_{ID}, x_{ID}) in the *KeyList*, and computes $d_{ID} = H_0(ID)^d$. Then, $\mathcal{A}_{II}$ returns (d_{ID}, x_{ID}) to the adversary $\mathcal{A}_{II}$. If $ID = ID^\star$, then β aborts it.

Table 1: Performance analysis of proposed RSA based certificateless signature scheme

Process	Running Time (s)	Energy Consumption (mJ)	ROM (KB)	RAM (Static + Stack) (KB)
Sign	1.45	34.8	1.7	2.3
Verify	1.37	32.88	1.7	2.3

5) **Sign Queries:** For each query on an input (m, ID), if $ID \neq ID^\star$, then β firstly obtains private key associated with ID by **Private-Key-Extract** queries on ID, then it produces a signature by using the obtained private key. If $ID = ID^\star$, then β computes as follows:

   a. β randomly choose $u_1 \in Z_n$ and $h \in \{0,1\}^l$, $u_2 \in Z_{\phi(n)}$.

   b. β computes $R_1 = {u_1}^e H_0(ID)^h$ and $R_2 = H_0(ID)^{u_2} PK_{ID}^h$.

   c. β searches whether $(R_1, R_2, ID, PK_{ID}, m)$ exists in the *H-list*. If it exists, then abort it. Otherwise, β sets $H(R_1$, R_2, ID, PK_{ID}, $m) = h$ and adds $H(R_1$, R_2, ID, PK_{ID}, m, $h)$ in the *H-list*.

   d. The resultant signature $\delta = (u_1,\ u_2,\ h)$ is returned to $\mathcal{A}_{II}$.

Output: After all the queries, $\mathcal{A}_{II}$ outputs a forgery $(ID^\star, PK_{ID^\star}, m^\star, \delta^\star = ({u_1}^\star, {u_2}^\star, h^\star))$ and win this game. It must satisfy the following conditions:

1) If $\delta^\star$ is a valid forgery, then $h^\star = H({R_1}^\star, {R_2}^\star, PK_{ID^\star}, ID^\star, m)$ which is in the *H-list*, where ${R_1}^\star = {u_1}^{\star e} H_0(ID^\star)^{h^\star}$ and ${R_2}^\star = H_0(ID^\star)^{{u_2}^\star} PK_{ID^\star}^{h^\star})$.

2) $ID^\star$ is the challenger's identity and $H_0()$ is queried by $ID^\star$.

By applying Forking Lemma [22], after replaying $\mathcal{A}_{II}$ with the same random tape but different choices of oracle H, β can obtain another valid certificateless signature $(ID^\star, PK_{ID^\star}, m^\star, \delta'^\star = ({u'_1}^\star, {u'_2}^\star, h'^\star))$. Then, they should satisfy ${R_2}^\star = H_0(ID^\star)^{u_2} PK_{ID^\star}^{h^\star}$ and ${R_2}^\star = H_0(ID^\star)^{u'_2} PK_{ID^\star}^{h'^\star}$. Thus, we have the following relation:

$$
\begin{aligned}
H_0(ID^\star)^{u_2} PK_{ID^\star}^{h^\star} &= H_0(ID^\star)^{u'_2} PK_{ID^\star}^{h'^\star} \\
(H_0(ID^\star))^{u_2 - u'_2} &= PK_{ID^\star}^{h'^\star - h^\star} \\
(g)^{t_{ID^\star}(u_2 - u'_2)} &= y^{h'^\star - h^\star} \\
(g)^{t_{ID^\star}(u_2 - u'_2)/h'^\star - h^\star} &= y.
\end{aligned}
$$

Obviously, the discrete logarithm of y to the base g is $t_{ID^\star}(u_2 - u'_2)/h'^\star - h^\star$. It denotes that the discrete problem can be solved by β. Obviously, it is in contradiction to the difficulty of solving the discrete logarithm problem.

□

4.2 Performance Analysis

The proposed RSA based CLS scheme has been evaluated for WSN based on few parameters like running time and energy consumption, ROM and RAM including static RAM and stack RAM. The results are shown in Table 1. The scheme has been implemented on MICAz platform [23] using TinyOS-2.1.1 [18] operating system for embedded devices and RELIC-0.3.3 [4] cryptographic library. The running time of the proposed scheme is 1.45 seconds and 1.37 seconds in sign and verify phase respectively. The energy consumption is 34.8 milliJoules and 32.88 milliJoules in sign and verify phase respectively. Further, the proposed scheme consumes 1.7 KB of ROM and 2.3 KB of RAM (static and stack) excluding the space used by cryptographic library.

5 Conclusion

RSA is a well defined industry implemented security approach. Also certificateless schemes have their own benefits. In this paper, we proposed an RSA-based efficient certificateless signature scheme and proved it to be secure under some well-studied assumptions. We believe the new scheme is more suitable for systems with low-bandwidth channels and/or low-computation power making it suitable for WSN, on the basis of implementation results on WSN environment.

References

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: A survey," *Computer Networks*, vol. 38, pp. 393–422, Mar. 2002.

[2] S. Al-Riyami and K. Paterson, "Certificateless public key cryptography," in *Advances in Cryptology (ASIACRYPT'03)*, LNCS 2894, pp. 452–473, Springer, 2003.

[3] F. Amin, H. Jahangir, and H. Rasifard, "Analysis of public-key cryptography for wireless sensor networks security," vol. 2, no. 5, pp. 403–408, 2008.

[4] D. Aranha and C. Gouvêa, "RELIC is an Efficient LIbrary for Cryptography," June 15, 2015. (`http://code.google.com/p/relic-toolkit/`)

[5] M. Bellare and G. Neven, "Identity-based multisignatures from rsa," in *Topics in Cryptology (CT-RSA 2007)*, LNCS 4377, pp. 145–162, Springer, 2006.

[6] X. Chen, K. Makki, K. Yen, and N. Pissinou, "Sensor network security: A survey," *IEEE Communications Surveys Tutorials*, vol. 11, no. 2, pp. 52–73, 2009.

[7] Y. Chen and Q. Zhao, "On the lifetime of wireless sensor networks," *IEEE Communications Letters*, vol. 9, no. 11, pp. 976–978, 2005.

[8] L. Chun-Ta, H. Min-Shiang, and C. Yen-Ping, "Improving the security of a secure anonymous routing protocol with authenticated key exchange for ad hoc networks," *International Journal of Computer Systems Science and Engineering*, vol. 23, no. 3, pp. 227–234, 2008.

[9] L. Chun-Ta, H. Min-Shiang, and C. Yen-Ping, "A secure and efficient communication scheme with authenticated key establishment and privacy preserving for vehicular ad hoc networks," *Computer Communications*, vol. 31, no. 12, pp. 2803–2814, 2008.

[10] H. Du and Q. Wen, "Efficient and provably-secure certificateless short signature scheme from bilinear pairings," *Computer Standards and Interfaces*, vol. 31, no. 2, pp. 390–394, 2009.

[11] P. Gong and P. Li, "Further improvement of a certificateless signature scheme without pairing," *International Journal of Communication Systems*, vol. 27, no. 10, pp. 2083–2091, Oct. 2014.

[12] M. Gorantla and A. Saxena, "An efficient certificateless signature scheme," in *Computational Intelligence and Security*, LNCS 3802, pp. 110–116, Springer, 2005.

[13] D. He, J. Chen, and R. Zhang, "An efficient and provably-secure certificateless signature scheme without bilinear pairings," *International Journal of Communication Systems*, vol. 25, no. 11, pp. 1432–1442, 2012.

[14] D. He, M. Khan, and S. Wu, "On the security of a rsa-based certificateless signature scheme," *International Journal of Network Security*, vol. 15, no. 6, pp. 408–410, 2013.

[15] B. Hu, D. Wong, Z. Zhang, and X. Deng, "Key replacement attack against a generic construction of certificateless signature," in *Information Security and Privacy*, LNCS 4058, pp. 235–246, Springer, 2006.

[16] X. Huang, Y. Mu, W. Susilo, D. Wong, and W. Wu, "Certificateless signatures: New schemes and security models," *The Computer Journal*, vol. 55, no. 4, pp. 457–474, 2012.

[17] X. Huang, W. Susilo, Y. Mu, and F. Zhang, "On the security of certificateless signature schemes from asiacrypt 2003," in *Cryptology and Network Security*, LNCS 3810, pp. 13–25, Springer, 2005.

[18] P. Levis, S. Madden, J. Polastre, R. Szewczyk, K. Whitehouse, A. Woo, D. Gay, J. Hill, M. Welsh, E. Brewer, and D. Culler, "Tinyos: An operating system for sensor networks," in *Ambient Intelligence*, pp. 115–148, 2005.

[19] C. Li, M. Hwang, and Y. Chu, "An efficient sensor-to-sensor authenticated path-key establishment scheme for secure communications in wireless sensor networks," *International Journal of Innovative Computing, Information and Control*, vol. 5, no. 8, pp. 2107–2124, 2009.

[20] K. McCurley, "Discrete logarithm problem," in *Proceedings of Symposia Applied Mathematics*, pp. 49–74, 1990.

[21] S. Olariu and Q. Xu, "Information assurance in wireless sensor networks," in *Proceedings of the IEEE International Symposium on Parallel and Distributed Processing*, vol. 13, pp. 236a, Los Alamitos, CA, USA, 2005.

[22] D. Pointcheval and J. Stern, "Security proofs for signature schemes," in *Advances in Cryptology (EUROCRYPT'96)*, LNCS 1070, pp. 387–398, Springer, 1996.

[23] A. Rev, "MPR-MIB series user manual," 2004. (`http://www-db.ics.uci.edu/pages/research/quasar/MPR-MIB%20Series%20User%20Manual%207430-0021-06_A.pdf`)

[24] A. Shamir, "Identity-based cryptosystems and signature schemes," in *Advances in Cryptology*, LNCS 196, pp. 47–53, Springer, 1985.

[25] G. Sharma and A. Verma, "Breaking the rsa-based certificateless signature scheme," *Information-An International Interdisciplinary Journal*, vol. 16, no. 11, pp. 7831–7836, 2013.

[26] J. Tsai, N. Lo, and T. Wu, "Weaknesses and improvements of an efficient certificateless signature scheme without using bilinear pairings," *International Journal of Communication Systems*, vol. 27, no. 7, pp. 1083–090, July 2014.

[27] R. Tso, X. Huang, and W. Susilo, "Strongly secure certificateless short signatures," *Journal of Systems and Software*, vol. 85, no. 6, pp. 1409–1417, 2012.

[28] R. Tso, X. Yi, and X. Huang, "Efficient and short certificateless signatures secure against realistic adversaries," *The Journal of Supercomputing*, vol. 55, no. 2, pp. 173–191, 2011.

[29] J. Walters, Z. Liang, W. Shi, and V. Chaudhary, *Security in Distributed, Grid, and Pervasive Computing*, Chap. 17 Wireless Sensor Network security: A survey, pp. 1–51, CRC Press, 2007.

[30] A. Wander, N. Gura, H. Eberle, V. Gupta, and S. Shantz, "Energy analysis of public-key cryptography for wireless sensor networks," in *Proceedings of the 3rd IEEE International Conference on Pervasive Computing and Communications (PERCOM'05)*, pp. 324–328, Washington, DC, USA, 2005.

[31] C. Wang, D. Long, and Y. Tang, "An efficient certificateless signature from pairings," *International Journal of Network Security*, vol. 8, no. 1, pp. 96–100, 2009.

[32] R. Watro, D. Kong, S. Cuti, C. Gardiner, C. Lynn, and P. Kruus, "Tinypk: Securing sensor networks with public key technology," in *Proceedings of the 2nd ACM Workshop on Security of Ad Hoc and Sensor Networks (SASN'04)*, pp. 59–64, New York, NY, USA, 2004.

[33] Z. Xu, X. Liu, G. Zhang, and W. He, "Mccls: Certificateless signature scheme for emergency mobile wireless cyber-physical systems," *International Journal*

of Computers Communications and Control, vol. 3, no. 4, pp. 395–411, 2008.

[34] Z. Xu, X. Liu, G. Zhang, W. He, G. Dai, and W. Shu, "A certificateless signature scheme for mobile wireless cyber-physical systems," in *Proceedings of the 8th International Conference on Distributed Computing Systems Workshops (ICDCS'08)*, pp. 489–494, 2008.

[35] W. Yap, S. Heng, and B. Goi, "An efficient certificateless signature scheme," in *Emerging Directions in Embedded and Ubiquitous Computing*, LNCS 4097, pp. 322–331, Springer, 2006.

[36] D. Yum and P. Lee, "Generic construction of certificateless signature," in *Information Security and Privacy*, LNCS 3108, pp. 200–211, Springer, 2004.

[37] F. Zhang, S. Li, S. Miao, Y. Mu, W. Susilo, and X. Huang, "Cryptanalysis on two certificateless signature schemes," *International Journal of Computers Communications and Control*, vol. 5, no. 4, pp. 586–591, 2010.

[38] J. Zhang and J. Mao, "An efficient rsa-based certificateless signature scheme," *Journal of Systems and Software*, vol. 85, no. 3, pp. 638–642, 2012.

[39] X. Zhang, H. Heys, and L. Cheng, "Energy efficiency of symmetric key cryptographic algorithms in wireless sensor networks," in *25th Biennial Symposium on Communications (QBSC'10)*, pp. 168–172, 2010.

[40] Z. Zhang, D. Wong, J. Xu, and D. Feng, "Certificateless public-key signature: Security model and efficient construction," in *Applied Cryptography and Network Security*, LNCS 3989, pp. 293–308, Springer, 2006.

[41] M. Zhou, M. Zhang, C. Wang, and B. Yang, "Cclas: A practical and compact certificateless aggregate signature with share extraction," *International Journal of Network Security*, vol. 16, no. 2, pp. 157–164, 2014.

Gaurav Sharma received his Ph.D and M.E degree in Computer Science & Engineering from Thapar University, Patiala, India. He had received M. Sc. as well as B. Sc. Degree from CCS University, Meerut, India. Presently he is working as an Asst. Professor at Galgotias University, India. His area of interests is routing and security in Ad hoc networks.

Suman Bala received her Ph.D and M.E degree in Computer Science & Engineering from Thapar University, Patiala, India. She had received B.Tech degree from Punjab Technical University, Jalandhar, India. Her areas of interest are: Wireless Sensor Networks, Security, Cryptography and Key Management.

Anil K. Verma is currently working as Associate Professor in the department of Computer Science and Engineering at Thapar University, Punjab (INDIA). He has more than 20 years of experience. He has published over 150 papers in referred journals and conferences (India and Abroad). He is member of various program committees for different International/National Conferences and is on the review board of various journals. He is a senior member (ACM), LMCSI (Mumbai), GMAIMA (New Delhi). He is a certified software quality auditor by MoCIT, Govt. of India. His research interests include wireless networks, routing algorithms and securing ad hoc networks.

Secure and Efficient Identity-based Proxy Multi-signature Using Cubic Residues

Feng Wang[1,2], Chin-Chen Chang[2,3], Changlu Lin[4], and Shih-Chang Chang[5]

(Corresponding author: Chin-Chen Chang)

College of Mathematics and Physics, Fujian University of Technology[1]
Fuzhou, Fujian,350108, China
Department of Information Engineering and Computer Science, Feng Chia University[2]
100 Wenhwa Rd., Seatwen, Taichung 40724, Taiwan
(Email: alan3c@gmail.com)
Department of Computer Science and Information Engineering, Asia University[3]
Taichung 41354, Taiwan
School of Mathematics and Computer Science, Fujian Normal University[4]
Fuzhou, Fujian, 350117, China
Department of Computer Science and Information Engineering, Notional Chung Cheng University[5]
160 San-Hsing, Ming-Hsiung, Chiayi 621,Taiwan

(Received July 14, 2014; revised and accepted Jan. 16 & June 4, 2015)

Abstract

The term "proxy multi-signature" refers to the situation in which a proxy signer is authorized to sign a message on behalf of a group of original signers. Combined with identity-based cryptography, we proposed an efficient identity-based proxy multi-signature scheme using cubic residues without bilinear pairing. Our scheme is secure against existential forgery on adaptive chosen-message and identity attacks under the hardness of integer factorization assumption. Compared with elliptic curve or bilinear pairing, the integer factorization assumption is more reliable and easier to use because it has been developed 2500 years ago. Furthermore, our scheme is more efficient than previous schemes based on bilinear pairing.

Keywords: Cubic residues, identity-based signature, integer factorization, proxy multi-signature, random oracle model

1 Introduction

Shamir [15] introduced identity-based cryptography in 1984 in order to simplify the key-management procedure of traditional, certificate-based, public-key infrastructures. Shamir's approach allowed an entity's public key to be derived directly from her or his identity, such as an email address, and the entity's private key can be generated by a trusted third party which is called the private key generator (PKG).

The notion of proxy signatures was proposed by Mambo et al. [10] in 1996. They identified the signers into two entities, i.e., the original signer and the proxy signer. The latter can sign a message on behalf of the former with a warrant the former delegated. Proxy signatures have many practical applications, such as distributed systems, grid computing, mobile agent applications, distributed shared object systems, global distribution networks, and mobile communications [2]. Since 1996, the proxy signature has been paid significant attention [7] and various extensions of the proxy signature have been proposed [1, 9, 11, 19, 22], one of which is the proxy multi-signature [9, 19, 22].

In 2000, Yi et al. proposed the proxy multi-signature [22] in which a designated proxy signer can generate a valid signature on behalf of a group of original signers. Proxy multi-signature can be used in the following scenario, i.e., a university wants to release a document that several departments may be involved, for example, the Deans Office, the Student Affairs Office, and the Human Resources Department, etc.. The document must be signed by all of the above entities or by a proxy signer delegated by those entities. Combined with identity-based cryptography, Li and Chen [9] proposed the notion of identity-based proxy multi-signature (IBPMS) and constructed a scheme using bilinear pairings in 2005. However, most existing IBPMS schemes were based on bilinear pairing [4, 9, 14, 20], which required more computational cost than normal operations, such as modular exponentiations in finite fields. Therefore, there was a strong interest in determining how to construct a secure scheme without pairing. In 2011, Tiwari and Padhye [18] proposed a secure IBPMS scheme based on the elliptic curve discrete logarithm problem. Although they claimed that their scheme was more efficient and had a smaller key size than pairing-based schemes, the security on which their method was based on the elliptic curve discrete logarithm problem assumption which was only a few decades old [6].

In this paper, we propose a new identity-based proxy multi-signature (IBPMS) scheme using cubic residues without bilinear pairing. The security of our method is based on the integer factorization assumption which is 2500 years old. We briefly introduce our contributions. First, our scheme is the first identity-based proxy multi-signature scheme using the cubic residues problem. Second, our scheme has been proven to be secure in the random oracle model under the hardness of integer factorization problem assumption. Third, our scheme is made more efficient than Cao and Cao's IBPMS scheme [4] based on bilinear pairing.

The rest of the paper is organized as follows. In Section 2, we introduce the cubic residues problem and integer factorization problem assumption. In Section 3, we give the formal definition and security model of identity-based proxy multi-signature. In Section 4, we propose a new identity-based proxy multi-signature scheme using cubic residues. In Section 5, we give the formal security proof for the proposed scheme under the random oracle model. In Section 6, we compare the efficiency and performance of our scheme with Cao and Cao's IBPMS scheme. Finally, we present our conclusions in Section 7.

2 Preliminaries

In this section, we review cubic residues and the method of their construction mentioned in [21] and integer factorization problem assumption

2.1 Cubic Residues

Definition 1. *For a positive integer n, if there is some x that satisfies the expression $x^3 \equiv C \pmod{n}$, we say that C is a cubic residue modulo n, and x is called the cubic root of C modulo n.*

From [21], we have Lemma 1, Theorem 1, and Theorem 2.

Lemma 1. *Let p be a prime number, $3_p = gcd(3, p-1)$, and $C \in Z_p^*$. We say that C is a cubic residue modulo p if and only if $C^{\frac{(p-1)}{3_p}} \pmod{p} \equiv 1$.*

Obviously, if p is prime number and $p \equiv 2 \pmod{3}$, then every $C \in Z_p^*$ is a cubic residue modulo p.

If q is prime number, and $q \equiv 4$ or $7 \pmod{9}$, for every $h \in Z_p^*$, we can construct a cubic residue modulo q as follows.

Let a be a non-cubic modulo q, we compute $\eta = [(q-1) \pmod{9}]/3$, $\lambda = \eta \pmod{2} + 1$, $\beta = (q-1)/3$, $\xi = a^{\eta \cdot \beta} \pmod{q}$, $\tau \equiv h^{\lambda \cdot \beta} \pmod{q}$, and

$$b = \begin{cases} 0, & if\ \tau = 1 \\ 1, & if\ \tau = \xi \\ 2, & if\ \tau = \xi^2, \end{cases}$$

then $C = a^b \cdot h$ is a cubic residue modulo q.

Theorem 1. *Let p, q be as mentioned above and $n = p \cdot q$. Then $C = a^b \cdot h$ is a cubic residue modulo n, and $s \equiv C^{[2^{\eta-1}(p-1)(q-1)-3]/9} \pmod{n}$ is a cubic root of C^{-1}.*

Theorem 2. *Let $n = p \cdot q$. If there is $s_1^3 \equiv s_2^3 \equiv C \pmod{n}$, and $s_1 \not\equiv s_2 \pmod{n}$, then $gcd(s_1 - s_2, n)$ is a non-trivial divisor of n.*

2.2 Integer Factorization Problem Assumption

The integer factorization problem assumption is one of the fundamental hardness problems, which has been studied extensively and used to construct cryptographic schemes. We will analyze the security of our proposed scheme based on this assumption. From [23], we have Definition 2 and Definition 3.

Definition 2. *Given $n = p \cdot q$, where p and q are prime numbers and they are unknown publicly, the integer factorization problem is defined to output a prime number $p(1 < p < n)$ such that p can divide n.*

Definition 3 (Integer factorization problem assumption). *The integer factorization problem (IFP) is a (t', ϵ')-hard assumption, if there is no polynomial time algorithm in time at most t', can solve the integer factorization problem with probability at least ϵ'.*

3 Formal Definition and Security Model

We give a formal definition and security model of the identity-based proxy multi-signature scheme based on the works of Cao and Cao [4], Singh and Verma [16], and Sun et al. [17].

3.1 Formal Definition of the Identity-based Proxy Multi-signature Scheme

In an identity-based proxy multi-signature scheme, there are two entities named as a group of the original signers and the proxy signer. We use ID_i, for $i = 1, 2, \cdots, n$, to denote the identity of original signer i, and ID_{ps} to denote the identity of the proxy signer. From [4], we have Definition 4.

Definition 4. *An identity-based proxy multi-signature scheme (IBPMS) is a tuple of seven algorithms as IBPMS=(**Setup**, **Extract**, **DelGen**, **DelVeri**, **PMKGen**, **PMSign**, **PMVeri**).*

Setup. PKG takes a security parameter as input, and outputs public parameter PP and its master key MK.

Extract. PKG takes its master key MK and a user's identity ID_i as inputs, and outputs the user's public key and secret key pair (H_{ID_i}, s_{ID_i}).

DelGen. For $i = 1, 2, \cdots, n$, the original signer i takes her or his secret key s_{ID_i} and a warrant w as inputs, and outputs her or his delegation $D_{i \to ps}$ to the proxy signer.

DelVeri. For $i = 1, 2, \cdots, n$, the proxy signer takes delegation $D_{i \to ps}$ from the original signer i and her or his identity ID_i as inputs, and verifies whether or not the delegation is valid.

PMKGen. The proxy signer takes her or his secret key $s_{ID_{ps}}$ and delegations $D_{i \to ps}$, $i = 1, 2, \cdots, n$, as inputs, and generates her or his private signing key sk_{ps}.

PMSign. The proxy signer takes her or his signing key sk_{ps}, message m, and delegations $D_{i \to ps}$, $i = 1, 2, \cdots, n$, as inputs, and generates the proxy multi-signature σ of the message m.

PMVeri. The verifier takes the proxy multi-signature σ and the original signers' identities, ID_i, $i = 1, 2, \cdots, n$, and the proxy signer's identity ID_{ps} as inputs, and verifies whether or not the proxy multi-signature is valid.

3.2 Security Model

Compared with Cao and Cao's method [4], and Sun et al.'s method [17], we use the security model of the proxy multi-signature which is described in [17]. And, we extend Sun et al.'s model into an identity-based proxy multi-signature to prove the security of our scheme. The adversaries in their model can be classified into three types as follows:

Type 1. The adversary, A_1, knows nothing except the identities of the original signers and the proxy signer.

Type 2. The adversary, A_2, knows the secret keys of $n-1$ original signers and proxy signer in addition to what A_1 knows in Type 1.

Type 3. The adversary, A_3, knows the secret keys of all of the original signers in addition to what A_1 knows in Type 1, but does not know the secret key of the proxy signer.

Obviously, if an adversary in Type 1 can forge a valid signature of the scheme, the adversary in Type 2 or Type 3 also can forge a valid signature. So, we only consider the Type 2 and Type 3 adversaries in this paper.

With regard to the Type 2 adversary A_2, we can assume that she or he has all of the secret keys of the $n-1$ original signers, except for signer n. If she or he has a valid delegation, $D_{n \to ps}$, she or he can output a valid proxy multi-signature herself or himself with the secret keys of the other original signers and proxy signer. So, the objective of the Type 2 adversary is to output a valid delegation, $D_{n \to ps}$.

With regard to the Type 3 adversary A_3, since she or he has all of the secret keys of the original signers, she or he can output a valid delegation $D_{i \to ps}$, $i = 1, 2, \cdots, n$, herself or himself. So, the objective of the Type 3 adversary is to output a valid proxy multi-signature under delegations $D_{i \to ps}$, $i = 1, 2, \cdots, n$.

Let an adversary $A_t (t = 2$ or $3)$ be a probabilistic Turing machine, A_t takes public parameter PP and a random tape as inputs and performs an experiment with the algorithm B. Inspired from [17], we define the following two definitions.

Definition 5. *For an identity-based proxy multi-signature scheme, we define an experiment of the adversary $A_t (t = 2$ or $3)$ with the security parameter λ as follows:*

Step 1. *Algorithm B runs the Setup algorithm and returns public parameter PP to the adversary A_t.*

Step 2. *B maintains several lists, e.g., E_{list}, D_{list}, S_{list}, and initializes them as null.*

Step 3. *When the adversary A_t makes adaptive queries from the algorithm B, B maintains several oracles and answers as follows:*

- ***Extract oracle:*** *The oracle takes a user's identity ID_i as input, returns her or his private key s_{ID_i}, and puts the tuple (ID_i, s_{ID_i}) into E_{list}.*
- ***DelGen oracle:*** *The oracle takes the original signer's identity ID_i and the warrant w as inputs, returns the delegation $D_{i \to ps}$, and puts the tuple $(ID_i, w, D_{i \to ps})$ into D_{list}.*
- ***PMSign oracle:*** *The oracle takes the message m and the delegations $D_{i \to ps}$, $i = 1, 2, \cdots, n$ as inputs, returns a proxy multi-signature σ signed by the proxy signer and puts the tuple (m, w, σ) into S_{list}.*

Step 4. *Eventually, A_t outputs a forgery.*

- *If $t = 2$, then it is the Type 2 adversary A_2. The forgery is of the tuple $(ID_n, w, D_{n \to ps})$, and $(ID_n, w, D_{n \to ps})$ is valid delegation of ID_n with warrant w, and $ID_n \notin E_{list}$, $(ID_n, w) \notin D_{list}$.*
- *If $t = 3$, then it is the Type 3 adversary A_3. The forgery is of the tuple (m, w, σ), and (m, w, σ) is a valid proxy multi-signature, and $ID_p \notin E_{list}$, $(w, m) \notin S_{list}$.*

If the output satisfies one of the above two items, A_t's attack was successful.

Definition 6. *For any polynomial adversary A_t ($t = 2$ or 3), if the probability of A_t's success in the above experiment is negligible, then, the identity-based proxy multi-signature scheme is said to be secure against existential forgery on adaptive chosen-message and identity attacks.*

4 Our Proposed IBPMS Scheme

In this section, we describe a new identity-based proxy multi-signature scheme. We designed our scheme, which extends the identity-based signature [21], based on the cubic residues. The proposed scheme includes the following seven algorithms:

Setup. Given the security parameters k and l, PKG carries out the algorithm and returns public parameters PP and master key MK as follows:

1) Randomly generates two k-bits prime numbers p and q, satisfying $p \equiv 2 \pmod 3$ and $q \equiv 4$ or $7 \pmod 9$, respectively; then computes $n = p \cdot q$.
2) Computes $d = [2^{\eta-1}(p-1)(q-1)-3]/9$, $\eta = [(q-1) \pmod 9]/3$, $\lambda = \eta \pmod 2 + 1$, $\beta = (q-1)/3$.
3) Randomly selects a non-cubic residue a modulo q and computes $\xi \equiv a^{\eta \cdot \beta} \pmod q$.
4) Selects four hash functions $H_1 : \{0,1\}^* \rightarrow Z_n^*$,$H_2$, H_3 ,$H_4 : \{0,1\}^* \rightarrow \{0,1\}^l$.

PKG publishes $(n, a, \eta, \lambda, H_1, H_2, H_3, H_4)$ as the public parameter PP and keeps (p, q, d, β) secret as the master key MK.

Extract. Given public parameter PP, the master key MK, and identity ID_i of user i, for $i = 1, 2, \cdots, n$, PKG computes the corresponding secret key as follows:

1) Computes $\tau_i \equiv H_1(ID_i)^{\lambda \cdot \beta} \pmod q$.
2) Computes $b_i = \begin{cases} 0, & if\ \tau_i = 1 \\ 1, & if\ \tau_i = \xi \\ 2, & if\ \tau_i = \xi^2 \end{cases}$, and $C_i = a^{b_i} \cdot H_1(ID_i) \pmod n$, $s_{ID_i} \equiv (C_i)^d \pmod n$.

PKG transmits secret key (s_{ID_i}, b_i), for $i = 1, 2, \cdots, n$ to user i via a secure channel.

DelGen. Let ID_i, for $i = 1, 2, \cdots, n$, be the identity of the original signer i, and ID_{ps} be the identity of the proxy signer. The original signer i, for $i = 1, 2, \cdots, n$, wants to delegate the proxy signer to get a warrant w of message m, so she or he takes her or his secret key (s_{ID_i}, b_i), and warrant w as inputs and outputs the delegation $D_{i \rightarrow ps}$. Then, the original signer i, for $i = 1, 2, \cdots, n$, continues as follows:

1) Randomly selects $r_i \in Z_n^*$, computes $R_i \equiv r_i^3 \pmod n$, and broadcasts R_i to the other original signers.
2) Computes $R \equiv \prod_{i=1}^{n} R_i \pmod n$, $h_w = H_2(w,\ R)$, $V_i \equiv r_i \cdot s_{ID_i}^{h_w} \pmod n$.

Each original signer i sends her or his delegation $D_{i \rightarrow ps} = (ID_i, b_i, w, R_i, V_i)$ to the proxy signer.

DelVeri. To verify each delegation $D_{i \rightarrow ps}$ with warrant w, the proxy signer computes $R \equiv \prod_{i=1}^{n} R_i \pmod n$, $h_w = H_2(w,\ R)$, $C_i \equiv a^{b_i} \cdot H_1(ID_i) \pmod n$, and checks $V_i^3 \cdot {C_i}^{h_w} \equiv R_i \pmod n$ for $i = 1, 2, \cdots, n$. If the equation holds, she or he accepts $D_{i \rightarrow ps}$ as a valid delegation; otherwise, it is rejected.

PMKGen. If the proxy signer accepts all delegations $D_{i \rightarrow ps}$, for $i = 1, 2, \cdots, n$, she or he computes $h_{ps} = H_3(ID_{ps}, w, R)$, $V \equiv \prod_{i=1}^{n} V_i \pmod n$, $sk_{ps} \equiv s_{ID_{ps}}^{h_{ps}} \cdot V \pmod n$ and takes sk_{ps} as her or his private signing key.

PMSign. The proxy signer takes sk_{ps} as input and randomly selects $r_{ps} \in Z_n^*$, computes $R_{ps} \equiv r_{ps}^3 \pmod n$, $h_m = H_4(ID_{ps}, w, m, R_{ps})$, $V_{ps} \equiv r_{ps} \cdot sk_{ps}^{h_m} \pmod n$. The tuple $(ID_1, ID_2, \cdots, ID_n, ID_{ps}, b_1, b_2, \cdots, b_n, b_{ps}, m, w, R, R_{ps}, V_{ps})$ is the proxy signature of message m on behalf of all original signers i, for $i = 1, 2, \cdots, n$.

PMVeri. In order to verify the proxy multi-signature $(ID_1, ID_2, \cdots, ID_n, ID_{ps}, b_1, b_2, \cdots, b_n, b_{ps}, m, w, R, R_{ps}, V_{ps})$ of message m under warrant w, the verifier conducts the following: computes $h_{ps} = H_3(ID_{ps}, w, R)$, $h_w = H_2(w, R)$, $h_m = H_4(ID_{ps}, m, w, R_{ps})$, $C \equiv \prod_{i=1}^{n} (a^{b_i} \cdot H_1(ID_i)) \pmod n$, $C_{ps} \equiv a^{b_{ps}} \cdot H_1(ID_{ps}) \pmod n$, then checks $V_{ps}^3 \cdot {C_{ps}}^{h_{ps} \cdot h_m} \cdot C^{h_w \cdot h_m} \equiv R_{Ps} \cdot R^{h_m} \pmod n$; if the equation holds, then she or he accepts it; otherwise, it is rejected.

Our scheme is correct because the following equation holds:

$$\begin{aligned}
& V_{ps}^3 \cdot {C_{ps}}^{h_{ps} \cdot h_m} \cdot C^{h_w \cdot h_m} \\
\equiv\ & (r_{ps} \cdot sk_{ps}^{h_m})^3 \cdot {C_{ps}}^{h_{ps} \cdot h_m} \cdot C^{h_w \cdot h_m} \\
\equiv\ & \left(r_{ps} \cdot (d_{ID_{ps}}^{h_{ps}} \cdot V)^{h_m}\right)^3 \cdot {C_{ps}}^{h_{ps} \cdot h_m} \cdot C^{h_w \cdot h_m} \\
\equiv\ & \left(r_{ps} \cdot (d_{ID_{ps}}^{h_{ps}} \cdot \textstyle\prod_{i=1}^{n} r_i \cdot s_{ID_i}^{h_w})^{h_m}\right)^3 \cdot {C_{ps}}^{h_{ps} \cdot h_m} \cdot C^{h_w \cdot h_m} \\
\equiv\ & r_{ps}^3 \cdot \left((d_{ID_{ps}}^3)^{h_{ps}} \cdot \textstyle\prod_{i=1}^{n} r_i^3 \cdot \prod_{i=1}^{n} (s_{ID_i}^3)^{h_w}\right)^{h_m} \cdot {C_{ps}}^{h_{ps} \cdot h_m} \cdot C^{h_w \cdot h_m} \\
\equiv\ & r_{ps}^3 \cdot \left((d_{ID_{ps}}^3)^{h_{ps}} \cdot \textstyle\prod_{i=1}^{n} r_i^3 \cdot \prod_{i=1}^{n} (s_{ID_i}^3)^{h_w}\right)^{h_m} \cdot {C_{ps}}^{h_{ps} \cdot h_m} \cdot C^{h_w \cdot h_m} \\
\equiv\ & R_{ps} \cdot ({C_{ps}}^{-h_{ps}} \cdot R \cdot \textstyle\prod_{i=1}^{n} {C_i}^{-h_w})^{h_m} \cdot {C_{ps}}^{h_{ps} \cdot h_m} \cdot C^{h_w \cdot h_m} \\
\equiv\ & R_{ps} \cdot R^{h_m} \pmod n.
\end{aligned}$$

5 Security Proof of Our Proposed Scheme

In this section, we give the security proof of our proposed scheme. We show that our scheme is secure against existential forgery under adaptive chosen-message and identity attacks in the random oracle model. We prove our scheme against Type 2 adversaries and Type 3 adversaries, respectively.

If a Type 2 adversary A_2 has the ability to break our scheme, we can construct a polynomial time algorithm B, by interacting with A_2, to solve the integer factorization problem.

Theorem 3. *Given a pair of security parameters (k, l), if the integer factorization problem is (t', ϵ')-hard, then our identity-based proxy multi-signature scheme is $(t, q_{H_2}, q_D, \epsilon_2)$-secure against existential forgery under adaptive chosen-message and identity attacks for the Type 2 adversary A_2, which satisfies:*

$$\epsilon' \geq \frac{4}{9} \cdot \left(\frac{(\epsilon_2 - \delta_2)^2}{q_{H_2} + 1} - \frac{\epsilon_2 - \delta_2}{2^l} \right),$$

$$t' = 2t + O\left(k^2 \cdot l + k^3\right),$$

where q_{H_2} and q_D denote the number of queries that A_2 can ask to the random oracle H_2 and DelGen oracle, respectively, and $\delta_2 = \frac{q_D \cdot (q_{H_2} + q_D)}{3 \cdot 2^k}$.

Proof. Assuming that adversary A_2 breaks the proposed scheme, we can construct an algorithm B to resolve the integer factorization problem.

Given an integer $n = p \cdot q$ (for some unknown p and q), and a non-cubic residue $a \pmod n$, we will design an algorithm B to output p and q with non-negligible probability.

Step 1. Algorithm B sends (n, a) to adversary A_2 as public parameters.

Step 2. B maintains several lists, i.e., $H_{1,list}$, $H_{2,list}$,E_{list}, and D_{list} and initializes them as null.

Step 3. B responds to A_2's queries as follows:

- H_1**-oracle:** A_2 requests H_1 on ID_i, and B checks if ID_i existed in $H_{1,list}$. If not, B picks a random $s_i \in Z_n^*$ and $b_i \in \{0, 1, 2\}$, computes $h_{1,i} = H_1(ID_i) \equiv \frac{s_i^3}{a^{b_i}} \pmod n$, and adds the tuple $(ID_i, h_{1,i}, s_i, b_i)$ into $H_{1,list}$; then, B returns $h_{1,i}$ to A_2.
- H_2**-oracle:** A_2 requests H_2 on (w, R), and B checks if (w, R) existed in $H_{2,list}$. If not, B picks a random $e \in \{0, 1\}^l$, adds the tuple (w, R, e) into $H_{2,list}$, then, B returns e to A_2.
- **Extract oracle:** A_2 requests Extract algorithm on ID_i , and B checks if ID_i existed in E_{list}. If not, B returns to H_1-oracle and gets $(ID_i, h_{1,i}, s_i, b_i)$ of $H_{1,list}$; then, B returns (s_i, b_i) to A_2 and adds the tuple (ID_i, s_i, b_i) into E_{list}.
- **DelGen oracle:** A_2 requests delegation on (ID_n, w). According to the assumption, A_2 has the secret keys of the original signers i, $i = 1, 2, \cdots, n-1$, by requesting Extract oracle. For $i = 1, 2, \cdots, n-1$, A_2 randomly selects $r_i \in Z_n^*$, computes $R_i \equiv r_i^3 \pmod n$, and sends R_i, where $i = 1, 2, \cdots, n-1$, to B. B randomly selects $V_n, \tau \in \{0, 1\}^l$, computes $R_n \equiv V_n^3 \cdot (a^{b_n} \cdot H_1(ID_n))^\tau \pmod n$, and $R \equiv \prod_{i=1}^n R_i \pmod n$; if R already exists in $H_{2,list}$, failure is returned; else (ID_n, b_n, w, R_n, V_n) is returned as the original signer n's delegation to A_2; also, τ is returned for the sake of helping A_2 completing the delegation on (ID_i, w) for $i = 1, 2, \cdots, n-1$. B adds the tuple (ID_n, b_n, w, R_n, V_n) into D_{list} and adds (w, R, τ) into $H_{2,list}$.

Step 4. A_2 outputs a delegation forgery of warrant w^* and ID_n^* with $D_{n \to ps}^* = (ID_n^*, b_n^*, w^*, R_n^*, V_n^*)$, which (ID_n^*, w^*) is not requested on the DelGen oracle, and ID_n^* is not requested on the Extract oracle.

Step 5. Finally, we will show how B resolves the integer factorization problem with A_2's delegation forgery.

We apply the oracle replay technique describes in Forking Lemma [12, 13] to factor n, i.e., B resets A_2 two times. For the first time, B records all the transcripts that interacted with A_2. For the second time, B starts with the first time random tape and returns the same answers to A_2, except H_2-oracle. Each time, when A_2 asks H_2-oracle, B chooses different random numbers, e^*,e^{**}, as the answer, respectively.

After two rounds of interacting with B, A_2 forges two delegations $(ID_n^*, b_n^*, w^*, R_n^*, V_n^*)$, $(ID_n^*, b_n^*, w^*, R_n^*, V_n^{**})$, together with delegations of original signers $1, 2, \cdots, n-1$, sends them to B. Then, B executes as follows:

- B computes $R^* \equiv \prod_{i=1}^n R_i^* \pmod n$, returns to the previous three records of $H_{2,list}$ lists for (w^*, R^*), obtains, e^*, e^{**}, and checks whether or not they satisfy $(e^* - e^{**}) \equiv 0 \pmod 3$; if so, then B aborts it.
- Else B can obtain$(V_n^*)^3 \cdot (C_n^*)^{e^*} = R_n^*$, $(V_n^{**})^3 \cdot (C_n^*)^{e^{**}} \equiv R_n^* \pmod n$, where $C_n^* \equiv a^{b_n^*} \cdot H_1(ID_n^*) \pmod n$.
- B obtains $(V_n^*/V_n^{**})^3 \equiv (C_n^*)^{e^{**} - e^*} \pmod n$.
- If $(e^{**} - e^*) \equiv 1 \pmod 3$, there is some $x \in Z_p^*$ satisfies the equation $(e^{**} - e^*) = 3x + 1$. So we obtain $(V_n^*/V_n^{**})^3 \equiv (C_n^*)^{3x+1} \pmod n$, and therefore $C_n^* \equiv \left(\frac{V_n^*}{V_n^{**} \cdot (C_n^*)^x}\right)^3 \pmod n$.

- If $(e^{**}-e^{*}) \equiv 2 \pmod 3$, there is some $x \in Z_p^*$ satisfies the equation $(e^{**}-e^{*}) = 3x-1$. So we obtain $(V_n^*/V_n^{**})^3 \equiv (C_n^*)^{3x-1} \pmod n$, and therefore $C_n^* \equiv \left(\frac{V_n^{**}\cdot(C_n^*)^x}{V_n^*}\right)^3 \pmod n$.

Then, if $(e^{**}-e^{*}) \not\equiv 0 \pmod 3$, B obtains the cubic root of C_n^*. And B can look up the list $H_{1,list}$ and obtain another cubic root of C_n^*. Then, B obtains two cubic roots of C_n^*. If the two cubic roots are not equal, B can factor n according to Theorem 2.

Since e^*, e^{**} are picked randomly, the probability of $(e^{**}-e^{*}) \not\equiv 0 \pmod 3$ is $\frac{2}{3}$, and the probability that the two cubic roots of C_n^* are inequal is $\frac{2}{3}$.

Next, we will analyze the probability of A_2 successfully forging two valid delegations similar to [3].

Let ϵ_2^* denote the probability of A_2 forging a delegation in a single run, and ϵ_2 denote the probability of A_2 forging a delegation in the real attack.

In $H_{2,list}$, all the records (w, R, e) are filled by H_2-oracle query and DelGen oracle query. So there are, at most $q_{H_2} + q_D$, different R's. For every DelGen oracle, B randomly selects $V_n, \tau \in \{0,1\}^l$, computes $R_n = V_n^3 \cdot (a^{b_n} \cdot H_1(ID_n))^{\tau}$ and $R = \prod_{i=1}^{n} R_i$, therefore, R can be considered as the random cubic residue modulo n. Obviously, the number of elements in cubic residues modulo n is $(3 \cdot 2^k)$. So the probability that R is in the $H_{2,list}$ is, at most $\frac{q_{H_2}+q_D}{3\cdot 2^k}$. So the probability of A_2 forging a delegation in a single run is $\epsilon_2^* \geq \epsilon_2 - \frac{q_D\cdot(q_{H_2}+q_D)}{3\cdot 2^k}$.

Let p_i denote the probability of forgery based on the $i^{th} H_2$-oracle query in a single run; then

$$\epsilon_2^* = \sum_{i=1}^{q_{H_2}+1} p_i.$$

Let $p_{i,s}$ denote the probability of forgery together based on $i^{th} H_2$-oracle query with input s, where s is a specific random tape input of length m. Then

$$2^m \cdot p_i = \sum_{s\in\{0,1\}^m} p_{i,s}.$$

For a specific random tape s, since twice valid forgery need different outputs of H_2-oracle query, the probability of twice forgery based on the same $i^{th} H_2$-oracle query is $p_{i,s}\cdot(p_{i,s}-2^{-l})$. Let P_i denote the probability of twice forgery based on the same $i^{th} H_2$-oracle query in two runs; then

$$P_i = \sum_{s\in\{0,1\}^m} 2^{-m}\cdot p_{i,s}\cdot(p_{i,s}-2^{-l}) \geq p_i^2 - 2^{-l}\cdot p_i.$$

So, the probability of twice forgery based on the same H_2-oracle query in two runs is $\sum_{i=1}^{q_{H_2}+1} P_i$. We have

$$\sum_{i=1}^{q_{H_2}+1} P_i \geq \sum_{i=1}^{q_{H_2}+1} p_i^2 - \sum_{i=1}^{q_{H_2}+1} 2^{-l}\cdot p_i \geq \frac{(\epsilon_2^*)^2}{q_{H_2}+1} - \frac{\epsilon_2^*}{2^l}$$

$$\geq \frac{\left(\epsilon_2 - \frac{q_D\cdot(q_{H_2}+q_D)}{(3\cdot 2^k)}\right)^2}{q_{H_2}+1} - \frac{\epsilon_2 - \frac{q_D\cdot(q_{H_2}+q_D)}{(3\cdot 2^k)}}{2^l}.$$

Taking $(e^{**}-e^{*}) \not\equiv 0 \pmod 3$ and the difference of the two cubic roots of C_n^* into account, the probability of factoring n is $\epsilon' \geq \frac{4}{9}\sum_{i=1}^{q_{H_2}+1} P_i \geq \frac{4}{9}\cdot\left(\frac{(\epsilon_2-\delta_2)^2}{q_{H_2}+1} - \frac{\epsilon_2-\delta_2}{2^l}\right)$, where $\delta_2 = \frac{q_D\cdot(q_{H_2}+q_D)}{3\cdot 2^k}$. So, the theorem is proved. □

As to the running time, according to [3], B has to run A_2 twice and perform some other operations to factor n. So B should spend the time $t' = 2t + O(k^2\cdot l + k^3)$ to factor n.

Theorem 4. *Given a security parameter (k, l), if the integer factorization problem is (t', ϵ')-hard, then our identity-based proxy multi-signature scheme is $(t, q_{H_4}, q_S, \epsilon_3)$-secure against existential forgery under adaptive chosen-message and identity attacks for the Type 3 adversary A_3, which satisfies:*

$$\epsilon' \geq \frac{4}{9}\cdot\left(\frac{(\epsilon_3-\delta_3)^2}{q_{H_4}+1} - 2^{-l}\cdot(\epsilon_3-\delta_3)\right)$$

$$t' = 2t + O\left(k^2\cdot l + k^3\right),$$

where q_{H_4} and q_S denote the number of queries that A_3 can ask to the random oracle H_4 and PMSign, respectively, and $\delta_3 = \frac{q_S\cdot(q_{H_4}+q_S)}{3\cdot 2^k}$.

Proof. This proof is similar to that of Theorem 3. So, we just describe the main difference with Theorem 3 as follows:

Step 1. Algorithm B does the same as Step 1 of Theorem 3.

Step 2. B deletes D_{list} list and adds $H_{3,list}$, $H_{4,list}$, S_{list} lists, and initializes them as null.

Step 3. B deletes DelGen oracle and adds H_3, H_4 and PMSign oracle accordingly.

- **H_3-oracle:** A_3 requests H_3 on (ID_{ps}, w, R), B checks if (ID_{ps}, w, R) existed in $H_{3,list}$. If not, B picks a random $\mu \in \{0,1\}^l$ and adds the tuple (ID_{ps}, w, R, μ) into $H_{3,list}$; then B returns $H_3(ID_{ps}, w, R) = \mu$ to A_3.
- **H_4-oracle:** A_3 requests H_4 on $(ID_{ps},\ w,\ m,\ R_{ps})$, and B checks if $(ID_{ps}, w, m,\ R_{ps})$ existed in $H_{4,list}$. If not, B picks a random $\eta \in \{0,1\}^l$ and adds the tuple $(ID_{ps}, w, m,\ R_{ps}, \eta)$ into $H_{4,list}$; then, B returns $H_4(ID_{ps}, w, m,\ R_{ps}) = \eta$ to A_3.
- **PMSign oracle:** A_3 requests PMSign algorithm on (w, m). A_3 randomly selects $r_i \in Z_n^*$ and computes $R_i = r_i^3 \pmod n$, $R = \prod_{i=1}^{n} R_i \pmod n$, and requests H_2-oracle query and obtains $H_2(w, R) = e$. Since A_3 knows all the

Table 1: Comparison of security

Scheme	Security Proof Method	Mathematics Tool	Assumption*
Cao and Cao [4]	Random oracle	bilinear pairings	CDH
Our scheme	Random oracle	Cubic residues	IFP

*CDH stands for computational Diffie-Hellman assumption, and IFP stands for integer factorization problem.

Table 2: Comparison with other schemes

Scheme	Extract	DelGen	DelVeri	PMKGen	PMSign	PMVeri	Total	Total Time (ms)
Cao and Cao [4]	$1M_p$ $+1H_M$	$2M_p$ $+1H_M$	$2H_M$ $+3O_P$	$1M_p$	$2M_p$ $+1H_M$	$1M_p$ $+3H_M$ $+4O_P$	$7M_p$ $+8H_M$ $+7O_P$	209.26
Our scheme	$1E_n$	$1E_n$	$1E_n$	$1E_n$	$1E_n$	$3E_n$	$8E_n$	42.48

Table 3: Cryptographic running time (ms)

Modular Exponentiation	Pairing	Pairing-based Scalar Multiplication	Map-to-point Hash
5.31	20.04	6.38	3.04

secret keys of original signers, A_3 can compute $V_i \equiv r_i \cdot s_{ID_i}^e \pmod n$ and obtain all the delegation $D_{i\to ps} = (ID_i, b_i, w, R_i, V_i)$, $i = 1, 2, \cdots, n$. A_3 sends $D_{i\to ps}$, $i = 1, 2, \cdots, n$, to B to request PMSign algorithm on (w, m). B computes $R \equiv \prod_{i=1}^{n} R_i \pmod n$ and obtains $H_3(ID_{ps}, w, R) = \mu$ by looking up the list $H_{3,list}$ - in H_3-oracle. B picks random V_p, $\varsigma \in \{0,1\}^l$, and computes $C \equiv \prod_{i=1}^{n} (a^{b_i} \cdot h_{1,i}) \pmod n$, $C_{ps} \equiv a^{b_p} \cdot h_{1,ps} \pmod n$, $V = \prod_{i=1}^{n} V_i$, $R_{ps} \equiv V_{ps}^3 \cdot ((C_{ps})^{\mu} \cdot C^{\eta}/R)^{\varsigma} \pmod n$. If R_{ps} already exists in $H_{4,list}$, B returns failure, else returns $(ID_1, ID_2, \cdots, ID_n, ID_p, b_1, b_2, \cdots, b_n, b_p, m, w, R, R_p, V_p)$ as proxy multi-signature of (w, m) to A_3. B adds the tuple $(ID_1, ID_2, \cdots, ID_n, ID_p, b_1, b_2, \cdots, b_n, b_p, m, w, R, R_p, V_p)$ into S_{list}, and adds $(ID_{ps}, w, m, R_{ps}, \varsigma)$ into $H_{4,list}$.

Step 4. A_3 outputs a proxy multi-signature forgery of (w, m) with $\sigma^* = (ID_1^*, ID_2^*, \cdots, ID_n^*, ID_{ps}^*, b_1^*, b_2^*, \cdots, b_n^*, b_{ps}^*, m^*, w^*, R^*, R_{ps}^*, V_{ps}^*)$, which ID_{ps}^* has not be requested on the Extract oracle, and (m^*, w^*) has not be requested on the PMSign oracle.

Step 5. Similar with Theorem 3, B resets A_3 twice with the same random tape, and gives the different random number until A_3 asks H_4-oracle. And A_3 can forge two proxy multi-signatures with the same value R_{ps}. B can resolve integer factorization problem with A_3's proxy multi-signature forgery.

As to the probability and running time, both of them are similar with Theorem 3. □

Furthermore, by Theorems 3 and 4, we can conclude Theorem 5 easily.

Theorem 5. *Given a security parameter (k, l), if the factoring problem is (t', ϵ')-hard, then our identity-based proxy multi-signature scheme is $(t, q_{H_2}, q_{H_4}, q_D, q_S, \epsilon)$-secure against existential forgery under adaptive chosen-message and identity attacks, which satisfies:*

$$\epsilon' \geq \frac{4}{9} \cdot \left(\frac{(\epsilon - \delta)^2}{2 \cdot \max\{q_{H_2}+1, q_{H_4}+1\}} - 2^{-l} \cdot (\epsilon - \delta) \right)$$

$$t' = 2t + O\left(k^2 \cdot l + k^3\right),$$

where $\epsilon = \epsilon_2 + \epsilon_3$ and $\delta = \delta_2 + \delta_3$.

We conclude that our scheme is secure against existential forgery under adaptive chosen-message and identity attacks under integer factorization problem assumption.

6 Comparison and Performance

In this section, we compare our scheme with Cao and Cao's IBPMS scheme [4]. The two schemes are provable security based on different hardness assumptions in the random oracle model. We describe them in detail in Table 1.

In order to simplify the complexity, we used the method of [5], which considers only a single original signer. Let M_p, H_M, O_P, E_n denote one pairing-based scalar multiplication, map-to-point hash function, pairing operation, and modular exponentiation, respectively. In order to make our analysis clearer, we changed the total computation cost into running time in the last column of Table 2 according to Table 3, which is referred to reference [8].

According to Tables 1 and 2, our schemes total running time decreased drastically compared with Cao and Cao's scheme [4]. The security of our scheme is based on integer factorization problem assumption without bilinear pairing. We note that the integer factorization problem assumption is 2500 years old.

7 Conclusions

Identity-based proxy multi-signature has proposed for years, and several schemes have been proposed. However, most of the existing scheme is based on bilinear pairing or elliptic curve. In this paper, we propose an efficient identity-based proxy multi-signature scheme using cubic residues. The security of our scheme is based on the integer factorization problem assumption, which is more reliable and easier to use because it has been developed 2500 years ago. Our scheme is prove security against existential forgery under adaptive chosen-message and identity attacks. Furthermore, the efficiency of our scheme is higher than the existing scheme based on bilinear pairing such as Cao and Cao's scheme etc.

Acknowledgments

The authors gratefully acknowledge the anonymous reviewers for their valuable comments.

References

[1] M. R. Asaar, M. Salmasizadeh, and W. Susilo, "An identity-based multi-proxy multi-signature scheme without bilinear pairings and its variant," *The Computer Journal*, vol. 58 , no. 4, pp. 1021–1039, 2015.

[2] A. Boldyreva, A. Palacio, and B. Warinschi, "Secure proxy signature schemes for delegation of signing rights," *Journal of Cryptology*, vol. 25, no. 1, pp. 57–115, 2012.

[3] Z. C. Cai, X. L. Dong, and Z. F. Cao, "Identity based signature scheme based on quadratic residues," *Science in China Series F: Information Sciences*, vol. 39, no. 2, pp. 199–204, 2009.

[4] F. Cao, and Z. F. Cao, "A secure identity-based proxy multi-signature scheme," *Information Sciences*, vol. 179, no. 3, pp. 292–302, 2009.

[5] X. F. Cao, and W. D. Kou, "A pairing-free identity-based authenticated key agreement protocol with minimal message exchanges," *Information Sciences*, vol. 180, no. 15, pp. 2895–2903, 2010.

[6] Cryptography Stack Exchange, *Why Is Elliptic Curve Cryptography Not Widely Used, Compared to RSA?*, Nov. 15, 2011. (http://crypto.stackexchange.com/questions/1190/why-is-elliptic-curve-cryptography-not-widely-used-compared-to-rsa).

[7] M. L. Das, A. Saxena, and D. B. Phata, "Algorithms and approaches of proxy signature: A survey," *International Journal of Network Security*, vol. 9, no. 3, pp. 264–284, 2009.

[8] D. B. He, J. H. Chen, and R. Zhang, "Efficient and provably-secure certificateless signature scheme without bilinear pairings," *International Journal of Communication Systems*, vol. 25, no. 11, pp. 1432–1442, 2012.

[9] X. X. Li, and K. F. Chen, "ID-based multi-proxy signature, proxy multi-signature and multi-proxy multi-signature schemes from bilinear pairings," *Applied Mathematics and Computation*, vol. 169, no. 1, pp. 437–450, 2005.

[10] M. Mambo, K. Usuda, and E. Oamoto, "Proxy signatures: delegation of the power to sign messages," *IEICE Transactions on Fundamentals of Electronic Communications and Computer Science*, vol. E79-A, no. 9, pp. 1338–1354, 1996.

[11] C. H. Pan, S. P Li, Q. H. Zhu, C. Z. Wang, and M. W. Zhang, "Notes on proxy signcryption and multi-proxy signature schemes," *International Journal of Network Security*, vol. 17, no. 1, pp. 29–33, 2015.

[12] D. Pointcheval, and J. Stern, "Security proofs for signature schemes," in *Advances in Cryptology (Eurocrypt'96)*, LNCS 1070, pp. 387–398, Springer, May 1996.

[13] D. Pointcheval, and J. Stern, "Security arguments for digital signatures and blind signatures," *Journal of Cryptography*, vol. 13, no. 3, pp. 361–396, 2000.

[14] R. A. Sahu, and S, Padhye, "Provable secure identity-based multi-proxy signature scheme," *International Journal of Communication Systems*, vol. 28, no. 3, pp. 497–512, 2015.

[15] A. Shamir, "Identity based cryptosystems and signature schemes," in *Proceedings of Advances in Cryptology (CRYPTO'84)*, LNCS 196, pp. 47–53, Springer, 1984.

[16] H. Singh, and G. K. Verma, "ID-based proxy signature scheme with message recovery," *Journal of Systems and Software*, vol. 85, no. 1, pp. 209–214, 2012.

[17] Y. Sun, C. X. Xu, Y. Yu, and B. Yang, "Improvement of a proxy multi-signature scheme without random oracles," *Computer Communications*, vol. 34, no. 3, pp. 257–263, 2011.

[18] N. Tiwari, and S. Padhye, "An ID-based proxy multi signature scheme without bilinear pairings," in *Proceedings of First International Conference on Security Aspects in Information Technology*, LNCS 7011, pp. 83–92, Springer, 2011.

[19] N. Tiwari, S. Padhye, and D. He "Provably secure proxy multi-signature scheme based on ECC," *Information Technology And Control*, vol. 43, no. 2. pp. 198–203, 2014.

[20] Q. Wang, and Z. F. Cao, "Identity based proxy multi-signature," *Journal of Systems and Software*, vol. 80, no. 7, pp. 1023–1029, 2007.

[21] Z. W. Wang, L. C. Wag, S. H. Zheng, Y. X. Yang, and Z. M. Hu, "Provably secure and efficient identity-based signature scheme based on cubic residues". *International Journal of Network Security*, vol. 14, no. 1, pp. 104–109, 2012.

[22] L. J. Yi, G. Q. Bai, and G. Z. Xiao, "Proxy multi-signature scheme: A new type of proxy signature scheme," *Electronics Letters*, vol. 36, no. 6, pp. 527–528, 2000.

[23] Y. Yu, Y. Mu, W. Susilo, Y. Sun, and Y. F. Ji, "Provably secure proxy signature scheme from factorization," *Mathematical and Computer Modelling*, vol. 55, no. 3-4, pp. 1160–1168, 2012.

Feng Wang was born in Shandong province, China, in 1978. He received his B.S. degree in Mathematics from Yantai Normal University (now named Ludong University), Yantai, in 2000 and the M.S. degree in Applied Mathematics from the Guangzhou University, Guangzhou, in 2006. Currently, he is a Lecturer in the College of Mathematics and Physics at Fujian University of Technology and a visiting scholar in Department of Information Engineering and Computer Science at Feng Chia University. His research interests include computer cryptography and information security.

Changlu Lin received the BS degree and MS degree in mathematics from the Fujian Normal University, P.R. China, in 2002 and in 2005, respectively, and received the Ph.D degree in information security from the state key laboratory of information security, Graduate University of Chinese Academy of Sciences, P.R. China, in 2010. He works currently for the School of Mathematics and Computer Science, and the Fujian Provincial Key Laboratory of Network Security and Cryptology, Fujian Normal University. He is interested in cryptography and network security, and has conducted research in diverse areas, including secret sharing, public key cryptography and their applications.

Shih-Chang Chang received his B.S. degree in 2005 and his M.S. degree in 2007, both in Department of Information Engineering and Computer Science from Feng Chia University, Taichung, Taiwan. He is currently pursuing his Ph.D. degree in Computer Science and Information Engineering from National Chung Cheng University, Chiayi, Taiwan. His current research interests include electronic commerce, information security, computer cryptography, and mobile communications.

Chin-Chen Chang received his Ph.D. degree in computer engineering from National Chiao Tung University. His first degree is Bachelor of Science in Applied Mathematics and master degree is Master of Science in computer and decision sciences. Both were awarded in National Tsing Hua University. Dr. Chang served in National Chung Cheng University from 1989 to 2005. His current title is Chair Professor in Department of Information Engineering and Computer Science, Feng Chia University, from Feb. 2005. Prior to joining Feng Chia University, Professor Chang was an associate professor in Chiao Tung University, professor in National Chung Hsing University, chair professor in National Chung Cheng University. He had also been Visiting Researcher and Visiting Scientist to Tokyo University and Kyoto University, Japan. During his service in Chung Cheng, Professor Chang served as Chairman of the Institute of Computer Science and Information Engineering, Dean of College of Engineering, Provost and then Acting President of Chung Cheng University and Director of Advisory Office in Ministry of Education, Taiwan. Professor Chang has won many research awards and honorary positions by and in prestigious organizations both nationally and internationally. He is currently a Fellow of IEEE and a Fellow of IEE, UK. On numerous occasions, he was invited to serve as Visiting Professor, Chair Professor, Honorary Professor, Honorary Director, Honorary Chairman, Distinguished Alumnus, Distinguished Researcher, Research Fellow by universities and research institutes. His current research interests include database design, computer cryptography, image compression and data structures.

Practical Implementation of a Secure Email System Using Certificateless Cryptography and Domain Name System

Suresh Kumar Balakrishnan[1] and V. P. Jagathy Raj[2]

(Corresponding author: Suresh Kumar Balakrishnan)

School of Computer and Information Sciences, Indira Gandhi National Open University (IGNOU)[1]
New Delhi 110068, India
(Email: suresh@sctimst.ac.in)
School of Management Studies, Cochin University of Science and Technology[2]
Kochi, India
(Email: jagathy@cusat.ac.in)

(Received Oct. 25, 2014; revised and accepted Jan. 16 & June 21, 2015)

Abstract

Email is currently the most widely used communication system in daily life. To improve security and efficiency, most email systems adopt Public Key Infrastructure (PKI) as the mechanism to implement security, but PKI based systems suffer from expensive certificate management and problems in scalability. Identity Based Cryptography (IBC) is another method, but it has the inherent drawback of Key Escrow. This paper proposes an implementation of a practical, secure email system based on certificateless cryptography, which uses Domain Name System (DNS) as the infrastructure for public key exchange and a secure key token/fingerprint authentication system for user authentication. The message payload is encrypted by a per-email symmetric key generated from a secret value, the public and private keys of both the sender and the receiver. The proposed mailing system is secure against standard security model.

Keywords: Certificateless cryptography, domain name system, identity based cryptography, multi-factor authentication, public key infrastructure, secure email system

1 Introduction

The main reason for using email is probably the convenience and speed with which it can be transmitted, irrespective of geographical distances. Similar to a postcard, an email has open access to the systems on its path. If anyone wants to intercept, copy or alter information, they can easily do so. Confidential information, such as bank statements, trade secrets, and even national secret information, is being exchanged through emails. Therefore, the contents of emails are more important and valuable than ever, and their security has raised many concerns. The main reason for not using encryption in email communications is that current email encryption solutions require expensive operations and hard key management. Therefore, research on simple, highly secure and efficient email systems are in great need.

Current email systems that use symmetric and asymmetric cryptographic schemes [12] suffer from key management problems. Identity Based Cryptography (IBC) [4, 19, 30] systems, which have been proposed to address such key management issues, also suffer from the key escrow problem, which violates the non-repudiation feature that should be offered by security systems.

This paper proposes a practical implementation of a secure email system in an open standard as an alternative technology for eliminating the problems with PKI and identity-based cryptographic mailing systems. This system uses the certificateless public key cryptography scheme by Al-Riyami and Paterson [29], Domain Name System (DNS) [11, 25, 26], for publishing a user's public key details and multi-factor user authentication for secure user authentication with the system.

The rest of the paper is organized as follows. Related works on existing email security systems and an introduction to certificateless cryptography are described in Section 2. Section 3 describes the design of the proposed system. Section 4 describes the implementation of the system. The security features of the paper are described in Section 5. Finally, benchmarking and conclusions are given in Section 6.

2 Literature Review

2.1 Existing Schemes to Secure Email Systems

The majority of client-based email security systems are based on Identity-Based Cryptography or Public Key Infrastructure (PKI) [8, 9, 21] schemes. The above security functions are implemented by these solutions, of which the most competitive ones are S/MIME [27] and PGP [3, 32, 36]. PGP uses hash functions and public key encryption algorithms, for example, RSA [12, 28] and MD5 [12], to enable encryption for content-protection and digital signature for non-repudiation. RSA public keys are attached as PGP certificates along with the message. However, self-signed PGP certificates are used for most users and form a chain-based credential trust network. This trust mode of PGP is only suitable for small-scale groups and is not suitable for large-scale groups or anonymous user environments. Moreover, it is very difficult to notify other users in the network, if the private key of a PGP user has been compromised. S/MIME employs the PKI framework. Due to the difficulty of certificate management in PKI, S/MIME cannot ignore tedious operations, such as certificate revocation, verification, and so on. In addition, both S/MIME and PGP use RSA for encrypting and signing email contents. This results in lower efficiency compared with Elliptic Curve Cryptography (ECC) [2, 14] with the same level of security. In the IBC scheme, it is difficult to prove the self-identity of the Trust Authority (TA) or the Key Generation Center (KGC) [35]. The scheme also suffers from the problem of key escrow, where a central trusted authority issues a private key to a user. Because a central authority is responsible for private key generation, it is able to work as an authorized user and could maliciously decipher the incoming encrypted text or generate false signatures. Several methods [7, 22, 34] have been proposed to solve the key escrow problem in IBC, and they can be easily classified into two groups based on the private key generation technique: (i) Multiple authority approach and (ii) User chosen secret key information approach. As per our survey, numerous techniques [5, 14, 15, 16, 24, 31] follow the multiple authority approach, while very few techniques [17, 29] are based on the secret key information approach. In the multiple authority approach, the critical task of private key generation is distributed among several authorities, and as a result, no single authority can perform any unauthenticated work. Although these methods successfully solve the key escrow problem, they introduce extra overhead on systems and lack of central control on key issuing policy and are not suited for email security systems. User chosen secret information approaches are either certificate based or certificateless. The certificate based scheme completely overcomes key escrow; however, it loses the advantage of an ID based scheme. The secret key exchange protocol based system is also not suited for email systems because a receiver of the email system may not always be online.

Domain Keys Identified Mail (DKIM) [10] permits users to claim some responsibility for a message by associating it with a domain name that they are authorized to use. This claim is validated through a cryptographic signature and by querying the Signer's domain directly to retrieve the appropriate public key. The approach taken by DKIM differs from previous approaches to message signing such S/MIME and OpenPGP [3] in that:

- The message signature is written as a message header field instead of part of the message body, so that neither human recipients nor existing MUA (Mail User Agent) software are confused by signature-related content appearing in the message body.
- There is no dependency on well-known trusted authority public and private-key pairs.

A new concept called Lightweight Signatures for Email (LES) [1], proposed by Ben Adida, David Chau, Susan Hohenberger, and Ronald L. Rivest, is an extension to DKIM. In LES, individual users authenticate within a domain, without requiring additional user authentication infrastructure. LES allows a user to send emails via any outgoing mail server, not just the official outgoing mail server mandated by DKIM. LES also supports repudiable signatures to protect users' privacy. Both DKIM and LES focus only on email authentication. LES requires a modified email client for authentication.

The Proxy based email system [6, 13, 18, 23] is another scheme that has the key escrow problem.

The scheme described as "An End-to-End Secure Mail System Based on Certificateless Cryptography in the Standard Security Model" [20] based on of Certificateless Public Key Cryptography (CL-PKC) [29] is not suitable for practical implementation with different domains. None of these works are satisfactory. Therefore, efficient email security systems are in great need.

This paper proposes a practical implementation of a secure email system using certificateless cryptography as an alternative technology for eliminating the problems with PKI and IBC based mailing systems.

2.2 Certificateless Public Key Cryptography

The concept of Certificateless Public Key Cryptography (CL-PKC) is introduced by Al-Riyami and Paterson [29] in 2003, to overcome the key escrow problem of identity-based cryptography. In CL-PKC, a trusted third party, called the Key Generation Center (KGC), supplies a user with a partial private key. The user then obtains his/her full private key by combining the partial private key with a secret value that is defined by the user and is unknown to the KGC. Thus, the KGC does not know the user's private keys. Subsequently, the user combines the his/her secret value with the KGC's public parameters to compute his/her public key.

Compared to identity based public key cryptography (IDPKC), the trust assumptions regarding the trusted third party in this scheme are significantly reduced. In CL-PKC, users can generate any number of private-public key pairs for the same partial private key. To guarantee that the KGC does not replace a user's public keys, they proposed a binding technique to bind a user's public key with his/her private key. In their binding scheme, the user first fixes his/her secret value and generates his/her public key and supplies the public key to KGC. Then, the KGC redefines the user identity as the user's identity concatenated with his/her public key. Using this binding scheme, the replacement of a public key of a user in the system by the KGC is equivalent to certificate forgery by a CA in a traditional PKI system.

2.3 Al-Riyami and Paterson Scheme

In the proposed secure mailing system, the CL-PKC concept, as proposed by Al-Riyami and Paterson, is used. The general description of the algorithms introduced by Alriyami and Paterson is provided. These algorithms are Setup, Set-Secret-Value, Partial-Private-Key-Extract, Set-Private-Key and Set-Public-Key.

Let k be a security parameter given to the Setup algorithm and IG be a Bilinear Diffie-Hellman Problem(BDH) parameter generator with input k.

Setup. (This operation is performed by the KGC): This algorithm runs as follows:

1) Run IG on input k to generate output $< G_1, G_2, e >$, where G_1 and G_2 are groups of some prime order q and e: $G_1 \times G_1 \rightarrow G_2$ is a pairing.
2) Choose an arbitrary generator $P \in G_1$.
3) Select a master-key s uniformly, at random, from Z_q^*, and set $P_0 = sP$.
4) Choose cryptographic hash functions H_1: $\{0,1\}^* \rightarrow G_1^*$ and H_2: $G_2 \rightarrow \{0,1\}^n$, where n is the bit-length of the plaintexts taken from some message space $M = \{0,1\}^n$ with a corresponding cipher text space $C = G_1 \times \{0,1\}^n$. Then, the KGC publishes the system parameters $params = < G_1, G_2, e, n, P, P_0, H_1, H_2 >$, while the secret master-key s is securely saved by the KGC.

Set-Secret-Value. (performed by the user): This algorithm takes as inputs *params* and entity m's identifier ID_m. Entity m selects $x_m \in Z_q^*$ at random and outputs x_m as m's secret value. Then, it computes $X_m = x_m P$ and sends X_m to the KGC.

Partial-Private-Key-Extract. (performed by the KGC): This algorithm takes as inputs an identifier $ID_m \in \{0,1\}^*$ and X_m and carries out the following steps to construct the partial private key for entity m with identifier ID_m.

1) Compute $Q_m = H_1(ID_m || X_m)$.
2) Output the partial private key $D_m = sQ_m \in G_1$.

Entity m, when armed with its partial private key D_m, can verify the correctness of the partial private key D_m by checking $e(D_m, P) = e(Q_m, Q_m)$.

Set-Private-Key. (performed by the user): This algorithm takes as inputs *params*, entity m's partial private key D_m and m's secret value $x_m \in Z_q^*$. Entity m transforms the partial private key D_m to private key S_m by computing

$$S_m = x_m D_m = x_m s Q_m \in G_1.$$

Set-Public-Key. (performed by the user): This algorithm takes as input *params* and entity m's secret value $x_m \in Z_q^*$ and constructs m's public key as $P_m =< X_m, Y_m >$, where $X_m = x_m P$ and $Y_m = x_m Q_m = x_m sP$.

3 System Design

The proposed secure e-mail system should securely exchange e-mail messages, be easy to use, make use of the existing secure e-mail standards, and it should be applied without making significant changes to the structure of the email communication system. In order to achieve this goals, some decisions are made before designing the system:

- The first question to be answered is whether to apply security to both the e-mail client and server, or just one of them. Any change in the e-mail servers is not recommended, since this implies that all the e-mail servers around the world should be updated to implement the new changes. Hence, this design would apply security to email clients only, and this will allow any organization to apply this system without having to modify the underlying network architecture.

- The analysis of the current encryption schemes shows that different aspects of the key distribution technology have attracted criticism, and shows that most of these aspects are related directly to the digital certificates management complexity. On the other hand, CLPKC provides a comparable security and an equivalent functionality, and does not need any digital certificates. Thus, CLPKC represents an excellent replacement to the existing email security technology, and it will be adopted in the design of this system.

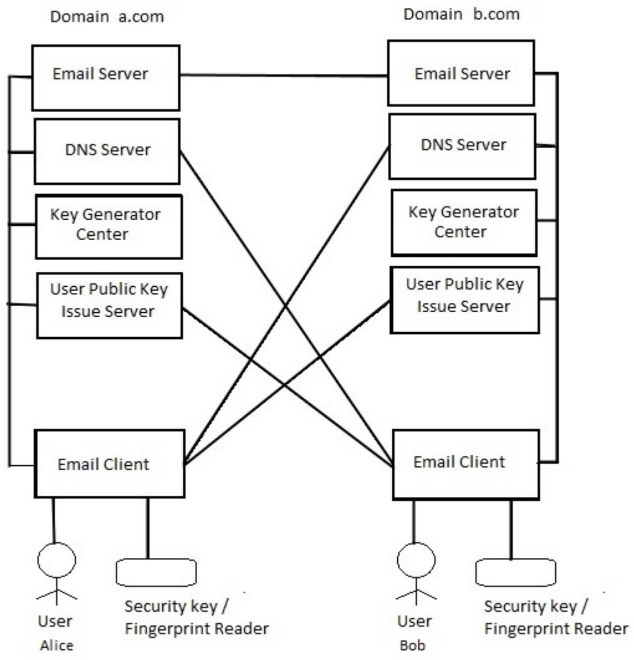


Figure 1: System architecture

3.1 Building Blocks

Figure 1 illustrates the architecture of the proposed system. Consider a user Alice at the domain *a.com* who wishes to send a secure email to user Bob at the domain *b.com*. Various components and operations of the proposed system are described as follows.

A user public key issue server issues public keys for users. The functionality of the user public key issue service may be implemented as part of the KGC server for small user base domains.

The user registration module is incorporated within the user public key issue server. During the user registration phase, the system asks for an email address, a password and a secret value (x_m). The user database can be linked to the email server. The registration module will contact KGC to obtain public parameters *params* and update the KGC with x_mP. The registration module also calculates the public key of the user and stores it in the user public key issue server for distribution. To keep the secret value with the user safely, we propose a usb security key token to store the secret with password protection.

An email plug-in attached to an email client is proposed for signing, verifying, encrypting and decrypting. If web-based email is used, all of the security functionalities should be incorporated in the email web server with the help of a client side code execution module (for example, java code under java virtual machine). During the security operation, the plugin module attached to the email client or client side code execution module reads the secret value stored in the USB security token with the help of software drivers.

Multi-factor (i.e., two or more factors) authentication is now a requirement for effective secure authentication. Multi-factor authentication is commonly defined as:

- Something the user knows (e.g., Password, PIN);
- Something the user has (e.g., ATM card, smart card, security token); and
- Something the user is (e.g., Biometric characteristic, such as a fingerprint, palm pattern).

For an efficient security system, it is recommended to use “authentication methods that depend on more than one” of these three factors (i.e., “multi-factor” authentication).

The same security token or a finger print system [33, 35] can be used along with a conventional username and password for authentication during registration or the update of a secret value. Biometrics, which refers to distinctive physiological and behavioral characteristics of human beings, are more reliable means of authentication than a traditional password based system. The fingerprint is the most widely used biometric because of its uniqueness and immutability. For the fingerprint system, the user has to be enrolled with the user public key server, while registration and a special software plug-in are required during authentication to verify the current captured image with the previously recorded fingerprint.

3.2 User Public-Key Distribution

DNS provides a well-established and trusted naming and routing infrastructure for domains. Hostname to IP address mappings and mail routing (using MX records) rely on it. Recently, DNS has been proposed as a public-key infrastructure with Domain Keys Identified Mail (DKIM) [2] by the Internet Engineering Task Force (IETF). In the DKIM scheme, public keys generated by RSA scheme are stored in specially named DNS records. Specifically, DKIM reserves the domainkey subdomain for every domain with an MX record. Any entries within this subdomain are public keys in base64-encoding with some associated parameters. By keeping each public-key record short (i.e., less than the size of a single UDP packet), this DNS-based key distribution mechanism is functional across a large portion of existing DNS servers.

Our proposed system uses the same technique as in DKIM for specifying the address of the user public key issue server of the domain.

3.3 Domain Setup

Figure 2 shows the domain setup steps. In the basic domain setup, Alice, with email address alice@a.com, will obtain her partial secret key from a key server (KGC) dedicated to her domain a.com.

The detailed setup procedure of the domain is defined as follows:

- Choose an identity-based signature scheme from the various schemes, e.g., the Boneh and Franklin method.

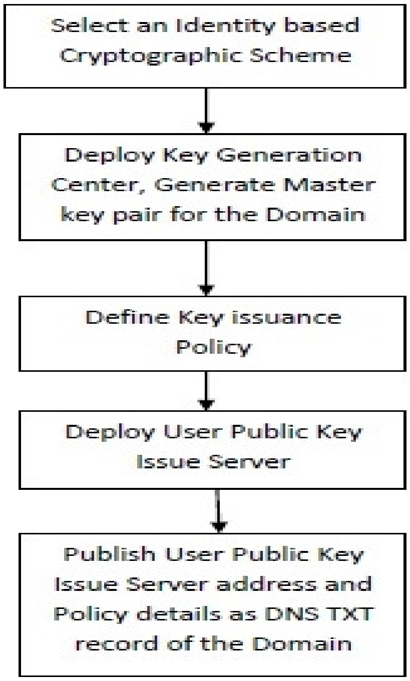

Select an Identity based Cryptographic Scheme
Deploy Key Generation Center, Generate Master key pair for the Domain
Define Key issuance Policy
Deploy User Public Key Issue Server
Publish User Public Key Issue Server address and Policy details as DNS TXT record of the Domain

Figure 2: Domain setup

- Generate a master key pair (public and master secret) for this scheme.
- Define key issuance policy *Policy* for the system. This includes definition on whether emails originating from this domain can, should, or must be signed.
- Publish the user public key issue server details and *Policy* as a DNS TXT record corresponding the MX record for a.com. The DNS record content is formatted as ibclses=<name of user public key issue server>, policy=<Policy>.

User Identities. A user has to first register to generate a secret value, store it in a USB security token and then pass the generated public key to the public key issue server. A user's public key can be derived by calling the user public key web service and passing the user's identity through *id_string*. We propose a standard format for *id_string* to specifically support email address authentication with domain policies and key expiration mechanisms.

Key Expiration. To provide simple key revocation capabilities, the user identity string includes key expiration information. Specifically, *id_string* includes the last date on which the key is valid - the expiration date -as a formatted character string. An id string is thus constructed as: <email>, <expiration date>.

For example, an identity string *id_string* for Bob would be: bob@b.com, 2014-07-10.

3.4 Domain Policies

Once an email domain decides to deploy an email security system, it simply needs to create a key server and a user public key distribution server for the domain and specify this server address in the appropriate DNS record. We propose that this record include a Policy parameter to specify how the domain chooses to participate in the secure email architecture. Policy determines the domain's requirements on its email users as well as its guarantees to any recipients. Three external Sign policy values are used:

None. Users of this domain may sign their emails. If the signature and verification fails, no warning header will be added by the recipient email signature verification system.

Basic. Users of this domain may sign emails with keys issued by this key server. If the signature and verification fails, a warning header will be added by the recipient email signature verification system.

Strong. Users of this domain are required to sign all of their emails with a key issued by this domain. The message will be rejected if verification fails.

Internal policies can also be implemented by adding some standing instructions to the email client. Examples include:

alice@a.com * E - Encrypt all messages from alice@a.com

alice@a.com bob@b.com S - Sign the message from Alice to Bob.

*@a.com tax@gov.gv ES - Sign and Encrypt all messages from domain a.com to tax@gov.gv

3.5 The Proposed System

Figure 3 shows the steps for sending secure email.

Assume that client Alice has a private key, $S_{Alice} = x_{Alice}D_{Alice}$, and a public key $P_{Alice} =< X_{Alice}, Y_{Alice} >$, while client Bob has a private key, $S_{Bob} = x_{Bob}D_{Bob}$, and a public key, $P_{Bob} =< X_{Bob}, Y_{Bob} >$. The public keys of all clients are available through the secure web service on the User public key issue server. The working of the security functionality is based on how the internal and external domain policies are specified for the domain.

The basic operation of the security functionality is as follows:

Encryption.

1) Assemble *id_string*$_{Bob}$, an identity string for Bob, using the current date+15 days as the expiration date to view the message within 15 days: bob@b.com, 2012-07-31.

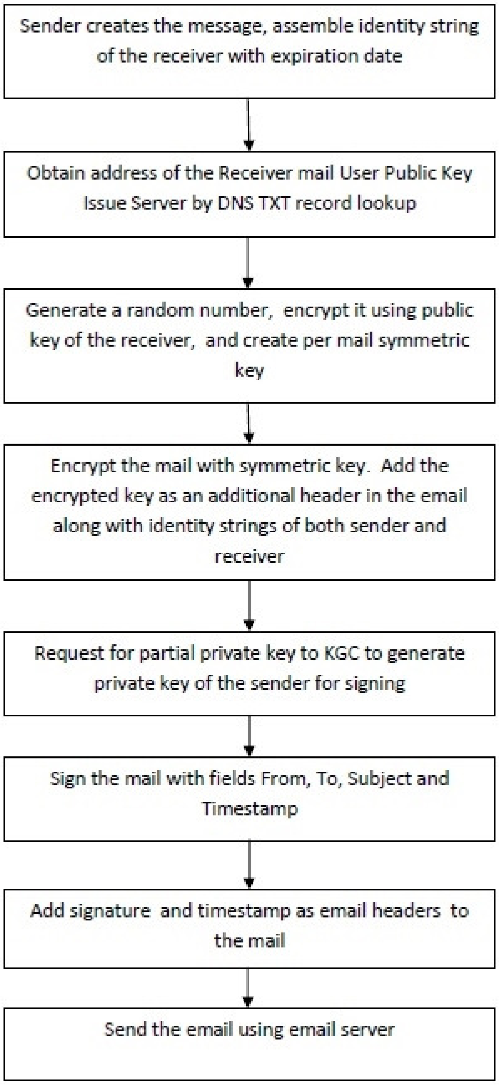

Sender creates the message, assemble identity string of the receiver with expiration date
Obtain address of the Receiver mail User Public Key Issue Server by DNS TXT record lookup
Generate a random number, encrypt it using public key of the receiver, and create per mail symmetric key
Encrypt the mail with symmetric key. Add the encrypted key as an additional header in the email along with identity strings of both sender and receiver
Request for partial private key to KGC to generate private key of the sender for signing
Sign the mail with fields From, To, Subject and Timestamp
Add signature and timestamp as email headers to the mail
Send the email using email server

Figure 3: Sending secure email

2) Obtain the address of the User public key issue server for the domain b.com using DNS TXT record lookup. Collect the public key of Bob P_{Bob} from the user key issue server for Domain b.com using secure web services. Then, check that $\hat{e}(X_{Bob}, Q_m) = \hat{e}(Y_{Bob}, P)$ to authenticate the public key of client Bob.

3) Generate a random number $t \in Z^*$ and encrypt it using the public key of client Bob as $t^* = EP_{Bob}(t)$. The random number has the feature of Perfect Forward and Backward Secrecy, which is always fresh and unrelated to any past or future sessions.

4) Compute $K_{AliceBob} = tx_{Alice}X_{Bob}$, and then compute the per-mail symmetric key

$$K = H_2(Q_{Alice}||Q_{Bob}||K_{AliceBob}).$$

5) Encrypt the mail M by using the AES algorithm with the symmetric key K as $M^* = E_K(M)$.

6) Add the encrypted value t^* at the beginning of the encrypted mail M^* (i.e., $M^* = M||t^*$) as additional headers as shown below:

   - The exact id_string_{Alice} in SMTP header *X-IBCLSES-Sender-IDString*.
   - The exact id_string$_{Bob}$ in SMTP header X-IBCLSES-Recipient-IDString.
   - t^* in base64-encoding in SMTP header X-IBCLSES-Encryption-Key

Signing.

1) Request a partial private key from the KGC to generate the private key for client Alice.

2) Sign the encrypted mail M^* (or M is encryption not required) along with the fields: From, To, Subject, and Timestamp, to produce the signature S using client Alice's private key.

3) Add additional headers to the mail messages as given below:

   - S in base64-encoding in SMTP header X-IBCLSES -Signature;
   - The exact id_stringAlice in SMTP header X-IBCLSES-Sender-IDString (if not already added);
   - The time stamp used in the signature in SMTP header X-IBCLSES-Timestamp.

The Email client sends Alice's mail using SMTP.

Figure 4 shows the steps involved in receiving the secure email. The detailed steps for signature verification and decryption is given below.

Signature Verification.

1) Download the secure mail from the mail server to the Email client of Bob using the IMAP/POP3 protocols.

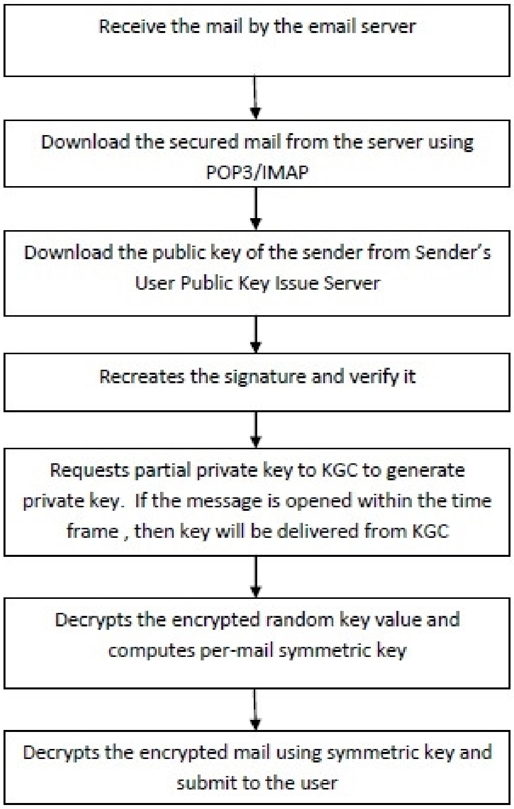

Receive the mail by the email server
Download the secured mail from the server using POP3/IMAP
Download the public key of the sender from Sender's User Public Key Issue Server
Recreates the signature and verify it
Requests partial private key to KGC to generate private key. If the message is opened within the time frame , then key will be delivered from KGC
Decrypts the encrypted random key value and computes per-mail symmetric key
Decrypts the encrypted mail using symmetric key and submit to the user

Figure 4: Receiving secure email

2) If the message header has the *X-IBCLSES-Signature* identifier, then the message is signed; determine the sender's email address, alice@a.com, and domain name, a.com, according to the email's From field and sender *id_string* email header *X-IBCLSES-Sender-IDString*.
3) Obtain the address of the User public key issue server for the domain a.com, by DNS TXT record lookup. Collect the public key of Alice P_{Alice}, from the user key issue server for Domain b.com using secure web services. Then, check that $\hat{e}(X_{Alice}, Q_m) = \hat{e}(Y_{Alice}, P)$ to authenticate the public key of client Alice.
4) Recreate the message M (or M^*) that was signed, using the declared From, To, and Subject fields, the email body, and the timestamp declared in *X-IBCLSES-Timestamp*.
5) Verify the signature.

Decryption.

1) If the message header has the *X-IBCLSES-Encryption-Key* identifier, then the message is in encrypted form.
2) Request a partial private key from the KGC to generate the private key for client Bob by passing the *id_string*. The system will not provide the partial key if the expiry date mentioned in the *id_string* is over.
3) Decrypt t^* using client Bob's private key
$$t = DS_{Bob}(t*).$$
4) Compute $K_{BobAlice} = tx_{Bob}X_{Alice}$.
5) Compute the per-mail symmetric key $K_{Bob} = H_2(Q_{Bob}||Q_{Alice}||K_{BobAlice})$. (i.e., $K_{Bob} = K_{Alice}$).
6) Decrypt the encrypted mail M^* using the symmetric key K as $M = D_K(M^*)$.

4 Implementation of the Proposed Secure Email System

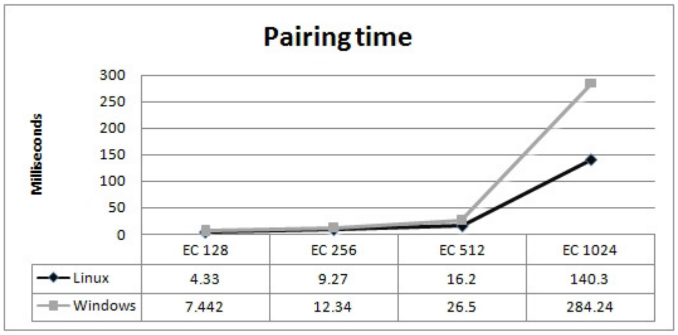
The prototype system was developed using the C++ programming language. To implement the IBC protocol, there is a need for a cryptographic library thatcan provide both Elliptic Curve Cryptography (ECC) and bilinear pairing functions. From the available literature, we selected the Miracl library. Miracl is an open source C/C++ Multiprecision Integer and Rational Arithmetic Cryptographic library. All of the basic encryption functions, such as setup, extract, encrypt and decrypt functions,were developed using the C++ language.

Security services for the email client were implemented as an extension to the Mozilla Thunderbird email client using JavaScript and the C++ library.

5 Security Discussion

The proposed system is secure against the standard security model because it is based on the Elliptic Curve Discrete Logarithm and Collision Resistant hash function standard cryptographic primitives. Other security properties provided by our solution follows.

1) End-to-end user authentication: Because the proposed mailing security system uses the CLPKC with the binding technique, both sender and receiver will authenticate each other using a pairing operation.
2) Key agreement between sender and receiver without interaction: The sender and receiver of the proposed system compute the shared secret key using their own secret values, the other party's public key and a randomly generated number, that is encrypted by the receiver's private key without any interaction between sender and receiver. Therefore, the proposed system is secure against the man-in-the-middle type attack.
3) Confidentiality of the message: The mail content is encrypted by a symmetric crypto system(such as AES), which guarantees the confidentiality of the message. The symmetric key can only be decrypted by the receiver.



	EC 128	EC 256	EC 512	EC 1024
Linux	4.33	9.27	16.2	140.3
Windows	7.442	12.34	26.5	284.24

Figure 5: Pairing time

4) Message integrity and non-repudiation: Because the sender signs the email message using his/her private key, the integrity and non-repudiation feature of the message can be verified, and the sender cannot disown the ownership of the email message.

5) Forward and backward secrecy: In the proposed system, each message session key during the message transmission is unique because both the sender and receiver use the random number *t*, which is generated by the sender and encrypted by the public key of the receiver in each session. The random number has the feature of Perfect Forward and Backward Secrecy, which is always fresh and unrelated to any past or future sessions.

6 Benchmark and Conclusion

The most costly procedure of the proposed system is the pairing operation. Benchmarking tests were performed on the operation, and the mean of 500 iterations was taken. The test only counts the time for a pairing, while the random point generation part is not considered. We conducted the test on an Intel Core i5-2400 CPU @3.10 GHz with 4GB RAM 1066MHz under the Windows 8 32 bit operating system and Oracle Linux 5.5 64-bit. For ECC 512 bits, we obtained an average speed of 16.2 milliseconds in Linux and 26.5 milliseconds in Windows, and for 1024 bits, we obtained 140.3 milliseconds in Linux and 284.24 milliseconds in Windows. Detailed performance analysis is given in Figure 5. From the test results, it is clear that we get better performance in 64 bit Linux system than 32-bit Windows and the proposed open standard system is very efficient and can be used in secure messaging as an alternative to the certificate based conventional PKI system.

This paper proposed an end-to-end secure mailing system based on certificate-less public key Cryptography, with DNS as the mechanism to publish a user's public key server address. The sender and receiver are able to compute the same secure secret symmetric key without any message exchange between them. This avoids a man-in-the-middle attack to obtain details of encryption/decryption key and hence the contents of the email.

Additionally, the proposed mail system is based on Elliptic Curve Cryptography, which is very efficient compared to conventional RSA based email systems and is also free from the heavy burden of certificate management of PKI. It avoids the key escrow problem of IBC based email systems by incorporating a partial private key generation system. The usability of IBE based systems are much better than PKI or PGP based systems. Moreover, the proposed system is based on standard cryptographic primitives, which makes it secure against the standard security model.

References

[1] B. Adida, D. Chau, S. Hohenberger, and R. L. Rivest, "Lightweight email signatures (extended abstract)," in *Proceedings of the 5th international conference on Security and Cryptography for Networks (SCN'06)*, pp. 288–302, 2006.

[2] D. Boneh and M. Franklin, "Identity-based encryption from the weil pairing," in *Advances in Cryptology (CRYPTO'01)*, LNCS 2139, Springer Verlag, pp. 213–229, 2001.

[3] J. Callas, L. Donnerhacke, H. Finney, and R. Thayer, *Open PGP Message Format*, Technical Report RFC 2440, Nov. 1998.

[4] S. Chatterjee and P. Sarkar, *Identity-based Encryption*, Springer Science+Business Media, LLC, 2011.

[5] L. Chen, K. Harrison, N. Smart, and D. Soldera, "Applications of multiple trust authorities in pairing based cryptosystems," in *International Conference on Infrastructure Security (InfraSec'02)*, LNCS 2437, Springer, pp. 260–275, 2002.

[6] T. Chen and S. Ma, "A secure email encryption proxy based on identity-based cryptography," in *International Conference on MultiMedia and Information Technology (MMIT'08)*, pp. 284–286, 2008.

[7] S. S. M. Chow, "Removing escrow from identity-based encryption," in *Public Key Cryptography (PKC'09)*, LNCS 5443, Springer, pp. 256–276, 2009.

[8] M. Cooper, *Internet X.509 Public Key Infrastructure (Latest Draft)*, IETF Internet Drafts, Jan. 2005.

[9] M. Crispin, *Internet Mail Access Protocol (ver. 4)*, Technical Report RFC 1730, Dec. 1994.

[10] D. Crocker, T. Hansen, and M. Kucherawy, *Domainkeys Identified Mail (DKIM) Signatures*, Technical Report RFC 6376, Sep. 2011.

[11] D. Eastlake, *Domain Name System Security Extensions*, Technical Report RFC 2535, Mar. 1999.

[12] B. A. Forouzan, *Cryptography and Network Security*, India: Tata McGraw-Hill Publishing Company Limited, 2007.

[13] Fortinet, *Forti Mail Identity Based Encryption*, Jan. 2014. (http://www. fortinet. com)

[14] M. Franklin and D. Boneh, "Identity based encryption from the weil pairing," *Journal of Computing*, vol. 32, no. 3, pp. 586–615, 2003.

[15] R. Gangishetti, M. C. Gorantla, M. Das, and A. Saxena, "Threshold key issuing in identity-based cryptosystems," *Computer Standards & Interfaces*, vol. 29, no. 2, pp. 260–264, 2007.

[16] R. Gangishetti, M. C. Gorantla, M. L. Das, A. Saxena, and V. P. Gulati, "An efficient secure key issuing protocol in id-based cryptosystems," in *Proceedings of the IEEE International Conference on Information Technology: Coding and Computing (ITCC 2005)*, pp. 674–678, Las Vegas, USA, Apr. 2005.

[17] C. Gentry, "Certificate-based encryption and the certificate revocation problem," in *Advances in Cryptology (EUROCRYPT'03)*, LNCS 2656, Springer, pp. 272–293, 2003.

[18] E. Gerck, *Secure Email Technologies X.509/PKI, PGP, IBE and Zmail: A Usability and Security Comparison*, ICFAI University Press, pp. 171–196, 2007.

[19] C. Gu and Y. Zhu, "New efficient searchable encryption schemes from bilinear pairings," *International Journal of Network Security*, vol. 10, no. 1, pp. 25–31, 2010.

[20] M. Hassouna, N. Mohamed, B. Barry, and E. Bashier, "An end-to-end secure mail system based on certificateless cryptography in the standard security model," *International Journal of Computer Science Issues*, vol. 10, no. 2, pp. 264–271, 2013.

[21] P. Hoffman, *SMTP Service Extension for Secure SMTP over Transport Layer Security*, Technical Report RFC 3207, Feb. 2002.

[22] M. P. Hoyle and C. J. Mitchell, "On solutions to the key escrow problem," in *State of the Art in Applied Cryptography*, LNCS 1528, Springer, pp. 277–306, 1998.

[23] HP, *HP Security Voltage*, Jan. 7, 2014. (http:// voltage. com)

[24] S. Kwon and S. H. Lee, "Identity-based key issuing without secure channel in a broad area," in *Information Security Applications*, LNCS 4298, Springer, pp. 30–44, 2007.

[25] P. Mockapetris, *Domain Names - Concepts and Facilities*, Technical Report RFC 1034, Nov. 1987.

[26] P. Mockapetris, *Domain Names - Implementation and Specification. IETF - Network Working Group, The Internet Society*, Technical Report RFC 1035, Nov. 1987.

[27] B. Ramsdell, *S/MIME Version 3 Message Specification*, Technical Report RFC 2633, June 1999.

[28] RSA Laboratories, "What is public-key cryptography," Jan. 20, 2014. (http://www. rsa. com/ rsalabs/)

[29] A. R. Sattam and P. Kenneth, "Certificateless public key cryptography a full version," in *Asiacrypt'03*, LNCS 2894, Springer, pp. 452–473, 2003.

[30] A. Shamir, "Identity based cryptosystems and signature schemes," *Computer Science*, vol. 196, pp. 47–53, 1984.

[31] D. Sharma and D. Jinwala, "Key generation protocol in IBC," *International Journal of Network Security*, vol. 15, no. 5, pp. 341–349, 2013.

[32] T. S. Sobh and M. I. Amer, "PGP modification for securing digital envelope mail using COM+ and web services," *International Journal of Network Security*, vol. 13, no. 2, pp. 79–91, 2011.

[33] J. Tian, L. Li, and X. Yang, "Fingerprint-based identity authentication and digital media protection in network environment," *Journal of Computer Science and Technology*, vol. 21, no. 5, pp. 861–870, 2006.

[34] J. Wang, X. Bai, J. Yu, and D. Li, "Protecting against key escrow and key exposure in identity-based cryptosystem," in *Theory and Applications of Models of Computation (TAMC'07)*, LNCS 4484, Springer, pp. 148–158, 2007.

Lecture Notes in Computer Science Volume 4484, 2007, pp 148-158

[35] Z. Wu, J. Tian, L. Li, C. P. Jiang, and X. Yang, "A secure email system based on fingerprint authentication," in *Scheme Intelligence and Security Informatics*, pp. 250–253, 2007.

[36] P. Zimmerman, *Pretty Good Privacy*, Jan. 20, 2014. (http://www. pgp. com)

Mr. Suresh Kumar Balakrishnan, currently working as an Engineer in Computer Division at Sree Chitra Tirunal Institute for Medical Sciences & Technology, Thiruvananthapuram, India is an M.Tech holder in Software Engineering from Cochin University of Science and Technology, Kochi, India. He has well-versed experience in implementing Computer Infrastructure, Network Security, Systems Management, Database Management and Hospital Information System. He has more than 17 years of experience in Computer Engineering fields. He is pursuing Ph.D in Computer Science at Indira Gandhi National Open University, New Delhi, India.

Dr. V. P. Jagathy Raj, currently working as Professor in Operations and Systems Management at School of Management Studies, Cochin University of Science and Technology, Kochi. He is equally well-versed in both Engineering and Management related areas. He has more than 24 years of teaching and research experience in Engineering and Management areas. He has more than 160 research publications in National and International Journals and Conference Proceedings to his credit in both Engineering and Management related areas. He has also presented number of research papers in National and International conferences.

Speeding up Pairing Computation Using Non-adjacent Form and ELM Method

Chao-Liang Liu[1], Gwoboa Horng[2], and Du-Shiau Tsai[3]
(Corresponding author: Chao-Liang Liu)

Department of Applied Informatics and Multimedia, Asia University[1]
500, Lioufeng Rd., Wufeng, Taichung 41354, Taiwan R.O.C.
Department of Computer Science and Engineering, National Chung-Hsing University[2]
250 Kuo Kuang Rd., Taichung, Taiwan R.O.C.
Department of Information and Networking Technology, Hsiuping University of Science and Technology[3]
No.11 Gongye Rd, Dali Dist., Taichung City 412-80,Taiwan, R.O.C
(Email: jliu@asia.edu.tw)

(Received Oct. 3, 2013; revised and accepted Jan. 15 & Mar. 26, 2014)

Abstract

The bilinear pairings such as Weil pairing and Tate pairing on elliptic curves have recently found many applications in cryptography. The first efficient algorithm for computing pairing was originally proposed by Miller and much subsequent research has been directed at many different aspects in order to improve efficiency. In 2003, Eisenträger, Lauter and Montgomery proposed a new point-double-addition method to speed up elliptic curve arithmetic computation and obtained a 7.8% performance improvement of the Miller algorithm of a general elliptic curve. In 2006, Blake et al. proposed a new concept based on the conjugate of a line to reduce the total number of lines in the Miller algorithm. In this paper we propose an enhancement of Eisenträger et al.'s algorithm for computing pairings. Our enhancement can further speed up the pairing computation by 5.9%.

Keywords: Elliptic curve cryptosystem, pairing-based cryptosystem, pairing computation

1 Introduction

Elliptic curve cryptograph, introduced by Miller [21, 22] and Koblitz [13] independently around 1985, provides the same level of security as the conventional public-key cryptography but with shorter keys. Numerous research efforts have been devoted to elliptic curve cryptography and a lot of cryptosystems have been proposed. By using Weil pairing, Menezes, Okamoto and Vanstone found some weak curves which contain cyclic groups that can be transformed into a finite field with small extension degree (MOV degree) [19]. Frey and Ruck extended their attack and found more weak curves with the Tate pairing [10]. Basically, the Weil/Tate pairing is a mapping with non-degenerate and bilinear properties, which will map a special pair of points on an elliptic curve to a certain multiplicative subgroup of a finite field. In recent years, bilinear pairings especially, Weil/Tate pairings, have found positive applications in cryptography. Indeed, many cryptographic applications based on pairings have been proposed, such as identity-based encryption systems [4], digital signatures [5, 6, 25, 26], signcryption [16, 24] and key agreement [12, 29]. As a result, the application of pairings plays an important role in modern cryptography. Therefore, efficiently implementation of pairing computation is an important issue due to being the most costly operation in these cryptosystems. The first efficient algorithm for computing pairing was proposed by Miller [21, 22]. The main idea of the Miller algorithm is to use lines to integrate the divisors, which the algorithm has processed (see Section 2, for details). A lot of research has been aimed in many different directions in order to improve efficiency [1, 2, 3, 7, 8, 10, 15, 17, 23, 27, 28, 33]. The research of Barreto, Kim, Lynn and Scott [1], and Galbraith, Harrison and Soldera [10] focuses particularly on the Tate pairing over some special curves. The research in [3, 8] can improve the performance of Weil/Tate pairing computation in general elliptic curves. We will continue in this direction.

It is well known that point subtraction and point addition on an elliptic curve have the same cost. Non-adjacent form (NAF for short) has been widely used for the scalar multiplication of nP for some point P on an elliptic curve [11]. Through this property, the efficiency of pairing computation can also be improved. For example, Eisenträger, Lauter and Montgomery gave a new point-addition/subtraction method (ELM method for short) to speed up scalar multiplication and pairing computation [8]. The majority of research in [8] literacy has focused on the double-addition/subtraction step when the bit of NAF representation of n is 1/-1. It is noticeable that the number of double step is twice the number of double-addition/subtraction step on average. With a parabola substitution, they get a 7.8% performance improvement of the Miller algorithm for a general elliptic curve (see Section 2, for details).

In 2006, Blake et al. proposed a new concept based on the conjugate of a line to reduce the total number of lines in the Miller algorithm [3]. Three different algorithms are proposed for three cases namely, BMX-1, BMX-2 and BMX-3. The first algorithm, $\log_2^n$ field multiplications are eliminated when there are relatively more zero bits (or average cases) of the binary representation of integer n. The second case is when there are relatively more one bits and $2H(n)$ field multiplications are removed where $H(n)$ is the number of bit 1. The third case saves $\log_3^n$ field multiplications when the characteristic of the field is three. Some successive works further improving Blake et al.'s algorithms [14, 18, 31, 32].

In this paper, we propose an algorithm to eliminate one more field multiplication in a double step, which the ELM method can not apply. Our new reduction method can reduce the number of lines, and hence improve the efficiency of pairing computation even further. The result can speed the computation of the Weil and Tate pairing by up to 5.9%, that is, combined with the ELM method, we can obtain a 13.3% performance improvement.

The rest of the paper is organized as follows. We briefly review some mathematical preliminaries, the Miller algorithm, the ELM method and Blake et al.'s formulae in Section 2. In Section 3 we describe our proposed algorithm. Its analysis is given in Section 4. Finally, some concluding remarks are given in Section 5.

2 Background

2.1 Weil/Tate Pairing and Miller Algorithm

Let E be an elliptic curve over a finite field F_q where q is a power of a prime p. We can express E as the Weierstrass equation

$$E : y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6,$$

where a_1, a_2, a_3, a_4, a_6 are all in F_q. If $\gcd(q, 6) = 1$, a nonsingular elliptic curve over the field F_q is given by an equation of the form

$$E_s : y^2 = x^3 + ax + b$$

with $a, b \in F_q$ and $4a^3 + 27b^2 \neq 0$. Let $E(F_q)$ denotes the set of points $(x, y) \in F_q^2$, satisfying E_s together with the point at infinity denoted as ∞. Then $E(F_q)$ together with point addition has a structure of an abelian group which is denoted as E. Explicit formulas for computing the coordinates of a point $R = P+Q$ from the coordinates of P and Q are well known [24, 26]. We give the formulae relative to $R = P + Q$ when $P \neq \pm Q$. That is, $R = (\lambda^2 - x_1 - x_2, \lambda x_1 - \lambda x_3 - y_1)$, where $P = (x_1, y_1)$, $Q = (x_2, y_2)$ and $\lambda = \frac{(y_2-y_1)}{(x_2-x_1)}$.

A divisor D is a formal sum of symbols from the set $\{(P)\colon\ P \in E\}$ with integer coefficients. That is $D = \sum_{P\in K} n_P(P)$. The set of all divisors, denoted by $Div(E)$, is a free abelian group generated by E. Define the degree of a divisor D, $\deg(D)$, to be $deg(D) = \sum_{P\in E} n_P$. We can define an important subgroup of $Div(E)$, denoted as $Div^0(E) = \{D \in Div(E)\colon\ deg(D) = 0\}$. The divisor of a nonzero rational function f is $div(f) = \sum_{P\in E} ord_P(f)(P)$, where $ord_P(f)$ is the order of f at P. It is well known that $div(f) \in Div^0(E)$ is called a principle divisor. If there exists a nonzero rational function f such that $D_1 = D_2 + div(f)$ then D_1 and D_2 are said to be equivalent, denoted as $D_1 \sim D_2$. The support of a divisor D is the set of points with nonzero coefficients, that is, $supp(D) = \{P \in E\colon\ n_P \neq 0\}$. If $div(f)$ and D have disjoint support, then we can evaluate $f(D) = \prod_{P\in E} f(P)^{n_P}$.

Let n be an integer relatively prime to q and $P, Q \in E[n]$, where $E[n]$ is the n-torsion subgroup of E. Then there exist divisors D_P, D_Q such that $D_P \sim (P) - (\infty)$ and $D_Q \sim (Q) - (\infty)$. Further, there exist functions f_P, f_Q such that $div(f_P) = nD_P$ and $div(f_Q) = nD_Q$. If D_P and D_Q have disjoint supports, then the Weil pairing is $e_n(P, Q) = \frac{f_P(D_Q)}{f_Q(D_P)}$. And the Tate pairing of order n is the map $\tau_n\colon\ E(F_q)[n]E(F_{q^k})/nE(F_{q^k}) \rightarrow F_{q^k}$, with $\tau_n(P, Q) = f_n(D_Q)^{(q^k-1)/n}$, where $div(f_n) = n(P) - n(\infty)$. Hence, computing the Weil/Tate pairing can be reduced to the evaluation of $f_P(S)$, where S is in the support of D_Q.

We briefly describe the main idea of the Miller algorithm as follows: Let $D_P = (P+R)-(R)$ with an auxiliary point R and $D_P^j = j(P+R)-j(R)-(jP)+(\infty)$, and then there is a rational function f_j such that $div(f_j) = D_P^j$, for each integer j, in particular, $f_n = f_P$. Hence

$$\begin{aligned}
div(f_{j+k}) &= (j+k)(P+R) - (j+k)(R) \\
&\qquad -((j+k)P) + (\infty) \\
&= [(j(P+R) - j(R) - (jP) + (\infty)] \\
&\qquad +[k(P+R) - k(R) - (kP) + (\infty)] \\
&\qquad +(jP) + (kP) - ((j+k)P) - (\infty) \\
&= div(f_j) + div(f_k) \\
&\qquad +[(jP) + (kP) + (-(j+k)P - 3(\infty)] \\
&\qquad -[((j+k)P) + (-(j+k)P) - 2(\infty)] \\
&= div(f_j) + div(f_k) + div(L_{jP,kP}) \\
&\qquad -div(L_{(j+k)P}),
\end{aligned}$$

where $L_{jP,kP}$ be a line through the points jP, kP and $-(j+k)P$. $L_{(j+k)P}$ be a vertical line through the points $(j+k)P$ and $-(j+k)P$. Then, $f_{j+k} = f_j f_k(\frac{L_{jP,kP}}{L_{(j+k)P}}$. As a result, we can obtain f_{j+k} from f_j and f_k with some "glue": the appropriate lines, $L_{jP,kP}$ and $L_{(j+k)P}$.

We can compute $f_n(S)$ recursively with initial values $f_0 = 1$ and $f_1 = \frac{L_{P+R}}{L_{R,R}}$. We describe the following algorithm, which is similar to the algorithm proposed in [1, 3]. Note that we can perform the Miller algorithm to compute Tate pairing by changing the initial value $f_1 = 1$, see [1] for details.

Algorithm 1 Miller algorithm

```
INPUT: Elliptic curve $E$, integer $n = \sum_{i=0}^{t} b_i 2^i$ with $b_i \in \{0,1\}$ and $b_t = 1$, and points $P$, $S \in E$ where $P$ has order $n$.
OUTPUT: $f = f_n(S)$.
$f \leftarrow f_1$; $Z \leftarrow P$;
for $j \leftarrow t-1$ down to 0 do
$f \leftarrow f^2 \frac{L_{Z,Z}(S)}{L_{2Z}(S)}$; $Z \leftarrow 2Z$;
for $j \leftarrow t-1$ down to 0 do
$f \leftarrow f^2 \frac{L_{Z,Z}(S)}{L_{2Z}(S)}$; $Z \leftarrow 2Z$;
if $b_j = 1$ then
$f \leftarrow f_1 f \frac{L_{Z,P}(S)}{L_{Z+P}(S)}$; $Z \leftarrow Z + P$;
return $f$;
End
```

2.2 ELM Method

In Algorithm 1, the cost of pairing computation consists of two main parts. One is a scalar multiplication of nP. The other is an exponential computation and multiplication with the glue. To decrease the cost of point's double-addition/subtraction of scalar multiplication, Eisenträger et al. eliminate two field multiplications through a new method to compute $2P + Q$ by computing $P + Q$ and $2P + Q = (P + Q) + P$. Note that, we do not care about the intermediate result $P + Q$. The explicit formulae are described as follows:

$$\begin{aligned}
\lambda_1 &= \frac{y_2 - y_1}{x_2 - x_1}, x_3 = \lambda_1^2 - x_1 - x_2 \\
\lambda_2 &= -\lambda_1 - \frac{2y_1}{x_3 - x_1}, x_4 = \lambda_2^2 - x_1 - x_3 \\
y_4 &= (x_1 - x_4)\lambda_2 - y_1,
\end{aligned}$$

where $P = (x_1, y_1)$, $Q = (x_2, y_2)$, $x_1 \neq x_2$, $P + Q = (x_3, y_3)$, and $2P + Q = (x_4, y_4)$ on an elliptic curve E_S. Moreover, λ_1 is the slope of $L_{P,Q}$ and λ_2 is the slope of $L_{P+Q,P}$.

To apply this point double-addition/subtraction method for the Miller algorithm, they construct a parabola to glue the Miller's divisors, whenever the corresponding bit is one, see [8] for detail.

Suppose we use the binary method in [11] to form nP, where n has t bits. There are $2t/3$ doubles and $t/3$ double-additions/subtractions. By way of estimating a division as 5.18 field multiplications, they compute the average cost of the standard algorithm as

$$\frac{(16.18 \times 2t/3) + (31.36 \times t/3)}{t} = 21.24$$

field multiplications per bit, and the average cost of their new method is

$$\frac{(16.18 \times 2t/3) + (26.36 \times t/3)}{t} = 19.57$$

field multiplications per bit. The performance improvement ratio for their new method is 7.8%. It is noticeable that these estimations are based on the computation of Tate pairing for which $f_n(Q_1)$ and $f_n(Q_2)$ are computed at the same time. Please see [8] or Section 4 for details.

We also need a divisor subtraction formula to use the NAF method to form $f_n(S)$ with respect to the Miller algorithm. Therefore, they proposed the first divisor subtraction formula:

$$\begin{aligned}
div(f_{j-k}) &= (j-k)(P+R) - (j-k)(R) \\
&\qquad -((j-k)P) + (\infty) \\
&= [(j(P+R) - j(R) - (jP) + (\infty))] \\
&\qquad -[k(P+R) - k(R) - (kP) + (\infty)] \\
&\qquad +(jP) - (kP) - ((j-k)P) + (\infty) \\
&= div(f_j) - div(f_k) \\
&\quad +[(jP) + (-jP) - 2(\infty)] \\
&\quad -[(-jP) + (kP) + ((j-k)P) - 3(\infty)] \\
&= div(f_j) - div(f_k) + div(L_{jP}) \\
&\qquad -div(L_{-jP,kP}),
\end{aligned}$$

Therefore,

$$f_{j-k} = \frac{f_j}{f_k} \cdot \frac{L_{jP}}{L_{-jP,kP}}. \quad (1)$$

2.3 Blake et al's Lemmas

From the analysis in [8], we know that if we can reduce one line then at least one field multiplication is saved in the Miller algorithm. For this reason, Blake et al. proposed three algorithms to reduce the number of lines. The first algorithm is suitable for every case. The second algorithm can work well if the Hamming weight of n is high. The third algorithm is proposed for fields of characteristic 3. These algorithms are based on the following two lemmas which were proved in [3].

Lemma 1. *If the line $L(x, y)$ intersects with E at points $P = (a, b)$, $Q = (c, d)$ and $-(P + Q) = (\alpha, \beta)$, then $L(x, y)\bar{L}(x, y) = -(x - a)(x - c)(x - \alpha)$, where $\bar{L}(x, y)$ is the conjugate of L with $L(R) = \bar{L}(-R)$ for $R \in E$.*

Lemma 2. *Let $Q \in E[n]$, $S \in E$ and $S \neq Q$, $2Q$, $\cdots$, nQ, then*

1) $\frac{L_{Q,Q}(S)}{L_Q^2(S)L_{2Q}(S)} = \frac{-1}{L_{Q,Q}(-S)}$.

2) For all integer k, we have $\frac{L_{(k+1)Q,kQ}(S)}{L_{(k+1)Q}(S)L_{(2k+1)Q}(S)} = -\frac{L_{kQ}(S)}{L_{(k+1)Q,kQ}(-S)}$.

3) $\frac{L_{Q,Q}(S)L_{2Q,Q}(S)}{L_{2Q}(S)L_{3Q}(S)} = -\frac{L_{Q,Q}(S)L_Q(S)}{L_{2Q,Q}(-S)}$.

They also remark that [3]:

1) Since $div(f) = fiv(cf)$ for any nonzero constant $c \in K$, the sign does not affect the pairing computation and therefore, minus signs will be omitted in the use of the above lemma.

2) The point $P \in E[n]$ will be fixed and Q is taken to be some multiple of P. In order to satisfy the condition of the lemma, it is sufficient to let $S \neq P$, $2P$, $\cdots$, nP. This is also the requirement of the original Miller algorithm.

3 A New Method for Computing Pairings

In Section 2, the ELM method concentrates on the double-addition/subtraction step in the point's scalar multiplication, however the number of double steps is twice the number of double-addition/subtraction steps. Therefore, we suggest a new algorithm to reduce one field multiplication when the corresponding bit of n is 0 in the Miller algorithm. Before expressing this new algorithm, we briefly describe the limitations of their method to compute the pairing.

To compute $2P+Q$ via $P+Q$, where P, Q, $P+Q$ and $2P+Q$ on an elliptic curve E_s but $P \neq \pm Q$. Then the capability and the limitations of their method are:

1) We have the x-coordinators of the points P, Q, $P+Q$ and $2P+Q$. But we do not have the x-coordinator of the point $2P$.

2) We have the y-coordinators of the points P, Q and $2P+Q$. But we do not have the y-coordinators of $P+Q$ and $2P$.

3) We have the slopes for the lines $L_{P,Q}$ and $L_{P+Q,P}$.

4) We can construct the linear functions $L_{P,Q}$, $L_{P+Q,P}$, L_P, L_Q, L_{P+Q}, and L_{2P+Q}. But we cannot construct $L_{2P,Q}$ and L_{2P}.

The detail description of the divisor subtraction formula (Equation (1)) with their point double-subtraction method in the Miller algorithm is

$$f \leftarrow \frac{f^2}{f_1} \cdot \frac{L_Z(S)L_{Z-P,Z}(S)}{L_{-Z,P}(S)L_{2Z-P}(S)}; Z \leftarrow 2Z - P;$$

when the bit of n is -1. Although the linear functions L_Z, $L_{Z-P,Z}$, L_{2Z-P}, and $L_{-Z,P}$ can be constructed, no parabola was revealed in [8]. It is well-known that there are no consecutive nonzero bits in the NAF representation such that there is always a zero bit before -1. In [3], they have examined the reduction formulae of bit 0 and bit 1, however, there were few studies of the relationship between bit 0 and bit -1. Therefore, we have to extend Lemma 2.2 to Lemma 3.1 in order to establish a reduction formula for this case.

Lemma 3. *Let $Q \in E[n]$, $S \in E$ and $S \neq Q, 2Q, \cdots, nQ$, then $\frac{L_(k-1)Q,kQ(S)}{L_{kQ}(S)L_{(2k-1)Q}(S)} = \frac{L_{(k-1)Q}(S)}{L_{(k-1)Q,kQ}(-S)}$.*

Proof. For a point $S \in E$, we write $S = (x_S, y_S)$, that is, x_S is the x-coordinate of S and y_S is the y-coordinate of S. By Lemma 2, we have:

$$\begin{aligned}
& \frac{L_(k-1)Q, kQ(S)}{L_{kQ}(S)L_{(2k-1)Q}(S)} \\
= \; & \frac{L_(k-1)Q, kQ(S)\bar{L}_{(k-1)Q,kQ}(S)}{L_{kQ}(S)L_{(2k-1)Q}(S)\bar{L}_{(k-1)Q,kQ}(S)} \\
= \; & \frac{-(x_S - x_{(k-1)Q})(x_S - x_{kQ})(x_S - x_{(2k-1)Q}}{(x_S - x_{kQ})(x_S - x_{(2k-1)Q})L_{(k-1)Q,kQ}(-S)} \\
= \; & \frac{L_{(k-1)Q}(S)}{L_{(k-1)Q,kQ}(-S)}.
\end{aligned}$$

□

Consider the NAF representation of $n = \sum_{i=0}^{t} b_i 2^i$ with $b_i \in \{0, 1\}$, $b_t = 1$ and $b_{i+1} \cdot b_i = 0$ for $0 \leq i < t$. We give the detail descriptions of the following three reduction formulae by applying Lemma 2.2 and Lemma 3.1. These formulae play a key role in our algorithm. Suppose that the Miller algorithm is performed by an addition/subtraction chain and glues the divisors in the trace of the point addition $(Z \pm P) + Z$ after $Z \pm P$ in the three cases. Note that, $L_Q(S)$ and $L_{-Q}(S)$ have the same value for points Q and S on an elliptic curve, and we can omit the minus signs as remarked in Section 2.3.

1) Case (0, 0) performs:

$$f \leftarrow f^2 \frac{L_{Z,Z}(S)}{L_{2Z}(S)}; Z \leftarrow 2Z;$$

$$f \leftarrow f^2 \frac{L_{2Z,2Z}(S)}{L_{4Z}(S)}; Z \leftarrow 2Z;$$

Putting together, we have:

$$f \leftarrow (f^2 \frac{L_{Z,Z}(S)}{L_{2Z}(S)})^2 \frac{L_{2Z,2Z}(S)}{L_{4Z}(S)} = f^4 \frac{-L^2_{Z,Z}(S)}{L_{2Z,2Z}(-S)}$$

Omitting the minus sign, we have:

$$f \leftarrow f^4 \frac{L^2_{Z,Z}(S)}{L_{2Z,2Z}(-S)}; Z \leftarrow 4Z.$$

2) Case (0,1) performs:

$$f \leftarrow f^2 \frac{L_{Z,Z}(S)}{L_{2Z}(S)}; Z \leftarrow 2Z;$$

$$f \leftarrow f_1 f^2 \frac{L_{Z,P}(S)L_{Z+P,Z}(S)}{L_{Z+P}(S)L_{2Z+P}(S)}; Z \leftarrow 2Z + P;$$

Putting together, we have:

$$\begin{aligned}
f &\leftarrow (f^2\frac{L_{Z,Z}(S)}{L_{2Z}(S)})^2 \cdot f_1\frac{L_{2Z,P}(S)}{L_{2Z+P}(S)}\frac{L_{2Z+P,2Z}(S)}{L_{4Z+P}(S)} \\
&= f_1 f^4\frac{L_{Z,Z}^2(S)}{L_{2Z}^2(S)}\frac{L_{2Z,P}(S)}{L_{2Z+P}(S)} \\
&\quad \cdot\frac{L_{2Z+P,2Z}(S)L_{2Z+P,2Z}(-S)}{L_{4Z+P}(S)L_{2Z+P,2Z}(-S)} \\
&= f_1 f^4\frac{L_{Z,Z}^2(S)}{L_{2Z}^2(S)}\frac{L_{2Z,P}(S)}{L_{2Z+P}(S)} \\
&\quad \cdot[-\frac{L_{2Z+P}(S)L_{2Z}(S)L_{4Z+P}(S)}{L_{4Z+P}(S)L_{2Z+P,2Z}(-S)} \\
&= f_1 f^4\frac{-L_{Z,Z}^2(S)L_{2Z,P}(S)}{L_{2Z}(S)L_{2Z+P,2Z}(-S)}.
\end{aligned}$$

Omitting the minus sign, we have:

$$f \leftarrow f_1 f^4\frac{L_{Z,Z}^2(S)L_{2Z,P}(S)}{L_{2Z}L_{2Z+P,2Z}(-S)}; Z \leftarrow 4Z + P.$$

3) Case (0, -1) performs:

$$\begin{aligned}
f &\leftarrow f^2\frac{L_{Z,Z}(S)}{L_{2Z}(S)}; Z \leftarrow 2Z; \\
f &\leftarrow \frac{f^2}{f_1}\frac{L_Z(S)}{L_{-Z,P}(S)}\frac{L_{Z-P,Z}(S)}{L_{2Z-P}(S)}; Z \leftarrow 2Z - P;
\end{aligned}$$

Putting together, we have:

$$\begin{aligned}
f &\leftarrow (f^2\frac{L_{Z,Z}(S)}{L_{2Z}(S)})^2 \cdot \frac{1}{f_1}\frac{L_{2Z}(S)}{L_{2Z,P}(S)}\frac{L_{2Z-P,2Z}(S)}{L_{4Z-P}(S)} \\
&= \frac{f^4}{f_1}\frac{L_{Z,Z}^2(S)}{L_{2Z}^2(S)}\frac{L_{2Z}(S)}{L_{-2Z,P}(S)} \\
&\quad \cdot\frac{L_{2Z-P,2Z}(S)L_{2Z-P,2Z}(-S)}{L_{4Z-P}(S)L_{2Z-P,2Z}(-S)} \\
&= \frac{f^4}{f_1}\frac{L_{Z,Z}^2(S)}{L_{2Z}^2(S)L_{-2Z,P}(S)} \\
&\quad \cdot[-\frac{L_{2Z-P}(S)L_{2Z}(S)L_{4Z-P}(S)}{L_{4Z-P}(S)L_{2Z-P,2Z}(-S)} \\
&= \frac{f^4}{f_1}\frac{-L_{Z,Z}^2(S)L_{2Z-P}(S)}{L_{-2Z,P}(S)L_{2Z-P,2Z}(-S)}.
\end{aligned}$$

Omitting the minus sign, we have:

$$f \leftarrow \frac{f^4}{f_1}\frac{L_{Z,Z}^2(S)L_{2Z-P}(S)}{L_{-2Z,P}(S)L_{2Z-P,2Z}(-S)}; Z \leftarrow 4Z - P.$$

From these formulae, there is only one line $L_{Z,Z}$ (or $L_{2Z,2Z}$) which needs to be evaluated at point S whence the relative bit of n is zero. That is, we can eliminate one field multiplication when we glue the divisors in the three cases (0, 0), (0, 1) and (0, -1). These detail descriptions of the three cases also provide the correctness of an improved Miller algorithm which we will describe in Algorithm 2.

Algorithm 2 The improved Miller algorithm

```
INPUT: Elliptic curve $E$, integer $n = \sum_{i=0}^{t} b_i 2^i$ with $b_i \in \{0, 1\}$, $b_t = 1$, $b_{i+1} \cdot b_i = 0$ for $0 \leq i < t$, and points $P$, $S \in E$ where $P$ has order $n$.
OUTPUT: $f = f_n(S)$.
$f \leftarrow f_1$; $Z \leftarrow P$; $i \leftarrow t - 1$;
while $i > 0$ do
if $(b_i, b_{i-1}) = (0, 0)$ then
$f \leftarrow f^4 \frac{L_{Z,Z}^2(S)}{L_{2Z,2Z}(-S)}$; $Z \leftarrow 4Z$; $i \leftarrow i - 2$;    {Case 0}
if $(b_i, b_{i-1}) = (0, 1)$ then
$f \leftarrow f_1 f^4 \frac{L_{Z,Z}^2(S)L_{2Z,P}(S)}{L_{2Z}(S)L_{2Z+P,2Z}(-S)}$; $Z \leftarrow 4Z + P$; $i \leftarrow i - 2$;    {Case 1}
if $(b_i, b_{i-1}) = (0, -1)$ then
$f \leftarrow \frac{f^4}{f_1} \frac{L_{Z,Z}^2(S)L_{2Z-P}(S)}{L_{-2Z,P}(S)L_{2Z-P,2Z}(-S)}$; $Z \leftarrow 4Z - P$; $i \leftarrow i - 2$;    {Case 2}
if $(b_i, b_{i-1}) = (1, 0)$ then
$f \leftarrow f_1 f^2 \frac{L_{Z,P}(S)L_Z(S)}{L_{Z+P,Z}(-S)}$; $Z \leftarrow 2Z + P$; $i \leftarrow i - 1$;    {Case 3}
if $(b_i, b_{i-1}) = (-1, 0)$ then
$f \leftarrow \frac{f^2}{f_1} \frac{L_Z(S)L_{Z-P,Z}(S)}{L_{-Z,P}(S)L_{2Z-P}(S)}$; $Z \leftarrow 2Z - P$; $i \leftarrow i - 1$;    {Case 4}
end-while
if $i = 0$ then
if $b_i = 1$ then
$f \leftarrow f^2 L_{Z,Z}(S)$; $Z \leftarrow 2Z$;
if $b_i = 1$ then
$f \leftarrow f_1 f^2 \frac{L_{Z,P}(S)L_Z(S)}{L_{Z+P,Z}(-S)}$; $Z \leftarrow 2Z + P$;
if $b_i = -1$ then
$f \leftarrow \frac{f^2}{f_1} \frac{L_Z(S)L_{Z_P,Z}(S)}{L_{-Z,P}(S)}$; $Z \leftarrow 2Z - P$;
return $f$;
End
```

4 Analysis

In this section, detailed analysis of the improvement is given. Additionally, the estimation of the cost of the improvement is in accordance with the rules which were discussed in [8]. The basic concept of the improvement is that it tries to find the maximum number of the pattern (0, 0) and only processes the first bit of the pattern (1, 0) in Case 3 and the pattern (-1, 0) in Case 4. It is noticeable that the methods of Case 3 and Case 4 can be replaced with the parabola substitution method which was described in [8]. As a result, only one line has to be evaluated for each zero bit of n in our improvement.

As indicated in [1, 3, 8], in the actual implementation of pairing computation, the operations in the numerator and denominator in each step are separated and perform one division at the very end. In [8], they estimate the total cost of pairing computation with the following specifications:

1) The pairing evaluates a quotient of the form $\frac{f_n(Q_1)}{f_n(Q_2)}$ for two points Q_1, Q_2 on E, where n is a t bits integer which consists of $2t/3$ zero bits and $t/3$ nonzero bits.

2) The cost of each bit is counted as the total number of field multiplications, but the cost of all field additions/subtractions are omitted.

3) The cost of a division is estimated as 5.18 field multiplications.

4) The cost of the standard algorithm is 16.18 field multiplications for each zero bit and 31.36 field multiplications for each nonzero bit.

5) The cost of the ELM method is 16.18 field multiplications for each zero bit and 26.36 field multiplications for each nonzero bit.

For simplicity, our estimation follows the analysis in [8] which counts the cost in each case separately. Plus, only three different cases need to be analyzed between our improvement and the ELM method. These are the cases of the cost of bit 0 and the cost of bit ±1 in (0, 1) and (0, -1):

1) The cost of bit 0 which appears in the patterns (0, 0), (0, 1) and (0, -1): In these cases, we must perform $f \leftarrow f^2 L_{Z,Z}(S)$ and $Z \leftarrow 2Z$ for each bit 0.

   a. A point doubling operation costs 3 field multiplications and a division.

   b. Evaluating $L_{Z,Z}$ at points Q_1 and Q_2 costs 2 field multiplications.

   c. Multiplying 4 fractions f_{nu}, f_{de}, $L_{Z,Z}(Q_1)$, and $L_{Z,Z}(Q_2)$ costs 4 field multiplications. Where f_{nu} is the numerator of f and f_{de} is the denominator of f. That is, we must compute $\frac{f_{nu} \cdot f_{nu} \cdot L_{Z,Z}(Q_1)}{f_{de} \cdot f_{de} \cdot L_{Z,Z}(Q_2)}$ in the improvement whence the relative bit is 0.

   The total cost of this case is $3 + 5.18 + 2 + 4 = 14.18$ field multiplications.

2) The cost of bit 1 of (0, 1): In this case, we must perform $f \leftarrow f_1 f^4 \frac{L^2_{Z,Z}(S) L_{2Z,P}(S)}{L_{2Z}(S) L_{2Z+P,2Z}(-S)}$ and $Z \leftarrow 4Z + P$. Then they can be separated as:

   $$f \leftarrow [f^2 L^2_{Z,Z}(S)]^2 f_1 \frac{L_{2Z,P}(S)}{L_{2Z}(S) L_{2Z+P,2Z}(-S)}$$

   and $Z \leftarrow (2Z+P)+2Z$. The cost of the first component is estimated in A. We estimate the cost of the second component as follows:

   a. A point double-addition costs 3 field multiplications and 2 divisions.

   b. Evaluating $L_{2Z,P}$ at points Q_1 and Q_2 costs 2 field multiplications. Evaluating $L_{2Z+P,2Z}$ at points $-Q_1$ and $-Q_2$ costs 2 field multiplications.

   c. Multiplying 10 fractions costs 10 field multiplications.

   The total cost of this case is $3+10.36+4+10 = 27.36$ field multiplications.

3) The cost of bit -1 which appear in the pattern (0, -1): In this case, we must perform $f \leftarrow \frac{f^4}{f_1} \frac{L^2_{Z,Z}(S) L_{2Z-P}(S)}{L_{-2Z,P}(S) L_{2Z-P,2Z}(-S)}$ AND $Z \leftarrow 4Z - P$. Then they can be separated as:

   $$f \leftarrow [F^2 l_{z,z}(s)]^2 \frac{L_{2Z-P}(S)}{f_1 \cdot L_{-2Z,P}(S) L_{2Z-P,2Z}(-S)}$$

   and $Z \leftarrow (2Z-P)+2Z$. The cost of the first component is estimated in A. We estimate the cost of the second component as follows:

   a. A point double-subtraction costs 3 field multiplications and 2 divisions.

   b. Evaluating $L_{-2Z,P}$ at points Q_1 and Q_2 costs 2 field multiplications. Evaluating $L_{2Z-P,2Z}$ at points $-Q_1$ and $-Q_2$ costs another 2 field multiplications.

   c. Multiplying 10 fractions costs 10 field multiplications.

   The total cost of this case is $3 + 10.36 + 2 + 2 + 10 = 27.36$ field multiplications.

Before we compute the average cost of our refinement, we define two sets, ODD and $EVEN$, for the pattern w, which appears in the NAF representation of n, where $n = \sum_{i=0}^{t} b_i 2^i$ with $b_i \in \{0, 1\}$, $b_t = 1$ and $b_{i+1} \cdot b_i = 0$ for $0 \leq i < t$. That is, $ODD = \{w = b_{i+r+1}(0, 0, \cdots, 0)b_i$: r is odd, $b_{i+r+1} \cdot b_i \neq 0$, $0 \leq i < i+r+1 \leq t\}$ and $EVEN = \{w = b_{i+r+1}(0, 0, \cdots, 0)b_i$: r is even, $b_{i+r+1} \cdot b_i \neq 0$, $0 \leq i < i + r + 1 \leq t\}$.

Without lost of generality, assume $|ODD| = |EVEN|$. Accordingly, the total number of Case 1 and Case 2 is estimated as the same as the total number of Case 3 and Case 4 in Algorithm 2. That is, in half of all nonzero bits, each bit costs 27.36 field multiplications and each bit of the rest costs 26.36 field multiplications. Therefore, the average cost of our improvement is $\frac{14.18 \times 2t/3 + 27.36 \times t/6 + 26.36 \times t/6}{t} = 18.41$ field multiplications per bit. Compared to the standard algorithm, the improvement is $\frac{21.24 - 18.41}{21.24} = 13.3\%$. In other words, we enhance the ELM method to obtain a $\frac{19.57 - 18.41}{19.57} = 5.9\%$ improvement in performance.

5 Concluding Remarks

An improvement in the computation of the pairings has been given and the corresponding performance has been analyzed. It is noticeable that this algorithm can be more efficient if more lines belonging to the nonzero bits are reduced. We can achieve this purpose by recoding the NAF representation of n into many patterns, such as (0^r), $(0, 1, 0)$ and $(0, -1, 0)$. But this is getting half the result with twice the effort. Therefore, we propose a concise algorithm which focuses on performance improvement of the zero bits and gives a simplified performance analysis. As a result, the proposed algorithm gains an improvement of 5.9% in performance when compared to the ELM method.

Acknowledgments

This study was supported by the Asia University, under grant No. 98-NSC-02 and 101-ASIA-34. The authors gratefully acknowledge the anonymous reviewers for their valuable comments.

References

[1] P. Barreto, H. Kim, B. Lynn, and M. Scott, "Efficient algorithms for pairing-based cryptosystems," in *Advance in Cryptography (Crypt'02)*, LNCS 2442, pp. 354–368, Springer-Verlag, 2002.

[2] S. Basu, "A new parallel window-based implementation of the elliptic curve point multiplication in multi-core architectures," *International Journal of Network Security*, vol. 14, no. 2, pp. 101–108, 2012.

[3] I. Blake, V. Murty, and G. Xu, "Refinement of Miller's algorithm for computing the Weil/Tate pairing," *Journal of Algorithms*, vol. 58, pp. 134–149, 2006.

[4] D. Boneh, and M. Franklin, "Identity-base encryption from the Weil pairing," in *Advance in Cryptography (Crypto'01)*, LNCS 2139, pp. 213–239, Springer-Verlag, 2001.

[5] D. Boneh, B. Lynn, and H. Shacham, "Short signature from the weil pairing," in *Advance in Cryptography (Asiacrypt'01)*, LNCS 2248, pp. 514–532, 2Springer-Verlag, 001.

[6] D. Boneh, X. Boyen, and H. Shacham, "Short group signatures," in *Advance in Cryptography (Crypto'04)*, LNCS 3152, pp. 41–55, Springer-Verlag, 2004.

[7] Y. Ding, K. W. Wong, and Y. M. Wang, "Joint sparse form of window three for Koblitz curve," *International Journal of Network Security*, vol. 2, no. 2, pp. 126–130, 2006.

[8] K. Eisentr?ger, K. Lauter, and P. L. Montgomery, "Fast elliptic curve arithmetic and improved Weil pairing evaluation," in *Topics in Cryptology (CT-RSA'03)*, LNCS 2612, pp. 343–354, Springer-Verlag, 2003.

[9] G. Frey and H. Ruck, "A remark concerning m-divisibility and the discrete logarithm in the divisor class group of curves," *Mathematics of Computation*, vol. 62, pp. 865–874, 1994.

[10] S. Galbraith, K. Harrison, and D. Soldera, "Implementing the tate pairing," in *Algorithm Number Theory Symposium*, LNCS 2369, pp. 324–337, Springer-Verlag, 2002.

[11] IEEE Computer Society, *IEEE Standard Specifications for Public-key Cryptography*, IEEE Standard 1363–2000, 2000.

[12] A. Joux, "A one round protocol for tripartite Diffie-Hellman," in *Algorithmic Number Theory*, LNCS 1838, pp. 385–393, Springer-Verlag, 2000.

[13] N. Koblitz, "Elliptic curve cryptosystems," *Mathematics of Computation*, vol. 48, pp. 203–209, 1987.

[14] D. P. Le and C. L. Liu, "Refinements of Miller's algorithm over weierstrass curves revisited," *The Computer Journal*, vol. 54, no. 10, pp. 1582–1591, 2011.

[15] J. Lee, H. Kim, Y. Lee, S. M. Hong, and H. Yoon, "Parallelized scalar multiplication on elliptic curves defined over optimal extension field," *International Journal of Network Security*, vol. 4, no. 1, pp. 99–106, 2007.

[16] X. Li and K. Chen, "Identity based proxy-signcryption scheme from pairings," in *Proceedings of 2004 IEEE International Conference on Services Computing (IEEE-SCC'04)*, pp. 494–497, 2004.

[17] T. C. Lin, "Algorithms on elliptic curves over fields of characteristic two with non-adjacent forms," *International Journal of Network Security*, vol. 9, no. 2, pp. 117–120, 2009.

[18] C. Liu, G. Horng, and T. Chen. "Further refinement of pairing computation based on Miller's algorithm," *Applied Mathematics and Computation*, vol. 189, no. 1, pp. 395–409, 2007.

[19] A. Menezes, T. Okamoto, and S. A. Vanstone, "Reducing elliptic curve logarithms to logarithms in a finite field," *IEEE Transaction on Information Theory*, vol. 39, pp. 1639–1646, 1993.

[20] A. Menezes, *Elliptic Curve Cryptosystems*, Kluwer Academic Publishers, 1993.

[21] V. Miller, "Use of elliptic curves in cryptosystems," in *Advance in Cryptography (Crypto'85)*, LNCS 218, pp. 417–426, Springer-Verlag, 1986.

[22] V. Miller, *Short Programs for Functions on Curve*, Sept. 2002. (http://www.researchgate.net/profile/Victor_Miller/publication/2551688_Short_Programs_for_functions_on_Curves/links/ 0c96052e065ca0bdbf000000.pdf)

[23] S. Moon, "A binary redundant scalar point multiplication in secure elliptic curve cryptosystems," *International Journal of Network Security*, vol. 3, no. 2, pp. 132–137, 2006.

[24] D. Nalla and K. Reddy, "Signcryption scheme for identity-based cryptosystems," *Cryptology ePrint Archive*, Report, 2003/066, 2003.

[25] H. Sahu and B. K. Sharma, "An MSS based on the elliptic curve cryptosystem," *International Journal of Network Security*, vol.11, no. 2, pp. 118, 2010.

[26] R. Sakai, K. Ohgishi, and M. Kasahara, "Cryptosystems based on pairing," in *SCIS'00*, pp. 26–28, 2000.

[27] Z. J. Shi and H. Yun, "Software implementations of elliptic curve cryptography," *International Journal of Network Security*, vol. 7, no. 1, pp. 141–150, 2008.

[28] S. M. Shohdy, A. B. El-Sisi, and N. Ismail, "Hardware implementation of efficient modified karatsuba multiplier used in elliptic curves," *International Journal of Network Security*, vol. 11, no. 3, pp. 155–162, 2010.

[29] N. P. Smart, "An identity based authenticated key agreement protocol based on weil pairing," *Electronics Letters*, vol. 38, pp. 630–632, 2002.

[30] J. H. Silverman, *The Arithmetic of Elliptic Curves*, Graduate Texts in Mathematics, vol. 106, Springer, New York, 1986.

[31] T. Wu, H. Du, M. Zhang, and R. Wang, "Improved algorithm of the tate pairing in characteristic three," in *The First International Symposium on Data, Privacy, and E-Commerce*, pp. 453–455, 2007.

[32] T. Wu, M. Zhang, X. Xu, and R. Wang, "Improved algorithm for tate pairing computation," in *Proceedings of The IEEE International Symposium on Electronic Commerce and Security (ISECS'08)*, pp. 41–45, 2008.

[33] D. Yong, Y. F. Hong, W. T. Wang, Y. Y. Zhou, and X. Y. Zhao, "Speeding Scalar Multiplication of Elliptic Curve over $GF(2^{mn})$," *International Journal of Network Security*, vol. 11, no. 2, pp. 70–77, 2010.

Chao-Liang Liu was born in Hsinchu, Taiwan, in 1966. He completed his B.Sc. degree in mathematics from National Cheng-Kung University, Taiwan, in 1989, and the Ph.D. degree in Institute of Computer Science from National Chung Hsing University, Taiwan, in 2007. He is currently an assistant professor at the Department of Applied Informatics and Multimedia, Asia University, Taiwan. His research interests include information security, cryptography, elliptic curve cryptosystem and pairing computation.

Gwoboa Horng received the B.S. degree in Electrical Engineering from National Taiwan University in 1981 and the M.S. and Ph.D. degrees from University of Southern California in 1987 and 1992 respectively, all in Computer Science. Since 1992, he has been on the faculty of the Department of Computer Science and Engineering at National Chung-Hsing University, Taichung, Taiwan, R.O.C. His current research interests include artificial intelligence, cryptography and information security.

Du-Shiau Tsai received the B.S. degree in Department of Computer Science and Information Management from Providence University, Taiwan, in 1996 and the M.S. degree in Institute of Computer Science, National Chung Hsing University, Taiwan, in 2003. Since 2005, he has been on the faculty at Hsiuping University of Science and Technology, Taiwan, ROC. He received the Ph.D. degree in the Institute of Computer Science, National Chung Hsing University, Taiwan, in 2007. His research interests include cryptography, information security, secret image sharing, image processing and digital watermarking.

A Novel and Provable Authenticated Key Agreement Protocol with Privacy Protection Based on Chaotic Maps towards Mobile Network

Hongfeng Zhu, Yifeng Zhang, Yan Zhang and Haiyang Li
(Corresponding author: Hongfeng Zhu)

Software College, Shenyang Normal University
No.253, HuangHe Bei Street, Huang Gu District, Shenyang, P.C 110034 - China
(Email:zhuhongfeng1978@163.com)

(Received Sept. 18, 2014; revised and accepted Jan. 16 & June 21, 2015)

Abstract

Key agreement is a crucial cryptographic primitive for building secure communication channels between two parties in a network. In the research literature a typical protocol aims for key secrecy and mutual authentication. However, there are many important practical scenarios where privacy protection is more desirable, especially for social network. Network privacy security means that the personal data and online data are not peep, intrusion, interference, illegal collection and utilization. In our paper, we propose a robust chaotic maps-based authentication key agreement scheme with privacy protection using smart cards. The key idea of our proposed scheme is to adopt chaotic maps for mutual authentication, not to encrypt/decrypt messages transferred between user and server, which can make our proposed scheme much more efficient. Next, we give the formal provable security under the random oracle model for our scheme. Finally, our proposed scheme can not only achieve privacy protection, but also avoid time-consuming modular exponentiation and scalar multiplication on elliptic curves. Meanwhile, it can resist various common attacks, and provide prefect forward secrecy and known-key secrecy. In brief, compared with related schemes, our proposed scheme is more secure, effective and practical.

Keywords: Chaotic maps, biometric, privacy protection, provable security, smart card

1 Introduction

1.1 Biometric Technology

At present, biometrics has widely used to certificate the identities of users. What is called biometrics is that through closely combining computer with the high-tech means of optical, acoustics, biological sensor and biological statistical principles, using physiological features (such as fingerprints, face, iris, etc) and behavior characteristics (such as voice, gait, etc) for certification of personal identity. Therefore, compared with traditional identification methods, biometric technology is safer and more convenient. It is not easy to forget, different to be stolen or counterfeit. In addition, it can be carry-on and available anytime and anywhere.

1.2 Chaotic System

Nowadays, chaos theory has widely used to cryptography. Compared with other related systems, chaotic system has numerous advantages, such as extremely sensitive to initial parameters, unpredictability, boundness, etc. Meanwhile, chaotic sequence generated by chaotic system has the properties of non-periodicity and pseudo-randomness. In a word, chaos theory and chaotic system have exploited a new way for cryptography.

1.3 Privacy Protection

In contemporary, with the rapid development of Internet, users can use personal computers or smart phones to login servers for a variety of services anytime and anywhere. However, in general, these intelligent terminals have automatic memory function. They can remember the passwords and identities of users. When these terminals of users are lost, stolen or being malicious attacked, the personal information of users is easy to leak. In consequence, it is a hot topic to protect the user privacy.

1.4 Relevant Work

In a client-server environment, authentication mechanism plays an important role in a secure protocol to certificate the identities of users. As everyone knows, in 1981, Lamport [12] firstly presented a remote authentication scheme based on password tables to certificate authored users over insecure channel. Form then on, many authentication schemes were presented and analyzed to improve the safety performance or the efficiency performance [4, 6, 9, 11, 13, 21]. Usually, alphanumeric passwords are widely used, and the security authentication of users is based on alphanumeric passwords. However, this kind of passwords is easily got by an adversary if he/she has enough time. Due to this reason, it is necessary to set up safer protection mechanisms to protect user information. Many existing schemes have been designed to solve this problem.

In 2000, Hwang et al. [9] firstly proposed the remote user authentication scheme using smart cards without a certification table to solve the problems of Lamport scheme [12]. But the passwords of users are maintained by the system. However, Chan et al. [3], Shen et al. [18] had pointed out that the scheme of [9] had flaws. In the last few years, many related schemes had been proposed, analyzed, and improved [1, 2, 7, 8, 10, 14, 15, 16, 19, 20, 22, 23, 24]. However, some of them still had defects. In 2009, Xu et al. [23] proposed a smart card based password authentication scheme with provable security. However, in 2010, Song [19] showed that the smart card authentication scheme [23] is vulnerable to internal and impersonation attacks, and proposed an efficient strong smart card authentication protocol. Unfortunately, Juan et al. [20] pointed out that the improved protocol by Song [19] cannot resist an off-line password guessing attack and also had some other weaknesses.

Then Juan et al. [20] proposed an advanced smart card based password authentication protocol in 2011. In the same year, Awasthi et al. [1] proposed a timestamp-based remote user authentication scheme using smart card without any verification table which can avoid potential risks of verification tables. In [1], remote server only kept a secret key for computing the passwords of users. Recently, many schemes based on chaos theory are proposed [2, 8, 14]. Compared with the related other schemes, these schemes based on chaotic maps avoid numerous complex operations. In 2013, Guo et al. [8] proposed a chaotic maps-based key agreement protocol which avoided modular exponential computing and scalar multiplication on elliptic curve.

Nowadays, with the fast development of Internet, privacy protection of users is a hot issue. In 2014, Liu et al. [16] proposed a multi-function password mutual authentication key agreement scheme with privacy preserving. However, this scheme was based on an elliptic curve. Its efficiency was lower than related scheme [15] based on chaotic maps because of modular exponential computing and scalar multiplication on elliptic curve. Considered the security and efficiency, we propose a robust mutual authentication key agreement scheme with privacy protection based on biometrics and chaotic maps using smart card.

1.5 Contributions

(1) Our scheme can avoid modular exponential computing and scalar multiplication and resist various attacks. (2) In our scheme, the identities of users are hidden in secure hash function. Users can anonymity login the server and do not leak any personal information. (3) Our scheme is based on chaotic maps. However, we do not use it to encrypt any message. It is only used to certificate users and server and establish a session key for their sessions. (4) Biometrics certification mechanism has many merits which can make our scheme faster and safer. According to these, we can show that our proposed scheme is more practical and effective.

1.6 Construction

The construction of our paper is organized as below: the theoretical concepts of one-way hash function and Chebyshev chaotic maps are explained in Section 2. Section 3 describes our proposed scheme in detail. Section 4 analyzes the security, functionality and efficiency of the proposed scheme. The paper is concluded in Section 5.

2 Theoretical Concepts

This section introduces the concepts of Chebyshev chaotic maps and biometrics authentication in detail.

2.1 Chebyshev Chaotic Maps

Chebyshev polynomial and Chebyshev chaotic maps [22] have the following properties:

1) Let n and x be an integer and a variable, respectively. The value of x belongs to the interval $[-1, 1]$. Chebyshev polynomial $T_n(x)$: $[-1, 1] \rightarrow [-1, 1]$ is defined as

$$T_n(x) = \cos(narccos(x)). \quad (1)$$

In terms of Equation (1), the recurrence relation of Chebyshev polynomial is defined as

$$T_n(x) = 2xT_{n-1}(x) - T_{n-2}(x), n \geq 2,$$

where $T_0(x) = 1$ and $T_1(x) = x$.

2) Chebyshev polynomial has two properties: The chaotic property: When $n \geq 1$, Chebyshev polynomial map $T_n(x)$: $[-1, 1] \rightarrow [-1, 1]$ of degree n is a chaotic map with its invariant density $f^*(x) = 1/(\pi\sqrt{1-x^2})$, for positive Lyapunov exponent $\ln n$. The semi-group property [25]: The semi-group property of Chebyshev polynomial defined on the interval $(-\infty, +\infty)$ holds, as follows:

$$T_n(x) \equiv (2xT_{n-1}(x) - T_{n-2}(x)) \bmod p$$

where $n \geq 2, x \in (-\infty, +\infty)$, and p is a large prime number. Evidently,

$$T_r(T_s(x)) \equiv T_{rs}(x) \equiv T_s(T_r(x)) \bmod p.$$

Besides, the following problems are assumed to be intractable within polynomial time.

3) Chaotic Maps-based Discrete Logarithm problem (CMDLP): Given two variables x and y, it is intractable to find the integer s, such that $T_s(x) = y$.

4) Chaotic Maps-Based Diffie-Hellman problem (CMDHP): Given x, $T_r(x)$, $T_s(x)$, it is intractable to find $T_{rs}(x)$, such that $T_r(T_s(x)) = T_{rs}(x)$ or $T_s(T_r(x)) = T_{rs}(x)$.

2.2 Biometrics certification

Figure 1 shows the flow chart of biometrics certification in detail. In the user registration phase, user inputs the biometrics in a biometric sensor, and then the system performs image processing, feature extraction, and generates template stored in the database. When performs the authentication phase, all the steps are the same until have been generated template. After that, the system draws on the database and compares the new generated template with the stored template, and then outputs the output result.

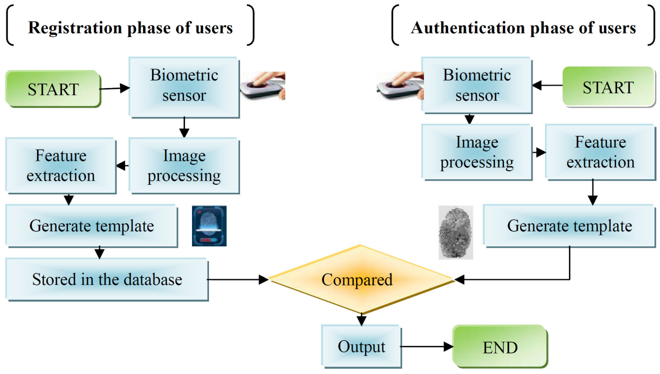


Figure 1: The composition of the proposed scheme

3 The Proposed Scheme

In this section, we introduce the proposed robust chaotic maps-based authentication key agreement scheme with privacy protection in detail. Firstly, we introduce the composition of the scheme. The proposed scheme is made up of four phases: the initialization phase, the user registration phase, the authentication key agreement phase, and the password and biometrics changing phase, respectively. Figure 2 shows the composition of the proposed scheme.

Next, we introduce the notations used in the proposed scheme. Notations are shown in Table 1.

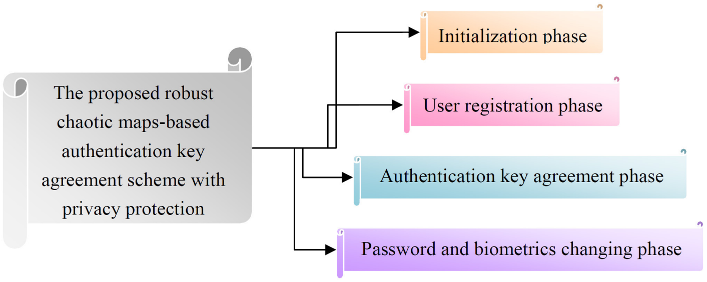

The proposed robust chaotic maps-based authentication key agreement scheme with privacy protection
Initialization phase
User registration phase
Authentication key agreement phase
Password and biometrics changing phase

Figure 2: The composition of the proposed scheme

Table 1: Notations

Notation	Definition
U_i, ID_i, PW_i	the *i*th user, the identity and password of the *i*th user, respectively
S	the server
B_i	the biometric sample of the *i*th user
τ	predetermined threshold for biometrics certification
$d(\cdot)$	symmetric parametric function
$(x, T_{k_i}(x)), k_i$	public key and secret key of the *i*th user maintained by server, respectively
m , n	random integer number
sk	session key
$h(\cdot)$	secure hash function
$\oplus$, $\|\|$	XOR operation, concatenation operation， respectively

3.1 Initialization Phase

In this phase, the server S chooses $(x, T_k(x)), k$ as its public key and secret key, and chooses a secure one-way hash function $h(\cdot)$; the ith user U_i chooses his/her identity ID_i, password PW_i and biometrics image sample B_i, respectively.

Additionally, U_i and S choose a symmetric parametric function $d(\cdot)$ and a predetermined threshold τ for biometrics certification. In each feature extraction, each different azimuth or origin of force will make the new extracted biometrics and the stored biometrics to have different degree of difference. $d(\cdot)$ is used to compute deviation degree between the results of feature extraction and the stored samples. The meaning of τ is the biggest deviation degree can be accepted.

3.2 User Registration Phase

1) U_i computes $M_i = h(ID_i || PW_i)$, $N_i = M_i \oplus h(B_i)$, and sends $N_i, h(B_i)$ to S via a secure channel.

2) S receives $N_i, h(B_i)$, stores the subscript i of N_i as an index. The subscript i is wrote in a form document which is made up of $< i, status - i >$ and $status - i$ means the login status of U_i. Then S computes $R_{U_i} = h(h(B_i)||k)$, $Z_i = R_{U_i} \oplus N_i$, stores $Z_i, N_i, h(\cdot), d(\cdot), \tau$ in a smart card, and gives the

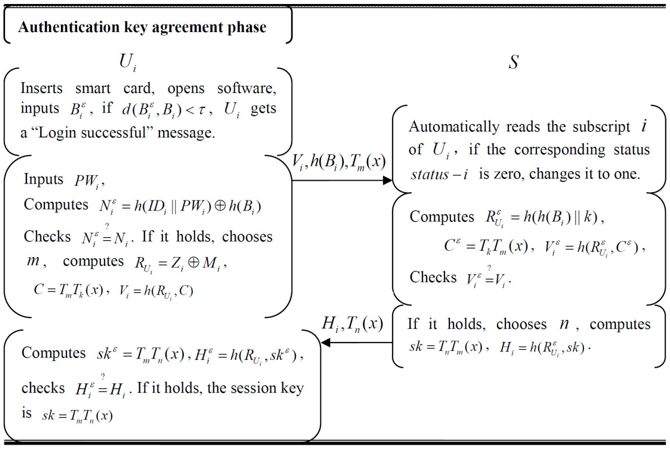
smart card to U_i via a secure channel. When U_i obtains the smart card, U_i stores B_i in it.

3.3 Authentication key agreement phase

Figure 3 shows the authentication key agreement phase as below:

1) U_i inserts the smart card into an intelligent card reader, opens the access software, inputs the biometrics B_i^ε via a sensor. Compared B_i^ε with the stored B_i, if $d(B_i^\varepsilon, B_i) \geq \tau$, U_i gets a Login failed message; if $d(B_i^\varepsilon, B_i) < \tau$, U_i gets a Login successful message.

2) After biometrics B_i login successful, U_i inputs his/her identity ID_i, password PW_i, the smart card computes $N_i^\varepsilon = h(ID_i||PW_i) \oplus h(B_i)$, and then checks whether $N_i^\varepsilon \overset{?}{=} N_i$ or not. If it does not hold, U_i gets a Wrong password message; If it holds, U_i computes $R_{U_i} = Z_i \oplus N_i$, and chooses a random integer number m, computes $C = T_m T_k(x)$, $V_i = h(R_{U_i}, C)$, and then sends $V_i, h(B_i), T_m(x)$ to S.



Inserts smart card, opens software, inputs B_i^ε, if $d(B_i^\varepsilon, B_i) < \tau$, U_i gets a "Login successful" message.

Inputs PW_i, Computes $N_i^\varepsilon = h(ID_i \| PW_i) \oplus h(B_i)$ Checks $N_i^\varepsilon \overset{?}{=} N_i$. If it holds, chooses m, computes $R_{U_i} = Z_i \oplus M_i$, $C = T_m T_k(x)$, $V_i = h(R_{U_i}, C)$

Automatically reads the subscript i of U_i, if the corresponding status $status - i$ is zero, changes it to one.

Computes $R_{U_i}^\varepsilon = h(h(B_i) \| k)$, $C^\varepsilon = T_k T_m(x)$, $V_i^\varepsilon = h(R_{U_i}^\varepsilon, C^\varepsilon)$, Checks $V_i^\varepsilon \overset{?}{=} V_i$.

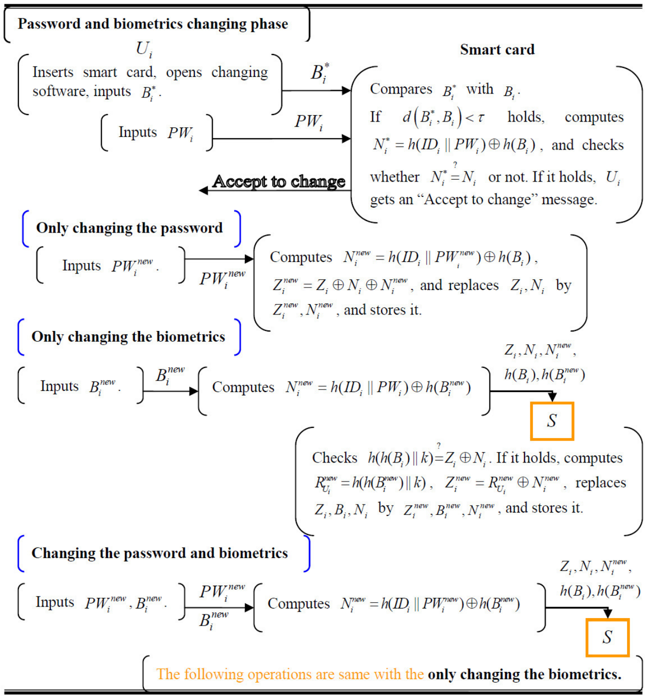
If it holds, chooses n, computes $sk = T_n T_m(x)$, $H_i = h(R_{U_i}^\varepsilon, sk)$.

Computes $sk^\varepsilon = T_m T_n(x)$, $H_i^\varepsilon = h(R_{U_i}, sk^\varepsilon)$, checks $H_i^\varepsilon \overset{?}{=} H_i$. If it holds, the session key is $sk = T_m T_n(x)$

Figure 3: The user registration and authenticated key agreement phase

3) S reads the subscript i of V_i. If the corresponding status $status-i$ of i is equal to one, S gives a Refused to login request message to U_i; if $status - i$ is equal to zero, S changes the status $status - i$ from zero to one, and then computes $R_{U_i}^\varepsilon = h(h(B_i)||k)$, $C^\varepsilon = T_k T_m(x)$, $V_i^\varepsilon = h(R_{U_i}^\varepsilon, C^\varepsilon)$, and then checks whether $V_i^\varepsilon \overset{?}{=} V_i$ or not. If it does not hold, S stops this phase; If it holds, S chooses a random integer number n, computes $sk = T_n T_m(x)$, $H_i = h(R_{U_i}^\varepsilon, sk)$, and then sends $H_i, T_n(x)$ to U_i.

4) U_i computes $sk^\varepsilon = T_m T_n(x)$, $H_i^\varepsilon = h(R_{U_i}, sk^\varepsilon)$, and then checks whether $H_i^\varepsilon \overset{?}{=} H_i$ or not. If it does not hold, U_i stop.s this phase; If it holds, U_i and S authenticate each other and the session key is $sk = T_m T_n(x)$.


Password and biometrics changing phase
U_i
Smart card
B_i^*
PW_i
Accept to change
Only changing the password
PW_i^{new}
Only changing the biometrics
B_i^{new}
Z_i, N_i, N_i^{new}, $h(B_i), h(B_i^{new})$
S
Changing the password and biometrics
PW_i^{new}
B_i^{new}
Z_i, N_i, N_i^{new}, $h(B_i), h(B_i^{new})$
S

Inserts smart card, opens changing software, inputs B_i^*.

Inputs PW_i

Compares B_i^* with B_i. If $d(B_i^*, B_i) < \tau$ holds, computes $N_i^* = h(ID_i \| PW_i) \oplus h(B_i)$, and checks whether $N_i^* \overset{?}{=} N_i$ or not. If it holds, U_i gets an "Accept to change" message.

Inputs PW_i^{new}.

Computes $N_i^{new} = h(ID_i \| PW_i^{new}) \oplus h(B_i)$, $Z_i^{new} = Z_i \oplus N_i \oplus N_i^{new}$, and replaces Z_i, N_i by Z_i^{new}, N_i^{new}, and stores it.

Inputs B_i^{new}.

Computes $N_i^{new} = h(ID_i \| PW_i) \oplus h(B_i^{new})$

Checks $h(h(B_i) \| k) \overset{?}{=} Z_i \oplus N_i$. If it holds, computes $R_{U_i}^{new} = h(h(B_i^{new}) \| k)$, $Z_i^{new} = R_{U_i}^{new} \oplus N_i^{new}$, replaces Z_i, B_i, N_i by $Z_i^{new}, B_i^{new}, N_i^{new}$, and stores it.

Inputs PW_i^{new}, B_i^{new}.

Computes $N_i^{new} = h(ID_i \| PW_i^{new}) \oplus h(B_i^{new})$

The following operations are same with the **only changing the biometrics.**

Figure 4: The password and biometrics changing phase

3.4 Password and Biometrics Changing Phase

Figure 4 shows the authentication key agreement phase as below:

1) U_i inserts the smart card into an intelligent card reader, opens the password and biometrics changing software, and inputs biometrics B_i^* at a sensor.

2) The biometrics certification process stored in the smart card compares B_i^* with B_i. If $d(B_i^*, B_i) \geq \tau$ holds, U_i gets a Refused to change message; if $d(B_i^*, B_i) < \tau$ holds, U_i inputs the password PW_i, the smart card computes $N_i^* = h(ID_i||PW_i) \oplus h(B_i)$, and then checks whether $N_i^* \overset{?}{=} N_i$ or not. If it does not hold, U_i gets a Refuse to change message; if it holds, U_i gets an Accept to change message.

   If only changing the password PW_i, U_i inputs the new password PW_i^{new}, the smart card computes $N_i^{new} = h(ID_i||PW_i^{new}) \oplus h(B_i)$, $Z_i^{new} = Z_i \oplus N_i \oplus N_i^{new}$, and then replaces Z_i, N_i by Z_i^{new}, N_i^{new}, and stores it.

   If only changing the biometrics B_i, U_i inputs the new biometrics B_i^{new}, and computes $N_i^{new} = h(ID_i||PW_i) \oplus h(B_i^{new})$, and then sends $Z_i, N_i, N_i^{new}, h(B_i), h(B_i^{new})$ to S. S checks whether $h(h(B_i)||k) \overset{?}{=} Z_i \oplus N_i$, if it holds, S computes $R_{U_i}^{new} = h(h(B_i^{new})||k)$, $Z_i^{new} = R_{U_i}^{new} \oplus N_i^{new}$, and then sends Z_i^{new} to the smart card. Then smart card replaces Z_i, B_i, N_i, by $Z_i^{new}, B_i^{new}, N_i^{new}$, and stores it.

If changing the password and biometrics in the same time, U_i inputs the new password PW_i^{new} and the new biometrics B_i^{new}, and computes $N_i^{new} = h(ID_i||PW_i^{new}) \oplus h(B_i^{new})$, and then sends $Z_i, N_i, N_i^{new}, h(B_i), h(B_i^{new})$ to S. The following operations are same with the only changing the biometrics.

4 Performance Analysis

4.1 Provable Security under the Random Oracle Model

The adversarial model of a mutual authentication and key agreement protocol is introduced as follows. Assume that a client-server environment contains two types of participants: n users U $= \{U_1, U_2, \cdots, U_i, \cdots, U_n\}$ and a server S. The ith instance of U_i is denoted by $\prod_U^i$, and the instance of the server is denoted by $\prod_S$. An adversary named A is a probabilistic polynomial time machine. Assume that A is able to potentially control all common communications in the proposed scheme via accessing to a set of oracles (as defined below). The public parameters are known by each participant.

1) *Extract*(ID_i) query:In Extract query model, A is able to obtain the private key of ID_i.

2) *Send*$(\prod_c^k, M)$ query: In Send query model, A can send a message M to the oracle $\prod_c^k$, where $c \in \{U, S\}$. When receiving the message M, $\prod_c^k$ responds to A according to the proposed scheme.

3) $h(m_i)$ query: In this query, when A makes this hash query with message m_i, the oracle $\prod_c^k$ returns a random number r_1 and records (m_i, r_1) into a list L_H. The list is initially empty.

4) *Reveal*$(\prod_c^k)$ query:In this query model, A can obtain a session key sk from the oracle $\prod_c^k$ if the oracle has accepted. Otherwise, $\prod_c^k$ returns a null to A.

5) *Corrupt*(ID_i) query: A can issue this query to ID_i and gets back its private key.

6) *Test*$(\prod_c^k)$ query: When A asks this query to an oracle $\prod_c^k$, the oracle chooses a random bit $b \in \{0, 1\}$. If $b = 1$, then $\prod_c^k$ returns the session key. Otherwise, it returns a random value. This query measures the semantic security of the session key.

In this model, A can make Send, Reveal, Corrupt and Test queries. Note that the capabilities of the adversary can make finite queries under adaptive chosen message attacks.

In next part, we show that the proposed scheme can provide the secure authentication and key agreement under the computational Chaotic Maps-Based DiffieHellman problem (CMDHP) assumption.

Theorem 1. *Assume that A can violate the proposed scheme with a non-negligible advantage ε and makes at most q_u, q_s, q_h queries to the oracle of the user $\prod_U^i$, oracle of the server S, and h, respectively. Then we can construct an algorithm to solve the CMDHP problem with a non-negligible advantage.*

Proof. We first assume the type of attack which is impersonating the user to communicate with server. Then we can construct an algorithm to solve the CMDHP problem.

For an instance of CMDHP problem $\{x, P_1 = T_{k_i}(x), P_2 = k_i\}$, B simulates the system initializing algorithm to generate the system parameters $\{x, P_{pub-u} = P_1, h\}$, is random oracles controlled by B. Then, B gives the system parameters to A. B interacts with A as follows.

h **- query:** B maintains a list L_h of tuples (str_i, h_i). When A queries the oracle h on (str_i, h_i), B responds as follows:

If str_i is on L_h, then B responds with h_i. Otherwise, B randomly chooses an integer h_i that is not found in L_h, and adds (str_i, h_i) into L_h, then responds with h_i.

Reveal - query: When the adversary A makes a *Reveal*$(\prod_c^m)$ query, B responds as follows. If $\prod_c^m$ is not accepted, B responds none. Otherwise, B examines the list L_h and responds with the corresponding h_i.

Send - query: When the adversary A makes a *Send*$(\prod_c^m, start)$ query, B responds as follows.

If $\prod_c^m = \prod_u^m$, B sets $T_m(x) \leftarrow P_1$, and randomly generates the values V_i and M_i. Otherwise, B generates a random number m^*, and computes $T_m(x) \leftarrow T_{m^*}(x)$, $C^* = T_{P_2}(T_{m^*}(x))$, $V_{i^*} = h(M_{i^*}, C^*)$, and responds with $\{V_{i^*}, M_{i^*}, T_{m^*}(x)\}$, where M_{i^*} is generated by B. The simulation works correctly since A cannot distinguish whether M_{i^*} is valid or not unless A knows the identity ID_i and the password PW_i.

When the adversary A makes a *Send*$(\prod_c^m$, $(V_{i^*}$, M_{i^*}, $T_{m^*}(x)))$ query, B responds as follows. If $\prod_c^m = \prod_u^m$, B cancels the game. Otherwise, B computes $C^* = T_{m^*}(T_{P_2}(x))$, then checks whether $V_i = h(M_{i^*}, C^*) \overset{?}{=} V_{i^*}$ to authenticate the U_i. If it holds, B computes the session key $sk = T_nT_m(x)$, $H_i = h(M_i, sk)$. Then B responds the corresponding message according to the description of the proposed scheme.

When the adversary A makes a *Send*$(\prod_c^m$, $(H_i$, $T_n(x)))$ query, B responds as follows. If $\prod_c^m = \prod_s$, B cancels the game. Otherwise, B computes $sk^* = T_n(T_{m^*}(x))$, $H_{i^*} = h(M_{i^*}, sk^*)$. If A can violate a user to the authentication, it means that A can get the values of sk and $M_i = h(h(ID_i||PW_i)||B_i)$ from the list L_h and then know the value of the session key $sk = T_mT_n(x)$. Therefore, if A an violate a user to the server authentication, then B is able to solve the CMDHP problem with non-negligible probability. It is a contradicting to the intractability of the CMDHP problem. From the above analysis, we can see that the probability that A can violate the user to the server authentication is negligible.

□

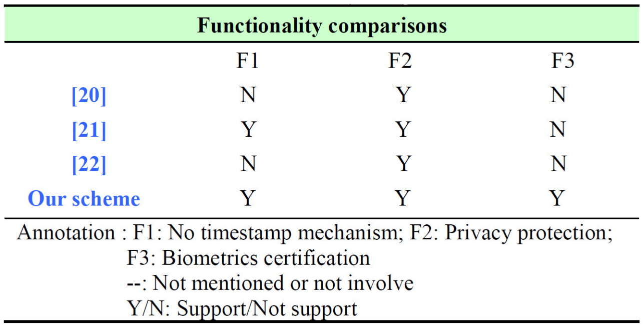
4.2 Functionality Analysis

In this subsection, Table 2 shows the functionality comparisons between our proposed scheme and related schemes about three aspects as below:

No timestamp mechanism.

Timestamp is a string produced by the current time of communication entities which can replace the random numbers at some nodes with a nonce. Unfortunately, if Alice delays delivery of the message, it may bring about the interval time for message transfer is equal or greater than ΔT,then the protocol will be stopped.

Table 2: Functionality comparisons



Functionality comparisons			
	F1	F2	F3
[20]	N	Y	N
[21]	Y	Y	N
[22]	N	Y	N
Our scheme	Y	Y	Y

Annotation : F1: No timestamp mechanism; F2: Privacy protection;
F3: Biometrics certification
--: Not mentioned or not involve
Y/N: Support/Not support

Privacy protection.

Usually, personal information of users is easy to leak. To solve this problem, our proposed scheme makes the sensitive information PW_i and ID_i hidden in a secure hash function, even if the message transferred over the insecure channel is intercepted by Alice, she cannot gain any useful information from the intercepted hash function.

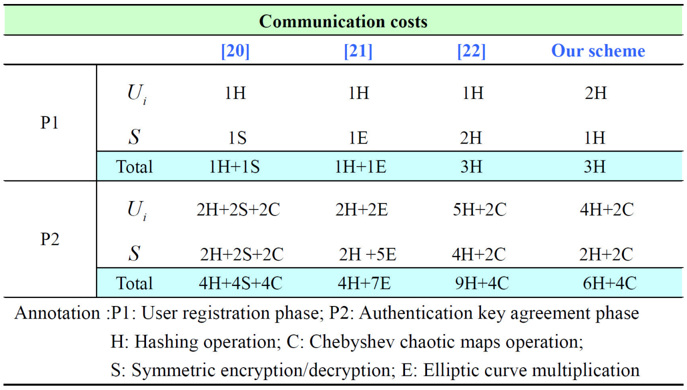
Biometrics certification.

Biometric technology is safer. It is not easy to forget, different to be stolen or counterfeit. In addition, it can be carry-on and available anytime and anywhere. In our proposed scheme, we use it as the first checkpoint of the authentication phase, not only can improve the security of our scheme, but also can increase the practicability of our scheme.

4.3 Efficiency Analysis

In this subsection, we analyze the efficiency of the proposed scheme, According to the required operations for communication entities in different phases, Table 3 summarizes the communication costs of our proposed scheme and related schemes in registration phase and authentication key agreement phase.

Table 3: Communication costs



Communication costs					
		[20]	[21]	[22]	Our scheme
P1	U_i	1H	1H	1H	2H
	S	1S	1E	2H	1H
	Total	1H+1S	1H+1E	3H	3H
P2	U_i	2H+2S+2C	2H+2E	5H+2C	4H+2C
	S	2H+2S+2C	2H +5E	4H+2C	2H+2C
	Total	4H+4S+4C	4H+7E	9H+4C	6H+4C

Annotation :P1: User registration phase; P2: Authentication key agreement phase
H: Hashing operation; C: Chebyshev chaotic maps operation;
S: Symmetric encryption/decryption; E: Elliptic curve multiplication

In Chang et al. [5] scheme, they showed that the average time for one time hash function operation was 0.605ms. In [14], Lee showed that one hash function operation was about one time faster than one Chebyshev chaotic maps operation. We can infer that the average time for one Chebyshev chaotic maps operation was about 1.21ms. In addition, according to [17], we can conclude that one hash function operation is about 10 times faster than a symmetric encryption/decryption. So a symmetric encryption/decryption operation was about 6.05ms.

According to Table 3, we can know that in registration phase, our proposed scheme only uses hash function operation, the execution time of registration phase is about 1.815ms; in the authentication phase, our proposed scheme uses hash operation and Chebyshev chaotic maps operation, the execution time of it is about 8.47ms. So compared with related schemes, the execution of our proposed scheme is acceptable, and our proposed scheme is more practical.

5 Conclusion

In the proposed scheme, we propose a robust chaotic maps-based authentication key agreement scheme with privacy protection using smart card. Our scheme has many practical merits: it refuses timestamp, modular exponentiation and scalar multiplication on an elliptic curve, and provides secure biometric authentication, chaotic maps-based authenticated key agreement, secure update protocol. Besides, chaos theory is only used to authenticate which can improve the efficiency of the proposed scheme. In the same time, the proposed scheme can resist various common attacks. In a word, compared with related schemes, the proposed scheme is safer and more practical.

References

[1] A. K. Awasthi, K. Srivastava, R. C. Mittal, "An improved timestamp-based remote user authentication scheme," *Computers and Electrical Engineering*, vol. 37, no. 6, pp. 69–874, 2011.

[2] K. Chain, W. C. Kuo, "A new digital signature scheme based on chaotic maps," *Nonlinear Dynamics*, vol. 74, no. 4, pp. 003–1012, 2013.

[3] C. K. Chan, L. M. Cheng, "Cryptanalysis of a remote user authentication scheme using smart cards," *IEEE Transactions on Consumer Electronics*, vol. 46, pp. 992–993, 2000.

[4] C. C. Chang, J. S. Lee, "An efficient and secure multi-server password authentication scheme using smart cards," in *IEEE Information Conference on Cyberworlds*, pp. 417–422, 2004.

[5] C. C. Chang, C. Y. Sun, "A Secure and Efficient Authentication Scheme for E-coupon Systems," *Wireless Personal Communications*, vol. 77, no. 4, pp. 2981–2996, 2014.

[6] F. Farhat, S. Salimi, A. Salahi, "An extended authentication and key agreement protocol of UMTS," in *Information Security Practice and Experience*, LNCS 5451, pp. 230–244, Springer, 2009.

[7] P. Gong, P. Li, W. B. Shi, "A secure chaotic maps-based key agreement protocol without using smart cards," *Nonlinear Dynamics*, vol. 70, no. 4, pp. 2401–2406, 2012.

[8] C. Guo, C. C. Chang, "Chaotic maps-based password-authenticated key agreement using smart cards," *Communications in Nonlinear Science and Numerical Simulation*, vol. 18, no. 6, pp. 1433–1440, 2013.

[9] M. S. Hwang, L. H. Li, "A new remote user authentication scheme using smart cards," *IEEE Transactions on Consumer Electronics*, vol. 46, no. 1, pp. 28–30, 2000.

[10] M. K. Khan, J. S. Zhang, X. M. Wang, "Chaotic hash-based fingerprint biometric remote user authentication scheme on mobile devices," *Chaos, Solitons and Fractals*, vol. 35, no. 3, pp. 519–524, 2008.

[11] J. Kim, S. Jun, "Authentication and key agreement method for home networks using a smart card," in *Computational Science and Its Applications (ICCSA'07)*, LNCS 4705, pp. 655–665, Springer, 2007.

[12] L. Lamport, "Password authentication with insecure communication," *Communications of the ACM*, vol. 24, no. 11, pp. 770–772, 1981.

[13] S. Laur, S. Pasini, "SAS-based group authentication and key agreement protocols," in *Public Key Cryptography (PKC'08)*, LNCS 4939, pp. 197–213, Springer, 2008.

[14] C. C. Lee, "A simple key agreement scheme based on chaotic maps for VSAT satellite communications," *International Journal of Satellite Communications and Networking*, vol. 31, no. 4, pp. 177–186, 2013.

[15] C. C. Lee, C. L. Chen, C. Y. Wu, S. Y. Huang, "An extended chaotic maps-based key agreement protocol with user anonymity," *Nonlinear Dynamics*, vol. 69, no. 1-2, pp. 79–87, 2012.

[16] T. H. Liu, Q. Wang, H. F. Zhu, "A multi-function password mutual authentication key agreement scheme with privacy preserving," *Journal of Information Hiding and Multimedia Signal Processing*, vol. 5, no. 2, pp. 165–178, 2014.

[17] B. Schneier, *Applied Cryptography, Protocols, Algorithms, and Source Code in C (2nd ed.)*, New York, Wiley, 1996.

[18] J. J. Shen, C. W. Lin, M. S. Hwang, "A modified remote user authentication scheme using smart cards," *IEEE Transactions on Consumer Electronics*, vol. 49, no. 2, pp. 414–416, 2003.

[19] R. Song, "Advanced smart card based password authentication protocol," *Journal of Computer Standards and Interfaces*, vol. 32, no. 5, pp. 321–325, 2010.

[20] J. E. Tapiador, J. C. Hernandez-Castro, P. Peris-Lopez, J. A. Clark, "Cryptanalysis of Song's advanced smart card based password authentication protocol," in *Cryptography and Securityin*, Nov. 11, 2011. (http://www. docin. com/ p-380824051.html)

[21] S. B. Wu, C. S. Li, "Identity-based SIP authentication and key agreement," *Advances in Intelligent and Soft Computing*, vol. 146, pp. 765–771, 2012.

[22] Q. Xie, J. M. Zhao, X. Y. Yu, "Chaotic maps-based three-party password-authenticated key agreement scheme," *Nonlinear Dynamics*, vol. 74, no. 4, pp. 1021–1027, 2013.

[23] J. Xu, W. T. Zhu, D. G. Feng, "An improved smart card based password authentication scheme with provable security," *Journal of Computer Standards and Interfaces*, vol. 31, no. 4, pp. 723–728, 2009.

[24] E. J. Yoon, K. Y. Yoo, "Robust biometrics-based multi-server authentication with key agreement scheme for smart cards on elliptic curve cryptosystem," *The Journal of Supercomputing*, vol. 63, no. 1, pp. 235–255, 2013.

[25] L. H. Zhang, "Cryptanalysis of the public key encryption based on multiple chaotic systems," *Chaos, Solitons & Fractals*, vol. 37, no. 3, pp. 669–674, 2008.

Hongfeng Zhu obtained his Ph.D. degree in Information Science and Engineering from Northeastern University. Hongfeng Zhu is a full associate professor of the Kexin software college at Shenyang Normal University. He is also a master's supervisor. He has research interests in wireless networks, mobile computing, cloud computing, social networks, network security and quantum cryptography. Dr. Zhu had published more than 50 international journal and international conference papers on the above research fields.

Yifeng Zhang, 24 years old, an undergraduate from Shenyang Normal University, major in information security management. During the four years of college, after completing her studies, he enjoys reading the book related to this major. Under the guidance of the teacher, he has published two articles in EI journals.

Yan Zhang, 23 years old, an undergraduate from Shenyang Normal University, major in information security management. In the four years of college, after completing her studies, she enjoys reading the book related to this major. Under the guidance of the teacher, she has published two articles in EI journals.

Haiyang Lee, graduate, graduated from Liaoning University Population Research Institute, Master demographic now at Shenyang Normal University Dean's Office Examination Management Division, lecturers title. He researches on labor and social security, wireless computer networks, network security.

A Secure Steganography Method Based on Integer Lifting Wavelet Transform

Seyyed Amin Seyyedi[1,3], Vasili Sadau[2], Nick Ivanov[3]
(Corresponding author: Seyyed Amin Seyyedi)

Department of Computer, Maku Branch, I.A.U, Maku, Iran[1]
5861993548, Maku, Iran
Department of Intelligent Systems, Belarusian State University[2]
No 4 St.Nezavisimosti, 220030, Minsk, Belarus
Department of Electronic Computing Machines, Belarusian State University of Informatics and Radioelectronics[3]
No 6 P.Brovki, 220013, Minsk, Belarus
(Email: amseyyedi@gmail.com)

(Received Nov. 18, 2013; revised and accepted July 20 & Nov 6, 2014)

Abstract

Steganography plays an important role in secret communication in digital worlds and open environments like Internet. Undetectability and imperceptibility of confidential data are major challenges of steganography methods. This article presents a secure steganography method in frequency domain based on partitioning approach. The cover image is partitioned into 8×8 blocks and then integer wavelet transform through lifting scheme is performed for each block. The symmetric RC4 encryption method is applied to secret message to obtain high security and authentication. Tree Scan Order is performed in frequency domain to find proper location for embedding secret message. Secret message is embedded in cover image with minimal degrading of the quality. Experimental results demonstrate that the proposed method has achieved superior performance in terms of high imperceptibility of stego-image and it is secure against statistical attack in comparison with existing methods.

Keywords: Cryptography, discrete wavelet transform, lifting scheme, steganography, statistical attack, tree scan order

1 Introduction

With development of the Internet and information processing techniques, data hiding has attracted lots of attention. Data hiding is a science of concealing information in a host medium that can be text, image, audio, video, etc without leaving any remarkable trace on the host medium [12, 35].Among the different kinds of digital media, the digital image is commonly used as a host image to convey side information in it. Hence, image hiding investigating is actual issue. Depending on the relationship between embedded information and the cover image, data hiding techniques are classified into steganography and watermark methods [15]. The major goal of steganography is to enhance communication security by inserting secret message into the digital image vs. copyright preserving; authentication and robustness are objectives of watermark techniques [9, 12].

Steganography is the art and science of transmission the secret message in such a way that the existence of information in container is undetectable [19, 21]. The word steganography is originally composed of two Greek words "steganos" and "graphia", which means "covered" and "writing" respectively. The notation of steganography was first introduced with the example of prisoners secret message by Simmons in 1983 [35].

There are a number of steganographic schemes hiding a secret message in an image file; these schemes can be classified according to the format of the cover image [13, 20] or the method of hiding.

Steganographic schemes in term of hiding method can be classified into two board categories namely spatial-domain techniques and frequency-domain techniques. In spatial domain techniques, the secret messages are embedded directly into cover image [8, 11, 14, 18, 22, 33, 34, 36]. The advantages of spatial domain methods are simple implementation, high payload and provide easy way to control, stego-image quality. The limitation of this approach is vulnerable to every slight steganalysis methods. Frequency domain techniques are popular data hiding approach [2, 6]. In frequency domain methods, the cover image converted into frequency domain coefficients before embedding the secret message in it. The most used transforms are the Fast Fourier Transforms (FFT), Discrete Cosines Transform (DCT), and Discrete Wavelet Transforms (DWT). Ability for high resistance against steganalysis methods and signal processing manipulations are advantages of frequency domain techniques to spatial domain ones. But transformations into frequency domain are computationally complex. Wavelets transform is a thriving branch of these methods. Some of these techniques try to achieve the high payload hiding and low distortion in cover image.

The effort to detect the presence of secret message is called steganalysis. The steganalyst is assumed to control the transmission channel and watch out for suspicious material [19]. A steganalysis method is considered as successful, and the respective steganographic system as broken, if the steganalyst be able to detect the existence of the secret message. The detection ability of statistical analysis scheme depends on the volume of hidden message [10, 23]. Hence, a secure transfer of secret message based on wavelet transform with appropriate payload without ruining the invisibility and detection by steganalyst is the aim of this study.

This study devoted to frequency domain issues; therefore it is necessary to mention relevant methods in this domain. Kang et al. [16] proposed a steganographic method based on wavelet and modulus function. First, an image is divided into blocks of prescribed size, and every block is decomposed into one-level wavelet. Then, the capacity of the hidden secret data is coordinated with the number of wavelet coefficients of larger magnitude. Finally, secret information is embedded by modulus function. Lai et al. [17] proposed an adaptive data hiding method based on Haar wavelet transform. The cover image is divided into 8×8 non-overlapping blocks, then Haar wavelet transform is performed on each blocks. A data hiding capacity function is used to determine the volume of embedding secret message in transformed sub bands. The secret message is embedded by LSB method. Safy et al. [27] to enlarge capacity of hidden data proposed the modification of Lai's method. Abdelwahab et al. [1] proposed data hiding technique in DWT where 1-level DWT is applied to both cover and secret images. Each of sub bands is divided into 4×4 non-overlapping blocks. Block of secret message is compared with cover blocks to determine the best match. The disadvantage of this method is that extracted data not totally identical to the embedded version. Raja et al. [25] proposed an adaptive steganography using integer wavelet transform. Their scheme embeds the payload in non-overlapping 4×4 blocks of the low frequency sub band. Two pixels at a time are chosen based on condition number of each block one on either side of principal diagonal. Low embedding capacity and not considering reliability of method against statistical attacks are disadvantages of this method. Bhattacharyya et al. [5] proposed a novel steganographic scheme based on Integer Wavelet Transform (IWT) through lifting scheme. The Pixel Mapping Method (PMM) is used to embed 2 bits of secret message into selected sub band to form the stego-image. The disadvantage of this method is low quality of stego-image and low payload size. Reddy et al. [26] proposed wavelet based non LSB steganography. The cover image is divided into 4×4 non-overlapping blocks, DWT/IWT applied to each block. The 2×2 cells of HH sub band are considered and manipulated with secret message bit pairs using identity matrix to generate stego-image. Seyedi et al. [28] proposed a new robust image adaptive steganography method in frequency domain. The proposed steganography method embeds the secret data in the blocks of an image that seems to be noisy based on the bit plane complexity of each block and does not destroy the co-occurrence matrix of wavelet coefficient. They used the one-third and rounding methods for embedding data in wavelet coefficients, and retain the co-occurrence matrix of wavelet coefficient. In comparison with methods mentioned earlier, our method provides better quality of image with reasonable payload, especially, high secrecy against steganalysis attacks.

This article presents frequency domain image steganography technique based on IWT through lifting scheme (LWT). In addition, to achieve higher security and authentication 56-bit key RC4 encryption method applied to the secret message before embedding procedure. Tree Scan Order (TSO) is performed in frequency domain to find the proper location of secret message. Secret message is embedded in cover image without degrading the quality of the original image.

The rest of this article organized as follows. Section two discusses the IWT based on lifting scheme. Section three presents a cryptography method. Section four presents the proposed image steganography technique. Section five presents experimental results and analysis. Conclusion is given in section six.

2 Integer Wavelet Transform Based on Lifting Scheme

Wavelets are special functional base for signal decomposition. As shown in Figure 1, applying two dimensional wavelet transform to an image represents it in four bands called LL, HL, HL, and HH. The LL band contains low pass coefficients and three other bands represent high pass coefficients of the image, including horizontal, vertical and diagonal features of the original image. The same decomposition can be applied to LL band.

Generally, wavelet filters have floating point coefficients, hence, when the input data consist of a set of integers (as in the case for images), the resulting filtered output has float point format, which does not allow exact reconstruction of the original image. However, exploiting wavelet transform with integer output provides exact inverse transform. Particularly, lifting scheme can be completely realised with integers. Above all, lifting scheme does not require temporary storage in the calculation steps [31].

In this paper biorthogonal Cohen-Daubechies Feauveau (CDF 2/2) lifting scheme was chosen as a case study.

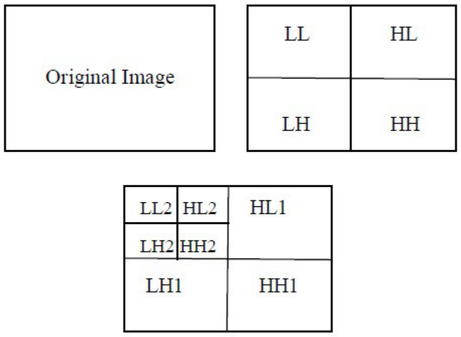


Figure 1: One level and two levels 2D wavelet transform

3 Cryptography and Steganography

One of the approaches to increasing security level of steganographic system is cryptography. Usually symmetric encryption method is recommended for steganographic methods. The symmetric encryption is a method that uses the identical key to encrypt and decrypt a secret message. In secure transmission of confidential data between parties, each party must agree on shared secret key. Based on Kerckhoffs principle [24], the security of encrypted data depends on the secrecy of the key. If attacker gains knowledge of the secret key, he can use the key to decrypt all the data. There are several algorithms for symmetric key encryption, one of them is RC4 [30].

In this paper symmetric encryption method RC4 with 56-bit key is utilized to encrypt secret message.

4 Proposed Method

In this article, a wavelet domain steganography is adopted for hiding reasonable amount of secret data with high security, good visibility and no loss of secret message. The cover image is partitioned into non-overlapping 8×8 blocks and 2D Integer LWT (IntLWT) is applied to each block. TSO is performed in transformed blocks to identify proper location of secret message. In addition, to achieve higher security and authentication 56-bit key RC4 encryption method is applied to the secret message before embedding it. The block diagram of proposed data embedding process is shown in Figure 3.

4.1 Embedding Region

The main idea behind the proposed algorithm is that secret message bits embed in proper frequency coefficients without visually degrading the quality of the original image. The TSO is applied to each transformed block to identify proper location of secret message.

4.2 Tree Scan Order

In the wavelet transform of an image, the energy in sub bands decreases as the level decomposition increases. Wavelet coefficients in upper sub bands have larger values [29]. It is possible be an edge. For human vision, the edge region has higher priority to embed the data than the smooth region. The trick is now to exploit the dependency between the wavelet coefficients across level decomposition. When 8×8 block of an image is wavelet transformed in three levels, ten sub bands will obtain as shown in Figure 3. In tree scan order upper sub (LL3) band is main root sub band. Figure 3 shows tree scanning order.

TSO use a series of decreasing thresholds and compares the wavelet coefficients with those thresholds. If the magnitude of a coefficient is smaller than a given threshold the node is called insignificant with respect to given threshold. Otherwise, the coefficient is significant. Initial threshold is calculated as:

$$T_1 = 2^{\lfloor log_2 max(B(i,j)) \rfloor} \quad (1)$$

$$T_n = \frac{T_n - 1}{2} \quad (2)$$

here $max(\cdot)$ signifies the maximum coefficient value in 8×8 block of an image and $B(i,j)$ denotes coefficients in 8×8 block.

In accordance to tree scan structure, there are spatially relation between lowest insignificant frequency coefficient at the node and children of each tree node in the next frequency sub band. The TSO is developed based on decreasing the wavelet coefficients with level of decomposition. The proposed method exploits this property for embedding secret message only into root coefficients. The coefficient is named as Root Coefficient (RC) if the value of the coefficient and its descendants are less than the threshold. RCs are the proper coefficients for embedding secret message. In proposed method threshold value for each block is calculated in three levels (T_3).

4.3 Embedding and Extracting Algorithms

Proposed secret message embedding algorithm for security point of view data transmission comprises the steps shown in Algorithm 1.

The extraction procedure consists of steps shown in Algorithm 2.

5 Experimental Results

Some experiments were conducted to assess the efficiency of the proposed method based on data payload, fidelity and security benchmarks [19]. The method has been simulated using the MATLAB 8.1 (R2013a) tools on Windows 7 version 6.1 platform. The secret message was generated randomly and RC4 of Microsoft encryption utility

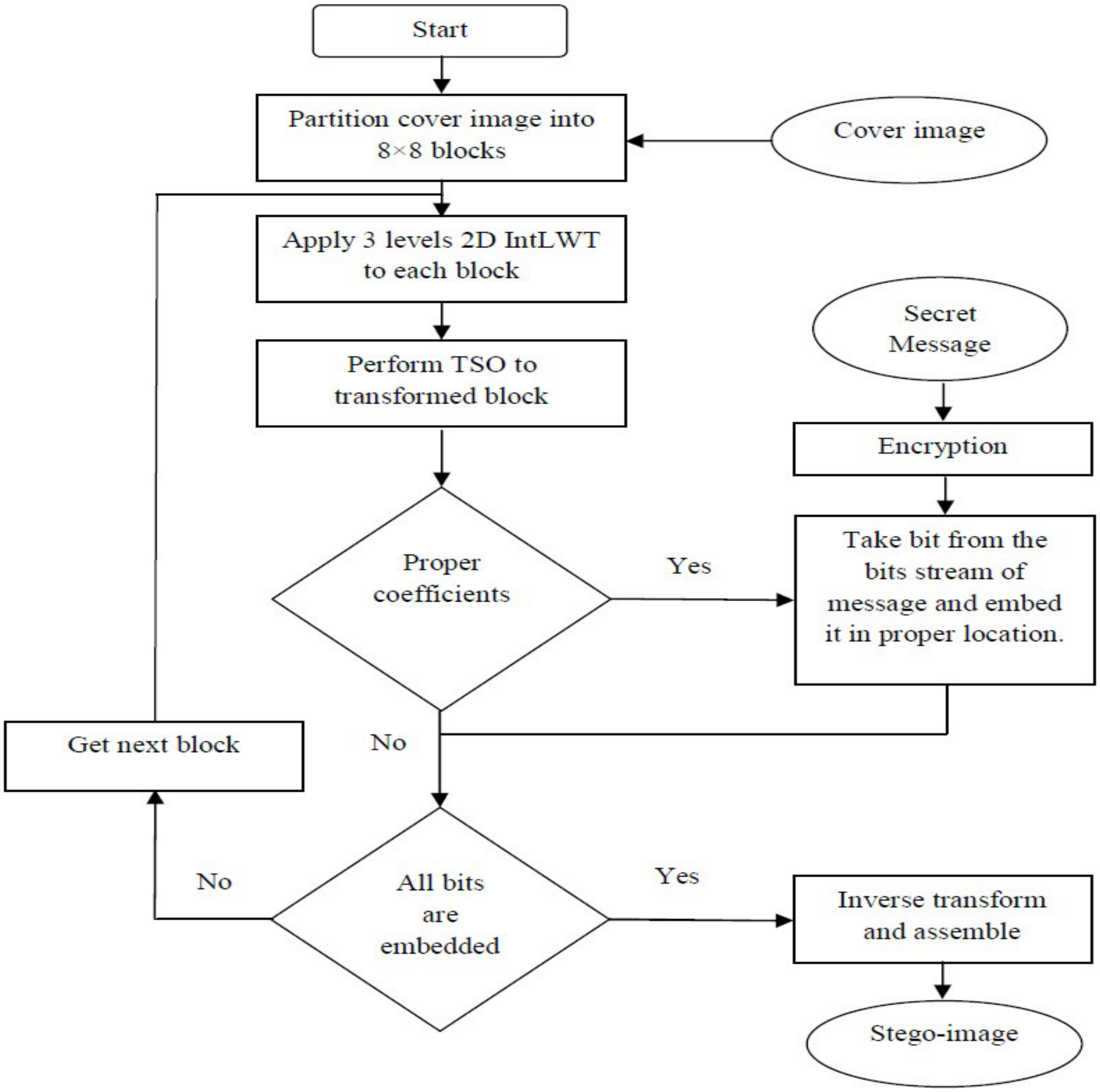

Start
Partition cover image into 8×8 blocks
Cover image
Apply 3 levels 2D IntLWT to each block
Perform TSO to transformed block
Secret Message
Encryption
Proper coefficients
Yes
Take bit from the bits stream of message and embed it in proper location.
Get next block
No
All bits are embedded
No
Yes
Inverse transform and assemble
Stego-image

Figure 2: Data embedding process

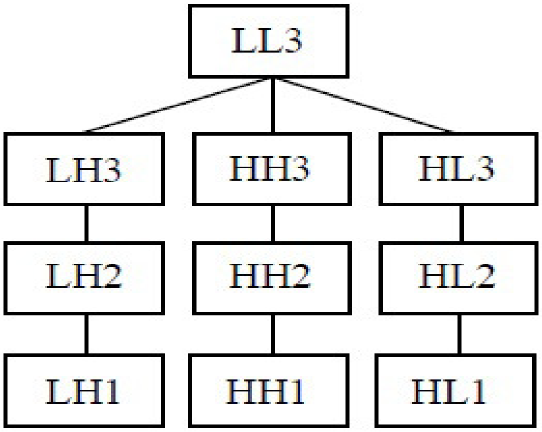


Figure 3: Tree scanning order

Table 1: Calculation of varies image similarity metrics of proposed method

Similarity metrics	Length of embedding message (byte)							
	500		1000		5000		10000	
512x512	Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.
PSNR	65.09	0.676	62.11	0.654	55.31	0.492	52.454	0.502
MSE	0.02	0.003	0.04	0.006	0.19	0.022	0.37	0.043

Algorithm 1 Embedding algorithm

```
Input: Cover image C of size $M \times N$ and secret message SE.
Output: Stego-image S.
Begin
Read cover image C.
Read the secret message SE and perform the RC4 encryption method on SE.
Partition C into $8 \times 8$ non-overlapping blocks.
Perform three levels IntLWT on each block.
Apply TSO to find RC coefficients for embedding secret message bits.
Embed SE bit by bit into RCs.
Perform inverse wavelet transform to each block.
Assemble stego-image S from blocks.
End
```

Algorithm 2 Extracting algorithm

```
Input: Stego-image S of size $M \times N$.
Output: Secret message SE.
Begin
Partition stego-image S into $8 \times 8$ non-overlapping blocks.
Apply three levels IntLWT to each block.
Find RCs and extract 1-LSB in each RC.
Decrypting extracted message by secret key.
End
```

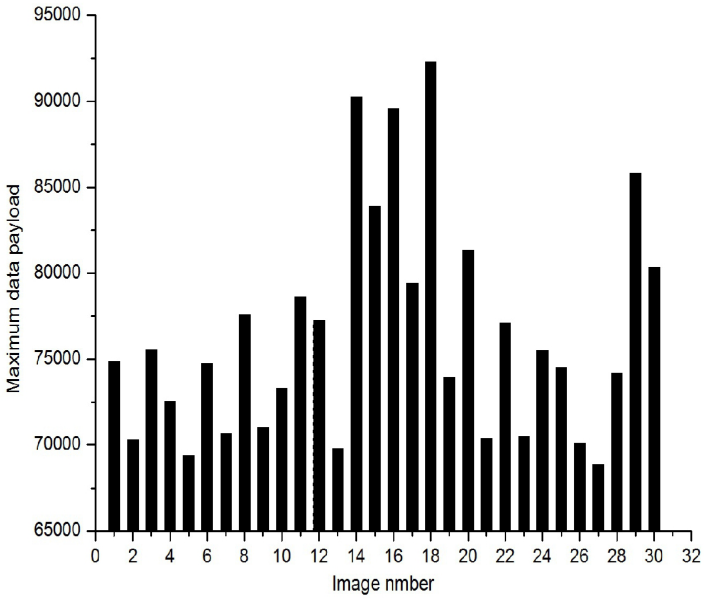


Figure 4: Maximum data payload of proposed method for several images

program was used to encrypt the secret message. All experiments were conducted on image database of Granada University [32].

Usually, data payload of steganographic method is one of the evaluation criteria. Data payload refers to the amount of information that can be hidden in the cover image. The embedding rate is usually given in absolute measurement such as the size of the secret message or in bits per pixel, etc. It depends on the embedding function, size of cover image, and may also depend on properties of the cover image. Figure 4 shows the maximum data payload of proposed method.

Fidelity (imperceptibility) refers to inability of human eyes to distinguish between cover image C and stego-image S. Usually the fidelity of stego-image measures by various image similarity metrics such as Mean Square Error (MSE), and Peak Signal to Noise Ratio (PSNR). Mean Square Error (MSE) is a simple non-perceptual error metric that is obtained from the cover image C and stego-image S where lower value are assumed to be indicative of lesser detectability. The is calculated using following formula:

$$MSE = \frac{1}{(M \times N)^2} \sum_{i=1}^{M} \sum_{j=1}^{N} (C_{ij} - S_{ij})^2. \tag{3}$$

The peak signal-to-noise ratio (PSNR) is calculated using following formula:

$$PSNR = 20log_{10} \frac{Max}{MSE} dB, \tag{4}$$

where *Max* denotes the maximum pixel value of the image. A higher PSNR value indicates the better quality of stego algorithm. Human visual system is unable to distinguish the images with PSNR more than 36 [22].

Table 1 shows the imperceptibility metrics of proposed method with various payload sizes. In order to compare proposed method with the Bhattacharyya, Reddy, and Lai methods several sizes of gray scale image Lena is used. Table 2 shows the various payloads vs. PSNR value of proposed method, Bhattacharyya method, Reddy method

Table 2: Comparison PSNR value after embedding in cover image of Lena image

Image size	Message (Byte)	CA(B)	CH(B)	CV(B)	CD(B)	Reddy Method	Lai Method (K=1)	Proposed Method
128x128	100	53.29	54.91	54.66	58.83	41.16	56.39	59.02
	200	50.66	52.35	51.994	56.90	36.29	52.76	56.19
	400	47.89	49.34	48.97	54.49	35.95	48.36	53.65
	500	47.03	48.36	48.03	48.02	35.61	47.51	52.65
256x256	100	59.77	62.36	62.34	58.83	57.93	64.83	64.75
	200	56.73	58.74	58.75	56.90	50.38	60.24	62.33
	400	53.69	55.43	55.55	54.49	43.43	57.81	59.31
	800	50.77	52.66	52.31	51.84	38.83	53.42	56.37
	1600	47.67	49.49	49.16	49.04	N/A	49.87	53.52
	2000	47.77	48.56	51.09	48.12	N/A	48.74	52.58

and Lai method for 128 × 128 and 256 × 256 cover image Lena. CA (B), CH (B), CV (B), CD (B) in Table 2 respectively denotes Approximation Coefficient, Horizontal Coefficient, Vertical Coefficient, Diagonal Coefficient of Bhattacharyya method. As shown, proposed method results are better related to the other methods in term of quality of stego-image on the same payload.

Security of steganographic system is defined in term of undetectability. There are many approaches in defining the security of a steganographic method [19]. Zollner [8] theoretically proved that a steganographic system is secure, if secret message has a random nature and is independent from the cover image and stego-image. Cachin [7] defined a steganographic method (by Kullback-Leibler KL divergence) to be ϵ-secure ($\epsilon \geq 0$), if the relative entropy between probability distribution of cover image (P_C) and stego-image (P_S) are at most ϵ. The detectability (D) is defined by:

$$D(P_C||P_S) = \int P_C log \frac{P_C}{P_S}. \qquad (5)$$

Thus, for a completely secure stego system D=0 and if $D \leq \epsilon$, then stego method is named ϵ-secure. In short, security of a stego method is defined in terms of undetectability. A steganography method is said to be undetectable or secure if the existence statistical tests cannot distinguish between the cover and the stego-image [17].

To compare the imperceptibility and security of proposed method with other methods, we did the experiments on the image data base [32]. Table 3 compares similarity and security metrics of proposed method with Lai and Reddy methods. According to it proposed method in same payload size is more imperceptible and secure than Reddy and Lai methods.

According to the results shown in Table 3, increasing the payload rate make conflict with imperceptibility metrics and security metrics.

5.1 Security Analysis of Proposed Method through Image Quality Metrics

Steganographic method is said to be undetectable or secure if the existence statistical tests cannot distinguish between the cover and the stego-images. During the embedding process in the cover image some statistical variations are arises. The stego-image is perceptually identical but is statistically different from the cover image. The attacker uses these statistical differences in order to detect the secret message. Recently various types of steganalysis methods for specific purposes have been developed in order to test the steganographic methods and detecting the stego-image from the cover image [23].

Avcibas et al. [3, 4] showed that embedding of secret message leaves unique artifacts, which can be detected using Image Quality Metrics (IQMs). There are twenty six different measures that are categorized into six groups as Pixel difference, Correlation, Edge, Spectral, Context, and Human visual system. Avcibas developed a discriminator for cover image and stego-image using a proper set of IQMs. In order to select appropriate set of IQMs, they used analysis of variance techniques. The selected IQMs for steganalysis are Minkowsky measures M1 and M2, Mean of the angle difference M4, Spectral magnitude distance M7, Median block spectral phase distance M8, Median block weight spectral distance M9, Normalized mean square HVS error M10. The IQMs scores are computed from images and their Gaussian filtered versions with $\alpha = 0.5$ and mask size 3 × 3.

The variations in IQMs for proposed method, Lai and Reddy methods with embedding the 4000 characters in cover images are computed. From experimental results it can be perceived that statistical difference between cover images and stego-images of proposed method is less than Lai and Reddy methods. Therefore, proposed method is more secured than Lai and Reddy methods. The warden cannot distinguish stego-image from cover image. The variations in IQMs for M7 and M9 are shown in Table 4.

Table 3: Comparison similarity and security metrics of proposed method with Lai and Reddy methods

Payload (Byte)	Metrics	Lai Method (K=1)		Reddy Method		Proposed Method	
		Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.
500	PSNR	59.85	2.49	46.28	6.16	69.09	0.68
	MSE	0.078	0.041	3.485	4.632	0.02	0.003
	D	3.81E-03	6.922E-03	1.460E-04	1.560E-04	1.02E-06	4.87E-07
1000	PSNR	56.62	2.47	42.40	5.55	62.12	0.65
	MSE	0.163	0.084	7.420	9.165	0.04	0.006
	D	3.82E-03	6.99E-03	2.76E-04	2.69E-04	2.19E-06	1.12E-06
2000	PSNR	53.38	2.28	38.70	4.96	59.18	0.6
	MSE	0.337	0.16	15.672	19.217	0.079	0.011
	D	3.83-03	6.989E-03	4.82E-04	3.69E-04	5.37E-06	4.3E-06
4000	PSNR	50.12	2.1	34.97	4.7	56.24	0.506
	MSE	0.7	0.311	35.932	45.768	0.115	0.018
	D	3.85E-03	6.98E-03	9.88E-04	6.6E-04	1.36E-05	1.34E-05

Table 4: IQMs variation for M7, M9

M7: Spectral magnitude distance				
Image number	Original	Proposed Method	Reddy Method	Lai Method
1	0.4723	0.4723	0.5089	0.4740
2	0.0685	0.0690	0.0753	0.0691
3	0.1258	0.1263	0.1368	0.1278
4	0.3477	0.3479	0.3781	0.3393
5	0.1207	0.1210	0.1296	0.1240
6	0.3208	0.3213	0.3420	0.3080
7	0.6414	0.6415	0.7169	0.6279
8	0.6084	0.6085	0.6584	0.6047
9	0.1862	0.1865	0.2008	0.1866
10	0.1566	0.1568	0.1714	0.1571
M9: Median block weight spectral distance				
Image number	Original	Proposed Method	Reddy MEthod	Lai Method
1	9.1477	9.1477	9.1479	9.1478
2	9.1045	9.1095	9.1112	9.1055
3	9.1269	9.1268	9.1304	9.1274
4	9.1400	9.1394	9.1404	9.1396
5	9.1191	9.1194	9.1214	9.1189
6	9.1392	9.1394	9.1398	9.1388
7	9.1481	9.1481	9.1486	9.1478
8	9.1472	9.1473	9.1482	9.1469
9	9.1485	9.1484	9.1505	9.1478
10	9.1236	9.1236	9.1284	9.1236

6 Conclusions

The major goal is addressed to security of stego algorithm. The proposed method exploited the property of tree scanning order under the integer wavelet transformation with lifting scheme and blocking approach of cover image. The confidential information could be sent in lossy channels using proposed method because it does provide sufficient secrecy and stability against statistical attack. Two layers of security are used to preserve secrecy of embedded message. Furthermore, if an attacker successes to extract secret message he will not be able to read it. The quality of stego-image occurs to be better in comparison with considered methods.

References

[1] A. A. Abdelwahab and L. A. Hassaan, "A discrete wavelet transform based technique for image data hiding," in *International Conference on Networking and Media Convergence*, pp. 1–9, Tanta, Egypt, Mar. 2008.

[2] A. Al-Ataby and A. Fawzi, "A modified high capacity image steganography technique based on wavelet transform," *The International Arab Journal of Information Technology*, vol. 7, no. 4, pp. 358–364, 2010.

[3] I. Avcibas, N. Memon, M. Kharrazi, and B. Sankur, "Image steganalysis with binary similarity measures," *EURASIP Journal on Advances in Signal Processing*, vol. 2005, no. 1, pp. 2749–2757, 2005.

[4] I. Avcibas, N. Memon, and B. Sankur, "Steganalysis using image quality metrics," *IEEE Transaction on Image Processing*, vol. 12, no. 3, pp. 221–229, 2003.

[5] S. Bhattacharyya and G. Sanyal, "Data hiding in images in discrete wavelet domain using PMM," *International Journal of Electrical and Computer engineering*, vol. 5, no. 6, pp. 597–606, 2010.

[6] K. Blossom, K. Amandeep, and S. Jasdeep, "Steganographic approach for hiding image in DCT domain," *International Journal of Advances in Engineering & Technology*, vol. 1, no. 3, pp. 72–78, 2011.

[7] C. Cachin, "An information theoretic model for steganography," *Information and Computation*, vol. 192, no. 1, pp. 41–56, 2004.

[8] C. K. Chan and L. M. Cheng, "Hiding data in images by simple LSB substitution," *Pattern Recognition*, vol. 37, no. 3, pp. 496–474, 2003.

[9] R. Chandramouli, M. Kharrazi, and N. Memon, "Image steganography and steganalysis concepts and practice," in *Digital Watermarking*, pp. 35–49, Seoul, Korea, Oct. 2003.

[10] R. Chandramouli and N. D. Memon, "Steganography capacity: A steganalysis perspective," *Security Watermarking Multimedia Contents*, SPIE 5020, pp. 173–177, 2003.

[11] C. C. Chang, J. Y. Hasiao, and C. S. Chan, "Finding optimal least significant bit substitution in image hiding by dynamic programming strategy," *Pattern Recognition*, vol. 36, no. 7, pp. 1583–1598, 2003.

[12] A. Cheddad, J. Condell, K. Curran, and P. M. Kevitt, "Digital image steganography: Survey and analysis of current methods," *Digital Signal Processing*, vol. 90, no. 3, pp. 727–752, 2010.

[13] M. Fallahpour, D. Megias, and M. Ghanbari, "Reversible and high-capacity data hiding in medical images," *IET Image Process*, vol. 5, no. 2, pp. 190–197, 2011.

[14] W. Hong, T. S. Chen, and C. W. Luo, "Data embedding using pixel value differencing and diamond encoding with multiple-base notational system," *The Journal of Systems and Software*, vol. 85, pp. 1166–1175, 2012.

[15] L. Ch Huang, L. Y. Tseng, and M. S. Hwang, "A reversible data hiding method by histogram shifting in high quality medical images," *Journal of Systems and Software*, vol. 86, no. 3, pp. 716–727, 2013.

[16] Z. Kang, J. Lin, and Y. He, "Steganography based on wavelet transform and modulus function," *Journal of Systems Engineering and Electronics*, vol. 18, no. 3, pp. 628–632, 2007.

[17] B. L. Lai and L. W. Chang, "Adaptive data hiding for images based on harr discrete wavelet transform," in *Advances in Image and Video Technology,Lecture Notes in Computer Science*, pp. 1085–1093, Hsinchu, Taiwan, December 2006.

[18] Y. P. Lee, J. C. Lee, W. K. Chen, K. C. Chang, I. J. Su, and C. P. Chang, "High-payload image hiding with quality recovery using tri-way pixel-value differencing," *Information Sciences*, vol. 191, pp. 214–225, 2012.

[19] B. Li, J. He, and J. Q. Shi, "A survey on image steganography and steganalysis," *Journal of Information Hiding and Multimedia Signal Processing*, vol. 2, no. 2, pp. 142–172, 2010.

[20] G. Liang, S. Wang, and X. Zhang, "Steganography in binary image by checking data-carrying eligibility of boundary pixels," *Journal of Shanghai University*, vol. 11, no. 3, pp. 272–277, 2007.

[21] C. S. Lu, *Multimedia Security: Steganography and Digital Watermarking Techniques for Protection of Intellectual Property.* USA: Idea Group, 2005.

[22] H. C. Lu, Y. P. Chu, and M. S. Hwang, "Image steganographic scheme based on pixel-value differencing and LSB replacement methods," *A new steganographic method of the pixel-value differencing*, vol. 50, no. 5, pp. 424–426, 2006.

[23] A. Nissar and A. H. Mir, "Classification of steganalysis techniques," *Digital Signal Processing*, vol. 90, no. 6, pp. 1758–1770, 2010.

[24] D. Omerasevic, N. Behlilovic, and S. Mrdovic, "Cryptostego a novel approach for creating cryptographic keys and messages," in *Signals and Image Processing (IWSSIP)*, pp. 83–86, Bucharest, Romania, July 2013.

[25] K. B. Raja, S. Sindhu, T. D. Mahalakshmi, S. Akshatha, B. K. Nithin, M. Sarvajith, K. R. Venugopal, and L. M. Patnaik, "Robust image adaptive steganography using integer wavelets," in *Communication Systems Software and Middleware (COMSWARE)*, pp. 614–621, Bangalore, India, Jan 2008.

[26] H. S. M. Reddy and K. B. Raja, "Wavelet based non LSB steganography," *International Journal Advanced Networking and Applications*, vol. 3, no. 3, pp. 1203–1209, 2011.

[27] R. O. El Safy, H. H. Zaye, and A. El Dessouki, "An adaptive steganographic technique based on integer wavelet transform," in *International Conference on Networking and Media Convergence*, pp. 111–117, Cario, Egypt, March 2009.

[28] S. H. Seyedi, H. Aghaeinia, and A. Sayadian, "A new robust image adaptive steganography method in wavelet transform," in *IEEE Electrical Engineering*, pp. 1–5, Tehran, Iran, May 2011.

[29] J. Shapiro, "Embedded image coding using zero tree of wavelet coefficients," *IEEE Transaction on Signal Processing*, vol. 41, no. 12, pp. 3445–3462, 1993.

[30] N. Smart, *Cryptography: An Introduction.* USA: McGraw-Hill College, 2004.

[31] W. Sweden, "The lifting scheme: A construction of second generation wavelets," *SIAM Journal on Mathematical Analysis*, vol. 29, no. 2, pp. 511–546, 1997.

[32] University of Granada, *Miscelaneous Gray Level Test Images (512 × 512)*, July 3, 2015. (http://decsai.ugr.es/cvg/dbimagenes/g512.php)

[33] C. M. Wang, N. I. Wu, C. S. Tsai, and M. S. Hwang, "A high quality steganographic method with pixel-value differencing and modulus function," *The Journal of Systems and Software*, vol. 81, no. 1, pp. 150–158, 2008.

[34] H. C. Wu, N. I. Wu, C. S. Tsai, and M. S. Hwang, "Image steganographic scheme based on pixel-value

differencing and LSB replacement methods," *IEEE Proceedings of Vision, Image and Signal Processing*, vol. 152, no. 5, pp. 611–615, 2005.

[35] N. Wu and M. S. Hwang, "Data hiding: Current status and key issues," *International Journal of Network Security*, vol. 4, no. 1, pp. 1–9, 2007.

[36] N. I. Wu, K. C. Wu, and C. M. Wang, "Exploring pixel-value differencing and base decomposition for low distortion data embedding," *Applied Soft Computing*, vol. 12, no. 2, pp. 942–960, 2012.

Seyyed Amin Seyyedi took his PhD degree in Methods and Systems of Information Protection, Information Security from Belarusian State University of Informatics and Radioelectronics in 2014. He is a member of computer department in Islamic Azad University. He has published one monograph and more than twenty publications in national and international scientific journals and conferences.His research interests include image steganography and watermark.

Vasili Sadau took his PhD degree in engineering science from National Academy of Belarus in 1984. Now he is Professor of Belarusian State University. He has published one monograph and more than sixty publications in national and international scientific journals and conferences.His research interests include the problems of information security in computer systems.

Nick Ivanov took his PhD degree in applied mathematics from National Academy of Belarus in 1978. Now he is Associate Professor of Belarusian State University Informatics and Radioelectronics. He was supervisor for several Graduate students. His research interests include discrete mathematics, image analysis, and image steganography.

A Survey of Public Auditing for Secure Data Storage in Cloud Computing

Wei-Fu Hsien[1], Chou-Chen Yang[1], and Min-Shiang Hwang[2,3]
(Corresponding author: Min-Shiang Hwang)

Department of Management Information System, National Chung Hsing University[1]
Department of Computer Science and Information Engineering, Asia University[2]
No. 500, Lioufeng Rd., Wufeng, Taichung 41354, Taiwan
(Email: mshwang@asia.edu.tw)
Department of Medical Research, China Medical University Hospital, China Medical University[3]
No. 91, Hsueh-Shih Road, Taichung 40402, Taiwan

(Received Mar. 15, 2015; revised and accepted May 11 & May 26, 2015)

Abstract

Cloud computing has been popular as the IT architecture. Cloud service providers offer many services based on cloud computing. Cloud storage service is the cloud services which can provide a huge storage space to solve the bottleneck of the storage space of local end users. However, cloud storage service may have data security because the users' data is not stored in their own storage. In this paper, we will focus on data integrity in the cloud storage service. Public auditability is a model of outsourcing data integrity verification, which can achieve efficiency and security. Therefore, we survey the previous researches of data integrity based on public auditability which includes collecting the basic requirements and evaluation metrics, providing the representative with approaches to analyze security and efficiency. Finally, we propose some future developments.

Keywords: CGA generation algorithm, hash functions, multithreading

1 Introduction

Cloud computing is a computing technology, and the Internet has grown in recent years. It can share the software and hardware resources, and provide resources to a user's computer or mobile device. The user can obtain a more efficient service because cloud computing can integrate resources. Therefore, in order to achieve cloud computing technology, it must satisfy five basic features: On-demand self-service, Broad network access, Resource pooling, Rapid elasticity and Measured service [10]. However, is very difficult for general users or small and medium enterprises to construct cloud environment because they cannot afford the huge costs. Therefore, many information technology companies are finding business opportunities to cloud services. Thus, cloud service providers have joined to build cloud environments and provide services to the user. Cloud service providers offer three services including Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS). The cost for users to rent cloud service is cheaper than the cost for users to build cloud environment.

Cloud storage service is the most common and popular service among many cloud services (e.g. Google Drive, Dropbox, Amazon S3 and Microsoft OneDrive) for general users. Users have a bottleneck in local storage space because there are more and more users to save data in cloud storage, so cloud storage service has high capacity which solves users' difficult problem. Besides, cloud storage service provides high capacity space, and, in order to achieve ubiquitous service, it also provides to access cloud services from web service or applications that utilize the application programming interface (API) by mobile devices (e.g. laptop, table computer and smart phones).

Although cloud storage service has many advantages, it brings a lot of challenging issues which include efficacy and security [5, 9]. One of the big challenges is verifying the integrity of the data because users cannot know how the cloud storage service handles their data. These cloud storage services are provided by commercial enterprises, so it cannot be fully trusted by users. Therefore, the cloud service provider may hide data loss and data errors in the service because their benefits. It is very serious when a user stores data in untrusted cloud storage, for example, a large size of the outsourced data and the client's limited resource capability, and the client how to find an efficient way to achieve integrity verifications without the local copy of data files.

In order to solve the problem of data integrity verification in the cloud storage service, many studies present different systems and security models [1, 2, 4, 6, 7, 8, 12, 13, 14, 15]. In these studies, the role of the verifier can fall into two categories: private auditability and public auditability. Private auditability implies the data owner directly verifying data in the cloud storage service is an efficient way. Public auditability implies the data owner allowing other to verify the data owner's data is inefficient. In general, the data owner may have a lot of data files which are stored in cloud storage service. However, the data owner cannot frequently verify their data because it will consume their resources which cannot process other action. In order to achieve an efficient verification of data integrity, the data owner can delegate a trusted third party auditor (TPA) to assist the validation data reduction to consume the data owner's computing resources.

The rest of paper is organized as follows: In Section 2, we review the related work of public auditability. We classify the basic requirements of function, security and efficiency in Section 3. In Section 4, we discuss the representative approaches of public auditability in detail. In Section 5, we analyze the basic requirement in the representative approaches. Finally, we summarize and discuss the future work in Section 6.

2 Related Work

In recent years, many of the literatures have pursued the context of remotely stored data verification [1, 2, 4, 6, 7, 8, 12, 13, 14, 15]. In 2007, Ateniese et al. [1] proposed the provable data possession (PDP) model which can provide public auditability and ensure possession of files on untrusted storage. They use RSA-based homomorphic verifiable tags to audit outsourced data. Their scheme first provides blockless verification and public verifiability at the same time. However, Ateniese et al.'s scheme cannot support dynamic data verification because their scheme only considers static data situation which means the client stores outsourced data and will not modify it. Therefore, Ateniese et al. [2] proposed a scalable PDP scheme to improve dynamic data verification in 2008. Nevertheless, their scheme cannot support fully dynamic data which cannot support block insertions because their scheme only allows simple block operation which implies partially dynamic data like block modification and block deletion. Wang et al. [14] proposed a challenge-response protocol which can determine the data correctness and locate possible errors. However, their scheme only supports partially dynamic data operation. Erway et al. [4] proposed a dynamic provable data possession which extends the PDP model to support fully dynamic data. They use another authenticated data structure which is a rank-based authenticated skip lists to prove and update the remote stored data. However, their scheme cannot support public verification because they only considers to achieve fully dynamic data.

Juels and Kaliski [6] proposed the proof of retrievability (POR) model, where spot-checking and error-correcting codes can make sure possession and retrievability of data files on remote archive service systems. However, their scheme only suits static data storage because the number of queries a client can perform is fixed a priori and embedding special blocks (call sentinels) which prevent the development of dynamic data updates. Shacham and Waters [12] proposed an improved POR scheme which uses BLS signature [3] to replace the RSA-based signature to reduce the proof size. They use public verifiable homomorphic linear authenticators that are built from BLS signature and secure random oracle model. They prove that it is secure in a polynomial extraction algorithm to reveal message. However, they only consider static data operation.

In order to satisfy public verification and dynamic data Wang et al. [15] proposed a new scheme in the Figure 1. Their scheme improves the index of data block which can support fully dynamic data. They extended their scheme to support batch auditing which can improve efficiency. Wang et al. [13] pointed out that Wang et al.'s scheme has data privacy issues which imply TPA can get the client's data information. Therefore, they use a random mask technology to avoid TPA learning knowledge on every verification process. Li et al. [7] consider that the client's resource-constrained device is simple and lightweight. Therefore, they propose a scheme which can delegate TPA to execute high computing process and solve the client's bottleneck. Liu et al. [8] think that previous studies are not efficient in dynamic data update because it is a fixed-size block update. Therefore, they propose a scheme which can support variable-size blocks in dynamic data update. In Section 4, we will describe these representative approaches in detail.

3 Basic Requirements and Evaluation Metrics

According to [1, 2, 4, 6, 7, 8, 12, 13, 14, 15] studies, where they provide the basic requirements of security and performance. In our paper, we classify and describe these requirements. Then we use these requirements to analyze the existing scheme in Section 4.

1) Security Evaluation:

Blockless Verification. The auditor can verify data blocks, and need not to retrieve all audited data blocks in the cloud storage service. Stateless Verification: the auditor needs not to maintain and update data situation because data situation is maintained by the client and cloud storage service together.

Batch Auditing. The auditor can verify the data of different clients at the same time because the auditor can be delegated by a lot of clients.

Dynamic Data. The data owner can insert, modify and delete data blocks in the cloud storage

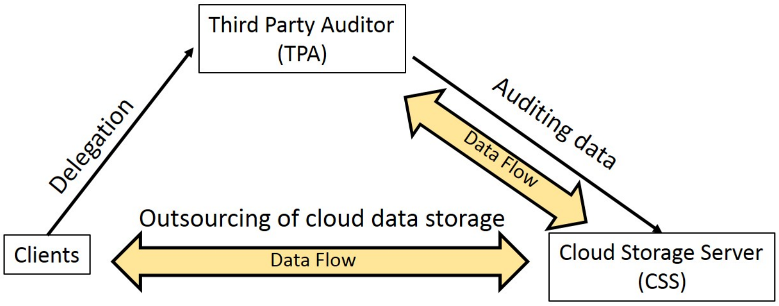

Third Party Auditor (TPA)
Delegation
Auditing data
Data Flow
Outsourcing of cloud data storage
Clients
Data Flow
Cloud Storage Server (CSS)

Figure 1: Public auditability in cloud data storage architecture

service because their data can be continuously updated at any time.

Privacy Presenting. The auditor cannot get knowledge which is the delegated data from the response of the cloud storage service.

2) Performance Evaluation:

Computing Cost. In order to achieve an efficient public auditing, we will analyze the client, TPA and cloud storage service cost on the computing resources.

Storage Cost. Because the client will upload data to the cloud storage service without the local copy of data files, we will analyze the client, TPA and cloud storage service cost on the storage spaces.

4 Representative Approaches

In the section we explain a preliminary concept and a system model before we introduce representative approaches.

4.1 Preliminary

Bilinear Pairing. Boneh et al. proposed a bilinear pairing mechanism to achieve a more efficient and secure verification. The mechanism will be explained as follows: Let G be additive group and G_T be a multiplicative group, all of prime order p. There exists a bilinear map $e : G \times G \to G_T$ and satisfies the following three properties [3]:

1) Bilinear: for all g_1, $g_2 \in G$ and $a, b \in Z_p$, $e(g_1^a, g_2^b) = e(g_1, g_2)^{ab}$.
2) Non-degenerate: if g is the generator of G, and $e(g, g)$ is the generator of G_T. It needs to satisfy $e(g, g) \neq 1$.
3) Computability: an efficient algorithm exists to compute $e(g_1, g_2)$ for any g_1, $g_2 \in G$.

Merkle Hash Tree. The Merkle Hash Tree (MHT) is an authenticated data structure intended to efficiently and securely prove that a set of elements are undamaged and unaltered [11]. It is constructed as a binary tree where the leaves in the MHT are the hashes of authentic data values.

The MHT is demonstrated in Figure 2, and the verifier wants to check whether a set of element are undamaged in outsourcing storage. First the verifier randomly chooses number of elements (where we assume a set of element have eight nodes and only chooses an element x_2) to send to the prover. The prover responses the verifier with the auxiliary authentication information (AAI) $\Omega_2 =< h(x_2), h_d, h_b >$. The verifier computes $h(x_2)$, $h_c = h(h(x_1)||h(x_2))$, $h_a = h(h_c||h_d)$ and $h_r = h(h_a||h_b)$ and then checks if the value h_r is the same as the authentic one. In public auditing, the MHT is used to authenticate both the values and the positions of data blocks. The root is constructed by the leaf nodes as the left-to-right sequence. Therefore, the leaf nodes positions can be uniquely determined by the way of computing the root in MHT.

4.2 System Model

Client. an individual consumer or organization has a lot of data files and needs to store in the cloud. It depends on the cloud to manage data and computation, so it can reduce storage cost.

Cloud Storage Service (CSS). A cloud service provider has huge storage space and computation resource to provide the clients' data.

Third Party Auditor (TPA). A trusted organization has expertise and capabilities that the clients do not have. It is responsible for assessing the clients' data on cloud storage service.

4.3 Public Auditing of Dynamic Data

Wang et al. [15] was the first to propose the scheme which can support public verification and fully dynamic data at the same time because previous studies only supported to modify and delete on a data file. They define public auditability which implies public verification is delegated by a trusted third party auditor (TPA) to verify.

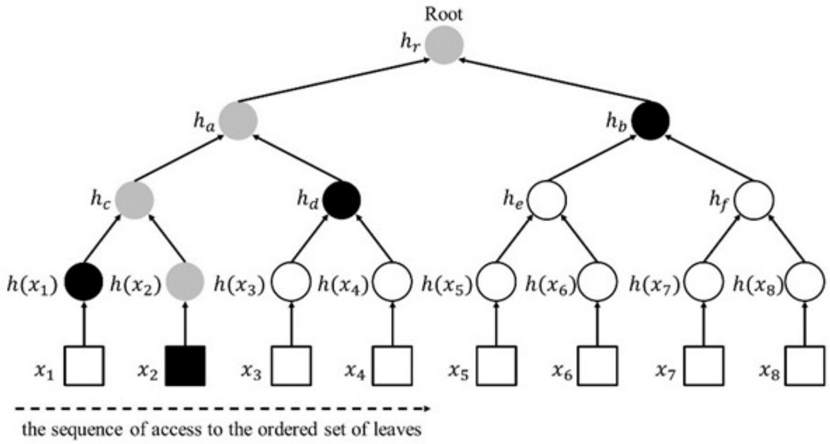


Figure 2: Merkle hash tree authentication of data elements. The leaf nodes $h(x_1), h(x_2), \cdots, h(x_n)$ is arranged in left-to-right sequence.

They propose a scheme to improve complex the file index information because this needs to consume a lot of computed resource. For example, when a file is inserted into the data block, this is required to recalculate the signature of the new index in the all files under the data block.

In order to solve the problem, they use $H(m_i)$ as the tag for block m_i instead of $H(name||i)$ [12] or $H(v||i)$ [6], and then single data operation on any file block will not affect the others. In the existing PDP or POR models $H(name||i)$ [12] or $H(v|i)$ [6] should be generated by the client in the verification process. However, in their scheme the client has no capability to compute $H(m_i)$ without the data file. In order to achieve blockless verification, the cloud storage service will process the computing $H(m_i)$ and then response it to the TPA. Because clients delegate audit services more and more, how to perform efficient audit service is a big problem. Therefore, in order to enhance the efficiency of audit services, they proposed a batch auditing protocol which can audit different client data files simultaneously. Their scheme is as follows:

1) Setup:

Setup 1. The client generates a signing key pair which is composed of signing public key (spk) and signing secret key (ssk). Then chooses a random number $\alpha \leftarrow Z_p$ and computes $v \leftarrow g^{\alpha}$. Thus, client's key pair is that the secret key is $sk = (\alpha, ssk)$ and the public key is $pk = (v, spk)$.

Setup 2. The client selects a file F which is split n blocks as $F = (m_1, m_2, \cdots, m_n)$, chooses a random element $u \leftarrow G$ and computes the file tag $t = name||n||u||SSig_s sk(name||n||u)$. The client computes signature $\sigma_i = (H(m_i) \cdot u^{m_i})^{\alpha}$ for each block and collects a signature set of $\phi = \{\sigma_i\}_{1 \le i \le n}$.

Setup 3. The client generates a root R from each hash value $H(m_i)$ of block i as a leaf node by the construction of the MHT. Then signs the root R as $Sig_{sk}(H(R)) \leftarrow (H(R))^{\alpha}$.

Setup 4. The client sends $\{F, t, \phi, sig_{sk}(H(R))\}$ to CSS. If CSS has received, the client will delete $\{F, \phi, sig_{sk}(H(R))\}$ from local storage.

2) Default Integrity Verification:

Setup 1. TPA selects c elements as a subset $I = \{s_1, s_2, \cdots, s_c\}$ from the auditing file, and chooses a random element $v_i \leftarrow Z_p$ for each block in I. Then TPA sends the challenged message $\{(i, v_i)\}_{(i \in I)}$ to CSS.

Setup 2. CSS receives the challenge of the client before computes $u = \sum_{i=s_1}^{s_c} v_i m_i \in Z_p$ and $\sigma = \Pi_{i=s_1}^{s_c} \sigma_i^{v_i} \in G$ from the stored block m_i with corresponding v_i. Then, CSS sends the proof $\{\{u, \sigma, H(m_i), \omega_i\}_{s_1 \le i \le s_c}, sig_{sk}(H(R))\}$ to TPA.

Setup 3. TPA generates the root R using $\{H(m_i), \omega_i\}_{s_1 \le i \le s_c}$ by checking

$$e(sig_{sk}(H(R)), g) \overset{?}{=} e(H(R), g^{\alpha}).$$

If the result is true, TPA verifies

$$e(\sigma, g) \overset{?}{=} e(\Pi_{i=s_1}^{s_c} H(m_i)^{v_i} \cdot u^{U}, v)$$

using the challenge message. Finally, if all results are true, TPA can make sure the client's data integrity in CSS.

3) Dynamic Data Operation with Integrity Assurance:

Setup 1. The client wants to modify the ith block m_i to m_i' and generates a new signature of block $\sigma_i' = (H(m_i') \cdot u^{m_i'})^{\alpha}$. Then the client sends the update request message (M, i, m_i', σ_i') to CSS.

Setup 2. CSS receives modification request from the client, and replaces (m_i, σ_i) with (m'_i, σ'_i). CSS computes a new root R', and sends $\{\omega_i, H(m_i), sig_{sk}(H(R)), R'\}$ to the client.

Setup 3. The client generates root R using $\{\omega_i, H(m_i)\}$ and verifies

$$e(sig_{sk}(H(R)), g) \overset{?}{=} e(H(R), g^{\alpha}).$$

If the result is true, the client generates new root R_{new} using $\{\omega_i, H(m'_i)\}$ and compares it with R'. If the result is true, the client signs R' as $sig_{sk}(H(R'))$ and sends it to CSS.

Setup 4. CSS receives the new root signature and updates it on the client file.

4) Batch Auditing:
Assume there are K clients delegate TPA to audit their data in CSS, and each client k has data files $F_i = (m_{(k,1)}, m_{(k,2)}, \cdots, m_{(k,n)})$, where $k \in \{1, 2, \cdots, K\}$. Their batch auditing protocol is as follows. In the setup and signature phase, one of the clients k chooses a random number $x_k \leftarrow Z_p$, and computes $v_k = g^x$, then the client's secret key is $sk = (x_k)$ and the public key $pk = (v_k)$. Client k chooses a random element $u_k \leftarrow G$ and computes signature $\sigma_{k,i} = (H(m_{(k,i)}) \cdot u_k^{m_{(k,i)}})^{x_k} \in G$. In the proof phase, CSS receives the challenge message $\{(i, v_i)\}_{s_1 \leq i \leq s_c}$ and computes $u_k = \sum_{(i,v_i)_{s_1 \leq i \leq s_c}} v_i m_{k,i} \in Z_p$ and $\sigma = \Pi_{i=1}^{k}(\Pi_{\{(i,v_i)\}_{s_1 \leq i \leq s_c}} \sigma_{k,i}^{v_i})$ for each client k ($k \in \{1, 2, \cdots, K\}$). CSS sends the proof $\{\sigma, \{u_k\}_{1 \leq k \leq K}, \{\omega_{k,i}\}, \{H(m_{k,i})\}\}$ to TPA. In the verification phase, first TPA generates the roots using $\{\{\omega_{k,i}\}, \{H(m_{k,i}\}\}$ and verifies the roots for each client's file. If the result is true, TPA verifies $e(\sigma, g) \overset{?}{=} \Pi_{k=1}^{K} e(\Pi_{(i,v_i)_{s_1 \leq i \leq s_c}}(H(m_{k,i}))^{v_i} \cdot (u_k)^{u_k}, v_k)$ using the challenge message to combine the bilinear map. Finally, if all results are true, TPA can make sure the clients' data integrity in CSS.

4.4 Public Auditing of Privacy-Preserving Data

Wang et al. [13] proposed a privacy protection scheme which is considered user's data privacy in the public auditability. Data privacy implies personally identifiable information or sensitive information whether they can be shared with third parties. As far as users are concerned what they depend on TPA just for the outsourced storage security of their data. However, most studies do not consider the protection of clients' private information in the auditing phase. This is a serious problem because an auditor may leak information without the client's authorization. Besides, there are legal regulations, such as the Health Insurance Portability and Accountability Act (HIPPA), it guarantees patient confidentiality for all healthcare-related data and demands the outsourced data not to be leaked to external parties.

Because public auditing model allows third-party auditors to assist clients to verify their data integrity, TPA obtains partly data blocks and learns by each sample to collect information in the auditing phase. For instance, Wang et al. [15] proposed a scheme where TPA sends the challenged message $\{(i, v_i)\}_{i \in I}$ to CSS and CSS responses the proof $\{\{u, \sigma, H(m_i), \omega_i\}_{s_1 \leq i \leq s_c}, sig_{sk}(H(R))\}$ to TPA. Therefore, TPA uses Homomorphic Linear Authenticator (HLA) characteristic to combine the blocks $u = \sum_{i=s_1}^{s_c} v_i m_i$, and it can potentially reveal user's data. TPA can gather the same set of c block $(m_1, m_2, \cdots, m_c)$ with corresponding random coefficients $\{v_i\}$. TPA can get the user's data $(m_1, m_2, \cdots, m_c)$ by computing different linear combinations $(u_1, u_2, \cdots, u_c)$. Thus it can be seen that it infringes the privacy-preserving guarantee.

Wang et al. proposed to integrate the homomorphic linear authenticator with random masking technique, and it achieves privacy-preserving public auditing. Because the random masking technique affects TPA learning knowledge, it can avoid TPA getting user's data. We are not going to elaborate on their scheme because it is similar to Wang et al.'s scheme on dynamic data and batch auditing operation.

Their scheme is as follows:

1) Setup:

Setup 1. The client chooses a random signing key pair (spk, ssk), a random number $x \leftarrow Z_p$, a random element $u \leftarrow G$, and computes $v \leftarrow g^x$. The secret key is $sk = (x, ssk)$ and the public key is $pk = (spk, v, g, u, e(u, v))$.

Setup 2. The client computes signature $\sigma_i = (H(W_i) \cdot u^{m_i})^x$ for each block where $W_i = name||i$ is combined with the user's identification $name$ and the block index i. Then, the client collects a signature set of $\phi = \{\sigma_i\}_{1 \leq i \leq n}$ and computes the file tag $t = name||SSig_{ssk}(name)$.

Setup 3. The client sends (F, ϕ, t) to CSS. If CSS has received, the client will delete (F, ϕ) from local storage.

2) Integrity Verification:

Setup 1. TPA selects c elements as a subset $I = \{s_1, s_2, \cdots, s_c\}$ from the auditing file, and chooses a random element $v_i \leftarrow Z_p$ for each block in I. Then TPA sends the challenged message $\{(i, v_i)\}_{i \in I}$ to CSS.

Setup 2. CSS receives challenge $\{(i, v_i)\}_{i \in I}$ and generates a response proof of data storage correctness. First, CSS computes $u'_i = \sum_{i=s_1}^{s_c} v_i m_i$ and $\sigma = \Pi_{i=s_1}^{s_c} \sigma_i^{v_i}$ which are corresponding to m_i and σ_i in the CSS's storage. Second, CSS randomly chooses an element $r \leftarrow Z_p$ and computes $R = e(u, v) \in G_T$ and $\gamma = H(R) \in Z_p$.

Finally, CSS computes $u = r + \gamma$ which is random masking technique, and sends $\{u, \sigma, R\}$ to TPA.

Setup 3. TPA receives the response proof of storage correctness, and computes $\gamma = H(R)$ to verify the equation $R \cdot e(\sigma^{\gamma}, g) \overset{?}{=} e((\Pi_{i=s_1}^{s_c} H(W_i^{v_i})^{\gamma} \cdot u^{u}, v)$. If the result is true, TPA can make sure the the client's data integrity, and cannot learn any knowledge about the data content stored in CSS.

4.5 Public Auditing of Resource-constrained Devices

Li et al. [7] propose a public auditability scheme in resource-constrained devices. Li et al.'s cloud data storage architecture is shown in Figure 3. Resource-constrained device is a simple and lightweight composition. Thus, these devices have low computation and storage capacity. However, these devices can achieve high mobility which allows users to carry and easily to use. Because the client may require repeatedly modified data in cloud storage service, this operation needs to compute in every update. Therefore, in the public audit model, the client needs a high burden of computing resources to operate dynamic data (such as signature $\sigma_i = (H(m_i) \cdot u^{m_i})^{\alpha}$ [15]) which is required to perform exponentiation and multiplication operation. In order to reduce the client's computation Li et al. propose a scheme which delegates trusted TPA to generate key, signature and delete file tag function. The clients can effectively reduce the computing resources because they only upload data to TPA. Therefore, the client will not have to compute signature on data update every time. Their scheme is as follows:

1) Integrity Verification:

Setup 1. TPA selects c elements as a subset $I = \{s_1, s_2, \cdots, s_c\}$ from the auditing file, and chooses a random element $v_i \leftarrow Z_p$ for each block in I. Then TPA sends the challenged message $\{(i, v_i)\}_{i \in I}$ to CSS.

Setup 2. CSS receives the challenge of the client before computes $u_j = \sum_{1 \leq i \leq c} v_i M_{ij} \in Z_p$ for $j = 1, 2, \cdots, s$ and $\sigma = \Pi_{1 \leq i \leq c} \sigma_i^{v_i} \in G$. CSS also provides some relevant information to TPA to verify client's data, and it includes $\{H(M_i, \Omega_i)\}_{1 \leq i \leq c}$ and $sig_{sk}(H(R))$. Finally, CSS responses $P = \{\{u_j\}_{1 \leq i \leq s}, \sigma, \{H(M_i, \Omega_i)\}_{1 \leq i \leq c}, sig_{sk}(H(R))\}$ to TPA.

Setup 3. TPA receives the response proof of storage correctness. First TPA generates the root R' using $\{H(M_i, \Omega_i)\}_{1 \leq i \leq c}$ and checks $sig_{sk}(H(R')) \overset{?}{=} sig_{sk}(H(R))$. Second TPA checks $e(t, g) \overset{?}{=} e(H(R), v)$. Finally, TPA checks whether $e(\sigma, g) \overset{?}{=} e(\Pi_{1 \leq i \leq c} H(M_i)^{v_i} \cdot \Pi_{j=1}^{s} u_j^{u_j}, v)$. If all results are true, TPA can make sure the clients' data integrity in CSS.

4.6 Authorized Public auditing of Fine-Grained Update

These schemes can support public auditing and dynamic data update. However, these schemes [15, 13, 7] support to insert, delete and modify operation in a fixed-size block which is later termed as coarse-grained updates. For instance, when a data block is partially modified, the block will be completely modified in coarse-grained updates. Therefore, this will cost additional resource. Liu et al. [8] propose a variable-size block scheme which is later termed as fine-grained updates in the public auditing. Their scheme can reduce an additional operation in partially modified block update. They also consider an authentication process to improve between the client and TPA because Wang et al.'s scheme [13] proposes challenge issues where TPA may learn the client's data by the verification process. Their scheme is as follows:

1) Integrity Verification:

Setup 1. TPA selects c elements as a subset $I = \{s_1, s_2, \cdots, s_c\}$ from the auditing file, and chooses a random element $v_i \leftarrow Z_p$ for each block in I. In order to achieve authentication, they add sig_{AUTH} and $\{VID\}_{PK_{CSS}}$ where $sig_{AUTH} = Sig_{ssk}(AUTH||t||VID)$ is include the client and TPA information and $\{VID\}_{PK_{CSS}}$ is means use CSS's public key to encrypt TPA's identification. Then TPA sends the challenged message $\{sig_{AUTH}, \{VID\}_{PK_{CSS}}, (i, v_i)\}_{i \in I}$ to CSS.

Setup 2. CSS receives the challenge of the client before verifies sig_{AUTH} with $AUTH, t, VID$ and the client's public key. If these verification are false, CSS reject it. Otherwise, CSS will compute $u_k = \sum_{i \in I} v_i m_{ik}$, ($k \in [1, w]$ and $w = max\{s_i\}_{i \in I}$) and compute $\sigma = \Pi_{i \in I} \sigma_i^{v_i}$. CSS provides the client's signature information sig from cloud storage. Finally, CSS responses $P = \{\{u_k\}_{k \in [1,w]}, \{H(m_i, \Omega_i)\}_{i \in I}, sig\}$ to TPA.

Setup 3. TPA receives the response proof of storage correctness. First TPA generates the root R' using $\{H(m_i, \Omega_i)\}_{i \in I}$ and checks $e(sig, g) \overset{?}{=} e(H(R'), v)$. Second TPA checks $e(\sigma, g) \overset{?}{=} e(w, v)$. Finally, TPA checks whether $e(\sigma, g) \overset{?}{=} e(\Pi_{i \in I} H(m_i)^{v_i} \cdot \Pi_{k \in [1,w]} u_k^{u_k}, g^{\alpha})$. If all results are true, TPA can make sure the clients' data integrity in CSS.

2) Dynamic Data Operation with Integrity Assurance:

Setup 1. The client wants to partial modify the ith block m_i to m_{new}. Therefore, the client computes update length in the ith block m_i. Then

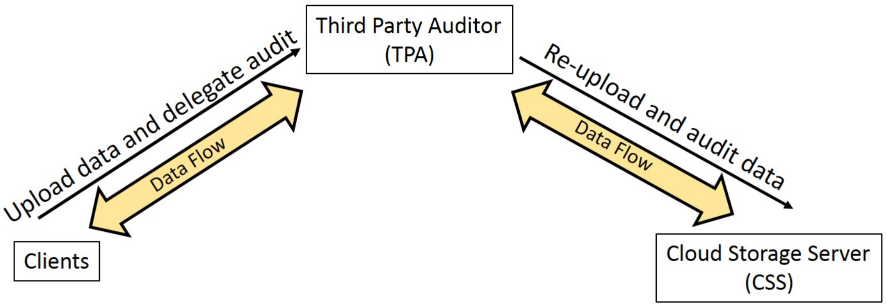


Figure 3: Li et al.'s cloud data storage architecture

the client sends the update request message $\{PM, i, o, m_{new}\}$ to CSS.

Setup 2. CSS receives the request from the client. First, CSS can ensure that the request is the partial modification (PM). Second, CSS uses $\{o, m_{new}\}$ to gather the sectors not involved in this update, which denote as $\{m_{ij}\}_{j\in M}$. Third, CSS will perform the update to get m'_i and use $\{m'_i, \Omega_i\}$ to compute R'. Finally, CSS responses the proof $P = \{\{m_ij\}_{j\in M}, H(m_i), \Omega_i, R', sig\}$ to the client.

Setup 3. The client generates root R using $\{\Omega_i, H(m_i)\}$ and verifies the signature $sig \overset{?}{=} (H(R))^\alpha$. If it success, then computes m'_i using $\{m_{ij}\}_{j\in M}, m_{new}\}$. Thus, CSS can compute R_new using $\{m'_i, \Omega_i\}$ and verifies $R_{new} \overset{?}{=} R'$. If all results are true, the client computes the new signature block $\sigma'_i = (H(m'_i)\Pi_{j=1}^{s_i} u_j^{m'_{ij}})^\alpha$ and the new signature root $sig' = (H(R'))^\alpha$, and then returns update message $\{\sigma'_i, sig'\}$ to CSS.

Setup 4. CSS receives the message, and then updates it on the client file.

5 Analysis

In the section, we will analyze these schemes [7, 8, 13, 15] which contain functional requirement, security and performance. And we also use the tables to present a corresponding requirement in each scheme.

5.1 Functional Requirement

In order to raise efficiency in verification, every scheme can support blockless verification. The comparison of functional requirement with related schemes is shown as Table 1. Because Li et al.'s scheme [7] needs TPA to assist the client's data file, their scheme does not satisfy stateless verification. Although Li et al. [7] and Liu et al. [8] did not explain whether their scheme support batch audit, we analyze whether their scheme can be extended to achieve it. In the dynamic data, because these scheme [7, 13, 15] do not consider partially modified data update, Liu et al. [8] only considered to update variable-size blocks. Wang et al. [13] only considered privacy presenting using random mask technology because other schemes assume that TPA can be fully trusted.

5.2 Performance Evaluation

We will analyze three phases: setup phase, auditing phase and dynamic data update phase. Before we analyze the performance evaluation, first we introduce the notations in Table 2. In Tables 3, 4, 5, we analyze the computation cost in setup, auditing, and dynamic data phases, respectively.

In the setup phase, Wang et al.'s scheme [15] is better than these schemes [7, 8, 13] because their scheme does not compute the number of sectors of a block. However, Li et al's scheme [7] is best on the client's point of view because the client delegates the whole operation process to TPA.

In the auditing phase, Wang et al.'s scheme [13] is better because the auditor reduces computation which cannot construct the root in the auditing phase. However, Liu et al.'s scheme [8] requires costly computing, but their scheme is the only way to achieve between TPA and CSS authentications.

In the dynamic data update phase, Liu et al.'s scheme [8] is better because their scheme can support partially modified data update which can reduce computing.

In the Table 6, we analyze storage cost in public auditing. Liu et al.'s scheme [12] requires a large storage space because their scheme can support partially modified data update and authentication. Li et al.'s scheme [7] needs to store some information on the TPA because their scheme make the client delegate TPA to perform signature.

6 Conclusions and Future Work

Because users' data is stored in the cloud storage service, it brings users' data security issues. In the public auditability model, users can delegate the third party auditor to verify their data is efficient. According to the literature, we sort out the basic requirements in public

Table 1: Comparison of functional requirements

	Wang et al. [15]	Wang et al. [13]	Li et al. [7]	Liu et al. [12]
Blockless verification	Yes	Yes	Yes	Yes
Stateless verification	Yes	Yes	No	Yes
Batch auditing	Yes	Yes	Yes	Yes
Dynamic data	Partial	Partial	Partial	Yes
Privacy presenting	No	Yes	No	No

Table 2: Notations

Notation	Description
T_E/T_D	The computing time of asymmetric encryptions;
T_{Ge}	The computing time of exponentiation in group operation;
T_{BLS}	The computing time of BLS signature;
T_B	The computing time of bilinear pairing;
T_M	The computing time of multiplication;
T_A	The computing time of addition;
T_{GM}	The computing time of multiplication in group operation;
T_h	The computing time of hash function;
n	The number of block in a file;
i	The number of verified block;
l	The number of inside node that needed in MHT;
o	The number of auxiliary authentication information (AAI);
s	The number of sectors of a block;
s_{max}	The maximum number of sectors a block.

Table 3: Comparison of computation in Setup phase

	Client	TPA
Wang et al. [15]	$(n+3)T_{BLS}+n(T_{GM}+T_{Ge})+(n+l)T_h$	No
Wang et al. [13]	$(n+3)T_{BLS}+n(T_{GM}+T_{Ge})+(2n+l)T_h$	No
Li et al. [7]	No	$(n+3)T_{BLS}+n(T_{GM}+s_{max}T_{Ge}))+(n+l)T_h$
Liu et al. [8]	$(n+3)T_{BLS}+n(T_{GM}+s_{max}T_{Ge}))+(n+l)T_h$	No

Table 4: Comparison of computation in Auditing phase

	TPA	CSS
Wang et al. [15]	oT_h+2T_B	$i(T_M+T_A+T_{Ge}+T_{GM})+(i+o)T_h$
Wang et al. [13]	T_h+T_B	$(i+1)(T_{GM}+T_A+T_{Ge})+iT_M+T_h$
Li et al. [7]	oT_h+2T_B	$i(T_M+T_A+T_{Ge}+T_{GM})+(o+i)T_h$
Liu et al. [8]	$T_E+oT_h+2T_B$	$T_D+i(T_M+T_A+T_{Ge}+T_{GM})+(o+i)T_h$

Table 5: Comparison of computation in Dynamic Data phase

	Client	TPA	CSS
Wang et al. [15]	$(2o+1)T_h + 2T_{BLS} + T_B$ $+(s_{max}+1)T_{GM} + s_m ax T_{Ge}$	No	$(o+l+2)T_h$
Wang et al. [13]	$(2o+1)T_h + 2T_{BLS} + T_B$ $+(s_{max}+1)T_{GM} + s_{max}T_{Ge}$	No No	$(o+l+2)T_h$ $(o+l+2)T_h$
Li et al. [7]	No No	$(2o+1)T_h + 2T_{BLS} + T_B$ $++(s_{max}+1)T_{GM} s_{max}T_{Ge}$	$(o+l+2)T_h$ $(o+l+2)T_h$
Liu et al. [8]	$(2o+1)T_h + 2T_{BLS} + T_B + T_{GM}$ $+(s+1)T_{GM} + sT_{Ge}$	No No	$(o+l+2)T_h$ $(o+l+2)T_h$

Table 6: Comparison of storage

Storage cost	Wang et al. [15]	Wang et al. [13]	Li et al. [7]	Liu et al.[8]
Auditor	No	No	$sig_{sk}(R)$	No
Cloud Storage Service	$F, t, \phi, sig_{sk}(H(R))$	F, t, ϕ	F, ϕ	$F, T, t, \phi, R, sig_{sk}(H(R))$

auditability, which can be classified to the case for your application.

For future development, with big data generation, data verification will be more and more difficult. Because big data have three characteristics including volume, velocity and variety, these characteristics will affect the implementation of data verification. Therefore, it will be a major challenge how to efficiently verify data integrity in big data. However, this scheme must also satisfy basic requirements.

References

[1] G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson, and D. Song, "Provable data possession at untrusted stores," in *Proceedings of the 14th ACM Conference on Computer and Communications Security*, pp. 598–609, Virginia, USA, 2007.

[2] G. Ateniese, R. D. Pietro, L. V. Mancini, and G. Tsudik, "Scalable and efficient provable data possession," in *Proceedings of the 4th International Conference on Security and Privacy in Communication Netowrks*, pp. 9:1–9:10, Istanbul, Turkey, 2008.

[3] D. Boneh, B. Lynn, and H. Shacham, "Short signatures from the weil pairing," in *Proceedings of the 7th International Conference on the Theory and Application of Cryptology and Information Security: Advances in Cryptology (ASIACRYPT'01)*, pp. 514–532, Gold Coast, Australia, 2001.

[4] C. Erway, A. K, C. Papamanthou, and R. Tamassia, "Dynamic provable data possession," in *Proceedings of the 16th ACM Conference on Computer and Communications Security*, pp. 213–222, Illinois, USA, 2009.

[5] I. A. T. Hashem, I. Yaqoob, N. B. Anuar, S. Mokhtar, A. Gani, and S. U. Khan, "The rise of big data on cloud computing: Review and open research issues," *Information Systems*, vol. 47, no. 6, pp. 98–115, 2015.

[6] A. Juels and J. Burton S. Kaliski, "Pors: Proofs of retrievability for large files," in *Proceedings of the 14th ACM Conference on Computer and Communications Security*, pp. 584–597, Virginia, USA, 2007.

[7] J. Li, X. Tan, X. Chen, D. Wong, and F. Xhafa, "OPoR: Enabling proof of retrievability in cloud computing with resource-constrained devices," accepted and to be publish in *IEEE Transactions on Cloud Computing*, Oct. 2014.

[8] C. Liu, J. Chen, L. T. Yang, X. Zhang, C. Yang, R. Ranjan, and K. Ramamohanarao, "Authorized public auditing of dynamic big data storage on cloud with efficient verifiable fine-grained updates," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 9, pp. 2234–2244, 2014.

[9] C. Liu, C. Yang, X. Zhang, and J. Chen, "External integrity verification for outsourced big data in cloud and iot: A big picture," *Future Generation Computer Systems*, vol. 49, no. 6, pp. 58–67, 2015.

[10] P. M. Mell and T. Grance, "The nist definition of cloud computing," Technical Report: SP 800-145, 2011.

[11] R. C. Merkle, "Protocols for public key cryptosystems," in *IEEE Symposium on Security and Privacy*, pp. 122–134, California, USA, 1980.

[12] H. Shacham and B. Waters, "Compact proofs of retrievability," in *Proceedings of the 14th International Conference on the Theory and Application of Cryptology and Information Security (ASIACRYPT'08)*, pp 90–107, Melbourne, Australia, 2008.

[13] C. Wang, S. S. M. Chow, Q. Wang, K. Ren, and W. Lou, "Privacy-preserving public auditing for se-

cure cloud storage," *IEEE Transactions on Computers*, vol. 62, no. 2, pp. 362–375, 2013.

[14] C. Wang, Q. Wang, K. Ren, and W. Lou, "Ensuring data storage security in cloud computing," in *Proceedings of the 17th International Workshop on Quality of Service (IWQoS'09)*, pp. 1–9, South Carolina, USA, 2009.

[15] Q. Wang, C. Wang, K. Ren, W. Lou, and J. Li, "Enabling public auditability and data dynamics for storage security in cloud computing," *IEEE Transactions on Parallel and Distributed Systems*, vol. 22, no. 5, pp. 847–859, 2011.

Wei-Fu Hsien received his B. S. in Department of Information Management from National Kaohsiung Marine University, Kaohsiung, Taiwan, ROC, in 2013. He is currently pursuing the M.S. degree with the Department of Management Information Systems from National Chung Hsing University. His research interests include security and privacy of cloud computing, and applied cryptography.

Chou-Chen Yang received his B.S. in Industrial Education from the National Kaohsiung Normal University, in 1980, and his M.S. in Electronic Technology from the Pittsburg State University, in 1986, and his Ph.D. in Computer Science from the University of North Texas, in 1994. From 1994 to 2004, Dr. Yang was an associate professor in the Department of Computer Science and Information Engineering, Chaoyang University of Technology. Currently, he is a professor in the Department of Management Information Systems, National Chung Hshing University. His research interests include network security, mobile computing, and distributed system.

Min-Shiang Hwang received the B.S. in Electronic Engineering from National Taipei Institute of Technology, Taipei, Taiwan, Republic of China, in 1980; the M.S. in Industrial Engineering from National Tsing Hua University, Taiwan, in 1988; and the Ph.D. in Computer and Information Science from National Chiao Tung University, Taiwan, in 1995. He also studied Applied Mathematics at National Cheng Kung University, Taiwan, from 1984-1986. Dr. Hwang passed the National Higher Examination in field "Electronic Engineer" in 1988. He also passed the National Telecommunication Special Examination in field "Information Engineering", qualified as advanced technician the first class in 1990. From 1988 to 1991, he was the leader of the Computer Center at Telecommunication Laboratories (TL), Ministry of Transportation and Communications, ROC. He was also a project leader for research in computer security at TL in July 1990. He obtained the 1997, 1998, and 1999 Distinguished Research Awards of the National Science Council of the Republic of China. He is a member of IEEE, ACM, and Chinese Information Security Association. His current research interests include database and data security, cryptography, image compression, and mobile communications.

3C-Auth: A New Scheme for Enhancing Security

Narasimhan Harini and Tattamangalam R. Padmanabhan
(Corresponding author: Narasimhan Harini)

Department of Computer Science and Engineering & Amrita University
Amritanagar P.O., Ettimadai, Coimbatore 641 112, India
(Email: nharini2003@gmail.com)

(Received Aug. 1, 2013; revised and accepted May 20 & Nov. 3, 2014)

Abstract

A multi factor authentication scheme called '3C-Auth' is proposed in this paper. The scheme carries out a comprehensive authentication process using the smart card, secret-pin, registered finger print, and registered mobile number of the user. The user's password is neither transmitted in plaintext form nor revealed to the authentication server. The scheme is shone to be proof against phishing, password guessing, replay, or stolen-verifier attacks. Resistance to parallel session and denial of service attacks and the use of QR-Code in preference to SMS for OTP transfer together, make the scheme attractive for operation under peak loads. Integration of the "3C-Auth" into Multi-Layered Filtering (MLF) scheme leads to secure handling of peak loads on the server ensuring concurrency and availability as well. This clearly enhances the QoS in terms of making right admittance to right resources.

Keywords: Authentication, peak load, QR-code, smart-card

1 Introduction

As Internet services become more popular and pervasive, a serious problem that arises is managing the performance of services under intense load. One of the most challenging problems for public Internet is the delivery of performance targets to users given the randomness of Web accesses. Internet has become indispensable for business and more and more people rely on it for their day to day activities; in turn it evolves continuously and is subject to more and more cyber security threats. Analysis of security breaches and other cyber security issues with particular focus on personal privacy and data security have been active research issues over the past two decades. A multifactor authentication scheme named "3C-Auth" is presented in this paper that uses true authentication to protect resources with high security requirements; it expects the user to possess all the tokens (smart-card, secret-pin, registered finger print and registered mobile phone) to prove his/her identity.

Rest of the paper is organized as follows: Relevant research in literature which forms the motivation for the present work is reviewed in Section 2. Sections 3 and 4 detail the proposed scheme and analyze its performance. Integration of the scheme with MLF *(*Multi Layer Filtering) architecture [2, 3] is presented in Section 5 and conclusions are in Section 6.

2 Related Work

2.1 Internet Architecture

The changeover from the academic Internet to a multi-functional business Internet puts much higher requirements on the architectural supports to control and balance the interests of all stake holders (like users, service providers, data owners, *etc.*). Their hopes and expectations for new applications and services demand new architectures that overcome the fundamental limitations of Internet like lack of data identity, lack of methods for reliable processing, real-time dispensation, scaling to deal with flash crowds, and so on. Since its creation, the Internet is driven by a small set of fundamental design principles rather than being based on a proper formal architecture that is created on a white board by a standardization or research group. The architectural principles and design model of the Internet are all about processing, storing, transmitting and controlling data. This trend is bound to escalate in the future, pointing to a clear need for extensions, enhancements, and re-engineering in Internet architecture. While improvements are needed in each dimension, these should be cohesive demanding a holistic approach. The architecture can be generalized to suit different categories of applications by integrating the admittance control policies that provide metric based differentiation and consecutively maximize the profit earned for having serviced a certain class of requests [16]. Research in this area has identified some key approaches to face overload, such as admission control (per request, per session), request scheduling, service differentiation, service degradation, and resource management.

2.2 Current State of Internet Services and Authentication Requirement

The following form the key features of the state of art of internet services:

- Generic nature;
- Accommodation of technological innovations;
- Robustness at times of overload.

As access to more and more services is pushed online, the range of sensitive information that a user must protect widens with time. It is also equally important to understand that complicated security schemes will not achieve widespread adoption among Internet users. Today hackers have the option of using many techniques to steal passwords such as shoulder surfing, snooping, sniffing, guessing, *etc.* – Implying that businesses should use commensurate secure approach. The challenge here is to balance strong security with usability. One Time Passwords (OTPs) for single session/transaction usage have been identified as the best way of protecting online transactions.

2.3 Authentication Schemes

In 1981, Lamport proposed a scheme to authenticate a remote user on a remote server over an insecure network. The requirement for storing verification tables used by the scheme was overcome by a scheme proposed in by [7]. Later [10] proposed a new remote authentication based on ElGamal crypto scheme exploiting tamper resistance property of smart cards. Most of the remote identity based remote authentication schemes proposed by researchers [7, 10, 13, 16] rely primarily on passwords for security. The schemes are vulnerable to dictionary attacks [6]. To overcome this problem random cryptographic secret key could be used [14]. However, such large key values (are difficult to be remembered and hence) require to be stored somewhere. Further these strong passwords and secret keys fail to provide non repudiation. An authentication scheme of Khan et al and Li et al uses biometric keys with advantages like "cannot be lost", "difficult to forge", "cannot be guessed" *etc.* [10] proposed an efficient biometric based smart card authentication scheme. [7] showed that the scheme makes two assumptions to ensure its correctness and security that may restrict its use for real time applications. [8] proposes a generic framework for preserving security in distributed systems. The three factor authentication scheme in [7] is based on password, smart card and biometric characteristics. The authors claim two benefits in the usage of fuzzy extractor:

- Elimination of the assumption of Li-Hwangs scheme that stores the hash of biometric template;
- Use of biometric authentication that supports reasonable tolerance.

In the analysis of their scheme the authors have shown how their protocol is secured against attackers of Type I (smart card and biometric), Type II (password and Biometric), and Type III (smart card and Password). Although the generic construction proposed by Huang et al satisfies the security requirements of three factor authentication the system may fail to secure resources that require very high degree of security the reason being biometric systems that are fast with the false rejection rate under 1% (together with a reasonably low false acceptance rate) are rare even today.

2.4 Security with OTP

Authentication of users in a distributed environment is an increasingly difficult task. As network and software grow in sophistication so do means and methods of malicious attackers. Today computer crackers use enormous resources to obtain information necessary to impersonate other users. Authentication systems based on one time passwords [5] provide more reliability than those based on remembered/stored ones. Hence, security sensitive industries (banks, government *etc.*) deploy one time password systems to reduce the damage of phishing and spyware attacks.

2.4.1 SMS-OTP

Most of the two factor authentication schemes authenticate users based on what they know and what they have, incorporating token-less second factor *(e.g. mobile)*. Each method has a reason to exist based on design criteria for the overall usage.

Online banking is a good example where strong remote authentication is guaranteed using two-factors as de facto standard. In practice, the first factor is usually in the form of PIN or password that the user types (for instance) into a web-based Internet application. The second factor is usually in the form of mobile phone that is known to be able to receive OTP as SMS directed to a particular mobile phone number. If the user successfully retypes this OTP into the web application, the second authentication factor is regarded as successfully verified *(i.e. the user has the mobile phone)*.

Security of the aforesaid scenario relies on the practical difficulty for an attacker to simultaneously compromise the operating environment of both the particular phone and the web browser where the user part of the serving application runs.

2.4.2 Problems with SMS-OTP

The main problems with the SMS-OTP design are under overloaded situations. These are:

- Delay in delivery of SMS;
- Low Coverage Areas;
- Non-availability of Mobile Phone;

- Downtime with SMS Gateway;
- Non-availability of service for roaming user;
- High Cost for roaming user;
- Complexity associated with sequence of operations in obtaining OTP from SMS when mobile phone is used for connecting to the Internet.

2.4.3 Authentication Using QR Code

In 2002, Clarke et al. suggested the usage of camera-based devices as an alternative but more secured authentication method for critical transactions with un-trusted computers. With the explosive growth in the amount of camera-equipped smart phones around us mobile based authentication [11] may become a popular authentication method in the near future. QR-code (a two-dimensional barcode) - as introduced by Japanese company Denso-Wave in 1994 is a more effective alternative. Its error correction capability facilitates data restoration even under conditions when substantial parts of the code are damaged. Modern cellular phones are natively equipped with the QR-code decoding software. Fortunately, for camera phones that are not equipped with QR-code readers, Quick-Mark and i-nigma are free tools that are available for many manufactured models and devices to decode QR-Codes free of cost. Depending on the data recognized and the nature of the application.

2.4.4 Summary of Findings

Internet has become the most important platform for business relations and social interactions. The rapid growth of Internet of Things and Services clearly shows that the ever increasing amount of physical items of our daily life which become addressable through a network could be made more easily manageable and usable through the use of Internet services. This course of exposed resources along with the level of privacy and value of the information they hold, together with increase in their usage, has led to the escalation in the number of security threats and violation attempts that existing systems do not appear robust enough to address. Internet architecture of tomorrow must meet the changing requirements of the Internet, ISPs (Internet Service Providers), Users etc. Perhaps one of the most compelling problems of the modern Internet is the lack of a comprehensive and unifying approach to deal with service concurrency, security, and availability particularly at times of overloads. It is also important to understand that the internet and its users are under continuous attacks *i.e.*, security is the underlying problem for many of the Internet services. One has to clearly understand that the impact of an attack can be major, and can include costly and embarrassing service disruptions, down-time, lost productivity, stolen data, regulatory fines, and irritated customers. Strong authentication has no precise definition; it is not a strictly mathematical concept with quantitative measurements but rather a qualitative measure that is evaluated using a relative scale. The present sophistication level of hackers, demands authentication schemes to be based on more than one factor. Evaluating multi-factor authentication solutions calls for a look into the following measures:

- Security and scalability of the technology;
- Hurdles to user adoption;
- Cost;
- Deployability.

3 Proposed Scheme

The primary goal here is to enhance the performance of Multi-layered Filtering (MLF) scheme and enable real world applications to take advantage of this added functionality.

A scheme that performs admission control with enhanced multi factor authentication “called 3CAuth”, is proposed in this paper and the same evaluated for efficiency. The scheme provides true authentication by expecting the user to possess all the relevant tokens (smart card, secret-pin, registered finger print, and registered mobile phone) to prove his/her identity.

The benefits of the scheme include:

- NOT revealing users password to the server;
- NOT transmitting passwords in plaintext over the Internet, and at the same time;
- RESISTING the major possible attacks like replay attack, password guessing attack, stolen-verifier attack, and phishing attack.

The scheme operates in two phases namely registration and login-authentication. Table 1 is the notations used in the two phases.

3.1 Registration Phase

Figure 1 depicts the activities in the registration process. As shown in Figure 1 it involves the steps/activities in Algorithm 1.

The sequence of operations for registering ONE user is illustrated in Figure 2.

3.2 Login Phase

When U_i wishes to login to server (S), he/she must insert the smart card into a card reader, provide biometric data BF_i' , capture the QR code displayed on the web page, decrypt it using the software installed in the mobile, and present the OTP for authentication purpose. The sequence is shown in block diagram form in Figure 3.

Table 1: Table of notations

Notations	Description
U_i	i^{th} User
ID_i	Unique Identifier of i^{th} User
PWD_i	Password of the i^{th} User
d	Private key in RSA
e	Public key in RSA
n	Computed as product of chosen prime numbers (p and q)
g	Generator element primitive to $GF(p)$ and $GF(q)$
SID_i	Smart card Identifier of i^{th} User
IMEI	International Mobile Station Equipment Identity
IMSI	International mobile subscriber identity
MID_i	Unique Key for mobile of i^{th} user
R_1 and R_2	Random numbers chosen for verification
T_s	Time at which the request is generated
BF_i	Biometric feature of i^{th} user
R_c	Random Challenge (in this context - One Time Password)

Figure 1: Registration process

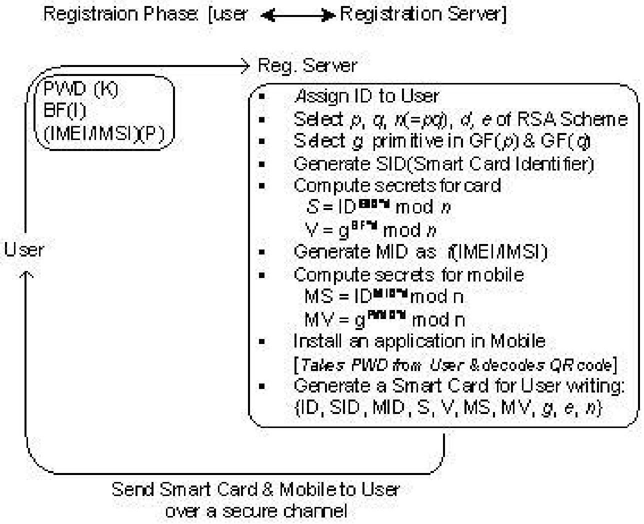

Registraion Phase: [user ⟷ Registration Server]
Reg. Server
PWD (K)
BF(I)
(IMEI/IMSI)(P)
User
Assign ID to User
Select p, q, n(=pq), d, e of RSA Scheme
Select g primitive in GF(p) & GF(q)
Generate SID(Smart Card Identifier)
Compute secrets for card
S = ID mod n
V = g mod n
Generate MID as f(IMEI/IMSI)
Compute secrets for mobile
MS = ID mod n
MV = g mod n
Install an application in Mobile
[Takes PWD from User & decodes QR code]
Generate a Smart Card for User writing:
{ID, SID, MID, S, V, MS, MV, g, e, n}
Send Smart Card & Mobile to User
over a secure channel

Figure 2: Registration process

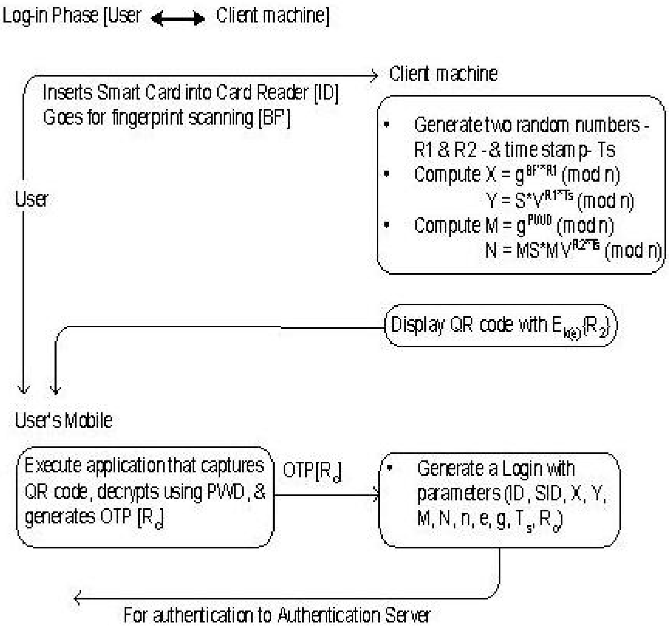

Log-in Phase [User ⟷ Client machine]
Client machine
Inserts Smart Card into Card Reader [ID]
Goes for fingerprint scanning [BF']
User
Generate two random numbers - R1 & R2 - & time stamp- Ts
Compute X = g^{BF'*R1} (mod n)
Y = S*V^{R1*Ts} (mod n)
Compute M = g^{PWD} (mod n)
N = MS*MV^{R2*Ts} (mod n)
Display QR code with E_{k(e)}(R_2)
User's Mobile
Execute application that captures QR code, decrypts using PWD, & generates OTP [R_c]
OTP[R_c]
Generate a Login with parameters (ID, SID, X, Y, M, N, n, e, g, T_s, R_c)
For authentication to Authentication Server

Figure 3: Login phase

Algorithm 1 Steps in registration process

```
Begin
The user $U_i$ chooses a password $PWD_i$ and provides his/her biometric feature $BF_i$.
The registration server sequences through the following further steps:
while More users to Register do
    Assigns an $ID_i$ for the user and generates two large prime numbers $p$ and $q$, and computes
        $n = p * q.$
    For security reasons, the lengths of $p$ and $q$ are recommended to be 512 bits at least.
    Chooses integers $e$ and $d$ which satisfy
        $e * d \bmod ((p-1)*(q-1)) \equiv 1$
    Further it also finds an integer $g$ which is a primitive element in both $GF(p)$ and $GF(q)$.
    Generates a smart card identifier $SID_i$ for the user $U_i$. In addition it generates a mobile phone identifier: $MID_i = (IMEI, IMSI)$ for the user $U_i$.
    Calculates $U_i$'s secret information as
        $S_i \equiv ID_i^{(SID_i*d)} \bmod n$
        $V_i \equiv g^{(d*BF_i)} \bmod n$
        $MS_i \equiv ID_i^{(MID_i*d)} \bmod n$
        $MV_i \equiv g^{(d*PWD_i)} \bmod n.$
    Stores ($ID_i$, $SID_i$, $MID_i$, $S_i$, $V_i$, $MS_i$, $MV_i$, $n$, $e$, $g$) in the smart card, installs an application (for capturing and decoding QR code after obtaining secret pin from the user) in user's mobile and issues the smart card to the user $U_i$ over a secure channel.
end while
End
```

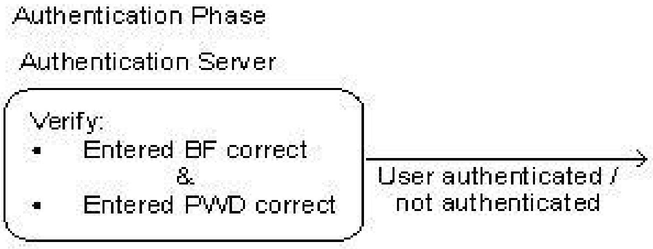


Figure 4: Authentication phase

Algorithm 2 Verification of possession of biometric characteristics and smart card

```
Begin
Check if $ID_i$ is a valid user identity and $SID_i$ is a legal smart card identity; if not reject the login request.
Check if $T_s$ is within the legal time interval limit due to transmission delay (may be initialized in SLA-service level agreement); if not reject the login request.
Verify if $Yi^e \overset{?}{\equiv} (ID_i^{(SID_i)} * Xi^{Ts} \bmod n)$;
The above equation holds iff $BF_i = BF_i'$.
i.e. the correct biometric value is provided during login phase. That is because
    $Yi^e \equiv (S_i * V_i^{(R1*Ts)} \bmod n)^e$
          $\equiv ID_i^{(SID_i*d*e)} * g^{(d*BF_i*e*R1)} \bmod n.$
          (Since $d * e \equiv 1 \bmod n$ we have)
          $\equiv ID_i^{(SID_i)} * g^{(BF_i*R1)} \bmod n$
and
    $ID_i^{(SID_i)} * (Xi^{Ts}) \equiv ID_i^{(SID_i)} * g^{(BF'i*R1)} \bmod n;$
End
```

3.3 Authentication Phase

The authentication phase (Figure 4) is executed by the remote host to determine whether U_i is allowed to login or not. The steps in login process are shown in Figure 3.

The authentication server upon receiving the login request from the user verifies the possession of smart card, biometric feature, and mobile phone as described in Algorithms 2 and 3.

4 Strengths of 3C-Auth

Resistance of the proposed method to different possible security attacks is explained here.

4.1 Parallel Session Attack

Here an attacker impersonates a legitimate user by intercepting the login request (ID_i, SID_i, S_i, V_i, M_i, N_i, n, e, g, T_s) and attempting to modify it to succeed in authentication. However the attacker has no way of obtaining the Biometric feature BF_i, PWD_i, and the random numbers R_1 and R_2; hence he/she cannot compute M_i, N_i, X_i, and Y_i which are dependent on PWD_i and R_1; a valid request cannot be created and the attempt fails. Hence it follows that the proposed scheme is secured against this type of attack.

4.2 Password Guessing Attack

The attacker attempts to guess user's secret parameters here. Although, one can extract parameters (n, e, g, S_i, V_i, SM_i, VM_i) from the user's smart card, obtaining BF_i or PWD_i from the smart card without the knowledge of d from $g^{(d*BF_i)}$ and $g^{(d*BF_i)}$, is not possible. Thus the difficulty of obtaining the discrete logarithm secures the scheme from password guessing attack even under stolen smart card situations.

Algorithm 3 Verification of possession of secret-pin and mobile phone

```
Begin
Confirm if $ID_i$ is a valid user identity and $MID_i$ is a legal mobile phone identity; if not reject the login request.
Confirm if $T_s$ is within the legal time interval limit due to transmission delay (may be initialized in SLA-service level agreement) , if not, reject the login request.
Check whether the following equation holds: $Ni^e = ID_i^{(MID_i)} * Mi^{(Rc*Ts)} \bmod n$
    The equation here holds iff $R_2$ = Rc i.e the correct OTP is provided by the user during login phase. The correct OTP can be obtained only in the mobile on which the application software is installed during registration phase and only if the password provided to it is correct . This is because
    $Ni^e \equiv (MS_i * MV_i^{(R2*Ts)})^e \bmod n$
    $\equiv ID_i^{(MID_i*d*e)} * g^{(d*PWD_i*e*R2*Ts)} \bmod n$
    $\equiv ID_i^{(MID_i)} * g^{(PWD_i*R2*Ts)} \bmod n$
    and
    $ID_i^{(MID_i)} * Mi^{(Rc*Ts)} \bmod n$
    $\equiv ID_i^{(MID_i)} * g^{(PWD_i*Rc*Ts)} \bmod n;$
    If the login request is rejected three times the user account is locked. He/She has to contact registration server to unlock the account.
End
```

4.3 Resistance to Replay Attack

Intercepting the login request message (ID_i, SID_i, S_i, V_i, M_i, N_i, n, e, T_s) of a user U_i and replaying the same message to the server becomes useless because the card reader puts a new timestamp in each new login request. The equations

$$\begin{aligned} Y_i^e &\equiv ID_i^{(CID_i*X_i)} \bmod n \quad \text{and} \\ N_i^e &\equiv ID_i^{(MID_i)} * M_i^{r_1} \bmod n \end{aligned}$$

will fail during the authentication phase.

4.4 Denial of Service Attack

In the proposed scheme an adversary can use invalid ID, PWD and BFs and overload the server by continuously keeping it busy. Even though an initial filtering for this type of attacks takes place in Stage I of MLF architecture non-legitimate requests that pass Stage I of MLF are blocked by the proposed scheme. This is obvious from the fact that a valid login request cannot be created (as discussed in Section 4.1). Further after three unsuccessful attempts the scheme automatically locks the user's account; the same can be unlocked only with the help of registration server.

4.5 Resistance to Phishing Attacks

The aim of phishing is mainly to collect private information that can be used to impersonate victims. A possible reading of the QR code (and extracting the OTP) by a hacker yields only the encrypted value of R_1; even if he manages to access the data typed by user, the private key remains inaccessible thanks to the strength of the RSA scheme. Thus the phishing attempt fails.

5 Integration of 3C-Auth with MLF

MLF is a practical (secured-concurrent-available) end-to-end framework based on admission control policies - a strategy that achieves robust performance on a wide range of Internet services subject to huge variation in load. MPAC (Multi Phase Admission Control) [3] enhances MLF to maximize the reward earned for having serviced a particular class of requests. 3C-Auth scheme described here can be integrated with this enhanced MLF framework to make it more comprehensive by adding security assurance. The integration involves two steps namely:

1) Enhancing SLA to include new features specific to authentication;

2) Modifying admission control policy to support 3C-Auth.

The 3C-Auth process is to be inserted at Stage II of the MLF framework at the Access Node component. Request processing in Stage II of the comprehensive MLF scheme is illustrated in Figure 5.

5.1 Service Level Agreement

The SLA that spells out the scope of service providers' allotment to the e-commerce in terms of resource capacity and time commitment used in MPAC is shown as follows:

Contract: Ecommerce System
{
Service : Classification of Customers
Customer Class = {Premium, Ordinary, New}
Inter-session States = {Home, Browse, Item, Addcart, BuyReq, BuyConfirm}
} { Service : Processing Requests (Peakload)
{
Availability >$minAvailability;
TimeBound <$maxDelay;
Throughput >$minThroughput;
Utilization <$maxUtilization;
WeightAdjustment(Forward) = $weight(Positive)

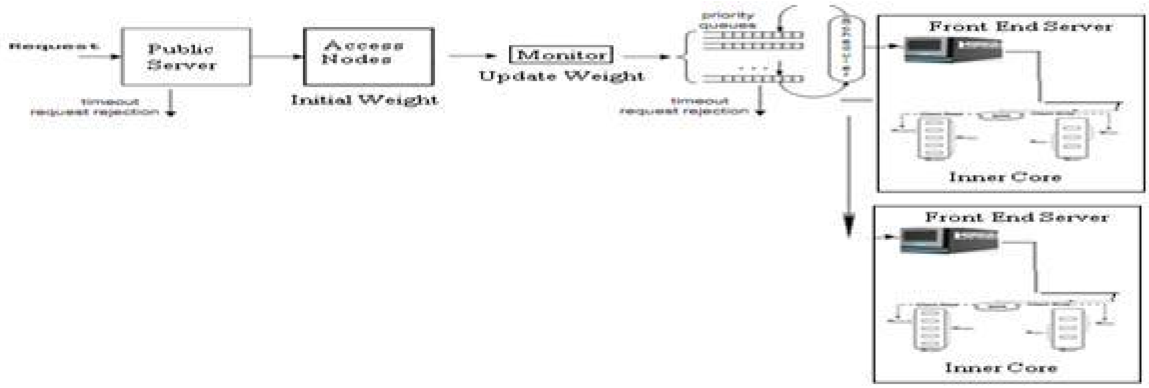


Figure 5: Request processing in Stage II of enhanced MLF (MLF+MPAC)

```
WeightAdjsutment(Backward) = $weight(zero,negative)
}
```

The template can be modified to support 3C-Auth by including the following specifications:

```
Security: Authservice
{
Service : Servicename /* Generic */
Resource Class = {Unclassified, secret,topsecret}
Factors Class = {Password/PIN, Hardwaretoken, Bio-Hard, Softtoken}
Protocol Class {None, 2CAuth,3C-Auth}
{ Service : News /* instance*/
Resource Class Unclassified
Registration{ }
Factors Class{}
Protocol Class {None}
} Service : Internet Banking
{ /* Announcements regarding new schemes for loans
{
Resources Class {Unclassified}
Registration { }
Factors Class { }
Protocol Class { }
} // Online banking
{
Resources Class {topsecret}
Registration {REQUIRED }
Factors Class {PIN, Hardwaretoken (Smart card), Softtoken}
Protocol Class {2CAuth[3]}
}
}
}
}
```

5.2 Admission Control Policy

MPAC uses a reward function defined by the application/service provider to improve the QoS using service differentiation. It computes the Expected Reward and the Cost Incurred in servicing the request and uses them as basic parameters to prioritize customers for E-commerce applications. The scheme can be made more comprehensive by re-computing priorities with authentication factors as well, for better security assurances. This is achieved in the comprehensive MLF framework (MLF + MPAC + 3C-Auth) as follows:

- Resources added to the pool are tagged with weights (based on SLA) that specify the number of factors required to access it;
- Incoming requests from Stage I are directed to the Access nodes by the public server for authentication;
- The access nodes proceed to authenticate users by assigning an initial weight of 0 to each request and updating it as per the credentials (number of factors) validated;
- Possessed weights are compared with the tagged weights associated with resources and in case of match in their weights access to the resource is permitted.

5.3 Results of Integration

The performance analysis demonstrated that the presented scheme performs a comprehensive authentication process satisfying the important requirements including friendliness, resistance to various kinds of sophisticated attacks, and stolen credentials. Further resistance offered by the scheme to parallel session attack and denial of service attack made the scheme more suitable for operation under peak loads. QR-Code based OTP has been found to show improved performance at peak load times compared to SMS-OTP method. With instantaneous SMS delivery the performance was on par with that of SMS-OTP scheme. The vulnerability associated with the in-absentia verification of the user is effectively handled by the scheme. Moreover, the scheme was found to be more user-friendly without sacrificing security assurances. With all these benefits contribution of the scheme towards improvement in QoS in terms of granting right access to resources can be considered significant.

6 Discussions and Conclusions

The proposed protocol is simple, fast and efficient if the user provides valid credentials for authentication. A detailed analysis of the proposed scheme has clearly brought out its advantages over authorization methods that use SMS to thwart attacks. Moreover, the scheme aptly fits at the access nodes in the enhanced MLF architecture making it more user-friendly without sacrificing security assurances. Improving in computational efficiency of the scheme is an interesting area of work; it can add substantially to its effectiveness.

References

[1] T. H. Feng, C. H. Ling, and M. S. Hwang, "Cryptanalysis of Tan's improvement on a password authentication scheme for multi-server environments", *International Journal of Network Security*, vol. 16, no. 4, pp. 318–321, 2014.

[2] N. Harini and T. R. Padmanabhan, "A secured-concurrent-available architecture for improving performance of web servers", in *Proceedings of 6th International Conference on Information Processing (ICIP 2012)*, pp 621-631, Bangalore, India, Aug. 10-12, 2012.

[3] N. Harini and T. R. Padmanabhan, "Admission control and request scheduling for secured-concurrent-available Architecture", *International Journal of Computer Applications*, vol. 63, no. 6, pp. 24–30, 2013.

[4] N. Harini and T. R. Padmanabhan, "2CAuth: A new two factor authentication scheme using QR-code", *International Journal of Engineering and Technology (IJET)*, vol. 5, no. 2, pp. 1087–1094, 2013.

[5] N. Harini, T. R. Padmanabhan and C. K. Shyamala, *Cryptography and Security*, Wiley India, First Edition, 2011.

[6] D. He, W. Zhao, and S. Wu, "Security analysis of a dynamic ID-based authentication scheme for multi-server environment using smart cards", *International Journal of Network Security*, vol. 15, no. 5, pp. 350–356, 2013.

[7] C. H. Huang, J. S. Chou, Y. Chen, "Improved multi-server authentication protocol", *International journal of Security and Communication Networks*, vol. 5, no. 3, pp. 331–341, 2012.

[8] X. Huang, Y. Xiang, A. Chonka, J. Zhou, and R. H. Deng, "A generic framework for three-factor authentication: Preserving security and privacy in distributed systems", *IEEE Transactions on Parallel and Distributed Systems*, vol. 22, no. 8, pp. 1390–1397, 2011.

[9] Qi Jiang, J. Ma, G. Li, and Li Yang, "Robust two-factor authentication and key agreement preserving user privacy", *International Journal of Network Security*, vol. 16, no. 3, pp. 229–240, 2014.

[10] C. T. Li and M. S. Hwang, "An efficient biometrics-based remote user authentication scheme using smart cards", *Journal of Network and Computer Applications*, vol. 33, no. 1, pp. 1–5, 2010.

[11] K. C. Liao and W. H. Lee, "A novel user authentication scheme based on QR-code", *Journal of Networks*, vol. 5, no. 8, pp. 937–941, 2010.

[12] J. J. Shen and P. W. Hsu, "A fragile associative watermarking on 2D barcode for data authentication", *International Journal of Network Security*, vol. 7, no. 3, pp. 301–309, 2008.

[13] J. J. Shen, C. W. Lin and M. S. Hwang, "Security enhancement for the timestamp-based password authentication", *Computers and Security*, vol. 22, no. 7, pp. 591–595, 2003.

[14] H. Tang, X. Liu, L. Jiang, "A robust and efficient timestamp-based remote user authentication scheme with smart card lost attack resistance", *International Journal of Network Security*, vol. 15, no. 6, pp. 446–454, 2013.

[15] A. Totok, V. Karamcheti, "RDRP: Reward-driven request prioritization for e-Commerce Web sites", *Electronic Commerce Research and Applications*, vol. 9, pp. 549–561, 2010.

[16] Li Yang, J. F. Ma, and Qi Jiang, "Mutual authentication scheme with smart cards and password under trusted computing", *International Journal of Network Security*, vol. 14, no. 3, pp. 156–163, 2012.

[17] X. Zhuang, C. C. Chang, Z. H. Wang, Y. Zhu, "A simple password authentication scheme based on geometric hashing function", *International Journal of Network Security*, vol. 16, no. 4, pp. 271–277, 2014.

N. Harini is an Assistant Professor in the department of Computer Science and Engineering, Amrita Vishwa Vidyapeetham. She has 3 years of industrial and 15 years of teaching and research experience. She is currently pursuing her Ph D in Security. Her research interests include cryptography, security. She has currently co-authored a book on Cryptography and Security.

T. R. Padmanabhan with MTech and PhD at the IIT Kharagpur, was in the faculty there from 1964 to 1979. With 20 years of development experience in the industry and an equal period in academic institutions, he is currently a Professor Emeritus at the Amrita School of Engineering. His research interests are security, digital communication, and VLSI design. He has (co)authored five books.

The Development of Deniable Authentication Protocol Based on The Bivariate Function Hard Problem

Normahirah Nek Abd Rahman[1] and Muhammad Rezal Kamel Ariffin[1,2]

(Corresponding author: Normahirah Nek Abd Rahman)

Al-Kindi Cryptography Research Laboratory, Institute for Mathematical Research, Universiti Putra Malaysia[1]
Department of Mathematics, Faculty of Science, Universiti Putra Malaysia[2]
43400 UPM Serdang, Selangor, Malaysia.
(Email: mahirah_mayrah@yahoo.com, rezal@upm.edu.my)

(Received Jan. 28, 2015; revised and accepted Apr. 25 & June 30, 2015)

Abstract

A deniable authentication protocol enables a receiver to identify the true source of a given message but not to prove the identity of the sender to the third party. Non-interactive protocol is more efficient than interactive protocol in terms of communication overhead, and thus several non-interactive deniable authentication protocols have been proposed. So, it is very necessary to design a deniable authentication protocol which is non-interactive, secure and efficient. This paper proposes a deniable authentication protocol based on the bivariate function hard problem (BFHP) cryptographic primitive. An improvement based on the BFHP is suggested since the problem of the BFHP provides the needed security elements plus its fast execution time. At the same time, the proposed protocol has properties of completeness, deniability, security of forgery attack, security of impersonation attack and security man-in-the-middle attack also has been proved.

Keywords: Bivariate function hard problem, deniable authentication protocol, non-interactive protocol

1 Introduction

Deniability is a privacy property that ensures protocol participants can later deny taking part in a particular protocol run while authentication is used to ensure that users are who they say they are. So, a deniable authentication protocol is a protocol that enables a receiver to identify the true source of a given message, but not to prove the identity of the sender to a third party. There are many interactive and non-interactive deniable authentication protocols have been proposed. However, the interactive manner makes deniable protocols inefficient.

Deniable authentication has two characteristics that differ from traditional authentication. The first one is only the intended receiver can identify the true source of a given message (i.e. able to identify the signature of the sender) and the second one is the receiver cannot prove the source of the message to a third party (i.e. unable to prove the signature of the sender to a third party that the signature belongs to the sender). In other words, once the receiver has obtained and authenticated the message from the sender, the receiver cannot impersonate as the sender to a third party. Because of these two characteristics, the deniable authentication protocol is very useful for providing secure negotiation over internet.

For example, suppose that a customer wants to order an item from a merchant, so the customer should make an offer to the merchant and create an authenticator for the offer because the merchant must be sure that this offer really comes from the customer. However, the merchant wants to be able to prevent the customer from showing this offer to another party in order to elicit a better deal. Therefore, we need a protocol that enables a receiver to identify the source of a given message, but prevents a third party from learning the sender's identity.

In 1998, Dwork et al. [4] proposed an interactive deniable authentication protocol based on concurrent zero knowledge proof while Aumann and Rabin [2] proposed an interactive deniable authentication protocol based on the integer factorization problem (IFP). Later, Deng et al. (2001) [3] introduced two interactive deniable authentication protocols based on the discrete logarithm problem (DLP) and IFP respectively. In 2002, Fan et al. [5] introduced another simple interactive deniable authentication protocol based on Diffie-Hellman Key Distribution Protocol. However, there is a common weakness in the four previous protocols which the sender does not know to whom he proves the source of a given message. That is, a third party can impersonate the intended receiver to identify the source of a given message. Meanwhile, these four protocols are interactive and less efficient.

This scenario has led many cryptographers to come up with non-interactive deniable authentication protocol in order to enhance the efficiency. Shao (2004) [12] proposed a non-interactive deniable authentication protocol based on generalized ElGamal signature scheme. Lu and Cao (2005) [10, 11] proposed two deniable authentication protocols based on bilinear pairing and IFP respectively but their protocol is still unable to achieve the second characteristic of being a deniable authentication protocol.

Later, in 2008, Hwang and Ma [8] proposed deniable authentication protocol with anonymous sender protection. The sender's anonymity is also used to protect the sender's privacy. Though the sent message is forgeable by the receiver, but the sender can provide evidence to prove the message was really sent by him. Hence, to reduce the computational cost of proposed protocols with anonymous sender protection, Hwang and Chao (2010) [7] proposed a new deniable authentication protocol with anonymous sender protection in an efficient way based on Schnorr signature scheme.

Then, Zhang et al. (2011) [13] proposed a new non-interactive deniable authentication protocol based on generalized ElGamal signature scheme, which is more efficient than the previous two protocols (Shao 2004, Lee et al. 2007) [9, 12] both in computation and communication. To authenticate the source of a message, although the proposed protocol needs one more modular exponentiation than Shao's protocol, but as to the length of the communicated messages, just 2|h| are required to be transmitted compared to 3|h| in Shao's protocol. Lee et al.'s protocol needs five exponentiation computations altogether compared to proposed protocol which needs only four. The transmitted bits of the proposed protocol are reduced to 320 bits compared to Lee et al.'s protocol which is 1184 ∼ 2208 bits.

In this paper, we propose a new non-interactive deniable authentication protocol based on the Bivariate Function Hard Problem (BFHP) (Ariffin et al. 2013) [1]. We prove our protocol is secure against forgery attack, impersonation attack and man-in-the-middle attack and prove the properties of completeness and deniability of this protocol. With its guaranteed security, we also show that the performance of the protocol requires reasonable numbers of operation in both sign and verify phases.

The layout of the paper is as follows. In Section 2, we will first review the definition of the BFHP. Proof will be given on the uniqueness and intractability of the BFHP. We will also review in this section, deniable authentication protocol in the standard model. In Section 3, we propose the standard model of the deniable authentication protocol followed by the security analysis in which proof is given. In Section 4, we provide efficiency analysis and comparison of the protocol. In Section 5, the conclusion about our deniable authentication protocol is made.

2 Preliminaries

2.1 Linear Diophantine Equations with Infinitely Many Solutions

Definition 1. *The successful process of prf-solving a Diophantine equation which has infinitely many solutions is the process of determining a preferred solution from a set of infinitely many solutions for the Diophantine equation.*

To further understand and obtain the intuition of Definition 1, we will now observe a remark by Herrmann and May (2008) [6]. It discusses the ability to retrieve variables from a given linear Diophantine equation. But before that we will put forward a famous theorem of Minkowski that relates the length of the shortest vector in a lattice to the determinant.

Theorem 1. *In an ω-dimensional lattice, there is exists a non-zero vector with*

$$\|v\| \leq \sqrt{\omega}\det(L)^{\frac{1}{\omega}}$$

We now put forward the remark.

Remark 1. There is a method for finding small roots of linear modular equations $a_1x_1+a_2x_2+...+a_nx_n \equiv 0 \pmod N$ with known modulus N. It is further assumed that $\gcd(a_i, N) = 1$. Let X_i be upper bound on $|x_i|$. The approach to solve linear modular equation requires to solve the shortest vector in a certain lattice. We assume that there is only one linear independent vector that fulfills Minkowski bound (Theorem 1) for the shortest vector. Herrmann and May (2008) [6] showed that under heuristic assumption that the shortest vector yields the unique vector $(y_1, ..., y_n)$ whenever

$$\prod_{i=1}^{n} X_i \leq N.$$

If in turn we have

$$\prod_{i=1}^{n} X_i \geq N^{1+\epsilon}.$$

Then the linear equation usually has N^ϵ many solutions, which is exponential in the bit-size of N. So, there is no hope to find efficient algorithms that in general improve on this bound, since one cannot even output all roots in polynomial time. We now put forward a corollary.

Corollary 1. *A linear Diophantine equation*

$$\begin{aligned} f(x_1, x_2, ..., x_n) &= a_1x_1 + a_2x_2 + ... + a_nx_n \\ &= N \end{aligned}$$

with

$$\prod_{i=1}^{n} x_i \geq N^{1+\epsilon}$$

is able to ensure secrecy of the preferred sequence $x = \{x_i\}$.

Remark 2. In fact if one were to try to solve the linear Diophantine equation $N = a_1x_1 + a_2x_2 + ... + a_nx_n$, where

$$\prod_{i=1}^{n} x_i \geq N^{1+\epsilon}$$

any method will first output a short vector x= $\{x_i\}$ as the initial solution. Then there will be infinitely many values from this initial condition that is able to reconstruct N.

2.2 Bivariate Function Hard Problem

The following proposition gives a proper analytical description of the Bivariate Function Hard Problem (BFHP).

Definition 2. *We define $\mathbb{Z}^+_{(2^{m-1},2^m-1)}$ as a set of positive integers in the interval as a set of positive integers in the interval $(2^{m-1}, 2^m - 1)$. In other words, if $x \in \mathbb{Z}^+_{(2^{m-1},2^m-1)}$, then x is a m-bit positive integer.*

Proposition 1. (Ariffin et.al (2013)) Let $F(x_1, x_2, ..., x_n)$ be a multiplicative one-way function that maps $F : \mathbb{Z}^n \to \mathbb{Z}^+_{(2^{m-1},2^m-1)}$. Let F_1 and F_2 be such function (either identical or non-identical) such that $A_1 = F(x_1, x_2, ..., x_n), A_2 = F(y_1, y_2, ..., y_n)$ and $\gcd(A_1, A_2) = 1$. Let $u, v \in \mathbb{Z}^+_{(2^{n-1},2^n-1)}$. Let (A_1, A_2) be public parameters and (u, v) be private parameters. Let

$$G(u, v) = A_1u + A_2v \tag{1}$$

with the domain of the function G is $\mathbb{Z}^2_{(2^{n-1},2^n-1)}$ since the pair of positive integers $(u, v) \in \mathbb{Z}^2_{(2^{n-1},2^n-1)}$ and $\mathbb{Z}^+_{(2^{m+n-1},2^{m+n}-1)}$ is the codomain of G since $A_1u + A_2v \in \mathbb{Z}^+_{(2^{m+n-1},2^{m+n}-1)}$.

If at minimum $n - m - 1$, where (n, m) is chosen such that the value k results in 2^k to be accepted as exponentially large for any probabilistic polynomial time (PPT) adversary to sieve through all possible answers, it is infeasible to determine (u, v) over $\mathbb{Z}$ from $G(u, v)$. Furthermore (u, v) is unique for $G(u, v)$ with high probability.

Remark 3. We remark that the preferred pair (u, v) in $\mathbb{Z}$, is *prf*-solution for Equation (1). The preferred pair (u, v) is one of the possible solutions for Equation (1) given by

$$u = u_0 + A_2t \tag{2}$$

and

$$v = v_0 - A_1t \tag{3}$$

for any $t \in \mathbb{Z}$.

Remark 4. Before we proceed with the proof, we remark here that the Diophantine equation given by $G(u, v)$ is solved when the preferred parameters (u, v) over $\mathbb{Z}$ are found. That is the BFHP is *prf*-solved when the preferred parameters (u, v) over $\mathbb{Z}$ are found.

Proof. We begin by proving that (u, v) is unique for each $G(u, v)$ with high probability. Let $u_1 \neq u_2$ and $v_1 \neq v_2$ such that

$$A_1u_1 + A_2v_1 = A_1u_2 + A_2v_2 \tag{4}$$

We will then have

$$Y = v_2 - v_1 = \frac{A_1(u_1 - u_2)}{A_2}$$

Since $\gcd(A_1, A_2) = 1$ and $A_2 \approx 2^m$, the probability that $Y = v_1 - v_2$ is an integer solution not equal to zero is 2^{-m}. Thus, we have $v_1 = v_2$ with probability $1 - \frac{1}{2^m}$. (i.e. $1 - \frac{1}{2^m}$ is the probability that A_2 divides $u_1 - u_2$).

Next we proceed to prove that to *prf*-solve the Diophantine equation given by Equation (1) is infeasible to be *prf*-solved. From the general solution for $G(u, v)$ is given by Equation (2) and Equation (3) for some integer t to find u within the stipulated interval $u \in (2^{n-1}, 2^n - 1)$ we have to find the integer t such that the inequality $2^{n-1} < u < 2^n - 1$ holds. This gives

$$\frac{2^{n-1} - u_0}{A_2} < t < \frac{2^n - 1 - u_0}{A_2}.$$

Then, the difference between the upper and the lower bound is

$$\frac{2^n - 1 - 2^{n-1}}{A_2} = \frac{2^{n-1} - 1}{A_2} \approx \frac{2^{n-2}}{2^m} = 2^{n-m-2}.$$

Since $n - m - 1 = k$ where 2^k is exponentially large for any probabilistic polynomial time (PPT) adversary to sieve through all possible answers, we conclude that the difference is very large and finding the correct t is infeasible. This is also the same scenario for v. □

Example 1. *Let $A_1 = 191$ and $A_2 = 229$. Let $u = 41234$ and $v = 52167$. Then $G = 19821937$. Here we take $m = 8$ and $n = 16$. We now construct the parametric solution for this BFHP. The initial points are $u_0 = 118931622$ and $v_0 = -99109685$. The parametric general solution are $u = u_0 + A_2t$ and $v = v_0 - A_1t$. There are approximately $286 \approx 2^9$ (i.e. $\frac{2^{16}}{229}$) values of t to try (i.e. difference between upper and lower bound), while at minimum the value is $t \approx 2^{16}$. In fact, the correct value is $t = 519172 \approx 2^{19}$.*

Case 1. For $(t', v' \notin \mathbb{Z})$, we can find the value of t' which $u' = u_0 + A_2t'$ such that $u \approx 2^n$. Let $u' = 43571 \approx 2^8$. Then $t' = 519161.7948$ and the value of $v' \notin \mathbb{Z}$ since $v' = 50217.79913$ which clearly results v will not be integer if u is not the *prf*-solution.

Case 2. For $(t', v' \in \mathbb{Z})$, we will obtain $v' \in \mathbb{Z}$ with probability $\frac{1}{2^m}$ which $u' = u_0 + A_2t'$ such that $t' = t_0$ and $u \approx 2^n$. Let $u' = 44211 \approx 2^8$. Then the value of $v' \in \mathbb{Z}$ since $v' = 49684$ for $t' = 519159$. Even we get $(t', u', v' \in \mathbb{Z})$ but $u' \neq u$ and $v' \neq v$. In fact there are 2^{19} choices in the example.

2.3 Deniable Authentication Protocol in Standard Model

A deniable authentication protocol in standard model consists of four phases (Setup, Key Generation, Signing, Verifying) which are defined as follows:

1) **Setup:** The authority determines the parameters that can be used by sender and the receiver to generate their private and public key.

2) **Key Generation:** An algorithm that generates private and public key. The private key which is randomly chosen and remain secret, to be used to generate the public key that will be published in public.

3) **Signing:** An algorithm that generates message authentication code (MAC) from the original message which involves hash function.

4) **Verifying:** An algorithm that involves verification of the new MAC generated with the MAC that has been sent by the sender. If both hold, the original message is authentic and has not been altered.

3 The Standard Model of Deniable Authentication Protocol Based on the BFHP

3.1 Proposed Deniable Authentication Protocol

Setup. The authority randomly chooses the following public parameters:

1) p is a large prime number of n-bit size.
2) g is a primitive root in $\mathbb{Z}_p$.
3) $H(\cdot)$ is a collision free hash function with an output is n bits.

Key Generation. When a user wishes to join the system, he chooses a random number $t \in \mathbb{Z}_p$ as his private key and compute $v = g^t (\text{mod } p)$ as his public key. The public key of each user is certificated by certification authority. The sender, S chooses his secret key $t_s \in \mathbb{Z}^+_{(2^{2n-1}, 2^{2n}-1)}$ and computes $v_s = g^{t_s} (\text{mod } p)$ as his public key. The reason why t_s is chosen out of $\mathbb{Z}_p$ can be observe in step 2(i) of signing phase in order for BFHP to hold.

The receiver, R chooses his secret key $t_R \in \mathbb{Z}_p$ and computes $v_R = g^{t_R} (\text{mod } p)$ as his public key.

Signing. When S wants to deniably authenticate a message M to the intended receiver R, he computes the following protocol:

1) Chooses randomly value $\alpha \in \mathbb{Z}^+_{(2^{2n-1}, 2^{2n}-1)}$.
2) Computes
   a. $\sigma = H_1(M)t_s + H_2(M)\alpha$;
   b. $k_1 = (v_R)^{-H_1(M){t_s}^2} (\bmod\, p)$;
   c. $k_0 = (v_R)^{\alpha H_2(M) t_s} (\bmod\, p)$;
   d. $MAC = H(k_0 \| M)$.

Then, S sends (k_1, σ, MAC) together with message M to R.

Verifying. After receiving (k_1, σ, MAC) together with message M from S, receiver, R computes

1) ${k_1}^* = (v_s)^{\sigma t_R}$;
2) ${k_0}' = k_1 \cdot {k_1}^*$;
3) $MAC = H({k_0}' \| M)$.

R verifies whether $H(k_0 \| M) = H({k_0}' \| M)$. If two equations hold, R accepts the received information. Otherwise, R rejects it. Note that $\|$ is the concatenate operator of strings.

Proposition 2. (Completeness) *If the sender and the receiver follow the protocol, the receiver is able to calculate ${k_0}'$ and then identify the source of the message.*

Proof. From the proposed protocol, we have

$$\begin{aligned} {k_0}' &= k_1 \cdot k_1^* \\ &= (v_R)^{-H_1(M){t_s}^2} \cdot (v_s)^{\sigma t_R} (\bmod\, p) \\ &= g^{-t_R H_1(M){t_s}^2} \cdot g^{{t_s}^2 H_1(M) t_R} \cdot g^{H_2(M) t_s \alpha t_R} (\bmod\, p) \\ &= g^{H_2(M) t_s \alpha t_R} (\bmod\, p) \\ &= (v_R)^{H_2(M) t_s \alpha} (\bmod\, p) \\ &= k_0 \end{aligned}$$

So, $H({k_0}' \| M) = H(k_0 \| M)$. □

Example 2. *The authority randomly chooses $p = 137$ and $g = 101$ as a primitive root in $\mathbb{Z}_p$. The sender, S chooses his secret key $t_s = 781$ and computes $v_s = 118$ as his public key. The receiver, R chooses his secret key $t_R = 157$ and computes $v_R = 11$.*

When S wants to deniably authenticate a message $M = 888$ to the intended receiver R, he chooses randomly value of $\alpha = 813$. He computes $\sigma = 35228$ since $H_1(M) = 17$ and $H_2(M) = 27$. Then, he computes $k_1 = 37$ and $k_0 = 36$. Next, he generates MAC by applying the hash function to the concatenation between M and k_0. He gets $MAC = 1dfc4f553a94cfbf96633b16b2b6e1b5$. Then, S sends (k_1, σ, MAC) together with message M to R.

After receiving (k_1, σ, MAC) together with message M from S, receiver, R computes ${k_1}^* = 38$, ${k_0}' = 36$ and $MAC = 1dfc4f553a94cfbf96633b16b2b6e1b5$ R verifies whether $H(k_0 \| M) = H({k_0}' \| M')$. If two equations hold, R accepts the received information. Otherwise, R rejects it.

3.2 Security Analysis of Deniable Authentication Protocol

Proposition 3. *The proposed protocol is deniable.*

Proof. If the receiver can simulate all the transmitted information between him and the sender, then he cannot prove to any third party where the message is from because the third party cannot identify whether the message is from the sender or is forged by receiver himself.

So, if the receiver tells a third party that the data is from the sender, then the sender can deny it and claims that the receiver himself forge the data. Hence the third party cannot identify who tells the truth.

After receiving (k_1, σ, MAC), the receiver can identify the source of the (k_1, σ, MAC) with his own private key, t_R. However, he cannot prove the source of the message to any party because the receiver can calculate k_0, so he can select any other message M' and construct $MAC' = H({k_0}'\|M')$ and tells the third party (k_1, σ, MAC') is the information he gets from S.

Without the randomly selected $\alpha \in \mathbb{Z}^+_{(2^{2n-1}, 2^{2n}-1)}$, the secret key t_s of S and secret key t_R of R, the third party cannot derive k_0 and k'_0. So he cannot prove whether the receiver is telling the truth. □

Proposition 4. *If the attacker cannot personate as the sender by using another pair of $(\alpha', {t_s}')$ in order to communicate with the intended receiver, then the proposed protocol can withstand forgery attack.*

Proof. The attacker chooses his secret key ${t_s}' \in \mathbb{Z}^+_{(2^{2n-1}, 2^{2n}-1)}$. When attacker wants to deniably authenticate a message M' to the intended receiver R, he computes as follows:

1) Chooses randomly value $\alpha' \in \mathbb{Z}^+_{(2^{2n-1}, 2^{2n}-1)}$.

2) Computes

   a. $\sigma = H_1(M){t_s}' + H_2(M)\alpha'$;

   b. $\overline{k_1} = (v_R)^{-H_1(M)({t_s}')^2} (\bmod\, p)$;

   c. $\overline{k_0} = (v_R)^{\alpha' H_2(M)({t_s}')} (\bmod\, p)$;

   d. $MAC = H(\overline{k_0}\|M')$.

Then, attacker sends $(\overline{k_1}, \sigma, MAC)$ together with message M' to R. After receiving $(\overline{k_1}, \sigma, MAC)$ together with message M' from attacker, receiver, R computes

1) ${k_1}^* = (v_s)^{\sigma t_R}$;

2) ${k_0}' = \overline{k_1} \cdot {k_1}^*$;

3) $MAC = H({k_0}'\|M')$.

Hence, $H(\overline{k_0}\|M') \neq H({k_0}'\|M')$. The message authentication code, $H(\overline{k_0}\|M') \neq H({k_0}'\|M')$ since $\overline{k_0} \neq {k_0}'$ and the receiver always uses the sender?s public key v_s to calculate ${k_1}^*$ and identify the source of the message as follows:

$$
\begin{aligned}
{k_0}' &= k_1 \cdot k_1^* \\
&= (v_R)^{-H_1(M)({t_s}')^2} \cdot (v_s)^{\sigma t_R} (\bmod\, p) \\
&= g^{-t_R H_1(M)({t_s}')^2} \cdot g^{t_s(H_1(M){t_s}' + H_2(M)\alpha')t_R} (\bmod\, p) \\
&= g^{-t_R H_1(M)({t_s}')^2} \cdot g^{({t_s}')t_s H_1(M)t_R} \cdot g^{H_2(M)t_s\alpha' t_R} (\bmod\, p) \\
&= g^{-t_R H_1(M)({t_s}')^2} \cdot g^{({t_s}')t_s H_1(M)t_R} \cdot (v_R)^{H_2(M)t_s\alpha'} (\bmod\, p) \\
&\neq \overline{k_0}
\end{aligned}
$$

The session secret key $k_0 = (v_R)^{H_2(M)t_s\alpha} (\bmod\, p)$ is protected by BFHP. That is, the pair (α, t_s) is protected by BFHP on σ. If the BFHP surrounding σ is *prf*-solved, then both (α, t_s) are found. Hence, no third party can forge a valid k_0 to cheat the receiver although he uses another pair of (α, t_s). □

Remark 5. On the other hand, if the DLP is solved, $t_s \in \mathbb{Z}_p$ would be found. However, the corresponding preferred α would not be obtained. In fact, both the preferred integers (α, t_s) is still not obtained.

Observed from $v_s = g^{t_s} (\bmod\ p)$. Solving the DLP, we will get $t_{s0} \in \mathbb{Z}_p$. If $t_s \equiv t_{s0} (\bmod\ p)$, then the attacker may initiate search for t_s since $t_s = t_{s0} + pj$ for some $j \in \mathbb{Z}$. Observe that since $t_{s0}, p \sim 2^n$ and $t_s \sim 2^{2n}$, we have $j \sim 2^n$ Hence the probability to obtain the correct j is $\dfrac{1}{2^n}$.

If $t_s \not\equiv t_{s0} (\bmod\ p)$, the attacker may not initiate search for t_s since he cannot find $j \in \mathbb{Z}$ as j is the number of time t_{s0} is reduced by p until t_s is obtained.

The following is an example continued from Example 2 in which we illustrated an attacker utilities attacked parameters (t'_s, α') as depicted in Proposition 4.

Example 3. *The authority randomly chooses $p = 137$ and $g = 101$ as a primitive root in $\mathbb{Z}_p$. The attacker, A chooses his secret key ${t_s}' = 727$. The receiver, R chooses his secret key $t_R = 157$ and computes $v_R = 11$.*

When A wants to deniably authenticate a message $M' = 555$ to the intended receiver R, he chooses randomly value of $\alpha = 847$. He computes $\sigma = 35228$ since $H_1(M) = 17$ and $H_2(M) = 27$. Then, he computes $\overline{k_1} = 37$ and $\overline{k_0} = 126$. Next, he generates MAC by applying the hash function to the concatenation between M' and k_0. He gets $MAC = 7306b18193e101e4e2b9a5bff79241e1$. Then, A sends $(\overline{k_1}, \sigma, MAC)$ together with message M' to R.

After receiving $(\overline{k_1}, \sigma, MAC)$ together with message M' from A, receiver, R computes $k_1^* = (v_S)^{\sigma t_R}$ using sender's public key, $v_s = 118$. He gets $k_1^* = 38$ and ${k_0}' = 36$. Then he computes $MAC = 4eca496522032ec8a7132e441c6725d1$. R verifies that $H(\overline{k_0}\|M') \neq H({k_0}'\|M')$. Then, R does not accept the information he gets from attacker.

Proposition 5. *If an attacker wants to impersonate as the intended receiver in order to identify the source of a given message, then the proposed protocol can withstand such an impersonation attack.*

Table 1: The comparison among deniable authentication protocols

	Fan et al. protocol		Zhang et al. protocol		The proposed protocol	
	S	R	S	R	S	R
Exponentiation	2+1	2+2	2	3	2	1
Hashing Computation	1+1	1+1	2	2	3	1
Data Transmission Overhead	$2\|n\| + 2\|h\|$		$2\|h\|$		$3\|n\| + \|r\|$	
Interactive	Yes		No		No	

Proof. In our protocol, any third party want to impersonate as the intended receiver cannot identify the source of the message even if he obtains (k_1, σ, MAC). If he can verify the message authenticator, he must find k_0 and ${k_0}'$. As we prove above, he cannot forge k_0 and ${k_0}'$ as

$$\begin{aligned} {k_0}' &= k_1 \cdot k_1^* \\ &= (v_R)^{-H_1(M){t_s}^2} \cdot (v_s)^{\sigma t_R} (\bmod p) \\ &= g^{-t_R H_1(M){t_s}^2} \cdot g^{{t_s}^2 H_1(M) t_R} \cdot g^{H_2(M) t_s \alpha t_R} (\bmod p) \\ &= g^{H_2(M) t_s \alpha t_R} (\bmod p) \\ &= (v_R)^{H_2(M) t_s \alpha} (\bmod p). \end{aligned}$$

It is shown that t_R is required in each step to calculate ${k_0}'$. Without the receiver's private key, t_R, it is impossible for the attacker to forge ${k_0}'$. □

Proposition 6. *The proposed protocol is secure against man-in-the-middle attack if man-in-the-middle cannot establish any session key with either the sender or the receiver.*

Proof. Objective of the man-in-the-middle attack is to pretend to be the sender and cheat the receiver. In order to pretend as a sender, he needs to compute σ' for the corresponding M'. But this is infeasible because the pair (α, t_s) is protected by BFHP within the initial σ. On the other hand, the man-in-the-middle cannot pretend to be the receiver to cheat the sender because he needs to obtain the receiver's private key, t_R to compute ${k_1}^* = (v_S)^{\sigma t_R}$. This is also infeasible because t_R is protected by the DLP within v_R. Therefore, the attacker is unable to pretend to be the sender or the receiver. □

4 Comparison

To study the performance of the proposed protocol, we compare it with some previous proposed deniable authentication protocols. We make comparison against the most known efficient interactive protocol (Fan et al. 2002) and non-interactive protocol (Y. Zhang et al. 2011). The comparison is summarized as in Table 1.

To authenticate the source of a message in Fan et al.'s interactive protocol, two modular exponentiation computation and one hashing computation are required by both sender and receiver. In addition, the sender needs to compute a signature with a message recovery which requires one modular and one hash function computation. The receiver needs to verify the signature which requires two modular exponentiation computation and one hash function computation. The data transmission overhead for Fan et al.'s protocol is $2|n| + 2|h|$ bits which $2|n|$ is the modular size and $2|h|$ is output size of hash function.

Our proposed protocol is non-interactive so that the communication process is shorter than in any interactive protocol. In signing phase, the sender needs two modular exponentiation computation and three hash function computation. The receiver needs one modular exponentiation computation and one hash function computation in verifying phase. Data transmission overhead for our proposed protocol is $3|n| + |r|$ bits, $|r|$ denotes the size of α and t_s while Y. Zhang et al.'s protocol is $2|h|$ bits.

5 Conclusion

A new deniable authentication protocol based on the bivariate function hard problem has been developed. One can observe from the Table 1 that the number of exponentiation computation needed is less that known efficient deniable authentication schemes. This suggested that the proposed method has better computational complexity on both the sender and the receiver's end.

The proposed protocol is proved to have the following characteristics which only intended receiver can be authenticated and it is deniable. Some possible attacks have also been considered and we showed that our proposed protocol is secure against forgery attack, impersonation attack and man-in-the-middle attack. Hence, our proposed deniable authentication protocol is more desirable than existing schemes. In the future studies, we will focus to improve the efficiency while still maintain the security of the protocol.

Acknowledgment

The author gratefully acknowledged the financial support of My PhD from the Ministry of Education Malaysia.

References

[1] M. R. K. Ariffin, M. A. Asbullah, N. A. Abu, and Z. Mahad, "A new efficient asymmetric cryptosystem based on the integer factorization problem of $n =$

p^2q," *Malaysian Journal of Mathematical Sciences*, vol. 7, pp. 19–37, 2013.

[2] Y. Aumann and M. O. Rabin, "Authentication, enhanced security and error correcting codes," in *Advances in Cryptology (CRYPTO'98)*, LNCS 1462, pp. 299–303, Springer-Verlag, 1998.

[3] X. Deng, C. H. Lee, and H. Zhu, "Deniable authentication protocol," *IEE Proceedings on Computer and Digital Techniques*, vol. 148, no. 2, pp. 101–108, 2001.

[4] C. Dwork, M. Noar, and A. Sahai, "Concurrent zero-knowledge," in *Proceedings of 30th Annual ACM Sympossium of Theory of Computing*, pp. 409–418, 1998.

[5] L. Fan, C. X. Xu, and J. H. Li, "Deniable authentication protocol based on diffie hellman algorithm," *Electronic Letters*, vol. 38, no. 4, pp. 705–706, 2002.

[6] M. Herrmann and A. May, "Solving linear equation modulo divisors: on factoring given any bits," *LAdvances in Cryptology (ASIACRYPT'08)*, LNCS 5350, pp. 406–424, Springer, 2008.

[7] S. J. Hwang and C. H. Chao, "An efficient non-interactive deniable authentication protocol with anonymous sender protection," *Journal of Discrete Mathematical Sciences and Cryptography*, vol. 13, no. 3, pp. 219–231, 2010.

[8] S. J. Hwang and J. C. Ma, "Deniable authentication protocol with anonymous sender protection," in *International Computer Symposium*, pp. 412–419, Tamsui, Taiwan, 2008.

[9] W. B. Lee, C. C. Wu, and W. J. Tsaur, "A novel deniable authentication protocol using generalized ElGamal signature scheme," *Information Sciences*, vol. 177, pp. 1376–1381, 2007.

[10] R. Lu and Z. Cao, "Non-interactive deniable authentication protocol based form bilinear pairings," *Applied Mathematics and Computation*, vol. 168, pp. 954–961, 2005.

[11] R. Lu and Z. Cao, "Non-interactive deniable authentication protocol based on factoring," *Computer Standard & Interfaces*, vol. 27, no. 4, pp. 401–405, 2005.

[12] Z. Shao, "Efficient deniable authentication protocol based on generalized elgamal signature scheme," *Computer Standard & Interface*, vol. 26, no. 5, pp. 449–454, 2002.

[13] Y. Zhang, Q. Xu, and Z. Liu, "A new non-interactive deniable authentication protocol based on generalized elgamal signature scheme," in *2011 6th IEE Joint International Conference on Information Technology and Artificial Intelligence Conference (ITAIC'11)*, pp. 193–197, Aug. 2011.

Normahirah Nek Abd Rahman received a Bachelor's degree and a Master's degree in Mathematics from Universiti Kebangsaan Malaysia in 2011 and 2012 respectively, and currently doing a PhD in Mathematical Cryptography in Universiti Putra Malaysia since 2013. Her current research topics include Diophantine equation, new attack on RSA cryptosystem using continued fraction and Coppersmith's method.

Muhammad Rezal Kamel Ariffin received his PhD in Mathematics from Universiti Kebangsaan Malaysia (National University of Malaysia) in 2009. He is a lecturer in the Department of Mathematics, Faculty of Science, Universiti Putra Malaysia (UPM). He is also an associate researcher at the Institute for Mathematical Research, UPM conducting research mainly in the mathematical aspects of cryptography. His current research interest is designing and analyzing number theoretic based cryptosystems.

Analysis of Second Order Matrices Construction in MFE Public Key Cryptosystem

Xuyun Nie[1,2], Chuanyong Hou[1], Zhaohu Xu[1] and Gang Lu[1]
(Corresponding author: Xuyun Nie)

School of Information and Software Engineering & University of Electronic Science and Technology of China[1]
Section 2, North Jianshe Road, Chengdu 610054, China
State Key Laboratory of Information Security & Institute of Information Engineering[2]
Beijing 100093, China
(Email: xynie@uestc.edu.cn)

(Received July 20, 2014; revised and accepted Mar. 20 & July 4, 2015)

Abstract

Medium Field Equations (MFE), which is a type of multivariate public key encryptions scheme proposed by Wang et al., was broken by Ding et al. using high order linearization equation (HOLE) attack. Recently, many people attempt to modify the second order matrices structure in the central map of MFE to resist HOLE attack. In this paper, we gave deeply analysis of all possible constructions by products of the second order matrices and their variants with transpose and adjoint in the central map of MFE. We proved that any modification with transpose and adjoint would satisfy the First Order Linearization Equations or the Second Order Linearization Equations. As an example, we gave a practical cryptanalysis of an improved MFE scheme.

Keywords: Linearization equation, MFE, multivariate public key cryptosystem, second order matrix

1 Introduction

Public key cryptosystem played an important role in our modern communication system. But with the rapid development of the quantum computer, the traditional public key cryptosystems based on the number theory hard problem, such as RSA and ElGamal cryptosystems, are all insecure under the quantum computer attack. Multivariate public key cryptosystem (MPKC) is one of the promising alternatives to the traditional public key cryptosystem against the quantum computer attack [8]. The security of the MPKC relies on the difficulty of solving a random system of nonlinear polynomial equations on a finite field, which is an NP-hard problem in general.

Let $\mathbb{K}$ be a finite field and m, n be two positive integers. The public key of MPKC is a set of multivariate polynomials, which are the expressions of the following map,

$$\begin{aligned}(y_1, \cdots, y_m) &= \bar{F}(x_1, \cdots, x_n)\\ &= T \circ F \circ S\\ &= (\bar{f_1}, \cdots, \bar{f_m}),\end{aligned}$$

where $\{y_1, \cdots, y_m\}$ are ciphertext variables and $\{x_1, \cdots, x_n\}$ are plaintext variables. The two invertible affine transformations T and S are the private keys of the MPKC, which are defined on $\mathbb{K}^m$ and $\mathbb{K}^n$ respectively. The map F is called central map. The key point in constructing an secure MPKC is to design a proper central map.

Medium Field Equation (MFE) [12] is a type of multivariate public key cryptosystem proposed by Wang et al. in 2006. The inventor of MFE used products of second order matrices to derive quadratic polynomials in its central map. To avoid the Paratin relation or linearization equations of form

$$\sum_{i=1,j=1}^{n,m} a_{ij}x_iy_j + \sum_{i=1}^{n} b_ix_i + \sum_{j=1}^{m} c_jy_j + d = 0,$$

the inventors used a transposed matrix instead of the original one in the central map of MFE. But the original MFE was broken by High Order Linearization Equation (HOLE) attack [2] in 2007. Given a public key, the attack can successfully recover the plaintext corresponding to a valid ciphertext.

In order to resist existing attack, many modifications of MFE were proposed. In 2009, Wang et al. [13] modified MFE and raised the public key from quadratic to quartic equations. It is indeed this case can avoid HOLEs attack. However, from their quartic public key, many so-called Quadratization Equations (QEs) can be found and can be used to break them [1]. In 2009, Tao et al. gave an improvement of MFE [9]. They introduced a new rational map in composition of the improvement and claimed that the new scheme can resist SOLEs attack. But there are still many SOLEs existing in this new scheme. Given a public key and a valid ciphertext, we can recover its corresponding plaintext [14]. In 2009, Huang et al. gave an improvement of MFE by redesigning the central map with transpose matrix and adjoint matrix [3]. After theoretical analysis, we found that it satisfied both Second Order Linearization Equations (SOLEs) and First Order Linearization Equations (FOLEs) [6].

In this paper, we summarize the steps of HOLEs attack. And then, we analyzed the construction based on the second order matrices in the central map of MFE. We found that if one want to remain degree two polynomials in the public key and ensure successfully decryption, one could only use the transpose matrices and the adjoint matrices. Given a second order matrix M over a finite field of characteristic 2, there are only 8 second order matrices with the same determinant of M. And these 8 matrices can be separated into two equivalent class with the matrix M and its transpose M^T. We list all possible constructions with a matrix M and its transpose M^T in the form of multiplication of two matrices. We found that all constructions will satisfy the SOLEs or FOLEs. So it is impossible to improve MFE by changing the form of second order matrices with their transpose and adjoint.

At last, we show how to find FOLEs satisfied by an improvement of MFE scheme [3] proposed by Jiasen Huang et al. After finding all the FOLEs, we use linearization equation attack breaking this improved version.

This paper is organized as follows. We introduce the MFE scheme, the idea of HOLEs attack on it and an improvement of MFE in Section 2. In Section 3, we give an analysis of the structure of the second order matrices in MFE scheme. Then we present a FOLEs attack on an improvement of MFE in Section 4. Finally, we conclude this paper in Section 5.

2 Preliminaries

In this section, we will introduce the MFE public key cryptosystem and the previous attack on MFE. Then, we will introduce one modification of MFE.

2.1 MFE Public Key Cryptosystem

We use the same notations as in [12]. Let $\mathbb{K}$ be a finite field of characteristic 2 and $\mathbb{L}$ be its degree r extension field. In MFE, we always identify $\mathbb{L}$ with $\mathbb{K}^r$ by a $\mathbb{K}$-linear isomorphism π: $\mathbb{L} \to \mathbb{K}^r$. Namely we take a basis of $\mathbb{L}$ over $\mathbb{K}$, $\{\theta_1, \cdots, \theta_r\}$, and define π by $\pi(a_1\theta_1+\cdots+a_r\theta_r) = (a_1, \cdots, a_r)$ for any $a_1, \cdots a_r \in \mathbb{K}$. It is natural to extend π to two $\mathbb{K}$-linear isomorphisms π_1: $\mathbb{L}^{12} \to \mathbb{K}^{12r}$ and π_2: $\mathbb{L}^{15} \to \mathbb{K}^{15r}$.

In MFE, its encryption map F: $\mathbb{K}^{12r} \to \mathbb{K}^{15r}$ is a composition of three maps ϕ_1, ϕ_2, ϕ_3. Let

$$(u_1, \cdots, u_{12r}) = \phi_1(x_1, \cdots, x_{12r}),$$

$$(v_1, \cdots, v_{15r}) = \phi_2(u_1, \cdots, u_{12r}),$$

$$(y_1, \cdots, y_{15r}) = \phi_3(v_1, \cdots, v_{15r}),$$

where ϕ_1 and ϕ_3 are invertible affine maps, ϕ_2 is its central map, which is equal to $\pi_1 \circ \bar{\phi}_2 \circ \pi_2^{-1}$.

ϕ_1 and ϕ_3 are taken as the private key, while the expression of the map $(y_1, \cdots, y_{15r}) = F(x_1, \cdots, x_{12r})$ is the public key.The map $\bar{\phi_2}$: $\mathbb{L}^{12} \to \mathbb{L}^{15}$ is defined as follows.

$$\begin{cases} Y_1 = X_1 + X_5X_8 + X_6X_7 + Q_1; & \\ Y_2 = X_2 + X_9X_{12} + X_{10}X_{11} + Q_2; & \\ Y_3 = X_3 + X_1X_4 + X_2X_3 + Q_3; & \\ Y_4 = X_1X_5 + X_2X_7; & Y_5 = X_1X_6 + X_2X_8; \\ Y_6 = X_3X_5 + X_4X_7; & Y_7 = X_3X_6 + X_4X_8; \\ Y_8 = X_1X_9 + X_2X_{11}; & Y_9 = X_1X_{10} + X_2X_{12}; \\ Y_{10} = X_3X_9 + X_4X_{11}; & Y_{11} = X_3X_{10} + X_4X_{12}; \\ Y_{12} = X_5X_9 + X_7X_{11}; & Y_{13} = X_5X_{10} + X_7X_{12}; \\ Y_{14} = X_6X_9 + X_8X_{11}; & Y_{15} = X_6X_{10} + X_8X_{12}, \end{cases}$$

where Q_1, Q_2, and Q_3 form a triple (Q_1, Q_2, Q_3) which is a triangular map from $\mathbb{K}^{3r}$ to itself, more detail please see [12].

The method of computing $\bar{\phi}_2^{-1}$ is listed as follows:
Write $X_1, \cdots, X_{12}, Y_4, \cdots, Y_{15}$ as six 2×2 matrices:

$$\begin{aligned} M_1 &= \begin{pmatrix} X_1 & X_2 \\ X_3 & X_4 \end{pmatrix} \\ M_2 &= \begin{pmatrix} X_5 & X_6 \\ X_7 & X_8 \end{pmatrix} \\ M_3 &= \begin{pmatrix} X_9 & X_{10} \\ X_{11} & X_{12} \end{pmatrix} \\ Z_3 = M_1M_2 &= \begin{pmatrix} Y_4 & Y_5 \\ Y_6 & Y_7 \end{pmatrix} \\ Z_2 = M_1M_3 &= \begin{pmatrix} Y_8 & Y_9 \\ Y_{10} & Y_{11} \end{pmatrix} \\ Z_1 = M_2^TM_3 &= \begin{pmatrix} Y_{12} & Y_{13} \\ Y_{14} & Y_{15} \end{pmatrix}. \end{aligned}$$

Then

$$\begin{cases} \det(M_1) \cdot \det(M_2) = \det(Z_3), \\ \det(M_1) \cdot \det(M_3) = \det(Z_2), \\ \det(M_2) \cdot \det(M_3) = \det(Z_1). \end{cases}$$

When M_1, M_2, and M_3 are all invertible, we can get values of $\det(M_1)$, $\det(M_2)$, and $\det(M_3)$ from $\det(Z_1)$, $\det(Z_2)$, and $\det(Z_3)$, for instance, $\det(M_1) = \big(\det(Z_2) \cdot \det(Z_3)/\det(Z_1)\big)^{1/2}$.

With the values of $\det(M_1)$, $\det(M_2)$, and $\det(M_3)$, we can use the triangular form of the central map to get $X_1, X_2, \cdots, X_{12}$ in turn. Then we can recover the plaintext corresponding the given ciphertext. More detail of decryption are presented in [12].

2.2 High Order Linearization Equation

High Order Linearization Equation (HOLE) is an type of equation of the following form:

$$\begin{aligned}\sum_{i=1,j=1}^{n,t} a_{ij} x_i g_j(y_1, y_2, \cdots, y_m) \\ + \sum_{k=1}^{l} c_k h_k(y_1, y_2, \cdots, y_m) + d = 0,\end{aligned} \tag{1}$$

where h_k, $1 \le k \le l$, g_j, $1 \le j \le t$, are polynomial functions in the ciphertext variables. The highest degree of g_j, $1 \le j \le l$ and h_k, $1 \le k \le l$ is called the order of the HOLE.

For example, the First Order Linearization Equation (FOLE) and the Second Order Linearization Equation (SOLE) are of the following forms, respectively.

$$\sum_{i=1,j=1}^{n,m} a_{ij} x_i y_j + \sum_{i=1}^{n} b_i x_i + \sum_{j=1}^{m} c_j y_j + d = 0.$$

$$\begin{aligned}\sum_i x_i \left(\sum_{j \le k} a_{ijk} y_j y_k + \sum_j b_{ij} y_j + c_i \right) \\ + \sum_{j \le k} d_{jk} y_j y_k + \sum e_j y_j + f = 0.\end{aligned}$$

Note that, given a valid ciphertext $\mathbf{y}' = (y_1', y_2', \cdots, y_m')$, we can substitute it into Equation (1) to get a linear equation in the plaintext variables. By finding all these equations we get a linear system in the plaintext variables, which can be solved by Gaussian Elimination. After having found a solution, we can do elimination on the public key or solve System (2).

$$\left\{ \begin{array}{ll} F_1(x_1, \cdots, x_n) & = y_1'; \\ \cdots & \\ F_m(x_1, \cdots, x_n) & = y_m'. \end{array} \right. \tag{2}$$

Then, we can also check whether there are some HOLEs satisfied by the eliminated public key and the form of HOLEs.

The steps of LE attack are listed in Algorithm 1.

Algorithm 1 Steps of LE Attack

```
1: Input: public key $F$ of a MPKC, ciphertext $\mathbf{y}' \in \mathbb{K}^m$
2: Output: corresponding plaintext $\mathbf{x}' \in \mathbb{K}^n$
3: Check whether there are some LEs satisfied by public key.
4: Determine the form of LEs and find all the LEs.
5: Substitute the ciphertext $\mathbf{y}'$ into the linearization equations and find all linear equations in the plaintext variables. Solve the system to find linear relations between plaintext variables. In other words, some plaintext variables can be written as linear expressions in the remaining variables.
6: Substitute the linear expressions of plaintext variables into the public key polynomials to get a "eliminated" public key expression (it is in fewer unknown plaintext components).
7: Check whether there are some LEs satisfied by the eliminated public key. If there are, goto Step 2.
8: Directly solve the last eliminated System (2).
9: Use the linear relations between plaintext variables to get the values of remained plaintext components.
```

2.3 Previous Attack on MFE

In designing the MFE scheme, the inventors have taken into account the LE attack. They used M_2^T instead of M_2 to avoid the FOLEs.

But Ding et al. found that there are many SOLEs satisfied by the MFE scheme. Denote by M^* the adjoint matrix of a second order matrix. From

$$Z_3 = M_1 M_2, Z_2 = M_1 M_3,$$

we have

$$M_3 M_3^* M_1^* M_1 M_2 = M_3 Z_2^* Z_3 = \det(Z_2) M_2. \tag{3}$$

Expanding Equation (3), we get four equations of the form

$$\sum a_{ijk}' X_i Y_j Y_k = 0. \tag{4}$$

In [2], 24 equations of this form can be found.

Substituting $(X_1, \cdots, X_{12}) = \pi_1^{-1} \circ \phi_1(x_1, \cdots, x_{12r})$ and $(Y_1, \cdots, Y_{15}) = \pi_2^{-1} \circ \phi_3^{-1}(y_1, \cdots, y_{15r})$ into Equation (4), we get $24r$ equations of the form

$$\begin{aligned}\sum_i x_i \left(\sum_{j \le k} a_{ijk} y_j y_k + \sum_j b_{ij} y_j + c_i \right) \\ + \sum_{j \le k} d_{jk} y_j y_k + \sum_j e_j y_j + f = 0.\end{aligned}$$

These equations are SOLEs.

Given a public key and a valid ciphertext, after finding all the SOLEs, one can recovered the corresponding plaintext efficiently.

2.4 Improvement of MFE

To avoid the SOLE, Jiasen Huang et al. proposed a modification of MFE. They modified only the matrix equations as follows.

M_1, M_2 and M_3 are defined as same as the origin MFE, while Z_1, Z_2 and Z_3 are defined as follows:

$$\begin{aligned} Z_3 &= M_1 M_2^* = \begin{pmatrix} Y_4 & Y_5 \\ Y_6 & Y_7 \end{pmatrix}, \\ Z_2 &= M_1^* M_3 = \begin{pmatrix} Y_8 & Y_9 \\ Y_{10} & Y_{11} \end{pmatrix}, \\ Z_1 &= M_2^T M_3^* = \begin{pmatrix} Y_{12} & Y_{13} \\ Y_{14} & Y_{15} \end{pmatrix}, \end{aligned}$$

where M_i^* $(1 \le i \le 3)$ are the adjoint matrices of M_i^*.

These matrices are also satisfied

$$\left\{ \begin{array}{l} \det(M_1) \cdot \det(M_2) = \det(Z_3), \\ \det(M_1) \cdot \det(M_3) = \det(Z_2), \\ \det(M_2) \cdot \det(M_3) = \det(Z_1). \end{array} \right.$$

so the decryption process is very similar to the original MFE. See [3] for more detail.

3 Analysis of the Structure Based on Second Order Matrices

In this section, we consider the second order matrices over a finite field $\mathbb{K}$ of characteristic 2.

In order to resist HOLE, many people try to improve the MFE scheme by modifying the second order matrices of the central map. To ensure the decryption successfully, they need keep the determinants unchanged.

Now we give two Propositions on the constructions by using the second order matrices.

Proposition 1. *Given a square matrix* $M = \begin{pmatrix} X_1 & X_2 \\ X_3 & X_4 \end{pmatrix}$, *where* $X_1, X_2, X_3, X_4 \in \mathbb{K}$, *there are eight square matrices which satisfy:*

1) *Components in these matrices are all constituted of* $X_1, X_2, X_3, X_4 \in \mathbb{K}$;

2) *The determinants of these matrices are equal to* $det(M)$.

And all matrices above can be transformed by M *or* M^T *through row transformations and column transformations.*

Proof: Given $X_1, X_2, X_3, X_4 \in \mathbb{K}$ of characteristic 2, there are 24 different matrices. We can calculate their determinate one by one. Clearly, there are eight matrices (including the matrix M) whose determinate equal to $det(M)$. We list as follows:

$$\begin{pmatrix} X_1 & X_2 \\ X_3 & X_4 \end{pmatrix}, \begin{pmatrix} X_2 & X_1 \\ X_4 & X_3 \end{pmatrix}, \begin{pmatrix} X_3 & X_4 \\ X_1 & X_2 \end{pmatrix},$$

$$\begin{pmatrix} X_4 & X_3 \\ X_2 & X_1 \end{pmatrix}, \begin{pmatrix} X_1 & X_3 \\ X_2 & X_4 \end{pmatrix}, \begin{pmatrix} X_2 & X_4 \\ X_1 & X_3 \end{pmatrix},$$

$$\begin{pmatrix} X_4 & X_2 \\ X_3 & X_1 \end{pmatrix}, \begin{pmatrix} X_3 & X_1 \\ X_4 & X_2 \end{pmatrix}.$$

Among the matrices above, the first four matrices can be easily derived from the matrix M through row transformation and column transformation. And the last four matrices can be gotten from M^T. □

Let us consider the following equations:

$$\begin{cases} Y_4 = X_1X_5 + X_2X_7; \\ Y_5 = X_1X_6 + X_2X_8; \\ Y_6 = X_3X_5 + X_4X_7; \\ Y_7 = X_3X_6 + X_4X_8. \end{cases} \quad (5)$$

The Equation (5) can be expressed by the following four matrices equations.

$$\begin{pmatrix} Y_4 & Y_5 \\ Y_6 & Y_7 \end{pmatrix} = \begin{pmatrix} X_1 & X_2 \\ X_3 & X_4 \end{pmatrix} \begin{pmatrix} X_5 & X_6 \\ X_7 & X_8 \end{pmatrix},$$

$$\begin{pmatrix} Y_4 & Y_5 \\ Y_6 & Y_7 \end{pmatrix} = \begin{pmatrix} X_2 & X_1 \\ X_4 & X_3 \end{pmatrix} \begin{pmatrix} X_7 & X_8 \\ X_5 & X_6 \end{pmatrix},$$

$$\begin{pmatrix} Y_6 & Y_7 \\ Y_4 & Y_5 \end{pmatrix} = \begin{pmatrix} X_3 & X_4 \\ X_1 & X_2 \end{pmatrix} \begin{pmatrix} X_5 & X_6 \\ X_7 & X_8 \end{pmatrix},$$

$$\begin{pmatrix} Y_6 & Y_7 \\ Y_4 & Y_5 \end{pmatrix} = \begin{pmatrix} X_4 & X_3 \\ X_2 & X_1 \end{pmatrix} \begin{pmatrix} X_7 & X_8 \\ X_5 & X_6 \end{pmatrix}.$$

So, we can say that the matrices $\begin{pmatrix} X_2 & X_1 \\ X_4 & X_3 \end{pmatrix}$, $\begin{pmatrix} X_3 & X_4 \\ X_1 & X_2 \end{pmatrix}$, $\begin{pmatrix} X_4 & X_3 \\ X_2 & X_1 \end{pmatrix}$ are equivalent to the matrix $\begin{pmatrix} X_1 & X_2 \\ X_3 & X_4 \end{pmatrix}$.

Similarly, the matrices $\begin{pmatrix} X_2 & X_4 \\ X_1 & X_3 \end{pmatrix}$, $\begin{pmatrix} X_4 & X_2 \\ X_3 & X_1 \end{pmatrix}$, $\begin{pmatrix} X_3 & X_1 \\ X_4 & X_2 \end{pmatrix}$ are equivalent to the matrix $\begin{pmatrix} X_1 & X_3 \\ X_2 & X_4 \end{pmatrix}$.

Notice that the matrix $\begin{pmatrix} X_4 & X_2 \\ X_3 & X_1 \end{pmatrix}$ is the adjoint matrix of the matrix $\begin{pmatrix} X_1 & X_2 \\ X_3 & X_4 \end{pmatrix}$. So, we can only consider a matrix and its transpose in the matrices form of the central map in MFE.

Proposition 2. *Given a square matrix* $M = \begin{pmatrix} X_1 & X_2 \\ X_3 & X_4 \end{pmatrix}$, *where* $X_1, X_2, X_3, X_4 \in \mathbb{K}$. $M_i, i = 1, \cdots, 4$ *are random second order matrices on finite field* $\mathbb{K}$, *define a set as follows:*

$$Q = \{MM_1, M_2M, M^TM_3, M_4M^T\},$$

then any two elements in Q can be deduced high order linearization equations in constructing the central map in MFE.

Proof: There are 6 forms of combination (Z_1, Z_2) in Q, we analysis of them respectively.

1) If $Z_1 = MM_1, Z_2 = M_2M$, we can derive
$$Z_2M_1 = M_2Z_1;$$

2) If $Z_1 = MM_1, Z_2 = M^TM_3$, we can derive
$$Z_2^TM_1 = M_3^TZ_1;$$

3) If $Z_1 = MM_1, Z_2 = M_4M^T$, we can derive
$$\det(Z_2)M_1 = M_4^T(Z_2^T)^*Z_1;$$

4) If $Z_1 = M_2M, Z_2 = M^TM_3$, we can derive
$$\det(Z_1)(M_3^T)^* = M_2^*Z_1(Z_2^T)^*;$$

5) If $Z_1 = M_2M, Z_2 = M_4M^T$, we can derive

$$Z_1M_4^T = M_2Z_2^T;$$

6) If $Z_1 = M^TM_3, Z_2 = M_4M^T$, we can derive

$$Z_2M_3 = M_4Z_1;$$

In Cases 1), 2), 5), and 6), we can derive FOLEs. In Cases 3) and 4), we can derive SOLEs.

The original MFE scheme satisfied Case 3) and 4) in the proof of Proposition 2. □

As to the improved MFE, the matrices equation $Z_2 = M_1^*M_3$ can be changed into

$$\begin{pmatrix} Y_{10} & Y_{11} \\ Y_8 & Y_9 \end{pmatrix} = \begin{pmatrix} X_1 & X_3 \\ X_2 & X_4 \end{pmatrix}\begin{pmatrix} X_{11} & X_{12} \\ X_9 & X_{10} \end{pmatrix}.$$

This equation and $Z_3 = M_1M_2^*$ satisfy Case 2). Similarly, according to the Proposition 1, we can deduce that the central map of the improved MFE scheme satisfy Cases 1), 5) and 6).

From Proposition 1 and Proposition 2 above, we can make sure that all the modifications of MFE by changing the form of matrices in MFE with their transpose and adjoint will fail to resist the HOLEs attack.

4 Linearization Equation Attack on Improvement of MFE

In this section, we give an example of Linearization Attack on Improvement of MFE. This work was presented on The 10th International Conference on Cryptology and Network Security (CANS 2011). The authors of [3] claimed their improvement of MFE can resist SOLEs attack. But according to Section 3, we know that this scheme satisfied the FOLEs. In this section, we will describe how to get the FOLEs and present the whole FOLE attack on this improvement.

4.1 Finding FOLEs

Note that, for any square matrices M_1 and M_2, we have

$$\begin{aligned} (M_1^*)^* &= M_1, \\ (M_1M_2)^* &= M_2^*M_1^*, \\ (M_1^*)^T &= (M_1^T)^*. \end{aligned}$$

From

$$Z_3 = M_1M_2^*, Z_2 = M_1^*M_3$$

we can derive

$$M_3^*Z_3 = M_3^*M_1M_2^* = (M_1^*M_3)^*M_2^* = Z_2^*M_2^*$$

and hence,

$$Z_2^*M_2^* = M_3^*Z_3$$

Expanding it, we have

$$\begin{pmatrix} Y_{11} & -Y_9 \\ -Y_{10} & Y_8 \end{pmatrix}\begin{pmatrix} X_8 & -X_6 \\ -X_7 & X_5 \end{pmatrix} = \begin{pmatrix} X_{12} & -X_{10} \\ -X_{11} & X_9 \end{pmatrix}\begin{pmatrix} Y_4 & Y_5 \\ Y_6 & Y_7 \end{pmatrix}.$$

That is,

$$\begin{cases} X_8Y_{11} + X_7Y_9 = X_{12}Y_4 - X_{10}Y_6 \\ -X_6Y_{11} - X_9Y_5 = X_{12}Y_5 - X_{10}Y_7 \\ -X_8Y_{10} - X_7Y_8 = -X_{11}Y_4 + X_9Y_6 \\ X_6Y_{10} + X_5Y_8 = -X_{11}Y_5 + X_9Y_7. \end{cases} \quad (6)$$

Applying $(X_1, \cdots, X_{12}) = \pi_1 \circ \phi_1(x_1, \cdots, x_{12r})$ and $(Y_1, \cdots, Y_{15}) = \pi_2^{-1} \circ \phi_3^{-1}(y_1, \cdots, y_{15r})$ into Equation (6), we get $4r$ equations of the form

$$\sum_{i,j} a_{ij}x_iy_j + \sum_i b_ix_i + \sum_j c_jy_j + d = 0, \quad (7)$$

where the coefficients $a_{ij}, b_i, c_j, d \in \mathbb{K}$, and the summations are respectively over $1 \leq i \leq 12r$ and $1 \leq j \leq 15r$. These equations are FOLEs. Apparently, these $4r$ equations are linearly independent.

Using the same technique, we can derive other $8r$ SOLEs. Note that

$$\begin{aligned} Z_1M_1 &= M_2^TM_3^*M_1 = M_2^TZ_2^* \\ Z_1^*M_1^T &= (M_2^TM_3^*)^*M_1^T = M_3(M_2^T)^*M_1^T \\ &= M_3(M_2^*)^TM_1^T = M_3Z_3^T. \end{aligned}$$

That is,

$$\begin{aligned} Z_1M_1 &= M_2^TZ_2^* \\ Z_1^*M_1^T &= M_3Z_3^T. \end{aligned}$$

Expanding them and substituting $(X_1, \cdots, X_{12}) = \pi_1 \circ \phi_1(u_1, \cdots, u_{12r})$ and $(Y_1, \cdots, Y_{15}) = \pi_2^{-1} \circ \phi_3^{-1}(z_1, \cdots, z_{15r})$ into them, we get another linearly independent $8r$ FOLEs.

To find all the FOLEs, we randomly generate sufficient plaintext/ciphertext pairs and substitute them into the FOLE to get a system of linear equations on the unknown coefficients $a_{1,1}, \cdots, a_{12r,15r}, b_1, \cdots, b_{12r}, c_1, \cdots, c_{15r}, d$. In this case, the number of unknown coefficients in these equations is equal to

$$12r \times 15r + 12r + 15r + 1 = 180r^2 + 27r + 1.$$

Suppose we derive D linearly independent FOLEs. Let $E_k(1 \leq k \leq D)$ denote these equations:

$$\sum_{i=1,j=1}^{12r,15r} a_{ij}^{(k)}x_iy_j + \sum_{i=1}^{12r} b_i^{(k)}x_i + \sum_{j=1}^{15r} c_j^{(k)}y_j + d^{(k)} = 0.$$

We used computer experiments to find all linearization equations. In one of our experiments, we choose $\mathbb{K} = GF(2^{16})$, $r = 4$. In this case, the number of unknown coefficients is equal to 2989.

Our experiments show that it take about 22 minutes on the execution of this step. And $D = 48$.

Note that, this step is independent of the value of the ciphertext and can be done once for a given public key.

4.2 Ciphertext-only Attack

Now we have derived all FOLEs. Our goal is to find corresponding plaintext $(x'_1, \cdots, x'_{12r})$ for a given valid ciphertext $(y'_1, \cdots, y'_{15r})$.

Substitute $(y'_1, \cdots, y'_{15r})$ into basis equations E_k, we can get k equations in following form:

$$\begin{cases} \sum\limits_{i,j} a_{ij}^{(k)} x_i y'_j + \sum\limits_i b_i^{(k)} x_i + \sum\limits_j c_j^{(k)} y'_j + d^{(k)} = 0 \\ 1 \leq k \leq D. \end{cases} \tag{8}$$

Suppose the dimension of the basis of System (8) solution space is s. Then, we can represent s variables of $x_1, \cdots, x_{12r}$ by linear combinations of other $12r - s$. Denote $w_1, \cdots, w_{12r-s}$ are remainder variables. Our experiments show $s = 32$, when $r = 4$.

Now substitute the expressions obtained above into $F_j(x_1, \cdots, x_{12r})$, we can get $15r$ new quadratic functions $\tilde{F}_j(w_1, \cdots, w_{12r-s})$, $j = 1, \cdots, 12r$. Then, our attack turn to solve the following system:

$$\begin{cases} \tilde{F}_i(w_1, \cdots w_{12r-s}) = y'_i \\ 1 \leq i \leq 15r. \end{cases} \tag{9}$$

There are $4r$ unknowns and $15r$ equations in System (9). We can solve this system by Gröbner basis method and recover the corresponding plaintext.

Our experiments show that it takes about 6 second to solve System (9) and our experiments recover the corresponding plaintext successfully.

All of our experiments were performed on a normal computer, with Genuine Intel(R) CPU T2300@1.66GHz, 504MB RAM by magma.

5 Conclusion

In this paper, we verified that all modifications of MFE by changing the form of matrices with transpose and adjoint will satisfy the SOLEs or FOLEs. Hence, they are all insecure.

In order to enhance the security of MPKCs, many enhancement methods were proposed such as Piece in hand [10], Extended Multivariate public key Cryptosystems (EMC) [11] etc. All of these methods are subjected to different levels of attacks [4, 5]. Recently, Qiao proposed three security enhancement methods on MPKC [7]. The security of their methods will be considered in the future.

Acknowledgments

This work was supported by the National Key Basic Research Program of China under grant 2013CB834203, the National Natural Science Foundation of China under grant No. 61370026, 61472064. The authors gratefully acknowledge the anonymous reviewers for their valuable comments.

References

[1] W. W. Cao, X. Y. Nie, L. Hu, X. L. Tang, and J. T. Ding, "Cryptanalysis of two quartic encryption schemes and one improved mfe scheme," in *Proceedings of The Third International Workshop (PQCrypto'10)*, pp. 41–60, Darmstadt, Germany, May 2010.

[2] J. T. Ding, L. Hu, X. Y. Nie, J. Y. Li, and J. Wagner, "High order linearization equation (hole) attack on multivariate public key cryptosystems," in *Proceedings of The 10th International Conference on Practice and Theory in Public-Key Cryptography (PKC'07)*, pp. 233–248, Beijing, China, Apr. 2007.

[3] J. S. Huang, B. D. Wei, and H. Y. Ou, "An improved MFE scheme resistant against sole attacks," in *Proceedings of Asia Pacific Conference on Postgraduate Research in Microelectronics and Electronics (PrimeAsia'09)*, pp. 157–160, Shanghai, China, Jan. 2009.

[4] X. Y. Nie, A. Petzoldt, J. Buchmann, and F. G. Li, "Linearization equation attack on 2-layer nonlinear piece in hand method," *IEICE Transactions on Fundamentals*, vol. E97-A, no. 9, pp. 1952–1961, 2014.

[5] X. Y. Nie, Z. H. Xu, and J. Buchmann, "Cryptanalysis of hash-based tamed transformation and minus signature scheme," in *Proceedings of The 5th International Workshop on Post-Quantum Cryptography (PQCrypto'13)*, pp. 115–164, Limoges, France, June 2013.

[6] X. Y. Nie, Z. H. Xu, L. Lu, and Y. J. Liao, "Security analysis of an improved MFE public key cryptosystem," in *Proceedings of The 10th International Conference on Cryptology and Network Security (CANS'11)*, pp. 118–125, Sanya, China, Dec. 2011.

[7] S. T. Qiao, W. B. Han, Y. F. Li, and L. Y. Jiao, "Construction of extended multivariate public key cryptosystems," *International Journal of Network Security*, vol. 18, no. 1, pp. 60–67, 2016.

[8] P. W. Shor, "Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer," *SIAM Review*, vol. 41, no. 2, pp. 303–332, 1999.

[9] H. W. Tao and Y. X. Chen, "An improved medium-field multivariate public-key encryption scheme," in *Proceedings of The International Conference on Computational Intelligence and Software Engineering*, pp. 1–4, Wuhan, China, Dec. 2009.

[10] S. Tsujii, K. Tadaki, R. Fujita, M. Gotaishi, and T. Kaneko, "Security enhancement of various mpkcs by 2-layer nonlinear piece in hand method," *IEICE Transactions on Fundamentals*, vol. E92-A, no. 10, pp. 2438–2447, 2009.

[11] H. Z. Wang, H. G. Zhang, Z. Y. Wang, and M. Tang, "Extended multivariate public key cryptosystems with secure encryption function," *Science China Information Sciences*, vol. 54, no. 6, pp. 1161–1171, 2011.

[12] L. C. Wang, B. Y. Yang, Y. H. Hu, and F.P. Lai, "Medium-field multivariate public key encryption scheme," in *Proceedings of The Cryptographers' Track at the RSA Conference*, pp. 132–149, San Jose, CA, USA, Feb. 2006.

[13] X. Wang, F. Feng, X. M. Wang, and Q. Wang, "A more secure mfe multivariate public key encryption scheme," *International Journal of Computer Science and Applications*, vol. 6, no. 3, pp. 1–9, 2009.

[14] Z. H. Xu, X. Y. Nie, H. Wang, and Y. J. Liao, "Cryptanalysis of an improved mfe public key cryptosystem," *International Journal of Security and Networks*, vol. 7, no. 3, pp. 174–180, 2012.

Xuyun Nie received the Ph.D. degree in Information security from Graduate university of Chinese Academy of Sciences, Beijing, China, in 2007. He is presently an Associate Professor at University of Electronic Science and Technology of China (UESTC). His research interests include cryptography and information security.

Chuanyong Hou received his Master Degree from University of Electronic Science and Technology of China in 2015. His research interests include cryptography and network security.

Zhaohu Xu received his Master Degree from University of Electronic Science and Technology of China in 2013. His research interests include multivariate public key cryptography and network security.

Gang Lu is a PH.D candidate in University of Electronic Science and Technology of China now. His research interests include cryptography and security of big data.

Cryptanalysis of an Identity Based Signcryption Scheme in the Standard Model

Yang Ming[1], Yumin Wang[2]
(Corresponding author: Yang Ming)

School of Information Engineering, Chang'an University[1]
Xi'an, Shaanxi 710064, China
State Key Laboratory of ISN, Xidian University[2]
Xi'an, Shaanxi 710071, China
(Email: yangming@chd.edu.cn, ymwang@xidian.edu.cn)

(Received Sept. 9, 2013; revised and accepted Jan 10 & May 25, 2014)

Abstract

Identity based signcryption (IBSC) is a novel cryptographic primitive that simultaneously provides the authentication and encryption in a single logic step. The IBSC has been shown to be useful in many applications, such as electronic commerce, mobile communications and smart cards. Recently, Li et al. (2013) [16] proposed a new identity based signcryption scheme and claimed that their scheme was provably secure in the standard model, i.e. (IND-IBSC-CCA2) semantically secure under adaptively chosen-ciphertext attack and (EUF-IBSC-CMA) existential unforgeable under adaptively chosen-message. However, in this paper, by giving concrete attacks, we show that Li et al's scheme is not secure in their security model. Additionally, we further indicate that Li et al's scheme also does not satisfy strongly existential unforgeability.

Keywords: Existential unforgeability, identity-based signcryption, semantically security, signcryption, standard model

1 Introduction

Confidentiality, integrity, non-repudiation and authentication are the important requirements for cryptographic applications. A traditional approach to achieve these requirements is to sign-then-crypt the message. The concept of signcryption was first proposed by Zheng [31]. The idea of this kind of primitive is to perform signature and encryption simultaneously in order to reduce the computational costs and communication overheads.

The concept of identity-based (simply ID-based) public key cryptography (ID-PKC) was introduced by Shamir [22] in 1984, which simplifies key management procedure of traditional certificate-based public key cryptography. The main idea of ID-PKC is that the user's public key can be calculated directly from his/her identity such as email addresses rather than being extracted from a certificate issued by a certificate authority (CA). Private keys are generated for the users by a trusted third party, called Private Key Generator (PKG) using some master key related to the global parameters for the system. The direct derivation of public keys in ID-PKC eliminates the need for certificates and some of the problems associated with them.

Lee present the first identity based signcryption (IBSC) scheme [18]. Since then, many identity based signcryption schemes were proposed [1, 3, 6, 7, 8, 15, 17]. To offer strong security guarantee, provable security is very essential for IBSC schemes. However, the early schemes [1, 3, 6, 7, 8, 15, 17, 18, 23, 27] use random oracle model to achieve the security requirement. The random oracle model was introduced by Bellare and Rogaway in [2]. The model is a formal model in analyzing cryptographic schemes, where a hash function is considered as a black-box that contains a random function. Although the model is efficient and useful, it has received a lot of criticism that the proofs in the random oracle model are not proofs. Canetti et al. [5] have shown that security in the random oracle model does not imply the security in the real world in that a scheme can be secure in the random oracle model and yet be broken without violating any particular intractability assumption, and without breaking the underlying hash functions.

Recently many efforts have been made to design provably secure IBSC scheme in the standard model (without using random oracles). In 2009, based on Waters scheme [26], Yu et al. [28] proposed the first identity based signcryption scheme without random oracles. However, in 2010, Wang and Qian [24], Jin et al. [10], Zhang [29] and Zhang et al. [30] independently pointed out that Yu et al.'s scheme [28] cannot achieve indistinguishability against chosen plaintext attacks. To remedy the security problem, Jin et al. [10] and Zhang [29] proposed improved IBSC schemes, respectively. Meanwhile, Ren and Gu [19] proposed a signcryption scheme based on Gentry's IBE [9] but it was shown by Wang et al. [25] that it had neither confidentiality nor existential unforgeability. In 2011, Li et al. [11] showed that the scheme in [10] satisfies neither confidentiality nor existentially unforgeability. Li and Takagi [14] further pointed out that the IBSC scheme in [10, 29] did not have the IND-CCA2 property (not even chosen plaintext attacks (IND-CPA)) and then present a fully secure IBSC scheme in the standard model. Li et al. also proposed anther two IBSC schemes [12, 13] in the standard model. But Selvi et al. [20] have also shown that Li et al's schemes [12, 13, 14] are not secure in the standard model. In 2012, Selvi et al. [21] presented the first provably secure ID based signcryption scheme in the standard model. This scheme satisfied the strongest notions of security available for the ID based signcryption schemes. In 2013, Li et al. [16] proposed a new identity-based signcryption scheme and claimed that their scheme is proven to be semantically secure under chosen-ciphertext attack and unforgeable under chosen-message attack in the standard model.

In this paper, using concrete attacks, we show that the Li et al's ID-based signcryption scheme [16] is not semantically secure under chosen-ciphertext attack and unforgeable under chosen-message attack. In addition, we indicate that this scheme is not strongly existentially unforgeable also.

2 Preliminaries

In this section, we briefly review the basic concepts on bilinear pairings, the formal definition and security model of identity based signcryption scheme.

2.1 Bilinear Pairings

Let $\mathbb{G}$ and $\mathbb{G}_T$ be two multiplicative cyclic groups of prime order p and let g be a generator of $\mathbb{G}$. The map e: $\mathbb{G}\times\mathbb{G} \rightarrow \mathbb{G}_T$ is said to be an admissible bilinear pairing with the following properties:

1) **Bilinearity:** $e(u^a, v^b) = e(u, v)^{ab}$ for all $u, v \in \mathbb{G}$ and for all $a, b \in \mathbb{Z}_p$.
2) **Non-degeneracy:** $e(g, g) \neq 1_{\mathbb{G}}$.
3) **Computability:** There exists an efficient algorithm to compute $e(u, v)$ for all $u, v \in \mathbb{G}$.

We note the modified Weil and Tate pairings associated with supersingular elliptic curves as examples of such admissible pairings.

2.2 Definition of Identity Based Signcryption

An identity based signcryption scheme consists of the following four functions:

Setup. Given a security parameter k, the private key generator (PKG) generates system parameters *params* and a master key *msk*. *params* is made public while *msk* is kept secret.

Extract. Given an identity u, the PKG computes the corresponding private key d_u and transmits it to u via a secure channel.

Signcrypt. Given a message M, the sender's private key d_s, and the receiver's identity u_r, the sender computes $\textbf{Signcrypt}(M, d_s, u_r)$ to obtain the ciphertext σ.

Unsigncrypt. When receiving σ, the receiver with identity u_r computes $\textbf{Unsigncrypt}(\sigma, d_r, u_s)$ and obtains the plaintext M or the symbol $\perp$ if σ is an invalid ciphertext between identities u_s and u_r.

2.3 Security Model of Identity Based Signcryption

Based on Malone-Lee model [18], Li et al. [16] defined the security notions for identity based signcryption scheme. The notions are semantically secure (i.e. indistinguishability against adaptive chosen ciphertext attacks, IND-IBSC-CCA2) and existentially unforgeable against adaptive chosen messages attacks (EUF-IBSC-CMA).

Confidentiality Game: For confidentiality, we consider the following game played between a challenger $\mathcal{C}$ and an adversary $\mathcal{A}$.

Setup. The challenger $\mathcal{C}$ takes a security parameter k and runs **Setup** algorithm to generate system parameters *params* and the master key *msk*. Then $\mathcal{C}$ sends *params* to $\mathcal{A}$ and keeps *msk* secret.

Phase 1. The adversary $\mathcal{A}$ can perform a polynomially bounded number of the following queries. These queries may be made adaptive, i.e. each query may depend on the answers to the previous queries.

Extract Queries. The adversary $\mathcal{A}$ chooses an identity u, $\mathcal{C}$ computes $d_u = \textbf{Extract}(u)$ and sends d_u to $\mathcal{A}$.

Signcrypt Queries. The adversary $\mathcal{A}$ produces a sender's identity u_s, the receiver's identity u_r and a plaintext M. $\mathcal{C}$ computes $d_s = \textbf{Extract}(u_s)$ and $\sigma = \textbf{Signcrypt}(M, d_s, u_r)$ and sends σ to $\mathcal{A}$.

Unsigncrypt Queries. The adversary $\mathcal{A}$ produces a sender's identity u_s, the receiver's identity u_r and a ciphertext σ. $\mathcal{C}$ generates the private key $d_r = \textbf{Extract}(u_r)$ and sends the result of $\textbf{Unsigncrypt}(\sigma, d_r, u_s)$ to $\mathcal{A}$.

Challenge. The adversary $\mathcal{A}$ decides when phase 1 ends. $\mathcal{A}$ chooses two equal length plaintexts M_0 and M_1, a sender's identity u_s^* and the receiver's identity u_r^* on which to be challenged. The identity u_r^*

should not appear in any extract queries in phase 1. $\mathcal{C}$ chooses randomly a bit b, computes $\sigma^* = \textbf{Signcrypt}(M_b, d_s^*, u_r^*)$ and sends σ^* to $\mathcal{A}$.

Phase 2. The adversary $\mathcal{A}$ makes a polynomial number of queries adaptively again as in phase 1 with the restriction that it cannot make extract query on u_r^* and cannot make an unsigncrypt query on σ^* under u_r^*.

Guess. The adversary $\mathcal{A}$ produces a bit b' and wins the game if $b' = b$.

The advantage of $\mathcal{A}$ is defined as $Adv^{Enc}(\mathcal{A}) = 2|\Pr[b' = b] - 1|$, where $\Pr[b' = b]$ denotes the probability that $b' = b$.

Definition 1. *(Confidentiality): An identity based signcryption scheme is said to have the indistinguishability against adaptive chosen ciphertext attacks (IND-IBSC-CCA2) or semantically security if no polynomially bounded adversary has a non-negligible advantage in the confidentiality game.*

Unforgeability Game: For unforgeability, we consider the following game played between a challenger $\mathcal{C}$ and an adversary $\mathcal{A}$.

Setup. The challenger $\mathcal{C}$ runs the **Setup** algorithm with a security parameter k obtains system parameters *params* and the master secret key *msk*. $\mathcal{C}$ sends *params* to $\mathcal{A}$.

Queries. The adversary $\mathcal{A}$ performs polynomially bounded number of queries adaptively just like in the confidentiality game.

Forgery. Finally, the adversary $\mathcal{A}$ produces a forgery (σ^*, u_s^*, u_r^*). We say $\mathcal{A}$ wins the game if the following are satisfied.

1) The ciphertext σ^* is valid.
2) The private key of u_s^* was not asked in the extract queries.
3) The ciphertext σ^* is not returned during the signcrypt queries.

The advantage of $\mathcal{A}$ is defined as the probability of success in winning the above game.

Definition 2. *(Unforgeability) An identity based signcryption scheme is said to have the existentially unforgeable against adaptive chosen message attacks (EUF-IBSC-CMA) if no polynomially bounded adversary has a non-negligible advantage in the unforgeability game.*

3 Review of Li et al. Identity Based Signcryption Scheme

In this section, we review Li et al.'s identity based signcryption scheme [16]. This scheme consists of the following four functions.

Setup. Let $(\mathbb{G}, \mathbb{G}_T)$ be bilinear groups such that $|\mathbb{G}| = |\mathbb{G}_T| = p$ for some prime p, and let g be a generator of $\mathbb{G}$. Given a bilinear map $e : \mathbb{G} \times \mathbb{G} \rightarrow \mathbb{G}_T$ and a collision-resistant hash function $H : \{0,1\}^* \rightarrow \{0,1\}^{n_m}$, the private key generator (PKG) randomly chooses $\alpha \in \mathbb{Z}_p^*$ and computes $g_1 = g^\alpha$. In addition, the PKG randomly picks up $g_2, u', m' \in \mathbb{G}$ and two vectors $\overrightarrow{u} = (u_i)$, $\overrightarrow{m} = (m_i)$ of length and n_u, n_m, respectively. The system parameters are $params = (\mathbb{G}, \mathbb{G}_T, e, H, g, g_1, g_2, u', m', \overrightarrow{u}, \overrightarrow{m})$ and the master key is $msk = g_2^\alpha$.

Extract. Let $U \subset \{1, \cdots, n_u\}$ be the set of indices such that $u[i] = 1$, where $u[i]$ is the i-th bit of u. Given an identity u, PKG randomly picks up $k_u \in \mathbb{Z}_p^*$ and computes

$$d_u = (d_{u1}, d_{u2}) = \left(g_2^\alpha (u' \prod_{i \in U} u_i)^{k_u}, g^{k_u} \right).$$

Suppose that the strings u_s and u_r of n_u bits are the identities of the sender and the receiver respectively. Let $U_s, U_r \subset \{1, \cdots, n_u\}$ be the set of indices that $u_s[i] = 1$, $u_r[i] = 1$, where $u_s[i]$, $u_r[i]$ are the i-th bit of u_s, u_r respectively. Therefore, the private keys for the sender and the receiver are

$$d_s = (d_{s1}, d_{s2}) = \left(g_2^\alpha (u' \prod_{i \in U_s} u_i)^{k_s}, g^{k_s} \right)$$

$$d_r = (d_{r1}, d_{r2}) = \left(g_2^\alpha (u' \prod_{i \in U_r} u_i)^{k_r}, g^{k_r} \right).$$

Singcrypt. On input $M \in \mathbb{G}_T$, the receiver's identity u_r, the sender with identity u_s uses his private key $d_s = (d_{s1}, d_{s2})$ to do the following steps:

1) Randomly choose $k \in \mathbb{Z}_p$;
2) Compute $\sigma_1 = M \cdot e(g_1, g_2)^k$;
3) Compute $\sigma_2 = g^k$;
4) Compute $\sigma_3 = (u' \prod_{i \in U_r} u_i)^k$;
5) Compute $\sigma_4 = d_{s2}$;
6) Compute $m = H(\sigma_1, \sigma_2, \sigma_3, \sigma_4, u_s, u_r)$ and let $M \subset \{1, \cdots, n_m\}$ be the set of indices j such that $m[j] = 1$;
7) Compute $\sigma_5 = d_{s1} \cdot (m' \prod_{j \in M} m_j)^k$;
8) Output the ciphertext $\sigma = (\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5)$.

Unsigncrypt. On input the ciphertext $\sigma = (\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5)$, the sender's identity u_s, the receiver with private key $d_r = (d_{r1}, d_{r2})$ decrypts the ciphertext as follows:

1) Compute $m = H(\sigma_1, sigma_2, \sigma_3, \sigma_4, u_s, u_r)$ and let $M \subset \{1, \cdots, n_m\}$ be the set of indices j such that $m[j] = 1$, where $m[j]$ is the j-th bit of m.

2) Check whether the following equality holds:

$$\begin{aligned} e(\sigma_5, g) &= e(g_1, g_2) \cdot e\left(u' \prod_{i \in U_s} u_i, \sigma_4\right) \\ &\quad \cdot e\left(m' \prod_{i \in M} m_j, \sigma_2\right). \end{aligned}$$

If holds, output $M = \sigma_1 \cdot \frac{e(d_{r2}, \sigma_3)}{e(d_{r1}, \sigma_2)}$ and $\perp$ otherwise.

4 Cryptanalysis of Li et al.'s Identity Based Signcryption Scheme

Although Li et al. [16] proved that their scheme is both semantically secure against adaptive chosen-ciphertext attacks (IND-IBSC-CCA2) and existentially unforgeable against adaptive chosen message attacks (EUF-IBSC-CMA). However, we will disprove their claims by giving three concrete attacks.

4.1 Attack Against Semantical Security

Li et al. [16] claimed that their scheme is semantically secure against adaptive chosen-ciphertext attack in the standard model, given that decisional bilinear Diffie-Hellman problem is hard. Unfortunately, this is not true. We show that his conclusion does not hold.

There exists a polynomial time adversary $\mathcal{A}$ who can always win IND-IBSC-CCA2 game as follows:

Setup. An adversary $\mathcal{A}$ generates master key msk and system parameters $params$ for challenger $\mathcal{C}$. In particular, $\mathcal{A}$ randomly chooses x', y', $x_1, \cdots, x_{n_u}$, $y_1, \cdots, y_{n_m} \in \mathbb{Z}_p$ and defines parameters $u', m', \overrightarrow{u}, \overrightarrow{m}$ as follows:

$$\begin{aligned} u' = g^{x'}, u_1 = g^{x_1}, \cdots, u_{n_u} = g^{x_{n_u}} \\ m' = g^{y'}, m_1 = g^{y_1}, \cdots, m_{n_m} = g^{y_{n_m}}. \end{aligned}$$

Phase 1. $\mathcal{A}$ need not issue any query.

Challenge. $\mathcal{A}$ generates two equal length plaintexts M_0 and M_1, and two identities u_s^* and u_r^* on which it wants to be challenged. When $\mathcal{A}$ receives the challenge ciphertext $\sigma^* = \mathbf{Signcrypt}(M_b, d_s^*, u_r^*)$, where b is the randomly bit chosen by the challenger. Recall that $\mathcal{A}$'s goal is to correctly guess the value b.

According to **Signcrypt** algorithm, the challenge ciphertext $\sigma^* = (\sigma_1^*, \sigma_2^*, \sigma_3^*, \sigma_4^*, \sigma_5^*)$ is of the following forms:

$$\begin{aligned} \sigma_1^* &= M_b \cdot e(g_1, g_2)^{k^*}, \\ \sigma_2^* &= g^{k^*}, \\ \sigma_3^* &= (u' \prod_{i \in U_r^*} u_i)^{k^*}, \\ \sigma_4^* &= d_{s2}^*, \\ \sigma_5^* &= d_{s1}^* \cdot (m' \prod_{j \in M^*} m_j)^{k^*}, \end{aligned}$$

where $U_r^* \subset \{1, \cdots, n_u\}$ be the set of indices i such that $u_r^*[i] = 1$, $M^* \subset \{1, \cdots, n_m\}$ be the set of indices j such that $m^*[j] = 1$ and $m^* = H(\sigma_1^*, \sigma_2^*, \sigma_3^*, \sigma_4^*, u_s^*, u_r^*)$.

Phase 2. Firstly, the adversary $\mathcal{A}$ randomly picks $\bar{k} \in \mathbb{Z}_p^*$ and defines another ciphertext $\bar{\sigma} = (\bar{\sigma}_1, \bar{\sigma}_2, \bar{\sigma}_3, \bar{\sigma}_4, \bar{\sigma}_5)$ as follows:

$$\begin{aligned} \bar{\sigma}_1 &= \sigma_1^* \cdot e(g_1, g_2)^{\bar{k}}, \\ \bar{\sigma}_2 &= \sigma_2^* \cdot g^{\bar{k}}, \\ \bar{\sigma}_3 &= \sigma_3^* \cdot (u' \prod_{i \in U_r^*} u_i)^{\bar{k}}, \\ \bar{\sigma}_4 &= \sigma_4^*, \\ \bar{\sigma}_5 &= \frac{\sigma_5^*}{(\sigma_2^*)^{y' + \sum_{j \in M^*} y_j}} \cdot (\sigma_2^*)^{y' + \sum_{j \in \bar{M}} y_j} \cdot (m' \prod_{j \in \bar{M}} m_j)^{\bar{k}}, \end{aligned}$$

where $\bar{M} \subset \{1, \cdots, n_m\}$ be the set of indices j such that $m^*[j] = 1$ and $\bar{m} = H(\bar{\sigma}_1, \bar{\sigma}_2, \bar{\sigma}_3, \bar{\sigma}_4, u_s^*, u_r^*)$.

Indeed, $\bar{\sigma} = (\bar{\sigma}_1, \bar{\sigma}_2, \bar{\sigma}_3, \bar{\sigma}_4, \bar{\sigma}_5)$ is a valid ciphertext under the same message M_b, the same sender with identity u_s^* and the receiver with identity u_r^*.

Correctness.

$$\begin{aligned} \bar{\sigma}_1 &= \sigma_1^* \cdot e(g_1, g_2)^{\bar{k}} \\ &= M_b \cdot e(g_1, g_2)^{k^*} \cdot e(g_1, g_2)^{\bar{k}} \\ &= M_b \cdot e(g_1, g_2)^{k^* + \bar{k}} \\ \bar{\sigma}_2 &= \sigma_2^* \cdot g^{\bar{k}} \\ &= g^{k^*} \cdot g^{\bar{k}} \\ &= g^{k^* + \bar{k}}, \\ \bar{\sigma}_3 &= \sigma_3^* \cdot (u' \prod_{i \in U_r^*} u_i)^{\bar{k}} \\ &= (u' \prod_{i \in U_r^*} u_i)^{k^*} \cdot (u' \prod_{i \in U_r^*} u_i)^{\bar{k}} \\ &= (u' \prod_{i \in U_r^*} u_i)^{k^* + \bar{k}} \\ \bar{\sigma}_4 &= \sigma_4^* \\ &= d_{s2}^*, \end{aligned}$$

$$\begin{aligned}
\bar{\sigma}_5 &= \frac{\sigma_5^*}{(\sigma_2^*)^{y'+\sum_{j\in M^*} y_j}} \cdot (\sigma_2^*)^{y'+\sum_{j\in\bar{M}} y_j} \cdot (m'\prod_{j\in\bar{M}} m_j)^{\bar{k}} \\
&= \frac{d_{s1}^* \cdot (m' \prod_{j\in M^*} m_j)^{k^*}}{(\sigma_2^*)^{y'+\sum_{j\in M^*} y_j}} \cdot (\sigma_2^*)^{y'+\sum_{j\in \bar{M}} y_j} \\
&\qquad \cdot (m'\prod_{j\in\bar{M}} m_j)^{\bar{k}} \\
&= \frac{d_{s1}^* \cdot (m' \prod_{j\in M^*} m_j)^{k^*}}{(g^{k^*})^{y'+\sum_{j\in M^*} y_j}} \cdot (g^{k^*})^{y'+\sum_{j\in\bar{M}} y_j} \\
&\qquad \cdot (m'\prod_{j\in\bar{M}} m_j)^{\bar{k}} \\
&= \frac{d_{s1}^* \cdot (m' \prod_{j\in M^*} m_j)^{k^*}}{\left(g^{y'+\sum_{j\in M^*} y_j}\right)^{k^*}} \cdot \left(g^{y'+\sum_{j\in\bar{M}} y_j}\right)^{k^*} \\
&\qquad \cdot (m'\prod_{j\in\bar{M}} m_j)^{\bar{k}} \\
&= \frac{d_{s1}^* \cdot (m' \prod_{j\in M^*} m_j)^{k^*}}{(m' \prod_{j\in M^*} m_j)^{k^*}} \cdot (m'\prod_{j\in\bar{M}} m_j)^{k^*} \\
&\qquad \cdot (m'\prod_{j\in\bar{M}} m_j)^{\bar{k}} \\
&= d_{s1}^* \cdot (m'\prod_{j\in\bar{M}} m_j)^{k^*} \cdot (m'\prod_{j\in\bar{M}} m_j)^{\bar{k}} \\
&= d_{s1}^* \cdot (m'\prod_{j\in\bar{M}} m_j)^{k^*+\bar{k}}
\end{aligned}$$

Then, the adversary $\mathcal{A}$ issues an unsigncrypt query by submitting the ciphertext $\bar{\sigma}$ under the sender with identity u_s^* and the receiver with identity u_r^*. According to the restrictions in IND-IBSC-CCA2 game, it is legal for $\mathcal{A}$ to issue this query on $\bar{\sigma}$ since $\bar{\sigma} \neq \sigma^*$. So the challenger $\mathcal{C}$ has to return the underlying message M_b to $\mathcal{A}$. Finally, $\mathcal{A}$ can certainly know the value b from the value M_b and win the IND-IBSC-CCA2 game with probability 100%.

In conclusion, Li et al.'s scheme is not semantically secure against chosen-message attacks.

4.2 Attack Against Existential Unforgeability

In this subsection, we show that Li et al.'s scheme [16] is not existentially unforgeable against chose message attacks. Given a ciphertext from the sender, the adversary $\mathcal{A}$ can generate the private key of the sender. Thus, $\mathcal{A}$ can arbitrarily forge the ciphertext on any message on behalf of the sender.

There exists a polynomial time adversary $\mathcal{A}$ who can always win EUF-IBSC-CMA game as follows:

Setup. The adversary $\mathcal{A}$ generates the master key msk and the system parameters $params$ for challenger $\mathcal{C}$. In particular, $\mathcal{A}$ randomly chooses $y', y_1, \cdots, y_{n_m} \in \mathbb{Z}_p$ and defines parameters $m', \overrightarrow{m}$ as follows:

$$m' = g^{y'}, m_1 = g^{y_1}, \cdots, m_{n_m} = g^{y_{n_m}}$$

Query phase. $\mathcal{A}$ can issue a signcrypt query by submitting a sender's identity u_s, a receiver's identity u_r and a message M. According to the EUF-IBSC-CMA game, the challenger $\mathcal{C}$ returns the ciphertext $\sigma = (\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5) = \textbf{Signcrypt}(M_b, d_s, u_r)$. The ciphertext has following forms:

$$\begin{aligned}
\sigma_1 &= M \cdot e(g_1, g_2)^k, \\
\sigma_2 &= g^k, \\
\sigma_3 &= (u' \prod_{i\in U_r} u_i)^k, \\
\sigma_4 &= d_{s2}, \\
\sigma_5 &= d_{s1} \cdot (m' \prod_{j\in M} m_j)^k,
\end{aligned}$$

where $U_r \subset \{1, \cdots, n_u\}$ be the set of indices i such that $u_r[i] = 1$, $M \subset \{1, \cdots, n_m\}$ be the set of indices j such that $m[j] = 1$ and $m = H(\sigma_1, \sigma_2, \sigma_3, \sigma_4, u_s, u_r)$.

From $\sigma_2 = g^k$, $\sigma_4 = d_{s2}$ and $\sigma_5 = d_{s1} \cdot (m' \prod_{j\in M} m_j)^k$, we can obtain the private key $d_{s1} = \frac{\sigma_5}{(\sigma_2)^{y'+\sum_{j\in M} y_j}}$ and $d_{s2} = \sigma_4$.

Correctness.

$$\begin{aligned}
\frac{\sigma_5}{(\sigma_2)^{y'+\sum_{j\in M} y_j}} &= \frac{d_{s1} \cdot (m' \prod_{i\in M} m_j)^k}{(\sigma_2)^{y'+\sum_{j\in M} y_j}} \\
&= \frac{d_{s1} \cdot (m' \prod_{j\in M} m_j)^k}{\left(g^{y'+\sum_{j\in M} y_j}\right)^k} \\
&= \frac{d_{s1} \cdot (m' \prod_{j\in M} m_j)^k}{(m' \prod_{j\in M} m_j)^k} \\
&= d_{s1}.
\end{aligned}$$

Then, $\mathcal{A}$ can forge the ciphertext for any message on behalf of this sender and win the EUF-IBSC-CMA game with the probability 100%.

Therefore, Li et al. scheme is not existential unforgeable against chosen-message attacks.

4.3 Attack Against Strongly Existential Unforgeability

Strongly existential unforgeability [4] means that the adversary cannot forge any signature different from those generated by the challenger. In practice, given a signature on some message, no one can derive other signatures on the same message.

Similar to Subsection 4.2, the adversary $\mathcal{A}$ first obtains a valid ciphertext $\sigma = (\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5)$ through issuing a signcrypt query on any message M under the sender with identity u_s and the receiver with identity u_r. Then, we can easily obtain another valid ciphertext $\bar{\sigma} = (\bar{\sigma}_1, \bar{\sigma}_2, \bar{\sigma}_3, \bar{\sigma}_4, \bar{\sigma}_5)$ on the same message M under (u_s, u_r) using the same method in Step 4 of Subsection 4.1.

Therefore, the $\sigma = (\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5)$ and $\bar{\sigma} = (\bar{\sigma}_1, \bar{\sigma}_2, \bar{\sigma}_3, \bar{\sigma}_4, \bar{\sigma}_5)$ are both valid ciphertexts of message M. So, Li et al. scheme is also not strongly existentially unforgeable.

5 Conclusion

Li et al. [16] proposed the provably secure identity based signcryption scheme in the standard model. However, in this paper, we show that their scheme still has security weaknesses. By giving concrete attacks on their security model, we prove that Li et al.'s scheme is neither semantically secure against adaptive chosen ciphertext attack nor existential unforgeable against adaptive chosen message attack. Finally, we demonstrate that this scheme is not secure against strongly existential unforgeable model.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No.61202438), the China Postdoctoral Science Foundation (No.2011M501427), the Key Project of Industry Science and Technology of Shaanxi Province (No.2015GY014) and the Project of Technology Transfer Promoting Engineering of Xi'an City (No. CXY1437(10)).

References

[1] P. S. L. M. Barreto, B. Libert, N. McCullagh, and J. J. Quisquater,"Efficient and provably-secure identity based signatures and signcryption from bilinear maps," in *Advance in Cryptology (Asiacrypt'05)*, LNCS 3788, pp. 515–532, Springer-Verlag, 2005.

[2] M. Bellare and P. Rogaway, "The exact security of digital signatures-how to sign with RSA and Rabin," in *Advances in Cryptology (Eurocrypt'96)*, LNCS 950, pp. 399–416, Springer-Verlag, 1996.

[3] X. Boyen, "Multipurpose identity based signcryption: a Swiss army knife for identity based cryptography," in *Advance in Cryptology (Crypt'03)*, LNCS 2792, pp. 383–399, Springer-Verlag, 2003.

[4] D. Boneh, E. Shen, and B. Waters, "Strongly unforgeable signatures based on computational Diffie-Hellman," in *Proceedings of Public Key Cryptography (PKC'05)*, LNCS 3958, pp. 229–240, Springer-Verlag, 2005.

[5] R. Canetti, O. Goldreich, and S. Halevi, "The random oracle methodology, revisited," in *Proceedings of the ACM Symposium on the Theory of Computing (STOC'98)*, pp. 209–218, 1998.

[6] H. Chen, Y. Li, and J. Ren, "A practical identity-based signcryption scheme," *International Journal of Network Security*, vol. 15, no. 6, pp. 484–489, 2013.

[7] L. Chen and J. Malone-Lee, "Improved identity-based signcryption," in *Proceedings of Public Key Cryptography (PKC'05)*, LNCS 3386, pp. 362–379, Springer-Verlag, 2005.

[8] S. S. M. Chow, S. M. Yiu, L. C. K. Hui, and K. P. Chow, "Efficient forward and provably secure ID-based signcryption scheme with public verifiability and public ciphertext authenticity," in *Proceedings of Information Security and Cryptology (ICISC'03)*, LNCS 2971, pp. 352–369, Springer-Verlag, 2004.

[9] C. Gentry, "Practical identity-based encryption without random oracles," in *Advance in Cryptology (Eurocrypt'06)*, LNCS 4004, pp. 445–464 Springer-Verlag, 2006.

[10] Z. Jin, Q. Wen, and H. Du, "An improved semantically-secure identity-based signcryption scheme in the standard model," *Computers and Electrical Engineering*, vol. 36, pp. 545–552, 2010.

[11] F. Li, Y. Liao, and Z. Qin, "Analysis of an identity based signcryption scheme in the standard model," *IEICE Transations on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 94-A, no. 1, pp. 268–269, 2011.

[12] F. Li, Y. Liao, Z. Qin, and T. Takagi, "Further improvement of an identity-based signcryption scheme in the standard model," *Computers and Electrical Engineering*, vol. 38, pp. 413–421, 2012.

[13] F. Li, F. B. Muhaya, M. Zhang, and T. Takagi, "Efficient identity-based signcryption in the standard model," in *Proceedings of International Conference on Provable Security (ProvSec'11)*, LNCS 6980, pp. 120–137, Springer-Verlag, 2011.

[14] F. Li and T. Takagi, "Secure identity-based signcryption in the standard model," *Mathematical and Computer Modelling*, vol. 57, no. 11-12, pp. 2685–2694, 2013.

[15] F. Li, X. Xin, and Y. Hu, "ID-based signcryption scheme with (t,n) shared unsigncryption," *International Journal of Network Security*, vol. 3, no. 2, pp. 155–159, 2006.

[16] X. Li, H. Qian, J. Weng, and Y. Yu, "Fully secure identity-based signcryption scheme with shorter signcryptext in the standard model," *Mathematical and Computer Modelling*, vol. 57, pp. 503–511, 2013.

[17] B. Libert and J. J. Quisquator, "A new identity based signcryption scheme from pairings," in *Proceedings of IEEE information theory workshop (ITW'03)*, pp. 155–158, Elsevier, 2003.

[18] J. Malone-Lee, "Identity based signcryption," Cryptology ePrint Archive, Report 2002/098, 2002.

[19] Y. Ren and D. Gu, "Efficient identity based signature/signcryption scheme in the standard model," in *Proceedings of The IEEE First International Symposium on Data, Privacy, and E-Commerce (ISDPE'07)*, pp. 133–137, 2007.

[20] S. S. D. Selvi, S. S. Vivek, D. Vinayagamurthy, and C. P. Rangan, "On the security of ID based signcryption schemes," Cryptology ePrint Archive, Report 2011/664, 2011.

[21] S. S. D. Selvi, S. S. Vivek, D. Vinayagamurthy, and C. P. Rangan, "ID-based signcryption scheme in standard model," in *Proceedings of International Conference on Provable Security (ProvSec'12)*, LNCS 7496, Springer-Verlag, pp. 35–52, 2012.

[22] A. Shamir, "Identity-based cryptosystems and signature schemes," in *Advances in Cryptology (Crypto'84)*, LNCS 196, pp. 47–53, Springer-Verlag, 1984.

[23] M. Toorani and A. A. B. Shirazi, "Cryptanalysis of an elliptic curve-based signcryption scheme," *International Journal of Network Security*, vol. 10, no. 1, pp. 51–56, 2010.

[24] X. Wang and H. Qian, "Attacks against two identity-based signcryption schemes," in *Proceedings of IEEE International Conference on Networks Security, Wireless Communications and Trusted Computing (NSWCTC'09)*, pp. 24–27, 2009.

[25] X. A. Wang, W. Zhong, and H. Luo, "Cryptanalysis of efficient identity based signature/signcryption schemes in the standard model," in *Proceedings of IEEE International Symposium on Intelligence Information Processing and Trusted Computing (IPTC'10)*, pp. 622–625, 2010.

[26] R. Waters, "Efficient identity based encryption without random oracles," in *Advance in Cryptology (Eurocrypt'05)*, LNCS 3494, pp. 114–127, Springer-Verlag, 2005.

[27] H. Xiong, J. Hu, and Z. Chen, "Security flaw of an ECC-based signcryption scheme with anonymity," *International Journal of Network Security*, vol. 15, no. 4, pp. 317–320, 2013.

[28] Y. Yu, B. Yang, Y. Sun, and S. Zhu, "Identity based signcryption scheme without random oracles," *Computer Standards and Interfaces*, vol. 31, pp. 56–62, 2009.

[29] B. Zhang, "Cryptanalysis of an identity based signcryption scheme without random oracles," *Journal of Computational Information Systems*, vol. 6, no, 6, pp. 1923–1931, 2010.

[30] M. Zhang, P. Li, B. Yang, H. Wang, and T. Takagi, "Towards confidentiality of ID-based signcryption schemes under without random oracle model," in *Proceedings of Pacific Asia Workshop on Intelligence and Security Informatics (PAISI'10)*, LNCS 6122, pp. 98–104, Springer-Verlag, 2010.

[31] Y. Zheng, "Digital signcryption or how to achieve cost (signature encryption)? cost (signature) + cost (encryption)," in *Advances in Cryptology (Crypto'97)*, LNCS 1294, pp. 165–179, Springer-Verlag, 1997.

Yang Ming was born in Shaanxi Province, China in 1979. He received the B.S. and M.S. degrees in mathematics from Xian University of Technology in 2002 and 2005 respectively, and the Ph.D. degree in cryptography from Xidian University in 2008. Currently he is a supervisor of postgraduate and associate professor of Chang'an University. His research interests include cryptography and digital signature.

Yumin Wang was born in Beijing, China in 1936. He received the B.S. degree from the Department of Telecommunication Engineering, Xidian University in 1959. In 1979-1981, he was a visiting scholar in Department of Electronic Engineering, Hawaii University. Currently he is a doctoral supervisor and professor of Xidian University. He is a fellow member of the Board of Governors of the Chinese Institute of Cryptography (preparatory committee) and a Senior Member (SM) of IEEE. His research interests include information theory, coding, and cryptography.

Cryptanalysis and Efficient Dynamic ID Based Remote User Authentication Scheme in Multi-server Environment Using Smart Card

Ruhul Amin

Computer Science and Technology, Jakir Hossain Institute of Polytechnic

Aurangabad, Murshidabad, West Bengal, India

(Email: amin_ruhul@live.com)

(Received Sept. 17, 2013; revised and accepted July 5, 2014 & Nov. 6, 2014)

Abstract

Resembling the single server environment, if the multi-server environment using smart card provides the users to access the different servers after registering once with the registration center and uses the same password and identity for all the service provider's servers, then security would be the matter of great concern. So, remote user authentication scheme becomes necessary to provide a better security. In this regard, many dynamic ID-based remote user authentication schemes in multi-server environment using smart card have been proposed in the literature. In 2012, Sood proposed Dynamic Identity Based Authentication Protocol for Two-Server Architecture and claimed that his scheme is more efficient in terms of security. But it is pointed out that Sood's scheme is insecure against off-line identity guessing attack, off-line password guessing attack, privileged insider attack, user impersonation attack, session key recovery attack and many logged in users' attack. In 2012, Li et al.'s proposed a scheme for providing better performance than Sood's scheme. But unfortunately Li et al.'s scheme also is insecure against off-line identity guessing attack, off-line password guessing attack, user impersonation attack and many logged in users' attack. To overcome the above mention attacks for both the schemes and related attacks on remote user authentication like (identity and password guessing attack, user impersonation attack, server masquerading attack, insider attack, session key discloser attack, smart card stolen attack, replay attack, many logged in users' attack and stolen verifier attack etc.), we proposed an efficient dynamic ID-Based remote user authentication scheme in multi-server environment using smart card. After performance analysis, the proposed scheme has lower computation complexity, better communication cost and higher security that makes the authentication system more secure and efficient than both Sood's and Li et al.'s schemes published earlier.

Keywords: Authentication, dynamic ID, multi-server

1 Introduction

It is terribly inefficient and difficult for the users to remember different identities and passwords for accessing various remote servers repetitively, when users used many single-server environments. However, users can login the control server only once and then access numerous different remote service providing servers, if they use the multi-server authentication scheme [1, 4, 5, 7, 14]. In 2000, first Ford and Kaliski [6] proposed password based multi-server authentication protocol that splits password among different servers but the protocol has high computation due to use of public keys by the servers. Then in 2001, Jablon [10] improved Ford and Kaliski's protocol, which do not use public keys. In 2003, Lin et. al.'s [16] proposed a multi-server authentication protocol based on the ElGamal digital signature scheme. But the use of public keys makes this protocol computation intensive. In 2004, Juang [11] proposed a smart card based multi-server authentication protocol using asymmetric encryption algorithm without using any verification table. In the same year, Chang and Lee [3] proposed an improved scheme over Juang [11] scheme. In 2007, Hu et al. [9] proposed an efficient password authentication key agreement protocol for multi-server architecture in which user can access multiple servers using smart card and one weak password and also provides mutual authentication and secret session key for secure communication. In 2008, Tsai [21] proposed a multi-server authentication protocol using smart cards based on the nonce and one-way hash function that does not require storing any verification table on the server and the registration center. This protocol does not use any symmetric key or asymmetric key algorithm for implementation. In 2009, Liao and Wang [15] proposed a dynamic identity based remote user authentication protocol using smart cards to achieve users' anonymity. This protocol uses only cryptographic one-way hash function for the implementation. In the same year, Hsiang and Shih [8] found that Liao and Wangs protocol is susceptible to insider attack, masquerade attack, server spoofing attack, registration center spoofing attack and does not provide mutual authentication as well. To overcome these drawbacks, they proposed an improved scheme over Liao and Wang's [15] scheme. Then in 2010, Sood et al. [20] showed that Hsiang and Shih's [8] scheme is insecure against replay attack, impersonation attack and stolen smart card attack.

The remainder of this paper is organized as follows: Section 2. briefly reviews the Sood's [19] scheme. Section 3. shows cryptanalysis of Sood's [19] scheme. Section 4. briefly reviews the Li et el.'s scheme. Section 5. describes cryptanalysis of Li et al.'s [13] scheme. Section 6. describes the proposed scheme. Section 7. shows the cryptanalysis of the proposed scheme. Section 8. compares the performance analysis with related schemes published earlier. We conclude the paper in Section 9. Finally, references are given in Section 10.

1.1 Contribution

In this paper, we have briefly reviewed Sood's and Li et al.'s authentication protocol for multi-server environment. Then, we demonstrated that both schemes suffers from several attacks described in Section 3 and Section 5 respectively. Afterward, we proposed a remote user authentication protocol for multi-server environment. After cryptanalysis of the proposed protocol, it can be claimed that the proposed protocol has no security weaknesses and takes minimum computational and communication cost than related scheme.

1.2 Preliminaries

In this section, a briefly review the basic concepts of cryptographic one-way hash function and a related mathematical problem are introduced.

Cryptographic One-way Hash Function: A cryptographic hash function maps a string of arbitrary length to a string of fixed length called the hashed value. It can be symbolized as: $h : X \to Y$, where $X = \{0,1\}^*$, and $Y = \{0,1\}^n$. X is binary string of arbitrary length and Y is a binary string of fixed length n. It is used in many cryptographic applications such as digital signature, random sequence generators in key agreement, authentication protocols and so on. Cryptographic one-way hash function satisfies the following properties:

1) *Preimage Resistant:* It is hard to find m from given y, where $h(m) = y$.

2) *Second-Preimage Resistant:* It is hard to find input $m' \in X$ such that $h(m) = h(m')$ for given input $m \in X$ and $m' \neq m$.

3) *Collision Resistant:* It is hard to find a pair $(m, m') \in X \times X$ such that $h(m) = h(m')$, where $m \neq m'$.

4) *Mixing-Transformation:* On any input $m \in X$, the hashed value $y = h(m)$ is computationally indistinguishable from a uniform binary string in the interval $\{0, 2^n\}$, where n is the output length of hash $h(\cdot)$.

Factorization Problem [18]: It is computationally infeasible to find two large primes p and q each of length at least 1024-bits from given n $(= p \times q)$.

2 Brief Review of Sood's Scheme

This section presents a brief description of Sood's [19] dynamic ID-based remote user authentication scheme in multi-server environment using smart card. The notations used throughout this paper are summarized in Table 1.

Table 1: Notation used

CS	$\longrightarrow$	Control Server
S_k	$\longrightarrow$	$k-th$ Service Provider Server
U_i	$\longrightarrow$	$i-th$ user
ID_i	$\longrightarrow$	Identity of U_i
PW_i	$\longrightarrow$	Password chosen by U_i
x	$\longrightarrow$	Secret key of Control Server CS
$H(\cdot)$	$\longrightarrow$	Cryptographic one-way hash function
SK	$\longrightarrow$	Shared secret session key
$\oplus$	$\longrightarrow$	Bitwise xor operation
$\|$	$\longrightarrow$	Concatenate operation
$(\cdot)$	$\longrightarrow$	Multiplication operation

Sood's [19] scheme consists of the following phases: Registration Phase, Login Phase, Authentication and Session Key Agreement Phase and Password Change Phase.

2.1 Registration Phase

In this phase, user U_i submits identity ID_i and password PW_i to the Control server over secure channel for registration. After receiving ID_i and PW_i, Control server computes $Z_i = H(ID_i \parallel PW_i) \oplus H^2(x)$, $V_i = y_i \oplus ID_i \oplus H(x)$, $B_i = H(ID_i, PW_i) \oplus PW_i \oplus y_i$ and $C_i = H(y_i) \oplus ID_i \oplus x$, where x is the secret key of the control server and y_i is the random number chosen by the CS such that $y_i \oplus x$ will be unique for each user. Then, Control server CS stores $y_i \oplus x$ in its database corresponding to C_i and issues a smart card for the user U_i by storing the security parameter $Z_i, V_i, B_i, H(\cdot)$ into the memory of smart card.

All service provider servers have to register themselves with the control server CS and CS agrees on unique secret key SK_k with each service provider S_k. Then, S_k remembers secret key SK_k and CS store SK_k by computing $SK_k \oplus H(x \parallel SID_k)$ corresponding service provider identity SID_k. The CS sends ID_i and $H(y_i)$ corresponding to newly registered user U_i to all service provider. Then, all the service provider store ID_i and $H(y_i)$ in the database for further use.

2.2 Login Phase

In the login phase, user U_i insert his/her smart card into cardreader and submits ID_i^* and password PW_i^* and choose the identity of service provider server SID_k. Then, smart card computes $y_i = B_i \oplus H(ID_i^* \parallel PW_i^*) \oplus PW_i^*$, $H(x) = V_i \oplus y_i \oplus ID_i^*$, $Z_i^* = H(ID_i^* \parallel PW_i^*) \oplus H^2(x)$ and verifies whether computed Z_i^* is equals with stored Z_i. If the verification holds, smart card generates random nonce N_1 and computes $CID_i = V_i \oplus y_i \oplus H(y_i) \oplus N_1$, $M_i = H^2(x) \oplus N_1$ and $E_i = H(y_i \parallel H(x) \parallel N_1 \parallel ID_i \parallel SID_k)$. Then, smart card sends login message $\{SID_k, CID_i, M_i, E_i\}$ to the service provider S_k through public channel.

2.3 Authentication and Session Key Agreement Phase

After receiving login request message $\{SID_k$, CID_i, M_i, $E_i\}$, server S_k generates random nonce N_2 and computes $G_i = SK_k \cdot N_2$. Then, service provider server S_k sends login request message $\{SID_k, CID_i, M_i, E_i, G_i\}$ to the control server. After receiving it, control server computes $N_1 = M_i \oplus H^2(x)$, $N_2 = G_i \oplus SK_k$ and $C_i^* = CID_i \oplus N_1 \oplus H(x) \oplus x$. Then, CS checks the condition whether computed C_i^* is identical with the stored C_i in its database or not. If the condition does not hold, control server rejects the login request otherwise extract y_i from $y_i \oplus x$ stored in the database. Then, control server further computes $ID_i = C_i \oplus H(y_i) \oplus x$, $E_i^* = H(y_i \parallel H(x) \parallel N_1 \parallel ID_i \parallel SID_k)$ and compares E_i^* with the received E_i to verify the legitimacy of the user U_i and service provider S_k. If the condition holds, control server extracts SK_k from $SK_k \oplus H(x \parallel SID_k)$ stored in the database. Then, control server generates random nonce N_3 and computes $A_i = N_1 \oplus N_3 \oplus H(SK_k)$, $D_i = ID_i \oplus H(N_1 \oplus N_2 \oplus N_3)$, $F_i = H[H(N_1 \oplus N_2 \oplus N_3) \parallel ID_i \parallel H(y_i)]$, $T_i = N_2 \oplus N_3 \oplus H(y_i \parallel ID_i \parallel H(x) \parallel N_1)$ and sends message $\{A_i, D_i, F_i, T_i\}$ to the service provider server S_k.

Service provider server S_k then computes $N_1 \oplus N_3 = A_i \oplus H(SK_k)$, $ID_i = D_i \oplus H(N_1 \oplus N_2 \oplus N_3)$ and extracts $H(y_i)$ corresponding ID_i from its database. Afterward, server S_k computes $F_i^* = H[H(N_1 \oplus N_2 \oplus N_3) \parallel ID_i \parallel H(y_i)]$ and compares F_i^* with F_i to verify the legitimacy of control server. If the above condition holds, server S_k sends F_i and T_i to smart card of user U_i.

After receiving F_i and T_i, smart card computes $N_2 \oplus N_3 = T_i \oplus H(y_i \parallel ID_i \parallel H(x) \parallel N_1)$ and $F_i^* = H[H(N_1 \oplus N_2 \oplus N_3) \parallel ID_i \parallel H(y_i)]$. Then, compares computed F_i^* with received F_i to verify the legitimacy of control server CS and service provider server S_k. If the above condition holds, then login request is accepted, otherwise rejects the session. Finally, user U_i, control server CS and service provider server S_k agree on the common secret session key as $SK = H(ID_i \parallel (N_1 \oplus N_2 \oplus N_3) \parallel H(y_i))$.

2.4 Password Change Phase

This phase is invokes when U_i wants to change the password. U_i inserts the smart card into the card reader and submits ID_i^* and PW_i^*. Then, card reader computes $y_i = B_i \oplus H(ID_i^* \parallel PW_i^*) \oplus PW_i^*$, $H(x) = V_i \oplus y_i \oplus ID_i^*$, $Z_i^* = H(ID_i^* \parallel PW_i^*) \oplus H^2(x)$ and compares the computed value of Z_i^* with stored value of Z_i. If identifies, U_i enters a new password PW_i^{new}. Then, card reader computes $Z_i^{new} = Z_i \oplus H(ID_i \parallel PW_i) \oplus H(ID_i \parallel PW_i^{new})$ and $B_i^{new} = B_i \oplus H(ID_i \parallel PW_i) \oplus PW_i \oplus H(ID_i \parallel PW_i^{new}) \oplus PW_i^{new}$. Then, stores Z_i^{new} and B_i^{new} instead of Z_i and B_i into memory of smart card.

3 Security Analysis of Sood's Scheme

In this section, the cryptanalysis of Sood's [19] scheme is presented. To analyze the security weaknesses of Sood's scheme, our assumptions are given as follows.

Assumption 1. *It can be assumed that an attacker could obtain the secret values stored in the smart card by monitoring the power consumption [12, 17] and an attacker can intercept all communicating messages between the users, the service provider servers S_k and control server CS.*

Assumption 2. *Due to the low entropy of ID_i and PW_i selected by U_i, we assume an adversary is able to off-line guess U_i's identity ID_i and password PW_i individually. However, he/she cannot off-line guess ID_i and PW_i simultaneously in polynomial time as pointed out by Sood et al. [20].*

Assumption 3. *It can also be assumed that a valid user can provide secret information of the control server CS to an attacker or a valid user can acts as an attacker after deriving secret information of the control server.*

Under these assumptions, it can be explained various attacks on Sood's [19] scheme such as off-line identity guessing attack, off-line password guessing attack, privileged insider attack, user impersonation attack, session key recovery attack and many logged in users' attack.

3.1 Off-line Identity Guessing Attack

User's identity can be either name, phone number, birthday or some meaningful text which can be easily guessed because of the low entropy. To successfully launch off-line identity guessing attack, an attacker has to keep control server's secret information $H^2(x)$ and $H(x)$ which can easily obtain under Assumption 3. After that, off-line identity guessing attack can be launched successfully as follows.

Step 1. From login phase of the protocol, an attacker can derives $N_{1a} = M_i \oplus H^2(x)$;

Step 2. Attacker computes $T_a = V_i \oplus H(x) = y_i \oplus ID_i$. So $y_i = T_a \oplus ID_i$;

Step 3. Then, Attacker computes $Z_a = CID_i \oplus V_i \oplus N_{1a} = y_i \oplus H(y_i) = T_a \oplus ID_i \oplus H(T_a \oplus ID_i)$;

Step 4. Now, attacker guess user's identity ID_i^{guess} separately and verifies the correctness $Z_a = T_a \oplus ID_i^{guess} \oplus H(T_a \oplus ID_i^{guess})$;

Step 5. Continue the above step until correct identity is obtained. After some guessing attacker can easily find the correct user identity. Thus, an attacker can successfully launch off-line identity guessing attack.

3.2 Off-line Password Guessing Attack

After launching successfully off-line identity guessing attack, an attacker can easily guess user's password from the smart card parameters Z_i in following steps:

Step 1. Attacker chooses PW_i^{guess} for the user U_i to find the correct password PW_i.

Step 2. Attacker then verifies the correctness of $Z_i = H(ID_i \parallel PW_i^{guess}) \oplus H^2(x)$ where $H^2(x)$ is known parameter to the attacker.

Step 3. The above steps will continue until the correct password obtained. After some guessing the attacker can easily find out the correct password. Thus, Sood's scheme can not resists off-line password guessing attack.

3.3 Privileged Insider Attack

Generally, many users use the same password for their convenience of remembering and easy of use whenever required. However, if the system manager or privileged insider of the server knows user's password, he/she may try to access user's U_i other accounts in other server. In Sood's scheme, user U_i provides his/her password PW_i to the remote server. As a result, Sood's scheme is insecure against insider attack, because system manager or privileged insider of the server may try to access the user's other accounts in other server using password PW_i.

3.4 User Impersonation Attack

To impersonate as a legitimate user, an attacker attempts to make a forged login request message which can be authenticated to a server. Under our assumption, Sood's scheme can not resist user impersonation attack as follows.

Step 1. Attacker can compute $y_i = V_i \oplus ID_i \oplus H(x)$. Then, attacker can easily compute $H(y_i)$.

Step 2. Now, Attacker generates a random number N_a and can easily compute login message $CID_i^* = V_i \oplus y_i \oplus N_a$, $M_i^* = H^2(x) \oplus N_a$ and $E_i^* = H(y_i \parallel H(x) \parallel N_a \parallel ID_i \parallel SID_k)$. Then attacker sends $\{CID_i^*, M_i^*, E_i^*\}$ to the service provider server S_k to proof himself as a valid user.

Step 3. It can be easily proved that the sending login message by an attacker is valid to the service provider server S_k. Then service provider server sends reply messages F_i and T_i to the attacker by computing $F_i = H[H(N_a \oplus N_2 \oplus N_3) \parallel ID_i \parallel H(y_i)]$ and $T_i = N_2 \oplus N_3 \oplus H(y_i \parallel ID_i \parallel H(x) \parallel N_a)$, where N_2 and N_3 is random number chosen by service provider server S_k and control server CS respectively.

Step 4. After receiving reply message from service provider server S_k, attacker computes $N_2 \oplus N_3 = T_i \oplus H(y_i \parallel ID_i \parallel H(x) \parallel N_a)$. Then, attacker and service provider server agree on the valid session key by computing $SK = H(ID_i \parallel (N_1 \oplus N_2 \oplus N_3) \parallel H(y_i))$ which is used for secure communication.

3.5 Many Logged In Users' Attack

Many logged in users' attack can be successfully launched after successful performance of off-line identity guessing attack and off-line password guessing attack as described in Section 3. After getting correct password of user U_i, an attacker can easily compute the value y_i which is different for all users. Then, attackers or non-registered user can successfully access the service of the service provider server S_k as follows.

Step 1. Attacker or non-registered user choose his/her desired password PW_i^a and computes $Z_i^a = H(ID_i \parallel PW_i^a) \oplus H^2(x)$, $V_i^a = y_i \oplus ID_i \oplus H(x)$ and $B_i^a = H(ID_i \parallel PW_i^a) \oplus P_i \oplus y_i$, where ID_i is the valid user's identity which remains unchanged.

Step 2. Then, attacker or non-registered users stores $\{Z_i^a, V_i^a, B_i^a, H(\cdot)\}$ into memory of new smart card and it can be used by many attackers or non-registered users as a valid user.

The above attack proves that Sood's scheme can not be used for practical implementation in terms of security because without stealing user's smart card, many non-registered users can act as a valid user.

3.6 Session Key Recovery Attack

In Sood's scheme, user's U_i, service provider server S_k and control server CS agree on the common session key SK which is based on the difficulty of cryptographic one-way hash function. The Common secret session key SK depends on the secret parameter ID_i, y_i and random nonce N_1, N_2, N_3. In the user impersonation attack in Section 3 shows that an attacker can easily obtain ID_i, y_i and random nonce N_1, N_2, N_3. So, after obtaining all these secret parameters attacker can compute secret session key for every transaction of user U_i. As a result, Sood's scheme is insecure against session key recovery attack.

4 Brief Review of Li et al.'s Scheme

This section presents brief description of Li et al.'s [13] dynamic ID based remote user authentication scheme in multi-server environment using smart card. Li et al.'s [13] scheme consists of following phases: Registration phase, Login phase, Authentication and Session Key Agreement phase.

4.1 Registration Phase

Whenever a new user wants to get services from the remote server, he/she must have to register with control server CS as follows:

User U_i chose his/her desired identity ID_i and password PW_i and generates a random nonce b. Then, U_i computes $PWB_i = H((ID_i \parallel PW_i) \oplus b)$ and sends ID_i, PWB_i to the control server CS through secure channel. After receiving registration messages from user U_i, CS first verifies U_i's personal information and credit and if it is valid then computes $TID_i = (T_i \parallel ID_i)$, $\sigma_i = H(TID_i \parallel x) \oplus H((ID_i \parallel PW_i) \oplus b)$, where T_i is the registration time of user U_i. Afterward, CS stores $\{\sigma_i, H(TID_i), T_i, H(\cdot)\}$ into memory of the smart card and issues it for the user U_i. After getting smart card, user U_i stores b into memory of the smart card and keeps it secret for personal use.

4.2 Login Phase

Whenever user U_i wants to get service from server S_k, then user U_i inserts his/her smart card into the card reader and submits ID_i and password PW_i and chooses the identity of service provider server SID_k. Then, smart card computes $TID_i^* = (T_i \parallel ID_i)$, $PWB_i = H((ID_i \parallel PW_i) \oplus b)$ and checks whether $TID_i^* = TID_i$ or not. If it does not hold, the smart card terminates this login, otherwise generates a random number N_1 and computes $\alpha_1 = \sigma_i \oplus PWB_i \oplus N_1$, $\alpha_2 = H((TID_i \parallel SID_k) \oplus N_1)$. Then, sends $\{TID_i,\ \alpha_1,\ \alpha_2\}$ to the service provider server S_k.

After receiving login messages $\{TID_i,\ \alpha_1,\ \alpha_2\}$ from user U_i, server S_k computes $\beta_1 = H(SID_k \parallel x) \oplus N_2$ and $\beta_2 = H((SID_k \parallel TID_i) \oplus N_2)$, where N_2 is the random number generated by service provider server S_k. Then, S_k sends $\{TID_i,\ \alpha_1,\ \alpha_2,\ SID_k,\ \beta_1,\ \beta_2\}$ to the control server CS through public channel.

4.3 Authentication and Session Key Agreement Phase

After receiving login request messages $\{TID_i,\ \alpha_1,\ \alpha_2,\ SID_k,\ \beta_1,\ \beta_2\}$ from server S_k, control server CS checks the validity of user's TID_i and server's SID_k. If both does not hold, rejects the connection, otherwise CS computes $N_1^* = \alpha_1 \oplus H(TID_i \parallel x)$ and verifies the freshness of N_1^*. If it does hold, CS further computes $\alpha_2^* = H((TID_i \parallel SID_k) \oplus N_1^*)$ and further compares with computed α_2^* equals with received α_2. if it is holds, then CS believes that user U_i is authentic, otherwise terminates the connection. Then, CS computes $N_2^* = \beta_1 \oplus H(SID_k \parallel x)$ and checks the freshness of N_2^*. If it does hold, CS further computes $\beta_2^* = H((SID_k \parallel TID_i) \oplus N_2^*)$ and further compares with computed β_2^* equals with received β_2. If it holds, then CS believes that service provider server S_k is authentic, otherwise terminates the connection. CS generates a random number N_3 and computes $\alpha' = H(N_1^*) \oplus N_2^* \oplus N_3$, $\gamma_u = H(H(TID_i \parallel x) \oplus SK)$, $\beta' = H(N_2^*) \oplus N_1^* \oplus N_3$ and $\gamma_s = H(H(SID_k \parallel x) \oplus SK)$, where SK is a common secret session key which is constructed by computing $SK = H(N_1^* \oplus N_2^* \oplus N_3)$. Finally, CS sends $\{\alpha',\ \gamma_u,\ \beta',\ \gamma_s\}$ to service provider server S_k.

After receiving the message from CS, server S_k computes $\beta'' = \beta' \oplus H(N_2)$, $SK_s = H(\beta'' \oplus N_2)$, $\gamma_s' = H(H(SID_k \parallel x) \oplus SK_s)$ and compares computed γ_s' with received γ_s. If it is invalid, server S_k terminates the connection, otherwise server S_k believes that control server CS is authentic and sends $\{\alpha',\ \gamma_u\}$ to the smart card user U_i. It can be easily shown that $SK_s = SK$ common secret session key between user U_i, server S_k and CS.

After receiving the response message from server S_k, smart card computes $\alpha'' = \alpha' \oplus H(N_1)$, $SK_u = H(\alpha'' \oplus N_1)$, $\gamma_u' = H(H(TID_i \parallel x) \oplus SK_u)$ and compares computed γ_u' with received γ_u. If it is not valid, terminates the connection, otherwise user believes that server S_k and control server CS is authentic participants. Finally, three participants user, service provider server and control server agree with a common secret session key $SK = SK_s = SK_u$ which can be used in future for secure communication.

5 Cryptanalysis of Li et al.'s Scheme

In this section, the cryptanalysis of Li et al.'s [13] scheme is presented. To analyze the security weaknesses of Li et al.'s scheme, we assume Assumptions 1 and 2 which are described in Section 3 of this paper.

5.1 Off-line Identity Guessing Attack

During the registration phase, user U_i usually chooses an identity which is easily remembered for his/her convenience. These easy to remember identities are low entropy and thus attacker can easily guess user's identity. Generally user's identity is static and often confined to a predefined format, so it is more easily guessed by the attacker than the password. Li et al.'s scheme suffers from identity guessing attack as follows:

Step 1. An attacker extracts information $H(TID_i)$, T_i from the valid user's smart card by monitoring power consumption.

Step 2. Then, attacker chooses user's identity ID_i^a and verifies the correctness $H(TID_i) = H(T_i \parallel ID_i^a)$.

Step 3. Continue step 2 until correct identity is obtained. After some guessing, an attacker can find out the correct user's identity ID_i.

Thus, Li et al.'s scheme is insecure against off-line identity guessing attack.

5.2 Off-line Password Guessing Attack

In remote user authentication schemes, for the sake of user-friendliness, a user is often allow to select his/her desired password during the registration phase. Generally, the user chooses his/her password which is easy to remember. But these easy to remember passwords are of low entropy and an attacker can guess the user's password. After launching successfully off-line identity guessing attack, an attacker can easily guess user's valid password using stored smart card's parameters σ_i, b and service provider server's login message $\{TID_i,\ \alpha_1,\ \alpha_2,\ SID_k,\ \beta_1,\ \beta_2\}$ as follows:

Step 1. Attacker computes $N_1 = \alpha_1 \oplus \sigma_i \oplus PWB_i$. Now,

$$\begin{aligned}\alpha_2 &= H((TID_i \parallel SID_k) \oplus N_1)\\ &= H((TID_i \parallel SID_k) \oplus \alpha_1 \oplus \sigma_i \oplus PWB_i)\\ &= H((TID_i \parallel SID_k) \oplus \alpha_1 \oplus \sigma_i\\ &\qquad \oplus H((ID_i \parallel PW_i) \oplus b)).\end{aligned}$$

Step 2. Now, Attacker chooses password PW_i^{guess} for user U_i to find the correct password PW_i. Then, attacker checks the correctness whether $\alpha_2 = H((TID_i \parallel SID_k) \oplus \alpha_1 \oplus \sigma_i \oplus H((ID_i \parallel PW_i^{guess}) \oplus b))$, where ID_i is the correct user identity by using identity guessing attack and all other parameters of α_2 is known to the attacker except password PW_i.

Step 3. An attacker then repeats the above process until the correct password is obtained. After some guessing, an attacker can find out the correct password. Thus, Li et al.'s scheme is vulnerable to off-line password guessing attack.

5.3 User Impersonation Attack

To impersonate as a legitimate user, an attacker attempts to make a forged login request message which can be authenticated to a server. Under our assumption, Li et al.'s scheme can not resist user impersonation attack as follows:

Step 1. Attacker can compute $PWB_i^a = H((ID_i \parallel PW_i) \oplus b)$, where ID_i, PW_i is the user's correct identity and password by using off-line identity and password guessing attack respectively. Attacker further computes $\alpha_1^a = \sigma_i \oplus PWB_i^{'} \oplus N_1^a$ and $\alpha_2^a = H((TID_i \parallel SID_k) \oplus N_1^a)$, where N_1^a is the random number generated by the attacker and attacker knows σ_i, b from user's smart card memory by monitoring power consumption.

Step 2. Then, attacker sends forged login message $\{TID_i,\ \alpha_1^a,\ \alpha_2^a\}$ to the service provider server S_k. It can be easily proved that the sending login message by an attacker is valid to the service provider server S_k. Then, service provider server S_k sends login message $\{TID_i,\ \alpha_1^a,\ \alpha_2^a,\ SID_k,\ \beta_1,\ \beta_2\}$ to the control server CS after computing β_1, β_2.

Step 3. After receiving login message from service provider server S_k, control server checks the validity of user's TID_i and server's SID_k. If both hold, CS computes, $N_1^a = \alpha_1^a \oplus H(TID_i \parallel x)$ and checks the freshness of N_1^*. If it holds, CS further computes α_2^* = $H((TID_i \parallel SID_k) \oplus N_1^a)$ and compares with computed α_2^* equals with received α_2^a. If it holds, then CS believes that the sending messages are authentic, otherwise terminates the connection. Then, CS sends $\alpha^{'}, \gamma_u$ to the smart card user U_i through service provider server S_k, where $\alpha^{'} = H(N_1^a) \oplus N_2^* \oplus N_3$ and $\gamma_u = H(H(TID_i \parallel x) \oplus SK)$.

Step 4. After receiving $\alpha^{'}, \gamma_u$ from CS through service provider server S_k, attacker derives $N_2^* \oplus N_3$ = $H(N_1^a) \oplus \alpha^{'}$ and computes session key $SK_a = H(N_1 \oplus N_2^* \oplus N_3)$ which is used for secure communication. Thus, Li et al.'s scheme is insecure against user impersonation attack.

5.4 Many Logged in Users' Attack

Many logged in users' attack can be successfully launched after successful performance of off-line identity guessing attack and off-line password guessing attack as described in Section 5. After getting correct password and identity of user U_i, attackers or non-registered user can successfully access the service of the server S_k as follows:

Step 1. Attacker can compute $PWB_i^a = H((ID_i \parallel PW_i) \oplus b)$, where ID_i, PW_i is the user's correct identity and password by using off-line identity and password guessing attack respectively. Then, computes $H(TID_i \parallel x) = \sigma_i \oplus PWB_i^a$.

Step 2. Now, Attacker chooses password PW_i^a and computes $\sigma_i^a = H(TID_i \parallel x) \oplus H((ID_i \parallel PW_i^a) \oplus b)$, where attacker keeps unchanged $H(TID_i), T_i$ which can be extracted from memory of smart card by monitoring power consumption.

Step 3. Then, attacker or non-registered users stores $\{\sigma_i^a, H(TID_i), T_i, H(\cdot)\}$ into memory of the smart card and it can be used by many attacker or non-registered users as a valid user.

Above attack proves that Li et al.'s scheme can not be used for practical implementation in terms of security. It is because, without stealing user's smart card, many non-registered users can acts as valid users.

6 Proposed Scheme

In this paper, We have shown that Sood's scheme and Li et al.'s scheme are insecure against various attacks. To overcome these weaknesses, in this section, we proposed an efficient dynamic identity based remote user authentication in multi-server environment using smart card. It can be assumed that control server CS is a trusted authority. The proposed scheme consists of four phases namely registration phase, login phase, authentication and session key agreement phase and password change phase. All these proposed phases are discussed as below:

6.1 Registration Phase

This phase is divided into two sub-phases: Server Registration phase and User Registration phase.

Server Registration Phase. In this phase, service provider server S_k selects his/her desired identity SID_k and submits it to control server CS over a secure channel. After receiving SID_k from S_k, CS computes $P_k = H(SID_k \parallel x)$ and sends it to the server S_k through secure channel and S_k keeps it as secret, where x is the secret key of control server CS.

User Registration Phase. Whenever a new user wants to get services from the service provider server, first he/she has to register with the control server CS. So, the user chooses his/her desired identity ID_i and password PW_i and generates a random nonce b. Afterwards, user computes $PWR_i = H(PW_i \oplus b)$, where $H(\cdot)$ is the secure one-way hash function like secure MD5 and sends ID_i, PWR_i to control server through secure channel for the registration. After receiving a registration message from user U_i, CS generates a random nonce y_i for each user U_i and computes $CID_i = H(ID_i \oplus y_i \oplus x)$ such that CID_i will be unique for each user U_i like bank account number. So, after computing CID_i for user U_i, CS checks whether the value of CID_i is exist in CS's database or not. If exists, CS chooses another random nonce y_i^* and computes again $CID_i = H(ID_i \oplus y_i^* \oplus x)$. CS again verifies whether CID_i is exist or not in the database. If exists, then again computes CID_i with the new random nonce until CID_i will be unique, otherwise control server CS computes $REG_i = H(ID_i \parallel PWR_i \parallel CID_i)$ and $T_i = H(CID_i \parallel x) \oplus PWR_i$ and issues a smart card for U_i after storing $\{CID_i,\ REG_i,\ T_i,\ y_i,\ H(\cdot)\}$ into memory of user's smart card. After getting smart card, user U_i stores b into memory of smart card and uses it securely for taking services from S_k.

6.2 Login Phase

Whenever an existing user U_i wants to get the service(s) from the server S_k, first inserts his/her smart card into the card reader and submits ID_i^* and PW_i^*; and chooses server identity SID_k. Then, card reader computes $PWR_i^* = H(PW_i^* \oplus b)$ and $REG_i^* = H(ID_i^* \parallel PW_i^* \parallel CID_i)$; and checks whether REG_i^* equals stored REG_i holds or not. If the verification holds, it implies $ID_i^* = ID_i$ and $PW_i^* = PW_i$. Then, smart card derives $L_1 = T_i \oplus PWR_i^*$ and generates random numbers N_1, N_2 and further computes $N_3 = N_1 \oplus N_2$, $L_2 = N_2 \oplus PWR_i^*$ and $L_3 = H(L_1 \parallel SID_k \parallel N_1 \parallel L_2 \parallel N_3)$ and sends login request message $\{CID_i,\ SID_k,\ T_i,\ L_3,\ L_2,\ N_3\}$ to control server CS.

6.3 Authentication Phase

After receiving the login request message $\{CID_i,\ SID_k,\ T_i,\ L_3,\ L_2,\ N_3\}$, control server first checks the format of CID_i and SID_k. If it is valid then computes $A_1 = H(CID_i \parallel x)$ and derives $PWR_i^{'} = T_i \oplus A_1$, $N_2^{'} = L_2 \oplus PWR_i^{'}$ and $N_1^{'} = N_3 \oplus N_2^{'}$. Further computes $L_3^{'} = H(A_1 \parallel SID_k \parallel N_1^{'} \parallel L_2 \parallel N_3)$ and verifies whether computed $L_3^{'}$ equals with received L_3. If it does not hold, CS terminates the session, otherwise CS believes that the user U_i is authentic and also believes that SID_k is the registered identity of service provider server S_k. Then, control server generates a random number N_4 and computes $A_2 = H(SID_k \parallel x)$, $A_3 = A_2 \oplus N_4$, $N_5 = N_1{'} \oplus N_4$ and $A_4 = H(A_2 \parallel N_4 \parallel N_1 \parallel CID_i)$. Then, CS sends $\{CID_i,\ A_4,\ A_3,\ N_5\}$ to the service provider server S_k of the corresponding identity SID_k through public channel.

After receiving messages $\{CID_i,\ A_4,\ A_3,\ N_5\}$ from CS, server S_k derives $N_4^{'} = P_k \oplus A_3$ and $N_1^{'} = N_4^{'} \oplus N_5$ and computes $A_4^{'} = H(P_k \parallel N_4^{'} \parallel N_1^{'} \parallel CID_i)$. Then, server S_k compares $A_4^{'}$ with received A_4. This equivalency authenticates the legitimacy of the control server CS and user U_i. Further, server S_k generates random number N_6 and computes $N_7 = N_1^{'} \oplus N_6$, $SK_s = H(SID_k \parallel CID_i \parallel N_6 \parallel N_1^{'})$, $A_5 = H(SK_s \parallel N_6)$ and sends $\{SID_k,\ A_5,\ N_7\}$ to the smart card user U_i through public channel.

After receiving messages $\{SID_k,\ A_5,\ N_7\}$ from the server S_k, the smart card derives $N_6^{'} = N_7 \oplus N_1$ and computes $SK_u = H(SID_k \parallel CID_i \parallel N_6^{'} \parallel N_1)$, $A_5^{'} = H(SK_u \parallel N_6^{'})$. Then, smart card compares computed $A_5^{'}$ equals with received A_5. This equivalency authenticates the legitimacy of the service provider server S_k. It can be easily shown that $SK_u = SK_s$ which is the common secret session key between user U_i and service provider server S_k.

6.4 Password Change Phase

This phase invokes when user U_i wants to change his/her password. U_i inserts the smart card into the card reader and submits ID_i^* and PW_i^*. Then, card reader computes $PWR_i^* = H(PW_i^* \oplus b)$ and $REG_i^* = H(ID_i^* \parallel PW_i^* \parallel CID_i)$; and checks whether REG_i^* equals stored REG_i holds or not. If it holds positively, U_i enters a new password PW_i^{new}. Then card reader computes $PWR_i^{new} = H(PW_i^{new} \oplus b)$, $REG_i^{new} = H(ID_i^* \parallel PWR_i^{new} \parallel CID_i)$ and $T_i^{new} = T_i \oplus PWR_i^* \oplus PWR_i^{new}$ and stores REG_i^{new}

and T_i^{new} instead of REG_i and T_i respectively into the memory of the smart card. Thus, U_i can change the password without taking any assistance from control server or service provider server S_k.

7 Cryptanalysis of The Proposed Scheme

This section describes cryptanalysis of the proposed scheme. To cryptanalyze the proposed scheme, it can be assumed that an attacker could obtain the secret values stored in the smart card by monitoring the power consumption [12, 17] and can intercept all communicating messages between the user, service providing server and the control server. Under these assumption, we will show that the proposed scheme resists different possible attacks related to remote user authentication.

7.1 Off-line Identity Guessing Attack

After getting secret values $\{CID_i,\ REG_i,\ T_i,\ y_i,\ H(\cdot)\}$ from user's smart card memory and login request message $\{CID_i,\ SID_k,\ T_i,\ L_3,\ L_2,\ N_3\}$, an attacker attempts to derive or guess user's identity ID_i. To obtain user's correct identity ID_i from CID_i, attacker has to guess x and ID_i simultaneously which is not possible in polynomial time, where x is the secret key of the control server. So, the proposed scheme resists off-line identity guessing attack.

7.2 Off-line Password Guessing Attack

After getting secret values $\{CID_i,\ REG_i,\ T_i,\ y_i,\ H(\cdot)\}$ from user's smart card memory and login request message $\{CID_i,\ SID_k,\ T_i,\ L_3,\ L_2,\ N_3\}$, an attacker attempts to derive or guess user's password PW_i in off-line mode. To get user's correct password, attackers has to guess either two secret parameters at a time which is not possible in polynomial time or has to solve inversion of cryptographic hash function which is also computationally hard. So, the proposed scheme is secure against off-line password guessing attack.

7.3 Privileged Insider Attack

The proposed scheme is secure against privileged insider attack because, user U_i provides PWR_i which equals with $H(PW_i \oplus y)$ instead of PW_i to the control server CS. As a result, system manager or privileged insider of the server can not derive valid user's password. So, the proposed scheme resists privileged insider attack.

7.4 User Impersonation Attack

To impersonate as a legitimate user, an attacker attempts to make a forged login request message which can be authenticated to a server. However, the attacker cannot impersonate as the legitimate user by forging the login request message even if the attacker can extract the secret values $\{CID_i,\ REG_i,\ T_i,\ y_i,\ H(\cdot)\}$ stored in the users smart card, because the attacker cannot compute the valid login request message $\{CID_i,\ SID_k,\ T_i,\ L_3,\ L_2,\ N_3\}$ without knowing the secret password PW_i of valid user U_i, control server secret key x and valid user identity ID_i. If the attacker wants to get these secret parameters, he/she must have to solve the inversion of cryptographic hash function which is computationally hard. So, the proposed scheme is secure against user impersonation attack.

7.5 Many Logged-in Users' Attack

The proposed scheme is secure against many logged-in users' attack because even if an attacker gets user's smart card then he/she has no way to derive or guess user's correct password PW_i, user's identity ID_i and server secret key x as described in Section 7. If the attacker wants to get the control server secret key x, user's password PW_i and user's identity ID_i, he/she must have to solve the inversion of cryptographic one-way hash function which is computationally hard. So the proposed scheme resists many logged-in users' attack.

7.6 Smart Card Stolen Attack

We assume that the user U_i has either lost his/her smart card or stolen by an attacker. After getting the smart card, an attacker can extract the secret information $\{CID_i,\ REG_i,\ T_i,\ y_i,\ H(\cdot)\}$ from the user's smart card. We also assume that attacker stores the $i-th$ login message $\{CID_i,\ SID_k,\ T_i,\ L_3,\ L_2,\ N_3\}$ of the user U_i. After getting all these parameters such as login message and smart card parameters, it is hard to derive user's password PW_i, identity ID_i and server secret key x by the attacker. As a result, attacker can not create the valid login message even after getting the valid user's smart card parameters. So, the proposed scheme is secure against smart card stolen attack.

7.7 Session Key Recovery Attack

In the proposed scheme, session key depends upon the difficulty of cryptographic one-way hash function and the random number N_1 and N_6. There is no way for an attacker to compute random number N_1 and N_6 from the known parameters that is from all communicating message of the proposed scheme. So the proposed scheme resists session key recovery attack.

Table 2: Comparison of computation cost of proposed scheme with related schemes

	[19]	[13]	Proposed protocol
Login Phase	$4T_h$	$3T_h$	$3T_h$
Authentication Phase	$17T_h + 1T_m$	$15T_h$	$9T_h$
Total	$21T_h + 1T_m$	$18T_h$	$12T_h$

Table 3: Comparison of communication and storage cost of proposed scheme with related schemes

	[19]	[13]	Proposed protocol
Storage Cost	512 bits	640 bits	768 bits
Communication Cost	1920 bits	1920 bits	1664 bits

Table 4: Security attack comparison of the proposed scheme with related schemes

	[19]	[13]	Proposed protocol
Off-line Identity Guessing Attack	YES	YES	NO
Off-line Password Guessing Attack	YES	YES	NO
Privileged Insider Attack	YES	NO	NO
User Impersonation Attack	YES	YES	NO
Many Logged In Users' Attack	YES	YES	NO
Session Key Recovery Attack	YES	NO	NO

8 Performance Analysis of the Proposed Scheme

In this section, we evaluated the performance of proposed scheme comparing with both the Sood's scheme and Li. et al's scheme. We have compare login and authentication phases of proposed scheme with both Sood's scheme and Li et al's scheme, because these phases are used frequently. Table 2 shows the computation over head and Table 3 shows the communication and storage cost of proposed scheme and both the related [13, 19] scheme. In Table 2, T_h is the time required for hashing operation and T_m is the time required for multiplication operation. Though, proposed scheme resists different possible attacks of both the related schemes, in spite of the proposed scheme which provides better computation cost than the related schemes.

It can be reasonably assumed that the length of ID_i, PW_i, SID_j, $h(\cdot)$ and random nonce returns 128 bits. The communication cost (capacity of transmitting message) of proposed scheme, Sood's [19] scheme and Li et al.'s [13] scheme are 1664 $bits = (13 \times 128)$, 1920 $bits = (15 \times 128)$ and 1920 $bits = (15 \times 128)$ respectively for each transaction. Also the storage cost (stored into the memory of smart card) takes almost same bits of proposed scheme and related schemes that is 768 *bits*, 512 *bits* and 512 *bits* respectively. Table 4 shows that their scheme is insecure against different possible attacks. Further proposed scheme provides strong authentication against different attacks described in Section 7. After resisting all possible attacks of related scheme, the proposed scheme provides low computational and communication cost than others related schemes. Hence the proposed scheme is more efficient and secure than both Sood's scheme and Li. et al's scheme.

9 Conclusion

We have shown that both Sood's [19] and Li et al.'s [13] schemes have security weaknesses described in Section 3 and Section 5 respectively. To overcome these weaknesses, we have proposed an Efficient Dynamic ID Based Remote User Authentication Scheme in Multi-server Environment using smart card. Further, we have shown that the proposed scheme using smart card which is more efficient in terms of computational and communication cost than related schemes. Additionally, the proposed scheme provides password change phase without taking any assistance of the control server and also provides strong mutual authentication. Cryptanalysis of the proposed scheme shows that the authentication system is more authentic, secure and efficient than related schemes published earlier. In future, we can incorporate biometric features with password to provide high security system and also try to analyze the security analysis of the proposed protocol using BAN logic.

References

[1] R. Amin, T. Maitra, D. Giri, "An improved efficient remote user authentication scheme in multi-server environment using smart card", *International Journal of Computer Applications*, vol. 69, no. 22, pp. 1–6, 2013.

[2] C. C. Chang and T. F. Cheng, "A robust and efficient smart card based remote login mechanism for multi-server architecture", *International Journal of Innovative Computing, Information and Control*, vol. 7, no. 8, pp. 4589–4602, 2011.

[3] C. C. Chang, J. S. Lee, "An efficient and secure multi-server password authentication scheme using

smart cards", in *Proceedings of the International Conference on Cyberworlds*, pp. 417–422, 2004.

[4] Te-Yu Chen, C. C. Lee, M. S. Hwang, J. Ke Jan, "Towards secure and efficient user authentication scheme using smart card for multi-server environments," *The Journal of Supercomputing*, vol. 66, no. 2, pp. 1008–1032, Nov. 2013.

[5] T. H. Feng, C. H. Ling, and M. S. Hwang, "Cryptanalysis of Tan's improvement on a password authentication scheme for multi-server environments," *International Journal of Network Security*, vol. 16, no. 4, pp. 318–321, 2014.

[6] W. Ford and B. S. Kaliski, "Server-assisted generation of a strong secret from a password", in *Proceding of IEEE 9th International Workshop Enabiling Technology*, pp. 176–180, Washington, June 2000.

[7] D. He, W. Zhao, and S. Wu, "Security analysis of a dynamic id-based authentication scheme for multi-server environment using smart cards", *International Journal of Network Security*, vol. 15, no. 5, pp. 350–356, 2013.

[8] H. C. Hsiang and W. K. Shih, "Improvement of the secure dynamic id based remote user authentication scheme for multi-server environment", *Computer Standards and Interface*, vol. 31, no. 6, pp. 1118–1123, 2009.

[9] L. Hu, X. Niu, and Y. Yang, "An efficient multi-server password authenticated key agreement scheme using smart cards", in *Proceedings of International Conference on Multimedia and Ubiquitous Engineering (MUE'07)*, pp. 903–907, Apr. 2007.

[10] D. P. Jablon, "Password authentication using multiple servers", in *Proceedings of RSA Security Conference*, pp. 344–360, London, Apr. 2001.

[11] W. S. Juang, "Efficient multi-server password key agreemented key exchange using smart cards", *IEEE Transactions on Consumer Electronics*, vol. 50, no. 1, pp. 251–255, 2004.

[12] P. Kocher, J. Jaffe, and B. Jun, "Differential power analysis", in *Proceedings of Advances in Cryptology*, pp. 388–397, 1999.

[13] C. T. Li, C. Y. Weng, and C. I Fan, "Two-factor user authentication in multi-server networks", *International Journal of Security and Its Applications*, vol. 6, no. 2, pp. 261–267, 2012.

[14] H. Li, I. C. Lin, M. S. Hwang, "A remote password authentication scheme for multi-server architecture using neural networks", *IEEE Transactions on Neural Networks*, vol. 12, no. 6, pp. 1498–1504, 2001.

[15] Y. P. Liao and S. S. Wang, "A secure dynamic ID based remote user authentication scheme for multi-server environment", *Computer Standard and Interface*, vol. 31, no. 1, pp. 24–29, 2009.

[16] I. C. Lin, M. S. Hwang and L. H. Li, "A new remote user authentication scheme for multi-server archicture", *Future Generation Computer System*, vol. 19, no. 1, pp. 13–22, 2003.

[17] T. S. Messerges, E. A. Dabbish, and R. H. Sloan, "Examining smart-card security under the threat of power analysis attacks", *IEEE Transactions on Computers*, vol. 51, no. 5, pp. 541–552, 2002.

[18] R. L. Rivest, A. Shamir, and L. Adleman, "A method for obtaining digital signatures and public-key cryptosystems", *Communications of the ACM*, vol. 21, no. 2, pp. 120–126, Feb. 1978.

[19] S. K. Sood, "Dynamic identity based authentication protocol for two-server architecture", *Journal of Information Security*, vol. 3, pp. 326–334, 2012.

[20] S. K. Sood, A. K. Sarje, and K. Singh, "A secure dynamic identity based authentication protocol for multi-server architecture", *Journal of Network and Computer Applications*, vol. 34, no. 2, pp. 609–618, 2011.

[21] J. L. Tsai, "Efficient multi-server authentication scheme based on one-way hash function without verification table", *Computers and Security*, vol. 27, pp. 115–121, 2008.

Ruhul Amin received his B.Tech and M.Tech Degree from West Bengal University of Technology in computer science and engineering department in 2009 and 2013 respectively. He has qualified GATE in 2011 in computer science. Currently, he is a lecturer of a polytechnic college. He has published three (3) international journal on the topic of remote user authentication. His research interest are remote user authentication and security in wireless sensor network.

Malicious Behavior Analysis for Android Applications

Quan Qian, Jing Cai, Mengbo Xie, Rui Zhang
(Corresponding author: Quan Qian)

School of Computer Engineering & Science, Shanghai University
99 Shangda Rd., Baoshan District, Shanghai, China
(Email: qqian@shu.edu.cn)

(Received Oct.31, 2013; revised and accepted Jan. 10 & Mar. 4, 2015)

Abstract

Android, as a modern popular open source mobile platform, makes its security issues more prominent, especially in user privacy leakage. In this paper, we proposed a two-step model which combines static and dynamic analysis approaches. During the static analysis, permission combination matrix is used to determine whether an application has potential risks. For those suspicious applications, based on the reverse engineering, embed monitoring Smali code for those sensitive APIs such as sending SMS, accessing user location, device ID, phone number, etc. From experiments, it shows that almost 26% applications in Android market have privacy leakage risks. And our proposed method is feasible and effective for monitoring these kind of malicious behavior.

Keywords: Android security, malicious behavior monitoring, permissions filtering, privacy leakage

1 Introduction

With the rapid development of network technology, the mobile Internet has been the development trend of the information age. According to the market research company Canalys released data, the global intelligent mobile phone shipments in 2011 has outpaced PC, reached 487,700,000 [6]. About the proportional share of the smartphone, Android OS has been in a rising trend. The first quarter shipments report of smartphone from Canalys showed that Android OS reached 75.6%, and there has been some increase compared with 69.2% of the previous quarter [5]. With the popularity and rapid development of Android OS, its security issues are also increasingly prominent. For instance, the security report from NetQin Company shows that they detected more than 65,227 new malware in 2012, a 263% increase over 2011. And the vast majority of malicious software is designed to attack Android and Symbian devices. Moreover, Android devices accounted for the number of devices being attacked 94.8%, and software for the purpose of stealing user's privacy data reached as high as 28%, ranking first in all types of malicious behavior [18].

The main purpose of this paper is to analyze the Android applications accurately and comprehensively based on combining static and dynamic method to reveal the malicious behaviors of applications leaking user's privacy data. Privacy leakage mentioned in this paper refers to Android applications using sensitive permissions granted by user during the installation to collect user's privacy data, including user's device ID, IMEI, phone number, contacts, call records, location information, etc., and send user's privacy data via SMS or network.

Currently, the method for detecting user privacy data leakage in intelligent mobile phone platform mainly has two categories, static and dynamic. Static analysis methods mainly focused on the control flow, data flow and structural analysis [15]. But Android application mostly written with Java, the program will inevitably exist a large number of implicit function calls, and the static analysis methods cannot effectively handle it. At the same time, static method can obtain the concrete execution path of the application without executing the source code, but it does not determine whether the path will actually be performed, which can only be verified by dynamic method. Concerning about the dynamic method, there are traditional sandbox and dynamic taint tracking technology. Sandbox technology is a kind of isolated operating mechanism, is currently widely used in software testing, virus detection and other software security related areas [2]. Some background research on Android security is briefly introduced as follows.

Kui Luo proposed an byte code converter for malicious code of leakage privacy, converting DVM (Dalvik Virtual Machine) byte code into Java code, and putting the Java code into the Indus (a static analysis of Java code and slice tool) to analyze [17]. Leonid Batyuk proposed a method by decompiling sample applications, in the premise of not affecting the program core function, through modifying the binary code to separate the malcode [1]. Although this method can analyze the sample malcode effectively, it is unsatisfactory when the target program has been obfuscated. Enck implemented a Dalvik decompiler, DED, by using the static analysis package tool. The tool use Fortify SCA (a kind of white-box source code security testing software) to analyze the application's control flow, data flow, structure and semantics [19]. Qian et al also depends on Dalvik decompiling and gives a basic two-step-famework for Android malware behavior monitoring [22]. ComDroid analyzes the DEX byte code disassembled by Dedexer, and checks the Intent creation and transmission to identify the program broadcast hijacking vulnerabilities [7]. ScanDroid extracts the security specification from configuration files of Android application and checks the consistency between the application data flow and the specification [10]. ScanDroid is based on the WALA analysis framework, can only evaluate the open source applications.

Static analysis method can help to identify Android applications that applied unnecessary extra permissions or opened some interfaces for outer access without any protection. However, this method is easily confused by a variety of technologies, for instance obfuscation. While dynamic technology can make up for this. Enck proposed TaintDroid [8], which is a tracking framework to detect privacy leakage using dynamic taint. It modified the DVM layer of Android to complete the function of tainting data, and add Hooks to API interface of tainted sources to achieve infecting the private data accessed by applications, and finally got private data leakage through monitoring the socket of the network interface. In [3], it places a LKM module (Loadable Kernel Module) in the Android simulator, builds a sandbox system, intercepts and records all the underlying system calls from the kernel layer. However, modifying Linux kernel will lead to the Android emulator running extremely unstable, and the paper only uses automatic tools to simulate user interactions, not used in actual environment. Isohara uses a kernel based behavior analysis which depends on a log collector in the Linux layer of Android device and logs analysis application deployed on remote server [16]. However, the behavior understandability based on kernel level log is not good. Also, the server side uses regular expressions based signatures to detect the malware, but the signature maintenance and definition are difficult. Similarly, Peng et al. analyzes Binder IPC data from the server side to identify behaviors of different applications. By calculating the TP (threat point) value of different applications to evaluate whether an application is a malware or not [21]. However, Binder IPC based monitoring is a kernel based method and the TP threshold is hard to pre-defined. Asaf Shabtai designs SELinux [24] and deploys in Android system, which makes up for the defects of high level process and the experiments in HTC G1 show the feasibility of running SELinux in Android. However, SELinux policy maintenance is relatively cumbersome, and mobile phones with limited computing and storage capacity are not very suitable for deploying this kind of system.

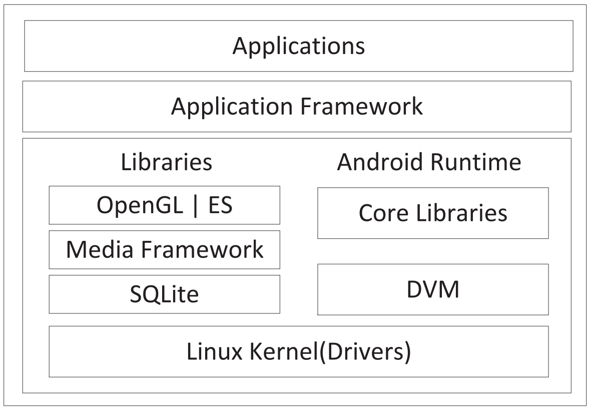


Figure 1: A brief architecture of Android system

The organization of the paper are as follows: Section 2 is about the Android basic framework and its security issues. Section 3 is the security analysis mechanism we proposed for android applications. Experiments are shown in Section 4. Section 5 summarizes the whole paper including the contributions and some future work.

2 Android Security Related Issues

2.1 Android Basic Architecture

Android, as a mobile operating system platform based on Linux kernel, was developed by the Google Open Handset Alliance [12]. Android has a layered architecture, including the Linux kernel layer, middleware layer and application layer, which can provide a uniform service for the upper layer, masks the differences of the current lay and lower layer [12]. The core functions of a smartphone are provided by the middleware layer, implemented by Java or C/C++. Applications running on Android are written in Java, and then multiple *.class* files are converted to *.dex* format by the Android DX tool. Each Android application is as a separate instance to run in DVM, and has a unique process identification number. Figure 1 gives a brief architecture of an Android system.

Among different components of Android, DVM [4], is the core part of Android platform. It can support Java applications, which are converted to *.dex* (Dalvik Executable) format. The *.dex* format is designed for a compressed format of Dalvik, suitable for memory and processor speed limited system. Dalvik is responsible for process isolation and threads management. Each Android application corresponds to a separate instance of Dalvik virtual machine, and can be executed in a virtual machine under its interpretation.

2.2 Android Security Mechanisms

Android security mechanisms are similar to Linux [13]. Android itself provides a series of mechanisms for the protection of privacy data. The core of Android security mechanism mainly includes the sandbox, application signature and permission mechanism. The permission mechanism limits applications to access user's privacy data (i.e. telephone numbers, contacts etc.), resources (i.e. log files) and system interface (i.e. Internet, GPS etc.). In permission mechanism, the phone's resources are organized by different categories, and each category corresponds to one kind of accessed resource. If an application requires access to certain resources, it needs to have the corresponding permissions. Android permission mechanism is coarse-grained and belongs to a kind of stated permissions before installing. Although this mechanism is simple, it also has some defects that cannot protect the user's privacy information adequately. Early in the conference of the ACSAC in 2009, Ontang et al questioned Android security model, and pointed out that the current Android permissions model cannot meet certain security requirements [20]. Enck proposed Kirin [9], a detection tool, to enhance existing Android permissions model. Based on a set of policy, Kirin is used to determine whether to grant the requested permissions to applications, and through the analysis of the Android application's Manifest file to ensure the granted permission in accordance with system strategy. Android permissions mechanism is coarse-grained and inflexible [13]. The application required permissions must be granted all before installed and cannot be changed after installation. This permission model leads to certain potential security threats. On the one hand, permissions to access private data will be decided by users. For those non-security awareness users, the permission granting process is casual and blind. During the installation phase, if the program obtains permissions to access privacy information, then can be arbitrary abuse of user privacy sensitive data at any time; On the other hand, the mechanism cannot effectively prevent permission elevation attacks. Applications can take advantage of a combination of permissions to steal the user's sensitive data.

In order to reveal Android apps leaking user privacy information behavior, according to the Android OS security mechanism, this paper proposed a malicious behavior analysis model combining the dynamic and static method, which will be discussed in detail in the next sections.

3 Android Malicious Behavior Analysis Framework

Generally speaking, methods for malware analysis mainly include static and dynamic approach. Static analysis is a kind of method based on program's source code. It has the advantages of being wide coverage and can analyze the source code comprehensively. However static method is based on source code. And if we cannot get the target source code, through decompiling or reverse engineering, it is hard to analyze the program accurately, especially in the occasion that the target program has been obfuscated. Dynamic analysis refers to the tracking and monitoring its run-time behavior through running the program. This kind of method is more accurate for capturing the actual malicious program behavior. Meanwhile, the dynamic method has its own disadvantages because of its limited execution coverage, that is to say we cannot guarantee all of the running paths have been trigged during the test.

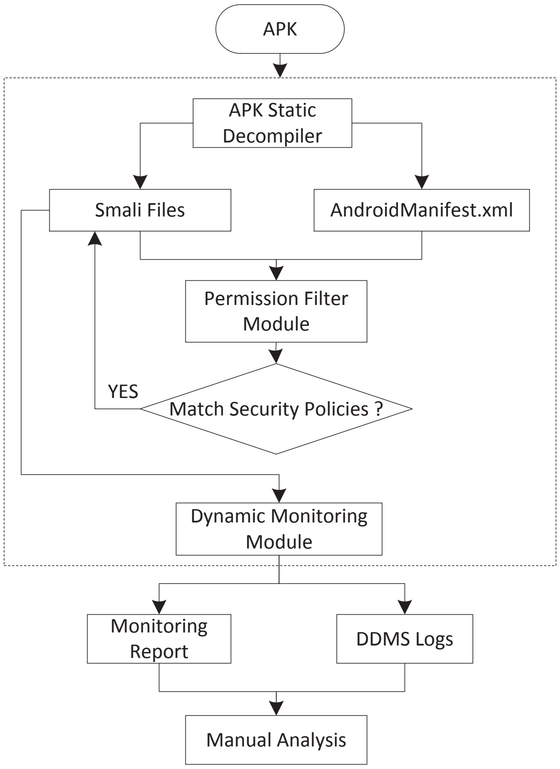


Figure 2: Malicious behavior analysis framework for android App

In this paper, we present a combination of static and dynamic security analysis model that can make up for their shortcomings with each other, enable the analysis of malicious behavior more comprehensively and accurately. Figure 2 shows the whole steps.

Before analyzing the Android application, APK (android application package) needs to be statically decompiled to get the corresponding configuration and Smali [14] files. Among them, the configuration file with the format of AndroidManifest.xml is mainly used for permissions filtering stage, and the Smali files are mainly applied to dynamic monitoring module. First of all, we choose those suspicious applications with great potential to leak user's privacy. Then if a program is suspicious, enter into the dynamic monitoring module, where input the target Smali codes, embed some tracking code, repackage and re-sign the APK. In future, once the APK is running, we can dynamically monitor the behavior of privacy leakage and give immediate alarm for users. And those alerts or logs can be used for further detailed analysis manually or automatically. Next, we will discuss the three core components of the framework: APK Static Decompiler, Permission Filtering Module and Dynamic Monitoring Module.

3.1 APK Static Decompiler

Before permission filtering and dynamic monitoring, we need to extract the Android application's AndroidManifest.xml file and Smali files corresponding to the target APK. The Android application is an installation package ended with suffix *.apk* (an acronym for Android Package). APK is similar to *.exe* executable file in PC, after installed can be executed in Android OS immediately. APK is actually a compressed file compliance with the ZIP format, which can be extracted by popular *.zip* compatible decompression tools. In addition, it must be noted that most applications are code-obfuscated, and the unzipped file is not able to analyze directly. It should be decompiled to extract its resource, permissions, the intermediate representation files. In this paper, we use the apktool [23] for decompiling. The file structure of Android application after decompiled is shown in Table 1.

Table 1: The file structure after APK decompiled

Directory/File	Description
res	Application's resource file, including pictures, sound, video and etc.
smali	Dalvik register bytecode files of APK
AndroidManifest.xml	The global configuration file of APK including the package name, permissions, referenced libraries and other related information of the application.
Apktool.yml	The configuration file of Apktool

3.2 Permission Filtering Module

Some permissions may not exist risks by itself, but if there are some permissions combined there may exist a security risk. For example, an application applies for permissions to read phone state and sending messages, then there may exist the threat of sending the phone number or IMEI out. Permissions filtering module is based on a set of security policies to determine whether an application has some special risk permission combinations. For Android permissions, there are four different security levels. Those are $Normal, Dangerous, SignatureandSignatureOrSystem$.

- **Normal** lower-risk permissions that present minimal risk to Android apps and will be granted automatically by the Android platform without asking for user's explicit approval.
- **Dangerous** higher-risk permissions that would give access to the user's personal sensitive data and even control over the phone device that can negatively impact the user. Applications requesting dangerous permissions can only be granted if the user approves the permission explicitly.
- **Signature** permissions that the system grants only if the requesting application is signed with the same certificate as the application declared the permission. Signature permissions are automatically grant without user explicit approval if the certificates match.
- **SignatureOrSystem** permissions are only granted to applications that are in the Android system image or are signed with the same certificate as the application that declared the permission. Permissions in this category are used for certain special situation where multiple vendors have applications built into a system image and need to share specific features explicitly because they are being built together.

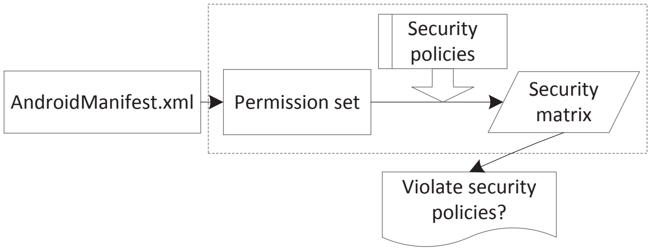


Figure 3: The procedure of permissions filtering module

From the above four permission levels, we mainly concern the *Dangerous* level which has great potential risks for leaking user privacy data. Moreover, through analyzing the malware samples, we find the process of privacy leakage has two steps: read the privacy information and send out. Accordingly, the potential causing privacy leakage permissions are also divided into two categories. One is mainly used to read the privacy data, such as *android.permission.*$READ\_PHONE\_STATE$, which allows to read phone state, such as SIM card, phone numbers, phone's IMEI (International Mobile Equipment Identity) and some others. The other is mainly used to send out privacy information. At present, we are only focused on two leakage ways, one is SMS (Short Message Service), and the other is network transmission, namely *android.permission.*$SEND\_SMS$ and *android.permission.*$INTERNET$. Figure 3 shows the procedure of permission filtering module.

In Figure 3, the security policy is the core part, where each security policy is a cross combination of the above two kinds of permission set. The first one is $READ\_P = \{a\_1, a\_2, ..., a\_n\}, n \in N$ and the second one $SEND\_P =$

$\{b\_1, b\_2, ..., b\_n\}, n \in N$. The security policy is $S\_i = \{a\_i, b\_i\}, i \in N$. The set of all security policies are $SECURITY\_P = \{S\_1, S\_2, ..., S\_n\}, n \in N$. After the first step of static decompile, we can extract the application permissions set $APP\_P = \{p\_1, p\_2, ..., p\_n\}, n \in N$ from the App's configuration file *AndroidManifest.xml*. We define a permission matrix, the column for accessing privacy data permissions and the row for sending permissions. Through the values of matrix we can determine whether some risky combination of two permissions exists. Matrix model can represent a combination of permissions, not only can reflect the presence or absence of permissions, but also can demonstrate the relationship between permissions in detail. Matrix model is shown in Table 2.

Table 2: An example of permissions matrix model

Read Permission	**Send Permission**	
	SEND_SMS	INTERNET
$ACCESS\_FINE\_LOCATION$	1	0
$READ\_CALENDAR$	0	0
$READ\_PHONE\_STATE$	0	1
$READ\_OWNER\_DATA$	-	-
$READ\_SMS$	-	-
...	...	...

During the static decompiling phase, we extract the permissions set APP_P form $AndroidManifest.xml$, and then classify APP_P into two categories, *read* and *send* defined above. We assume that if APP_P set on matrix has a valid value (here is 1), that is to say, the APK requested permissions have the higher-risk, and then the app can be regarded as suspicious. For example, the second row of table 2, "0" means that the application did not violate the security policy of leaking user calendar data risk because there are no permission combinations of $(READ\_CALENDAR, SEND\_SMS)$ and $(READ\_CALENDAR, SEND\_INTERNET)$. Conversely, "1" indicates that the APP_P set of an application holds the risk of phone state and user location leakage.

3.3 Dynamic Monitoring Module

This module is to monitor the call information of sensitive APIs in APK. We implement dynamic real-time monitoring by inserting monitoring code to the decompiled APK. The Android developers write the application in Java, compiles it into Java bytecode, and finally transfers to the Dalvik bytecode which can be executed in DVM. So it is straightforward to do the monitoring reversely by converting the Dalvik bytecode to Java bytecode, then rewrite the Java bytecode, and finally convert the rewritten Java bytecode back to Dalvik bytecode. However, this kind of approach sometimes does not work. First of all, there are several important differences between JVM(Java Virtual Machine) and DVM. The most obvious one is that JVM is based on stack whereas DVM is register based. Several tools, such as *dex2jar* [14] and *ded* [19], attempt to convert Dalvik bytecode back to Java bytecode. However this is not a lossless converting, that is to say, some information from the Java bytecode is lost when being converted to Dalvik. These tools try to infer the missing details based on the context, but sometimes the inference is unreliable (as described by Reynaud et al. [23]). Even though these errors may not prevent static analysis on the converted Java bytecode, in our experience they often lead to invalid Java bytecode or later invalid Dalvik bytecode. In other words, after we converted an application's Dalvik bytecode to Java bytecode (e.g. *dex2jar*) and then back to Dalvik bytecode, the resulting application sometimes failed to run. So the feasible way is to directly use the Dalvik bytecode.

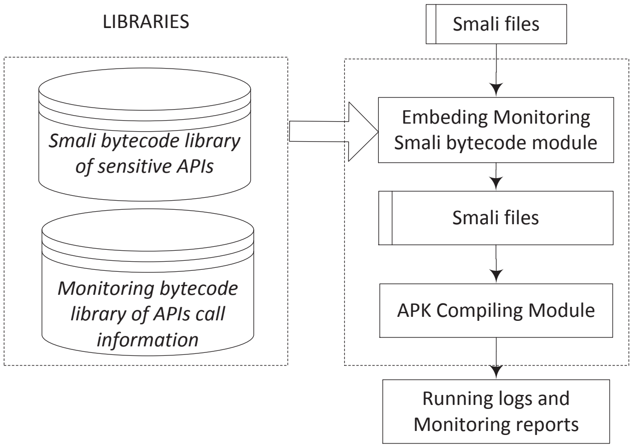

LIBRARIES
Smali bytecode library of sensitive APIs
Monitoring bytecode library of APIs call information
Smali files
Embeding Monitoring Smali bytecode module
Smali files
APK Compiling Module
Running logs and Monitoring reports

Figure 4: The procedure of dynamic monitoring module

Smali and baksmali are an assembler and disassembler respectively for the dex format used by the DVM. Its syntax is loosely based on Jasmin's syntax [11]. Smali is an intermediate representation of Dalvik bytecode. Smali can fully realized all the features of dex format (annotations, debug information, thread information, etc.). Moreover, dex and Smali can convert lossless between each other. So, in this paper in order to avoid the differences between JVM and DVM, we try to directly rewrite Dalvik bytecode, insert the monitoring Smali bytecode into the decompiled Smali files. The procedure of dynamic monitoring module is shown in Figure 4.

In Figure 4, we can obtain Smali files from the static decompiling. Then locate the concrete position of the sensitive API, and embed monitoring Smali bytecode to each different sensitive API. After that we use apktool to repackage the modified Smali bytecode to create a new APK and use the signature tool to re-sign it. Running the new APK on Android emulator, we can use logcat to view the runtime logs. It can generate a log on SD card which records the detailed call information of those

Table 3: Descriptoin of Smali syntax

Type		**Syntax**	**Meaning**
Primitive Types		V	void
		Z	boolean
		B	byte
		S	short
		C	char
		I	int
		J	long (64 bits)
		F	float
		D	double (64 bits)
Reference types	Object	Lpackage/name/ObjectName	Package.name.ObjectName
	Array	[primitive type signature	[I, represents a array of int, like int[] in java
	Array of Objects	Lpackage/name/ObjectName	[Ljava/lang/String, represents a array of String Objects

sensitive APIs.

1) **Smali Bytecode.** Smali [11] is an Intermediate Representation(IR) for Dalvik Bytecode. Smali code is a kind of register based language which can shield the source code level differences. For instance, malware sometimes use source code obfuscation to avoid detection. But in Smali code, the core sensitive APIs are inevitably exposed. So, we can monitor these sensitive APIs to track the behavior of those suspicious programs.

   In DVM, although all the register is 32 bits, it can support any data types. In order to represent a 64 bits type (Long/Double), it uses two registers. Dalvik bytecode has two types, primitive types and reference types. Reference types are only Objects and Arrays. The Smali syntax is indicated briefly in Table 3.

   Objects take the form Lpackage/name/ObjectName, where the leading L indicates that it is an Object type, package/name is the package that the object is in, ObjectName is the name of the object. For example, Ljava/lang/String; is equivalent to java.lang.String. Arrays take the form $[I$, i.e. int[] in java. Methods take the form as below:

   Lpackage/name/ObjectName;→
   MethodsName(III)Z.

   In this example, *MethodsName* is obviously the name of the method. *(III)Z* is the signature of the method. *III* are the parameters (in this case, three integers), and Z is the return type . For example, $method(I;\ [[II;\ Ljava/lang/String;\ [Ljava/lang/Object;)\ Ljava/lang/String;$ is equivalent to a string $method(int,\ int[\cdot][\cdot],\ String,\ Object[\cdot])$ in java.

2) **Smali bytecode library for sensitive APIs.** The Smali bytecode library stores sensitive APIs and their corresponding Smali bytecode. The main function of the library is to locate the detailed position of sensitive APIs in Smali files after the target APK was decompiled. According to the typical sensitive APIs that malwares often used for leaking Android user's privacy data, we choose five and their class name, function name, and Smali bytecode are indicated in Table 4.

3) **Monitoring bytecode library for Sensitive APIs.** The monitoring bytecode library is to store the sensitive APIs calling information when the APK is running. For different APIs, monitoring information to be recorded are different. Such as SMS sending text messages, we need to record the message recipients as well as the content of the message. The unique part of each API is its input and output. According to API's function prototypes and register naming principles in Smali syntax, we can obtain the Smali register number of each API parameters. According to Smali syntax, there are two ways to determine a method that how many registers are available, which can be shown in Table 5 .

   When a method is invoked, its parameters will be placed in the last N available registers. For example, supposing a method has two parameters and five available registers ($v0 \backsim v4$), then the parameters will be placed in the last two registers ($v3$ and $v4$). Moreover, the first argument of the non-static method is always the object which call the method, and for static methods except there is no implicit this parameter, others are the same. For example, the method for sending text messages is as follows:

   public $sendTextMessage$ (String destinationAddress, String scAddress, String text, PendingIntent sentIntent, PendingIntent deliveryIntent).

   The above method has 5 parameters, defined as "*public*" which means it is a non-static method and the first register $v0$ is the object used to call the

Table 4: Smali bytecode library for sensitive API

Class Name	**Function Name**	**Description**	**Smali Bytecode**
android. telephony. SmsManager	sendTextMessage (String, String, String, PendingIntent, PendingIntent)	Send messages	Landroid/telephony/SmsManager; →sendTextMessage(Ljava/lang/String; Ljava/lang/String;Ljava/lang/String; Landroid/app/PendingIntent; $Landroid/app/PendingIntent;)V$
android.location. LocationManager	getLastKnownLocation (String)	Get location	Landroid/location/LocationManager; → getLastKnownLocation(Ljava/lang/String)
android. telephony. TelephonyManager	getDeviceId()	Get ID, IMEI of phone	Landroid/telephony/TelephonyManager; → getDeviceId()Ljava/lang/String
android.location. LocationManager	getSimSerialNumber()	Get SIM serial Number	Landroid/telephony/TelephonyManager; → getSimSerialNumber () Ljava/lang/String
android.telephony. TelephonyManager	getLine1Number()	Get phone Number	Landroid/telephony/TelephonyManager; → getLine1Number () Ljava/lang/String

method, namely, it stores "*this*" parameter. Meanwhile, for $sendTextMessage$ we need to record the SMS destination address (corresponding to the destinationAddress parameter) and the SMS content (text parameter). According to the principle of register allocation in Smali syntax, the first parameter destinationAddress stored in registers $v1$ and the third parameter text stored in the register $v3$.

For other sensitive APIs, sometimes we need to record the return value of the method. And the instruction for the return value is the last instruction in the method. The basic bytecode for return instruction is $return$ and there are four return instructions in total, which are shown in Table 6.

Table 5: Dalvik bytecode for registers

Instruction	**Description**
.registers	Specifies the total number of registers in a method
.locals	Indicates the number of nonparameter register in a method, which appears in the first line of the method

After we have got the syntax of Smali, the monitoring code for sensitive APIs in Table 4 can be provided in Table 7.

So far, we have introduced the mechanisms of our static-dynamic analysis method. Next, detailed experiments will show how it works.

4 Experiments

Before experiment, some necessary tools such as Eclipse, JDK6, JRE6, Android SDK, Python2.7, and other tools will be installed. APK static decompiler and permissions

Table 6: Dalvik bytecode for return value

Instruction	**Description**
return-void	Return from a void method
return vAA	Return a 32-bit non-object value and the return value register is an 8-bit register, vAA
return-wide vAA	Return a 64-bit non-object value and the return value register is an 8-bit register, vAA
return-object vAA	Return a Object value and the return value register is an 8-bit register, vAA

Table 7: Monitoring code for sensitive APIs of Table 4

Return type	**Function Name**	**Register**
void	sendTextMessage(String destinationAddress, String scAddress, String text, PendingIntent sentIntent, PendingIntent deliveryIntent)	v1,v3
Location	getLastKnownLocation(String provider)	vAA
string	getDeviceId()	vAA
string	getSimSerialNumber()	vAA
string	getLine1Number()	vAA

filtering module were implemented with Java. Among them, APK static compiling module is to call apktool [23]. Permissions filtering module mainly implements security policy settings, extract permissions features from Androidmanifest.xml that generated by static decompiling module. Dynamic monitoring module is to scan the Smali files generated by APK static decompiling module, embed monitoring bytecode, repackage and re-sign the Smali files to generate a new APK. Then we run the new APK in Android emulator, it will generate some running logs or monitoring report to show what has happened.

4.1 Android Markets Sensitive API Analysis

To illustrate Android users facing the growing threat of information leakage, we choose 642 popular applications to conduct experiments in the permissions filtering module. These 642 applications mainly come from Android online markets such as shouji.com.cn, appchina.com, market.goapk.com, and eoemarket.com. APK samples chose in this paper are popular applications coming from different app markets, and they mostly have more than half of Android users.

Meanwhile, in these experiments we only concern user's privacy data including: Location, SMS, Contacts, Address book, phone Number, IMEI (International Mobile Equipment Identity) and ICCID (Integer Circuit Card Identity). Leakage ways includes sending messages and network. We handled 642 APK samples by permissions filtering module and found that almost 26% apps have security risks for leakage user's sensitive data. Here, revealing user's privacy information refers to the app handled by permissions filtering, and it is considered to be suspicious. The specific statistical data is shown in Table 8.

Table 8: Analysis of suspicious Apps

Market	**App Number**	**Suspicious Number**	**Ratio (%)**
shouji.com.cn	59	6	10.17
appchina.com	283	53	18.73
market.goapk.com	66	25	37.89
eoemarket.com	234	85	36.32
Total	642	169	26.32

Then we analyze the permissions requested by APK through permissions filtering module. According to the security policies matrix proposed in section 3, we count the number of applications corresponding to each type of privacy information. The specific statistical results are shown in Table 9.

From Table 9, the security policy violated by most of those 642 apps is about IMEI permission combinations. Namely, the most common information leakage is IMEI. The reason may be the IMEI can determine phone type and device parameters, and can provide accurate user identity information for developers and advertisers. The next is phone Number, Contacts, and Location. If these sensitive information are used illegally, it will possibly bring huge losses to users.

Table 9: Analysis of privacy information leakage

Leakage corresponding to security policies	**App amounts**	**Ratio(%)**
Location	15	2.34
SMS text	1	-
Contacts	17	2.64
PhoneNumber	61	9.50
IMEI	199	31.01
ICCID	7	1.09

4.2 Sensitive API Monitoring

To verify effectiveness and feasibility of dynamic monitoring module, we did experiments on the Android emulator in Windows7. Android source code version is Android 2.3.4_$r1$, and the kernel is Linux kernel 2.6.29 goldfish. The monitoring report generated by dynamic monitoring module is TXT, and the output position is in Android emulator SD card. We designed an APK, *showLog.apk* (also can use message box or phone ringing), to show the monitoring log. We chose an APK, $SendSMS\_example.apk$, that will automatically send text message in the background. *showLog.apk* and $SendSMS\_example.apk$ successfully installed in Android emulator are shown in Figure 5.

For different sensitive API, through context we need to find its relative registers, the return value and then call different log functions to record the information while the app runs. For send text message API, $sendTextMessage$, the Smali code is shown in Figure 6.

From Figure 6, it shows the $sendTextMessage$ function has 5 parameters. Among them we only focus on the recipient number and the message contents. From the context, it is obvious that register $v1$ stores the recipient number and $v3$ stores the message contents. The other 3 registers $v2, v4, v5$ are null. So, we just need to pass the $v1$ and $v3$ to log function (Smali format) to record the SMS information while it runs. After embedding the log monitoring bytecode, we again used the apktool to repackage the modified Smali files, and then call signapk.jar to re-sign the new APK. The monitoring report was stored in external SD card. In order to view the log, we designed showLog.apk to show the recipient number, messages contents and timestamps. An example of detailed monitoring report on Android emulator is shown in Figure 7.

The above experiments show that the dynamic monitoring module was successfully embedded into the Smali files of original APK. The log records the detailed infor-

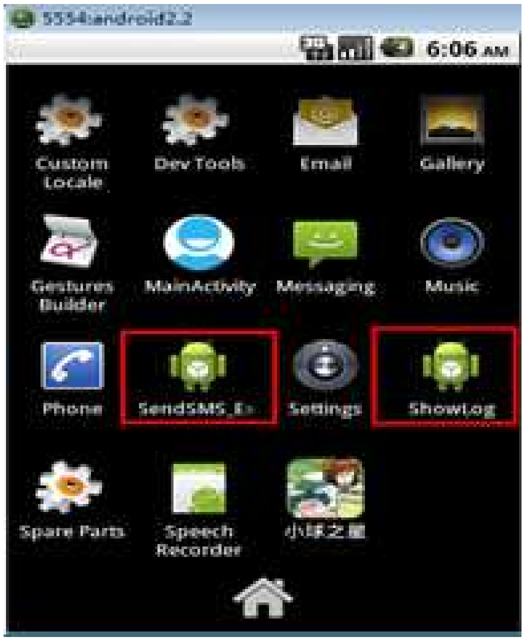
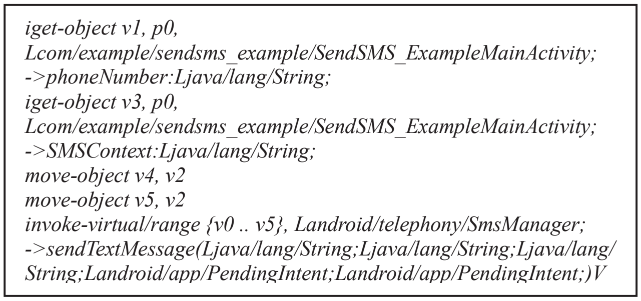
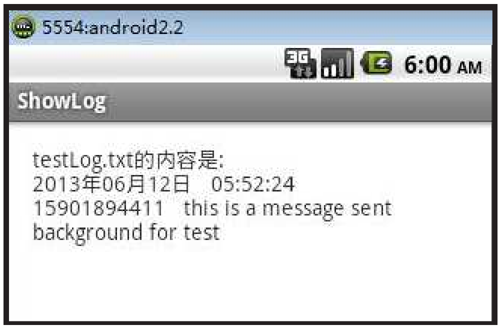


Figure 5: Running interface of installed *showLog.apk* and *SendSMS_example.apk*



```
iget-object v1, p0,
Lcom/example/sendsms_example/SendSMS_ExampleMainActivity;
->phoneNumber:Ljava/lang/String;
iget-object v3, p0,
Lcom/example/sendsms_example/SendSMS_ExampleMainActivity;
->SMSContext:Ljava/lang/String;
move-object v4, v2
move-object v5, v2
invoke-virtual/range {v0 .. v5}, Landroid/telephony/SmsManager;
->sendTextMessage(Ljava/lang/String;Ljava/lang/String;Ljava/lang/
String;Landroid/app/PendingIntent;Landroid/app/PendingIntent;)V
```

Figure 6: An example of Smali code for *sendTextMessage*



```
testLog.txt的内容是:
2013年06月12日   05:52:24
15901894411   this is a message sent
background for test
```

Figure 7: An Example of monitoring log for sendTextMessage

mation related to the sensitive API. Once we have found the suspicious behavior of an app, any further deep analysis, such as DDMS (Dalvik Debug Monitor Server), can analyze it more accurately and comprehensively.

5 Conclusion

A two-step malicious Android application detection method was proposed in this paper. First of all, we use permission combination matrix to discover those potential risk applications. And then those suspicious applications are further sent into the dynamic monitoring module to track the call information of the sensitive APIs while it is running. As a conclusion, it shows some advantages of our approach:

1) Using Smali bytecode, it is based on intermediate language, which shows some advantages over Java source code method and it possesses anti-obfuscation to a certain degree.

2) The method is simple, just insert some monitoring Smali bytecode, and the performance influence can be ignored.

3) This method can be used in a wide scale, which can deploy remotely and provide monitoring service automatically.

Further research directions include considering more sensitive APIs and provide a real App on Android market for fans to use. Also, we need integrate others malware detection method, such as dynamic taint analysis to conduct some cross-field deep research.

Acknowledgments

This study was supported by Shanghai Municipal Natural Science Foundation under grant No.13ZR1416100. The authors gratefully acknowledge the anonymous reviewers for their valuable comments.

References

[1] L. Batyuk, M. Herpich, S. A. Camtepe, and K. Raddatz, "Using static analysis for automatic assessment and mitigation of unwanted and malicious activities within android applications," in *Proceeding of the 6th International Conference on Malicious and Unwanted Software*, pp. 66–72, Fajardo, Oct. 2011.

[2] A. R. Beresford, A. Rice, N. Skehin, and R. Sohan, "Mockdroid: Trading privacy for application functionality on smartphones," in *Proceedings of the 12th Workshop on Mobile Computing Systems and Applications*, pp. 49–54, Phoenix, USA, Mar. 2011.

[3] T. Blasing, L. Batyuk, A. D. Schmidt, S. A. Camtepe, and S. Albayrak, “An android application sandbox system for suspicious software detection,” in *Proceeding of the 5th International Conference on Malicious and Unwanted Software*, pp. 55–62, Nancy,France, Oct. 2010.

[4] D. Bornstein, *Dalvik VM Internals*, 2008. (https://sites.google.com/site/io/dalvik-vm-internals)

[5] Canalys, *Mobile Device Shipments Survey Report in the First Session of 2013*, May 2013. (http://cn.engadget.com/tag/canalys)

[6] Canalys, *Smartphone Shipments Survey Report in the Fourth Session of 2011*, May 2013. (http://cn.engadget.com/tag/canalys)

[7] E. Chin, A. P. Felt, K. Greenwood, and D. Wagner, “Analyzing inter-application communication in android,” in *Proceedings of the 9th International Conference on Mobile Systems, Applications and Service*, pp. 239–252, Washington, USA, June 2011.

[8] W. Enck, P. Gilbert, B. G. Chun, L. P. Cox, J. Jung, P. McDaniel, and A. N. Sheth, “Taintdroid: An information-flow tracking system for realtime privacy monitoring on smartphones,” in *Proceedings of the 9th USENIX Conference on Operating Systems Design and Implementation*, pp. 1–6, Vancouver, Canada, Oct. 2010.

[9] W. Enck, M. Ongtang, and P. McDaniel, “On lightweight mobile phone application certification,” in *Proceedings of the 16th ACM Conference on Computer and Communications Security*, pp. 235–245, Chicago, USA, Nov. 2009.

[10] A. P. Fuchs, A.Chaudhuri, and J. S. Foster, “Scandroid: automated security certification of android applications,” *Technical Report of University of Maryland*, 2009. (http://www. cs. umd. edu/ ãvik/ projects/ scandroidascaa)

[11] Google, *Smali*, July 11, 2015. (http://code. google. com/ p/ smali/)

[12] Google, *Android Home Page*, 2009. (http://www. android. com)

[13] Google, *Android Security and Permissions*, 2013. (http://d.android.com/guide/topics /security /security.html)

[14] Google, *Dex2jar: Tools to Work with Android .dex and java .classfiles*, 2013. (http://code. google. com/ p/ dex2jar/)

[15] P. Hornyack, S. Han, J. Jung, S. schechter, and D. Wetherall, “These aren’t the droids you’re looking for: Retrofitting android to protect data from imperious applications,” in *Proceedings of the 18th ACM Conference on Computer and Communications Security*, pp. 639–652, Chicago, USA, Oct. 2011.

[16] T. Isohara, K. Takemori, and A. Kubota, “Kernel-based behavior analysis for android malware detection,” in *Seventh International Conference on Computational Intelligence and Security*, pp. 1011–1015, Hainan, China, Dec. 2011.

[17] K. Luo, “Using static analysis on android applications to identify private information leaks,” *Master Dissertation of Kansas State University*, 2011.

[18] Netqin, *Mobile Security Report in 2012*, May 2013. (http://cn.nq.com/anquanbobao)

[19] D. Octeau, W. Enck, and P. McDaniel, *The DED Decompiler*, 2011. (http://siis. cse. psu. edu/ ded/ papers/ NAS-TR-0140-2010.pdf)

[20] M. Ongtang, S. McLaughlin, W. Enck, and P. McDaniel, “Semantically rich application-centric security in android,” in *Proceedings of the 25th Annual Computer Security Applications Conference*, pp. 340–349, Honolulu, USA, Dec. 2009.

[21] G. Peng, Y. Shao, T. Wang, X. Zhan, and H. Zhang, “Research on android malware detection and interception based on behavior monitoring,” *Wuhan University Journal of Natural Sciences*, vol. 17, no. 5, pp. 421–427, 2012.

[22] Q. Qian, J. Cai, and R. Zhang, “Android malicious behavior detection based on sensitive api monitoring,” in *2nd International Workshop on Security*, pp. 54–57, Nov. 2013.

[23] D. Reynaud, D. Song, T. Magrino, E. Wu, and R. Shin, “Freemarket:shopping for free in android applications,” in *19th Annual Network & Distributed System Security Symposium*, Hilton San Diego, USA, Feb. 2012.

[24] A. Shabtai, Y. Fledel, and Y. Elovici, “Securing android-powered mobile devices using selinux,” *IEEE Security & Privacy*, vol. 8, no. 3, pp. 36–44, 2010.

Quan Qian is a Professor in Shanghai University, China. His main research interests concerns computer network and network security, especially in cloud computing, data privacy protection and wide scale distributed network environments. He received his computer science Ph.D. degree from University of Science and Technology of China (USTC) in 2003 and conducted postdoc research in USTC from 2003 to 2005. After that, he joined Shanghai University and now he is the lab director of network and information security.

Jing Cai is a master degree student in the school of computer science, Shanghai University. Her research interests include Android security and software security analysis.

Mengbo Xie is a master degree student in the school of computer science, Shanghai University. Her research interests include data privacy, privacy based data mining, computer and network security.

Rui Zhang received her B.E. and Ph.D. degree from Department of Electronic Engineering & Information Science, University of Science and Technology of China, in 2003 and 2008, respectively. After that, she joined the School of Computer Engineering and Science, Shanghai

University. Now, she is an associate professor and her main research interests include computer networks, network coding for wireless networks and wireless communication, etc.

Anonymous Network Information Acquirement Protocol for Mobile Users in Heterogeneous Wireless Networks

Guangsong Li[1], Qi Jiang[2], Yanan Shi[1], and Fushan Wei[1]

(Corresponding author: Guangsong Li)

State Key Laboratory of Mathematical Engineering and Advanced Computing[1]
No. 62 of Science Road, Zhengzhou 450002, P. R. China
School of Computer Science and Technology, Xidian University[2]
No. 2 of Taibai Road, Xi'an 710071, P. R. China
(Email: lgsok@163.com)

(Received Dec. 10, 2013; revised and accepted Nov. 15 & Dec. 28, 2014)

Abstract

Media independent information service is one of the important parts of the IEEE 802.21 standard to optimize vertical handover in wireless heterogeneous networks. In this paper, an anonymous network information acquirement protocol is proposed for a mobile user, which can be used to establish a secure channel between the mobile user and the information server. Security and performance analysis shows that the proposed protocol is very suitable for mobile environments.

Keywords: Anonymity, heterogeneous network, media independent information service

1 Introduction

Communication in next generation wireless networks will use multiple access technologies, creating a heterogeneous network environment. Practically, a single network cannot cater for all different user needs or provide all services. Nowadays the availability of multimode mobile devices capable of connecting to different wireless technologies provides users with the possibility to switch their network interfaces to different types of networks. Vertical handovers among heterogeneous networks should be supported to guarantee the service continuity. To achieve a seamless handover, a mobile user needs to obtain information of existing networks nearby, in order that he can choose a suitable target network and do some preparations for possible handover. However, the neighbor information discovery is the most time-consuming phase in the handover process [15].

The IEEE 802.21 working group defines the Media Independent Handover (MIH) services [4] to facilitate handover between heterogeneous networks. Media Independent Information Service (MIIS) is a very important part of MIH services, which specifies information about nearby networks and the query/response mechanism that allows mobile nodes to get that information from information servers. MIH messages will be exchanged over various wireless media between mobile nodes and access networks in future heterogeneous networks. Thus the MIH services may be a new target to attackers, which will be the main concerns for equipment vendors and service providers. Some typical threats about MIIS are listed in [9], which includes identity spoofing, tampering, replay attack, denial of service and information disclosure. Note that an attacker may be able to trace a user's movements or predict future movements by inspecting MIIS messages. Thus, it is desirable to hide the roaming user's identity and movements from eavesdroppers. However, security mechanisms are not within the scope of the IEEE 802.21 standard.

IEEE 802.21a task group was set up to address security issues of MIH services. As to MIH security, two frameworks about MIH service access control were proposed [5, 8]: (i) 3-party case, the access control is applied through EAP process (for instance, EAP-TLS [13]) with an EAP server, where the information server plays a part of authenticator; (ii) 2-party case, the access control is based on a pre-shared key or public key certification, where the user and the information server execute a mutual authentication and key establishment procedure like TLS [2]. Saadat et al. [11] describe the main technical requirements to establish a secure channel between the user and the information server. They also propose that the user should be authenticated by an authentication server and a shared key between the user and the information server should be generated by the authentication server. However, the specific authentication method is not referred. Saha et al. [12] propose a PLA-MIH scheme to transport 802.21 messages over a secure network layer protocol in a hop-by-hop manner. It has the advantage that ensures very strong security of the signaling framework. However, it adds much overhead to all entities involved, for it needs every packet in MIH signaling to be signed. Won et al. propose another secure MIH message transport solution called MIHSec [14], which computes the MIH keys by utilizing the keys generated from the data link layer authentication procedure. Though MIHsec has a good performance for MIH message transportation, it introduces other issues. First, it is closely integrated with date link layer authentication, thus it is not media independent. Second, the access router may know the key for MIH messages encryption, which degrades the level of security.

We note that user anonymity is not addressed in all above schemes. It is very important for a roaming user to keep his identity secret and movements untraceable. In [7], we propose an access authentication scheme with user anonymity. The scheme provides an anonymous access authentication of MIIS considering that the access control for information is applied through an access authentication controller. The protocol can be used to establish a secure channel between the mobile node and the information server. The solution has the advantages of lightweight computation and easy implementation, However, it has the following weak points: first, it needs the mobile user to require a service ticket from his home server every time before accessing MIIS, which may take a long time if the user is far away from his home network; second, the home server has to be always online and available, so it is easy for the home server to become the bottleneck.

In order to achieve an efficient network information discovery process with more security properties, this paper proposes a new Anonymous Network Information Acquirement (ANIA) protocol using an Schnorr like ID-based signature scheme [3]. The anonymous authentication process does not involve the home server, which resulting a very short authentication latency. We also give a rigorous formal analysis of its security using a modular approach.

Our contribution mainly includes:

- Quick mutual authentication with user anonymity between the user and the information server;
- A shared session key established for MIIS information secure transmitting;
- Lightweight computation and low communication cost in the proposed protocol.

The rest of this paper is organized as follows. Section 2 gives a quick review of eCK model. In Section 3 we present our new approach in detail. Section 4 gives a formal security proof of our protocol under the ECK model. Section 5 includes performance analysis. Finally, conclusions are drawn in Section 6.

2 Related Work

The extended Canetti-Krawczyk (eCK) model [6] is described as an experiment between an adversary Δ and a challenger Σ. Initially, Δ selects the identities of n honest parties, for whom Σ generates static private key/public key pairs.

Execution of an Authenticated Key Exchange (AKE) protocol by one of these parties is called an AKE session. A session identifier *sid* is defined as

$sid = (role, \Phi, \Psi, comm)$,

where $role = \{I, R\}$ is the role (initiator/responder) of the owner of the session, Φ is the identity of the owner, Ψ is the identity of the other party in the session, and *comm* is the concatenation of communication messages between the two parties. Two sessions $sid = (role, \Phi, \Psi, comm1)$ and $sid^{*} = (role, \Phi, \Psi, comm2)$ are matching sessions if role is the complement of $role^{*}$ and $comm1 = comm2$. A protocol execution between Φ and Ψ without the intervention of an adversary produces two matching sessions.

In the experiment, Δ controls all communications between the parties, and can reveal the static private key of a party, the ephemeral private key in a session, and the session key of a session. Δ can make any sequence of the following queries, which Σ needs to answer accordingly:

- Send(Φ, Ψ, *comm*). Δ sends a message *comm* to Φ on behalf of Ψ. Σ returns Φ's response.
- StaticKeyReveal(Φ). Σ returns the static private key of Φ.
- EphemeralKeyReveal(*sid*). Σ returns the ephemeral private key of the session *sid.*
- SessionKeyReveal(*sid*). Σ returns the session key of the session sid.
- Establish(Φ). Using this query, the adversary registers an arbitrary public key on behalf of an adversary controlled party Φ. Σ only checks the validity of the public key, but does not need to check the possession of the corresponding private key.

  A session *sid* (*role*, Φ, Ψ, *comm*) is fresh if the following conditions hold:

- Both Φ and Ψ are honest parties.
- Δ did not query the session key of sid or its matching session sid^{*} (if the matching session exists).
- Δ did not query both the static private key of Φ and the ephemeral private key of Φ in this session.
- If sid^{*} exists, then Δ did not query both the static private key of Ψ and the ephemeral private key of Ψ in this session.
- If sid^{*} does not exist, then Δ did not query the static private key of Ψ.

Security of an AKE is defined as follows. In an eCK experiment, Δ issues Send, StaticKeyReveal, EphemeralKeyReveal, SessionKeyReveal, and Establish queries polynomial times (in a security parameter) in any sequence. Then Δ selects a completed session *sid*, and makes a query Test(*sid*). To answer Test(*sid*), Σ chooses a bit $b\in$ $\{0, 1\}$uniformly at random. If $b = 1$, then Σ sets the session key of *sid* as $\mathbf{K}$. Otherwise, Σ selects $\mathbf{K}$ from the key space uniformly at random. Σ then returns $\mathbf{K}$ as the answer of Test(*sid*). Δ continues to query Send, StaticKeyReveal, EphemeralKeyReveal, SessionKeyReveal, and Establish polynomial times. At last, Δ outputs a bit b^*, and terminates the game. If the selected test session is fresh and $b^* = b$,then Δ wins the game.

The advantage of the adversary Δ in the eCK experiment with AKE protocol Π is defined as $\mathrm{ADV}_\Pi(\Omega) = \Pr\{\Delta \text{ wins}\}-1/2$.

eCK Security

An AKE protocol is secure (in the eCK model) if no efficient adversary Δ has more than a negligible advantage in winning the above experiment, i.e., $\mathrm{ADV}_\Pi(\Delta) < 1/Q(\mu)$ for any polynomial $Q(\cdot)$ when μ sufficiently large.

3 Network Information Acquirement Protocol with User Anonymity

This section focuses on a new proposal for anonymous network information acquirement using an efficient Schnorr like ID-based signature scheme [3].

3.1 Network Initialization

We consider a network model as depicted in Figure 1. A mobile user (MU) roams into a visited network (V) and he wants to get network information nearby for possible handover. We assume the MU registers with a home authentication server (HAS) in his home network (H) and has a long term shared key k_{MH} with the HAS. The MIIS is provided by an information server (IS) in the Internet. Suppose there is an agreement between the IS and the HAS for MUs using MIIS to optimize handover. We also assume there is a time synchronization mechanism in the system.

In this phase, the HAS runs a setup algorithm and generates the system parameters, including a master secret key (s), and the corresponding master public key (PK_{HAS}), by using a security parameter k. The HAS performs the following steps:

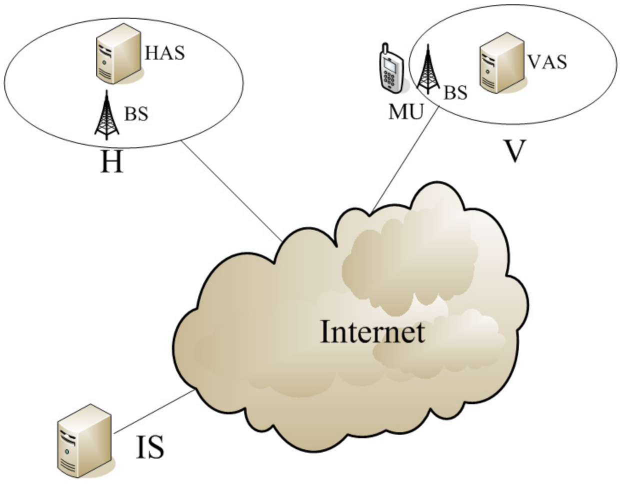


Figure 1: Network model

1) Specifies q, p, E/F_p, P and G where q is a large prime number and p is the field size, E/F_p is an elliptic curve E over a finite field F_p, P is a base point of order q on the curve E and G is a cyclic group of order q under the point addition "+" generated by P.

2) For the randomly chosen master secret key $s\in Z_q^*$, computes PK_{HAS} as sP.

3) Chooses two hash function H_1: $\{0,1\}^* \longrightarrow Z_q^*$, H_2: $\{0,1\}^*\times G\times\{0,1\}^* \longrightarrow Z_q^*$.

4) Chooses one key derivation function f: $G\longrightarrow\{0,1\}^k$.

5) Outputs system parameters $\{q$, p, E/F_p, P, G, PK_{HAS}, H, $f\}$, and keeps s secret.

Later, the HAS computes the private keys of all users and the IS. This algorithm takes the master secret key s and an identifier (ID) as input and generates a private key corresponding to that ID. In order to achieve MIIS access anonymity, the HAS selects a pseudo-ID (PID) for each MU and based on the PID a private key is generated. The HAS works as follows for each MU with identifier PID_{MU}. It chooses at random $r_{MU} \in_R Z_q^*$,compute $R_{MU} = r_{MU}P$ and $h_{MU} = H_1(PID_{MU}, R_{MU})$. Then it computes $s_{MU} = r_{MU}+h_{MU}s$. The MU's private key is the tuple (s_{MU}, R_{MU}) and is transmitted to the MU via a secure channel, namely encrypted by the key shared between the HAS and the MU. The MU's public key is defined as $PK_{MU} =s_{MU}P$, which can also be computed with R_{MU}, PID_{MU}, and PK_{HAS} from the equation: $PK_{MU}=R_{MU}+H_1(PID_{MU}, R_{MU})PK_{HAS}$. The HAS also generates a private key for the IS as above procedure using ID_{IS}. The private key and public key of the IS are denoted as (s_{IS}, R_{IS}) and $PK_{IS}=s_{IS}P$, respectively.

3.2 Anonymous Secure Channel Establishment

When a MU moves to a new place, it should contact the IS to get information about neighbor networks. Suppose that the MU is now in coverage area of network V and he is already connected with the network. Then an anonymous authentication and key establishment process will be conducted between the MU and the IS. The flow chart of our scheme is depicted in Figure 2.

1) MIIS Authentication Request (MU⟶IS):

   PID_{MU}, A, t_{MU}, σ.

   The MU selects a random number $a \in Z_q^*$, and computes $A = aP$. He sends a MIIS authentication request message to the IS. The message content is as the following, $\{PID_{MU}$, A, t_{MU}, $\sigma\}$, where t_{MU} is the timestamp of the MU, and σ is a signature generated by the schnorr like ID-based signature using s_{MU}. Denote $\{PID_{MU}$, A, t_{MU}, $\sigma\}$as m, then σ is generated as follows [3]: The MU selects a random number $x \in Z_q^*$,and computes xP, y=x+$s_{MU}H_2(PID_{MU}$, xP, $m)$, then he outputs the signature σ=$\{xP$, y, $R_{MU}\}$.

2) MIIS Authentication Response (IS⟶MU)

   ID_{IS}, R_{IS}, B, A, c, MAC.

   Upon receiving the request message from the MU, first the IS checks the time stamp t_{MU}. If it is fresh, the IS computes the MU's public key by the equation PK_{MU}= $H_1(PID_{MU}$, $R_{MU})$ PK_{HAS}+R_{MU} (Note R_{MU} can be extracted from σ). Then the IS verifies the signature σ using PK_{MU} by checking the following equation: yP=xP+ $H_2(PID_{MU}$, xP, $m)PK_{MU}$. Successful signature verification implies the message is actually sent by a valid user of the HAS. Hence, the IS accepts the message. Otherwise the protocol is terminated at this stage. Next the IS selects a random number $b \in Z_q^*$, and computes $B = bP$. Then it computes the shared secret k_{IM} as follows: K_{IM} =$(b$+$s_{IS})(PK_{MU}$+$A)$, k_{IM} =$f(K_{IM}$, PID_{MU}, $ID_{IS})$. The IS randomly chooses a temporary ID (TID_{MU}) for the MU and stores an item $\{TID_{MU}$, PID_{MU}, $R_{MU}\}$. The IS generates a ciphertext c by encrypting TID_{MU} using k_{IM} and a symmetric cryptographic algorithm. Later it sends a MIIS authentication response message to the MU. The message content is as the following, $\{ID_{IS}$, R_{IS}, B, A, c, $MAC\}$, where MAC is a value computed using a secure message authentication function λ by the equation MAC=$\lambda(ID_{IS}$, R_{IS}, B, A, c, $k_{IM})$.

On receiving the response message from the IS, the MU computes as bellow.

$PK_{IS} = H_1(ID_{IS}, R_{IS})PK_{HAS} + R_{IS}$; $K_{MI} = (a + s_{MU})(PK_{IS} + B)$. Then the shared session key k_{MI} is derived from the equation: k_{MI} =$f(K_{MI}$, PID_{MU}, $ID_{IS})$.

Next the MU checks whether the MAC equals to $\sigma(ID_{IS}$, R_{IS}, B, A, c, $k_{MI})$. If it does not hold, the IS fails to pass the authentication. Otherwise, the IS passes the authentication and a secure channel between the IS and the MU is established using the shared key. The MU decrypts c and stores TID_{MU}. Then neighbor network information of the MU can be acquired from the IS through the secure channel.

Notes.

The MU authentication is achieved by verifying the signature of the user. On the other side, the MU authenticates the IS by MAC generated using the shared key. It is easy to see that K_{MI}=K_{IM}.

Later, if the MU moves to another place and wants to access the IS again, the MU will uses TID$_{MU}$ as his identity. The ANIA protocol will be performed except that the message sent in Step (1) consists of $\{TID_{MU}$, A, t_{MU}, xP, $y\}$. Note that R_{MU} (a part of the MU's signature σ composed of $\{xP$, y, $R_{MU}\}$) is not sent in the message, since the PID_{MU} and $R_{MU}$are stored in the IS. The IS identifies the MU by the TID_{MU}, and it generates a new temporary identity $TID^*_{MU}$for the MU during the authentication procedure.

4 Security Analysis

We assume that the cryptography suites employed in our protocol are all secure, such as, hash function, message authentication function and ID-based signature scheme. Then our protocol is secure under the extended Canetti-Krawczyk (eCK) model [6].

Computational Diffie-Hellman (CDH) Assumption.

> Let G be a cyclic group generated by P, whose order is a prime q. View G as an additive group. The CDH assumption states that, given (P, aP, bP), for randomly chosen a, b∈$\{0, 1, 2, \ldots, q\text{-}1\}$, it is computationally intractable to compute the value abP [6].

Theorem 1. *Under the CDH assumption in the cyclic group G of prime order q, using a signature scheme sig and a message authentication function λ that are both existentially unforgeable under adaptively chosen-message attacks, the ANIA protocol is a secure authenticated key-exchange protocol with respect to the eCK model, when hash functions H_1, H_2 and key derivation function f are modeled as random oracles.*

Proof. Let Δ be any adversary against the ANIA protocol. We start by observing that since the session key sk is computed as sk=$f(\theta)$ for some 3-tuple θ, the adversary Δ has only two ways to distinguish sk from a random string:

1) Forging attack. At some point Δ queries f on the same 3-tuple θ.

2) Key-replication attack. Δ succeeds in forcing the establishment of another session that has the same session key as the test session.

If random oracles produce no collisions, the key-replication attack is impossible as equality of session keys implies equality of the corresponding 3-tuples (which are

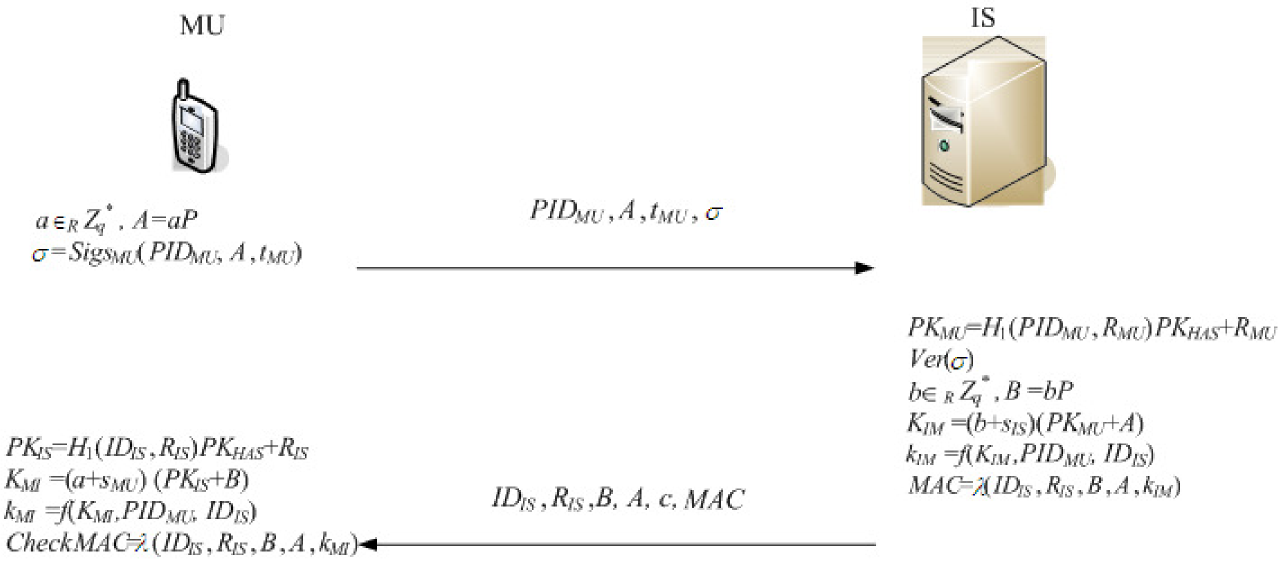

MU
IS
$a \in_R Z_q^*$, $A=aP$
$\sigma=Sigs_{MU}(PID_{MU}, A, t_{MU})$
$PID_{MU}, A, t_{MU}, \sigma$
$PK_{MU}=H_1(PID_{MU}, R_{MU})PK_{HAS}+R_{MU}$
$Ver(\sigma)$
$b \in_R Z_q^*$, $B=bP$
$K_{IM}=(b+s_{IS})(PK_{MU}+A)$
$k_{IM}=f(K_{IM}, PID_{MU}, ID_{IS})$
$MAC=\lambda(ID_{IS}, R_{IS}, B, A, k_{IM})$
$ID_{IS}, R_{IS}, B, A, c, MAC$
$PK_{IS}=H_1(ID_{IS}, R_{IS})PK_{HAS}+R_{IS}$
$K_{MI}=(a+s_{MU})(PK_{IS}+B)$
$k_{MI}=f(K_{MI}, PID_{MU}, ID_{IS})$
$CheckMAC\overset{?}{=}\lambda(ID_{IS}, R_{IS}, B, A, k_{MI})$

Figure 2: Anonymous authentication and key establishment with IS for MU

used to produce session keys). In turn, distinct key exchange sessions must have distinct 3-tuples. Therefore, if random oracles produce no collisions, Δ must perform a forging attack.

Case 1: Active attack. The adversary could break the security of the protocol via insertion of a message of its choice. In this case we will construct an adversary Ξ against the signature scheme *sig* or the message authentication function λ.

We only take the adversary against *sig* for example. The construction of attacker against λ is very similar. The input to Ξ consists of the parameters of the signature scheme, which includes access to a signature oracle. Ξ selects at random one party as MU. For session executed by MU, instead of using the MU's private key to compute the signatures, Ξ will make use of the signature oracle that it has access to in the signature security game that Ξ is simultaneously playing. Therefore, if the active attack occurs, Ξ succeeds in breaking the unforgeability of *sig*. By assumption forging a valid signature can only occur with negligible probability, the protocol is resilient against active attacks.

Case 2: Passive attack. In this case, the Test session has a matching session owned by another honest party. We show that if the adversary performs a successful forging attack, the CDH problem could be solved by a solver Ξ. The input to the Ξ is a CDH problem instance ($U = uP$, $V = vP$), where u, $v\in Z_q^*$ and U, $V\in G$. The goal of Ξ is to compute $\mathrm{CDH}(U, V) = uvP$. For simplicity, we use γ, ω and Γ, Ω denote the static secret keys s_{MU}, s_{IS} and public keys PK_{MU}, PK_{IS} respectively.

The adversary Δ is allowed to reveal a subset of (γ, a, ω, b), but it is not allowed to reveal both (γ, a) or both (ω, b). We only take the subcase for example that (γ, b) is revealed by Δ. Other subcases are similar.

Ξ selects random matching sessions executed by MU and IS, and modifies the experiment as follows. Ξ sets the ephemeral public keys of MU in the test session to be U, and sets the static public key of IS in the matching session to be V (namely, $A{=}U$, $\mathit{\Omega} = V$). If Δ wins the game, it must queries f on the same 3-tuple θ, thus it successfully forges $K{=}(\gamma{+}u)(v{+}b)P$. Then Ξ can solve the CDH problem as below: $\mathrm{CDH}(U, V){=}K{-}\gamma bP{-}\gamma V{-}bU$. With the hardness of the CDH assumption, the adversary could not win the experiment and hence the protocol is secure.

□

In the following, we further discuss some security properties of our protocol.

User Anonymity. In our scheme, the pseudo ID, instead of the MU's real identity, is used in access MIIS for privacy protection.

Key Freshness. The session key k_{MI} is computed from a function using random numbers from the MU and the IS respectively, which assures the freshness of session key.

Forward Secrecy. The random numbers used in session key generation are unpredictable for any party except the MU or the IS. Even if the intruder attacks long term secret information of the MU and the IS, he can not compromise the past random numbers and the past session keys.

Resistance to Replay Attack. Replay attack involves the passive capture of data and its subsequent retransmission to produce an unauthorized effect. A replay attack can be prevented by checking the timestamp or the MAC.

5 Performance Analysis

Computation and communication overheads are considered as two important metrics of authentication protocols. We present performance comparison of 802.21a proposal [5], SAM protocol [7], and ANIA protocol according to the metrics.

The computation overhead is the time cost of all the cryptography operations. Since the MU is always resource-constraints, we primarily take the MU's computation overhead into account. We take EAP-TLS [13] and TLS [2] as 802.21a proposal instances for 3-party case and 2-party case respectively. Here, TLS handshake is based on public key certificate and Diffie-Hellman key exchange. And public key related algorithms of 802.21a and SAM are all considered based on ECC, where ECDSA and ECDH for 802.21a, and ECDH for SAM.

To evaluate computation overhead of the mobile node, we implemented all cryptographic operations required in the two schemes using the Crypto++ Library (version 5.6.2) [1]. The cryptographic experiments were executed on a laptop with PIII 1.0 GHz CPU and 128MB RAM. Here the key length of the ECC system is set as 160 bits. In the experiment, SHA-160(or its variation) is introduced to implement hash functions and key derivation function, and AES-128 is introduced as the symmetric cipher used in the protocols. The mainly results are listed in Table 1.

Table 1: Mainly cryptographic operations and computation costs

Computation operations	**Notation**	**Time (ms)**
point multiplication	T_{PM}	1.532
random number generation	T_{RG}	0.072
symmetric encryption	T_{SE}	0.106
symmetric decryption	T_{SD}	0.106
hash value computation	T_{HC}	0.031
key derivation	T_{KD}	0.031

Table 2 shows the MU's computation costs of the four schemes during the handover authentication procedure. In the ANIA protocol, the MU needs: $1T_{HC}$ and $1T_{PM}$ for computing the IS's public key; $1T_{HC}$, $1T_{RG}$, and $1T_{PM}$ for message signature; $1T_{RG}$, $2\mathrm{T}_{PM}$, $1T_{KD}$ for key exchange; $1T_{HC}$ for MAC verification; $1T_{SD}$ for T_{ID}. From the table, we can conclude that the ANIA protocol is more efficient than 802.21a proposal, since 3 costly point multiplication operations are saved; and it is a little more complex than the SAM protocol because of one additional costly point multiplication operation.

As to communication performance, the HAS is not involved during the authentication between the MU and the IS in both 802.21a proposal 2-party case and the ANIA scheme. The SAM protocol and 802.21a proposal 3-party case both need the HAS to acts as an anchor for trust establishment. Since the MU now roams to a visited network which may be far away from his home network, communication between the MU and the HAS could take a long latency. Table 3 shows the message numbers needed between the related entities. From Table 3, we can see that ANIA performs better than other schemes.

We carried out some simulation experiments of the four schemes using OPNET 10.5 [10] to verify analysis above. For simplicity, only 2 WLANs (denoted as H and V) are used as the access network in the topology, and two ASs and one IS are deployed, where the servers are connected to the Internet as in Figure 1. The simulations run with 20˜100MUs and 10 APs uniformly distributed in each WLAN area for 5 minutes of simulation time. For the MIIS authentication request pattern, assume 20 percents of the MUs in one WLAN move into the other WLAN, and each roaming MU makes 10 requests randomly distributed over the whole simulation period. The simulation parameters are listed in Table 4. Here we mainly focus on the measurements of average authentication latency and the number of messages delivered in the network. The computation costs of MUs are considered in the simulation, while the computation costs of the servers are neglected because of their powerful processing abilities.

Figure 3 shows the average authentication latency of the four schemes as the number of MUs changes. We can see that the average authentication latency of those schemes become larger as the number of MUs increases. The reason is that the number of packets generated in the network increases as the number of MUs increases, which makes packets collision and retransmission happen more often. The ANIA protocol gets the shortest average authentication latency among those protocols in all scenarios. This suggests that the ANIA protocol is highly effective in authentication latency. Figure 4 shows the changes of the number of messages delivered in the network when the number of MUs changes. As we can see from the results, the number of messages delivered of 802.21a-3 increases sharply while that of other protocols increases smoothly as the number of MUs increases. It shows that the ANIA protocol delivers the smallest messages in the network in all scenarios. The simulation results indicate that the ANIA protocol has advantages in communication performance compared with other protocols.

6 Conclusion

In this paper, we focus specifically on security of MIIS, and propose a new anonymous access authentication protocol for MIIS. We apply an identity-based Schnorr like

Table 2: Message numbers between the related entities

Computation costs	**802.21a (2-party)**	**802.21a (3-party)**	**SAM**	**ANIA**
	$3T_{HC} + 7T_{PM}$ $+1T_{RG} + 1T_{KD}$	$3T_{HC} + 7T_{PM}$ $+1T_{RG} + 1T_{KD}$	$4T_{HC} + 3T_{PM} + 1T_{RG}$ $+1T_{KD} + 1T_{SE} + 1T_{SD}$	$3T_{HC} + 4T_{PM} + 2T_{RG}$ $+1T_{KD} + 1T_{SD}$
Total (ms)	10.917	10.917	5.101	6.502

Table 3: MUs computation costs of the four schemes

Message numbers	**802.21a (2-party)**	**802.21a (3-party)**	**SAM**	**ANIA**
Between MU and HAS	0	6	2	0
Between MU and IS	4	9	2	2

Table 4: Simulation parameters

WLAN area	300m∗300m
The number of APs in each WLAN	10
Coverage of AP	100m
The number of MUs in each WLAN	20∼ 100
The number of MIIS request for each MU	10
Simulation time	5 minutes

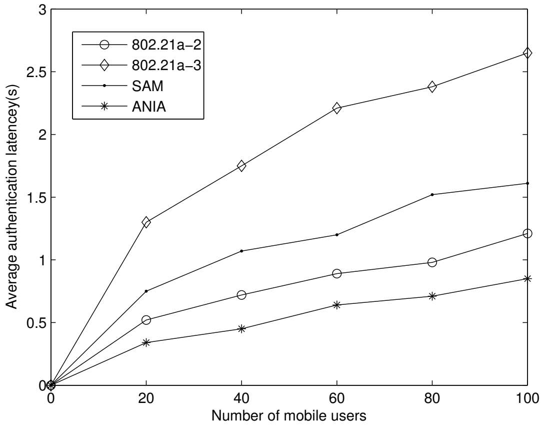


Figure 3: Comparison about average authentication latency

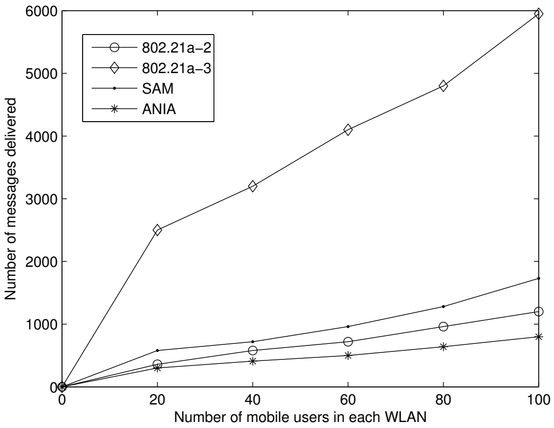


Figure 4: Comparison about number of messages delivered

signature for user authentication with a PID. The security and performance analysis shows that the proposed scheme has excellent performance. We will further analyze the performance of the proposed scheme in the future. Now we are making an effort to put up a real test-bed to evaluate performance of our protocol.

Acknowledgments

The authors would like to thank the anonymous reviewers and the editor for their comments that will help them to improve this paper. This work is supported by the National Natural Science Foundation of China (61201220, 61202389, 61309016, 61379150), and the Fundamental Research Funds for the Central Universities (Program No. JB140302).

References

[1] Cryptopp, *Crypto++ Library*, July 11, 2015. (`http://www.cryptopp.com/`)

[2] T. Dierks and E. Rescorla, *The Transport Layer Security (TLS) Protocol Ver. 1.2*, Technical Report RFC 5246, 2008.

[3] D. Galindo and F. D. Garcia, "A schnorr-like lightweight identity-based signature scheme," in *Proceedings of The Second International Conference on Cryptology in Africa (Africacrypt 2009)*, pp. 135–148, 2009.

[4] IEEE, *Media Independent Handover Services*, IEEE 802.21 Standard, 2009.

[5] IEEE, *IEEE Standard for Local and Metropolitan Area Networks: Media Independent Handover Services - Amendment for Security Extensions to Media Independent Handover Services and Protocol*, IEEE Standard, May 2012.

[6] B. LaMacchia, K. Lauter, and A. Mityagin, "Stronger security of authenticated key exchange," in *Proceedings of The First International Conference on Provable Security (ProvSec 2007)*, pp. 1–16, 2007.

[7] G. Li, J. Ma, and Q. Jiang, "SAM: Secure access of media independent information service with user anonymity," *EURASIP Journal on Wireless Communications and Networking*, vol. 2010, pp. 12, Apr. 2010.

[8] R. Marin-Lopez, F. Bernal-Hidalgo, S. Das, and et al., "A new standard for securing media independent handover: IEEE 802.21A," *IEEE Wireless Communications*, vol. 20, no. 6, pp. 82–90, 2013.

[9] T. Melia, G. Bajko, S. Das, N. Golmie, and J. Zuniga, *IEEE 802.21 Mobility Services Framework Design (MSFD)*, Technical Report RFC 5677, 2009.

[10] Opnet, *Opnet*, July 11, 2015. (`http://www.opnet.com/`)

[11] I. Saadat, F. Buiati, D. Rupèrez Cañas, L. Javier, and G.Villalba, "Overview of IEEE 802.21 security issues for mih networks," in *Proceedings of International Conference on Information Technology (ICIT'11)*, pp. 196–214, 2011.

[12] S. Saha and D. Lagutin, "PLA-MIH: A secure IEEE 802.21 signaling scheme," in *Proceedings of IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob'09)*, pp. 252–257, 2009.

[13] D. Simon, B. Aboba, and R. Hurst, *The EAP TLS Authentication Protocol*, Technical Report RFC 5216, 2008.

[14] J. Won, M. Vadapalli, C. Cho, and V. C. M. Leung, "Secure media independent handover message transport in heterogeneous networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2009, pp. 15, 2009.

[15] S. Yoo, D. Cypher, and N. Golmie, "Timely effective handover mechanism in heterogeneous wireless networks," *Wireless Personal Communications*, vol. 52, no. 3, pp. 449–475, 2010.

Guangsong Li received the B.S. degree in applied mathematics from Information Engineering University, Zhengzhou, P. R. China in 1999, and M. S. degree in applied mathematics from Information Engineering University in 2002, and the Ph. D. degree in Cryptography from Information Science and Technology Institute, Zhengzhou, P. R. China, in 2005. Now he is an associate professor of State Key Laboratory of Mathematical Engineering and Advanced Computing, Zhengzhou. His current interests include mobile communication, wireless security, and digital rights management.

Qi Jiang received the B.S. degree in Computer Science from Shaanxi Normal University in 2005 and Ph.D. degree in Computer Science from Xidian University in 2011. He is now an associate professor of the School of Computer Science and Technology, Xidian University. His research interests include security protocols and wireless network security, etc.

Yanan Shi received his M.S. degrees in applied mathematics from the Zhengzhou Information Science and Technology Institute, China, in 2008. She is currently a lecturer of State Key Laboratory of Mathematical Engineering and Advanced Computing, Zhengzhou, China. Her research fields include cryptography and information security.

Fushan Wei received his M.S. and Ph.D. degrees in applied mathematics from the Zhengzhou Information Science and Technology Institute, China, in 2008 and 2011, respectively. He is currently a lecturer of State Key Laboratory of Mathematical Engineering and Advanced Computing, Zhengzhou, China. His research fields include cryptography and information security.

Guide for Authors
International Journal of Network Security

IJNS will be committed to the timely publication of very high-quality, peer-reviewed, original papers that advance the state-of-the art and applications of network security. Topics will include, but not be limited to, the following: Biometric Security, Communications and Networks Security, Cryptography, Database Security, Electronic Commerce Security, Multimedia Security, System Security, etc.

1. Submission Procedure

Authors are strongly encouraged to submit their papers electronically by using online manuscript submission at http://ijns.jalaxy.com.tw/.

2. General

Articles must be written in good English. Submission of an article implies that the work described has not been published previously, that it is not under consideration for publication elsewhere. It will not be published elsewhere in the same form, in English or in any other language, without the written consent of the Publisher.

2.1 Length Limitation:

All papers should be concisely written and be no longer than 30 double-spaced pages (12-point font, approximately 26 lines/page) including figures.

2.2 Title page

The title page should contain the article title, author(s) names and affiliations, address, an abstract not exceeding 100 words, and a list of three to five keywords.

2.3 Corresponding author

Clearly indicate who is willing to handle correspondence at all stages of refereeing and publication. Ensure that telephone and fax numbers (with country and area code) are provided in addition to the e-mail address and the complete postal address.

2.4 References

References should be listed alphabetically, in the same way as follows:

For a paper in a journal: M. S. Hwang, C. C. Chang, and K. F. Hwang, ``An ElGamal-like cryptosystem for enciphering large messages," *IEEE Transactions on Knowledge and Data Engineering*, vol. 14, no. 2, pp. 445--446, 2002.

For a book: Dorothy E. R. Denning, *Cryptography and Data Security*. Massachusetts: Addison-Wesley, 1982.

For a paper in a proceeding: M. S. Hwang, C. C. Lee, and Y. L. Tang, ``Two simple batch verifying multiple digital signatures," in *The Third International Conference on Information and Communication Security (ICICS2001)*, pp. 13--16, Xian, China, 2001.

In text, references should be indicated by [number].

Subscription Information

Individual subscriptions to IJNS are available at the annual rate of US$ 200.00 or NT 7,000 (Taiwan). The rate is US$1000.00 or NT 30,000 (Taiwan) for institutional subscriptions. Price includes surface postage, packing and handling charges worldwide. Please make your payment payable to "Jalaxy Technique Co., LTD." For detailed information, please refer to http://ijns.jalaxy.com.tw or Email to ijns.publishing@gmail.com.