IJNS

International Journal of Network Security

ISSN 1816-353X (Print)　　Vol. 21, No. 2 (Mar. 2019)
ISSN 1816-3548 (Online)

INTERNATIONAL JOURNAL OF NETWORK SECURITY

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

23-75, P.O. Box, Taichung, Taiwan 40199, R.O.C.

International Journal of Network Security

Vol. 21, No. 2 (Mar. 1, 2019)

Provably Secure Searchable Attribute-Based Authenticated Encryption Scheme

Zhenhua Liu and Yaqing Fan
(Corresponding author: Zhenhua Liu)

School of Mathematics and Statistics, Xidian University
No. 2 South Taibai Road, Xi'an, Shaanxi 710071, P.R. China
(Email: zhualiu@hotmail.com)

(Received Sept. 26, 2017; revised and accepted Mar. 27, 2018)

Abstract

In cloud storage system, attribute-based encryption can support fine-gained access control over encrypted data. Furthermore, searchable attribute-based encryption can allow data users to retrieve encrypted data from a cloud storage system. However, these encryption algorithms cannot provide authenticity. In this paper, we propose a new concept——searchable attribute-based authenticated encryption and establish its security model framework. Then, we embed ingeniously search mechanism into key-policy attribute-based signcryption, and present a concrete searchable attribute-based authenticated encryption scheme. Finally, according to the proposed framework, our scheme is proven to achieve (1) Ciphertext indistinguishability under the Decisional Bilinear Diffie-Hellman Exponent hardness assumption; (2) Existential unforgeability based on the hardness assumption of Computational Diffie-Hellman Exponent problem; (3) Selective security against chosen-keyword attack under the Decisional Linear hardness assumption; (4) Keyword secrecy based on the one-way hardness of hash function.

Keywords: Attribute-Based Encryption; Authenticated Encryption; Searchable Encryption; Signcryption

1 Introduction

With the development of cloud computing, many data owners store their data in the cloud server for simplifying local IT management and reducing the cost. Although cloud services have various advantages, they will bring security and privacy concerns to upload sensitive information to the cloud server [14]. Therefore, it is essential to encrypt sensitive data before uploading them to the remote server.

Traditional public key encryption technology can protect the confidentiality of data, but cannot provide the data sharing service. However, as a kind of "one-to-many" public key encryption, attribute-based encryption can solve this problem. Thus, attribute-based encryption is considered as one of the most appropriate encryption technology to achieve the data confidentiality and expressive fine-grained access control in cloud system.

Sahai and Waters [17] first introduced the concept of attribute-based encryption, and proposed a concrete attribute-based encryption scheme which only supports threshold access policy. In order to support more flexible access policy, Goyal *et al.* [7] first presented a key-policy attribute-based encryption (KP-ABE) scheme in 2006. In key-policy attribute-based encryption, a private or decryption key is associated with access control policy and a ciphertext is computed with respect to a set of attributes. On the contrary, if a ciphertext specifies an access control policy and a private or decryption key is associated with a set of attributes, such an attribute-based encryption is called as ciphertext-policy attribute-based encryption (CP-ABE). Bethencourt *et al.* [1] proposed the first ciphertext-policy attribute-based encryption scheme in 2007. Due to the suitable application for cloud computing, a number of attribute-based encryption schemes [3, 4, 11, 12, 23, 24] have been proposed to obtain better expressive, efficiency and security.

Attribute-based encryption provides the data confidentiality and expressive fine-grained access control. Nevertheless, encryption may prevent the ciphertext from being searched quickly. To solve this problem, Song *et al.* [19] first proposed the concept of searchable encryption which provides a fundamental approach to search over encrypted cloud data. Further, Boneh *et al.* [2] presented the first public key encryption scheme with keyword search scheme, in which one can search the encrypted data by a keyword. Since then, a number of searchable public key encryption schemes [8–10, 21] have been proposed to enrich the search feature of scenarios and improve security and efficiency.

However, users are considered to be legitimate in the above search schemes, which is not suitable for more practical scenarios. The reason is that without access control, all users in these systems have not been restricted access to the entire database. Thus, the confidentiality of sensitive data may be compromised. In order to eliminate this threat, in 2013, Wang *et al.* [22] constructed a ciphertext-policy attribute-based encryption scheme supporting keyword search function. In their scheme, data owners encrypt data with an access policy, generate the index for the corresponding keyword collection, and then store them to the cloud server. It is only when an authorized user's certificate meets the access policy that she or he can search and decrypt the encrypted data. In the recent years, attribute-based searchable encryption schemes [5, 13, 20, 27] have been made great development.

Furthermore, in 2014, to assure the cloud server to execute faithfully the search operations on behalf of the data users, Zheng *et al.* [28] proposed a novel cryptographic concept of verifiable attribute-based keyword search, and constructed a verifiable key-policy attribute-based searchable encryption scheme and a verifiable ciphertext-policy attribute-based searchable encryption scheme based on the access tree. Such search encryption schemes can support fine-grained access control and search, and verify the cloud server's faith. But these schemes can neither authenticate data owners nor guarantee the availability of the shared data. Thus, it is necessary to provide the authentication for searchable attribute-based encryption schemes.

It is well known that encryption can guarantee the confidentiality, signature can provide the authenticity, and signcryption can support these two functionalities in public key cryptosystems. In 1997, Zheng *et al.* [29] first introduced the concept of signcryption, which is a logical incorporation of signature and encryption. Subsequently, in order to ensure fine-grained access control for the data in the cloud and authenticate data owners, Gagné *et al.* [6] presented the first attribute-based signcryption (ABSC) scheme. Since then, a number of attribute-based signcryption [15, 25, 26] have been proposed. Though attribute-based signcryption schemes can ensure that data owners share the encrypted and authenticated data with data users, these schemes cannot take the retrieval of the signcryption data into account. As a result, it is important to support the searchability for attribute-based signcryption schemes.

1.1 Our Contribution

The main contribution of this paper can be summarized as follows:

- We introduce a novel concept of searchable attribute-based authenticated encryption (SAAE), which can achieve expressive fine-grained access control, efficient data retrieval and authentication, simultaneously.

- We replace access tree structure with linear secret sharing scheme to modify Zheng *et al.*'s attribute-based keyword search method [28], embed such a search mechanism into Rao *et al.*'s key-policy attribute-based signcryption scheme [16], and propose a searchable attribute-based authenticated encryption scheme. More to the point, we use an identical secret key to generate a trapdoor and decrypt a ciphertext, which promotes the logical combination of the above two schemes [16, 28]. Hence, the cost is significantly reduced than the cumulative cost of signcryption and search.

- Our scheme is proven to achieve selective encryption attribute set secure against chosen-keyword attack and keyword secrecy in the standard model rather than random model [28].

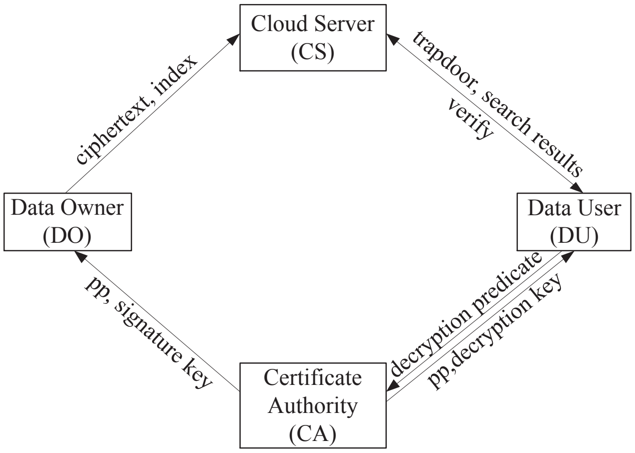


Figure 1: The system framework

2 Generic Framework and Its Security Models

2.1 System Framework

We consider a cryptographic cloud storage system supporting fine-grained access control, data retrieval and data authentication over encrypted data.

As shown in Figure 1, the system framework consists of four entities: Certificate Authority (CA), Data Owners (DO), Data Users(DU), and Cloud Server (CS).

CA: Certificate authority is a global trusted authority. It is responsible for generating the public parameters and the master secret key. Meanwhile, CA is also in charge of generating private/secret keys for participants in the system, such as signing keys for data owners and decryption keys for data users.

DO: Data owners take charge of providing the shared data. They first obtain signing keys from CA according to their roles. Then they select an attribute set to signcrypt their personal data and generate corresponding keyword index. Finally, data owners upload the signcrypted data and the encrypted index to the cloud server.

DU: Data users are some entities who want to access the encrypted data in the cloud server. In order to search data of interest, data users first need to generate a trapdoor and send it to the cloud server. Then, with the help of the cloud server, data users complete the data retrieval. Finally, data users check the validity of the returned results and decrypt the valid ciphertext.

CS: Cloud server is honest-but-curious. It is responsible for storing the ciphertext and index for data owners, and providing the data retrieval service for data users.

2.2 Generic Framework

Let $\mathcal{M}$ be a message space, $\mathcal{G}$ be an access structure space, $\mathcal{U}_s$ be a space of signing attributes, and $\mathcal{U}_e$ be a space of encryption attributes. A generic searchable attribute-based authenticated encryption scheme consists of the following eight algorithms:

- **Setup**$(k)\rightarrow(pp, msk)$: Given a security parameter k, CA creates the public parameters pp and the master secret key msk, where msk is owned by CA.
- **sExtract**$(msk, \mathbb{A}_s)\rightarrow sk_{\mathbb{A}_s}$: Taking as input the master secret key msk and a signing access structure $\mathbb{A}_s$ over a set $W_s \subset \mathcal{U}_\text{s}$, CA outputs a signing key $sk_{\mathbb{A}_s}$ to a legitimate data owner over a secret channel.
- **dExtract**$(msk, \mathbb{A}_d)\rightarrow sk_{\mathbb{A}_d}$: On input the master secret key msk and a decryption access structure $\mathbb{A}_d$, CA outputs a decryption key $sk_{\mathbb{A}_d}$ to the user over a secret channel.
- **Signcrypt**$(pp, m, sk_{\mathbb{A}_s}, W_e, W_s) \rightarrow CT$: Data owner performs this algorithm to signcrypt a message $m \in \mathcal{M}$ with the input the public parameters pp, a signing key $sk_{\mathbb{A}_s}$ with the signing attribute set W_s, and an encryption attribute set W_e.
- **GenIndex**$(W_e, pp, kw) \rightarrow I$: Data owner chooses a keyword kw for the message m, then encrypts the keyword kw with the input pp and W_e, then outputs a keyword ciphertext or index I.
- **GenToken**$(sk_{\mathbb{A}_d}, pp, kw') \rightarrow T$: Given an interested keyword kw', data user executes this algorithm with the input $sk_{\mathbb{A}_d}$ and pp to generate a keyword ciphertext or corresponding token T.
- **Search**$(I, T) \rightarrow CT'$: After gaining the token T, CS first matches it with the index I, then returns the relevant search results CT' to data user.
- **Unsigncrypt**$(pp, CT', sk_{\mathbb{A}_d}, W_s) \rightarrow m$: The algorithm first checks the validity of ciphertext CT' with the input signing attribute set W_s and public parameters pp. If CT' can pass the verification, the algorithm continues with the input $sk_{\mathbb{A}_d}$ and returns to data user the message m corresponding to CT'.

2.3 Security Model

The security of the proposed scheme is considered from two aspects. On the one hand, the security of the authenticated encryption or signcryption consists of the indistinguishability of data ciphertext against selective encryption attribute set and chosen-ciphertext attacks and the unforgeability of signature against selective signature attribute set and chosen-message attacks [16].

On the other hand, the security of searchable encryption is composed of the indistinguishability of keyword ciphertext or index against selective encryption attribute set and chosen-keyword attacks, and the secrecy of keyword against chosen-token attacks [28].

1) **Indistinguishability of Data Ciphertext**
We formalize the indistinguishability of data ciphertext against selective encryption attribute set and adaptive chosen-ciphertext attacks by the following game between a challenger $\mathcal{C}$ and an adversary $\mathcal{A}$:

Init: The challenger $\mathcal{C}$ gives the space of signature attributes $\mathcal{U}_s$ and the space of encryption attributes $\mathcal{U}_e$. Next, the adversary $\mathcal{A}$ chooses a challenge encryption attribute set W_e^* and send it to $\mathcal{C}$.

Setup: Given a security parameter k, $\mathcal{C}$ runs the **Setup**(k) algorithm to output the public parameters pp, while keeps the master secret key msk to himself.

Phase 1: $\mathcal{A}$ queries the following oracles for polynomial times:

- $\mathcal{O}_{sE}(\mathbb{A}_s)$: On input a signature access structure $\mathbb{A}_s$, $\mathcal{C}$ runs **sExtract**$(msk, \mathbb{A}_s)$ to generate a signing key $sk_{\mathbb{A}_s}$ and sends it to $\mathcal{A}$.
- $\mathcal{O}_{dE}(\mathbb{A}_d)$: On input a decryption access structure $\mathbb{A}_d$ such that $W_e^* \notin \mathbb{A}_d$, $\mathcal{C}$ runs **dExtract**$(msk, \mathbb{A}_d)$ to generate a decryption key $sk_{\mathbb{A}_d}$ for $\mathcal{A}$.
- $\mathcal{O}_{SC}(m, W_e, W_s)$: Given a message m, an encryption attributes set W_e, and a signature attribute set W_s, $\mathcal{C}$ selects a signature access structure $\mathbb{A}_s$ with the restriction $W_s \in \mathbb{A}_s$ and computes $sk_{\mathbb{A}_s}$ through algorithm **sExtract**$(msk, \mathbb{A}_s)$. Finally, $\mathcal{C}$ sends the ciphertext CT generated by running algorithm **Signcryption**$(pp, m, sk_{\mathbb{A}_s}, W_e, W_s)$ to $\mathcal{A}$.
- $\mathcal{O}_{US}(pp, CT, sk_{\mathbb{A}_d}, W_s)$: Taking as input ciphertext CT and decryption access structure $\mathbb{A}_d$, $\mathcal{C}$ runs **dExtract**$(msk, \mathbb{A}_d)$ to obtain the decryption key $sk_{\mathbb{A}_d}$ and runs **Unsigncrypt**$(pp, CT, sk_{\mathbb{A}_d}, W_s)$ to output the message m.

Challenge: $\mathcal{A}$ sends two equal length messages m_0, m_1 and a signature attribute set W_s to $\mathcal{C}$.

Then, $\mathcal{C}$ chooses a signature access structure $\mathbb{A}_s$ such that $W_s \in \mathbb{A}_s$ and runs the algorithm **sExtract**$(msk, \mathbb{A}_s)$ to generate the signing key $sk_{\mathbb{A}_s}$. Finally, $\mathcal{C}$ selects $b \in \{0,1\}$ and runs **Signcrypt**$(pp, m_b, sk_{\mathbb{A}_s}, W_e, W_s)$ to output the ciphertext CT_b.

Phase 2: $\mathcal{A}$ continues to query the oracles as in Phase 1. The restriction is that $W_e^* \notin \mathbb{A}_d$ for any $\mathbb{A}_d$ in $\mathcal{O}_{US}(pp, CT, sk_{\mathbb{A}_d}, W_s)$.

Guess: $\mathcal{A}$ returns a guess $b' \in \{0,1\}$. It wins the game if $b' = b$.

The advantage of $\mathcal{A}$ in the above game is defined as

$$Adv\left(1^k\right) = \left|\Pr\left[b = b'\right] - \frac{1}{2}\right|$$

Definition 1. *A searchable attribute-based authenticated encryption scheme can achieve the ciphertext indistinguishability under selective encryption attribute set and adaptive chosen-ciphertext attacks if an adversary $\mathcal{A}$ wins the above game with a negligible advantage.*

2) **Unforgeability of Signature**

We formalize the unforgeability of signature against selective signature attribute set and adaptive chosen-message attacks by the following game between a challenger $\mathcal{C}$ and an adversary $\mathcal{A}$.

Init: The challenger $\mathcal{C}$ gives the space of signature attributes $\mathcal{U}_s$ and the space of encryption attributes $\mathcal{U}_e$. Then, the adversary $\mathcal{A}$ chooses a signature attribute set W_s^* and send it to $\mathcal{C}$.

Setup: Given a security parameter k, $\mathcal{C}$ runs the algorithm **Setup**(k) to output the public parameter pp and keeps the master secret key msk to himself.

Phase: $\mathcal{A}$ can query the following oracles for polynomial times:

- $\mathcal{O}_{sE}(\mathbb{A}_s)$: Given a signature access structure $\mathbb{A}_s$ such that $W_s^* \notin \mathbb{A}_s$, $\mathcal{C}$ computes $sk_{\mathbb{A}_s}$ by running **sExtract**$(msk, \mathbb{A}_s)$ and sends it to $\mathcal{A}$.
- $\mathcal{O}_{dE}(\mathbb{A}_d)$: Taking as input a decryption access structure $\mathbb{A}_d$, $\mathcal{C}$ runs the algorithm **dExtract**$(msk, \mathbb{A}_d)$ to output a decryption key $sk_{\mathbb{A}_d}$.
- $\mathcal{O}_{SC}(m, W_e, W_s)$: Given a message m, an encryption attribute set W_e and a signature attribute set W_s, $\mathcal{C}$ selects a signature access structure $\mathbb{A}_s$ with the restriction $W_s \in \mathbb{A}_s$, obtains $sk_{\mathbb{A}_s}$ by the oracle $\mathcal{O}_{sE}(\mathbb{A}_s)$, and runs **Signcrypt**$(m, pp, sk_{\mathbb{A}_s}, W_e)$ to generate CT for $\mathcal{A}$.
- $\mathcal{O}_{US}(CT, \mathbb{A}_d)$: On input ciphertext CT and decryption access structure $\mathbb{A}_d$, $\mathcal{C}$ obtains a decryption key $sk_{\mathbb{A}_d}$ by the oracle $\mathcal{O}_{dE}(\mathbb{A}_d)$ and runs the algorithm **Unsigncrypt**$(pp, CT, sk_{\mathbb{A}_d}, W_s)$ to output the message m.

Forgery: $\mathcal{A}$ sends a forgery $CT^*_{(W_s^*, W_e)}$ of message m^* and a decryption access structure $\mathbb{A}_d$. If the ciphertext $CT^*_{(W_s^*, W_e)}$ is valid and cannot be gained from $\mathcal{O}_{SC}(m, W_e, W_s)$, $\mathcal{A}$ wins the game.

The advantage of $\mathcal{A}$ in the above game is defined as

$$Adv_{\mathcal{A}} = \Pr[\mathcal{A} \ \ wins].$$

Definition 2. *A searchable attribute-based authenticated encryption scheme can achieve the unforgeability of signature against selective signature attribute set and adaptive chosen-message attacks if adversary $\mathcal{A}$ wins the above game with a negligible advantage.*

3) **Indistinguishability of Keyword Ciphertext** We formalize the indistinguishability of keyword ciphertext or index against selective encryption attribute set and adaptive chosen-keyword attacks by the following game between a challenger $\mathcal{C}$ and an adversary $\mathcal{A}$.

Init: $\mathcal{C}$ gives the space of encryption attributes $\mathcal{U}_e$ and $\mathcal{A}$ selects a challenge encryption attribute set W_e^* and sends it to $\mathcal{C}$.

Setup: Given a security parameter k, $\mathcal{C}$ runs algorithm **Setup**(k) to output the public parameters pp, while the master secret key msk is owned by himself.

Phase 1: $\mathcal{A}$ queries the following oracles for polynomial times. Meanwhile, $\mathcal{C}$ maintains a keyword list L_{kw}, which is initially empty.

- $\mathcal{O}_{dE}(\mathbb{A}_d)$: Taking as input a decryption access structure $\mathbb{A}_d$ such that $W_e^* \notin \mathbb{A}_d$, $\mathcal{C}$ runs **dExtract**$(msk, \mathbb{A}_d)$ to output a decryption key $sk_{\mathbb{A}_d}$. Otherwise, $\mathcal{C}$ outputs $\perp$.
- $\mathcal{O}_{GT}(\mathbb{A}_d, kw)$: $\mathcal{C}$ first runs $\mathcal{O}_{dE}(\mathbb{A}_d)$ to output a decryption key $sk_{\mathbb{A}_d}$ and runs **GenToken**$(sk_{\mathbb{A}_d}, kw)$ to return to $\mathcal{A}$ a token T. If the attribute set W_e^* satisfies the access structure $\mathbb{A}_d$, $\mathcal{C}$ adds kw to L_{kw}.

Challenge: $\mathcal{A}$ sends two equal length keywords kw_0 and kw_1 such that $kw_0, kw_1 \notin L_{kw}$ to $\mathcal{C}$. Then, $\mathcal{C}$ randomly picks a bit $b \in \{0,1\}$ and runs the algorithm **GenIndex**(W_e^*, kw_b) to generate I for $\mathcal{A}$.

Phase 2: $\mathcal{A}$ issues a series of queries as in Phase 1. The restriction is that if $W_e^* \in \mathbb{A}_d$, $\mathcal{A}$ cannot query $\mathcal{O}_{GT}$ with $(\mathbb{A}_d, kw_0)$ or $(\mathbb{A}_d, kw_1)$.

Guess: $\mathcal{A}$ outputs a guess $b' \in \{0,1\}$. If $b' = b$, $\mathcal{A}$ wins the game.

The advantage of $\mathcal{A}$ in breaking the above scheme is defined as

$$Adv\left(1^k\right) = \left|Pr\left[b = b'\right] - \frac{1}{2}\right|$$

Definition 3. *A searchable attribute-based authenticated encryption scheme can achieve the indistinguishability of keyword ciphertext or index against selective encryption attribute set and adaptive chosen-keyword attacks if adversary $\mathcal{A}$ wins the above game with a negligible advantage.*

4) **Keyword Secrecy**

Finally, we formalize the secrecy of keyword against adaptive chosen-token attacks by the following game between a challenger $\mathcal{C}$ and an adversary $\mathcal{A}$.

Setup: Given a security parameter k, $\mathcal{C}$ runs the algorithm **Setup(k)** to generate the public parameters pp and master secret key msk.

Phase: $\mathcal{A}$ issues the following queries for $|q|$ times.

- $\mathcal{O}_{dE}(\mathbb{A}_d, msk)$: On input a decryption access structure $\mathbb{A}_d$, $\mathcal{C}$ returns to $\mathcal{A}$ a corresponding secret key $sk_{\mathbb{A}_d}$ and adds $\mathbb{A}_d$ to the list $L_{\mathbb{A}_d}$, which is initially empty.
- $\mathcal{O}_{GT}(\mathbb{A}_d, kw)$: Given a decryption access structure $\mathbb{A}_d$, $\mathcal{C}$ generates a secret key $sk_{\mathbb{A}_d}$ by running $\mathcal{O}_{dE}(\mathbb{A}_d, msk)$ and returns to $\mathcal{A}$ a token T by the algorithm **GenToken($sk_{\mathbb{A}_d}, kw$)**.

Challenge: $\mathcal{A}$ submits a challenge W_e^* to $\mathcal{C}$. Then, $\mathcal{C}$ selects kw^* and $\mathbb{A}_d^*$ such that $W_e^* \in \mathbb{A}_d^*$. $\mathcal{C}$ runs **GenIndex(pp, W_e^*, kw^*)** and **GenToken($sk^*_{\mathbb{A}_d}, kw^*$)** to return $\mathcal{A}$ index I^* and token T^*. Note that $\forall \mathbb{A}_d \in L_{\mathbb{A}_d}, W_e^* \notin \mathbb{A}_d$.

Guess: After issuing q distinct keywords, $\mathcal{A}$ outputs a keyword kw' and wins the keyword secrecy game if $kw' = kw^*$.

Definition 4. *A searchable attribute-based authenticated encryption scheme can achieve the secrecy of keyword against adaptive chosen-token attacks if the advantage of $\mathcal{A}$ in breaking the above keyword secrecy game is at most $\frac{1}{|\mathcal{M}|-|q|} + \varepsilon$, where $|q|$ denotes the number of queried keywords, ε is a negligible probability in security parameter l, and $\mathcal{M}$ is the message space.*

3 Our Concrete Construction

By using Rao *et al.*'s key-policy attribute-based signcryption scheme [16] and modifying Zheng *et al.*'s attribute-based keyword search method [28], we construct a concrete scheme for searchable attribute-based authenticated encryption as follows:

3.1 The Proposed Scheme

Setup: Given the secure parameter k, CA outputs two cyclic groups $\mathbb{G}_1, \mathbb{G}_2$ of an prime order p, a generator g of $\mathbb{G}_1$, and a map $e : \mathbb{G}_1 \times \mathbb{G}_1 \to \mathbb{G}_2$. Let $\mathcal{U}_e = \{att_j\}$ and $\mathcal{U}_s = \{att'_j\}$ be the set of encryption and signature attributes, respectively. CA chooses four one-way, collision-resistant hash functions $H_1 : \mathbb{G}_2 \times \mathbb{G}_1 \times \{0,1\}^{l_\tau} \to \{0,1\}^*$, $H_2 : \{0,1\}^* \to \{0,1\}^l$, $H_3 : \mathbb{G}_1 \to \mathbb{Z}_p$ and $H_4 : \{0,1\}^* \to \mathbb{Z}_p$, where l is large enough so that the hash functions are collision resistant and $l_\tau \approx 40$. CA randomly chooses $a, b, c \leftarrow \mathbb{Z}_p$, $T_0, K_0, \delta_1, \delta_2, y_0, y_1, \cdots, y_l \in \mathbb{G}_1$, $T_j \in \mathbb{G}_1$ for each signature attribute $att'_j \in \mathcal{U}_s$, $K_j \in \mathbb{G}_1$ for each encryption attribute $att_j \in \mathcal{U}_e$, and sets public parameters pp and master secret key msk as follows:

$$\begin{aligned} pp &= \left\{ \begin{array}{l} e(g,g)^{ac}, g^a, g^b, g^c, \delta_1, \delta_2, \{H_i\}_{i=1}^{i=4}, \{y_i\}_{i\in[l]}, \\ T_0, K_0, y_0, \left\{T_j \left| att'_j \in \mathcal{U}_s \right.\right\}, \left\{K_j \left| att_j \in \mathcal{U}_e \right.\right\} \end{array} \right\}, \\ msk &= \{a, b, c\} \end{aligned}$$

sEtract: On input a signature predicate(S, ρ), where S is an $l_s \times n_s$ matrix and i-th row (i.e $\vec{S}_i$) is associated with an attribute $att'_{\rho(i)}$. Then, CA chooses a random vector $\vec{v}_s = (ac, v_2, \cdots, v_{n_s}) \in \mathbb{Z}_p^{n_s}$ and sets $\left\{\lambda_{\rho(i)} = \vec{S}_i \cdot \vec{v}_s | i \in [l_s]\right\}$. For each row $i \in [l_s]$, CA selects $r_i \in \mathbb{Z}_p$ at random and calculates

$$\begin{aligned} D_{s,i} &= g^{\lambda_{\rho(i)}}\left(T_0 T_{\rho(i)}\right)^{r_i} \\ D'_{s,i} &= g^{r_i} \\ D''_{s,i} &= \left\{D''_{s,i,j} = T_j^{r_i}, \forall att'_j \in \mathcal{U}_s \backslash att'_{\rho(i)}\right\} \end{aligned}$$

Finally, the signature key is set as:

$$sk_{(S,\rho)} = \left\{(S, \rho), \left\{D_{s,i}, D'_{s,i}, D''_{s,i}\right\}_{i\in[l_s]}\right\}.$$

dExtract: Take as input a decryption predicate (D, φ), where D is an $l_e \times n_e$ matrix and i-th row (i.e $\vec{D}_i$) is associated with an attribute $att_{\varphi(i)}$. Then, CA chooses a random vector $\vec{v}_e = \left(ac, v'_2, \cdots, v'_{n_e}\right) \in \mathbb{Z}_p^{n_e}$ and sets $\left\{\lambda_{\varphi(i)} = \vec{D}_i \cdot \vec{v}_e \left| i \in [l_e]\right.\right\}$. For each row $i \in [l_e]$, CA selects $\tau_i \in \mathbb{Z}_p$ and computes

$$\begin{aligned} D_{e,i} &= g^{\lambda_{\varphi(i)}}\left(K_0 K_{\varphi(i)}\right)^{\tau_i} \\ D'_{e,i} &= g^{\tau_i} \\ D''_{e,i} &= \left\{D''_{e,i,j} = K_j^{r_i}, \forall att_j \in \mathcal{U}_e \backslash att_{\varphi(i)}\right\} \end{aligned}$$

Finally, CA sets the decryption key as follows:

$$sk_{(D,\varphi)} = \left\{(D, \varphi), \left\{D_{e,i}, D'_{e,i}, D''_{e,i}\right\}_{i\in[l_e]}\right\}.$$

Signcrypt: For each legitimate data owner, he holds an authorized signature attribute set W_s, which satisfies the signature predicate (S, ρ). Therefore, data owner can find a coefficient set

$\{w_i : i \in I_s\}$ such that $\sum_{i \in I_s} w_i \lambda_{\rho(i)} = ac$, where $I_s = \left\{ i \in [l_s] \left| att'_{\rho(i)} \in W_s \right. \right\}$. Note that the secret shares $\left\{ \lambda_{\varphi(i)} \right\}_{i \in [l_s]}$ and the secret value ac are not explicitly known to the data owner [16]. To signcrypt a message $m \in \{0,1\}^{l_m}$, data owner chooses an encryption attribute set W_e which describes the target users. Then, he randomly selects $\theta, \vartheta \in \mathbb{Z}_p$ and calculates

$$\begin{aligned} C_1 &= g^{\theta}, C_2 = \left(\prod_{att_j \in W_e} K_0 K_j \right)^{\theta}, \sigma_1 = \left(g^{\theta}\right)^{\vartheta}, \\ C_3 &= H_1\left(e(g,g)^{ac\theta}, \sigma_1, \tau \right) \oplus m, \mu = H_3\left(C_1\right) \end{aligned}$$

where $\tau \in \{0,1\}^{l_\tau}$. Next, he picks $\xi \in \mathbb{Z}_p$ at random, computes

$$\begin{aligned} \sigma_2 &= g^{\xi} \prod_{i \in I_s} \left(D'_{s,i}\right)^{w_i}, C_4 = \left(\delta_1^{\mu} \delta_2\right)^{\theta} \\ (k_1, \cdots, k_l) &= H_2\left(\sigma_2 || \tau || W_s || W_e\right) \\ \beta &= H_4\left(\sigma_1 || C_2 || C_3 || C_4 || W_s || W_e\right) \\ \sigma_3 &= \prod_{i \in I_s} \left(D_{s,i} \cdot \prod_{att'_j \in W_s, j \neq \rho(i)} D''_{s,i,j} \right)^{w_i} \\ &\quad \cdot \left(T_0 \prod_{att'_j \in W_s} T_j \right)^{\xi} \cdot \left(y_0 \prod_{i \in [l]} y_i^{k_i} \right)^{\theta} \cdot C_4^{\beta\vartheta} \end{aligned}$$

and sets $\sigma_4 = \tau$. Finally, the signcryption of m is set as:

$$CT = \{W_s, W_e, C_1, C_2, C_3, C_4, \sigma_1, \sigma_2, \sigma_3, \sigma_4\}$$

GenIndex: In order to quickly search the encrypted data when needed, data owner extracts and encrypts a keyword kw before outsourcing CT to the cloud sever. he first picks $r_1, r_2 \in \mathbb{Z}_p$ and computes

$$W_0 = g^{cr_1}, W_1 = g^{a(r_1+r_2)} g^{br_1 H_4(kw)}, W_2 = g^{r_2}$$

For each $att_j \in W_e$, he sets $W_j = (K_0 K_j)^{r_2}$. Finally, data owner uploads a index I of keyword kw and the signcryption CT of message m to CS, where

$$I = \left\{ W_0, W_1, W_2, \{W_j\}_{att_j \in W_e} \right\}$$

GenToken: Data user inputs an interested keyword kw' and runs algorithm **GenToken**$(sk_{(D,\varphi)}, kw')$ to generate a token T. He first selects a random element $\beta \in \mathbb{Z}_p$ and computes

$$tk_1 = \left(g^a g^{bH_4(kw')} \right)^{\beta}, tk_2 = g^{c\beta}$$

For each $i \in [l_e]$, calculate

$$D'_i = \left(D_{e,i}\right)^{\beta}, K'_i = \left(D'_{e,i}\right)^{\beta}$$

A token of keyword kw' is set as:

$$T = \left\{ (D, \varphi), tk_1, tk_2, \left\{ \left(D'_i, K'_i\right) | i \in [l_e] \right\} \right\}$$

Finally, he sends the token T of keyword kw' to CS.

Search: After verifying that the attribute set W_e satisfies the decryption predicate (D, φ), CS executes algorithm **Search**(I, T) to return the relevant ciphertext CT'. If the attribute set W_e satisfies the decryption predicate $\mathbb{D}$, CS can find a coefficient set $\{w'_i | i \in I_e\}$ such that $\sum_{i \in I_e} w'_i \vec{D}_i = (1, 0, \cdots, 0)$, where $I_e = \left\{ i \in [l_e] \left| att_{\varphi(i)} \in W_e \right. \right\}$. And thus, $\sum_{i \in I_e} w'_i \lambda_{\varphi(i)} = ac$. Then, CS computes

$$E = \prod_{i \in I_e} \left(\frac{e\left(D'_i, W_2\right)}{e\left(K'_i, W_j\right)} \right)^{w'_i} = e(g,g)^{ac\beta r_2}$$

and checks whether the following Equation (1) holds or not.

$$e\left(W_0, tk_1\right) \cdot E = e\left(W_1, tk_2\right) \qquad (1)$$

If the equation holds, CS returns the relevant ciphertext CT' to data user. Otherwise, output $\perp$.

Unsigncrypt: After obtaining the search results CT', data user first computes

$$\begin{aligned} \mu &= H_3\left(C_1\right), \\ (k_1, \cdots, k_l) &= H_2\left(\sigma_2 || \sigma_4 || W_s || W_e\right), \\ \beta &= H_4\left(\sigma_1 || C_2 || C_3 || C_4 || W_s || W_e\right) \end{aligned}$$

and checks the validity of CT' according to the following Equation (2).

$$\begin{aligned} e\left(\sigma_3, g\right) &= e(g,g)^{ac} \cdot e\left(T_0 \prod_{att'_j \in W_s} T_j, \sigma_2 \right) \qquad (2) \\ &\quad \cdot e\left(y_0 \prod_{i \in [l]} y_i^{k_i}, C_1 \right) \cdot e\left(\left(\delta_1^{\mu} \delta_2\right)^{\beta}, \sigma_1 \right) \end{aligned}$$

If Equation (2) does not hold, output $\perp$. Otherwise, data user decrypt CT' as follows.

1) Reconstruct a set of coefficient $\{v_i : i \in I_e\}$ such that $\sum_{i \in I_e} \lambda_{\varphi(i)} v_i = ac$.

2) Set

$$\begin{aligned} E_1 &= \prod_{i \in I_e} \left(D_{e,i} \prod_{att_j \in W_e} D''_{e,i,j} \right)^{v_i} \\ E_2 &= \prod_{i \in I_e} \left(D'_{e,i}\right)^{v_i} \end{aligned}$$

3) Compute

$$e(g,g)^{ac\theta} = \frac{e\left(C_1, E_1\right)}{e\left(C_2, E_2\right)} \qquad (3)$$

4) Recover $m = C_3 \oplus H_1\left(e(g,g)^{ac\theta}, \sigma_1, \sigma_4 \right)$.

3.2 Correctness

In this section, we illustrate the correctness of the above equations.

3.2.1 Correctness of Equation(1)

W_e satisfies (D,φ), so there is $\sum_{i\in I_e} w_i'\lambda_{\varphi(i)} = ac$. Then,

$$E = \prod_{i\in I_e}\left(\frac{e\left(D_i', W_2\right)}{e\left(K_i', W_j\right)}\right)^{w_i'} = e(g,g)^{ac\beta r_2}$$

$$e\left(W_0, tk_1\right) = e(g,g)^{ac\beta r_1}\cdot e(g,g)^{bc\beta r_1 H_4\left(kw'\right)}$$
$$e\left(W_1, tk_2\right) = e(g,g)^{ac\beta(r_1+r_2)}\cdot e(g,g)^{bc\beta r_1 H_4(kw)}$$

Therefore, we have $e\left(W_0, tk_1\right)\cdot E = e\left(W_1, tk_2\right)$.

3.2.2 Correctness of Equation(2)

When W_s satisfies (S,ρ), there exists $\sum_{i\in I_s}\lambda_{\rho(i)}w_i = ac$. Then,

$$\begin{aligned}
&\prod_{i\in I_s}\left(D_{s,i}\prod_{att_j'\in W_s, j\neq\rho(i)} D''_{s,i,j}\right)^{w_i}\\
=\ &\prod_{i\in I_s}\left(g^{\lambda_{\rho(i)}}\left(T_0T_{\rho(i)}\right)^{r_i}\prod_{att_j'\in W_s, j\neq\rho(i)} T_j^{r_j}\right)^{w_i}\\
=\ &g^{ac}\left(T_0\prod_{att_j'\in W_s} T_j\right)^{\sum_{i\in I_s} r_iw_i}
\end{aligned}$$

So,

$$\begin{aligned}
e\left(\sigma_3, g\right) = e(g,g)^{ac}\cdot e&\left(T_0\prod_{att_j'\in W_s} T_j, \sigma_2\right)\\
\cdot e&\left(y_0\prod_{j\in[l]} y_i^{k_i}, C_1\right)\cdot e\left(\left(\delta_1^{\mu}\delta_2\right)^{\beta}, \sigma_1\right)
\end{aligned}$$

3.2.3 Correctness of Equation(3)

Since W_e satisfies (D,φ), we have $\sum_{i\in I_e}\lambda_{\varphi(i)}v_i = ac$. Next,

$$\begin{aligned}
E_1 &= \prod_{i\in I_e}\left(g^{\lambda_{\varphi(i)}}\left(K_0K_{\varphi(i)}\right)^{\tau_i}\cdot\prod_{att_j\in W_e, j\neq\varphi(i)} K_j^{\tau_j}\right)^{v_i}\\
&= g^{ac}\left(K_0\prod_{att_j\in W_e} K_j^{\tau_j}\right)^{\sum_{i\in I_e}\tau_iv_i}\\
E_2 &= \prod_{i\in I_e}\left(D'_{e,i}\right)^{v_i} = \prod_{i\in I_e}\left(g^{\tau_i}\right)^{v_i} = g^{\sum_{i\in I_e}\tau_iv_i}
\end{aligned}$$

Hence,

$$\begin{aligned}
\frac{e\left(C_1, E_1\right)}{e\left(C_2, E_2\right)} &= \frac{e\left(g^{\theta}, g^{ac}\left(K_0\prod_{att_j\in W_e} K_j^{\tau_j}\right)^{\sum_{i\in I_e}\tau_iv_i}\right)}{e\left(\left(K_0\prod_{att_j\in W_e} K_j\right)^{\theta}, g^{\sum_{i\in I_e}\tau_iv_i}\right)}\\
&= e(g,g)^{ac\theta}
\end{aligned}$$

Therefore, we prove that our scheme is correct.

4 Security Proof

Based on Rao *et al.*'s scheme [16] and Zheng *et al.*'s scheme [28], the security of the proposed scheme can be guaranteed through the following four theorems. Due to space constraints, these proofs will be shown in **Appendix A-D**.

Theorem 1. *Our scheme can achieve the indistinguishability of data ciphertext under selective encryption attribute set and adaptive chosen-ciphertext attacks based on the hardness assumption of n-DBDHE problem without using any random oracle.*

Theorem 2. *Our scheme is unforgeable under selective signature attribute set and adaptive chosen-message attacks based on the hardness assumption of the n-CDHE problem without using the random oracle.*

Theorem 3. *Our scheme can achieve the indistinguishability of keyword ciphertext or index under selective encryption attribute set and chosen-keyword attacks based on the hardness assumption of DL problem in the standard model.*

Theorem 4. *Given the given one-way hash function H_4, our scheme can guarantee the secrecy of keyword against adaptive chosen-token attacks.*

4.1 Efficiency Analysis

Table 1 shows the computation cost and size in different phases. For convenience, we use the following notations.

- e: Time cost of an exponentiation;
- p: Time cost of a bilinear pairing;
- $|G_1|(|G_2|)$: Size of a group element in group $\mathbb{G}_1(\mathbb{G}_2)$;
- $u_s(u_e)$: Number of signature(decryption) attributes in $\mathcal{U}_s(\mathcal{U}_e)$;
- $l_s(l_e)$: Number of signature(decryption) attributes in $(S,\rho)((D,\varphi))$;
- ϕ_s: Number of signature attributes required in the signcryption;

- ϕ_e: Number of decryption attributes required in the unsigncryption;

- l: Minimum value that the hash functions are collision resistant.

Table 1: Efficiency analysis of the proposed scheme

Algorithm	Computation Cost	Size
Setup	-	$(u_s + u_e + l)\,\lvert G_1\rvert + \lvert G_2\rvert$
Signing key	$u_s l_s \cdot e$	$u_s l_s$
Decryption key	$u_e l_e \cdot e$	$u_e l_e$
Signcryption	$(\phi_s + 10) \cdot e$	6
Index	$(\phi_e + 4) \cdot e$	$\phi_e + 3$
Token	$(2l_e + 2) \cdot e$	$2l_e + 2$
Search	$(2\phi_e + 2) \cdot p + \phi_e \cdot e$	-
Unsigncryption	$6p + 2\phi_e \cdot e$	-

In the system **Setup** phase, CA takes charge of generating public parameters, whose size is $u_s + u_e + l$ group elements in group $\mathbb{G}_1$ and one group element in group $\mathbb{G}_2$. When a data owner wants to upload join the system, he needs to require the CA to generate the signing key, which needs $u_s l_s$ exponentiation operations to output $u_s l_s$ group elements in group $\mathbb{G}_1$. Similar to the singing key, decryption key contains $u_e l_e$ group elements in group $\mathbb{G}_1$ and takes $u_e l_e$ exponentiation operations. Before upload some data, the data owner needs to encrypt the data and its key. At the **Signcrypt** phase, data ciphertext only contains 6 group elements in group $\mathbb{G}_1$, which needs $(\phi_s + 10)$ exponentiation operations. In addition, in the **GenIndex** phase, the size of the keyword ciphertext or index is $\phi_e + 3$ group elements in group $\mathbb{G}_1$, whose generation needs $(\phi_e + 4)$ exponentiation operations. With the decryption key, a data user can run the **GenToken** algorithm to generate the token, which needs $(2l_e + 2)$ exponentiation operations to generate $2l_e + 2$ group elements in group $\mathbb{G}_1$. To search over an index, the main computation cost is $2\phi_e + 2$ bilinear pairing operations and ϕ_e exponentiation operations. In the **Unsigncryption** phase, there exits 6 bilinear pairing operations and $2\phi_e$ exponentiation operations.

5 Conclusions

In this paper, we have proposed a new concept of searchable attribute-based authenticated encryption and constructed a concrete scheme. The proposed scheme is more appealing on account of its supporting for expressive fine-grained access control, data retrieval and authentication. In our scheme, data owners are allowed to share their data with flexible access policy and authorize legitimate data users to issue search queries according to their token. To avoid receiving illegal data, the users can check the validity of the ciphertext. Furthermore, security proof shows that the proposed scheme is selectively secure in the standard model and has keyword secrecy.

Acknowledgment

This work is supported by the National Key R&D Program of China under Grant No. 2017YFB0802000, the National Natural Science Foundation of China under Grants No.61472470 and 61807026, and the Scientific Research Plan Project of Education Department of Shaanxi Province under Grant No.17JK0362.

References

[1] J. Bethencourt, A. Sahai, and B. Waters, "Ciphertext-policy attribute-based encryption," in *IEEE Symposium on Security and Privacy (SP'07)*, pp. 321–334, 2007.

[2] D. Boneh, G. D. Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search," in *Eurocrypt*, vol. 3027, pp. 506–522, 2004.

[3] Z. Cao, L. Liu, Z. Guo, "Ruminations on attribute-based encryption," *International Journal of Electronics and Information Engineering*, vol. 8, no. 1, pp. 9–19, 2018.

[4] P. S. Chung, C. W. Liu, and M. S. Hwang, "A study of attribute-based proxy re-encryption scheme in cloud environments", *International Journal of Network Security*, vol. 16, no. 1, pp. 1-13, 2014.

[5] Q. Dong, Z. Guan, and Z. Chen, "Attribute-based keyword search efficiency enhancement via an online/offline approach," in *IEEE 21st International Conference on Parallel and Distributed Systems (ICPADS'15)*, pp. 298–305, 2015.

[6] M. Gagné, S. Narayan, and R. Safavi-Naini, "Threshold attribute-based signcryption," in *SCN*, vol. 6280, pp. 154–171, 2010.

[7] V. Goyal, O. Pandey, A. Sahai, and B. Waters, "Attribute-based encryption for fine-grained access control of encrypted data," in *Proceedings of the 13th ACM Conference on Computer and Communications Security*, pp. 89–98, 2006.

[8] C. Gu and Y. Zhu, "New efficient searchable encryption schemes from bilinear pairings," *International Journal Network Security*, vol. 10, no. 1, pp. 25–31, 2010.

[9] S. T. Hsu, C. C. Yang, and M. S. Hwang, "A study of public key encryption with keyword search," *International Journal Network Security*, vol. 15, no. 2, pp. 71–79, 2013.

[10] F. G. Jeng, S. Y. Lin, B. J. Wang, C. H. Wang, and T. H. Chen, "On the security of privacy-preserving keyword searching for cloud storage services," *International Journal Network Security*, vol. 18, no. 3, pp. 597–600, 2016.

[11] J. Lai, R. H. Deng, C. Guan, and J. Weng, "Attribute-based encryption with verifiable outsourced decryption," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 8, pp. 1343–1354, 2013.

[12] C. C. Lee, P. S. Chung, and M. S. Hwang, "A survey on attribute-based encryption schemes of access control in cloud environments," *International Journal Network Security*, vol. 15, no. 4, pp. 231–240, 2013.

[13] J. Li and L. Zhang, "Attribute-based keyword search and data access control in cloud," in *Tenth International Conference on Computational Intelligence and Security (CIS'14)*, pp. 382–386, 2014.

[14] L. Liu, Z. Cao, C. Mao, "A note on one outsourcing scheme for big data access control in cloud," *International Journal of Electronics and Information Engineering*, vol. 9, no. 1, pp. 29–35, 2018.

[15] X. Meng and X. Meng, "A novel attribute-based signcryption scheme in cloud computing environments," in *IEEE International Conference on Information and Automation (ICIA'16)*, pp. 1976–1979, 2016.

[16] Y. S. Rao and R. Dutta, "Efficient attribute-based signature and signcryption realizing expressive access structures," *International Journal of Information Security*, vol. 15, no. 1, pp. 81–109, 2016.

[17] A. Sahai, B. Waters, *et al.*, "Fuzzy identity-based encryption," in *Eurocrypt*, vol. 3494, pp. 457–473, 2005.

[18] S. Selvi, S. Vivek, D. Vinayagamurthy, *et al.*, "ID-based signcryption scheme in standard model," in *International Conference on Provable Security (ProvSec'12)*, vol. 7496, pp. 35–52, 2012.

[19] D. X. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in *Proceedings of IEEE Symposium on Security and Privacy*, pp. 44–55, 2000.

[20] W. Sun, S. Yu, W. Lou, Y. T. Hou, and H. Li, "Protecting your right: Verifiable attribute-based keyword search with fine-grained owner-enforced search authorization in the cloud," *IEEE Transactions on Parallel and Distributed Systems*, vol. 27, no. 4, pp. 1187–1198, 2016.

[21] B. Wang, S. Yu, W. Lou, and Y. T. Hou, "Privacy-preserving multi-keyword fuzzy search over encrypted data in the cloud," in *Proceedings IEEE*, pp. 2112–2120, 2014.

[22] C. Wang, W. Li, Y. Li, and X. L. Xu, "A ciphertext-policy attribute-based encryption scheme supporting keyword search function," in *CSS*, pp. 377–386, 2013.

[23] Y. Wang, "Lattice ciphertext policy attribute-based encryption in the standard model," *International Journal Network Security*, vol. 16, no. 6, pp. 444–451, 2014.

[24] B. Waters, "Ciphertext-policy attribute-based encryption: An expressive, efficient, and provably secure realization," in *Public Key Cryptography*, vol. 6571, pp. 53–70, 2011.

[25] J. Wei, X. Hu, and W. Liu, "Traceable attribute-based signcryption," *Security and Communication Networks*, vol. 7, no. 12, pp. 2302–2317, 2014.

[26] H. Xiong, J. Geng, Z. Qin, and G. Zhu, "Cryptanalysis of attribute-based ring signcryption scheme," *International Journal Network Security*, vol. 17, no. 2, pp. 224–228, 2015.

[27] J. Ye, J. Wang, J. Zhao, J. Shen, and K. C. Li, "Fine-grained searchable encryption in multi-user setting," *Soft Computing*, pp. 1–12, 2016.

[28] Q. Zheng, S. Xu, and G. Ateniese, "Vabks: Verifiable attribute-based keyword search over outsourced encrypted data," in *Proceedings IEEE (INFOCOM'14)*, pp. 522–530, 2014.

[29] Y. Zheng, "Digital signcryption or how to achieve cost (signature & encryption) ¡¡ cost (signature)+ cost (encryption)," *Advances in Cryptology (CRYPTO'97)*, pp. 165–179, 1997.

Appendix A

Proof of Theorem 1

Proof. In this phase, the encryption attribute space $\mathcal{U}_e$ is considered to have n attributes and the hash functions $\{H_i\}_{i=1}^{4}$ are collision resistant. Assume that the simulator $\mathcal{C}$ has a n-DBDHE instance $(\vec{y}_{a,\theta}, Z)$, where

$$\vec{y}_{a,\theta} = \left(g, g^{\theta}, \{g_i\}_{i=1,\cdots n,n+2,\cdots,2n}\right), g_i = g^{a^i}, a, \theta \in \mathbb{Z}_p.$$

Then, $\mathcal{C}$ attempts to distinguish Z is $e\left(g_{n+1}, g^{\theta}\right)$ or a random element of $\mathbb{G}_2$ through the following game. In addition, $\mathcal{C}$ plays the role of challenger in the game of the security model and interact with adversary $\mathcal{A}$.

Init: $\mathcal{C}$ gives the space of signature attributes $\mathcal{U}_s = \{att'_j\}$, the space of encryption attributes $\mathcal{U}_e = \{att_1, \cdots, att_n\}$ and the message space $\mathcal{M} = \{0,1\}^{l_m}$. Then, the adversary $\mathcal{A}$ chooses a challenged encryption attribute set $W_e^* \subset \mathcal{U}_e$ and send it to $\mathcal{C}$.

Setup: $\mathcal{C}$ generates public parameters for $\mathcal{A}$ as follows:

1) Select $\alpha' \in \mathbb{Z}_p$ and set $e(g,g)^{ac} = e(g,g)^{\alpha'} \cdot e\left(g_1, g_n\right)$ by implicitly setting $ac = \alpha' + a^{n+1}$.
2) For each $att_j \in \mathcal{U}_e$, select $\gamma_j \in \mathbb{Z}_p$ and set $K_j = g^{\gamma_j} g_{n+1-j}$. In addition, $K_0 = g^{\gamma_0} \prod_{att_j \in W_e^*} K_j^{-1}$, where $\gamma_0 \in \mathbb{Z}_\mathrm{p}$.
3) Choose $t_0 \in \mathbb{Z}_p$, sets $T_0 = g_1 g^{t_0}$, pick $t_j \in \mathbb{Z}_p$, and compute $T_j = g^{t_j}$ for each $att'_j \in \mathcal{U}_s$.
4) Let $C_1^* = g^{\theta}, \mu^* = H_3\left(C_1^*\right)$, and compute $\delta_1 = g_n^{1/\mu^*}, \delta_2 = g_n^{-1} g^d$, where $d \in \mathbb{Z}_p$.
5) Pick $x_0, \cdots, x_l \in \mathbb{Z}_p$ and compute $u_0 = g^{x_0}, \cdots, u_l = g^{x_l}$.

Phase 1: $\mathcal{A}$ queries the following oracles for polynomially times.

- $\mathcal{O}_{sE}(S, \rho)$: Given a signature LSSS access structure (S, ρ), $\mathcal{C}$ generates a signature key $sk_{(S,\rho)}$ for $\mathcal{A}$ as follows:
  1) Choose α' at random and set $ac = \alpha' + a^{n+1}$.

2) Let $S = (S_{i,j})_{l_s \times k_s}$, where $\vec{S_i} = (S_{i,1}, \cdots, S_{i,k_s})$ is the i-th row of S.
3) Pick $v_2, \cdots, v_{k_s} \in \mathbb{Z}_p$ and generate a vector $\vec{v}_s = \left(\alpha' + a^{n+1}, v_2, \cdots, v_{k_s}\right)$, which implies the secret value is $\alpha' + a^{n+1}$.
4) Let $\vec{v}_s = \vec{w}_s + \left(a^{n+1}, 0, \cdots, 0\right)$, where $\vec{w}_s = (\alpha', v_2, \cdots, v_{k_s})$. So $\lambda_{\rho(i)} = \vec{S}_i \vec{v}_s = \vec{S}_i \vec{w}_s + a^{n+1} S_{i,1}$.
5) Select $r'_i \in \mathbb{Z}_p$ and compute

$$\begin{array}{rcl} D_{s,i} & = & g^{\vec{S}_i \vec{w}_s} \left(T_0 T_{\rho(i)}\right)^{r'_i} g_n^{-\left(t_0 + t_{\rho(i)}\right) S_{i,1}} \\ D'_{s,i} & = & g^{r'_i} g_n^{-S_{i,1}} \\ D''_{s,i} & = & \left\{ D''_{s,i,j} = T_j^{r'_i} g_n^{-t_j S_{i,1}}, \forall att'_j \in \mathcal{U}_s \backslash att'_{\rho(i)} \right\} \end{array}$$

where r_i is implicitly set $r_i = r'_i - a^n S_{i,1}$.

Hence, the signature key is set as

$$sk_{(S,\rho)} = \left\{ (S, \rho), \left\{ D_{s,i}, D'_{s,i}, D''_{s,i} \right\}_{i \in [l_s]} \right\}.$$

- $\mathcal{O}_{dE}(D, \varphi)$: Given a decryption LSSS access structure (D, φ) such that $W_e^* \notin (D, \varphi)$, where $D = (D_{i,j})_{l_e \times k_e}$. $\mathcal{C}$ computes a decryption key $sk_{(D,\varphi)}$ as follows:
Due to $W_e^* \notin (D, \rho)$, there is a vector $\vec{w} = (-1, w_2, \cdots, w_{k_e}) \in \mathbb{Z}_p^{k_e}$ so that $\vec{D}_i \vec{w} = 0$ for $\forall i \in [l_e]$ where $\varphi(i) \in W_e^*$.
  1) Select $v'_2, \cdots, v'_{k_e} \in \mathbb{Z}_p$ at random and set $\vec{v}_e = -\left(\alpha' + a^{n+1}\right) \vec{w} + \vec{v}'$, in which $\vec{v}' = \left(0, v'_2, \cdots v'_{k_e}\right)$.
  2) If $att_{\varphi(i)} \in W_e^*$, we have $\vec{D}_i \vec{w} = 0$. So $\lambda_{\varphi(i)} = \vec{D}_i \vec{v}_e = \vec{D}_i \vec{v}'$.
  Select $\tau_i \in \mathbb{Z}_p$ and compute

$$\begin{array}{rcl} D_{e,i} & = & g^{\vec{D}_i \vec{v}'} \left(K_0 K_{\varphi(i)}\right)^{\tau_i} \\ D'_{e,i} & = & g^{\tau_i} \\ D''_{e,i} & = & \left\{ D''_{e,i,j} = K_j^{\tau_i}, \forall att_j \in \mathcal{U}_e \backslash att_{\varphi(i)} \right\} \end{array}$$

  3) Otherwise, we have

$$\lambda_{\varphi(i)} = \vec{D}_i \vec{v}_e = \vec{D}_i \left(\vec{v}' - \alpha' \vec{w}\right) - \left(\vec{D}_i \vec{w}\right) a^{n+1}.$$

  In this case, $g^{\lambda_{\varphi(i)}}$ contains g_{n+1} which is unknown to $\mathcal{C}$. $\mathcal{C}$ chooses $\tau'_i \in \mathbb{Z}_p$ and τ_i is implicity set as $\tau_i = \tau'_i + \left(\vec{D}_i \vec{w}\right) a^{\varphi(i)}$.
  Next, set

$$\begin{array}{rcl} D_{e,i} & = & g^{\vec{D}_i \left(\vec{v}' - \alpha' \vec{w}\right)} \cdot \left(K_0 K_{\varphi(i)}\right)^{\tau'_i} \cdot g_{\varphi(i)}^{\left(\gamma_0 + \gamma_{\varphi(i)}\right) \vec{D}_i \vec{w}} \\ & & \cdot \prod_{att_j \in W_e^*} \left(g_{\varphi(i)}^{\gamma_j} g_{n+1-j+\varphi(i)}\right)^{-\vec{D}_i \vec{w}} \\ D'_{e,i} & = & g^{\tau'_i} g_{\varphi(i)}^{\vec{D}_i \vec{w}} \\ D''_{e,i} & = & \left\{ D''_{e,i,j} = K_j^{\tau'_i} \left(g_{\varphi(i)}^{\gamma_j} g_{n+1-j+\varphi(i)}\right)^{\vec{D}_i \vec{w}}, \right. \\ & & \left. \forall att_j \in \mathcal{U}_e \backslash att_{\varphi(i)} \right\} \end{array}$$

The decryption key is set as

$$sk_{(D,\varphi)} = \left\{ (D, \varphi), \left\{ D_{e,i}, D'_{e,i}, D''_{e,i} \right\}_{i \in [l_e]} \right\}.$$

- $\mathcal{O}_{SC}(m, W_e, W_s)$: On input a message $m \in \mathcal{M}$, an encryption attribute set $W_e \in \mathcal{U}_e$ and a signature attribute set $W_s \in \mathcal{U}_s$, $\mathcal{C}$ chooses a signature access structure (S, ρ) such that $W_s \in (S, \rho)$, runs $\mathcal{O}_{sE}(S, \rho)$ to generate a signature key $sk_{(S,\rho)}$ and runs **Signcrypt**$(pk, m, sk_{(S,\rho)}, W_e, W_s)$ to return $\mathcal{A}$ the ciphertext CT.
- $\mathcal{O}_{US}(CT, (D, \varphi))$: Suppose that $CT = \{W_e, W_s, C_1, C_2, C_3, C_4, \sigma_1, \sigma_2, \sigma_3, \sigma_4\}$. $\mathcal{C}$ first checks whether $C_1 = C_1^*$ or not. If yes, $\mathcal{C}$ outputs $\perp$. Otherwise, $\mathcal{C}$ continues the following process.
  1) If $W_e^* \notin (D, \varphi)$, obtain the decryption key $sk_{(D,\varphi)}$ according to the oracle $\mathcal{O}_{dE}(D, \varphi)$, then run the algorithm **Unsgincrypt**$(CT, sk_{(D,\varphi)})$ to return the message m;
  2) Otherwise, compute $\mu = H_3(C_1)$ and set

$$e(g,g)^{ac\theta} = e\left(C_4 / C_1^d, g_1\right)^{\left(\frac{\mu}{\mu^*} - 1\right)^{-1}} e\left(C_1, g^{\alpha'}\right)$$

  Finally, output the message

$$m = C_3 \oplus H_1\left(e(g,g)^{ac\theta}, \sigma_1, \sigma_4\right)$$

  Note that $C_1 = g^\theta$ is random for $\mathcal{A}$, so the probability of $C_1 = C_1^*$ is at most $1/p$.

Challenge: $\mathcal{A}$ sends two messages $m_0^*, m_1^* \in \mathcal{M}$ and a signature attribute set W_s^* to $\mathcal{C}$. $\mathcal{C}$ picks a random bit $b^* \in \{0, 1\}$ and signcrypt the message m_b^* with the input W_e^* and W_s^* as follows:

1) Set $C_1^* = g^\theta$ and $\mu^* = H_3(C_1^*)$.
2) Select $v \in \mathbb{Z}_p$, and compute $C_2^* = \left(g^\theta\right)^{\gamma_0}$ and $\sigma_1^* = \left(g^\theta\right)^v$.
3) Choose $\tau^* \in \{0,1\}^{l_\tau}$, and compute $C_3^* = H_1\left(Z \cdot e\left(g^\theta, g^{\alpha'}\right), \sigma_1^*, \tau^*\right) \oplus m_b^*$.
4) Select $\xi \in \mathbb{Z}_p$, and set $\sigma_2^* = g^\xi g_n^{-1}$, which implies $\xi' = \xi - a^n$.
5) Set $C_4^* = \left(g^\theta\right)^d$.
6) Let

$$\begin{array}{rcl} (k_1^*, \cdots, k_l^*) & = & H_2\left(\sigma_2^* || \tau^* || W_s^* || W_e^*\right), \\ \beta^* & = & H_4\left(\sigma_1^* || C_2^* || C_3^* || C_4^* || W_s^* || W_e^*\right) \end{array}$$

and

$$\begin{array}{rcl} {\sigma_3}^* & = & g^{\alpha'} \left(T_0 \prod_{att'_j \in W_s^*} T_j\right)^\xi \cdot \left(g_n^{-t_0} \prod_{att'_j \in W_s^*} g_n^{-t_j}\right) \\ & & \cdot \left(g^\theta\right)^{d\beta^* v + x_0 + \sum_{j \in [l]} k_j^* x_j} \end{array}$$

where $k_i^* \in \{0, 1\}$, for all $i \in [l]$.

7) Set $\sigma_4^* = \tau^*$.

Phase 2: $\mathcal{A}$ continues to query the oracles as in Phase 1. The restriction is that $\mathcal{A}$ cannot query the $\mathcal{O}_{US}(CT, \mathbb{A}_d)$ for any $\mathbb{A}_d$ with $W_e^* \in \mathbb{A}_d$.

Guess: $\mathcal{A}$ returns a guess $b' \in \{0,1\}$. If $b' = b$, then $\mathcal{C}$ can guess that $Z = e\left(g_{n+1}, g^{\theta}\right)$ in the n-DBDHE instance.

We notice that $\mathcal{C}$ can distinguish $Z = e\left(g_{n+1}, g^{\theta}\right)$ or a random element in $\mathbb{G}_2$ if and only if $\mathcal{C}$ doesnot abort the game and $\mathcal{A}$ can output b' such that $b' = b$. Here, $\mathcal{C}$ aborts the game when $C_1 = C_1^*$. After querying q the $\mathcal{O}_{US}(\cdot)$, the possibility of $\mathcal{C}$ aborts is at most q/p. Hence, the possibility of $\mathcal{C}$ in solving the n-DBDHE problem is

$$\Pr\left[e\left(g_{n+1}, g^{\theta}\right) \leftarrow C\left(\vec{y}_{a,\theta}, Z\right)\right] > 1/2 + \varepsilon - q/p.$$

$\square$

Appendix B

Proof of Theorem 2

Proof. Given an n-CDHE problem instance $(g, g_1, \cdots, g_n, g_{n+2}, \cdots, g_{2n}) \in \mathbb{G}_1^{2n}$, where $a \in \mathbb{Z}_p$, g is the generator of $\mathbb{G}_1$ and $g_i = g^{a^i}$. The simulator $\mathcal{C}$ attempts to compute g_{n+1} though the following game. In which, $\mathcal{C}$ plays the role of challenger and interacts with adversary $\mathcal{A}$. Here, $\{H_i\}_{i=1}^4$ are four one-way, collision resistant hash functions.

Init: $\mathcal{C}$ specifies the encryption attribute space $\mathcal{U}_e = \{att_j\}$ and the signature attribute space $\mathcal{U}_s = \{att'_1, \cdots, att'_n\}$. Then, $\mathcal{A}$ chooses a challenge signature set $W_s^* \subseteq \mathcal{U}_s$ and sends it to $\mathcal{C}$.

Setup: Given the security parameter k, $\mathcal{C}$ generates public parameters as follows:

1) Sample $\alpha' \in \mathbb{Z}_p$, and set $e(g,g)^{ac} = e(g,g)^{\alpha'} \cdot e(g_1, g_n)$ by implicitly setting $ac = \alpha' + a^{n+1}$.

2) Select $t_0 \in \mathbb{Z}_p$, and set $T_0 = g^{t_0} \prod_{att'_j \in W_s^*} T_j^{-1}$. For $\forall att'_j \in \mathcal{U}_s$, pick $t_j \in \mathbb{Z}_p$ and set $T_j = g^{t_j} g_{n+1-j}$.

3) Choose $\gamma_0, \{\gamma_j\}_{att_j \in \mathcal{U}_e}$ and compute

$$K_0 = g_1 g^{\gamma_0}, \{K_j = g^{\gamma_j}\}_{att_j \in \mathcal{U}_e}.$$

4) Pick $d, d' \in \mathbb{Z}_p$ and set $\delta_1 = g^d, \delta_2 = g^{d'}$.

5) Let $\zeta = k$, where $\zeta(l+1) < p$ and l is the output size of the hash function H_2. Select an integer ϖ with the restriction $0 \leqslant \varpi \leqslant l$. Pick $(d_0, \cdots, d_l) \in \mathbb{Z}_\zeta^{l+1}$, $(x_0, \cdots, x_l) \in \mathbb{Z}_p^{l+1}$ and compute $y_0 = g_n^{p-\zeta\varpi+d_0} g^{x_0}$, $\left\{y_j = g_n^{d_j} g^{x_j}\right\}_{j\in[l]}$. For $\vec{k} = (k_1, \cdots, k_l) \in \{0,1\}^l$,let

$$\begin{aligned} F\left(\vec{k}\right) &= p - \zeta\varpi + d_0 + \sum_{j\in[l]} k_j d_j \\ J\left(\vec{k}\right) &= x_0 + \sum_{j\in[l]} k_j x_j \end{aligned}$$

So, $y_0 \prod_{j\in[l]} y_j^{k_j} = g_n^{F(\vec{k})} g^{J(\vec{k})}$. In addition, set

$$\mathcal{K}\left(\vec{k}\right) = \begin{cases} 0, & if\ d_0 + \sum_{j\in[l]} k_j d_j = 0 \mod \zeta \\ 1, & otherwise. \end{cases}$$

Due to $\zeta(l+1) < p$, when $\mathcal{K}\left(\vec{k}\right) = 1$, $F\left(\vec{k}\right) \neq 0$.

Phase: $\mathcal{A}$ queries the following oracles for polynomial times:

- $\mathcal{O}'_{sE}(S, \rho)$: Given a signature LSSS access structure (S, ρ) such that $W_s^* \notin (S, \rho)$, where $S = (S_{i,j})_{l_s \times n_s}$. $\mathcal{C}$ computes a signature key $sk_{(S,\rho)}$ as follow.
  Since $W_s^* \notin (S, \rho)$, there is a vector $\vec{w} = (-1, w_2, \cdots, w_{k_s})$ so that $\vec{S}_i\vec{w} = 0$ for $\forall i \in [l_s]$, where $\rho(i) \in W_s^*$.

  1) Pick $v'_2, \cdots, v'_{k_s} \in \mathbb{Z}_p$ and set $\vec{v}_s = -\left(\alpha' + a^{n+1}\right) \cdot \vec{w} + \vec{v}'$, where $\vec{v}' = \left(0, v'_2, \cdots, v'_{k_s}\right)$.

  2) If $att_{\rho(i)} \in W_s^*$, there exists $\vec{S}_i\vec{w} = 0$. So $\lambda_{\rho(i)} = \vec{S}_i\vec{v}_s = \vec{S}_i\vec{v}'$.

  3) Select $\tau_i \in \mathbb{Z}_p$, and compute

  $$\begin{aligned} D_{s,i} &= g^{\vec{S}_i\vec{v}'}\left(T_0 T_{\rho(i)}\right)^{\tau_i} \\ D'_{s,i} &= g^{\tau_i} \\ D''_{s,i} &= \left\{D''_{s,i,j} = T_j^{\tau_i}, \forall j \in [n] \setminus \{\rho(i)\}\right\} \end{aligned}$$

  4) Otherwise, $\lambda_{\rho(i)} = \vec{S}_i\vec{v}_s = \vec{S}_i\left(\vec{v}' - \alpha'\vec{w}\right) - \left(\vec{S}_i\vec{w}\right)a^{n+1}$. Select $\tau'_i \in \mathbb{Z}_p$ and set

  $$\begin{aligned} D_{s,i} &= g^{\vec{S}_i\left(\vec{v}-\alpha'\vec{w}\right)}\left(T_0 T_{\rho(i)}\right)^{\tau_i} g_{\rho(i)}^{\left(t_0+t_{\rho(i)}\right)\vec{S}_i\vec{w}} \\ &\quad \prod_{att'_j \in W_e^*}\left(g_{\rho(i)}^{t_j} g_{n+1-j+\rho(i)}\right)^{-\vec{S}_i\vec{w}} \\ D'_{s,i} &= g^{\tau'_i} g_{\rho(i)}^{\vec{S}_i\vec{w}} \\ D''_{s,i} &= \left\{D''_{s,i,j} = T_j^{\tau'_i}\left(g_{\rho(i)}^{t_j} g_{n+1-j+\rho(i)}\right)^{\vec{S}_i\vec{w}}, \right. \\ &\quad \left. \forall j \in [n] \setminus \rho(i)\right\} \end{aligned}$$

  where τ_i is implicity set

  $$\tau_i = \tau'_i + \left(\vec{S}_i\vec{w}\right)a^{\rho(i)}.$$

Hence, the signature key is set as

$$sk_{(S,\rho)} = \left\{ (S,\rho), \{D_{s,i}, D'_{s,i}, D''_{s,i}\}_{i\in[l_s]} \right\}.$$

- $\mathcal{O}'_{dE}(D,\varphi)$: Given a decryption LSSS access structure (D,φ), $\mathcal{C}$ generates a decryption key $sk_{(D,\varphi)}$ as follow:
  1) Pick $v_2, \cdots, v_{k_e} \in \mathbb{Z}_p$ and set
  $$\vec{v}_e = \vec{w}_e + \left(a^{n+1}, 0, \cdots, 0\right),$$
  where $\vec{w}_e = (\alpha', v_2, \cdots, v_{k_e})$. So
  $$\lambda_{\varphi(i)} = \vec{D}_i\vec{v}_e = \vec{D}_i\vec{w}_e + a^{n+1}D_{i,1}.$$
  2) Select $r'_i \in \mathbb{Z}_p$ and implicitly set $r_i = r'_i - a^n D_{i,1}$. Compute
  $$\begin{aligned} D_{e,i} &= g^{\vec{D}_i\vec{w}_e}\left(K_0K_{\varphi(i)}\right)^{r'_i} g_n^{-\left(\gamma_0+\gamma_{\varphi(i)}\right)D_{i,1}} \\ D'_{e,i} &= g^{r'_i} g_n^{-D_{i,1}} \\ D''_{e,i} &= \{D''_{e,i,j} = K_j^{r'_i} g_n^{-\gamma_j D_{i,1}}, \\ &\quad \forall att_j \in U_e \backslash att_{\varphi(i)}\} \end{aligned}$$

$\mathcal{C}$ outputs the decryption key

$$sk_{(D,\varphi)} = \left\{ (D,\varphi), \{D_{e,i}, D'_{e,i}, D''_{e,i}\}_{i\in[l_e]} \right\}.$$

- $\mathcal{O}'_{SC}(m, W_e, W_s)$: $\mathcal{C}$ constructs a signature access structure (S,ρ) such that $W_s \in (S,\rho)$. If $W_s^* \notin (S,\rho)$, $\mathcal{C}$ runs $\mathcal{O}'_{sE}(S,\rho)$ to obtain a signing key $sk_{(S,\rho)}$, then outputs a ciphertext CT by running the algorithm **Signcrypt**$(pk, m, sk_{(S,\rho)}, W_e, W_s)$. Otherwise, $\mathcal{C}$ outputs a ciphertext as follow.
  1) Pick $\theta', \xi' \in \mathbb{Z}_p$ and set $\sigma_2 = g^{\xi'}$.
  2) Select $\tau \in \{0,1\}^{l_\tau}$ such that $\mathcal{K}\left(\vec{k}\right) \neq 0$, sets $\sigma_4 = \tau$ and compute $C_1 = g^{\theta'} g_1^{-1/F\left(\vec{k}\right)}$, which implies $\theta = \theta' - a\big/F\left(\vec{k}\right)$.
  3) Compute
  $$\begin{aligned} C_2 &= \left(K_0 \prod_{att_j\in W_e} K_j\right)^{\theta'} \\ &\quad \cdot \left(g_2 g_1^{\gamma_0} \prod_{att_j\in W_e} g_1^{\gamma_j}\right)^{\frac{-1}{F\left(\vec{k}\right)}} \end{aligned}$$
  4) Select $\vartheta \in \mathbb{Z}_p$ and compute $\sigma_1 = g_1^{\vartheta}$.
  5) Compute
  $$\begin{aligned} C_3 &= H_1\left(e(g,g)^{ac\theta'} \cdot e(g,g_1)^{-\alpha'/F\left(\vec{k}\right)}\right. \\ &\quad \left. \cdot\, e(g_2,g_n)^{-1/F\left(\vec{k}\right)}, \sigma_1, \tau\right) \oplus m \end{aligned}$$
  6) Set $\mu = H_3(C_1)$, and compute
  $$C_4 = (\delta_1^{\mu}\delta_2)^{\theta'}\left(g_1^{\mu d + d'}\right)^{-1/F\left(\vec{k}\right)}.$$
  7) Set $\beta = H_4(\sigma_1||C_2||C_3||C_4||W_s||W_e)$.
  8) Compute
  $$\sigma_3 = g^{\alpha'}\left(T_0 \prod_{att'_j\in W_s} T_j\right)^{\xi'} \left(g_n^{F\left(\vec{k}\right)} g^{J\left(\vec{k}\right)}\right)^{\theta'}$$
- $\mathcal{O}'_{US}(CT, (D,\varphi))$: Given the ciphertext CT and the decryption access structure (D,φ), $\mathcal{C}$ first runs the oracle $\mathcal{O}'_{dE}(D,\varphi)$ to get a decryption key $sk_{(D,\varphi)}$ and outputs $\mathcal{A}$ the message m by the algorithm **Unsigncrypt** $(pp, CT, sk_{(D,\varphi)})$.

Forgery: For message m^*, such that (m^*, W_s^*, W_e^*) has never been queried in $\mathcal{O}'_{SC}(m, W_e, W_s)$, $\mathcal{A}$ sends a forgery $CT^* = \{W_s^*, W_e^*, C_1^*, C_2^*, C_3^*, C_4^*, \sigma_1^*, \sigma_2^*, \sigma_3^*, \sigma_4^*\}$ for a message m^* and a decryption access structure $\mathbb{A}_d^*$ to $\mathcal{C}$, where **Unsigncrypt**$(CT^*, sk_{\mathbb{A}_d^*}) = m^*$. Then, $\mathcal{C}$ computes

$$\vec{k}^* = (k_1^*, \cdots, k_l^*) = H_2(\sigma_2^*||\sigma_4^*||W_s^*||W_e^*)$$

and checks whether $F\left(\vec{k}^*\right) = 0$. If not, $\mathcal{C}$ aborts. Otherwise, $\mathcal{C}$ verifies the validity of CT^* though Equation (2).

If $\mathcal{A}$ wins the game, i.e. the CT^* can passe the verification, which means that

$$\begin{aligned} C_1^* &= g^{\theta}, \sigma_1^* = g^{\theta\vartheta}, \sigma_2^* = g^{\xi'}, \mu^* = H_3(C_1^*) \\ \beta^* &= H_4(\sigma_1^*||C_2^*||C_3^*||C_4^*||W_s^*||W_e^*), C_4^* = g^{\left(d\mu^* + d'\right)\theta} \\ \sigma_3^* &= g^{\alpha' + a^{n+1}}\left(T_0 \prod_{att'_j\in W_s^*} T_j\right)^{\xi'} \left(g_n^{F\left(\vec{k}^*\right)} g^{J\left(\vec{k}^*\right)}\right)^{\theta} (C_4^*)^{\beta^*\vartheta} \end{aligned}$$

In order to provide a perfect simulation, the game cannot abort in **Forgery** phase. Following Selvi *et al.* [18], we have

$$\Pr[\neg\, abort] = \frac{1}{\zeta} \cdot \frac{1}{l+1} = \frac{1}{k(l+1)}.$$

Suppose that $\mathcal{A}$ wins the game with an advantage ε, $\mathcal{C}$ can solve the n-CDHE problem with the advantage $\varepsilon' = \varepsilon/(k(l+1))$. □

Appendix C

Proof of Theorem 3

Proof. Given a DL instance $(g, h, f, f^{r_1}, g^{r_2})$, where $g, h, f \in \mathbb{G}_1$ and $r_1, r_2 \in \mathbb{Z}_p$. The simulator $\mathcal{C}$ attempts to compute $h^{r_1+r_2}$ through the following game, in which, $\mathcal{C}$ plays the role of challenger and interacts with adversary $\mathcal{A}$. Here, H_4 is a one-way hash function.

Init: $\mathcal{C}$ gives the encryption attribute space $\mathcal{U}_e = \{att_1, \cdots, att_n\}$. Then, $\mathcal{A}$ chooses a challenge $W_e^* \subseteq \mathcal{U}_e$ and sends it to $\mathcal{C}$.

Setup: Given the security parameter k, $\mathcal{C}$ generates public parameters as follows:

1) Set $g^a = h, g^c = f$, where $a, c \in \mathbb{Z}_p$ are unknown.
2) Select $d \in \mathbb{Z}_p$ and compute $g^b = f^d = g^{cd}$, which implies $b = cd$.
3) For each $att_j \in \mathcal{U}_e \backslash W_e^*$, pick $\alpha_j, \beta_j \in \mathbb{Z}_p$ and set $K_j = K_0^{-1} f^{\alpha_j} g^{\beta_j}$.
4) For each $att_j \in W_e^*$, set $K_j = K_0^{-1} g^{\beta_j}$.

$\mathcal{C}$ outputs the public parameters $pp = \left(e, g, h, f, f^d, K_0, \{K_j\}_{att_j \in \mathcal{U}_e}\right)$ and sets the master key as $msk = d$.

Phase 1: $\mathcal{A}$ queries the following oracles for polynomially times:

- $\mathcal{O}_{dE}(D, \varphi)$: $\mathcal{A}$ sends an encryption access structure (D, φ) to $\mathcal{C}$, where $D = (D_{i,j})_{l_e \times k_e}$. If $W_e^* \in (D, \varphi)$, $\mathcal{C}$ outputs $\perp$. Otherwise, $\mathcal{C}$ performs as follow.

  Due to $W_e^* \notin (D, \varphi)$, there exists a vector $\vec{w} = (-1, w_2, \cdots, w_{k_e}) \in \mathbb{Z}_p^{k_e}$ such that $\vec{D}_i \vec{w} = 0$ for all $i \in [l_e]$ where $\varphi'(i) \in W_e^*$.

  1) Pick $v_2', \cdots v_{k_e}' \in \mathbb{Z}_p^{k_e}$ and set $\vec{v}_e = a\vec{w} + \vec{v}'$, where $\vec{v}' = \left(0, v_2', \cdots, v_{k_e}'\right)$.
  2) If $att_{\varphi(i)} \in W_e^*$, we have $\vec{D}_i \vec{w} = 0$. Hence, $\lambda_{\varphi(i)} = \vec{D}_i \vec{v}'$. Select a random number $\tau_i \in \mathbb{Z}_p$ and compute
  $$\begin{aligned} D_{e,i} &= f^{\vec{D}_i \vec{v}'} \left(g^{\beta_{\varphi(i)}}\right)^{\tau_i} \\ &= g^{c\lambda_{\varphi(i)}} \left(K_0 K_{\varphi(i)}\right)^{\tau_i} \\ D_{e,i}' &= g^{\tau_i} \end{aligned}$$
  3) Otherwise, $\mathcal{C}$ selects $\tau_i' \in \mathbb{Z}_p$ and computes
  $$\begin{aligned} D_{e,i} &= \left(g^{\lambda_{\varphi(i)}}\right)^{\frac{-\beta_{\varphi(i)}}{\alpha_{\varphi(i)}}} \left(f^{\alpha_{\varphi(i)}} g^{\beta_{\varphi(i)}}\right)^{\tau_i'} \\ &= f^{\lambda_{\varphi(i)}} \left(f^{\alpha_{\varphi(i)}} g^{\beta_{\varphi(i)}}\right)^{\tau_i' - \frac{\beta_{\varphi(i)}}{\alpha_{\varphi(i)}}} \\ D_{e,i}' &= g^{\lambda_{\varphi(i)}\left(\frac{-1}{\alpha_{\varphi(i)}}\right)} g^{\tau_i'} \end{aligned}$$
     where, τ is implicity set as $\tau_i = \tau_i' - \frac{\lambda_{\varphi(i)}}{\alpha_{\varphi(i)}}$.

  $\mathcal{C}$ returns
  $$sk = \left\{(D, \varphi), \{D_{e,i}, D_{e,i}'\}_{i \in [l_e]}\right\}.$$
- $\mathcal{O}_{GT}(kw, (D, \varphi))$: $\mathcal{C}$ first runs the oracle $\mathcal{O}_{dE}(D, \varphi)$ to obtain $sk_{(D,\varphi)}$, then returns $\mathcal{A}$ a token T by the algorithm **GenToken**$(sk_{(D,\varphi)}, kw)$. If $W_e^* \in (D, \varphi)$, $\mathcal{C}$ adds kw to L_{kw}, a keyword list which is initially empty.

Challenge: $\mathcal{A}$ sends two keywords kw_0 and kw_1 to $\mathcal{C}$, where kw_0 and kw_1 do not belong to keyword set list L_{kw}. Then, $\mathcal{C}$ randomly selects a bit $b \in \{0, 1\}$ and runs the algorithm **GenIndex**(kw_b, W_e^*) to generate the index $I = \{W_0, W_1, W_2, \{W_j | att_j \in W_e^*\}\}$, where
$$W_0 = f^{r_1}, W_1 = T(f^{r_1})^{dH_4(kw_b)}, W_2 = g^{r_2}$$
For each $att_j \in W_e^*$, set $W_j = (g^{r_2})^{\beta_j}$.

Phase 2: This phase is performed as in Phase 1. The restriction is that if $W_e^* \in (D, \varphi)$, $\mathcal{A}$ cannot query the oracle $\mathcal{O}_{GT}$ with the input $((D, \varphi), kw_0)$ or $((D, \varphi), kw_1)$.

Guess: $\mathcal{A}$ outputs a guess $b' \in \{0, 1\}$. If $b' = b$, $\mathcal{C}$ outputs $T = h^{r_1 + r_2}$. Otherwise, $\mathcal{C}$ outputs $T = R \in \mathbb{G}_2$.

Suppose that $\mathcal{A}$ wins the above game with an advantage ε. In the challenge phase, if $T = h^{r_1 + r_2}$, the index I is valid, then $\mathcal{A}$ outputs $b' = b$ with the probability $\frac{1}{2} + \varepsilon$. Otherwise, T is a random element in $\mathbb{G}_2$, so the index I is not valid, $\mathcal{A}$ outputs $b' = b$ with the probability $\frac{1}{2}$.

Therefore, $\mathcal{C}$ can solve the DL problem with an advantage $\frac{\varepsilon}{2}$.

□

Appendix D

Proof of Theorem 4

Proof. We utilize a challenger $\mathcal{C}$ to conduct the following keyword secrecy game.

Setup: $\mathcal{C}$ first chooses random elements $a, b, c \in \mathbb{Z}_p$, $f \in \mathbb{G}_1$. For each $att_j \in \mathcal{U}_e$, $\mathcal{C}$ selects $\alpha_j \in \mathbb{Z}_p$ and sets $K_j = g^{\alpha_j}$. Then, public parameters are set as $pk = \left(e, g, g^a, g^b, g^c, \{K_j | att_j \in \mathcal{U}_e\}\right)$ and master key is set as $mk = (a, b, c)$.

Phase: The adversary $\mathcal{A}$ queries the following two oracles for polynomial times.

- $\mathcal{O}_{dE}(D, \varphi)$: $\mathcal{C}$ runs the algorithm **dExtract**$(msk, (D, \varphi))$ to gain a secret key $sk_{(D,\varphi)}$, sends it to $\mathcal{A}$, and adds (D, φ) to the list L_{dE}.
- $\mathcal{O}_{GT}(kw, (D, \varphi))$: $\mathcal{C}$ first runs the oracle $\mathcal{O}_{dE}(D, \varphi)$ to obtain secret key $sk_{(D,\varphi)}$, then calls **GenToken**$((sk_{(D,\varphi)}), kw)$ algorithm to generate token T for $\mathcal{A}$.

Challenge: $\mathcal{A}$ first chooses an attribute set W_e^*, then $\mathcal{C}$ selects an encryption access structure (D^*, φ^*) such that $W_e^* \in (D^*, \varphi^*)$ and computes $sk_{(D,\varphi)}^*$ according to the oracle $\mathcal{O}_{dE}\left(msk, (D^*, \varphi^*)\right)$. Next, $\mathcal{C}$ randomly selects a keyword kw^* and computes I^* and T^*, where $\forall\ (D, \varphi) \in L_{dE}$, $W_e^* \notin (D, \varphi)$.

Guess: $\mathcal{A}$ outputs a keyword set kw' to $\mathcal{C}$, then $\mathcal{C}$ computes I' by running the algorithm **GenIndex**(W_e^*, kw'). If **Search**$(T^*, I') = 1$, then $\mathcal{A}$ wins the game.

Suppose that $\mathcal{A}$ has issued q_{kw} different keyword sets before returning kw', and the probability of $\mathcal{A}$ winning the keyword secrecy game is at most $\frac{1}{|\mathcal{M}|-|q_{kw}|}+\varepsilon$, where $|q_{kw}|$ is denoted as the number of the different keywords. The size of remaining keyword set space is $|\mathcal{M}|-|q_{kw}|$, and H_4 is denoted as a one-way hash function which means recovering kw^* from $H_4(kw^*)$ has at most a negligible probability ε. Therefore, given $|q_{kw}|$ distinct keywords $\mathcal{A}$ has queried, $\mathcal{A}$ wins the keyword secrecy game with the probability at most $\frac{1}{|\mathcal{M}|-|q_{kw}|}+\varepsilon$. □

Biography

Zhenhua Liu received his B.S. degree from Henan Normal University, M.S., and Ph.D degrees from Xidian University, China, in 2000, 2003 and 2009, respectively. He is currently a professor with Xidian University. His research interests include cryptography and information security.

Yaqing Fan received her B.S. degree in 2015 from College of Mathematics and Information Sciences, Taiyuan Normal University. Now, she is a master degree student in Mathematics at Xidian University. Her research interests include cryptography and cloud security.

A Robust Authentication Protocol for Multi-server Architecture Using Elliptic Curve Cryptography

Xueqin Zhang, Baoping Wang, Wenpeng Zhang

(Corresponding author: Baoping Wang)

Software School, Nanyang Normal University
No. 1638, Wolong Road, Wolong District, Nanyang 473000, China
(Email: baoping_wang@outlook.com)

(Received July 17, 2017; revised and accepted Nov. 5, 2017)

Abstract

The multi-server architecture authentication scheme enables users access to the multiple distributed servers with only one single registration procedure. It provides a scalable solution for repeated registration issue in multi-server environment. In this paper, we present a secure remote authentication scheme for multiple servers architecture with elliptic curves cryptosystem (ECC). The proposed scheme could resolve many grave flaws and provide message authenticity, while preserving user anonymity. In the security analysis, we prove the completeness of the proposal BAN-logic, which one of the important formal methods for evaluating information exchange protocols. Noticeable, our scheme also shows impressive efficiency and practicability comparing with other related schemes.

Keywords: Anonymity; Authentication; BAN-logic; Elliptic Curve Cryptography; Multi-server

1 Introduction

In the digital information world, users can easily obtain the information services of the distributed networks anywhere and anytime such as online shopping, online bank, and pay-TV. Authentication plays an important part to construct a secure communication channel between participants in the information systems. To ensure the security of the communication between these participants, more robust remote authentication protocols are urgent needed.

In 1981, Lamport [14] proposed a well-known authentication protocol based on password for the insecure communication, since then, ample of remote user authentication protocols have been presented to improve security and efficiency [2,5,7,8,17,32,33]. However, these protocols are designed for single-server architecture. If conventional protocols are applied to the multi-server environment, the network users not only need to log into various remote servers with repetitive registrations, but also need to remember various identities and passwords. In this paper, we propose a comparatively robust remote user authentication protocol suiting for multiple servers environment, which guarantees better efficiency and achieves various of the security properties. specifically, we analyze the validity of the proposed protocol with formal proof BAN-logic, which is widely employed to validate the beliefs of the involved participants in information exchange protocol.

In the first eight years of the 21st century, many researchers have proposed authentication protocols for multi-server architecture, respectively [3, 10, 18, 23, 29, 30]. However, in these protocols, user's identity is transmitted in the form of plaintext through public communication channel. In order to resolve the privacy problems raised by static ID, Liao and Wang [22] proposed a dynamic ID based remote user authentication protocol for multi-server architecture, which could eliminate the risk of ID-theft and protect users' privacy. However, their protocol cannot withstand insider attack and masquerade attack. Besides, their scheme fails to provide mutual authentication. Later on, Hsiang and Shih [6] proposed an improved multi-server password authenticated key agreement protocol. In their scheme, only the registration center possesses master secret x and it uses it to issue the private keys for service provider and legal users. The solutions is seemingly to remedy these vulnerabilities of Liao and Wang's protocol, and the authors claim their protocol could resist masquerade attack, server spoofing, registration center spoofing attack and insider attack. Nevertheless, Sood *et al.* [27] pointed out their protocol was susceptible to replay attack, impersonation attack and stolen smart card attack, more over, the password change phase of their protocol was incorrect. Meanwhile, Sood *et al.* presented a multi-server authentication protocol with two-server paradigm, in which the service provider is exposed to users and the control server (Registration center) is not directly accessible to them between verification process. This strategy protect the control server is less likely to be attack. In 2012, Li *et al.* [21] demonstrated Sood *et al.*'s protocol was vulnerable to leak-of-verifier attack and stolen smart card attack. Furthermore, the authentication and session key agreement of the scheme was wrong. In order to tackle these problems, Li *et al.* proposed a more robust authentication protocol for multi-server environment using smart cards. The authors employed the verification strategy introduced in Sood *et al.*'s proposal and also inherited its critical vulnerabilities. Subsequently, Li *et al.*'s protocol was demonstrated that it failed to tackle the replay attack, the password guessing attack and the masquerade attack [11].

In 2013, Pippal *et al.* [26] introduced multiple servers authentication scheme without verification table. Furthermore, it allows the legal users could access multiple servers with no help of registration center (in other words, users and service servers could complete mutual authentication independently). Nevertheless, its verification method has a fatal problem that too much sensitive parameters are stored in users' smart card. Li *et al.* [20] demonstrated that their scheme was susceptible to off-line password guessing attack, impersonation attack and privileged insider attack. They also present their remediation with a flexible registration, which the number of servers is no longer fixed. In 2017, Srinivas *et al.* [28] showed that Li *et al.*'s protocol was vulnerable to a range of ignored security flaws and proposed a new authentication for multiple servers environment. Recently, a pile of multi-server authentication protocols are published for providing stronger robustness and better efficiency [15,16,19,25,34].

The structure of our paper is organized as follows. In Section 2, we present a robust multiple servers authentication schemes. Then, the security analysis of our protocol and the comparisons between our proposal and related protocols are presented in Sections 3 and 4, respectively. Finally, Section 5 presents the conclusion.

2 Our Scheme

The multiple servers system consists of three involved parties, registration center RC, authorized servers S_j and users U_i. RC is the trusted party and administrates the whole system. S_j has the jurisdiction to offer network services and U_i could access these services.

In this section, we present a new authentication scheme for multi-server architecture, which can be divided into five phases: initialization phase, server registration phase, user registration phase, authenticated key agreement phase and password change phase. The abbreviations and notions used in our protocol are listed in Table 1. The briefly steps are described as follows.

Table 1: Notations

Notations	Meaning
U_i	The ith user
S_j	The jth service providing server
RC	The registration center
ID_i	The identity of the user U_i
PW_i	The password of the user U_i
SID_j	The identity of the server S_j
x	The master secret key of the RC
P	The generator of G
P_{pub}	RC's public key, where $P_{pub} = xP$
SK	The session key shared among U_i, S_j
$H(\cdot)$	A one-way hash function
$Enc_{Key}(M)$	Encryption of messages M using Key
$Dec_{Key}(C)$	Decryption of ciphertext C using Key
$\oplus$	Exclusive-OR operation
$\|$	String concatenation operation

2.1 Initialization Phase

In this phase, RC chooses two large prime numbers p and q with $p = 2q + 1$. Subsequently, RC selects a generator P of order q on the elliptic curve $E_p(a, b)$, which possesses good security properties [9,12,31]. Finally, RC generates x as the master secret key, which is minimum of 1024 bits for security purpose.

2.2 Server Registration Phase

When a server S_j wants to register and become an authorized server, S_j and RC should execute the following interactions.

SR.1: S_j chooses its identity SID_j and transmits it to RC for registration via a secured communication channel.

SR.2: RC computes $s_j = H(SID_j \| x)$ and assigns it to S_j via secure channels.

SR.3: On receiving s_j, S_j stores it secretly and finishes the registration.

2.3 User Registration Phase

U_i and RC should execute the following interactions to finish the registration phase:

UR.1: U_i selects his/her identity ID_i, the password PW_i and random number r, then U_i computes $RPW_i = H(PW_i \| r)$ and sends $\{ID_i, RPW_i\}$ to RC for registration.

UR.2: Upon receiving U_i's registration request, RC calculates $A_i = H(ID_i \| RPW_i)$, $K_i = H(ID_i \| x)$, $B_i = K_i \oplus A_i$, where x is the master secret key of RC and kept by RC privately. Then RC stores $\{B_i, Enc(), P, P_{pub}, H(\cdot)\}$ on U_i's smart card and issues it to U_i via a secure channel.

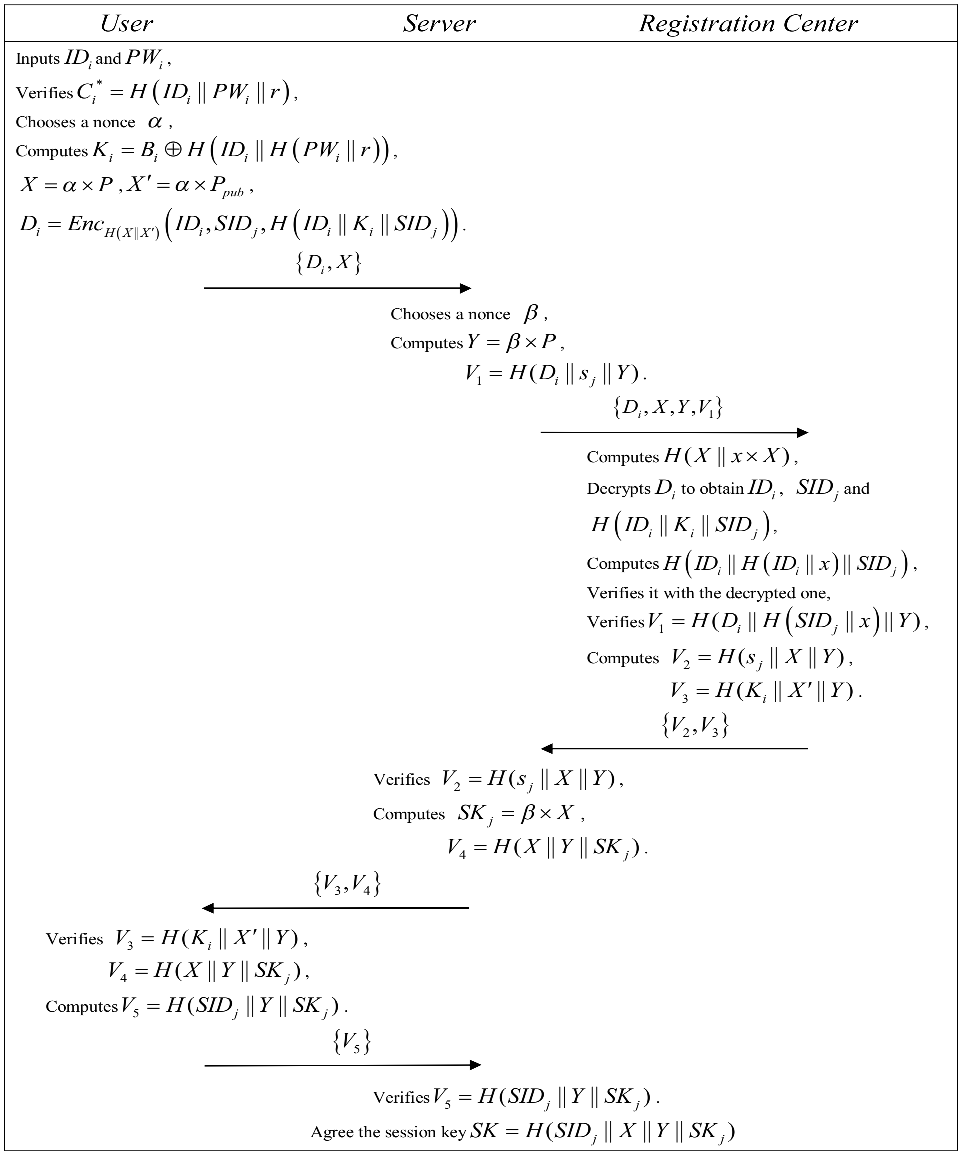

User
Server
Registration Center
Inputs ID_i and PW_i,
Verifies $C_i^* = H(ID_i \| PW_i \| r)$,
Chooses a nonce α,
Computes $K_i = B_i \oplus H(ID_i \| H(PW_i \| r))$,
$X = \alpha \times P$, $X' = \alpha \times P_{pub}$,
$D_i = Enc_{H(X\|X')}(ID_i, SID_j, H(ID_i \| K_i \| SID_j))$.
$\{D_i, X\}$
Chooses a nonce β,
Computes $Y = \beta \times P$,
$V_1 = H(D_i \| s_j \| Y)$.
$\{D_i, X, Y, V_1\}$
Computes $H(X \| x \times X)$,
Decrypts D_i to obtain ID_i, SID_j and
$H(ID_i \| K_i \| SID_j)$,
Computes $H(ID_i \| H(ID_i \| x) \| SID_j)$,
Verifies it with the decrypted one,
Verifies $V_1 = H(D_i \| H(SID_j \| x) \| Y)$,
Computes $V_2 = H(s_j \| X \| Y)$,
$V_3 = H(K_i \| X' \| Y)$.
$\{V_2, V_3\}$
Verifies $V_2 = H(s_j \| X \| Y)$,
Computes $SK_j = \beta \times X$,
$V_4 = H(X \| Y \| SK_j)$.
$\{V_3, V_4\}$
Verifies $V_3 = H(K_i \| X' \| Y)$,
$V_4 = H(X \| Y \| SK_j)$,
Computes $V_5 = H(SID_j \| Y \| SK_j)$.
$\{V_5\}$
Verifies $V_5 = H(SID_j \| Y \| SK_j)$.
Agree the session key $SK = H(SID_j \| X \| Y \| SK_j)$

Figure 1: Authenticated key agreement phase

UR.3: U_i stores the random number r and $C_i = H(ID_i \| PW_i \| r)$ into the issued smart card.

2.4 Authenticated Key Agreement Phase

Whenever U_i wants to access the services of R_j, the following operations will be performed during the authenticated key agreement phase.

A.1: U_i inserts his/her smart card into the card reader and inputs ID_i, PW_i. Then the smart card computes $C_i^* = H(ID_i \| PW_i \| r)$ and checks whether it is equal to the stored value C_i. If so, the smart card proceeds the following steps. Otherwise, the smart card aborts this procedure. Then, the smart card computes $K_i = B_i \oplus H(ID_i \| H(PW_i \| r))$, $X = \alpha \times P$, $X' = \alpha \times P_{pub}$, $D_i = Enc_{H(X\|X')}(ID_i, SID_j, H(ID_i \| K_i \| SID_j))$ with a chosen random nonce α. After that, U_i sends the login request message $M_1 = \{D_i, X\}$ to S_j.

A.2: Upon receiving M_1, S_j also generates a random integer number β and calculates $Y = \beta \times P$, $V_1 = H(D_i \| s_j \| Y)$. Then, S_j transmits $M_2 = \{D_i$, X, Y, $V_1\}$ to RC.

A.3: Upon receiving M_2, RC computes $X' = x \times X$ firstly. Then, RC can get U_i's secret value $\{ID_i, SID_j, H(ID_i \| K_i \| SID_j)\}$ of login request by calculating $Dec_{H(X\|X')}(D_i)$. Subsequently, RC computes $H(ID_i \| H(ID_i \| x) \| SID_j)$ and compares it with the retrieved one in D_i to validate U_i. If the computed one does not exist in the decrypted results from D_i, RC will terminate this session. Else, RC authenticates U_i successfully and will continue to verify the legitimacy of S_j. RC uses the aforementioned decrypted value SID_j from D_i to calculate $V_1^* = H(D_i \| H(SID_j \| x) \| Y)$ and checks whether $V_1^*? = V_1$. If the equation does not hold, RC rejects this request and terminates this session. Else, RC accepts this request and computes $V_2 = H(s_j\|X\|Y)$, $V_3 = H(K_i\|X'\|Y)$. Finally, RC sends the reply message $M_3 = \{V_2, V_3\}$ to S_j.

A.4: On receiving M_3, S_j computes $H(s_j\|X\|Y)$ and checks it with the received V_2. If they are not equal, S_j rejects these messages and terminates this session. Otherwise, S_j successfully authenticates RC, and then computes $SK_j = \beta \times X = \alpha\beta \times P$, $V_4 = H(X\|Y\|SK_j)$. After that, S_j submits $M_4 = \{V_3, V_4, Y\}$ to U_i.

A.5: Upon receiving the response M_4, the smart card checks whether the equation $V_3 = H(K_i\|X'\|Y)$ holds or not. If not, the smart card stops this session. Otherwise, the smart card calculates $SK_j = \alpha \times Y = \alpha\beta \times P$ and checks whether $H(X\|Y\|SK_j)$ is equal to received V_4. If not, the smart card stops this session. Otherwise, the smart card computes $V_5 = H(SID_j\|Y\|SK_j)$. Finally, the smart card sends the response message $M_5 = \{V_5\}$ to S_j.

A.6: S_j computes and checks $V_5? = H(SID_j\|Y\|SK_j)$ after receiving M_5. If this equation holds, S_j successfully authenticates U_i and mutual authentication is completed. Otherwise, the session will be terminated.

After finishing the mutual authentication of U_i, S_j and RC, U_i and S_j shares the common session key $SK = H(SID_j\|X\|Y\|SK_j)$.

2.5 Password Change Phase

Suppose U_i wants to select a new password PW_i^{new} to replace original password. Then the smart card should execute the following procedures.

Step 1: U_i makes a request to the smart card and enters ID_i and old password PW_i to the smart card.

Step 2: U_i's smart card checks $C_i? = H(ID_i\|PW_i\|r)$. If yes, U_i inputs a new password PW_i^{new}. Otherwise, the smart card rejects the password change request and terminates this procedure.

Step 3: The smart card computes $A_i^{new} = H(ID_i\|H(PW_i^{new}\|r))$, $B_i^{new} = B_i \oplus A_i \oplus A_i^{new}$, $C_i^{new} = H(ID_i\|PW_i^{new}\|r)$ and stores B_i^{new}, C_i^{new} into its memory to replace B_i, C_i.

3 Security Analysis and Discussion

In the following we will evaluated our scheme by BAN-logic and demonstrate it could withstand common network attacks.

3.1 Validity Proof Based on BAN-logic

In this section, the validity of our proposed scheme is evaluated by BAN-logic [1]. Specifically, BAN-logic helps each participants to trust the exchanged messages and it is a widely employed method for analyzing authentication protocol. We define ample of notations used in the following proof procedures are defined.

$\mathcal{P} \mid\equiv X$: The principal $\mathcal{P}$ believes X.

$\sharp(X)$: The formula X is fresh.

$\mathcal{P} \Rightarrow X$: The principal $\mathcal{P}$ has jurisdiction over X.

$\mathcal{P} \lhd X$: The principal $\mathcal{P}$ sees X.

$\mathcal{P} \mid\sim X$: The principal $\mathcal{P}$ once said the statement X.

(X, Y): The formula X or Y is the part of (X, Y).

$\langle X \rangle_Y$: The formula X is combined with Y.

$\{X\}_Y$: This represents the formula X is message and it is encrypted under the key Y.

$\mathcal{P} \xleftrightarrow{k} \mathcal{Q}$: The principals $\mathcal{P}$ and $\mathcal{Q}$ communicate with each other with the shared key k. Note that, k will never be known to any other principals.

$\mathcal{P} \overset{k}{\rightleftharpoons} \mathcal{Q}$: $\mathcal{P}$ and $\mathcal{Q}$ shared a secret k, which is possibly known to other principals trusted by them.

SK: the formula SK represents the session key used in the current session.

In the following, we present some logical postulates which used in the demonstration of our protocol:

- The message-meaning rule: $\frac{\mathcal{P}|\equiv\mathcal{Q}\overset{k}{\rightleftharpoons}\mathcal{P},\mathcal{P}\lhd\langle X\rangle_k}{\mathcal{P}|\equiv\mathcal{Q}|\sim X}$.
- The freshness-conjuncatenation rule: $\frac{\mathcal{P}|\equiv\sharp(X)}{\mathcal{P}|\equiv\sharp(X,Y)}$.
- The nonce-verification rule: $\frac{\mathcal{P}|\equiv\sharp(X),\mathcal{P}|\equiv\mathcal{Q}|\sim X}{\mathcal{P}|\equiv\mathcal{Q}|\equiv X}$.
- The jurisdiction rule: $\frac{\mathcal{P}|\equiv\mathcal{Q}\Rightarrow X,\mathcal{P}|\equiv\mathcal{Q}|\equiv X}{\mathcal{P}|\equiv X}$, $\frac{\mathcal{P}|\equiv(X,Y)}{\mathcal{P}|\equiv X}$, $\frac{\mathcal{P}\lhd(X,Y)}{\mathcal{P}\lhd X}$, $\frac{\mathcal{P}|\equiv\mathcal{Q}|\sim(X,Y)}{\mathcal{P}|\equiv\mathcal{Q}|\sim X}$.

Let present some authentication goals we should proved in the demonstration of the proposed authentication scheme.

Goal 1: $U_i \mid\equiv (U_i \xleftrightarrow{SK} S_j)$;

Goal 2: $S_j \mid\equiv (U_i \xleftrightarrow{SK} S_j)$.

Next, let present the corresponding idealised protocol.

Message 1: $U_i \rightarrow S_j$: $(\{ID_i, SID_j, \langle ID_i, SID_j \rangle_{K_i}\}_{X'}, X)$;

Message 2: $S_j \rightarrow RC$: $(X, Y, \{ID_i, SID_j, \langle ID_i, SID_j\rangle_{K_i}\}_{X'}, \langle\{ID_i, SID_j, \langle ID_i, SID_j\rangle_{K_i}\}_{X'}, Y\rangle_{s_j})$;

Message 3: $RC \rightarrow S_j$: $(\langle X, Y\rangle_{s_j}, \langle X', Y, S_j \mid\sim Y\rangle_{K_i})$;

Message 4: $S_j \rightarrow U_i$: $(\langle X', Y, S_j \mid\sim Y\rangle_{K_i}, \langle X, Y\rangle_{SK_j})$;

Message 5: $U_i \rightarrow S_j$: $\langle SID_j, Y\rangle_{SK_j}$.

We make the following assumptions about the initial state of the scheme to further analyze the proposed scheme:

Let present the following assumptions for analyzing our scheme:

Assumption 1: $U_i \mid\equiv (U_i \overset{K_i}{\rightleftharpoons} RC)$
Assumption 2: $S_j \mid\equiv (S_j \overset{s_j}{\rightleftharpoons} RC)$
Assumption 3: $RC \mid\equiv (S_j \overset{s_j}{\rightleftharpoons} RC)$
Assumption 4: $U_i \mid\equiv \sharp(X')$
Assumption 5: $S_j \mid\equiv \sharp(Y)$
Assumption 6: $S_j \mid\equiv RC \Rightarrow (X, Y)$
Assumption 7: $S_j \mid\equiv \beta$
Assumption 8: $U_i \mid\equiv RC \Rightarrow (X', Y, S_j \mid\sim Y)$
Assumption 9: $U_i \mid\equiv \alpha$
Assumption 10: $U_i \mid\equiv X$
Assumption 11: $U_i \mid\equiv SID_j$
Assumption 12: $S_j \mid\equiv Y$
Assumption 13: $S_j \mid\equiv SID_j$

With the above assumptions and logical postulates, we prove the completeness of our scheme as follows:

Upon RC obtaining Message 2, we can prove that:

$$RC \lhd (X, Y, \{ID_i, SID_j, \langle ID_i, SID_j\rangle_{K_i}\}_{X'}, \langle\{ID_i, SID_j, \langle ID_i, SID_j\rangle_{K_i}\}_{X'}, Y\rangle_{s_j}).$$

Based on the jurisdiction rule, we can prove that:

$$RC \lhd \langle\{ID_i, SID_j, \langle ID_i, SID_j\rangle_{K_i}\}_{X'}, Y\rangle_{s_j}.$$

Based on the Assumption 3 and the message-meaning rule, we can prove that:

$$RC \mid\equiv S_j \mid\sim (\{ID_i, SID_j, \langle ID_i, SID_j\rangle_{K_i}\}_{X'}, Y).$$

Based on the jurisdiction rule, we can prove that:

$$RC \mid\equiv S_j \mid\sim Y.$$

Upon S_j obtaining Message 3, we can prove that:

$$S_j \lhd (\langle X, Y\rangle_{s_j}, \langle X', Y, S_j \mid\sim Y\rangle_{K_i}).$$

Based on the jurisdiction rule, we can prove that:

$$S_j \lhd \langle X, Y\rangle_{s_j}.$$

Based on Assumption 2 and the message-meaning rule, we can prove that:

$$S_j \mid\equiv RC \mid\sim (X, Y).$$

Based on Assumption 5 and the freshness-conjuncatenation rule, we can prove that:

$$S_j \mid\equiv \sharp(X, Y).$$

Based on $S_j \mid\equiv RC \mid\sim (X, Y)$ and the nonce-verification rule, we can prove that:

$$S_j \mid\equiv RC \mid\equiv (X, Y).$$

Based on Assumption 6 and the jurisdiction rule, we can prove that:

$$S_j \mid\equiv (X, Y).$$

Based on the jurisdiction rule, we can prove that:

$$S_j \mid\equiv X.$$

Based on $SK_j = \beta \times X$ and Assumption 7, we can prove that:

$$S_j \mid\equiv SK_j.$$

Based on $SK = H(SID_j \| X \| Y \| SK_j)$, $S_j \mid\equiv SK_j$ and Assumption 12, 13, we can prove that:

$$S_j \mid\equiv (U_i \overset{SK}{\longleftrightarrow} S_j)(\textbf{Goal 2}).$$

Upon U_i receiving Message 4, we can prove that:

$$U_i \lhd (\langle X', Y, S_j \mid\sim Y\rangle_{K_i}, \langle X, Y\rangle_{SK_j}).$$

Based on the jurisdiction rule, we can prove that:

$$U_i \lhd \langle X', Y, S_j \mid\sim Y\rangle_{K_i}.$$

Based on the Assumption 1 and the message-meaning rule, we can prove that:

$$U_i \mid\equiv RC \mid\sim (X', Y, S_j \mid\sim Y).$$

Based on Assumption 4 and the freshness-conjuncatenation rule, we can prove that:

$$U_i \mid\equiv \sharp(X', Y, S_j \mid\sim Y).$$

Based on $U_i \mid\equiv RC \mid\sim (X', Y, S_j \mid\sim Y)$ and the nonce-verification rule, we can prove that:

$$U_i \mid\equiv RC \mid\equiv (X', Y, S_j \mid\sim Y).$$

Based on Assumption 8 and the jurisdiction rule, we can prove that:

$$U_i \mid\equiv (X', Y, S_j \mid\sim Y).$$

Based on the jurisdiction rule, we can prove that:

$$U_i \mid\equiv S_j \mid\sim Y,$$

$$U_i \mid\equiv Y.$$

Based on $SK_j = \alpha \times Y$ and Assumption 9, we can prove that:

$$U_i \mid\equiv SK_j.$$

Based on $SK = H(SID_j \| X \| Y \| SK_j)$, $U_i \mid\equiv SK_j$ and Assumption 10, 11, we can prove that:

$$U_i \mid\equiv (U_i \overset{SK}{\longleftrightarrow} S_j)(\textbf{Goal 1}).$$

3.2 Security Evaluation

In this section, we prove our protocol could eliminate some common security flaws and achieve several significative properties.

3.2.1 Preserve User Anonymity

Suppose that all of authentication messages $\{D_i, X, Y, V_1, V_2, V_3, V_4, V_5\}$ transmitted between U_i, S_j and RC are obtained by attackers. The chosen random numbers α and β have randomness property, and they guarantee these parameters are all session-variant. Accordingly, without knowing α and β, the adversary will have to solve the computation Diffie-Hellman problem to retrieve specific static element in the transmitted messages. Hence, our scheme could overcome the security flaw of user anonymity breach.

3.2.2 Forward secrecy

In the proposed protocol, random numbers α and β are used to compute the session key SK, which security is guaranteed by the computation Diffie-Hellman problem. Hence, the adversary need to solve the hard problem to generate the session key, in other words, our protocol provides the property of forward secrecy.

3.2.3 Off-line Password Guessing Attack

The non-tamper resistant smart cards no longer secure stored data, and the adversary can reveal the secret information $\{B_i, C_i, r, Enc(), P, P_{pub}, H(\cdot)\}$ in another legitimated user U_i's smart card [13,24]. Even after gathering these information, the attacker could not guess ID_i and PW_i from $C_i = H(ID_i \| PW_i \| r)$ at the same time. The impossibility of guessing two parameters correctly simultaneously in polynomial time demonstrated that our scheme could resist off-line password guessing attack with smart card security breach.

3.2.4 Forgery Attack

In our proposal, the adversary has to generate a valid message $\{D_i, X\}$ if he wants to forgery the legal user U_i, where $D_i = Enc_{H(X\|X')}(ID_i, SID_j, H(ID_i\|K_i\|SID_j))$. The adversary $\mathcal{A}$ could not generate D_i with the knowledge of K_i, which is secured by $A_i = H(ID_i\|RPW_i)$ and stored in the U_i's smart card. With the demonstrated identity and password confidentiality, we can obtain that our scheme could overcome forgery attack.

3.2.5 Server Impersonating Attack

In the proposal, the adversary $\mathcal{A}$ impersonates S_j to fool the remote user U_i with a forgery response message $\{V_3, V_4\}$, where $V_3 = H(K_i\|X'\|Y)$, $V_4 = H(X\|Y\|SK_j)$. Nevertheless, $SK_j = \beta \times X = \alpha\beta \times P$ is impossible for $\mathcal{A}$ to compute without the knowledge of α or β. Thus, $\mathcal{A}$ could not transmit to U_i a valid response message to fool U_i and our proposal is able to withstand server impersonating attack.

3.2.6 Replay Attacks

The replay attack is that attackers re-submit authentication messages transmitted between U_i, S_j, RC to tamper with the information. It is impossible for our proposal since the authentication messages are contributed to random nonce. Neither the replay of an old login message $\{X, D_i\}$ in the step A.1 nor the replay of the response message $\{V_3, V_4\}$ of the service providing server S_j in the step A.4 of the authenticated key agreement phase, due to the random numbers are updated for every session and $\mathcal{A}$ could not get the random numbers, as it will fail in step A.4 and step A.6 of authenticated key agreement phase. Therefore, our protocol can withstand replay attack.

3.2.7 Known Key Attack

Since neither the structure of session key SK is the same with any other authentication message, nor SK functions as part of any other authentication message, the leakage of SK does not affect other unexposed sessions. Thus, the known key attack is resisted effectively.

3.2.8 Proper Mutual Authentication

Our proposed authentication scheme for multiple servers architecture could offer proper mutual authentication. U_i transmits the login request $\{D_i, X\}$ to server S_j for service access. And then, S_j adds its computed values Y and V_1 for mutual authentication. The registration center RC employs these messages to validate U_i and S_j. If any one is unauthentic, RC rejects the login request. Otherwise, it distributes the parameters $\{V_2, V_3\}$ to S_j. S_j verifies the correction of V_2 and computes V_4 as the challenge message to U_i. Subsequently, U_i uses the received messages $\{V_3, V_4\}$ to validate both RC and S_j. Further, he/she responses V_5 for the final session key verification. Noticeable, any fabricated message in the whole process cannot pass the verification. Therefore, our scheme could offer proper mutual authentication.

4 Performance and Functionality Analysis

In this section, we will evaluate our protocol in the performance and functionality by making comparisons with some other related protocols [25,28,34]. In the following, let define some notations used to analyze the computational complexity for the aforementioned protocol: T_{sym} indicates the time complexity of symmetrical encryption and T_{asy} is the time complexity of the asymmetric encryption. Noticeable, executing exclusion-OR operation and string concatenation operation consume very few computation resources, in the evaluation of performance we usually neglect the computational complexity of them.

Table 2: Comparisons of functionality

	Srinivas et al.'s [28]	Zhu et al.'s [34]	Mishra's [25]	Ours
Preserving User anonymity	No	Yes	Yes	Yes
Prevention of forgery attack	Yes	No	No	Yes
Prevention of off-line dictionary attack	No	No	Yes	Yes
Prevention of server impersonating attack	Yes	No	No	Yes
Prevention of replay attack	Yes	Yes	Yes	Yes
Prevention of known key attack	Yes	Yes	Yes	Yes
Mutual authentication	Yes	Yes	Yes	Yes
Providing correct proof of BAN-logic	No	Yes	No	Yes

Table 3: Performance comparisons

	Srinivas *et al.*'s [28]	Zhu *et al.*'s [34]	Mishra's [25]	Ours
Computational cost	$11T_{sys} + 4T_{asy}$	$22T_{sys} + 8T_{asy}$	$9T_{sys} + 8T_{asy}$	$19T_{sys} + 6T_{asy}$

Table 2 lists the functionality comparisons of our proposed protocol and other related protocols [25,28,34]. Obviously, we can conclude that our protocol is more robust, due to it not only could prevent all known attacks, but also provides several security properties. Furthermore, we also provide the formal proof validated by BAN-logic.

Table 3 shows the performance comparisons of our protocol and other related protocols [25, 28, 34]. According to Table 3, we know that the cost of the proposed protocol is slightly higher than the [25, 28] and lower than the scheme in [34]. However, our protocol can achieve all security properties as mentioned in Table 2.

5 Conclusions

In this paper, we present a robust remote user authentication scheme for multi-server architecture using elliptic curve cryptosystem. The proposal not only could overcome a range of network flaws, but also achieves ample of security properties. In addition, we employed BAN-logic to validate the proposed scheme. The performance of our scheme also indicates relative excellent performance, which is more suitable for practical applications.

References

[1] M. Burrows, M. Abadi and R. Needham, "A logic of authentication," *ACM Transactions on Computer Systems*, vol. 8, no. 1, pp. 18-36, 1990.

[2] C. C. Chang and C. Y. Lee, "A smart card-based authentication scheme using user identify cryptography," *International Journal of Network Security*, vol. 15, no. 2, pp. 139-147, 2013.

[3] T. Y. Chen, C. C. Lee, M. S. Hwang, J. K. Jan, "Towards secure and efficient user authentication scheme using smart card for multi-server environments," *Journal of Supercomputing*, vol. 66, no. 2, pp. 1008–1032, Nov. 2013.

[4] 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.

[5] D. L. Guo and F. T. Wen, "A more robust authentication scheme for roaming service in global mobility networks using ECC," *International Journal of Network Security*, vol. 18, no. 2, pp. 217-233, 2016.

[6] H. C. Hsiang and W. K. Shih, "Improvement of the secure dynamic ID based remote user authentication scheme for multi-server environment," *Computer Standards & Interface*, vol. 31, no. 6, pp. 1118-1123, 2009.

[7] M. S. Hwang, C. C. Lee, Y. L. Tang, "An improvement of SPLICE/AS in WIDE against guessing attack", *International Journal of Informatica*, vol. 12, no. 2, pp.297–302, Apr. 2001.

[8] M. S. Hwang, Li-Hua Li, "A New Remote User Authentication Scheme Using Smart Cards", *IEEE Transactions on Consumer Electronics*, vol. 46, no. 1, pp. 28–30, Feb. 2000.

[9] M. S. Hwang, S. F. Tzeng, C. S. Tsai, "Generalization of proxy signature based on elliptic curves", *Computer Standards & Interfaces*, vol. 26, no. 2, pp. 73–84, 2004.

[10] W. S. Juang, "Efficient multi-server password authenticated key agreement using smart cards," *IEEE Trans on Consumer Electronics*, vol. 50, no. 1, pp. 251-255, 2004.

[11] M. K. Khan and D. B. He, "A new dynamic identity-based authentication protocol for multi-server environment using elliptic curve cryptography," *Security & Communication Networks*, vol. 5, no. 11, pp. 1260-1266, 2012.

[12] N. Koblitz, "Elliptic curve cryptosystem," *Mathematics of Computation*, vol. 48, no. 177, pp.203-209, 1987.

[13] P. Kocher, J. Jaffe and B. Jun, "Differential power analysis," *Proceedings of Advances in Cryptology*, pp. 388-397, 1999.

[14] L. Lamport, “Password authentication with insecure communication,” *Communications of the ACM*, vol. 24, no. 11, pp. 770-772, 1981.

[15] C. C. Lee, Y. M. Lai and C. T. Li, “An improved secure dynamic ID based remote user authentication scheme for multi-server environment, *International Journal of Security and Its Applications*, vol. 6, no. 2, pp. 203-210, 2012.

[16] C. C. Lee, T. H. Lin and R. X. Chang, “A secure dynamic ID based remote user authentication scheme for multi-server environment using smart cards,” *Expert Systems with Applications*, vol. 38, no. 11, pp. 13863-13870, 2011.

[17] C. C. Lee, C. H. Liu, and M. S. Hwang, “Guessing attacks on strong-password authentication protocol”, *International Journal of Network Security*, vol. 15, no. 1, pp. 64–67, 2013.

[18] L. H. Li, L. C. Lin and M. S. Hwang, “A remote password authentication scheme for multiserver architecture using neural networks,” *IEEE Trans on Neural Network*, vol. 12, no. 6, pp. 1498-1504, 2001.

[19] X. Li, J. Ma, W. D. Wang, Y. P. Xiong and J. Z. Zhang, “A novel smart card and dynamic ID based remote user authentication scheme for multi-server environments,” *Mathematical and Computer Modelling*, vol. 58, no. 1-2, pp. 85-95, 2013.

[20] X. Li, J. Niu, S. Kumari, J. Liao and W. Liang, “An enhancement of a smart card authentication scheme for multi-server architecture,” *Wireless Personal Communications An International Journal*, vol. 80, no. 1, pp. 175-192, 2015.

[21] X. Li, Y. P. Xiong, J. Ma and W. D. Wang, “An efficient and security dynamic identity based authentication protocol for multi-server architecture using smart cards,” *Journal of Network and Computer Applications*, vol. 35, no. 2, pp. 763-769, 2012.

[22] Y. P. Liao and S. S. Wang, “A secure dynamic ID based remote user authentication scheme for multi-server environment,” *Computer Standards & Interface*, vol. 31, no. 1, pp. 24-29, 2009.

[23] I. C. Lin, M. S. Hwang and L. H. Li, “A new remote user authentication scheme for multi-server architecture,” *Future Generation Computer System*, vol. 19, no. 1, pp. 13-22, 2003.

[24] T. S. Messerges, E. A. Dabbish, and R. H. Sloan, “Examining smart-card security under the threat of power analysis attacks,” *IEEE Trans Comput*, vol. 5, no.51, pp. 541-552, 2002.

[25] D. Mishra, “Design and analysis of a provably secure multi-server authentication scheme,” *Wireless Personal Communications*, vol. 86, no. 3, pp. 1-25, 2016.

[26] R. S. Pippal, C. D. Jaidhar and S. Tapaswi, “Robust smart card authentication scheme for multi-server architecture,” *Wireless Personal Communications*, vol. 72, no. 1, pp. 729-745, 2013.

[27] 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.

[28] J. Srinivas, S. Mukhopadhyay and D. Mishra, “A self-verifiable password based authentication scheme for multi-server architecture using smart card,” *Wireless Personal Communications*, 2017. DOI:10.1007/s11277-017-4476-9

[29] W. J. Tsaur, C. C. Wu and W. B. Lee, “A smart card-based remote scheme for password authentication in multi-server Internet services,” *Computer Standards & Interfaces*, vol. 27, no. 1, pp. 39-51, 2004.

[30] J. L. Tsai, “Efficient multi-server authentication scheme based on one-way hash function without verification table,” *Computers & Security*, vol. 27, no. 3-4, pp. 115-121, 2008.

[31] S. F. Tzeng, M. S. Hwang, “Digital signature with message recovery and its variants based on elliptic curve discrete logarithm problem”, *Computer Standards & Interfaces*, vol. 26, no. 2, pp. 61–71, 2004.

[32] Y. Y. Wang, J. Y. Liu, F. X. Xiao and J. Dan, “A more efficient and secure dynamic ID-based remote user authentication scheme,” *Computer Communications*, vol. 4, no. 32, pp.583-585, 2009.

[33] J. H. Wei, W. F. Liu and X. X. Hu, “Secure and efficient smart card based remote user password authentication scheme,” *International Journal of Network Security*, vol. 18, no. 4, pp. 782-792, 2016.

[34] H. F. Zhu, Y. F. Zhang and Y. Sun, “Provably Secure Multi-server Privacy-Protection System Based on Chebyshev Chaotic Maps without Using Symmetric Cryptography,” *International Journal of Network Security*, vol. 18, no. 5, pp. 803-815, 2016.

Biography

Xueqin Zhang received the B.S. and M.S. degrees in Computer Science and Technology from Henan Polytechnic University, Jiaozuo, Henan, China in 2005 and 2008, respectively. She is a lecturer in Nanyang Normal University, Nanyang, Henan, China. Her research interests include data mining and information security.

Baoping Wang received B.S. degree in Computer Application Technology from Henan University, Kaifeng, China in 1997 and M.S. in Computer Application Technology from Guizhou University, Guiyang, China in 2006. He is an associate professor in Nanyang Normal University, Henan, China. His research interests include computer network technology and information security.

Wenpeng Zhang received B.S. degree in Computer Application Technology from Henan University, Kaifeng, China, in 1997, and M.S. in Computer Application Technology from Wuhan University of Technology, Wuhan, China, in 2007. He is an associate professor in Nanyang Normal University, Henan, China. His research interests include computer network technology and Database technology.

A Secure Privacy-Preserving Cloud Auditing Scheme with Data Deduplication

Chen Li and Zhenhua Liu

(Corresponding author: Chen Li)

School of Mathematics and Statistics, Xidian University
No. 2 South Taibai Road, Xi'an, Shaanxi 710071, P.R. China
(Email: lichen0906@foxmail.com)

(Received July 18, 2017; revised and accepted Nov. 5, 2017)

Abstract

In this paper, we propose a privacy-preserving public auditing scheme supporting data deduplication. In the process of auditing, since the authentication tag of a message contains only one element, the storage and transmission cost of the tag can be significantly reduced. Meanwhile, by eliminating user's private key in the response, our scheme achieves unconditional anonymity against third party auditor. Moreover, during data deduplication, Bloom filter is utilized to efficiently check ownership of the data that a user claims to have. For public auditing, the proposed scheme is proven to be uncheatable and anonymous under the variant of the BDH hardness assumption in the random oracle model. And security analysis indicates that our scheme is unforgeable during deduplication. Compared to existing schemes with similar features, our scheme achieves higher security and better functionality through function evaluation and security analysis.

Keywords: Cloud Storage; Data Deduplication; Privacy Preservation; Public Auditing

1 Introduction

With the development of cloud computing, a rising number of enterprises and organizations choose to outsource their data to a third-party cloud service provider, who can provide resource-constrained users with convenient storage and computing services and thus reducing users' storage burden [15, 17]. Although cloud storage offers many advantages, it also brings some security challenges such as data integrity and storage efficiency.

Different from local data, cloud data is stored in an uncertain domain via Internet. Therefore, users surely can suspect the integrity of their data that stored in cloud due to the fact that their data is vulnerable to the attack from both outside and inside of the cloud [7,8]. Once their data is corrupted, cloud server might passively hide some data loss from users to maintain their reputation. Worse, due to the insufficiency of storage space or some other economic reasons, cloud server might even delete users' data and cheat users that their data still stores integrally. To cope with the conflict between resource-constrained users and large amounts of data, it is essential to consider how can users verify the integrity of data efficiently without retrieving them.

Since cloud service is increasingly used, data redundancy inevitably occurs in cloud storage. Research shows that 80% - 90% of cloud data is redundant [12, 22], and this rate is still increasing, which causes a big waste of cloud storage space. In order to save storage space in cloud, a technique called deduplication came into being, in which cloud server keeps only one single copy of data and sends a storage link to every user who possess the data. However, several security threats potentially exists during deduplication [9]. For instance, if a malicious user needs to gain access to a message that already exists in the cloud, he can pass the examination by only owing the hash value of the massage rather than the concrete message. It is obvious that cloud server cannot distinguish whether user indeed possess the data only through matching its hash value. Therefore, how to convince cloud server that user who upload a duplicate of the data indeed possess the data becomes another issue in cloud service.

In this paper, aiming at solving both data integrity and storage efficiency, we concentrate on how to design a secure and efficient public auditing scheme with data deduplication and users' anonymity. Inspired by a Proof of Ownership protocol that Blasco *et al.* [3] designed, we will propose a privacy-preserving auditing scheme with data deduplication which achieves a better trade-off between efficiency and security through improving Wu *et al.*'s auditing scheme [23].

The rest of this paper is organized as follows. A review about some related works is given in Section 2. Some preliminaries are presented in Section 3. The system model and security model for the proposed scheme are described in Section 4. The concrete construction of privacy-preserving auditing scheme with data deduplication is detailed in Section 5. We analyze the proposed scheme in Section 6. The performance evaluation and efficiency improvement are discussed in Section 7. Finally, some concluding remarks are given in Section 8.

2 Related Works

This section mainly consists of the research advance of three related works: integrity auditing, data deduplication, and auditing schemes with data deduplication. Moreover, a comparison among several related works that achieve both integrity auditing and data deduplication is shown below.

2.1 Integrity Auditing

In order to efficiently verify the integrity of stored cloud data, Ateniese *et al.* [2] came up with the notion of provable data possession (PDP) in 2007. More precisely, under the situation that cloud server could hide data errors for his own benefit, PDP allows cloud server to proof that users' data are completely stored without retrieving the entire data. Considering the size of users' data and users' limited computation resource, outsourced data are not suitable for users themselves to audit in many cases. Therefore, it is a preferable way to introduce a third party auditor (TPA) to ensure data's integrity and availability. Liu *et al.* [16] summarized some existing research situations and development trends of public auditing.

Although PDP can assist user verifying data integrity, TPA may reveal users' identities for personal benefits during public auditing [27]. To preserve users' privacy, Wang *et al.* [19] came up with a public auditing scheme supporting data sharing and privacy-preserving. In Wang *et al.*'s scheme, challenge generated by TPA utilizes all users' public keys, and thus the privacy of user's identity can be realized. Several constructions [11, 20, 21] were subsequently presented. The main solution in these constructions is the anonymity of ring signature or group signature techniques. However, the tag size of ring signature or group signature is significantly large, which causes a higher transmission and verification cost than many traditional signature schemes. Therefore, by reducing the size of authentication tag to only one element, Wu *et al.* [23] proposed an efficient auditing scheme, which can achieve users' identity privacy by eliminating user's private key during a challenge-and-response protocol. Besides, the authentication tag in the scheme is irrelevant to the number of users within the group.

2.2 Data Deduplication

For increasing storage efficiency, cloud server needs to identify and remove redundant data by retaining only one copy of each block (block-level deduplication) or file (file-level deduplication). And data deduplication can take place before data are uploaded to cloud server (client-side deduplication) or after they are uploaded (server-side deduplication) [10]. However, server-side deduplication only reduces storage cost of cloud server instead of reducing bandwidth. Hence, client-side deduplication is more widely used, since user does not need to upload data if a duplicate already exists, and thus reducing bandwidth cost between user and cloud server remarkably.

During a client-side deduplication system, user sends the hash value of data to cloud server and cloud server checks whether the duplicate exists in cloud storage. Nonetheless, Halevi *et al.* [9] explained several security attacks that may occur in client-side deduplication systems. For instance, if a malicious user needs to gain access to a message that already exists in the cloud, he can pass the examination by only owing the hash value of the massage rather than the concrete message. As a solution to these attacks, Halevi *et al.* [9] first introduced the concept of Proof of Ownership (PoW), which has been extended into a number of related works [18, 24], but all require a higher computational complexity. Therefore, based on Bloom filters, Blasco *et al.* [3] introduced a novel efficient PoW protocol that provides a flexible and scalable solution to the weaknesses of client-side deduplication.

2.3 Auditing Schemes with Data Deduplication

From the above discussions, privacy-preserving public auditing and data deduplication are two main branches of the research for efficient cloud storage [13]. So it becomes a significant matter to support these two functions simultaneously. However, a mechanical combination of privacy-preserving public auditing and efficient deduplication mechanisms cannot efficiently solve both data deduplication and integrity auditing. The reason is that storage efficiency contradicts with the authentication tags (*i.e.*, signatures) during public auditing.

To the best of our knowledge, only the following papers achieve both public auditing and data deduplication. There follows some analysis. Yuan and Yu [26] proposed a constant cost storage public auditing scheme supporting data deduplication, but they did not consider the privacy-preserving property. Then Alkhojandi and Miri [1] showed a privacy-preserving public auditing mechanism supporting a variant of client-side deduplication performed by a mediator, but the mediator may reveal users' data during deduplication. Besides, Alkhojandi and Miri's scheme failed to reduce user's bandwidth overhead. Subsequently, Li *et al.* [14] presented a scheme called SecCloud which aims to solve both data integrity and secure deduplication by using a MapReduce cloud to replace TPA. Nevertheless, uploading data to the MapReduce cloud violates the privacy of user's identity. Furthermore, Li *et al.*'s scheme [14] cannot reduce the bandwidth for users. To solve the bandwidth problem, Kardas and Kiraz [13] came up with a secure deduplication scheme that supports client-side deduplication along with privacy-preserving public auditing. However, when uploading a message, user needs to compute at least an asymmetric encryption to complete the PoW protocol,

Table 1: Comparison of auditing mechanisms with deduplication

Schemes	Anonymous	No extra entities	Public auditing	User-side bandwidth reduction (client-side deduplication)
Yuan *et al.* [26]	No	Yes	Yes	Yes
Naelah *et al.* [1]	Yes	No	Yes	No
Li *et al.* [14]	No	Yes	Yes	No
Kardas *et al.* [13]	Yes	No	Yes	Yes
Ours	Yes	Yes	Yes	Yes

which consequentially causes efficiency problem. Besides, the key server may recover part of message's encrypt key through the value received by user. Table 1 compares the function that these schemes [1,13,14,26] can realize.

3 Preliminaries

We now explain some preliminary notions that will form the foundations of our scheme.

3.1 Bilinear Pairings

Let $\mathbb{G}_1, \mathbb{G}_2$ be the cyclic groups of prime order p, g be a generator of $\mathbb{G}_1$, and $e : \mathbb{G}_1 \times \mathbb{G}_1 \rightarrow \mathbb{G}_2$ be a bilinear map [5] with the following properties:

1) **Bilinearity**: $e(g^a, g^b) = e(g, g)^{ab}, a, b \in \mathbb{Z}_p$;

2) **Non-degeneracy**: There exist $u, v \in \mathbb{G}_1$ such that $e(u, v) \neq 1$;

3) **Computability**: For all $u, v \in \mathbb{G}_1$, $e(u, v)$ can be efficiently computed.

3.2 Complexity Assumptions

Definition 1. *Given (g, g^a, g^b, g^c), for random $a, b, c \in \mathbb{Z}_p^*$, the Bilinear Diffie-Hellman(BDH) problem [5] is to compute $e(g, g)^{abc}$.*

A challenger $\mathcal{C}$ has advantage ε in solving the BDH problem if

$$\Pr\left[e(g, g)^{abc} \leftarrow \mathcal{A}(g, g^a, g^b, g^c)\right] \geq \varepsilon.$$

The (ε, t)-BDH assumption holds if no t-time algorithm has the advantage at least ε in solving the BDH problem.

Definition 2. *Given (g, g^a, g^b, g^{ac}), for random $a, b, c \in \mathbb{Z}_p^*$, the variant of the BDH (vBDH) problem [23] is to compute $e(g, g)^{bc}$.*

A challenger $\mathcal{C}$ has advantage ε in solving the vBDH problem if

$$\Pr\left[e(g, g)^{bc} \leftarrow \mathcal{A}(g, g^a, g^b, g^{ac})\right] \geq \varepsilon.$$

The (ε, t)-vBDH assumption holds if for any t-time algorithm, the advantage ε in solving the vBDH problem is negligible.

3.3 Bloom Filter

As a probabilistic data structure, Bloom filter [4] can approximately represent the elements of a set and verify the membership of elements. Since both the storage space and the insert or query time are constant, Bloom filter has the advantage of memory and time efficiency [3]. On the other hand, Bloom filter sacrifices a certain amount of accuracy, since an element that is not in the set may be recognized as being part of the set, which is called as false positives. But false negatives cannot occur in Bloom filter.

Bloom filter consists of k random hash functions $h_1(\cdot), h_2(\cdot), \cdots, h_k(\cdot)$ and an m bit array. When initializing the Bloom filter, all the positions of bit array are set to 0. To insert an element x into Bloom filter, we compute k addresses $a_1 = h_1(x) \mod m, a_2 = h_2(x) \mod m, \cdots, a_k = h_k(x) \mod m$ and set the position of corresponding bit array to 1. To determine whether the element is in the set, we need to compute k hash values $h'_1, h'_2, \cdots, h'_k$ and check if all the corresponding values are 1. With certain false positive rate, the element is in the set if and only if all bits are 1 in Bloom filter. In other words, the element is not in the set if not all the bits are 1.

4 Problem Statement

4.1 System Model

In this system, there are three main entities named cloud server, user and TPA, as shown in Figure 1.

- Cloud server (CS) provides users with cloud storage and computing service. Therefore, user can rent or buy storage space from CS to store their individual data and perform some specific computation with CS's help. The data format stored in CS is a tuple $(index, message, tag)$.

- User computes an index according to a message, and uses her or his secret key to compute the message's tag. Then, the user uploads the tuple $(index, message, tag)$ to CS. During data deduplication, the user does not upload the message when the duplicate exists.

- TPA can verify users' messages by challenging CS with a message index set and the corresponding challenge value. Afterwards, TPA checks the response from CS and sends the auditing result back to users.

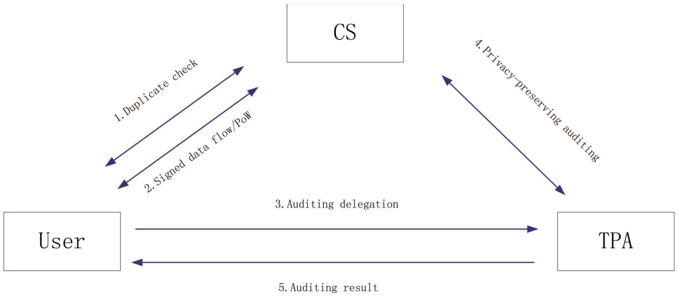


Figure 1: Architecture of a general scheme

Figure 1 shows an outline of the procedure for a general scheme that supports deduplication and public auditing.

1) Before uploading a message, user first checks if there is a duplicate one in CS.

2) When uploading a message, user either sends the message with a corresponding signature or passes the PoW challenge to avoid uploading the message.

3) If user needs to check the integrity of data, she or he sends an auditing delegation to TPA.

4) TPA and CS run the privacy-preserving auditing interactively.

5) After verifying the auditing result, TPA sends the result back to user.

4.2 General Scheme

A scheme supporting public auditing and deduplication [1, 13, 14, 26] is generally composed of these six algorithms, namely **Initialize**, **KeyGen**, **FileUpload**, **Challenge**, **Respond** and **Verify**. The detailed algorithms come as follows.

- **Initialize** (1^k): Take the security parameter 1^k as input, and output the public parameter *params*.

- **KeyGen**: User u_i's secret and public key pair (sk_i, pk_i) are generated by running the key generation algorithm.

- **FileUpload** (sk_i, id_j, m_j): If user u_i needs to upload a message m_j which is identified by the index id_j, then he computes $H(m_j)$ and sends $H(m_j)$ to CS for checking whether the message m_j has been stored in CS at first.

  Case 1: If there is no duplicate of m_j, user u_i computes m_j's authentication tag $\sigma_{i,j}$ using her or his secret key sk_i, generates Bloom filter BF_j by splitting m_j into n blocks $\{m_{j,1}, \cdots, m_{j,n}\}$, and then uploads the tuple $(id_j, m_j, \sigma_{i,j})$ along with the Bloom filter BF_j.

  Case 2: If a duplicate of m_j exists, for $1 \leq t \leq n$, CS randomly chooses t blocks and sends a corresponding identity set $K = \{k_1, \cdots, k_t\}$ to user u_i.

  According to the set K, user u_i computes a set of tokens $\{T_{j,k_q} | q = 1, \cdots, t\}$ and sends the set back to CS for checking whether the tokens are in the Bloom filter BF_j.

- **Challenge** $(pk_1, \cdots, pk_d, s, \mathcal{I})$: Taking public keys of d users, a secret value s and a random index subset $\mathcal{I}$ of the entire storage space as input, TPA computes and sends a challenge *chal* to CS.

- **Respond** $(chal, M, \Sigma)$: When CS receives the challenge, he computes the response (μ, σ_{res}) with a set of messages $M = \{m_j | id_j \in \mathcal{I}\}$ and a set of corresponding tags $\Sigma = \{\sigma_{i,j} | id_j \in \mathcal{I}\}$. Subsequently, the response (μ, σ_{res}) is sent to TPA.

- **Verify** $(\mu, \sigma_{res}, chal, s)$: Using challenge *chal* and secret value s as inputs, TPA checks whether the response (μ, σ_{res}) is correct and outputs "true" if (μ, σ_{res}) pass the validation or "false" otherwise.

4.3 Security or Threat Model

This subsection consists of three security aspects named uncheatable, information-theoretical anonymous and unforgeable of tokens.

Since CS is semi-honest, he may try to deceive users that their data are still securely stored when he is unable to recover the data due to some technical problems or storage devices damage. Moreover, with the risk that TPA can reveal users' identities during auditing process, the anonymity of user should be considered in a general scheme. Besides, a user might attempt to pass the PoW challenge so as to possess a message that the user does not.

Therefore, a general scheme should be uncheatable against adaptive chosen-message attack according to Wu *et al.*'s model [23], achieve information-theoretical anonymity which refers to Zhang and Zhao's model [28], and attain unforgeability of tokens based on Blasco *et al.*'s model [3]. In the rest of this subsection, we formalize the models mentioned above in the form of security model or threat model.

4.3.1 Uncheatability

During this part, CS is regarded as a semi-honest one who could attempt to cheat user. Therefore, a general scheme is uncheatable against adaptive chosen-message attack. In the following description, we consider a security game between an adversary $\mathcal{A}$ and a challenger $\mathcal{C}$.

Setup: $\mathcal{C}$ inputs the security parameter 1^k and runs the **Initialize** and **KeyGen** algorithms. Then $\mathcal{C}$ gives the public parameter *params* and the public keys $(pk_1, \cdots, pk_d)$ of all users to $\mathcal{A}$.

Sign Query: $\mathcal{A}$ could query the sign oracle adaptively for tag of a pair (id_j, m_j) under the public key pk_i that $\mathcal{A}$ chooses. $\mathcal{C}$ returns the corresponding tag $\sigma_{i,j}$ through running the **FileUpload** algorithm.

Challenge: $\mathcal{A}$ chooses set $\mathcal{I}^*$ from all the message indexes, and ensures that at least one index in $\mathcal{I}^*$ has not been queried in the sign oracle before. $\mathcal{C}$ generates a challenge *chal* of $\mathcal{I}^*$ from the **Challenge** algorithm and returns *chal* to $\mathcal{A}$.

Respond: Finally, $\mathcal{A}$ outputs the response (μ, σ_{res}).

We define the advantage of adversary $\mathcal{A}$ in cheating challenger $\mathcal{C}$ as

$$Adv(A) = \Pr\left[\begin{array}{c|c} \begin{array}{c}\textbf{Verify}\\ = \text{“}true\text{''}\end{array} & \begin{array}{c} (\textbf{params}, pk_1, \cdots pk_d) \\ \leftarrow \textbf{Setup}(1^k) \\ (pk_i, id_j, m_j) \leftarrow \mathcal{A} \\ \sigma_{i,j} \leftarrow \textbf{FileUp}(sk_i, id_j, m_j) \\ \mathcal{I}^* \leftarrow \mathcal{A} \\ chal \leftarrow \textbf{Chal}(pk_1, \cdots pk_d, \mathcal{I}^*) \\ (\mu, \sigma_{res}) \leftarrow \mathcal{A}(chal) \end{array}\end{array}\right]$$

Definition 3. *A general scheme is uncheatable against adaptive chosen-message attack if for any polynomial-time adversary $\mathcal{A}$, the advantage $Adv(\mathcal{A})$ is negligible.*

4.3.2 Information-Theoretical Anonymity

Suppose the challenge that TPA chooses only contains one message's index during **Challenge** and **Respond** algorithms, and TPA attempts to correctly determine the identity of user from the **Challenge** and **Respond** algorithms. Thus, TPA acts as a malicious one in this part.

A general scheme can achieve information-theoretical anonymity described by a game between an adversary $\mathcal{A}$ and a challenger $\mathcal{C}$.

Setup: $\mathcal{C}$ inputs the security parameter 1^k and runs the **Initialize** and **KeyGen** algorithms. $\mathcal{C}$ sends the public parameter *params* and all d users' secret and public key pairs $\{(sk_1, pk_1), \cdots, (sk_d, pk_d)\}$ to $\mathcal{A}$.

Challenge: $\mathcal{A}$ chooses a pair (id_j, m_j) and computes the challenge *chal* of this pair by running **Challenge** algorithm.

Respond: $\mathcal{C}$ picks $i \in \{1, \cdots, d\}$ at random and computes the tag $\sigma_{i,j}$ using i-th user's secret key sk_i through the **FileUpload** algorithm. Then, $\mathcal{C}$ generates the response (μ, σ_{res}) from the **Respond** algorithm and returns (μ, σ_{res}) to $\mathcal{A}$.

Guess: $\mathcal{A}$ checks whether the response is correct through the **Verify** algorithm and outputs $i' \in \{1, \cdots, d\}$ if the response passes the verification.

We define the advantage of the adversary $\mathcal{A}$ in distinguishing user's integrity of a pair $(message, tag)$ as

$$Adv(\mathcal{A}) = \left|\Pr(\cdot) - \frac{1}{d}\right|,$$

where

$$\Pr(\cdot) = \Pr\left[\begin{array}{c|c} i' = i & \begin{array}{c} (params, (sk_1, pk_1), \cdots, (sk_d, pk_d)) \\ \leftarrow \textbf{Setup}\,(1^k) \\ (id_j, m_j) \leftarrow \mathcal{A} \\ chal \leftarrow \textbf{Chal}\,(pk_1, \cdots pk_d, id_j) \\ i \leftarrow_R \{1, \cdots d\} \\ \sigma_{i,j} \leftarrow \textbf{FileUp}\,(sk_i, id_j, m_j) \\ (\mu, \sigma_{res}) \leftarrow \textbf{Res}\,(chal, m_j, \sigma_{i,j}) \\ i' \leftarrow \mathcal{A}\,(\mu, \sigma_{res}) \end{array}\end{array}\right]$$

Definition 4. *A general scheme achieves information-theoretical anonymity if for any polynomial-time adversary $\mathcal{A}$, the advantage $Adv(\mathcal{A})$ is negligible.*

4.3.3 Unforgeability of Tokens

The existing deduplication schemes [3, 24] did not construct a formal security model, but only give some threat models. Therefore, a threat model conducted by a malicious user is given below.

A malicious user's propose is to pass the PoW challenge for a message m_j he does not own. Suppose that the malicious user possesses several message blocks and the hash value of m_j. Therefore, the malicious user can attempt to forge tokens for passing the PoW challenge. Moreover, a general scheme cannot prevent a malicious user who almost obtains the whole message from passing the PoW challenge. In other words, the malicious user does not need to forge tokens if he already possess the message anyway.

5 Our Construction

In this section, we will construct a concrete scheme which mainly contains two parts. The first one is deduplication, which is conducted by user and CS. Before uploading a message, user divides the message into n blocks, and generates a Bloom filter utilizing a pseudorandom function. The other one is public auditing. Unlike existing auditing works [19, 20, 25] by adopting ring signature or group signature techniques, our scheme uses a constant-size tag generation algorithm, which is a variant of Boneh *et al.*'s signature scheme [6]. Therefore, the transmission and communication cost are less than these existing works [19, 20, 25]. The detailed algorithms of our scheme are shown below.

Initialize(1^k): Take the security parameter 1^k as input, and output the public parameter

$$params = \{\omega, g \in \mathbb{G}_1, H, H_1\},$$

where $H : \{0,1\}^* \to \mathbb{G}_1$, $H_1 : \{0,1\}^m \to \{0,1\}^l$ are two collision resistant hash functions, and m, l are the length of message and token respectively. Let $Prf : \{0,1\}^l \times \{0,1\}^* \to \{0,1\}^\kappa$ be a pseudorandom function [3], where κ is a positive integer.

KeyGen: User u_i selects $x_i \in \mathbb{Z}_p^*$ at random and computes g^{x_i}. Then user u_i's secret and public key pair are $(sk_i, pk_i) = (x_i, g^{x_i})$.

FileUpload(sk_i, id_j, m_j): Suppose that the j-th message is $m_j \in \mathbb{Z}_p$ with an index id_j. Before uploading m_j, user u_i computes and uploads $h_{i,j} = H(m_j)$ to CS. Upon receiving the upload request, CS first checks whether $h_{i,j}$ already exists.

Case 1: If there is no duplicate in CS, *i.e.* m_j does not exist in CS, CS returns "No Duplicate" to user u_i. Then user u_i computes the authentication tag of m_j as

$$\sigma_{i,j} = (H(id_j) \cdot \omega^{m_j})^{1/x_i}.$$

Furthermore, user u_i splits message m_j into n message blocks $\{m_{j,1}, \cdots, m_{j,n}\}$ with equal length. For each message block $m_{j,l} (l = 1, \cdots, n)$, user u_i computes its corresponding token $T_{j,l} = H_1(m_{j,l})$ and a pseudorandom value

$$P_{j,l} = Prf(T_{j,l}, l).$$

Then user u_i inserts every $P_{j,l}$ into Bloom filter BF_j and uploads the tuple

$$(index, message, tag) = (id_j, m_j, \sigma_{i,j})$$

along with the Bloom filter BF_j. After that, CS computes $H(m_j)$ and verifies whether

$$h_{i,j} = H(m_j),$$
$$e(\sigma_{i,j}, g) = e(H(id_j) \cdot \omega^{m_j}, pk_i)$$

hold. If these two euqations hold, CS stores the tuple $(id_j, m_j, \sigma_{i,j})$ along with the Bloom filter BF_j and returns a storage link of message m_j to user u_i. Otherwise, CS returns an error message to user u_i.

Case 2: If the duplicate of m_j exists, user u_i performs a PoW protocol by interacting with CS. Specifically speaking, CS chooses t message blocks at random and sends the identifier set of blocks $K = \{k_1, \cdots, k_t\}$ to user u_i, where $1 \le t \le n$. Upon receiving the set K, user u_i computes each token

$$T_{j,k_q} = H_1(m_{j,k_q}), \text{ for } q = 1, \cdots, t.$$

Then user u_i sends $\{T_{j,k_q} | q = 1, \cdots, t\}$ back to CS. For all t chosen message blocks, CS computes $P_{j,k_q} = Prf(T_{j,k_q}, k_q)$ with the returned $\{T_{j,k_q} | q = 1, \cdots, t\}$. Next, CS checks whether all P_{j,k_q} belong to the Bloom filter BF_j. If yes, a storage link of message m_j is sent to user u_i. Otherwise, returns an error message to user u_i.

Challenge$(pk_1, \cdots, pk_d, s, \mathcal{I})$: To check the integrity of users' data, TPA selects a random index subset $\mathcal{I}$ from the whole storage space $\mathcal{S}$, chooses $s \in \mathbb{Z}_p^*$, $h \in \mathbb{G}_1$, and $s_j \in \mathbb{Z}_p^*$ for every $id_j \in \mathcal{I}$ at random. Then TPA computes the challenge

$$chal = (Q, pk_{chal}),$$

where

$$Q = \{(id_j, s_j) | id_j \in \mathcal{I}\}$$

and

$$pk_{chal} = (pk_1^s, \cdots, pk_d^s, h, h^s).$$

Respond$(chal, M, \Sigma)$: Upon receiving the challenge $chal$ from TPA, CS checks whether

$$e(pk_i^s, h) = e(pk_i, h^s), \text{ for } i = 1, \cdots, d,$$

and computes the response (μ, σ_{res}) from a set of messages $M = \{m_j | id_j \in \mathcal{I}\}$ and a set of corresponding tags $\Sigma = \{\sigma_{i,j} | id_j \in \mathcal{I}\}$, where

$$\mu = \sum_{(id_j, s_j) \in Q} s_j \cdot m_j,$$
$$\sigma_{res} = \prod_{(id_j, s_j) \in Q} e(\sigma_{i,j}^{s_j}, pk_i^s).$$

Verify$(\mu, \sigma_{res}, chal, s)$: Finally, TPA checks whether

$$\sigma_{res} = e\left(\prod_{(id_j, s_j) \in Q} H(id_j)^{s_j} \cdot \omega^\mu, g^s\right),$$

and outputs the verification result.

6 Security Analysis

6.1 Consistency

In this part, we analyze the correctness mainly about the **Challenge** and **Respond** algorithms. During message uploading and integrity verification, assume that the tuple $(id_j, m_j, \sigma_{i,j})$ is stored in the whole storage space $\mathcal{S}$, and the tag $\sigma_{i,j}$ of each message m_j is signed by user u_i. Thus, for $id_j \in \mathcal{S}$, CS stores $(id_j, m_j, \sigma_{i,j})$, where

$$\sigma_{i,j} = (H(id_j) \cdot \omega^{m_j})^{1/x_i}.$$

To check the integrity of users' messages, TPA selects a random index subset $\mathcal{I} \subset \mathcal{S}$, chooses $s \in \mathbb{Z}_p^*$, $h \in \mathbb{G}_1$, and $s_j \in \mathbb{Z}_p^*$ for every $id_j \in \mathcal{I}$ at random.

Then TPA computes the challenge

$$chal = (Q, pk_{chal}),$$

where

$$Q = \{(id_j, s_j) | id_j \in \mathcal{I}\}$$

and

$$pk_{chal} = (pk_1^s, \cdots, pk_d^s, h, h^s).$$

With the challenge *chal* received from TPA, CS computes the response

$$\mu = \sum_{(id_j, s_j)\in Q} s_j \cdot m_j,$$

and

$$\begin{aligned}\sigma_{res} =& \prod_{(id_j,s_j)\in Q} e(\sigma_{i,j}^{s_j}, pk_i^s)\\ =& \prod_{(id_j,s_j)\in Q} e(H(id_j)^{s_j}, g^s)\cdot \prod_{(id_j,s_j)\in Q} e(\omega^{s_j\cdot m_j}, g^s).\end{aligned}$$

Finally, TPA checks the correctness of the response (μ, σ_{res}) by computing

$$\sigma_{res} = e\left(\prod_{(id_j,s_j)\in Q} H(id_j)^{s_j}\cdot \omega^{\mu}, g^s\right).$$

The above analysis indicates that

$$\begin{aligned}\sigma_{res} =& \prod_{(id_j,s_j)\in Q} e(H(id_j)^{s_j}, g^s)\cdot \prod_{(id_j,s_j)\in Q} e(\omega^{s_j\cdot m_j}, g^s)\\ =&e\left(\prod_{(id_j,s_j)\in Q} H(id_j)^{s_j}\cdot \omega^{\mu}, g^s\right).\end{aligned}$$

6.2 Uncheatability

In this section, we prove that our scheme is uncheatable in the random oracle model through the following theorem. The method of the proof is similar to the one in Wu *et al.*'s scheme [23].

Theorem 1. *If there exists an adversary $\mathcal{A}$ that has advantage ε in outputting a valid response of the challenge, then there is a simulation algorithm $\mathcal{C}$ that runs in polynomial time and has advantage at least ε/v in solving the vBDH problem through interacting with $\mathcal{A}$.*

Proof. Suppose that the simulator $\mathcal{C}$ receives an instance of vBDH problem as

$$(p, \mathbb{G}_1, \mathbb{G}_2, e, g, g^a, g^b, g^{ac}),$$

and the propose of $\mathcal{C}$ is to compute the solution $e(g, g)^{bc}$. By interacting with adversary $\mathcal{A}$ who runs in time t, queries hash oracle at most v times and could adaptively query the sign oracle, $\mathcal{C}$ computes the solution as the challenger in the following game.

Setup: $\mathcal{C}$ randomly chooses $r_0, r_1, \cdots, r_d \in \mathbb{Z}_p^*$, computes

$$(g^a)^{r_0}, (g^a)^{r_1}, \cdots, (g^a)^{r_d},$$

sets $\omega = (g^a)^{r_0}$ and selects hash function $H : \{0,1\}^* \to \mathbb{G}_1$, which can be regarded as the random oracle later. Then $\mathcal{C}$ returns the public parameter

$$params = \{g, \omega \in \mathbb{G}_1, H\}$$

and public keys of all d users

$$(pk_1, \cdots, pk_d) = ((g^a)^{r_1}, \cdots, (g^a)^{r_d})$$

to $\mathcal{A}$.

Hash Query: $\mathcal{A}$ can adaptively query the hash oracle for the hash values of messages' indexes. $\mathcal{C}$ maintains a list which is initially empty and randomly chooses $j^* \in \{1, \cdots, v\}$ and $t^* \in \mathbb{Z}_p^*$. If $\mathcal{A}$ queries the hash oracle for the hash value of index id_{j^*}, then $\mathcal{C}$ sets $h_{j^*} = (g^b)^{t^*}$, adds (id_{j^*}, t^*) to the list and returns h_{j^*} back to $\mathcal{A}$. Otherwise, $\mathcal{C}$ selects $t_j \in \mathbb{Z}_p^*$ at random, adds (id_j, t_j) to the list and returns $h_j = (g^a)^{t_j}$ back to $\mathcal{A}$.

Sign Query: $\mathcal{A}$ can adaptively query the sign oracle for the message m_j signed by i-th user's public key pk_i. Assume that $\mathcal{A}$ has queried the hash value of index id_j before, then $\mathcal{C}$ checks the list and finds the corresponding value (id_j, t_j). If $id_j = id_{j^*}$, $\mathcal{C}$ aborts. Otherwise, returns

$$\sigma_{i,j} = g^{t_j/r_i}\cdot g^{r_0\cdot m_j/r_i}$$

as the tag of the pair (id_j, m_j) under public key pk_i. It is obvious that if

$$\begin{aligned}\sigma_{i,j} =&g^{t_j/r_i}\cdot g^{r_0\cdot m_j/r_i}\\ =&(g^{a\cdot t_j})^{1/a\cdot r_i}\cdot (g^{a\cdot r_0\cdot m_j})^{1/a\cdot r_i}\\ =&(H(id_j)\cdot \omega^{m_j})^{1/sk_i},\end{aligned}$$

the tag is valid.

Challenge: $\mathcal{A}$ chooses an index set $\mathcal{I}^*$ of messages. Moreover, $\mathcal{A}$ should ensure that at least one index in set $\mathcal{I}^*$ has not been queried in the sign oracle before. Without loss of generality, we suppose that there is only one index $id_{j'}$ that has not been queried before. If $id_{j'} \neq id_{j^*}$, $\mathcal{C}$ aborts. Otherwise, $\mathcal{C}$ selects s_j, for $id_j \in \mathcal{I}^*$ and $y \in \mathbb{Z}_p^*$ at random, and computes

$$pk_{chal} = ((g^{ac})^{r_1}, \cdots, (g^{ac})^{r_d}, (g^a)^y, (g^{ac})^y).$$

Then the challenge $chal = (Q, pk_{chal})$, where

$$Q = \{(id_j, s_j) | id_j \in \mathcal{I}^*\}$$

is sent to $\mathcal{A}$. It is evident that *chal* is a valid challenge since

$$(g^{ac})^{r_i} = (g^{a \cdot r_i})^c = pk_i^c, \text{ and } (g^{ac})^y = (g^{ay})^c = h^c$$

for $c \in \mathbb{Z}_p^*$ and $h = g^{ay}$ at random.

Respond: Finally, $\mathcal{A}$ outputs the response (μ, σ_{res}).

If $\mathcal{A}$ wins the game, *i.e.* the response can successfully pass the verification, which means that (μ, σ_{res}) satisfies

$$\sigma_{res} = e\left(\prod_{(id_j, s_j) \in Q} H(id_j)^{s_j} \cdot \omega^{\mu}, g^c\right)$$

Therefore, $\mathcal{C}$ can output

$$\begin{aligned}&\left(\frac{\sigma_{res}}{e\left(\prod_{\substack{(id_j, s_j) \in Q \\ id_j \neq id_{j'}}} g^{t_j}, g^{ac}\right)^{s_j} \cdot e\left(g^{r_0 \cdot \mu}, g^{ac}\right)}\right)^{1/(s_{j'} \cdot t^*)}\\ =&e(H(id_{j'})^{s_{j'}}, g^c)^{1/(s_{i'} \cdot t^*)}\\ =&e(g, g)^{bc}\end{aligned}$$

as the solution of the vBDH problem.

It is obvious that if the simulator $\mathcal{C}$ does not abort and adversary $\mathcal{A}$ could output a valid response of the challenge, $\mathcal{C}$ can successfully output the correct solution of the vBDH problem.

Suppose that $\mathcal{A}$ is able to make **Hash Query** at most v times, and the index set $\mathcal{I}^*$ of messages contains at least one index that $\mathcal{A}$ has not queried before **Challenge** algorithm.

$\mathcal{C}$ selects $j^* \in \{1, \cdots v\}$ at random and sets $h_{j^*} = (g^b)^{t^*}$, which makes $\mathcal{A}$ unable to answer the **Sign Query** for index id_{j^*}. During **Challenge** algorithm, if $\mathcal{C}$ does not abort, then it can be shown that the index which satisfies $id_j = id_{j^*}$ has not been queried before. Since $\mathcal{C}$ could answer all the queries sent by $\mathcal{A}$ except for id_{j^*}, $\mathcal{C}$ does not abort in **Sign Query** during the case that $\mathcal{C}$ did not abort in the **Challenge** algorithm. All in all, the probability that $\mathcal{C}$ does not abort is at least

$$\Pr(\neg abort_{\mathcal{C}}) \geq 1/v.$$

Therefore, if the advantage for $\mathcal{A}$ to output a valid response is ε, $\mathcal{C}$ has at least

$$Adv_{vBDH}(\mathcal{C}) \geq \Pr(\neg abort_{\mathcal{C}}) \cdot Adv(\mathcal{A}) \geq \varepsilon / v$$

advantage solving the vBDH problem. □

6.3 Information-Theoretical Anonymity

Then, we prove that our scheme achieves information-theoretical anonymity [28].

Theorem 2. *Our scheme achieves information-theoretical anonymity,* i.e. *the advantage of any adversary $\mathcal{A}$ in distinguishing the user's identity of a pair $(message, tag)$ is negligible.*

Proof. Suppose that the adversary $\mathcal{A}$ needs to reveal the identity of user who signed the message m_j. Thus, $\mathcal{A}$ interacts with the simulator $\mathcal{C}$ to guess user's identity through the following game.

Setup: $\mathcal{C}$ runs the **Initialize** and **KeyGen** algorithms for all d users' secret and public key pairs. And then $\mathcal{C}$ sends the public parameter **params** to $\mathcal{A}$ along with all the secret and public key pairs.

Challenge: $\mathcal{A}$ chooses a pair (id_j, m_j) and computes

$$pk_{chal} = ((pk_1)^s, \cdots, (pk_d)^s, h, h^s)$$

for $s \in \mathbb{Z}_p^*$ and $h \in \mathbb{G}_1$ at random. Then, $\mathcal{A}$ sends the challenge $chal = (id_j, s_j, pk_{chal})$ to $\mathcal{C}$.

Respond: $\mathcal{C}$ checks whether

$$e(pk_i^s, h) = e(pk_i, h^s)$$

holds for $i = 1, \cdots, d$. If these equations hold, $\mathcal{C}$ randomly picks an $i \in \{1, \cdots, d\}$ and computes the tag

$$\sigma_{i,j} = (H(id_j) \cdot \omega^{m_j})^{1/sk_i}$$

using i-th user's secret key sk_i. Then, $\mathcal{C}$ generates the response (μ, σ_{res}) where

$$\mu = \sum_{(id_j, s_j) \in Q} s_j \cdot m_j,$$
$$\sigma_{res} = \prod_{(id_j, s_j) \in Q} e(\sigma_{i,j}^{s_j}, pk_i^s),$$

and returns the response (μ, σ_{res}) to $\mathcal{A}$.

Guess: $\mathcal{A}$ checks whether the response is valid, and outputs an $i' \in \{1, \cdots, d\}$.

Adversary $\mathcal{A}$ outputs the guess as $i' \in \{1, \cdots, d\}$ representing user's identity. If $\mathcal{A}$ wins the game, *i.e.* the identity is true. Assume $i \neq i'$, then we can easily generate two identical responses which stem from two tags produced by u_i and $u_{i'}$ for the same message. That is to say, the advantage for $\mathcal{A}$ to distinguish user's identity from the response is negligible. As a result, $\mathcal{A}$ can gain no more information from the response than randomly guessing the signer of the tag even if $\mathcal{A}$ holds all users' secret keys. □

6.4 Unforgeability of Tokens

When a user needs to upload a message to CS, the hash value of the message should be sent to CS to check whether a duplicate has been stored in CS or not. If there is no duplicate in CS, user needs to upload the tuple $(index, message, tag)$ and a Bloom filter. However, if there is a duplicate already stored in CS, an additional PoW protocol is needed in order to avoid the case that a malicious user only knows the message's hash value instead of the real message. Briefly speaking, the proposed scheme uses Bloom filters to match the tokens generated by user to conduct the PoW challenge.

Since the parameters in PoW protocol are independent from the main scheme and none of the available PoW schemes utilizing Bloom filter gives out a concrete security proof, security analysis is given in this part. According to a classical PoW scheme proposed by Blasco *et al.* [3], the security of our scheme focuses on the unforgeability of message's tokens.

The PoW challenge requires user to produce tokens for t message blocks at randomly chosen positions. Once received by CS, the tokens are processed with pseudorandom function and checked for membership in the corresponding Bloom filter.

Suppose a malicious user wants to gain access to message m_j. If the malicious user possesses several message blocks and attempts to pass the PoW challenge, he needs to generate all tokens of the t message blocks challenged by CS. Considered by Blasco *et al.* [3], if the malicious user possesses only a few message blocks, the parameters of Bloom filter are set to a proper range, and t is large enough, then the probability that the malicious user successfully forges message blocks' tokens and thus passes the PoW challenge is negligible. We should be aware that a user can pass the PoW protocol when he possesses almost all m_j's message blocks. However, this user can be regarded as a legitimate user for m_j since he almost possess the message.

7 Efficiency Evaluation

In this section, we evaluate the performance of our mechanism and provide a comparison among several schemes [1, 13, 14, 26].

7.1 Communication Overhead

According to the description in Section 4, our mechanism does not introduce communication overhead to users during **Initialize**, **KeyGen** and **Verify** algorithms. The size of the Bloom filter BF_i is $t \cdot |q|$ bits. The size of an auditing message (Q, pk_{chal}) is $k \cdot (|I| + |q|) + p \cdot |q|$ bits. The size of an auditing proof (μ, σ_{res}) is $2t \cdot |q|$ bits. Therefore, if the message is a new one for CS, the total communication overhead of an auditing task is $(k+p+2t+1)|q|+k|I|$ bits, otherwise the total communication overhead is $(t+k+p+2t+1)|q|+k|I|+|m|+t$ bits. It is obvious that if there is a duplicate in CS, the communication overhead is smaller than the case that there is no duplicate. Table 2 provides a comparison between some existing schemes [1, 13, 14, 26] about the communication cost.

7.2 Computation Overhead

As shown in the **FileUpload** algorithm of the proposed scheme, user generates a Bloom filter by splitting message into n blocks, and then computes the authentication tag on the file-level. Under condition that CS receives a new message, the computation cost of uploading a message is $2exp + (n + 4)hash + 2mul + 2pair + n \cdot prf$. Otherwise, the corresponding computation cost is $exp+(t+1)hash+t \cdot prf$. Moreover, the computation cost of auditing phase is $(n + 3k + 1)exp + k \cdot hash + 3k \cdot mul + (2n + k)pair$. Therefore, the total computation cost of our mechanism is $(n+3k+3)exp+(n+k+4)hash+(3k+2)mul+(2n+k+2)pair+n \cdot prf$ if the message is a new one for CS. Otherwise, the total computation cost is $3exp + (n + t + 5)hash+2mul+2pair+(n+t)prf$. Evaluating the existing schemes [1, 13, 14, 26], we provide a detail comparison in Table 3 along with some notations in Table 4.

As shown in Table 3, since our scheme can verify the integrity of several messages instead of one message during one challenge-and-response protocol, the efficiency of the proposed scheme is a little lower than Kardas and Kiraz's scheme [13] in **Challenge** and **Respond** algorithms.

However, the proposed scheme has advantage on efficiency during **KeyGen** and **FileUpload** due to the use of asymmetric encryption and signature in Kardas and Kiraz's scheme [13].

Furthermore, Alkhojandi and Miri's scheme [1] and our proposed scheme have almost the same efficiency, while the former has two security problem as discussed in Section 1. So, all these above indicate that the proposed scheme achieves a better trade-off for efficient and secure than the existing schemes.

By utilizing the Pairing Based Cryptography (PBC) Library, an efficiency experiment result is given under the Linux environment. The following experiments run on a personal computer with its configuration parameters as Intel Core i5 2.5 GHz Processor and 4 GB RAM. We assume that the size of element in $\mathbb{G}_1$ and $\mathbb{Z}_p$ is 160 bits, the size of one message block is 2 KB, the size of an element in set $\mathcal{I}$ is 20 bits. The experiment result given below comes from the average of 50 experiments.

Figures 2 and 3 show the communication time changes when k ranges from 100 to 300. Meanwhile, Figure 4 shows the computation time with t ranges from 50 to 250. Figures 2 and 3 indicate that with the increasing number of auditing message blocks, communication in Yuan and Yu's scheme [26], Li *et al.*'s scheme [14] and our scheme has an upward trend, which indicates these three schemes apply to smaller amount of data. However, our scheme has a higher efficiency than Yuan and Yu's scheme [26] and Li *et al.*'s scheme [14]. Though Alkhojandi and Miri's

Table 2: Communication overhead comparison

Scheme	There is a duplicate in CS	The message is a new one for CS
Yuan *et al.* [26]	$(k+t)\lvert I\rvert+(k+4)\lvert q\rvert+(\frac{t}{n}+1)\lvert m\rvert$	$(k+t)\lvert I\rvert+(k+4)\lvert q\rvert+(\frac{t}{n}+1)\lvert m\rvert$
Naelah *et al.* [1]	$k\lvert I\rvert+(n+t+6)\lvert q\rvert+(\frac{t}{n}+1)\lvert m\rvert+k+t$	$k\lvert I\rvert+(n+6)\lvert q\rvert+2\cdot\lvert m\rvert+k$
Li *et al.* [14]	$k\lvert I\rvert+(n+2k+t+2)\lvert q\rvert+\lvert m\rvert+k+t$	$k\lvert I\rvert+(3n+2k+3)\lvert q\rvert+\lvert m\rvert+k$
Kardas *et al.* [13]	$6\lvert q\rvert+7\lvert m\rvert+t$	$k\lvert I\rvert+(n+4)\lvert q\rvert+(n+2)\lvert m\rvert+n+k+t$
Ours	$k\lvert I\rvert+(k+p+3t+1)\lvert q\rvert+t$	$k\lvert I\rvert+(k+p+2t+1)\lvert q\rvert+\lvert m\rvert$

Table 3: Computation overhead comparison

Schemes	KeyGen	FileUpload	Challenge	Respond	Verify
Yuan *et al.* [26]	$(n+2)exp$	$(t+1)hash+$ $(2n+t+3)exp+$ $(2t+1)mul+4pair$	-	$k\cdot(n+1)mul+$ $(k+1)exp$	$khash+(k+1)$ $exp+4pair$
Naelah *et al.* [1]	$exp+pair$	$2n\cdot hash+2n\cdot$ $exp+t\cdot mul$	-	$hash+3k\cdot exp+$ $(3k+1)mul+$ $k\cdot pair$	$kt\cdot hash+$ $(kt+2t+1)exp+$ $(t+1)pair$
Li *et al.* [14]	exp	$(n+1)hash+$ $(nt+2)exp+$ $n\cdot(t+3)mul$	-	$(k+1)mul+kexp$	$(k+1)hash+$ $2pair+exp$
Kardas et al. [13]	$2hash+prf$	$(n+1)AsymEnc+$ $n\cdot hash+n\cdot$ $exp+n\cdot mul$	-	$k\cdot exp+2k\cdot$ $mul+pair+hash$	$2pair+(k+4)exp$
Ours	exp	$t\cdot hash+2mul+$ $2pair+exp$	$(n+1)exp$	$(2n+k)pair+$ $2k\cdot mul+k\cdot exp$	$k\cdot hash+2k\cdot exp$ $+k\cdot mul+pair$

Table 4: Notations

Notation	Significance
exp	One exponentiation operation
$hash$	One hashing operation
mul	One multiplication operation
$pair$	One pairing operation on $e:\mathbb{G}_1\times\mathbb{G}_1\rightarrow\mathbb{G}_2$
prf	One pseudorandom function operation
$AsymEnc$	An asymmetric encryption or signature
$\lvert I\rvert$	The size of an element of set $\mathcal{I}$
$\lvert q\rvert$	The size of an element of $\mathbb{Z}_p$ or $\mathbb{G}_1$
$\lvert m\rvert$	The size of the message m
p	The number of users within the group
k	The number of selected blocks during challenge
n	The number of blocks in one message m_i
t	The number of blocks CS choose to challenge users during PoW

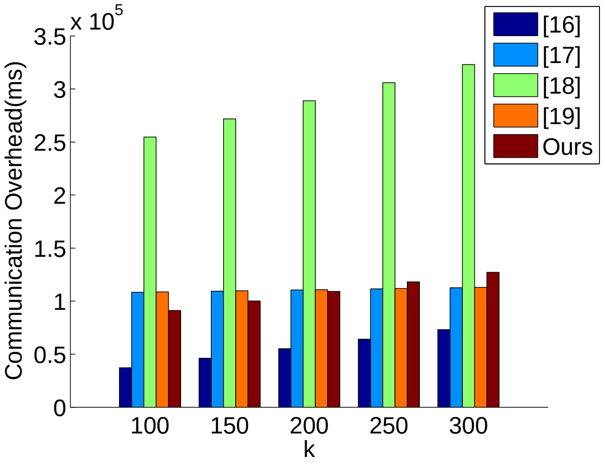


Figure 2: Communication overhead when there is no duplicate

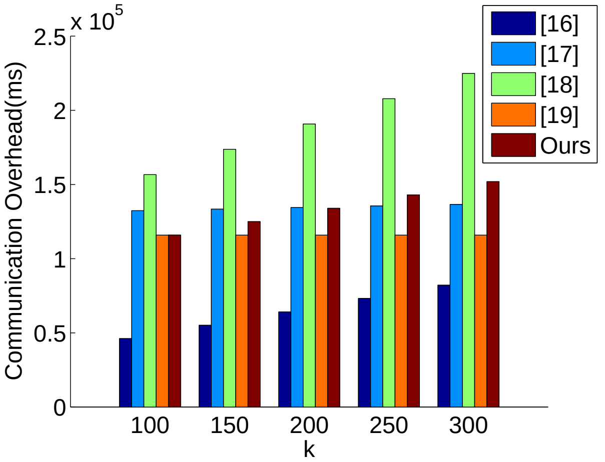


Figure 3: Communication overhead when there is a duplicate

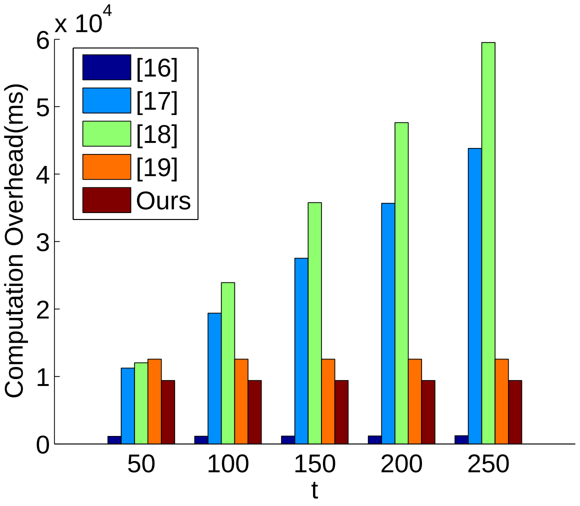


Figure 4: Computation overhead

scheme [1] and Kardas and Kiraz's scheme [13] both have the advantage of communication time compared with our scheme, Figure 4 indicates that the computation time of their schemes are higher than ours. Thus, the experimental results are in great agreement with the above theoretical analysis.

8 Conclusions

In this paper, we have proposed a privacy-preserving public auditing system with data deduplication, which provides data integrity and storage efficiency in cloud computing. Traditional privacy-preserving auditing schemes apply ring signature or group signature to achieve anonymity. And this kind of technology inevitably causes the tag size significantly large. Thus, we use another way to guarantee users' identity privacy against the TPA in order to reduce the tag size. In the proposed scheme, the tag generation algorithm reduces the tag size to only one element. On the other hand, our scheme uses Bloom filter to perform PoW protocol during deduplication, which can be more efficient than some state of the art solutions. Efficiency analysis indicates that the proposed scheme achieves a better trade-off between efficiency and function compared with existing schemes with similar features.

Acknowledgments

We are grateful to the anonymous reviewers for their invaluable suggestions. This work is supported by the National Key R&D Program of China under Grant No. 2017YFB0802000, the National Natural Science Foundation of China under Grants No.61472470 and 61572390, and the Scientific Research Plan Project of Education Department of Shaanxi Province under Grant No.17JK0362.

References

[1] N. Alkhojandi, and A. Miri, "Privacy-preserving public auditing in cloud computing with data deduplication," *International Symposium on Foundations and Practice of Security (FPS'14)*, pp. 35–48, Nov. 2014.

[2] 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 (CCS'07)*, pp. 598–609, Oct. 2007.

[3] J. Blasco, R. D. Pietro, A. Orfila, and A. Sorniotti, "A tunable proof of ownership scheme for deduplication using Bloom filters," in *IEEE Conference on Communications and Network Security (CNS'14)*, pp. 481–489, Oct. 2014.

[4] B. H. Bloom, "Space/time trade-offs in hash coding with allowable errors," *Communications of the ACM*, vol. 13, no. 7, pp. 422–426, 1970.

[5] D. Boneh, and M. Franklin, "Identity-based encryption from the Weil pairing," *Annual International Cryptology Conference (CRYPTO'01)*, pp. 213–229, Aug. 2001.

[6] D. Boneh, B. Lynn, and H. Shacham, "Short signatures from the Weil pairing," *International Conference on the Theory and Application of Cryptology and Information Security (ASIACRYPT'01)*, pp. 514–532, Dec. 2001.

[7] Z. Cao, L. Liu, O. Markowitch, "Analysis of one scheme for enabling cloud storage auditing with verifiable outsourcing of key updates," *International Journal of Network Security*, vol. 19, no. 6, pp. 950–954, 2017.

[8] P. S. Chung, C. W. Liu, and M. S. Hwang, "A study of attribute-based proxy re-encryption scheme in cloud environments", *International Journal of Network Security*, vol. 16, no. 1, pp. 1-13, 2014.

[9] S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg, "Proofs of ownership in remote storage systems," in *Proceedings of the 18th ACM Conference on Computer and Communications Security (CCS'11)*, pp. 491–500, Oct. 2011.

[10] D. Harnik, B. Pinkas, and A. Shulman-Peleg, "Side channels in cloud services: Deduplication in cloud storage," *IEEE Symposium on Security and Privacy (S&P'10)*, vol. 8, no. 6, pp. 40–47, May 2010.

[11] W. F. Hsien, C. C. Yang, and M. S. Hwang, "A survey of public auditing for secure data storage in cloud computing," *International Journal of Network Security*, vol. 18, no. 1, pp. 133–142, 2016.

[12] M. S. Hwang, T. H. Sun, C. C. Lee, "Achieving dynamic data guarantee and data confidentiality of public auditing in cloud storage service," *Journal of Circuits, Systems, and Computers*, vol. 26, no. 5, 2017.

[13] S. Kardas, and M. S. Kiraz, "Solving the secure storage dilemma: An efficient scheme for secure deduplication with privacy-preserving public auditing," *Cryptology ePrint Archive*, Report 2016/696, 2016.

[14] J. Li, J. Li, D. Xie, and Z. Cai, "Secure auditing and deduplicating data in cloud," *IEEE Transactions on Computers*, vol. 65, no. 8, pp. 2386–2396, 2016.

[15] L. Liu, Z. Cao, C. Mao, "A note on one outsourcing scheme for big data access control in cloud," *International Journal of Electronics and Information Engineering*, vol. 9, no. 1, pp. 29–35, 2018.

[16] C. W. Liu, W. F. Hsien, C. C. Yang, and M. S. Hwang, "A survey of public auditing for shared data storage with user revocation in cloud computing," *International Journal of Network Security*, vol. 18, no. 4, pp. 650–666, 2016.

[17] Y. Ming, Y. Wang, "On the security of three public auditing schemes in cloud computing," *International Journal of Network Security*, vol. 17, no. 6, pp. 795–802, 2015.

[18] W. K. Ng, Y. Wen, and H. Zhu, "Private data deduplication protocols in cloud storage," in *Proceedings of the 27th Annual ACM Symposium on Applied Computing (SAC'12)*, pp. 441–446, Mar. 2012.

[19] B. Wang, B. Li, and H. Li, "Oruta: Privacy-preserving public auditing for shared data in the cloud," *IEEE Transactions on Cloud Computing*, vol. 2, no. 1, pp. 43–56, 2014.

[20] B. Wang, B. Li, and H. Li, "Knox: Privacy-preserving auditing for shared data with large groups in the cloud," in *International Conference on Applied Cryptography and Network Security (ACNS'12)*, pp. 507–525, June 2012.

[21] B. Wang, H. Li, and M. Li, "Privacy-preserving public auditing for shared cloud data supporting group dynamics," in *IEEE International Conference on Communications (ICC'13)*, pp. 1946–1950, June 2013.

[22] Z. Wang, Y. Lu, and G. Sun, "A policy-based deduplication mechanism for securing cloud storage," *International Journal of Electronics and Information Engineering*, vol. 2, no. 2, pp. 70–79, 2015.

[23] G. Wu, Y. Mu, W. Susilo, and F. Guo, "Privacy-preserving cloud auditing with multiple uploaders," in *International Conference on Information Security Practice and Experience*, pp. 224–237, 2016.

[24] J. Xu, E. C. Chang, and J. Zhou, "Weak leakage-resilient client-side deduplication of encrypted data in cloud storage," in *Proceedings of 8th ACM SIGSAC Symposium on Information, Computer and Communications Security*, pp. 195–206, 2013.

[25] Y. Yu, Y. Mu, J. Ni, J. Deng, and K. Huang, "Identity privacy-preserving public auditing with dynamic group for secure mobile cloud storage," in *International Conference on Network and System Security (NSS'14)*, pp. 28–40, Oct. 2014.

[26] J. Yuan, and S. Yu, "Secure and constant cost public cloud storage auditing with deduplication," in *IEEE Conference on Communications and Network Security (CNS'13)*, pp. 145–153, May 2013.

[27] J. Zhang, P. Li, and M. Xu, "On the security of an mutual verifiable provable data auditing in public cloud storage," *International Journal of Network Security*, vol. 19, no. 4, pp. 605–612, 2017.

[28] J. Zhang, and X. Zhao, "Efficient chameleon hashing-based privacy-preserving auditing in cloud storage," *Cluster Computing*, vol. 19, no. 1, pp. 47–56, 2016.

Biography

Chen Li is a master degree student in the School of Mathematics and Statistics at Xidian University. Her interest focuses on cryptography and network security.

Zhenhua Liu is a professor in the School of Mathematics and Statistics at Xidian University, Xi'an, China. His research interests include public key cryptography, cryptographic theory and security protocols in cloud computing.

A Cloud-assisted Passenger Authentication Scheme for Japan Rail Pass Based on Image Morphing

Yanjun Liu and Chin-Chen Chang

(Corresponding author: Chin-Chen Chang)

Department of Information Engineering and Computer Science, Feng Chia University
No. 100, Wenhua Road, Xitun District, Taichung 40724, Taiwan
(Email: alan3c@gmail.com)

(Received July 27, 2017; revised and accepted Oct. 21, 2017)

Abstract

The Japan rail (JR) pass is the most economical means to travel around Japan through public transportations. Therefore, the passenger authentication for JR pass has become a very typical and popular smart-card based application. However, a traditional passenger authentication scheme for the JR pass has two major problems: 1) The passenger must present his/her passport to the attendant each time for the authentication, increasing the probability of losing the passport; 2) The passenger's passport number, which is printed on the JR pass for authentication, may reveal the passenger's personal information. To overcome these security and privacy weaknesses, we innovatively propose a cloud-assisted passenger authentication scheme for the JR pass based on image morphing technique. Analyses show that the method of our proposed scheme is quite simple to implement and can resist well-known attacks.

Keywords: Cloud-assisted; Image Morphing; Japan Rail (JR) Pass; Passenger Authentication; Security

1 Introduction

Nowadays, user authentication is used extensively in many applications in the field of information security. The remote user authentication mechanism was originally proposed by Lamport [8] in 1981. There are two parties in the authentication mechanism, *i.e.*, a user and a server. The user is allowed to set a password and transmits it to the server secretly for registration. When the user wants to gain some services from the server, he/she must be requested to provide the password to the server to authenticate his/her identity. Unfortunately, the server must maintain a verifier table [8] that preserves some personal information of the users. When the number of users increases drastically, it requires huge storage space for the verifier table. Moreover, the use of the verifier table risks a severe security problem such that the theft of the verifier table can induce a leak of users' private information. Given that a smart card can enhance security as well as efficiency, many authentication schemes [7,9,10] based on smart card have been proposed.

In a smart card-based authentication scheme, each user is assigned a smart card rather than central storage for confidential information as used in conventional schemes. Each smart card stores some holders' personal information, thereby significantly diminishing the risk of information leak and increasing the storage efficiency [10]. The smart card can also preserve additional information used for tamper-proof to increase security.

The smart card-based authentication consists of a card authentication process and a user authentication process [9]. After the smart card is registered to the server, the registered information is stored in the smart card and then transmitted to the terminal to judge the legality of the card. It is considered more secure since the card authentication process can be achieved only between the smart card and the terminal rather than involving the server. However, even if the card authentication is passed, it cannot ensure that the card holder is the true user. Thus, the user should be authenticated after the card authentication. Biometrics information [6,11] such as facial features, fingerprint and iris of the user is usually requested to provide for the user authentication. Because biometrics information of two different people cannot be identical, the illegal user can be immediately identified if he/she cannot offer the biometrics information unique to the true user's body. For convenience, the biometrics information of the true user is often stored or printed on the smart card. When the card is used, a human operator performs a face-to-face authentication of the card holder by comparing the biometrics information offered by the card holder with that of the true user. If the biometrics information are the same, the human operator can be convinced that the card holder is legal.

The concept of smart card-based authentication can be used in many applications in our real life thanks to its advantages in high security and low cost. The authentication for Japan rail (JR) pass [1,2] is a very typical and popular application among these applications. Therefore, inspired by the card user authentication, we will propose a novel method to authenticate users for the JR pass in this paper. The JR pass [1], provided by the Japan Railways Group, is the most economical means to travel around Japan through public transportations such as trains, buses and ferries. However, the JR pass can only be used by overseas tourists for sight-seeing and Japanese nationals living outside of Japan who meet some particular requirements. Only eligible passengers can purchase 7, 14 or 21 day length JR pass to plan their trip. In fact, a JR pass can be regarded as a smart card. Therefore, how to efficiently authenticate the eligible passenger when the JR pass is used becomes a crucial issue.

Now let us take a look at the way that a traditional passenger authentication scheme for the JR pass conducts [2]. An eligible passenger can purchase a JR pass at a sales office at a JR station or a travel agency by presenting his/her passport to the attendant. Then, the passenger is issued with a JR pass on which the passport number is printed. Before every boarding, the validity of the passenger should be authenticated by an attendant at a manned ticket gate at a JR station. For the authentication, the passenger is requested to present his/her JR pass and passport together to the attendant. The attendant then confirms that the passenger is the owner of the passport and the passport number on the JR pass is the same as that on the passport. The passenger is permitted to board only if the authentication is passed.

Although the above mentioned passenger authentication is simple to implement, it leads to two security problems. First, the passenger must present his/her passport to the attendant each time for the authentication, which means that the passenger needs to carry the passport all the time so that the passport is easy to get lost. Second, the passport number should be printed on the JR pass to perform the authentication. Unfortunately, the disclosure of the passenger's personal information (*i.e.*, passport number) may happen if the JR pass is stolen, lost or discarded when it expires. To overcome these weaknesses, we will innovatively propose a more secure passenger authentication scheme for the JR pass based on image morphing. Image morphing [3] is a technique which is originally used to produce special visual effects in movie industry. It creates a morphed image by using two images, one called source image and the other called target image. The created morphed image looks like both the source image and the target image if they have similar structures. This impressive feature makes image morphing a newly widespread approach in the area of authentication [12–16]. In this paper, a cloud-assisted passenger authentication scheme for the JR pass using image morphing is proposed to increase the security efficiently. To the best of our knowledge, the proposed scheme is the first authentication scheme for the JR pass that combines image morphing and cloud storage techniques. The unique characteristics of the proposed scheme are listed below:

1) To effectively protect the passenger's privacy, the passenger's passport number is no longer printed on the JR pass; instead, facial features of the passenger are employed for authentication. In particular, a morphed image is generated by the face image of the passenger and a pre-selected reference image through morphing, thereby hiding the face image of the passenger into the morphed image.

2) The generated morphed image is stored on the cloud storage, which efficiently avoids the disclosure of the passenger's personal information.

3) The authentication for the JR passenger is quite simple. When the JR pass is used, only an image demorphing process is performed to verify the validity of the passenger. An attendant at a manned ticket gate uses the morphed image stored on the cloud storage and the reference image to restore the face image of the passenger through de-morphing. This method can significantly increase both security and convenience since the passenger's passport never needs to be used during the authentication.

The rest of the paper is organized as follows. In Section 2, we offer an overview of image morphing and some user authentication schemes based on image morphing. Section 3 proposes a cloud-assisted passenger authentication scheme for the JR pass using image morphing. Section 4 analyzes the correctness and security of the proposed scheme. Ultimately, our conclusions are given in Section 5.

2 Preliminaries

In this section, we introduce some background knowledge before presenting the proposed scheme. We first give the basic knowledge of image morphing, and then describe some related authenticated scheme using image morphing.

2.1 Image Morphing and De-Morphing

Image morphing technique [3] uses a source image and a target image to create a morphed image that looks like both the source image and the target image if they have similar structures. Image de-morphing, as its name implies, is the reverse of image morphing that restores the source image (or the target image) from the morphed image and the target image (or the source image). In the following, we will briefly introduce how image morphing and de-morphing work, respectively.

The source image and the target image are denoted as I_s and I_t with the size of $N_1 \times N_2$, respectively. Now we present how to generate the morphed image I_{st}^m using I_s and I_t.

2.1.1 Image Morphing Process [3]

Step 1: Choose $n (n \in N)$ control pixel pairs. One pixel chosen from I_s and its corresponding pixel chosen from I_t constitute a control pixel pair. Assume that (x_i^s, y_i^s) and (x_i^t, y_i^t) for $i = 1, 2, \ldots, n$ denote the coordinates of i^{th} control pixel in I_s and I_t, respectively. The coordinates of all the selected control pixels in I_s and I_t, represented as matrices C_s and C_t, are shown as follows:

$$C_s = \begin{bmatrix} x_1^s & x_2^s & \ldots & x_n^s \\ y_1^s & y_2^s & \ldots & y_n^s \end{bmatrix} \quad (1)$$

$$C_t = \begin{bmatrix} x_1^t & x_2^t & \ldots & x_n^t \\ y_1^t & y_2^t & \ldots & y_n^t \end{bmatrix} \quad (2)$$

Step 2: Calculate horizontal and vertical distances of control pixel pairs in I_s and I_t. Horizontal-distance vector D_1 and vertical-distance vector D_2 are computed as follows:

$$D_1 = \begin{bmatrix} x_1^s - x_1^t & x_2^s - x_2^t & \cdots & x_n^s - x_n^t \end{bmatrix} \quad (3)$$

$$D_2 = \begin{bmatrix} y_1^s - y_1^t & y_2^s - y_2^t & \cdots & y_n^s - y_n^t \end{bmatrix} \quad (4)$$

Step 3: Calculate horizontal and vertical distances of all corresponding pixel pairs in I_s and I_t. In order to retrieve the distances of all corresponding pixel pairs from n control pixel pairs, linear interpolations are made on D_1 and D_2, and then interpolation matrices B_1 and B_2 with the size of $N_1 \times N_2$ are generated. Obviously, the component in matrix B_1, denoted as $b_1(x, y)$, represents the horizontal distance between the pixel (x, y) in I_s and its corresponding pixel in I_t. Accordingly, the component in matrix B_2, denoted as $b_2(x, y)$, represents the vertical distance between the pixel (x, y) in I_s and its corresponding pixel in I_t. For more detailed information about the implementation of interpolation, interested readers can refer to Ref. [14].

Step 4: Warp the source image and the target image. Assume that $\alpha (0 \leq \alpha \leq 1)$ is the morphing rate. To warp the source image I_s, the pixel (x, y) in I_s is shifted by $[\alpha b_1(x, y)]$ in the horizontal direction and by $[\alpha b_2(x, y)]$ in the vertical direction. The following equation accurately describes the warping process for the pixel (x, y) in I_s:

$$p_s^w(x, y) = p_s(x + [\alpha b_1(x, y)], y + [\alpha b_2(x, y)]), \quad (5)$$

where $p_s(x, y)$ and $p_s^w(x, y)$ are the gray values of pixel (x, y) in I_s before and after warping, and $[h]$ is the rounding of h. Then, the warped source image I_s^w is generated when the shift of all the pixels in I_s completes.

The target image I_t is warped by a similar means. To warp the target image I_t, the pixel (x, y) in I_t is shifted by $[(1-\alpha)b_1(x, y)]$ in the horizontal direction and by $[(1-\alpha)b_2(x, y)]$ in the vertical direction. The warping process for the pixel (x, y) in I_t is defined as follows:

$$\begin{aligned} p_t^w(x, y) = p_t(x + [(1-\alpha)b_1(x, y)], \\ y + [(1-\alpha)b_2(x, y)]), \end{aligned} \quad (6)$$

where $p_t(x, y)$ and $p_t^w(x, y)$ are the gray values of pixel (x, y) in I_t before and after warping. After that, the warped target image I_t^w is generated.

Step 5: Generate the morphed image. The morphed image I_{st}^m is created by using the warped source image I_s^w, the warped target image I_t^w and an appropriate morphing rate α as follows:

$$I_{st}^m = (1-\alpha)I_s^w + \alpha I_t^w. \quad (7)$$

Image de-morphing is the reverse of image morphing that restores the source image (or the target image) from the morphed image and the target image (or the source image). In the following, we take the reconstruction of the source image as an example to explain the process of de-morphing. It is noticed that the de-morphing operation is on the assumption that we have already known the coordinates matrix C_s, the target image I_t, the coordinates matrix C_t, the morphed image I_{st}^m and the morphing rate α. First, horizontal-distance vector D_1 and vertical-distance vector D_2 for control pixel pairs from the source image I_s and the target image I_t are computed by Equations (3) and (4). Then, interpolation matrices B_1 and B_2 are obtained after linear interpolations are made on D_1 and D_2. After that, the warped target image I_t^w is obtained by Equation (6). According to Equation (7), the warped source image I_s^w can be computed as:

$$I_s^w = \frac{I_{st}^m - \alpha I_t^w}{1-\alpha} \quad (8)$$

Finally, every pixel in I_s^w can easily return to the original location in I_s by the following operation:

$$p_s(x, y) = p_s^w(x - [\alpha b_1(x, y)], y - [\alpha b_2(x, y)]), \quad (9)$$

and then the restored source image $I_{st \to s}^{de}$ is generated immediately.

An example of the processes of image morphing and de-morphing is shown in Figure 1. The source image and the target image, selected from Yale face database [5], are shown in Figures 1 (a) and 1 (b), respectively. The Morphed image created by Figures 1 (a) and 1 (b) with the morphing rate $\alpha = 0.5$ is illustrated in Figure 1 (c). Figure 1 (d) illustrates the restored source images by de-morphing.

2.2 Related Work

In recent years, image morphing is used extensively in user authentication schemes [12–16]. Zhao and Hsieh [16] proposed a card user authentication scheme using the user's

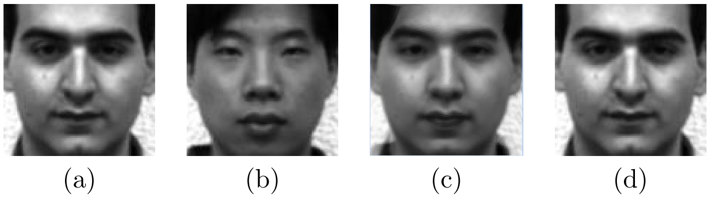


Figure 1: Image morphing and de-morphing: (a) source image; (b) target image; (c) morphed image ($\alpha = 0.5$); (d) recovered source image ($\alpha = 0.5$)

face image as important information for authentication. In their scheme, a morphed image is created by the face image of the card user and a reference image and then is printed on the card. Thus, the face image of the card user is actually concealed into the morphed image. When the card is used, a human operator should authenticate the identity of the card holder. The operator uses the morphed image and the reference image to recover a face image through de-morphing. If the recovered image is identical to the face image of the card user, it indicates that the card holder is legal and he/she can use the card to obtain required services. In addition, the authors extended the image morphing to the case when $l(l > 1)$ reference images are used and called it generalized image morphing. A user authentication scheme using generalized image morphing was also proposed and the security was analyzed thoroughly.

One of the most crucial issues in image morphing is to increase the visual effect of the morphed image [12, 15]. Thus, many algorithms focusing on the optimization of the selection of control pixel pairs from the source image and target image are proposed to achieve this goal. Also, these improved image morphing techniques are applied to the design of authentication schemes. In 2013, Mao *et al.* [12] proposed an edge directed automatic control pixel selection algorithm for better edge detection during morphing instead of manual selection. The experimental results showed that the new algorithm can increase both accuracy and efficiency of morphing. Zhao *et al.* [15] used an interactive genetic algorithm (IGA) to build an appropriate feature point set (FPS) to make more natural-looking morphed images. Thus, their previously proposed card user authentication scheme [16] can be enhanced by using the improved morphing method. Later, in 2015, Mao *et al.* [13, 14] innovatively proposed two authentication schemes based on image morphing. In [13], Mao *et al.* first presented a source-based image morphing (SBIM) algorithm that selects control pixels only in the source image. Accordingly, the de-morphing employs the coordinates of control pixels only in the source image to recover the original source image. Furthermore, a novel proxy user authentication (PUA) authentication scheme based on SBIM was proposed. The scheme can authenticate both a primary user and a proxy user who acts as a deputy of the primary user via image exchange. The key agreement scheme proposed in [14] is also implemented by image exchange. A communication user uses a pre-assigned secret image as the source image and another selected image as the target image to generate a morphed image, and then sends it to the receiver. The receiver reconstructs the target image from the morphed image and the source image by de-morphing. A secret session key can be established if the reconstructed target image is the same as the original target image. Moreover, the key agreement scheme can resist both active and passive attacks.

3 Proposed Scheme for JR Pass

In this section, we propose a novel passenger authentication scheme for the JR pass, which innovatively adopts image morphing and cloud storage techniques to significantly enhance the security compared to traditional JR pass authentication mechanisms. Our proposed scheme involves four entities, *i.e.*, a passenger, an attendant with a terminal at a sales office, an attendant with a terminal at a manned ticket gate and a set of cloud storage servers. An eligible passenger can purchase a JR pass with his/her passport from an attendant at a sales office. Only an assigned JR pass number appears on the pass for privacy protection and is used for authentication. Meanwhile, a morphed image that hides the face image of the passenger is generated and then stored on a cloud storage server. When the passenger wants to board a vehicle, an attendant at a manned ticket gate must authenticate the validity of the passenger. According to the number printed on the JR pass, the attendant retrieves the morphed image corresponding to the passenger from the cloud storage servers, and then recovers the face image of the passenger by de-morphing the morphed image. The passenger is permitted to board only if the authentication is passed.

Our proposed scheme mainly consists of the pass buying phase and the passenger authentication phase. The following subsections elaborate both phases of the proposed scheme.

3.1 Definition of Notations

Before the detailed description of both phases, the definition of the notations used in the proposed scheme is given in Table 1.

Table 1: Notations used in the proposed scheme

Notation	Definition
I_j	Original image j
I_j^w	Warped image j
I_{jk}^m	Morphed image using I_j as the source image and I_k as the target image
$I_{jk\to j}^{de}$	Restored source image j from I_{jk}^m via de-morphing
C_j	Coordinates of the control pixels of I_j
$\alpha_j (0 \leq \alpha_j \leq 1)$	Morphing rate corresponding to I_j
M	Morphing function
DM	De-morphing function
PG_j	Passenger j whose face image is I_j
PID_j	Number of the JR pass held by PG_j
IND_j	Index of PID_j in the set of all sorted JR pass numbers
ADS	Attendant at a sales office
TS	Terminal at a sales office
ADG	Attendant at a manned ticket gate
TG	Terminal at a manned ticket gate
CSS	Cloud storage servers
L_{jk}^m	Location on CSS where I_{jk}^m is stored

3.2 Pass Buying Phase

In this phase, an eligible passenger purchases a JR pass from an attendant at a sales office. Different from the traditional JR pass, the passenger's passport number is no longer printed on it so that the disclosure of private information can be avoided efficiently. Instead, only an assigned, unique number appears on each JR pass for sale. Moreover, all the JR pass numbers are sorted by value in ascending order and the index of each number in the sorted set is recorded. On the other hand, a morphed image that conceals the face image of the passenger is generated and then stored on a cloud storage server. The JR pass number and the morphed image are two essential elements for later authentication. This phase is described in detail as follows and demonstrated in Figure 2.

Step 1: Passenger PG_j presents his/her passport to an ADS.

Step 2: ADS verifies the submitted passport to confirm that PG_j is eligible.

Step 3: If PG_j is an eligible passenger, ADS takes a digital photograph I_j of PG_j as the face image of PG_j. Image I_j contains PG_j's distinct facial features and is used as the source image for morphing.

Step 4: ADS uses an authorized terminal TS to generate a morphed image. First, TS retrieves a preselected reference face image I_k as the target image for morphing. It should be noticed that the same target image I_k is used for face images (as source images) of different passengers. Then, TS selects coordinates of the control pixels C_j of I_j, coordinates of the control pixels C_k of I_k and the morphing rate α_j. Finally, TS generates a morphed image I_{jk}^m using I_j and I_k as

$$I_{jk}^m = M(I_j, I_k, C_j, C_k, \alpha_j). \tag{10}$$

Step 5: TS selects a JR pass with a printed number PID_j on it for PG_j. TS obtains the index IND_j of PID_j in the set of all sorted JR pass numbers. Then, TS sends IND_j and I_{jk}^m to cloud storage servers CSS.

Step 6: CSS chooses the IND_j^{th} vacant location L_{jk}^m in a specified space on itself to store I_{jk}^m. From this step, we can infer that the location of I_{jk}^m on CSS depends on the value of JR pass number PID_j. That is, a smaller PID_j leads to a smaller IND_j, thus a lower location of I_{jk}^m on CSS.

Step 7: TS stores C_j and α_j on the JR pass and then issues the pass to PG_j.

3.3 Passenger authentication phase

After the pass buying phase is completed, the passenger obtains a JR pass. When a JR pass holder wants to board a vehicle, an attendant at a manned ticket gate must authenticate that the holder is the true user of the JR pass. A morphed image is retrieved from CSS according to the number printed on the JR pass, and then it is used to restore a face image via de-morphing. If the restored face image is the same as that of the JR pass holder, it can be convinced that the holder is a legal passenger. The passenger authentication phase is described as follows.

Step 1: Passenger PG_j presents the JR pass that he/she holds to an ADG.

Step 2: ADG inserts the JR pass into an authorized terminal TG and TG verifies that the pass is legal. After that, TG extracts PID_j, C_j and α_j from the pass.

Step 3: TG obtains the index IND_j of PID_j and then sends IND_j to CSS.

Step 4: CSS uses IND_j to determine the location L_{jk}^m where the morphed image I_{jk}^m is stored, and then retrieves I_{jk}^m.

Step 5: CSS sends I_{jk}^m to TG.

Step 6: TG restores a face image via de-morphing. First, TG retrieves the reference image I_k that is used as the target image in previous morphing. Then, TG selects coordinates of the control pixels C_k of I_k.

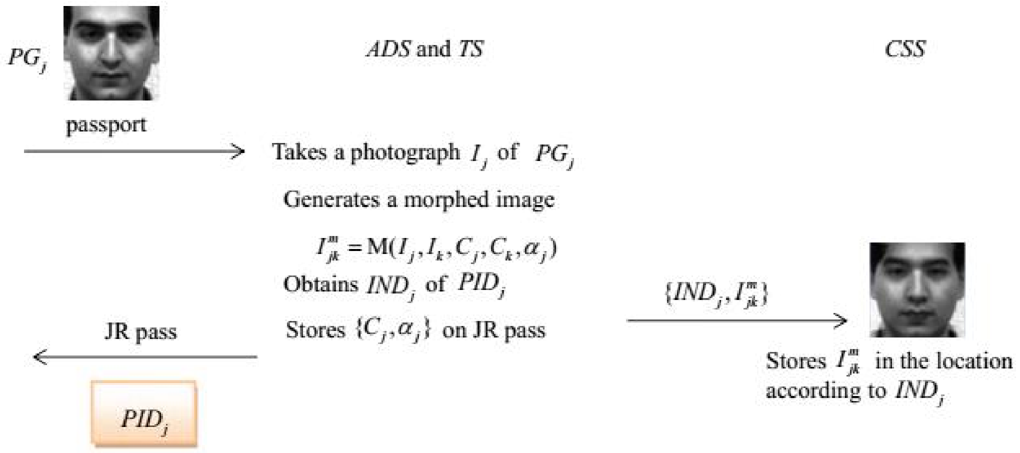

PG_j
passport
ADS and TS
CSS
Takes a photograph I_j of PG_j
Generates a morphed image
$I_{jk}^m = \mathrm{M}(I_j, I_k, C_j, C_k, \alpha_j)$
Obtains IND_j of PID_j
Stores $\{C_j, \alpha_j\}$ on JR pass
JR pass
PID_j
$\{IND_j, I_{jk}^m\}$
Stores I_{jk}^m in the location according to IND_j

Figure 2: Pass buying phase of our proposed scheme

Finally, TS restores a face image $I_{jk\to j}^{de}$ from the morphed image I_{jk}^m and the target image I_k by de-morphing as

$$I_{jk\to j}^{de} = \mathrm{DM}(I_{jk}^m, I_k, C_k, C_j, \alpha_j). \quad (11)$$

Step 7: ADG manually compares the restored face image $I_{jk\to j}^{de}$ with the face of passenger PG_j. If they are the same, PG_j is a legal passenger; otherwise, the authentication fails.

The passenger authentication phase is demonstrated in Figure 3.

4 Correctness and Security Analyses

4.1 Correctness Analysis

Correctness means that the face image of each passenger can be restored correctly in the passenger authentication phase of our proposed scheme. After the pass buying phase is completed, the passenger PG_j obtains a JR pass which has a unique, printed number PID_j and stores the coordinates of the control pixels C_j of PG_j's face image I_j and the morphing rate α_j. Meanwhile, PG_j's corresponding morphed image I_{jk}^m that conceals the information of I_j is stored on the cloud storage server CSS. Therefore, the keys for reconstructing PG_j's face image I_j are distributed among different places to enhance the security.

Let Key_1, Key_2 and Key_3 denote three different keys for reconstructing the image I_j through de-morphing. These three keys are described as follows: (1) $Key_1 = \{PID_j, C_j, \alpha_j\}$. Key_1 is stored on the JR pass held by PG_j. Among the three elements of Key_1, the JR pass number PID_j is printed explicitly on the pass. (2) $Key_2 = I_k$. Key_2 is a reference face image I_k which works together with I_j to generate a morphed image I_{jk}^m. It is secure enough to have a same Key_2 for different I_j because the combination of different I_j and a same Key_2 can generate distinguished morphed images. Thus, Key_2 can be stored on a local memory and is retrieved easily by all terminals TS and TG. (3) $Key_3 = I_{jk}^m$. Key_3 is stored on the CSS and the location where Key_3 is stored has a relationship with the value of PID_j in Key_1.

Actually, the pass buying phase is an image morphing process and the passenger authentication phase is a de-morphing process. Since de-morphing is the reverse of morphing, TG can successfully reconstruct the face image of the passenger through de-morphing as long as it can obtain the same parameters used in previously conducted morphing. In other words, if TG can obtain all the keys Key_1, Key_2 and Key_3, it can ensure that face image of the passenger restored by TG is correct. As shown in Figure 4, TG can directly retrieve Key_1 from the JR pass and Key_2 from the local memory, and immediately selects C_k of I_k by using Key_2. To obtain Key_3, TG must send the index IND_j of PID_j to CSS. Then, CSS uses IND_j to determine where I_{jk}^m is stored and then transmits I_{jk}^m to TG. Consequently, TG obtains the same parameters $\{C_j, \alpha_j, I_k, C_k\}$ employed in the previous morphing and the output image I_{jk}^m of morphing.

According to the obtained parameters, TG can restore the face image of the passenger. First, horizontal-distance vector D_1 and vertical-distance vector D_2 are computed by using C_j and C_k. Then, interpolation matrices B_1 and B_2 are obtained according to D_1 and D_2. Afterwards, the warped target image I_k^w is obtained by Equation (6) and the warped source image I_j^w is computed by Equation (8). Finally, all pixels in I_j^w are shifted to their original locations in I_j by Equation (9) and the restored source image $I_{jk\to j}^{de}$ is obtained. Therefore, $I_{jk\to j}^{de}$ is the same as the face image I_j of PG_j, which implies that the correctness of our proposed scheme is achieved.

To further prove the correctness, some experiments utilizing the field morphing method [3] are conducted and the results are shown in Table 2. Assume there is a passenger PG_1 whose face image is I_1. In the pass buying phase, the morphed images with different morphing rates ($\alpha_1 = 0.1, 0.3, 0.5, 0.7$ and 0.9) can be generated from

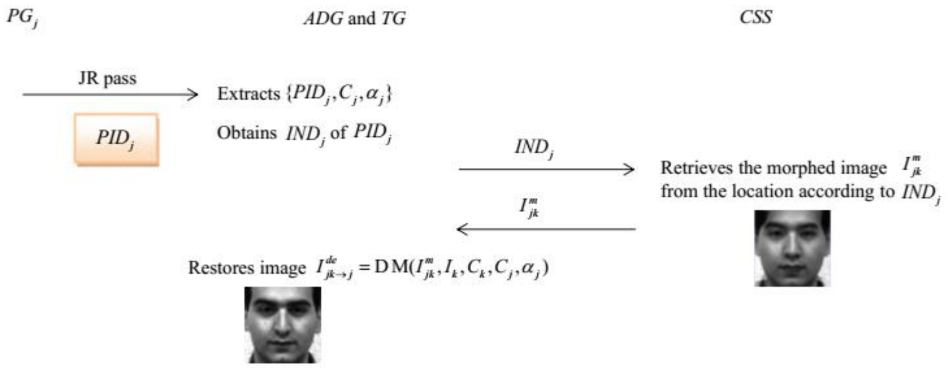

PG_j
ADG and TG
CSS
JR pass
PID_j
Extracts $\{PID_j, C_j, \alpha_j\}$
Obtains IND_j of PID_j
IND_j
Retrieves the morphed image I_{jk}^m
from the location according to IND_j
I_{jk}^m
Restores image $I_{jk \to j}^{de} = \mathrm{DM}(I_{jk}^m, I_k, C_k, C_j, \alpha_j)$

Figure 3: Passenger authentication phase of our proposed scheme

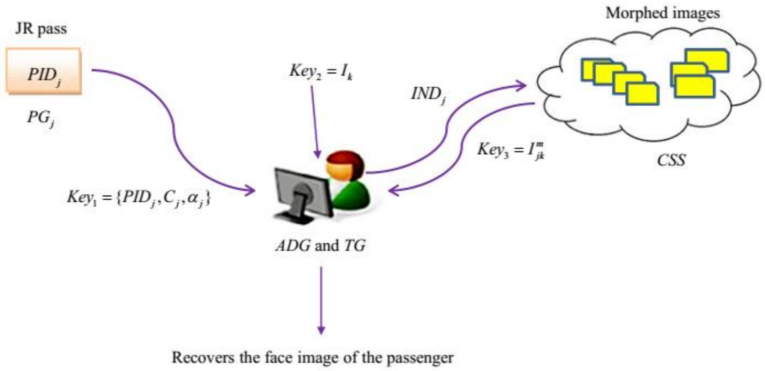
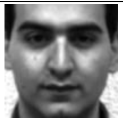
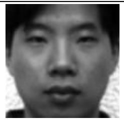
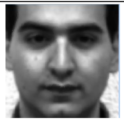
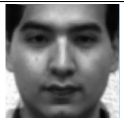
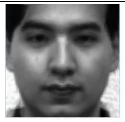
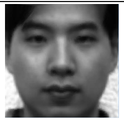
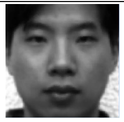
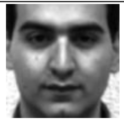
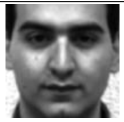
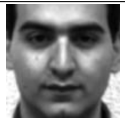
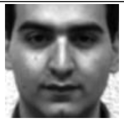
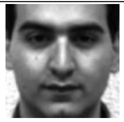


Figure 4: Keys for passenger authentication

Table 2: Experimental results

Source Image I_1 (for Passenger PG_1)			Target Image I_2		
Morphing Rate	$\alpha_1 = 0.1$	$\alpha_1 = 0.3$	$\alpha_1 = 0.5$	$\alpha_1 = 0.7$	$\alpha_1 = 0.9$
Morphed Image I_{12}^m					
Restored Image $I_{12 \to 1}^{de}$					
PSNR of $I_{12 \to 1}^{de}$	35.14 dB	34.83 dB	32.91 dB	31.67 dB	31.45 dB

the source image I_1 and the target/reference image I_2, as demonstrated in Table 2. It is observed that the morphing rate α_1 plays a very important role in the way the morphed image looks like for the given source image and target image. Therefore, the morphed image can be more similar to the source image when a smaller α_1 is set; on the contrary, the morphed image can be more similar to the target image when a larger α_1 is set. Table 2 also illustrates the restored images from different morphed images in the passenger authentication phase. The visual quality of the restored image is very good, thus it is impossible to visually perceive the slight difference between the original source image and the restored one.

4.2 Security Analysis

In this subsection, we analyze that our proposed scheme can enhance security significantly in terms of protecting the passenger's privacy and resisting well-known attacks.

First, the passenger's confidential information can be protected efficiently. No confidential information about the passenger but only a unique number is printed on the JR pass. This can guarantee that if the JR pass is stolen or discarded when it expires, no one can obtain the personal information from the pass. Moreover, although the face image of the passenger is used for authentication by our proposed scheme, it is not stored anywhere. Instead, a morphed image that hides the face image of the passenger is generated. To further protect the passenger's privacy, the morphed image is stored on CSS which can be retrieved only by authorized attendants in JR stations and the location for the storage of the morphed image is associated with the JR pass number. Even if an unauthorized person retrieves the morphed image from CSS, he/she is unable to know who the true passenger is since the morphed image is unlike the passenger's face image. In addition, the passport never needs to be carried with the passenger during the whole process of authentication, which extremely decreases the probability of losing the passport.

Next, we consider some attacking scenarios to analyze whether our proposed scheme is secure. Usually, a malicious attacker wants to impersonate the passenger to use the JR pass. The attacker can pass the authentication only if his/her face image can be restored by the attendant ADG who takes responsibility of the authentication. However, this impersonation attack will fail. Assume that the attacker steals or duplicates the JR pass and then presents it to an ADG. ADG conducts the operations of authentication in Subsection 3.3 and restores a face image via de-morphing. Because all the keys Key_1, Key_2 and Key_3 for de-morphing are identical to those used in morphing, ADG definitely restores the face image of the true passenger. Then, ADG can immediately distinguish the differences between the restored face image and that of the attacker. Therefore, the attacker cannot pass the identity authentication.

Since the attacker becomes aware that it is impossible for him/her to pass the authentication in the above situation, he/she tries to forge the data in Key_1 stored in the JR pass. If the attacker modifies the parameter PID_j in Key_1 and sends Key_1 to ADG, ADG will retrieve another morphed image I'^{m}_{jk} that was created from another passenger's face image, resulting in an incorrect Key_3. Then, ADG uses correct key Key_2 and incorrect keys Key_1 and Key_3 to restore a face image via de-morphing. However, the restored face image is unlike that of the attacker due to the incorrect keys so that the attack can be detected easily. The attacker may also modify the values of C_j and α_j in Key_1. Obviously, Key_1 becomes incorrect and it similarly leads to the fact that the restored face image is different from that of the attacker. Therefore, no matter how to forge the information stored in the JR pass, the attacker cannot make the face image of him/her look the same as the restored one.

Moreover, an attacker may launch an attack on ADG and obtain Key_2, *i.e*, the reference face image I_k. Then, the attacker can forge Key_3 by creating a morphed image using his/her face image and I_k. However, the attacker cannot put the forged Key_3 onto CSS since the attacker does not have the privilege of accessing CSS. Thus, ADG still retrieves correct Key_3 rather than the forged Key_3 and recovers a different face image from that of the attacker. In a word, the attacker fails to launch the impersonation attack on the true passenger due to the strategy that the keys for authentication are distributed among different places.

In the following, some experiments using the test images in Yale face database [5] and AT&T face database [4] are performed to further support the above analyses. Four attacking scenarios for impersonation attacks are considered, *i.e.*

1) The JR pass number PID_j in Key_1 is modified with the morphing rate $\alpha_j = 0.5$ while other elements in Key_1 are the same;

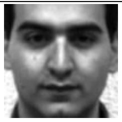
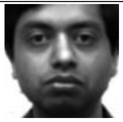
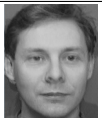
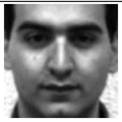
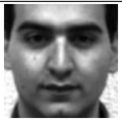
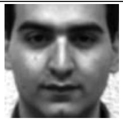
2) The coordinates of the control pixels C_j in Key_1 is modified with the morphing rate $\alpha_j = 0.5$ while other elements in Key_1 are the same;

3) α_j is modified to be 0.8 while other elements in Key_1 are the same;

4) Key_3 is forged by morphing I_k in Key_2 and the face image of the attacker I_α with the morphing rate $\alpha_j = 0.5$.

Table 3 demonstrates the attacking results on passenger PG_1 under the four scenarios mentioned above, using PSNR values to check whether the impersonation attacks are successful. From Table 3 we can see that all of the PSNR values are very small (less than 13 dB), which indicates that the difference between the restored image and the attacker's face image is very large, evaporating the attacker's attempt to restore a similar face image to that of him/her. Moreover, the PSNR values in the first and the last scenarios are extremely small (5.67 dB and 7.38 dB).

Table 3: Attack experiments

	Legal User PG_1 (Image I_1)		Attacker (Image I_α)	
	Attacking Scenario 1	Attacking Scenario 2	Attacking Scenario 3	Attacking Scenario 4
Recovered Image				
PSNR	5.67 dB	12.47 dB	12.25 dB	7.38 dB

This is because the modified parameters in these two scenarios have made more significant impact on restoration according to previous analyses in this subsection.

5 Conclusions

In this paper, we proposed a novel passenger authentication scheme for the JR pass, which innovatively adopts image morphing and cloud storage techniques to significantly enhance the security compared to traditional JR pass authentication mechanisms. The proposed scheme has the following contributions:

1) Only an assigned, unique number is printed on each JR pass to protect the passenger's privacy;

2) A morphed image is generated by the face image of the passenger and a pre-selected reference image through morphing, thereby hiding the face image of the passenger into the morphed image;

3) The generated morphed image is stored on the cloud storage, which efficiently avoids the disclosure of the passenger's personal information;

4) The authentication for the JR passenger is quite simple by performing an image de-morphing process. The proposed scheme can significantly increase both security and convenience according to our analyses.

References

[1] *Japan Railways Group. Explore with a Japan rail pass.* http://www.jrpass.com,accessed, Dec. 2016.

[2] *JRPass ltd. Japan rail pass: your sightseeing passport to Japan.* http://www.japanrailpass.net/en/index.html, Dec. 2016.

[3] T. Beier and S. Neely, "Feature-based image metamorphosis," *ACM SIGGRAPH Computer Graphics*, vol. 26, pp. 35–42, 1992.

[4] AT&T face database. http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html.

[5] Yale face database. http://cvc.yale.edu/projects/yalefaces/yalefaces.html.

[6] 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 Trans. Parallel Distrib. Syst*, vol. 22, pp. 1390–1397, 2011.

[7] W. S. Juang, S. T. Chen, and H. T. Liaw, "Robust and efficient password-authenticated key agreement using smart card," *IEEE Trans. Ind. Electron*, vol. 55, pp. 2551–2556, 2008.

[8] L. Lamport, "Password authentication with insecure communication," *Commun. ACM 24*, pp. 770–772, 1981.

[9] C. T. Li, "A new password authentication and user anonymity scheme based on elliptic curve cryptography and smart card," *IET Inf. Secur*, vol. 7, pp. 3–10, 2013.

[10] X. Li, J. Niu, M. K. Khan, and J. Liao, "An enhanced smart card based remote user password authentication scheme," *J. Netw. Comput. Appl*, vol. 36, pp. 1365–1371, 2013.

[11] X. Li, J. Niu, Z. Wang, and C. Chen, "Applying biometrics to design three-factor remote user authentication scheme with key agreement," *Secur. Commun. Netw*, vol. 7, pp. 1488–1497, 2014.

[12] Q. Mao, K. Bharanitharan, and C. C. Chang, "Edge directed automatic control point selection algorithm for image morphing," *IETE Tech. Rev*, vol. 30, pp. 343–243, 2013.

[13] Q. Mao, K. Bharanitharan, and C. C. Chang, "A proxy user authentication protocol using source-based image morphing," *Comput. J*, vol. 58, pp. 1573–1584, 2015.

[14] Q. Mao, C. C. Chang, L. Harn, and S. C. Chang, "An image-based key agreement protocol using the morphing technique," *Multimed. Tool. Appl*, vol. 74, pp. 3207–3229, 2015.

[15] Q. Zhao, M. Akatsuka, and C. H. Hsieh, "Generating facial images for steganography based on iga and image morphing," in *IEEE International Conference on Systems, Man, and Cybernetics*, pp. 364–369, Seoul, Korea, 2012.

[16] Q. Zhao and C. H. Hsieh, "Card user authentication based on generalized image morphing," in *The 3rd International Conference on Awareness Science and Technology*, pp. 117–122, Dalian, China, 2011.

Biography

Yanjun Liu received her Ph.D. degree in 2010, in School of Computer Science and Technology from University of Science and Technology of China (USTC), Hefei, China. She was an assistant professor serving in Anhui University in China from 2010 to 2015. She serves as a senior research fellow in Feng Chia University in Taiwan since 2015. Her specialties include E-Business security and electronic imaging techniques.

Chin-Chen Chang received his Ph.D. degree in computer engineering from National Chiao Tung University. His current title is Chair Professor in Department of Information Engineering and Computer Science, Feng Chia University, from February 2005. He is currently a Fellow of IEEE and a Fellow of IEE, UK. He consecutively won Outstanding Talent in Information Sciences of the R. O. C., AceR Dragon Award of the Ten Most Outstanding Talents, Outstanding Scholar Award of the R. O. C., Outstanding Engineering Professor Award of the R. O. C., Distinguished Research Awards of National Science Council of the R. O. C., Top Fifteen Scholars in Systems and Software Engineering of the Journal of Systems and Software, and so on. His current research interests include database design, computer cryptography, image compression, and data structures.

Privacy Preserving and Public Auditable Integrity Checking on Dynamic Cloud Data

Surmila Thokchom[1] and Dilip Kr. Saikia[2]
(Corresponding author: Surmila Thokchom)

Department of Computer Science and Engineering, National Institute of Technology Meghalaya[1]
Bijni Complex, Laitumkhrah, Shillong, India
Department of Computer Science and Engineering, Tezpur University[2]
Tezpur, Assam, India
(Email: surmila.thokchom@nitm.ac.in)

(Received Aug. 9, 2017; revised and accepted Nov. 18, 2017)

Abstract

The Cloud storage service allows data owner to outsource data storage at the Cloud which introduces security challenges requiring an auditor to check the integrity of stored data. This paper proposes an efficient auditing scheme for checking the integrity of dynamic data outsourced at untrusted Cloud storage. This scheme based on the Boneh and Boyen signature enables a third-party auditor to audit the client's data while preserving the privacy of the data. The scheme is found to be secure in the standard model. Complexity analysis shows the proposed scheme is efficient when compared to existing schemes.

Keywords: Batch Auditing; Cloud Computing; Dynamic Data; Privacy-Preserving Auditing

1 Introduction

Cloud computing introduces many attractive services. One such service is the cloud storage where the Cloud hosts the data and software of the client [21, 24]. Due to the economic advantage it offers and its elastic nature the service is very attractive for enterprises as well as individuals. However, the service introduces security concerns for the clients. The client loses control over their own data since it is stored in the Cloud servers. The data is put on risk from various threats in terms of privacy and integrity, both from within and outside the Cloud service provider. To overcome these threats the client cannot remain reliant on the assurance of the service provider alone. There is need for additional checks. To take care of the privacy issue the data can be transmitted and stored in encrypted form [1]. Provisions need to be made for verification of integrity of the data, preferably by an independent auditor without compromising on privacy of the data to the Cloud as well as the Auditor.

Various integrity checking schemes have been proposed over the years to check the integrity of the stored data in the Cloud. Integrity checking in these schemes is either performed by the data owner or by a third-party auditor who is employed by the data owner for verifying the integrity of the data on their behalf [16].

This paper presents a scheme for storage of dynamic Cloud data in the standard security model based on Boneh-Boyen signature [4, 5]. The scheme allows public auditability preserving data privacy. The scheme extends the Boneh-Boyen signature scheme that uses bilinear mapping to allow integrity checking of encrypted data so that the data privacy is not compromised while auditing. It is shown that the proposed scheme is efficient and secure through complexity analysis and simulation studies. The proposed scheme is designed to minimize the computational load on the auditor in the verification process. A mechanism is also presented for the scheme for batch auditing in multi-owner and multi-Cloud environment.

The rest of the paper is organized as follows. Section 2 discusses the existing related works. Section 3 presents the system model used and Section 4 presents the theoretical preliminaries. The proposed auditing protocol is presented in Section 5. Section 6 discusses the dynamic data operations with data integrity assurance support and Section 7 analyses the security issues of the model. The performance analysis is done in Section 8. Section 9 presents the multi-owner and multi-Cloud batch auditing support and Section 10 concludes the paper.

2 Related Works

Ateniese *et al.* [2] were the first one to propose provably secure schemes using RSA-based Homomorphic Verifiable Tag (HVT) to verify the integrity of stored data in the cloud without retrieving the data from the cloud. Their schemes have several drawbacks such as high overhead on server computation and communication cost and fails to provide fully secure data possession. Their scheme only supports static data. Zhu *et al.* [46] propose the Cooperative PDP (CPDP) scheme which is based on homomorphic verifiable response and hash index hierarchy. Their scheme emphasize on checking the integrity of client's data stored at multiple cloud service provider. Their scheme considers only static data. Hanser *et al.* [14] proposed a provable data possession based on elliptic curves cryptosystem which allows the data owner and the third party auditor to simultaneously audit the data outsource at the cloud storage. Their scheme only provides probabilistic guarantee of data possession and supports only static data.

Juels *et al.* [18] proposed the first Proof of Retrievability method where the sentinels are hidden among the regular data before outsourcing the data to the cloud storage. The verifier checks the data on the basis of the sentinels hidden among the data. Any changes in the data will affect the sentinels. The number of challenge is restricted by the number of embedded sentinels. Their scheme also supports only static data. Shacham *et al.* [28,29] proposed two compact PoR schemes. One is a public verifiable PoR scheme built using BLS signature [6] and the other one is a private verifiable PoR schemeusing the pseudo-random function. Schwarz *et al.* [27] propose a challenge-response scheme the data store remotely based on the algebraic signature properties in the peer-to-peer networks. Chen [10] extended the algebraic signatures for checking the possession of data in cloud storage. Their scheme has less overhead on the server and the client as compare to homomorphic based scheme. Their scheme also supports only static data. Wang [36] proposed a proxy provable data possession scheme using bilinear pairing technique. Their scheme uses a proxy for checking the integrity of the outsource data. Their scheme is also meant for static data only.

Ateniese *et al.* [3] updated the static Provable Data Possession methods called scalable Provable Data Possession based on symmetric key cryptography to make the scheme dynamic and to increase the efficiency and scalability. Their scheme supports data modification, data appending and data deletion operations. Their scheme restricts the number of updates and challenges and does not support data insertion.

The schemes proposed by Wang *et al.* [34,35] also supports only partial dynamic data operations. Erway *et al.* [12, 13] proposed Dynamic Provable Data Possession which support full dynamic data operations such as insertion, deletion, modification and appending. The authors presented two varieties of Dynamic PDO scheme. The first scheme utilizes rank based authenticated skip list for supporting fully dynamic data operations and the second scheme uses rank based RSA trees. Both the schemes use homomorphic block tag. The first scheme does not support privacy. Second scheme has high probability of possession guarantee but incurs high computation as compared to the first scheme and it lacks flexibility for data updates. Wang *et al.* [37] proposed a scheme which enables public auditing and supports full data dynamic operations. Their scheme uses Merkle Hash Tree and bilinear aggregate signature. Their scheme does not consider privacy of the data from the auditor and incurs high computation at the auditor side. The scheme proposed by Hu *et al.* in [15], support dynamic data. However the privacy of the data might be compromised as the server needs to send linear combinations of the data as proof to the auditor. Wang *et al.* [32, 33] proposed a scheme which solves data leakage problem to auditor by using homomorphic linear authenticator (HLA) and random masking before sending the proof to the auditor. Their scheme suffers from large storage overhead at the server side due to large number of data tags. The scheme is found to be vulnerable to attacks from a malicious CSP and an outside attacker. The vulnerability of this scheme is because of the inappropriate definition and the use of private/public parameters during signature generation. Worku *et al.* [38] proposed a scheme which is more efficient than the protocol in [33]. However, Liu *et al.* [22] has shown that a malicious Cloud service provider can still produce a proof to the challenge given by an auditor without being caught even after deleting all files of a data owner.

Li *et al.* [19] proposed a light weight integrity checking scheme meant for low performance end devices. The scheme is a privacy preserving public auditable and supports dynamic data operation. Since the scheme is for low end devices data uploading is only for very small set. Zhang *et al.* [45] propose a scheme taking into consideration if the auditor colludes with the cloud server. The scheme utilizes the Bitcoin to generate the challenge random blocks and a check log file is maintained. The data owner will check the check log file to confirm that the auditor has done a fair integrity checking on the data. This adds a computation overhead for the data owner.

The scheme proposed by Yang *et al.*. [39] supports dynamic data, batch auditing, and preserves data privacy in random oracle model. The scheme uses Bilinearity property of Bilinear pairing. The server presents the proof of the data possession to auditor in an encrypted form which auditor can only verify. The scheme uses index table which will incur storage overhead proportional to the file size on the auditor.

Chattopadhyay *et al.* [8], proposed a scheme using simple low cost Boolean based encryption and decryption for image-files only. The encrypted data files will be shared on the Cloud. A threshold (t, n)-secret sharing scheme is used for obtaining the symmetric key. Liu *et al.* [23] proposed a public auditing scheme for the regenerating-code-based Cloud storage. In the absence of the data owner a proxy which has a priviledge is used to regenerate athenticators thus allowing the data owner not to be online all the time. Privacy of the data is preserved by randomizing the encode coefficients with a pseudorandom function. Chen *et al.* [11] proposed a remote data possession checking (RDPC) scheme. Their scheme is based on homomorphic hashing and the Merkle hash tree is used for enabling the data dynamics operations. Yu *et al.* stated in their paper [43] that [11] schemes is vulnerable to forgery attack and replace attack launched by a malicious server and their proposed scheme has shown improvement to overcome this vulnerability.

Lin *et al.* [20] proposed a scheme for mobile provable data possession. The scheme uses a hash tree data structure and a Boneh-Lynn-Shacham [6] short signature scheme. To reduce the computation overhead of the mobile data owner for generating the tag of the data block, trusted third party is utilized. Many communications take place between the trusted third party and the data owner while generating the data tag and that leads to high communication overhead between the third party and the data owner. Yi et al. [40] proposed a multi-copy Provable Data Possession. The scheme considers that the data owners stores multiple copies of sensitive data in the cloud. The scheme gives assurance that multiple copies of data are consistent with the latest version.

The scheme by Chen *et al.* [9] is based on homomorphic network coding signature scheme. It does not support dynamic data. Ma *et al.* [26] proposes an efficient privacy preserving scheme based on homomorphic network coding and RSA for the standard model and supports dynamic data. The scheme does not consider batch auditing.

Various Cloud storage auditing exists [7, 41, 42] which deals in key exposer problems. There are many integrity checking schemes for shared data among group of users [17, 25, 30, 31, 44]. This paper does not go into details of the shared data schemes since shared data is not considered in the proposed scheme.

3 The System Model

The proposed auditing scheme considers the cloud storage architecture as illustrated in Figure 1. This storage architecture comprises three entities; The client or the data owner, the Cloud servers and the third-party auditor. The client creates data and stores it at the Cloud storage. Upon requirement the client can retrieve and update the data. The Cloud server stores and maintains the client data and gives access the data to the client. The auditor is a neutral trusted entity who has the expertise and resources to perform integrity checking on large data sets. The auditor periodically or upon request will challenge the Cloud server to provide proof of integrity of the outsourced data. Based on the proof provided by the Cloud server on the challenge sets, the third-party auditor will deliver unbiased audit reports to the client.

The data integrity threats of the client data from the server may be non-malicious or malicious. At any time, if the integrity of the client data is compromised, the Cloud servers may try to hide it so as to maintain its reputation. Therefore, the Cloud server is considered untrusted. The dynamic data stored on the cloud may face the following attack [39]:

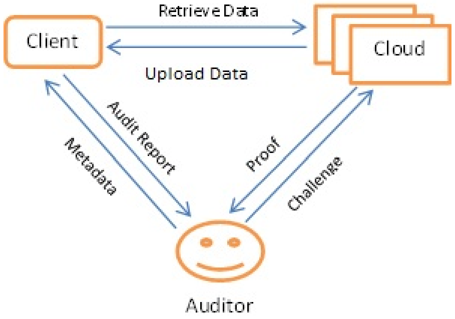


Figure 1: System model

1) *Replay attack*: The client's data may not be updated correctly on the server and to make up the mistake, the server may use the previous uncorrupted pair of data block and data tag, to replace the challenge pair of data and data tag so that the auditing will passed.

2) *Forging attack*: If the Cloud server has the information required for generating the data tags, it can forge the data when the client updates his data to a new version. The forging by the server may go undetected by the auditor if suitable provisions are not made.

On the other hand, the integrity of the auditor is not questioned. However unprotected data may lead to loss of privacy. Hence the data is to be transmitted and stored in encrypted form and the auditor should be able to verify the integrity of the stored data with zero knowledge of the content.

4 Preliminaries

The scheme proposed is primarily based on Bilinear Map and Boneh-Boyen signature scheme. These are briefly discussed below

4.1 Bilinear Maps

Let G_1, G_2 and G_T be multiplicative cyclic groups of order p. Let g_1 be the generator of G_1 and g_2 be the generator of G_2. A map $e : G_1 \times G_2 \rightarrow G_T$ will be a bilinear map if it satisfies the following properties:

- *Computable*: an efficient algorithm exists for computing map e;
- *Bilinear*: $e(u^a, v^b) = e(u, v)^{ab}$ for all $u \in G_1$, $v \in G_2$ and $a, b \in Z_p$;
- *Non-degeneracy*:$e(g_1, g_2)$ does not equal to 1.

4.2 Boneh-Boyen Signature Scheme

The Boneh-Boyen signature scheme is based on Bilinear Mapping and it comprises the three functions of *Key Generation*, *Signing* and *Verification*. These are defined as follows.

KeyGeneration: Generate the key pair *(PK, SK)* as follows. Choose random integers $x, y \rightarrow Z_p^*$ and compute, $A = g_2^x \in G_2$, $B = g_2^y \in G_2$, $z \leftarrow e(g_1, g_2)$ The public key is $PK{=}(g_1, g_2, A, B, z)$ and the private key is $SK{=}(g_1, x, y)$.

Signing: Given a message $m \in Z_p$ and a private key $SK{=}(g_1, x, y)$, select a random value $r \in Z_p$ and compute $S = g_1^{\frac{1}{(x+m+yr)}}$ with the inverse $\frac{1}{(x+m+yr)}$ computed modulo p. The signature is $\sigma = (S, r)$.

Verification: Given a message $m \in Z_p$ and a public key $PK{=}(g_1, g_2, A, B, z)$ and signature $\sigma = (S, r)$, it is verified by checking $e(\sigma, A, g^m, B^r) = z$ if is true.

Table 1: Notation used

A, B	Public Keys
α, x, y	Private Keys
σ	Tag
s	Intermediate value of Tag
r, h, t	Random values
n	Number of blocks
k	Number of Challenge blocks
C_{info}	Meta data

5 Proposed Cloud Storage Auditing Protocol

The proposed Cloud storage auditing protocol assumes that a data file F is split into n number of data blocks $(b_1, b_2, b_3 \ldots\ldots b_n)$ and these data blocks are encrypted individually with a suitable encryption algorithm to produce the encrypted data file $C{=}(c_1, c_2, c_3 \ldots\ldots c_n)$ before they are uploaded. The scheme uses the following five functions for the auditing as illustrated in Figure 2.

1) $GenKey(k) \rightarrow K$: This function is executed at the client side taking security parameter k as input and produces a secret key and public key pair, *K=(SK,PK)*;

2) $GenTag(C, SK, h, t) \rightarrow S$: This function takes an encrypted data file C and the secret key SK, random values h and t to compute a set of the data tags S, one for each of the data blocks in C. The encrypted data blocks in C and their corresponding tags in$S$ and t are uploaded to the Cloud storage and corresponding C_{info} which comprises the tuple $\langle$*Index, BlockName, Version number, h value for the data block uploaded in the cloud storage* $\rangle$. Here h is treated as a secret value between data owner and the auditor;

3) $GenChall(C_{info}) \rightarrow Q$: The auditor executes this function with the meta data,C_{info} as input to generate a challenge Q, to be sent to the Cloud server;

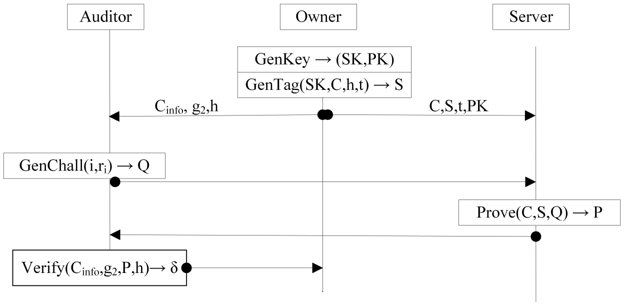


Figure 2: Work flow of the auditing scheme

4) $Prove(Q, C, S, PK) \rightarrow P$: The cloud server executes this function taking the challenge Q received from the auditor, the stored data file and its set of random t value *(C,t)* and the public key PK as inputs to produce a proof P;

5) $Verify(P, Q, C_{info}) \rightarrow V$: The auditor executes this function with inputs - the proof P provided by the server, the challenge Q, the metadata C_{info} and to produce the output value of *V =1*, if the proof is correct, otherwise produce.

5.1 Theoretical Basis

The original Boneh-Boyen signature scheme is adopted for the proposed scheme as follows.

- Let $(\alpha, x, y) \in Z_p$ be a randomly generated value used as private keys *SK=(α, x, y)* and *PK= (A,B)* the public keys computed as- $A = g_2^{\frac{y}{(\alpha+x)}}$, $B = g_2^{\frac{1}{(\alpha+x)}}$

- Let $\Sigma = (\sigma_1, \sigma_2 \ldots \sigma_n)$ be the set of n tags with $\sigma_i = (t_i, s_i)$ generated for the encrypted data block c_i in, $C{=}(c_1, c_2 \ldots c_n)$ where t_i and h_i are two random values and $S_i \in G_2$ computed as-

$$s_i = h_i^{\frac{\alpha+x}{yt_i+c_i}} \tag{1}$$

- Let $Q = \{i, r_i\}, i = 1, 2 \cdots k$ be a challenge set generated with random values $r_i \in Z_p^*$, one for each of the chosen k number of data blocks in set D.

- Let Z be a quantity computed using h value from the metadata C_{info}, for each of the k chosen data blocks in D.

$$Z = \prod_{d=1}^{k} e\left(h_d, g_2\right)^{r_d^2} \tag{2}$$

- Let P be the proof generated for each of the chosen data blocks as-

$$P = \prod_{d=1}^{k} e\left(s_d^{r_d}, A^{r_d t_d}, B^{C_d r_d}\right) \tag{3}$$

Theorem: As per the bilinearity property $P = Z$. i.e.,

$$\prod_{d=1}^{k} e(s_d^{r_d}, A^{r_d t_d}, B^{C_d r_d}) = \prod_{d=1}^{k} e\left(h_d, g_2\right)^{r_d^2} \tag{4}$$

Proof.

$$\begin{aligned} P &= \prod_{d=1}^{k} e\left(s_d^{r_d}, A^{r_d t_d}, B^{C_d r_d}\right) \\ &= \prod_{d=1}^{k} e\left(h_d^{\frac{r_d(\alpha+x)}{yt_d+C_d}}, g_2^{\frac{r_d y t_d}{\alpha+x}}, g_2^{\frac{r_d C_d}{\alpha+x}}\right) \\ &= \prod_{d=1}^{k} e\left(h_d^{\frac{\alpha+x}{yt_d+C_d}}, g_2^{\frac{yt_d+C_d}{\alpha+x}}\right)^{r_d r_d} \end{aligned}$$

By Bilinearity Property we can rewrite the expression as

$$\begin{aligned} &= \prod_{d=1}^{k} e(h_d, g_2)^{\frac{(\alpha+x)(yt_d+C_d)r_d^2}{(yt_d+C_d)(\alpha+x)}} \\ &= \prod_{d=1}^{k} e(h_d, g_2)^{r_d^2} \end{aligned}$$

□

6 Data Integrity Assurance Support for Dynamic Data Operation

The auditor and the client will be maintaining metadata which consists of block, version, h value for each of the blocks. The index number is the current block number. For tag generation of each of the blocks the h value is used. There are three types of operations the client can perform to update their data.

1) *Data block modification*: Consider the operation ofthe client modifyingthe data block, *Block-b to Block-b'*. First of all, the client will download this block from the cloud server and make the required modification on the data block. The client will compute a new tag for the modified block. To compute new tag, the client will generate a random value for h and also a random value for t and compute the new tag σ using Equation (1). The client will then update *version* and h value in the metadata table. The client will then upload the modified block and new tag value and the new t value to the cloud server. The client will communicate the auditor the new h value and the *version* number of the modified block.

2) *Data block insertion*: Now consider the operation of inserting a newdata block, *Block-0* after the k^{th} block. The client will generate a random value h and t for computing the tag of the block. The client will update the metadata table by moving down all items following the k^{th} block entry in the table by one block. The client will then upload the new block and its corresponding tag value and t value to the cloud server. The client will communicate the auditor the insertion of the new block and its h value.

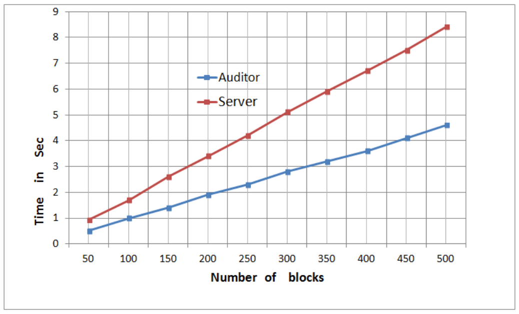


Figure 3: Computation costs of the auditor and the server

3) *Data block deletion*: If the client wants to delete the k^{th} block, it will send a request to the cloud server for removing the block and shall communicate the same to the auditor. In the metadata table, both the auditor and the client will delete the k^{th} entry from the table and shall move up the following entries in the table by one slot each.

7 Security Analysis of the Model

As discussed in Section 2, in a cloud storage system, when dynamic data are stored, the cloud server may carry out replay attack and forging attack. In the proposed scheme, h value is used while computing the tag for each block and as stated it is unique for each version of each block. Hence a replay attack will never pass the audit check.

On the other hand, if the server could forge the data tag, it can pass the audit using any data and its forge data tag. Forging a data tag in our scheme requires the server should be able to predict the value of h, which is a randomly selected unique value for each block. If any block is modified, this modified block will have a new h value. A forging attack by the server will therefore get detected in audit.

8 Performance Analysis

The Communication and Computation cost of the proposed scheme can be computed as follows.

- *Communication Cost*: The challenge and the proof parts will give the communication cost between the auditor and the server. In the challenge part, the cost depends on the number of blocks d which the auditor sent for audit. In the proof, it is the only the proof result. So the communication cost will be O(d).

Table 2: Comparison with different cloud data integrity auditing schemes

Schemes	**Computation cost**		**Communication cost**	**TPA**	**Privacy**	**Dynamic**	**Model**	**Batch**
	Server	**Auditor**						
[37]	O(d log n)	O(d log n)	O(d log n)	Yes	Yes	Yes	ROM	Yes
[33]	O(d log n)	O(d log n)	O(d log n)	Yes	Yes	Yes	ROM	Yes
[2]	O(d)	O(d)	O(d)	No	No	No	ROM	No
[39]	O(d)	O(d)	O(d)	Yes	Yes	Yes	ROM	Yes
[12]	O(d log n)	O(d log n)	O(d log n)	No	No	No	Standard	No
[9]	O(d)	O(d)	O(d)	Yes	Yes	No	Standard	No
[26]	O(d)	O(d)	O(d)	Yes	Yes	Yes	Standard	No
Our scheme	O(d)	O(d)	O(d)	Yes	Yes	Yes	Standard	Yes

- *Computation Cost*: The scheme in this paper involves three computations cost, namely at the owner side, the server side and the auditor. The simulation of scheme has been done on a Windows system with an Intel(R) Core(TM) i3 CPU at 3.60 GHz and 4.00GB RAM. The code for simulation of the scheme uses the pairing-based cryptography library version 0.5.12. An elliptic curve of MNT d159, with a base field size of 159 bits and embedding degree 6 is chosen. The d159 curve has a 160-bit group order, which means prime p is 160-bits long. The simulation is run multiple times and averaged to obtain stable results. Computation cost of the auditor and server versus the number of data blocks are compared in Figure 3. As shown in the Figure 3, data blocks are taken up to 500 blocks and block size is 2 KB. For 500 blocks (1000 KB) of data the Server requires around 8 seconds for providing the proof and the auditor requires around 4 seconds for verification.

  The graph in Figure 3, shows computation cost of the Auditor and the Server. The computation cost for the auditor consists of the time for auditing and verifying the data. Auditing time is just the generation of random numbers for the queried number of data blocks. The verifying time will compute Equation (3), which comprises mapping of each of the queried data blocks and then multiplying each of them. Each of the terms in Equation (3), takes *O(1)* time to be computed and the expression consists of product of challenged number of blocks. So time taken to compute will be *O(d)*, where d is the number of challenge blocks. The computation cost for the server is the time for proving the possession of the data given as a challenge. The server computes the expression in Equation (2) for providing the proof, which again takes *O(d)*, as each term in this equation takes *O(1)* to compute. From the given result, it is apparent that in the integrity checking protocol most of the computations are done at the server side for computation of the proof and thus minimizing the load on the auditor in the verification.

  Table 2 presents a comparison of the proposed scheme with other existing integrity schemes [2, 9, 12, 26, 33, 37, 39] in terms of computation complexity of the server and the verifier computation, communication cost, support for third party audit (TPA), preserving privacy, support of dynamic data [26, 39]. In the table, d represents the number of challenge blocks,n is the total number of blocks stored in the cloud.

9 Multiowner and Multicloud Batch Auditing Support

The proposed scheme can be extended to support batch auditing for multicloud and multiowner. Let n be the number of owner and l be the number of cloud service provider. Steps to be followed are as follows:

1) *Initialization*: Each of the owner will run the *GenKey(k)* algorithm to generate a pair of private and public keys. Each of the data owner can generate different private and public keys for different cloud servers. They will then run the *GenTag(c,sk)* algorithm to generate tag $\sigma_i = (t_i, S_i)$for each of their data blocks using Equation (1) as in *GenTag* algorithm in Section 4. The clients then uploads the data blocks and the corresponding tags to the chosencloud servers. Each owners then sends meta data to the auditor. The meta data consists of the cloud server name, block number and the h value for each of the datablocks.

2) *ChallengeBatchwise*: The auditor executes this function to generate challenge to l number of cloud severs for n number of data owners. It takes C_{info} as input to generate a challenge set Q, which consists of total m number of data blocks, to be sent to the cloud server. Out of m number of data blocks, each owner has k number of data blocks. The auditor generates a random number for each of the selected data blocks and sends these block numbersalong with the respective random numbers to the corresponding cloud servers.

3) *ProveBatch*: As a proof each cloud server will send the proof value to the auditor as:

$$P = \prod_{o=1}^{n} \prod_{d=1}^{k} e\left(s_d^{r_d}, A^{r_d t_d}, B^{C_d r_d}\right)$$

4) *VerifyBatch*: The auditor will first compute the product of all the proofs provided by individual cloud servers (cs).

$$\prod_{cs=1}^{l} P_{cs}$$

The auditor then computes the following product for the entire set of data blocks of all the owners in different cloud servers.

$$\prod_{d=1}^{m} e(h_d, g_2)^{r_d^2}$$

The auditor then verifies the integrity of the data by checking for the following equality:

$$\prod_{cs=1}^{l} P_{cs} = \prod_{d=1}^{m} e(h_d, g_2)^{r_d^2} \qquad (5)$$

10 Conclusion

The paper has presented an integrity checking scheme in standard model that supports dynamic data operations and public verifiability while preserving data privacy. The scheme is extended to support batch auditing for multi-owner and multi-cloud servers to increase the efficiency of the auditing scheme. It is shown that the scheme is resistant to replay and forge attacks by the server. In the scheme, the computation cost of verifying the data integrity is kept low for the auditor. The proposed scheme is shown to comply with the best existing schemes in terms of computational and communication complexity. The scheme also has the advantages of using the standard security model and supporting batch auditing in the multi-owner and multi-cloud environment.

The proposed scheme however suffers a small storage overhead at the auditor side in terms of requires storing the tag value h for each block to be stored at the auditor side thus creating some storage overhead at the auditor side. Future efforts to overcome this overhead and incorporating provision for sharing of the data by multiple clients may be worthwhile.

References

[1] D. S. AbdElminaam, "Improving the security of cloud computing by building new hybrid cryptography algorithms," *International Journal of Electronics and Information Engineering*, vol. 8, no. 1, pp. 40-48, 2018.

[2] 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 Commmunication Security (CCS'07)*, pp. 598–609, Oct. 2007.

[3] G. Ateniese, R. D. Pietro, L. V. Mancini, and G. Tsudik, "Scalable and efficient provable data possession," in *ACM Proceedings of the 4th International Conference on Security and Privacy in Communication Networks*, pp. 1–10, Sep. 2008.

[4] D. Boneh and X. Boyen, "Short signatures without random oracles," in *In Advances in Cryptology (EUROCRYPT'04)*, pp. 56–73, May 2004.

[5] D. Boneh and X. Boyen, "Short signatures without random oracles and the sdh assumption in bilinear groups," *Journal of Cryptology*, vol. 21, no. 2, pp. 149–177, 2008.

[6] D. Boneh, B. Lynn, and H. Shacham, "Short signatures from the weil pairing," in *Advances in Cryptology (ASIACRYPT'01)*, pp. 514–532, Dec. 2001.

[7] Z. Cao, L. Liu, and O. Markowitch, "Analysis of one scheme for enabling cloud storage auditing with verifiable outsourcing of key updates," *International Journal of Network Security*, vol. 19, no. 6, pp. 912–921, 2016.

[8] A. K. Chattopadhyay, A. Nag, and K. Majumder, "Secure data outsourcing on cloud using secret sharing scheme," *International Journal of Network Security*, vol. 19, no. 6, pp. 912–921, 2016.

[9] F. Chen, T. Xiang, Y. Yang, and S. S. M. Chow, "Secure cloud storage meets with secure network coding," in *Proceedings of IEEE Conference on Computer Communications (INFOCOM'14)*, pp. 673–681, Apr. 2014.

[10] L. Chen, "Using algebraic signatures to check data possession in cloud storage," *Future Generation Computer System*, vol. 29, no. 7, pp. 1709–1715, 2013.

[11] L. Chen, S. Zhou, X. Huang, and L. Xu, "Data dynamics for remote data possession checking in cloud storage," *Computers and Electrical Engineering*, vol. 39, no. -, pp. 2413–2424, 2013.

[12] C. Erway, A. Kupcu, C. Papamanthou, and R. Tamassia, "Dynamic provable data possession," in *Proceedings of ACM Conference on Computer and Commmunication Security*, pp. 213–222, Nov. 2009.

[13] C. Erway, A. Kupcu, C. Papamanthou, and R. Tamassia, "Dynamic provable data possession," *ACM Transactions on Information System Security*, vol. 17, no. 4, pp. 1–29, 2015.

[14] C. Hanser and D. Slamanig, "Efficient simultaneous privately and publicly verifiable robust provable data possession from elliptic curves," in *Proceedings of the 10th International Conference on Security and Cryptoghraphy*, pp. 1–10, July 2013.

[15] H. Hu and S. S. Yau, "Dynamic audit services for integrity verification of outsourced storages in clouds," in *Proceedings of ACM Symposium on Applied Computing*, pp. 1550–1557, Mar. 2011.

[16] M. S. Hwang, T. H. Sun, C. C. Lee, "Achieving dynamic data guarantee and data confidentiality of public auditing in cloud storage service," *Journal of Circuits, Systems, and Computers*, vol. 26, no. 5, 2017.

[17] T. Jiang, X. Chen, and J. Ma, "Public integrity auditing for shared dynamic cloud data with group user revocation," *IEEE Transactions on Computers*, vol. 65, no. 8, pp. 2363–2373, 2016.

[18] A. Juels, J. Burton, and S. Kaliski, "Pors: Proofs of retrievability for large files," in *Proceedings of the 14th ACM Conference on Computer and Commmunication Security (CCS'07)*, pp. 584–597, Oct. 2007.

[19] J. Li, L. Zhang, J. K. Liu, H. Qian, and Z. Dong, "Privacy-preserving public auditing protocol for low-performance end devices in cloud," *IEEE Transaction on Information Forensics And Security*, vol. 11, no. 11, pp. 2572–2583, 2016.

[20] C. Lin, Z. Shen, Q. Chen, and F. T. Sheldon, "A data integrity verification scheme in mobile cloud computing," *Journal of Network and Computer Applications*, vol. 77, no. -, pp. 146–151, 2017.

[21] C. W. Liu, W. F. Hsien, C. C. Yang, and M. S. Hwang, "A survey of public auditing for shared data storage with user revocation in cloud computing", *International Journal of Network Security*, vol. 18, no. 4, pp. 650–666, 2016.

[22] H. Liu, L. Chen, Z. Davar, and M. Pour, "Insecurity of an efficient privacy-preserving public auditing scheme for cloud data storage," *Journal of Universal Computer Science*, vol. 21, no. 3, pp. 473–482, 2015.

[23] J. Liu, K. Huang, H. Rong, H. Wang, and M. Xian, "Privacy-preserving public auditing for regenerating-code-based cloud storage," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 7, pp. 1513–1528, 2015.

[24] L. Liu, Z. Cao, C. Mao, "A note on one outsourcing scheme for big data access control in cloud," *International Journal of Electronics and Information Engineering*, vol. 9, no. 1, pp. 29–35, 2018.

[25] X. Liu, Y. Zhang, B. Wang, and J.Yan, "Mona: Secure multi-owner data sharing for dynamic groups in the cloud," *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 6, pp. 1182–1191, 2013.

[26] M. Ma, J. Weber, and J. Berg, "Secure public-auditing cloud storage enabling data dynamics in the standard model," in *Third International Conference on Digital Information Processing Data Mining and Wireless Communications*, pp. 170–175, July 2016.

[27] T. S. J. Schwarz and E. L. Miller, "Store, forget, and check: Using algebraic signatures to check remotely administered storage," in *Proceedings of the 26th IEEE International Conference on Distributed Computing Systems (ICDCS'06)*, pp. 12–22, July 2006.

[28] H. Shacham and B. Waters, "Compact proofs of retrievability," in *Advances in Cryptology (ASIACRYPT'08)*, pp. 90–107, Dec. 2008.

[29] H. Shacham and B. Waters, "Compact proofs of retrievability," *Journal of Cryptology*, vol. 26, no. 3, pp. 442–483, 2013.

[30] B. Wang, B. Li, and H. Li, "Oruta: Privacy-preserving public auditing for shared data in the cloud," *IEEE Cloud Computing*, vol. 2, no. 1, pp. 43–56, 2014.

[31] B. Wang, B. Li, and H. Li, "Panda: Public auditing for shared data with efficient user revocation in the cloud," *IEEE Trans Services Computing*, vol. 8, no. 1, pp. 92–106, 2015.

[32] C. Wang, S. chow, Q. Wang, K. Ren, and W. Lou, "Privacy-preserving public auditing for data storage security in cloud computing," in *Proceedings of the 26th IEEE International Conference on Distributed Computing Systems (ICDCS'06)*, pp. 1–9, Mar. 2010.

[33] C. Wang, S. chow, Q. Wang, K. Ren, and W. Lou, "Privacy-preserving public auditing for secure cloud storage in cloud computing," *IEEE Transactions on Computers*, vol. 62, no. 2, pp. 362–375, 2013.

[34] C. Wang, K. Ren, W. Lou, and J. Li, "Toward publicly auditable secure cloud data storage services," *IEEE Network*, vol. 24, no. 4, pp. 19–24, 2010.

[35] C. Wang, Q. Wang, K. Ren, and W. Lou, "Towards secure and dependable storage services in cloud computing," *IEEE Transactions on Service Computing*, vol. 5, no. 2, pp. 220–232, 2012.

[36] H. Wang, "Proxy provable data possession in public clouds," *IEEE Transaction on Services Computing*, vol. 6, no. 4, pp. 551–559, 2013.

[37] 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.

[38] S. Worku, C. Xu, J. Zhao, and X. He, "Secure and efficient privacy-preserving public auditing scheme for cloud storage," *Computers and Electrical Engineering*, vol. 40, no. 5, pp. 1703–1713, 2014.

[39] K. Yang and X. Jia, "An efficient and secure dynamic auditing protocol for data storage in cloud computing," *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 9, pp. 1717–1726, 2013.

[40] M. Yi, J. Wei, and L. Song, "Efficient integrity verification of replicated data in cloud computing system," *Computers and Security*, vol. 65, pp. 202–212, 2017.

[41] J. Yu, K. Ren, C. Wang, and V. Varadharajan, "Enabling cloud storage auditing with key-exposure resistance," *IEEE Transactions on Information Forensics Security*, vol. 10, no. 6, pp. 1167–1179, 2015.

[42] J. Yu, K. Ren, C. Wang, and V. Varadharajan, "Enabling cloud storage auditing with verifiable outsourcing of key updates," *IEEE Transactions on Information Forensics Security*, vol. 11, no. 6, pp. 1362–1375, 2016.

[43] Y. Yu, J. Ni, M. H. Au, H. Liu, H. Wang, and C. Xu, "Improved security of a dynamic remote data possession checking protocol for cloud storage," *Expert*

Systems with Applications, vol. 41, pp. 7789–7796, 2014.

[44] J. Yuan and S. Yu, "Public integrity auditing for dynamic data sharing with multiuser modification," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 8, pp. 1717–1726, 2015.

[45] Y. Zhang, C. Xu, H. Li, and X. Liang, "Cryptographic public verification of data integrity for cloud storage systems," *IEEE Cloud Computing*, vol. 3, no. 5, pp. 44–52, 2016.

[46] Y. Zhu, H. Hu, G. Ahn, and M. Yu, "Cooperative provable data possession for integrity verification in multi-cloud storage," *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 12, pp. 2231–2244, 2012.

Biography

Surmila Thokchom is an Assistant Professor of Computer Science and Engineering at the National Institute of Technology Meghalaya, India. Her research interest includes cryptography, information security, cloud computing.

Dr. Dilip Kr. Saikia is a Professor of Computer Science and Engineering at Tezpur University, India. Current areas of his research interest include Software Defined Networks, Network Function Virtualization, Wireless Sensor Networks and Network Security.

On Security Improvement of Adaptive Pixel Pair Matching with Modified Searching Mechanism

Wien Hong[1,2], Shuozhen Zheng[1], and Xiaoyu Zhou[1]
(Corresponding author: Wien Hong)

School of Electrical and Computer Engineering, Nanfang College of Sun Yat-Sen University[1]
882 Wenquan Rd. Conghua district, Guangzhou 510970, China
School of Computer and Software, Nanjing University of Information Science and Technology[2]
219 Niliu Rd. Nanjing, Jiangsu 210044, China
(Email: wienhong@gmail.com)

(Received Oct. 16, 2017; revised and accepted Mar. 27, 2018)

Abstract

This paper proposes a data hiding scheme that improves the adaptive pixel pair matching (APPM) method. Based on pixel pair matching, APPM employs a pixel pair as an embedding unit, and uses a specially designed reference table that minimizing the embedding distortion for data embedment. Although APPM has the capability to embed secret digit in any notational system, it is vulnerable to the detection by the RS scheme if digits in 4-ary notational system are embedded into images with large flat area (pixels having the similar grayscale values) such cartoon images. A modified version of APPM is proposed in this paper by using a revised pixel pair replacement mechanism (PPRM). With the proposed PPRM method, the stego image not only is totally undetectable by the RS scheme but also provides the equivalent image quality of the original APPM method.

Keywords: APPM; Data Hiding; Pixel Pair Matching

1 Introduction

The simple LSB substitution technique is a commonly used data hiding method in which least significant bits of pixels are replaced by secret data. The LSB method is easy to implement, and achieves an acceptable image quality. Therefore, it is widely used in many applications such as data hiding, watermarking, and image authentication [1,6,9,14,15,17,19,21–23]. However, during the LSB embedment, pixels with odd values remain unchanged or subtracted by one, and pixels with even values remain unchanged or add by one. As a result, the unbalanced replacement significantly increases the detectability by the steganalyzers such as RS scheme [7]. Moreover, The LSB method distorts the image significantly. Therefore, it is not suitable for applications where a high image quality is demanded [7,13,16].

In 2004, Chan *et al.* [2] proposed a simple but efficient data hiding method by using optimal pixel adjustment process (OPAP). When secret data are embedded into the rightmost r LSBs, the OPAP method employs a simple adjustment for the leftmost $8-r$ bits such that the stego pixel value is the closest to its original pixel value. The OPAP method has the same payload as the LSB method but provides better image quality. However, the OPAP method has the equivalent distortion compared to that of the LSB method when the payload is 1 bit per pixel (bpp).

Both LSB and OPAP employ a single pixel as an embedding unit for data embedment. Another type of data hiding method utilizes a pixel pair as an embedding unit to embed a n-ary digit. Data hiding method of this type are termed pixel pair matching (PPM). The PPM-based method uses a reference table as a guide, and embeds a digit into a pixel pair by modifying pixel values of this pair. For example, to embed a digit d_B in base B into a pixel pair (r, c) using a reference table R_T, the coordinate (r, c) in R_T is firstly located and obtain a searching region $\Omega(r, c)$. In this region, a coordinate (r', c') is found which satisfies $R_T(r', c') = d_B$ and is the closest to (r, c). The pixel pair (r, c) is then replaced by the new coordinate (r', c'). The embedded digits d_B can be extracted by locating the element at coordinate (r', c') of the given reference table R_T, *i.e.*, $d_B = R_T(r', c')$. Figure 1 shows the schematic diagram of the PPM-based method.

Mielikainen [18] in 2006 proposed a LSB matching revisited (LSBMR) method based on PPM. In his method, only one pixel in a pixel pair is changed by one grayscale unit and two bits (a 4-ary digit) can be embedded into this pixel pair. The mean square error (MSE) cause by data embedding using LSBMR is 0.375 [18], which is significantly smaller than that of LSB (0.5). In the same year, Zhang and Wang [24] proposed an exploiting modification direction (EMD) method to enhance the embedding efficiency of LSBMR. EMD embeds a 5-ary digit into a pixel pair but only modifies one pixel one grayscale unit at most. Although EMD provides a better embedding efficiency and lower detectability, the payload is limited to 1.161 bpp at most.

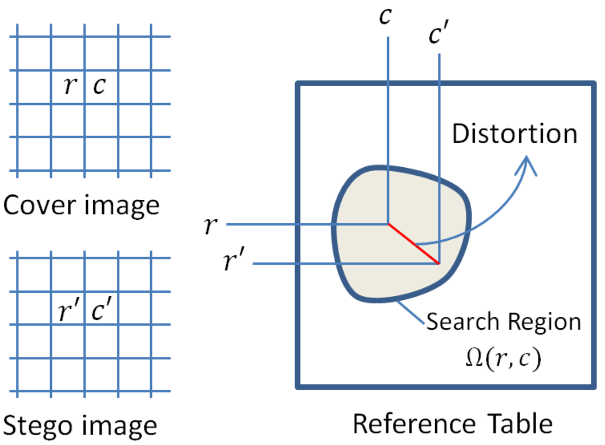


Figure 1: Illustration of the PPM method

In 2008, Chang *et al.* [3] proposed a data hiding method based on solutions of Sudoku tables to increase the payload of EMD. Later, Hong *et al.* [11, 12] modified the searching algorithm of [3] to further increase the image quality by 1.8 dB under the same payload. Chao *et al.* [4] in 2009 proposed a diamond encoding (DE) method with extensible payload. DE embeds a digit in B-ary notational system into a pixel pair, where $B = 2k^2 + 2k + 1$ and k is an integer. When $k = 1$, the embedding performance of DE is equivalent to that of EMD.

In 2012, Hong and Chen [10] proposed an adaptive pixel pair matching (APPM) method to further enhance the embedding performance of PPM based method. Compared to DE method, APPM embeds digits in any notational system but DE embeds digits only in base $2k^2 + 2k + 1$. Since the MSE caused by pixel pair replacement in APPM method are minimized, APPM always achieves a higher image quality under the same payload with lower detectability compared to other PPM-based methods. Although APPM embeds digit in any notational system, it is likely detectable when using RS scheme [7] if 4-ary secret digits are embedded into the image containing larger flat area such cartoon images. Some recent works [5, 8, 20] have exploited the merit of the APPM and developed state-of-the-art works for hiding data. However, the vulnerability to the detection by the RS scheme is yet unsolved.

In this paper, a data hiding method that modified APPM's embedding method by stochastically selecting embeddable positions is proposed. The proposed method not only maintains the same image quality but also provides a smaller detectability than that of APPM method. The rest of this paper is organized as follows. Section 2 is the proposed method, while Section 3 gives the experimental results and discussions. Concluding remarks are given in Section 4.

2 The Proposed Method

In the PPM-based method, the embedding performance is greatly influenced by the design of reference table. In general, a reference table can be constructed by patches or using a formula. APPM use the function

$$R_T(r, c) = (r + c_B \times c) \bmod B$$

to construct the reference table, where c_B is the embedding parameter for B-ary reference table. The reference table used in APPM can be divided into patches (search region) such that the MSE caused by pixel pair replacement is minimized. The embedding parameter used in APPM method for B-ary is listed in Table 1. More details about the obtaining of c_B can be seen in [10].

Although APPM performs the best compared to the existing PPM-based method, it is likely to be detectable by RS scheme [7] when cover images contain large flat area with 4-ary digits are embedded. The flat area in images represents pixels in that area having the similar grayscale values and the cartoon images are often of this type. When APPM are applied on pixels in flat area, APPM's embedding algorithm will confined a search region such that one pixel in a pixel pair will always be added or subtracted by one when embedding a certain 4-ary digit. In this circumstance, it is likely to be detected by the RS scheme.

In this section, a method to secure APPM's embedding method is proposed by evading the RS detection. To do this, if there are two candidates satisfying $R_T(r', c') = d_B$ and are both the nearest to (r, c), then one of them is randomly selected to replace the original pixels. With the aid of random selection, different candidates will be selected and thus the stego image can evade the RS detection. Note that the proposed method can be applied on both the natural images or artificial images such as cartoon images.

2.1 Embedding Procedures

Let I be the cover image of size $M \times M$, and S be the set of B-ary secret digits to be embedded. Firstly, a reference table R_T is constructed according to the extraction function. Then, the pixel pairs in the cover image are scanned and secret digits are embedded into the scanned pixel pairs. The detailed embedding procedures are listed below.

Input: Cover image of size $M \times M$, embedding parameter c_B, and B-ary secret digits S.

Output: Stego image.

Step 1: Construct the reference table R_T for embedding B-ary secret digits using the function $R_T(r, c) = (r + c_B \times c) \bmod B$.

c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	c_{11}	c_{12}	c_{13}	c_{14}	c_{15}	c_{16}	c_{17}	c_{18}
1	1	2	2	2	2	3	3	3	3	4	5	4	4	6	4	4
c_{19}	c_{20}	c_{21}	c_{22}	c_{23}	c_{24}	c_{25}	c_{26}	c_{27}	c_{28}	c_{29}	c_{30}	c_{31}	c_{32}	c_{33}	c_{34}	c_{35}
4	8	4	5	5	5	5	10	5	5	5	12	12	7	6	6	10
c_{36}	c_{37}	c_{38}	c_{39}	c_{40}	c_{41}	c_{42}	c_{43}	c_{44}	c_{45}	c_{46}	c_{47}	c_{48}	c_{49}	c_{50}	c_{51}	c_{52}
15	6	16	7	7	6	12	12	8	7	7	7	7	14	14	9	22
c_{53}	c_{54}	c_{55}	c_{56}	c_{57}	c_{58}	c_{59}	c_{60}	c_{61}	c_{62}	c_{63}	c_{64}					
8	12	21	16	24	22	9	8	8	8	14	14					

Figure 2: Embedding parameter c_B used in APPM

Step 2: Extract a secret digit s_B from S.

Step 3: Scan the pixels in the cover image using the raster scan order. Let the pixels pair (r, c) be the scanned pixels.

Step 4: In the reference table R_T, find all the coordinates (r', c') satisfying $R_T(r', c') = s_B$ and having the smallest L, where $L = (r'-r)^2+(c'-c)^2$. If there are more than one pixel pair satisfying the above two conditions, randomly choose a pair $(\hat{r}', \hat{c}')$ and then replace the original pair (r, c) by $(\hat{r}', \hat{c}')$.

Step 5: Repeat Steps 2–4 until all the secret data are embedded.

2.2 Extraction Procedures

To extract the embedded secret digits, the receiver obtains the information about c_B and B via a secret channel, and then performs the data extraction. The detailed extraction procedures are listed below.

Input: Stego image, the parameter c_B and B.

Output: Secret data S.

Step 1: Construct the reference table R_T which is identical to the one used in the embedding procedure.

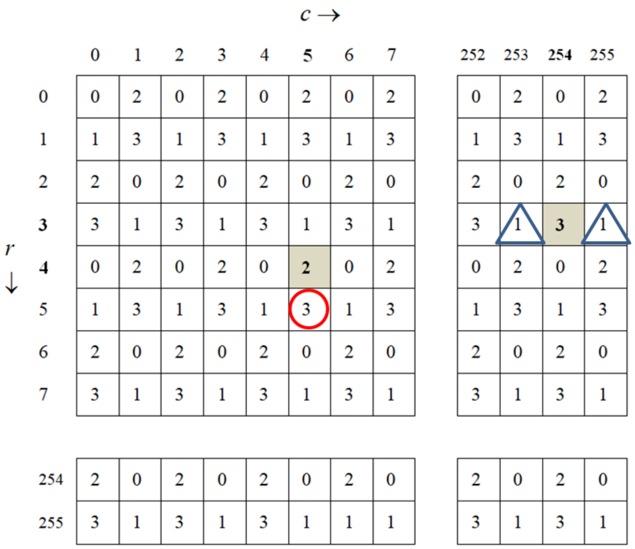
Step 2: Scan the pixel pairs in the stego image using the raster scan order. Let the scanned pixel pair be (r', c'). The embedded secret digit can be easily extracted by using the equation $s_B = R_T(r', c')$

Step 3: Repeat Step 2 until all the secret digits are extracted.

2.3 A Complete Example

In this section, an example is used to illustrate the proposed method. Let $\{(3,254),(4,5)\}$ be a set of cover pixel pairs and two 4-ary digits ($B = 4$) to be embedded into these pixel pair are $S = \{1_4, 3_4\}$. From Figure 2, the embedding base $c_B = 2$ can be obtained. The reference table R_T can be constructed using $f(r, c) = (r + 2 \times c) \bmod 4$. For example, the entity located in the zeroth row and the fifth column is $f(0, 5) = (0 + 2 \times 5) \bmod 4 = 2$, and the entity located in the fifth row and the third column is $f(5, 3) = (5 + 2 \times 3) \bmod 4 = 3$. The constructed table is partially shown in Figure 3. The first scanned cover pixel pair is $(3, 254)$ and the secret data to be embedded is $s_4 = 1_4$. The position located at $(3, 254)$ in the reference table is shaded gray, as shown in Figure 3. Since $(3, 253)$ and $(3, 255)$ (marked by triangles) are both the closest coordinate to $(3, 254)$ and $R_T(3, 253) = R_T(3, 255) = 1_4$, a pixel pair is randomly selected to replace $(3, 254)$. Suppose the selected pair is $(3, 253)$, and thus the pixel pair $(3, 253)$ is used to replace the original pixel pair $(3, 254)$. Next, the second pixel pair $(4, 5)$ is visited and the secret data to be embedded is $s_4 = 3_4$. Since $(5, 5)$ is the closet coordinate (marked by a circle) to $(4, 5)$ while satisfying $R_T(5, 5) = 3_4$, the cover pixel pair $(4, 5)$ is then replaced by $(5, 5)$. Therefore $(r', c') = (5, 5)$ is obtained. As a result, the set of stego pixel pair is $\{(3, 253), (5, 5)\}$.

To extract the embedded digits from the stego pixel pair $(3, 253), (5, 5)$, the reference table is firstly constructed, as in the embedding phase. Because $R_T(3, 253) = 1$, a secret digit $s_4 = 1_4$ is extracted. Similarly, Because $R_T(5, 5) = 3$, a secret digit $s_4 = 3_4$ is extracted.

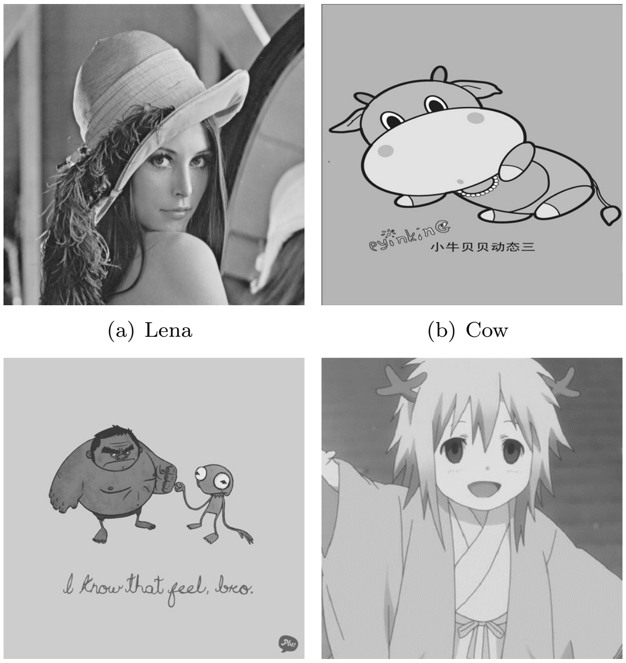


	0	1	2	3	4	5	6	7	252	253	254	255
0	0	2	0	2	0	2	0	2	0	2	0	2
1	1	3	1	3	1	3	1	3	1	3	1	3
2	2	0	2	0	2	0	2	0	2	0	2	0
3	3	1	3	1	3	1	3	1	3	1	**3**	1
4	0	2	0	2	0	**2**	0	2	0	2	0	2
5	1	3	1	3	1	3	1	3	1	3	1	3
6	2	0	2	0	2	0	2	0	2	0	2	0
7	3	1	3	1	3	1	3	1	3	1	3	1
254	2	0	2	0	2	0	2	0	2	0	2	0
255	3	1	3	1	3	1	1	1	3	1	3	1

Figure 3: Reference table used in the example

3 Experimental Results

In this section, several tests are performed to demonstrate the applicability of the proposed method and compared the results with those of the original APPM method. Four 8-bit grayscale test images, including Lean, Cow, Giant, and Kid, each of size 512 × 512, are used in the experiments, as shown in Figure 4. Among these test images, the Lena image is a natural image and others are cartoon images. These cartoon images all contain large flat areas in which pixels and their neighbors have similar grayscale values. The pseudo random number generator is used to generate 4-ary secret digits. The PSNR metric is used to measure the image quality.



(a) Lena (b) Cow

(c) Giant (d) Kid

Figure 4: Four test images

3.1 Image Quality Comparison

In this section, the image qualities obtained by the proposed and the APPM methods are compared. The results are shown in Table 1. Table 1 shows that proposed method does not degrade image quality comparing to that of APPM and still offers a very satisfactory image quality.

Table 1: Comparisons of PSNR

Image	APPM	PPRM
Lena	52.39	52.39
Cow	52.38	52.39
Giant	52.38	52.38
Kid	52.38	52.38

3.2 RS Scheme Steganalysis

In this section, the RS scheme is used to detect the stego images obtained from the proposed method and the APPM method. RS scheme partitions images into groups G of n consecutive pixels, and use a discrimination function and a mask M to classify G into three disjoint groups, namely regular, singular and unusable groups. The ratios of the regular groups R_{+M}, R_{-M} and singular groups S_{+M}, S_{-M} are then calculated. In general, if the LSBs of an image are not embedded, the relationships $R_{+M} \simeq R_{-M}$ and $S_{+M} \simeq S_{-M}$ generally hold. Otherwise, the difference between them will be increased as the embedding rate is increased. In the experiments,

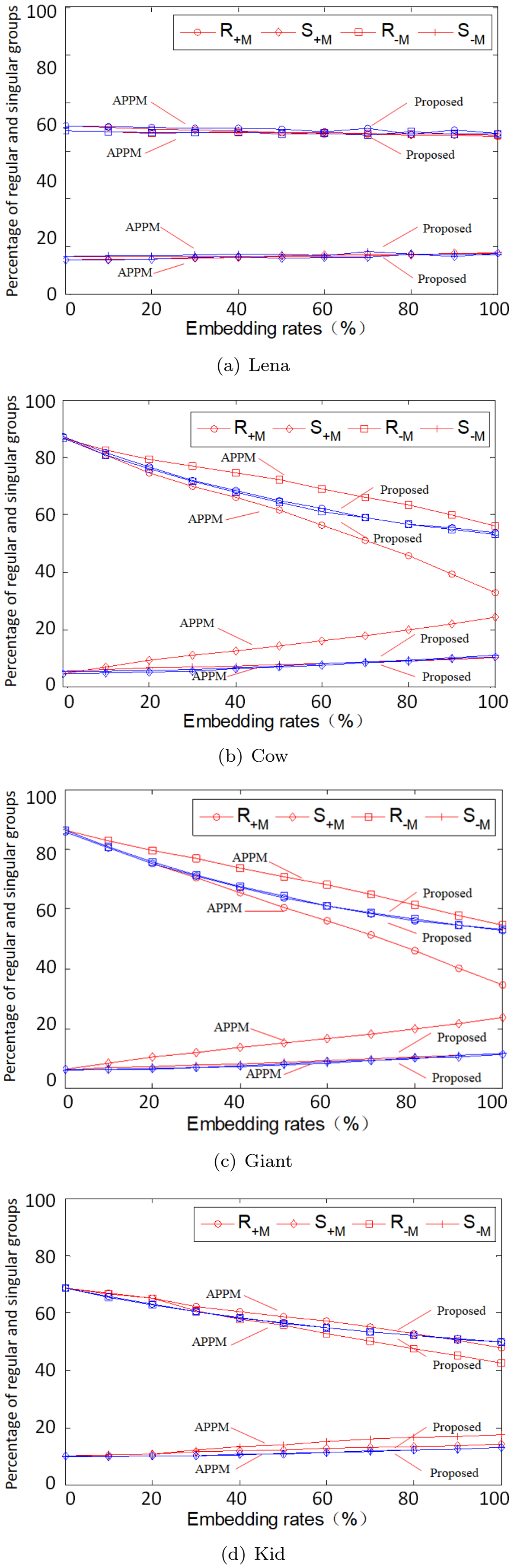
$$M = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

is used as the mask matrix. The results are shown in Figure 5.

Note that for the Lena image, the RS scheme cannot detect the presence of the embedment of the proposed PPRM and the APPM methods because R_{+M} and R_{-M} are indistinguishable and so do S_{+M} and S_{-M}. However, for the test images Cow, Giant and Kid, the differences between R_{+M} , R_{-M} and S_{+M} , S_{-M} increase in the APPM method as the embedding rates increases, indicating that the presence of embedment is more detectable at larger payload. For example, when the embedding rate is 100% (fully embedded), $R_{+M} \simeq 32\%$ and $R_{-M} \simeq 58\%$ are obtained. The difference between them is 26%. The large difference shows that the image is more likely an embedded one. On the other hand, R_{+M} and R_{-M} of the proposed method are both close to 57% even when the embedding rate is 100%, and S_{+M} and S_{-M} also have the similar trends. Therefore, the proposed method is more likely undetectable using the RS scheme. Experiments on other test images also show the similar results, indicating that the proposed method effectively resists the RS attack while providing a very satisfactory image quality.

4 Conclusions

In this paper, a more secure data hiding method by modifying the embedding method of APPM is proposed. During embedding, if the candidates of pixel pairs are more than one, one of them is randomly selected to replace the original pixel pair. The modified pixel pair selection successfully randomizes the replacement to avoid always selecting the same candidates. Compared to the original APPM, the proposed method is undetectable by RS scheme without sacrificing the image quality. The proposed work can be utilized as an embedding method for digital watermarking or image authentication, since the embedding distortion is lower than those of LSB or LSB matching while providing an adjustable payload. The future work will be extended to include more pixels as an embedding unit to conceal data bits while minimizing the distortion.



(a) Lena

(b) Cow

(c) Giant

(d) Kid

Figure 5: RS detection results of four test images

References

[1] J. Bai, C. C. Chang, T. S. Nguyen, C. Zhu, and Y. Liu, "A high payload steganographic algorithm based on edge detection," *Displays*, vol. 46, pp. 42-51, 2017.

[2] C. K. Chan and L. M. Cheng, "Hiding data in images by simple lsb substitution," *Pattern Recognition*, vol. 37, no. 3, pp. 469–474, 2004.

[3] C. C. Chang, Y. C. Chou, and T. D. Kieu, "An information hiding scheme using sudoku," in *3rd International Conference on Innovative Computing Information and Control*, pp. 17–17, June 2008.

[4] R. M. Chao, H. C. Wu, C. C. Lee, and Y. P. Chu, "A novel image data hiding scheme with diamond encoding," *EURASIP Journal on Information Security*, vol. 2009, pp. 658047, May 2009.

[5] J. Chen, "A PVD-based data hiding method with histogram preserving using pixel pair matching," *Signal Processing: Image Communication*, vol. 29, no. 3, pp. 375–384, 2014.

[6] H. Dadgostar and F. Afsari, "Image steganography based on interval-valued intuitionistic fuzzy edge detection and modified lsb," *Journal of Information Security and Applications*, vol. 30, pp. 94-104, 2016.

[7] J. Fridrich, M. Goljan, and R. Du, "Detecting lsb steganography in color, and gray-scale images," *IEEE Multimedia*, vol. 8, pp. 22–28, Oct. 2001.

[8] W. Hong, "Adaptive image data hiding in edges using patched reference table and pair-wise embedding technique," *Information Sciences*, vol. 221, pp. 473–489, 2013.

[9] W. Hong, M. Chen, and T. S. Chen, "An efficient reversible image authentication method using improved pvo and lsb substitution techniques," *Signal Processing: Image Communication*, vol. 58, pp. 111-122, 2017.

[10] W. Hong and T. S. Chen, "A novel data embedding method using adaptive pixel pair matching," *IEEE Transactions on Information Forensics and Security*, vol. 7, pp. 176–184, Feb. 2012.

[11] W. Hong, T. S. Chen, and C. W. Shiu, "A minimal euclidean distance searching technique for sudoku steganography," in *International Symposium on Information Science and Engineering*, vol. 1, pp. 515–518, Dec. 2008.

[12] W. Hong, T. S. Chen, and C. W. Shiu, "Steganography using sudoku revisited," in *Second International Symposium on Intelligent Information Technology Application*, vol. 2, pp. 935–939, Dec. 2008.

[13] L. C. Huang, L. Y. Tseng, 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, Mar. 2013.

[14] B. Jana, "Dual image based reversible data hiding scheme using weighted matrix," *International Journal of Electronics and Information Engineering*, vol. 5, no. 1, pp. 6–19, 2016.

[15] M. Juneja and P. S. Sandhu, "Improved LSB based steganography techniques for color images in spatial domain," *International Journal of Network Security*, vol. 16, pp. 452–462, 2014.

[16] A. D. Ker, "Steganalysis of lsb matching in grayscale images," *IEEE Signal Processing Letters*, vol. 12, no. 6, pp. 441–444, 2005.

[17] S. Manoharan, D. RajKumar, "Pixel value differencing method based on CMYK colour model," *International Journal of Electronics and Information Engineering*, vol. 5, no. 1, pp. 37–46, 2016.

[18] J. Mielikainen, "LSB matching revisited," *IEEE Signal Processing Letters*, vol. 13, pp. 285–287, May 2006.

[19] S. Rajendran and M. Doraipandian, "Chaotic map based random image steganography using LSB technique," *International Journal of Network Security*, vol. 19, pp. 593–598, 2017.

[20] S. Y. Shen and L. H. Huang, "A data hiding scheme using pixel value differencing and improving exploiting modification directions," *Computers & Security*, vol. 48, pp. 131–141, 2015.

[21] Y. L. Wang, J. J. Shen, and M. S. Hwang, "An improved dual image-based reversible hiding technique using lsb matching," *International Journal of Network Security*, vol. 19, pp. 858–862, 2017.

[22] Y. L. Wang, J. J. Shen, M. S. Hwang, "A novel dual image-based high payload reversible hiding technique using LSB matching", *International Journal of Network Security*, vol. 20, no. 4, pp. 801–804, 2018.

[23] H. C. Wu, N. I Wu, C. S. Tsai, M. S. Hwang, "Image steganographic scheme based on pixel-value differencing and LSB replacement methods", *IEE Proceedings Vision, Image and Signal Processing*, vol. 152, no. 5, pp. 611–615, 2005.

[24] X. Zhang and S. Wang, "Efficient steganographic embedding by exploiting modification direction," *IEEE Communications Letters*, vol. 10, pp. 781–783, Nov. 2006.

Biography

Wien Hong received his M.S. and Ph.D. degree from the State University of New York at Buffalo, USA in 1994 and 1997, respectively. He is currently a researcher at Taiwan development Institute, and Nanjing University of Information Science and Technology. He is also a professor in the School of Electrical and Computer Engineering at Nanfang College of Sun Yat-Sen University. His research interests include steganography, watermarking and image compression.

Shuozhen Zheng is a senior researcher and engineer at School of Electrical and Computer Engineering, Nanfang College of Sun Yat-Sen University since 2013. He has incorporate several major projects about digital signal processing (DSP) and applications of embedded system. His research interests include development of single-chip microcomputer, image compression, data hiding, and image authentication.

Xiaoyu Zhou jointed the research team of information security in Nanfang College of Sun Yat-Sen University since 2016. She has independently completed several projects in the field of electrical and computer engineering, including digital watermarking and pattern recognition. Her research interests include computer vision and image authentication.

Secure Multiple-Antenna Ultrawideband System: A Wireless Physical-Layer Security Perspective

Tanit Somleewong and Kiattisak Maichalernnukul
(Corresponding author: Kiattisak Maichalernnukul)

College of Information and Communication Technology, Rangsit University
52/347 Phaholyothin Road, Pathumthani 12000, Thailand
(Email: kiattisak.m@rsu.ac.th)

(Received Oct. 12, 2017; revised and accepted Jan. 10, 2018)

Abstract

Recently, it has been suggested that the cryptographic security of wireless communication systems can be improved by exploiting characteristics of ultrawideband (UWB) signals or spatial diversity in multiple-antenna channels. In this paper, a multiple-antenna prerake UWB system which can achieve robust physical-layer security is proposed. The security performance of the proposed system is analytically evaluated in terms of the probability of an adversary correctly determining a secret key versus the decoding error probability of a legitimate receiver. Numerical results based on a standardized UWB channel model show how the number of antennas and that of prerake fingers affect the security performance.

Keywords: Multiple Antennas; Physical Layer; Prerake; Ultrawideband

1 Introduction

The broadcast nature of wireless channels necessitates securing the message transmission over such medium. While this need can be satisfied by using some kind of powerful encryption algorithms, low-power wireless systems such as radio frequency identification (RFID) systems may not even have enough power and resources to operate them [1,2,7,22,23]. Recent research on communication theory indicates that characteristics of ultrawideband (UWB) signals can be exploited to complement the levels of cryptographic security of wireless systems [5,11]. Specifically, the extremely large bandwidth of UWB signals makes their transmissions more robust to interference than narrow band transmissions. Moreover, since the transmit power of UWB devices is limited by relevant regulatory authorities such as the Federal Communications Commission (FCC) in the USA and the European Commission (EC) in Europe [9], these low-power devices are rather difficult to eavesdrop. UWB signaling such as impulse radio (IR) [24] can also be deliberately designed to achieve some level of encryption at the physical layer.

In the above design, a time-hopping sequence is adopted as a secret parameter for the UWB communication link [5,11]. That is, only a legitimate receiver who knows this sequence can successfully decode the overall message. In evaluating the physical-layer security performance of IR-UWB systems in [5,11], it is assumed that the transmitter, legitimate receiver, and adversary (*i.e.*, eavesdropper) are equipped with a single antenna. On the other hand, it is well known that the use of multiple antennas is capable of achieving spatial diversity. Many works such as [15,21,25,27] have then focused on this deployment for UWB systems, resulting in significant performance improvement. To the best of our knowledge, however, the ability of multiple-antenna IR-UWB systems to support higher-layer cryptographic protocols has been investigated only in [26], where the authors presented a secure space-time coding scheme which uses channel state information (CSI) as the secret key in multiple-antenna links.[1] Unfortunately, an adversary still can employ the blind deconvolution method [6] to estimate its corresponding CSI, making it difficult for this scheme to attain the perfect communication secrecy.

In this paper, an alternative multiple-antenna IR-UWB system which can effectively provide the physical-layer security is proposed. Specifically, the transmitter employs multiple antennas to perform prerake filtering[2] with spatial transmit diversity, leading to temporal focusing (*i.e.*, the received signal is compressed in time domain) as well as spatial focusing (*i.e.*, the received signal is focused on the intended receiver, or more precisely, the legitimate one) [9]. Accordingly, the legitimate receiver who shares a common key with the transmitter requires only a simple matched filter to decode data. The security performance

[1] For the case of multiple-antenna narrow band systems, readers are referred to [10] and references therein.

[2] Another similar pre-filtering technique, called time reversal, has also been used in UWB communications [13,27]. Nevertheless, this technique can be continuous-time processing based on physical waveform recording, and has a wider variety of applications such as electromagnetic imaging [14,19] and underwater acoustic communications [4].

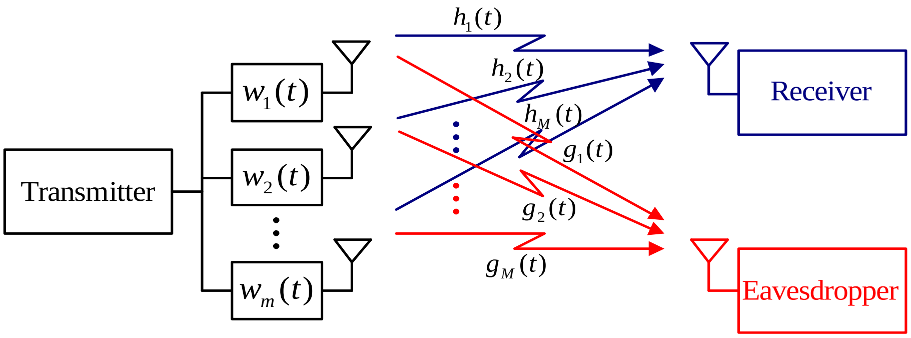


Figure 1: Multiple-antenna prerake UWB system

of the proposed system is also analytically evaluated in terms of the probability of the adversary correctly determining the key versus the decoding error probability of the legitimate receiver.

The rest of the paper is organized as follows. Section 2 describes the system model. In Section 3, the security performance of the multiple-antenna prerake UWB system is analyzed. Section 4 provides the numerical results, followed by the conclusion given in Section 5.

2 System Model

We consider a UWB system where the transmitter is equipped with M antennas, and the legitimate receiver as well as the eavesdropper are equipped with one antenna,[3] as shown in Figure 1. The M transmit antennas are spatially spaced in such a way that the transmitted signals undergo statistically independent channel fading.[4] In general, a UWB link channel can be modeled by the stochastic tapped-delay-line propagation model [9]. The channel impulse response (CIR) for a UWB transmission link from the transmitter to the legitimate receiver is thus described by

$$h_m(t) = \sum_{l=0}^{L_t-1} \alpha_{m,l}\delta(t - lT_p), \qquad (1)$$

where $m \in \{1, 2, \ldots, M\}$ is the index of the transmit antenna, L_t is the number of multipath components, l is the path index, $\alpha_{m,l}$ is the energy-normalized path gain with $\sum_{l=0}^{L_t-1} |\alpha_{m,l}|^2 = 1$, and T_p is the minimum multipath resolution. The minimum T_p is equal to the width of the unit-energy monocycle pulse $p(t)$, since any two paths whose relative delay is less than the pulse width are not resolvable. Similarly, the CIR for a UWB transmission link from the transmitter to the eavesdropper can be written as Equation (1) with $h_m(t)$ and $\alpha_{m,l}$ replaced by $g_m(t)$ and $\beta_{m,l}$, respectively.

As in [11], perfect timing and synchronization among the transmitter, legitimate receiver, and eavesdropper are assumed. Also, we suppose that a randomly generated b-bit secret key K is shared by the transmitter and the legitimate receiver, and it is divided into n parts, *i.e.*, $K = (\kappa_1, \kappa_2, \ldots, \kappa_n)$, to make use of the limited key bits.[5] The transmitter employs a time-hopping method and binary pulse amplitude modulation, and then uses each key part κ_j which consists of b/n bits, $j \in \{1, 2, \ldots, n\}$, to select a position index in $\{0, 1, \ldots, 2^{b/n} - 1\}$ that is shared by the pulses in the corresponding N_f/n frames (see Figure 2). Without loss of generality, we will consider an IR-UWB signal carrying the first binary data bit $b_0 \in \{-1, 1\}$ with equal probability in the first symbol period. To apply prerake filtering with spatial transmit diversity, the channel reciprocity is assumed to be satisfied,[6] and partial CSI of the links between the transmitter and legitimate receiver, *i.e.*, $\{\alpha_{m,l}\}_{m=1,l=0}^{M,L-1}$ with $L < L_t$, is assumed to be known at the transmitter.[7] Therefore, a partial-prerake filter [20] with L taps (also called prerake fingers) can be used at the m-th antenna of the transmitter, and the transmitted signal is represented by

$$s_m(t) = \sqrt{\frac{E_s}{N_f}} \sum_{k=0}^{N_f-1} b_0 z_m(t - kT_f - c_{0,\lfloor \frac{nk}{N_f} \rfloor} T_p), \qquad (2)$$

where E_s is the energy per symbol, N_f is the number of frames in one symbol period (denoted by $T_s := N_f T_f$), T_f is the frame period, $\{c_{0,\lfloor \frac{nk}{N_f} \rfloor}\}_{k=0}^{N_f-1}$ is the time-hopping sequence, with $\lfloor \cdot \rfloor$ denoting the integer floor, and $z_m(t)$ is formed by passing the UWB pulse $p(t)$ through the

[3] The use of multiple antennas at the legitimate receiver and the eavesdropper is beyond the scope of this work and will be considered in our future work.

[4] In practice, such antenna spacing is expected to be on the order of a few ten centimeters [3].

[5] For simplicity, b is assumed to be divisible by n.

[6] The experimental results in [18] show that the reciprocal theorem is indeed valid for a UWB multipath environment.

[7] The reason behind the partial CSI consideration is as follows. In typical indoor UWB environments, the number of multipath components can be on the order of from several tens to a hundred more [16]. From a practical point of view, only a subset of the multipath components can be exploited at the transmitter or receiver side.

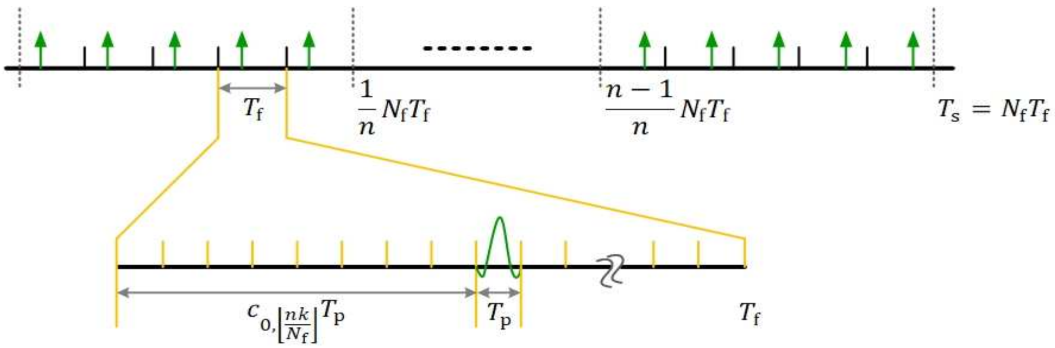


Figure 2: IR-UWB signal with secure time-hopping

partial-prerake filter $w_m(t)$. Mathematically speaking,

$$\begin{aligned} z_m(t) &= w_m(t) * p(t) \\ &= \frac{1}{\sqrt{M\sum_{l=0}^{L-1}|\alpha_{m,l}|^2}} \sum_{l=0}^{L-1} \alpha^*_{m,L-1-l}p(t-lT_p), \end{aligned} \quad (3)$$

where $*$ and $(\cdot)^*$ denote the convolution and complex conjugate, respectively, and the normalization factor $\sqrt{M\sum_{l=0}^{L-1}|\alpha_{m,l}|^2}$ is introduced in order to keep the total power transmitted from the M antennas constant [12]. The purpose of such filtering is to produce a strong peak of the received per-frame signal at the legitimate receiver, and then a matched filter (*i.e.*, single correlator) can be used to receive this path.

3 Performance Analysis

In the following, we will derive the error probabilities of the legitimate receiver and eavesdropper for the above system.

3.1 Legitimate Receiver

The IR-UWB signal received at the legitimate receiver can be expressed as

$$\begin{aligned} r(t) &= \sum_{m=1}^{M} h_m(t) * s_m(t) + u(t) \\ &= \sum_{m=1}^{M}\sum_{l=0}^{L_t-1} \alpha_{m,l}s_m(t-lT_p) + u(t), \end{aligned} \quad (4)$$

where $u(t)$ is the additive white Gaussian noise (AWGN) with zero mean and double-sided power spectral density (PSD) $N_0/2$. Based on Equations (2)-(4), it can be shown that the received signal $r(t)$ includes N_f strong paths (corresponding to N_f frames), and the delay of each relative to the first arrival path is $(L-1)T_p$. To perform matched filtering to these peaks, the legitimate receiver who knows the secrete key K (or more specifically, the time-hopping sequence $\{c_{0,\lfloor \frac{nk}{N_f} \rfloor}\}_{k=0}^{N_f-1}$) generates the template signal $v(t) = \frac{1}{\sqrt{N_f}}\sum_{k=0}^{N_f-1} p(t-kT_f-c_{0,\lfloor \frac{nk}{N_f} \rfloor}T_p)$. Thus, the decision variable for b_0 is given by

$$\begin{aligned} y =& \text{Re}\Bigg(\frac{1}{\sqrt{N_f}}\sum_{k=0}^{N_f-1}\Bigg[\int_{kT_f+c_{0,\lfloor \frac{nk}{N_f} \rfloor}T_p+(L-1)T_p}^{(k+1)T_f+c_{0,\lfloor \frac{nk}{N_f} \rfloor}T_p+(L-1)T_p} r(t) \\ & \times p(t-kT_f-c_{0,\lfloor \frac{nk}{N_f} \rfloor}T_p-(L-1)T_p)\,\mathrm{d}t\Bigg]\Bigg), \end{aligned}$$

where $\text{Re}(\cdot)$ denotes the real part. Following the same procedure as in [8], the bit error probability of the legitimate receiver conditioned on $\{\alpha_{m,l}\}_{m=1,l=0}^{M,L-1}$ is obtained as

$$P_b = Q\left(\sqrt{\frac{2E_s\sum_{m=1}^{M}\sum_{l=0}^{L-1}|\alpha_{m,l}|^2}{MN_0}}\right), \quad (5)$$

where $Q(\cdot)$ denotes the Gaussian Q-function.

3.2 Eavesdropper

The IR-UWB signal received at the eavesdropper can be expressed as

$$\begin{aligned} \tilde{r}(t) &= \sum_{m=1}^{M} g_m(t) * s_m(t) + \tilde{u}(t) \\ &= \sum_{m=1}^{M}\sum_{l=0}^{L_t-1} \beta_{m,l}s_m(t-lT_p) + \tilde{u}(t), \end{aligned} \quad (6)$$

where $\tilde{u}(t)$ is the zero-mean AWGN with the same PSD as $u(t)$. In general, the eavesdropper is not likely to be very close to the legitimate receiver and then the received signal $\tilde{r}(t)$ in Equation (6) tends to be almost immersed in the background noise (see, *e.g.*, Figure 8 in [27], for an illustration). As a result, it is very difficult for the eavesdropper to find the data pulses without knowledge of their locations.

To establish a lower bound on the security performance, we focus on the worse-case scenario where the eavesdropper knows the transmitted data bit [11] in the sequel. Let us consider the first N_f/n frames, where data pulses are located at the identical time slot in each frame. To use the matched filtering technique (similarly to the legitimate receiver), the eavesdropper generates the template signal $\tilde{v}_i(t) = \frac{b_0}{\sqrt{N_\mathrm{f}}} \sum_{k=0}^{N_\mathrm{f}/n-1} p(t-kT_\mathrm{f}-iT_\mathrm{p})$ when the data pulse is in the i-th time slot, $i \in \{0, 1, \ldots, 2^{b/n}-1\}$. The decision statistic for the first N_f/n frames is therefore given by

$$\tilde{y}_i = \mathrm{Re}\Bigg(\frac{b_0}{\sqrt{N_\mathrm{f}}} \sum_{k=0}^{N_\mathrm{f}/n-1} \Bigg[\int_{kT_\mathrm{f}+iT_\mathrm{p}+(L-1)T_\mathrm{p}}^{(k+1)T_\mathrm{f}+iT_\mathrm{p}+(L-1)T_\mathrm{p}} \tilde{r}(t) \times p(t-kT_\mathrm{f}-iT_\mathrm{p}-(L-1)T_\mathrm{p})\, \mathrm{d}t \Bigg] \Bigg).$$

Note in Figure 2 that there are many slots in each frame due to the extreme bandwidth expansion, but only one of them contains a data pulse. Hence, the eavesdropper inevitably has to deploy the template at various delays, and then picks the output with the largest value. Following the approach outlined in [9, Chapter 6], it is straightforward to show that[8]

$$\tilde{y}_i \sim \mathcal{N}(\mu_i, \sigma^2),$$

where

$$\mu_i = \begin{cases} \mathrm{Re}\left(\frac{E_\mathrm{s}}{n} \sum_{m=1}^{M} \frac{\sum_{l=0}^{L+i-c_{0,0}-1} \alpha^*_{m,l-i+c_{0,0}} \beta_{m,l}}{\sqrt{M \sum_{l=0}^{L-1} |\alpha_{m,l}|^2}} \right), \\ \qquad (c_{0,0}-L+1)\, U[c_{0,0}-L+1] \le i \le c_{0,0} \\ \mathrm{Re}\left(\frac{E_\mathrm{s}}{n} \sum_{m=1}^{M} \frac{\sum_{l=i-c_{0,0}}^{L+i-c_{0,0}-1} \alpha^*_{m,l-i+c_{0,0}} \beta_{m,l}}{\sqrt{M \sum_{l=0}^{L-1} |\alpha_{m,l}|^2}} \right), \\ \qquad c_{0,0}+1 \le i \le L_\mathrm{t}-L+c_{0,0} \\ \mathrm{Re}\left(\frac{E_\mathrm{s}}{n} \sum_{m=1}^{M} \frac{\sum_{l=i-c_{0,0}}^{L_\mathrm{t}-1} \alpha^*_{m,l-i+c_{0,0}} \beta_{m,l}}{\sqrt{M \sum_{l=0}^{L-1} |\alpha_{m,l}|^2}} \right), \\ \qquad L_\mathrm{t}-L+c_{0,0}+1 \le i \le L_\mathrm{t}+c_{0,0}-1 \\ 0, \ \text{otherwise}, \end{cases} \tag{7}$$

and

$$\sigma^2 = \frac{N_0}{2}. \tag{8}$$

In Equation (7), $c_{0,0}$ is the actual time-hopping subsequence for the first N_f/n frames, and $U[\,\cdot\,]$ denotes the discrete-time unit step function. Applying the result of optimal detection for orthogonal signaling in [17, Chapter 4], the probability of finding the correct pulse position in the first N_f/n frames conditioned on $\{\alpha^*_{m,l}\}_{m=1,l=0}^{M,L-1}$ and $\{\beta_{m,l}\}_{m=1,l=0}^{M,L_\mathrm{t}-1}$ is obtained as

$$\begin{aligned} &\Pr\{\tilde{y}_i < \tilde{y}_{c_{0,0}}, \forall i \neq c_{0,0}\} \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \left[\prod_{i=0, i\neq c_{0,0}}^{2^{b/n}-1} \left(1 - Q\left(\frac{x-\mu_i}{\sigma}\right)\right)\right] \\ &\quad \times \exp\left(-\frac{(x-\mu_{c_{0,0}})^2}{2\sigma^2}\right) \mathrm{d}x. \end{aligned}$$

Because the time-hopping subsequence for each group of N_f/n frames is independently assigned by the corresponding key part, the conditional probability of error for finding the entire key at the eavesdropper is given by

$$P_\mathrm{e} = 1 - \left(\Pr\{\tilde{y}_i < \tilde{y}_{c_{0,0}}, \forall i \neq c_{0,0}\}\right)^n. \tag{9}$$

4 Numerical Results

A description of the security performance of the proposed system can be obtained by plotting the average probability of the eavesdropper correctly determining the key (*i.e.*, $1 - \bar{P}_\mathrm{e}$) versus the average bit error probability of the legitimate receiver (*i.e.*, $\bar{P}_\mathrm{b}$) on a log-log scale, as shown in Figures 3-5. For these plots, the parameters are set as follows: $b = 30$, $n = 5$, $N_\mathrm{f} = 25$, $T_\mathrm{f} = 400$ ns, and $T_\mathrm{p} = 125$ ps. Furthermore, the received signal-to-noise (SNR) ratio is assumed to be the same at the legitimate receiver and the eavesdropper, while $\bar{P}_\mathrm{b}$ and $\bar{P}_\mathrm{e}$ are obtained, respectively, by averaging Equations (5) and (9) over 10,000 channel realizations generated from one of the IEEE 802.15.4a channel models, namely CM3 for an office line-of-sight environment [16]. The label "ideal" refers to the case in which $L = L_\mathrm{t}$.

Figure 3 compares the security performance of our prerake UWB system and the rake UWB (or more specifically, baseline) system proposed in [11] when both systems use a single transmit antenna ($M = 1$). The results in this figure show that, with the same number of fingers (L), the former system outperforms the latter one. This may be explained, for example, by considering the mean (μ_i) and variance (σ^2) of $\tilde{y}_i$ for the two systems. From Equations (7) and (8), we have $\mu_{c_{0,0}} = \mathrm{Re}\left(\frac{E_\mathrm{s}}{n} \frac{\sum_{l=0}^{L-1} \alpha^*_l \beta_l}{\sqrt{\sum_{l=0}^{L-1} |\alpha_l|^2}}\right)$ and $\sigma^2 = \frac{N_0}{2}$ for the prerake UWB system with $M = 1$,[9] while its rake counterpart (see [11, Section IV-B]) yields $\mu_{c_{0,0}} = \frac{E_\mathrm{s}}{n} \sum_{l=0}^{L-1} |\beta_l|^2$ and $\sigma^2 = \frac{N_0}{2} \sum_{l=0}^{L-1} |\beta_l|^2$. In our simulation trials, we find that the first mean $\mu_{c_{0,0}}$ tends to be less than the second one. Meanwhile, the first variance σ^2 is obviously larger than the second one. For these reasons, the prerake UWB system generally has a lower value of $\Pr\{\tilde{y}_i < \tilde{y}_{c_{0,0}}, \forall i \neq c_{0,0}\}$ and thus a higher probability of error P_e.

Figures 4 and 5 show the security performance of the prerake UWB system with two transmit antennas ($M = 2$) and that with four transmit antennas ($M = 4$), respectively. As seen in these figures, the system performance improves when the number of antennas or the number of fingers is increased. This improvement results from the temporal and spatial focusing described in Section 1. In addition, by varying those two numbers while keeping their product constant, increasing the number of antennas is found to be more beneficial to the system performance than increasing the number of fingers. For example, when the average bit error probability of the legitimate receiver for $M = 1$ and $L = 20$ and that for

[8] To obtain this closed-form expression, we assume that $T_\mathrm{f} > (2L_\mathrm{t} + 2^{b/n} - 1)T_\mathrm{p}$.

[9] For notational simplicity, we omit the antenna index m.

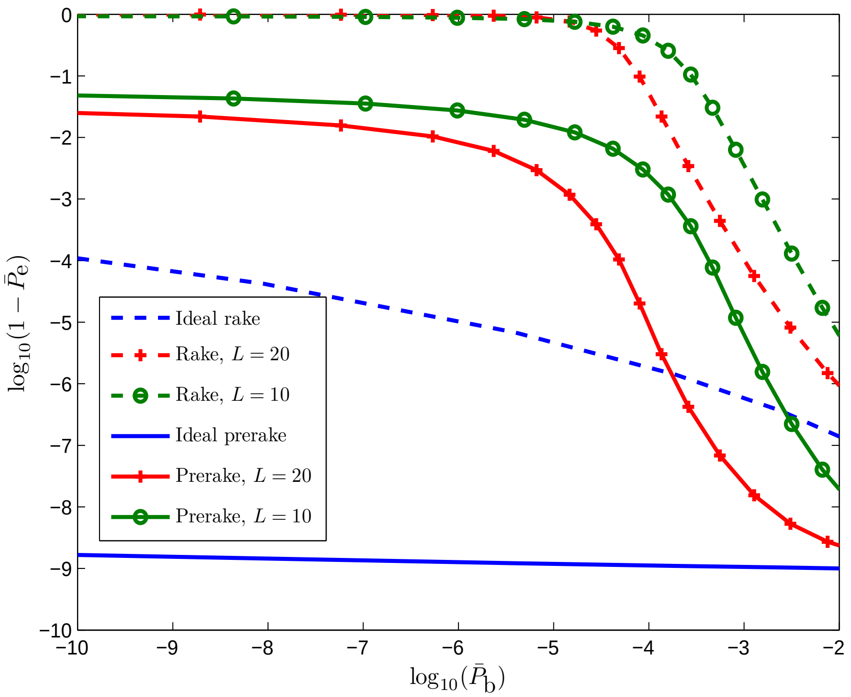


Figure 3: Performance comparison of rake and prerake UWB systems ($M = 1$)

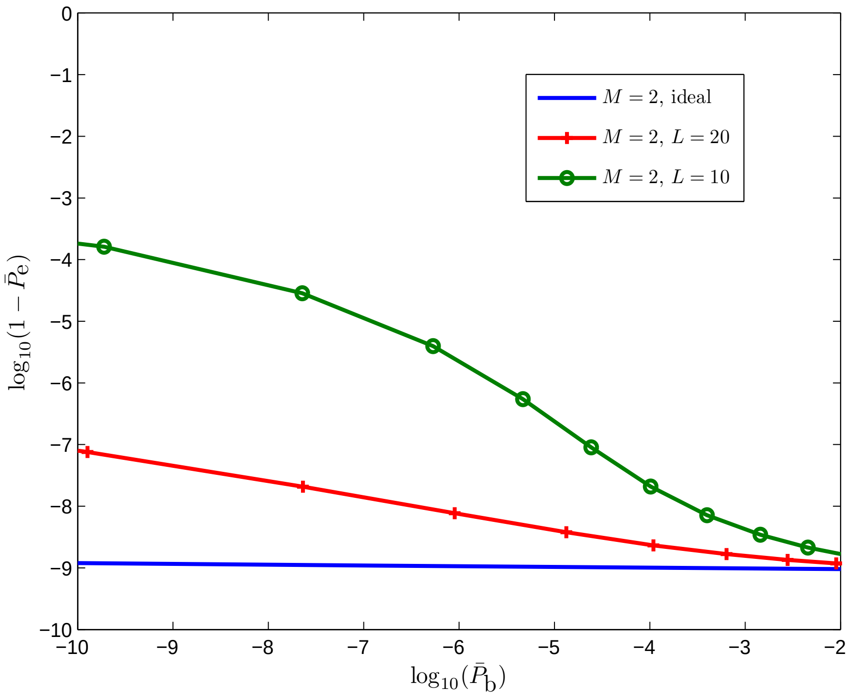


Figure 4: Security performance for different number of prerake fingers ($M = 2$)

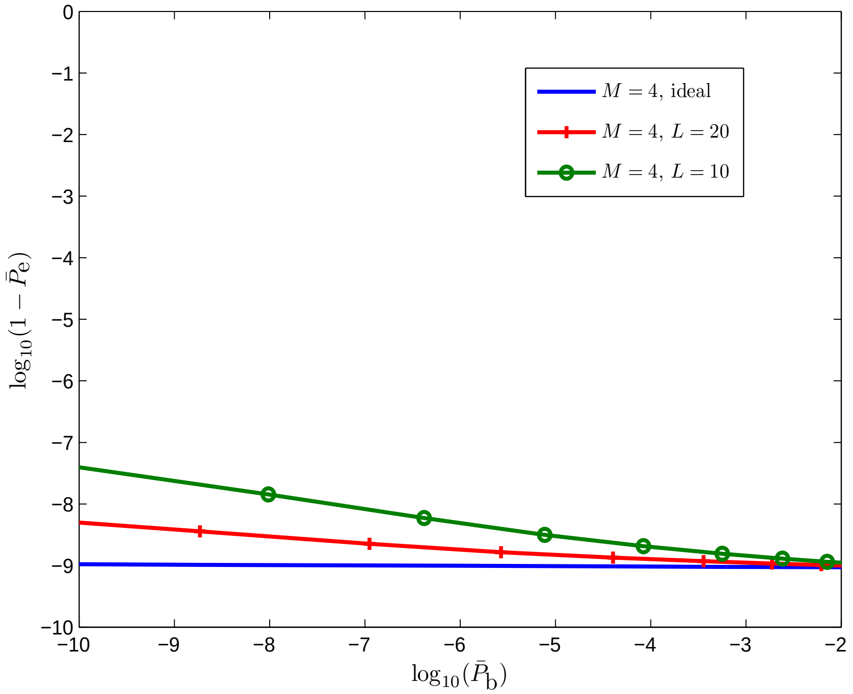


Figure 5: Security performance for different number of prerake fingers ($M = 4$)

$M = 2$ and $L = 10$ are 10^{-4}, the average probabilities of the eavesdroppers correctly determining the key are approximately 10^{-5} and 10^{-8}, respectively. This is due to the fact that the power delay profile of the considered UWB channel model is exponentially decaying [16].

5 Conclusion

We have presented a multiple-antenna prerake UWB system which enables the key-and-location-based security and can satisfactorily thwart the adversary in eavesdropping. The bit error probability of the legitimate receiver and the probability of the adversary finding the correct positions for data pulses have been derived. The performance results have suggested that deploying multiple antennas can save the numbers of prerake fingers, which is required to achieve a high physical-layer security. As our results do not take account of spatial correlation between the transmitter-to-legitimate-receiver and transmitter-to-adversary links, the effect of this correlation on the security performance will be examined in future work.

References

[1] Y. C. Chen, W. L. Wang, and M. S. Hwang, "RFID authentication protocol for anti-counterfeiting and privacy protection," in *Proceedings of The 9th International Conference on Advanced Communication Technology*, pp. 255–259, 2007.

[2] P. Cui, "An improved ownership transfer and mutual authentication for lightweight RFID protocols," *International Journal of Network Security*, vol. 18, no. 6, pp. 1173–1179, 2016.

[3] R. D'Errico, A. Sibille, A. Giorgetti, and M. Chiani, "Antenna diversity in UWB indoor channel," in *Proceedings of IEEE International Conference on Ultra-Wideband*, pp. 13–16, Sep. 2008.

[4] G. F. Edelmann, H. C. Song, S. Kim, W. S. Hodgkiss, W. A. Kuperman, and T. Akal, "Underwater acoustic communications using time reversal," *IEEE Journal of Oceanic Engineering*, vol. 30, no. 4, pp. 852–864, 2005.

[5] D. S. Ha and P. R. Schaumont, "Replacing cryptography with ultra wideband (UWB) modulation in secure RFID," in *Proceedings of The First IEEE Conference on Radio Frequency Identification*, pp. 23–29, Mar. 2007.

[6] Y. Hua, S. An, and Y. Xiang, "Blind identification of FIR, MIMO channels by decorrelating subchannels," *IEEE Transactions on Signal Processing*, vol. 51, no. 5, pp. 1143–1155, 2003.

[7] A. Juels and S. A. Weis, "Authenticating pervasive devices with human protocols," in *Proceedings of The 25th Annual International Cryptology Conference*, pp. 293–308, Aug. 2005.

[8] M. Jun and T. Oh, "Performance of pre-rake combining time hopping UWB system," *IEEE Transactions*

on Consumer Electronics, vol. 50, no. 4, pp. 1033–1037, 2004.

[9] T. Kaiser and F. Zheng, *Ultra Wideband Systems with MIMO*, 2010. ISBN:9780470740019.

[10] J. Kitchen, *On MIMO Wireless Eavesdrop Information Rates*, 2011. (`http://hdl.handle.net/11343/36890`)

[11] M. Ko and D. L. Goeckel, "Wireless physical-layer security performance of UWB systems," in *Proceedings of IEEE Military Communications Conference*, pp. 2143–2148, Nov. 2010.

[12] P. Kyritsi, G. Papanicolaou, P. Eggers, and A. Oprea, "MISO time reversal and delay-spread compression for FWA channels at 5 GHz," *IEEE Transactions on Antennas and Propagation*, vol. 3, no. 11, pp. 96–99, 2004.

[13] X. Liu, B. Z. Wang, S. Xiao, and J. Deng, "Performance of impulse radio UWB communications based on time reversal technique," *Progress in Electromagnetics Research*, vol. 79, pp. 401–413, 2008.

[14] N. Maaref, P. Millot, X. Ferrieres, C. Pichot, and O. Picon, "Electromagnetic imaging method based on time reversal processing applied to through-the-wall target localization," *Progress in Electromagnetics Research M*, vol. 1, pp. 59–67, 2008.

[15] K. Maichalernnukul, T. Kaiser, and F. Zheng, "On the performance of coherent and noncoherent UWB detection systems using a relay with multiple antennas," *IEEE Transactions on Wireless Communications*, vol. 8, no. 7, pp. 3407–3414, 2009.

[16] A. F. Molisch, D. Cassioli, C. C. Chong, S. Emami, A. Fort, B. Kannan, J. Karedal, J. Kunisch, H. G. Schantz, K. Siwiak, and M. Z. Win, "A comprehensive standardized model for ultrawideband propagation channels," *IEEE Transactions on Antennas and Propagation*, vol. 54, no. 11, pp. 3151–3166, 2006.

[17] J. G. Proakis and M. Salehi, *Digital Communications*, 2008. (`https://www.amazon.com/Digital-Communications-5th-John-Proakis/dp/0072957166`)

[18] R. C. Qiu, C. Zhou, N. Guo, and J. Q. Zhang, "Time reversal with MISO for ultrawideband communications: Experimental results," *IEEE Antennas and Wireless Propagation Letters*, vol. 5, no. 1, pp. 269–273, 2006.

[19] A. B. Ruffin, J. Van Rudd, J. Decker, L. Sanchez-Palencia, L. Le Hors, J. F. Whitaker, and T. B. Norris, "Time reversal terahertz imaging," *IEEE Journal of Quantum Electronics*, vol. 38, no. 8, pp. 1110–1119, 2002.

[20] K. Usuda, H. Zhang, and M. Nakagawa, "Pre-rake performance for pulse based UWB system in a standardized UWB short-range channel," in *Proceedings of IEEE Wireless Communications and Networking Conference*, pp. 920–925, Mar. 2004.

[21] L. C. Wang, W. C. Liu, and K. J. Shieh, "On the performance of using multiple transmit and receive antennas in pulse-based ultrawideband systems," *IEEE Transactions on Wireless Communications*, vol. 4, no. 6, pp. 2738–2750, 2005.

[22] C. H. Wei, M. S. Hwang, and A. Y. H. Chin, "A mutual authentication protocol for RFID," *IEEE IT Professional*, vol. 13, no. 2, pp. 20–24, Mar. 2011.

[23] C. H. Wei, M. S. Hwang, and A. Y. H. Chin, "A secure privacy and authentication protocol for passive RFID tags," *International Journal of Mobile Communications*, vol. 15, no. 3, pp. 266–277, 2017.

[24] M. Z. Win and R. A. Scholtz, "Impulse radio: How it works," *IEEE Communications Letters*, vol. 2, no. 2, pp. 36–38, 1998.

[25] L. Yang and G. B. Giannakis, "Analog space-time coding for multiantenna ultra-wideband transmissions," *IEEE Transactions on Communications*, vol. 52, no. 3, pp. 507–517, 2004.

[26] Y. Zhang and H. Dai, "A real orthogonal space-time coded UWB scheme for wireless secure communications," *EURASIP Journal on Wireless Communications and Networking*, vol. 2009, pp. 1–8, 2009.

[27] C. Zhou, N. Guo, and R. C. Qiu, "Time reversed ultra-wideband (UWB) multiple-input multiple-output (MIMO) based on measured spatial channels," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 6, pp. 2884–2898, 2009.

Biography

Tanit Somleewong received the B.E. degree from Mahanakorn University of Technology, Thailand, in 1997 and the M.S. degree from Walailak University, Thailand, in 2004. He is currently working toward the Ph.D. degree with College of Information and Communication Technology, Rangsit University, Thailand.

Kiattisak Maichalernnukul received the Dr.-Ing. degree (summa cum laude) in electrical engineering from University of Hannover, Germany, in 2010. From 2005 to 2006, he was a Research Assistant with the National Electronics and Computer Technology Center, Thailand. From 2007 to 2011, he was a Scientific Assistant with the Institute of Communications Technology, University of Hannover. Since July 2011, he has been a Lecturer with Rangsit University, Thailand. His research interests are in the areas of communication theory and signal processing for wireless communications. He received an Information Technology Society (ITG) Dissertation Prize from the German Association for Electrical, Electronic, and Information Technologies (VDE) in 2011 and a Young Scientist Award from the URSI Asia-Pacific Radio Science Conference (AP-RASC) in 2013.

A Secure and Efficient Data Aggregation Scheme for Cloud-Assisted Wireless Body Area Network

Huaijin Liu, Yonghong Chen, Hui Tian, and Tian Wang
(Corresponding author: Huaijin Liu)

Huaqiao University, College of Computer Science and Technology
Xiamen 361021, China
(Email: lhjhqdx@163.com)

(Received Oct. 10, 2017; revised and accepted Mar. 27, 2018)

Abstract

The development of healthcare system has been greatly facilitated by the use of cloud-assisted wireless body area network (WBAN), which provides a more convenient and intelligent medical service for the users. However, how to establish a secure channel between WBAN and the cloud service and efficient transmission of WBAN data to the cloud service is a great challenge. In this paper, we propose a secure and efficient data aggregation scheme for cloud-assisted wireless body area network. First, we use the privacy homomorphism technique to encrypt the user data, so that the aggregation of data without decryption, to ensure the security and privacy of user data. Then use the base station to help users forward data to the cloud service and allow users to select the best relay node according to the proposed greedy forwarding model, improve the transmission efficiency of user data. The security analysis and experimental results show that the proposed scheme has high security and lower loss ratio, smaller delay and less energy consumption.

Keywords: Aggregation; Cloud-assisted WBAN; Privacy Homomorphism; Wireless Body Area Network

1 Introduction

Wireless body area network (WBAN) appears very promising for healthcare service system, as if can monitor the user's physiological parameters in a timely manner, leading to enhanced efficiency of medical services. Due to the limited computing and storage resources of WBAN, a cloud service to help deal with and store large amounts of user data can provide users with more reliable and intelligent healthcare services [15, 16]. However, there lies a challenge in designing cloud services to be combined with WBANs. The main challenge is how to ensure user identity and data privacy while improving the efficiency of user data transmission. On the one hand, in order to ensure the security and privacy of user data, the literature [2, 9, 13] proposes to establish secure communication between users and cloud services through bilinear mapping, reducing key management and storage overhead. The literature [5–7] proposed the use of chaotic public key cryptography to encrypt user data against external and internal attacks. However, these schemes are not suitable for WBAN with limited computing and storage capacity. In order to reduce the computation and communication overhead, the literature [3, 4, 12] proposed to use the time-varying human physiological signal to establish secure channel. But this method is limited to the symmetric network topology. On the other hand, in order to improve the transmission efficiency of user data, Liang *et al.* [8] proposed a privacy-preserving emergency call scheme PEC. The scheme protects the security of healthcare service system through an attribute-based ciphertext strategy and to transmit the emergency data to the cloud service by broadcasting. Although it can resist cloud service compromise attacks and reduce the transmission delay, the scheme has a large energy consumption. Chen *et al.* [1] proposed a privacy-preserving data aggregation scheme to reduce the communication overhead of the whole system. However, the scheme can not resist compromise attacks from users or cloud servers. Subsequently, Zhang *et al.* [14] proposed a priority-based data aggregation scheme PHDA for cloud-assisted WBAN. The scheme encrypts the user data through the Paillier public key cryptography and uses the base station to help the user forward data to the cloud service, reducing the delay and increasing the packet arrival rate. However, the scheme has a large computational complexity and can not resist compromise attacks. Therefore, how to protect the user data security while maintaining the efficient transmission of data on cloud assisted WBAN is still an important challenge.

In this paper, our goal is to design a secure and efficient cloud-assisted WBAN to address secure communication and efficient transmission issues. First, we propose a lightweight data aggregation scheme that encrypts data through privacy homomorphic technology, making the cloud service aggregate data without decrypting and

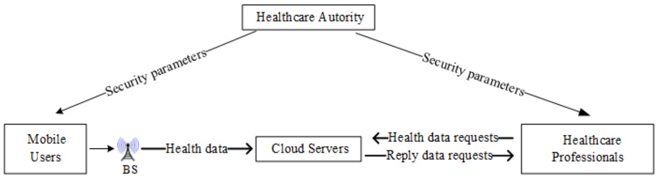


Figure 1: A generic telemedicine service architecture

ensuring data confidentiality. Second, in order to ensure the integrity of the data, we aggregate user data labels for batch authentication, reducing the overhead of user data authentication. Finally, in order to improve the transmission efficiency of user data, we propose a greedy forwarding model to forward user data. The performance evaluation shows that our scheme can meet the requirement of delivery rate and delay, while lower energy is consumed.

2 Network Architecture and Attack Model

2.1 Network Architecture

In this section, we present a generic telemedicine service architecture, as shown in Figure 1. The architecture consists of three main components:

1) The WBAN which collects user health data;

2) The cloud service which allow medical professionals to access to stored data;

3) The healthcare authority which designate and enforce security policies.

First, the healthcare organization generates and sends his security parameters to each user and medical personnel, which is used to enforce the security policy of the medical institution. Then, WBAN collects the user's health data and uploads it to the cloud service through the base station. The cloud service to the user's data aggregation, storage and delivery to the medical personnel for diagnosis and analysis.

2.2 Greedy Forwarding Model

In WBAN, we need to consider the low power requirements of the user equipment. When the user equipment and the base station communication distance is relatively large, the use of multi-hop relay mode can reduce the total power consumption of WBAN. Then, in order to reduce the number of hops between the user equipment and the base station, we present a greedy forwarding model, as shown in Figure 2. In this model, the sending node selects the node closest to the destination node in the communication range as the next hop forwarding node.

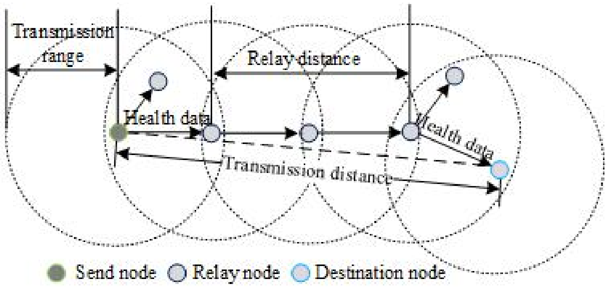


Figure 2: A greedy forwarding model

2.3 Attack Model

Attackers may exist in the network and launched attacks to threaten the user's identity and privacy data, reduce network performance. In addition to eavesdropping, intercepting all network transmission messages, it is possible to replay the previous legitimate messages or fake legitimate users to send false messages to the base station. In addition, the user equipment and the base station are semi-trusted, and attackers are likely to launch a compromise attack on base stations and user equipment, access to the corresponding secret information to cause a greater threat.

3 The Proposed Scheme

In this section, we present a secure and efficient data aggregation scheme (SEDA). SEDA involves privacy homology [10] basic technology, in order to facilitate the later description, we first briefly introduce it.

3.1 Basic Technology

Privacy homomorphic technology does not require the aggregator to decrypt the received privacy data and can directly perform the aggregation operation. Its main principle is: take a small integer $d \geq 2$ and a large integer g as the public key, take a small divisor $g^{'}$ of g and $r \epsilon Z_g$ as the secret key; Randomly divide the message m into d parts $m_1, \cdots, m_d$, satisfy $m = \sum_{i=1}^{d} m_i mod g^{'}$, compute $E_k(m) = (m_1 r mod g, m_2 r^2 mod g, \cdots, m_d r^d mod g)$; Compute the $i-th$ coordinate by $r^{-i} mod g$ to obtain $m_i mod g$, and compute $D_k\left(m\right) = \sum_{i=1}^{d} m_i mod g^{'}$ restore m.

3.2 SEAD Scheme

This scheme is divided into four phases: initialization phase, data encryption phase, data aggregation phase and data decryption phase. The verification work is carried out in the data decryption phase.

3.2.1 Initialization Phase

The healthcare authority TA first selects a positive integer $d \geq 2$ and a large integer g as the public key, and selects a small divisor $g^{'}$ of g and $r\epsilon Z_g$ as the private key. Then randomly generates a shared key SK_i for each user u_i and selects two one-way function: $H_1 : \{0,1\}^* \times Z_g \to Z_g$, $H_2 : Z_g \to Z_g$. Finally, TA preloads $\left\{r, g^{'}, SK_i, H_1, H_2\right\}$ into u_i.

3.2.2 Data Encryption Phase

The user u_i needs to encrypt m_i before uploading the physiological data m_i to base station BS. Before u_i encrypts, first it needs to calculate seed mask value $r_i^1 = H_1\left(ID_{u_i} \parallel SK_i\right)$, and then calculate $r_i^u = H_2\left(r_i^{u-1}\right)$ before each encryption. Using the seed mask value r_i^u to mask the encrypted data m_i : $\widehat{m}_i = \left(m_i + r_i^u\right) modg$, and then divide $\widehat{m}_i$ into d part $m_{i1}, \cdots, m_{id}$, satisfy $\widehat{m}_i = \sum_{j=1}^{d} m_{ij} modg^{'}$. Calculate C_i:

$$\begin{aligned} C_i &= [C_{i1}, C_{i2}, \cdots, C_{id}] \\ &= \left[m_{i1} rmodg, m_{i2} r^2 modg, \cdots, m_{id} r^d modg\right] \end{aligned} \tag{1}$$

and calculate the label $dgt_i^u = \left(r_i^u + sk_i\right) modg$. Then forwards $C_i \parallel dgt_i^u$ to the nearby BS through the greedy forwarding model.

3.2.3 Data Aggregation Phase

When the cloud service receives the n messages sent by BS in the time period t, it needs to aggregate the n messages:

$$\begin{aligned} C_{12\cdots n} = & \sum_{i=1}^{n} C_i = \left[\sum_{i=1}^{n} C_{i1}, \cdots, \sum_{i=1}^{n} C_{id}\right] \\ = & \left[\sum_{i=1}^{n} m_{i1} rmodg, \cdots, \sum_{i=1}^{n} m_{id} r^d modg\right] \end{aligned} \tag{2}$$

$$Dgt_i^u = \sum_{i=1}^{n} dgt_i^u \tag{3}$$

and then send $C_{12\cdots n} \parallel Dgt_i^u$ to the medical personnel.

3.2.4 Data Decryption Phase

After receiving the aggregated data $C_{12\cdots n} \parallel Dgt_i^u$, the medical personnel calculate $r_i^u = H_2\left(r_i^{u-1}\right), i = 1, 2, \cdots, n$, and then verifies:

$$\left(Dgt_i^u - \sum_{i=1}^{n} sk_i\right) modg \overset{?}{=} \sum_{i=1}^{n} r_i^u \tag{4}$$

If equal, decrypt $C_{12\cdots n}$ to obtain the aggregated data m:

$$m = \left(\sum_{i=1}^{n} C_{i1} r^{-1} + \cdots + \sum_{i=1}^{n} C_{id} r^{-d} - \sum_{i=1}^{n} r_i^u\right) modg^{'} \tag{5}$$

4 Security Analysis and Proof

In this section, we discuss the security performance of our proposed SEAD scheme. We focus on the attack model in Section 2.3.

Theorem 1. *The proposed scheme can resist eavesdropping/tampering attacks.*

Proof. In our scheme, user data needs to be encrypted and signed before uploading. The attacker can not obtain the user privacy data and tamper with the communication data without knowing the private key SK_i and r of the user and the medical personnel. Therefore, our scheme can resist eavesdropping/tampering attacks. □

Theorem 2. *The proposed scheme can resist user compromise attacks.*

Proof. Our scheme involves two types of secret keys: SK_i and r, SK_i is the shared key between the user u_i and the medical personnel, and r is the shared key of all users and the medical personnel. The attacker compromises one or some network users to obtain the secret key r, it also can not get other user's privacy information. Because the compromised user can not obtain the shared key SK_i of the uncompromised user u_i and the medical personnel. Therefore, our scheme can resist user compromise attacks. □

Theorem 3. *The proposed scheme can resist cloud service compromise attacks.*

Proof. In our scheme, the attacker compromises that the cloud service can not obtain the user's privacy data. Because the cloud service is only responsible for aggregating user data, there is no shared key SK_i between the user and the medical personnel, can not decrypt the user data. Therefore, our scheme can resist cloud service compromise attacks. □

Theorem 4. *The proposed scheme can resist replay attacks.*

Proof. In our scheme, the user will send $C_i \parallel dgt_i^u$ to the base station every time, and the label $dgt_i^u = \left(r_i^u + sk_i\right) modg$ will be updated by updating $r_i^u = H_2\left(r_i^{u-1}\right)$. If the attacker replays the previous interactive message, it will not be able to pass the detection of the medical personnel. Therefore, our scheme can resist replay attacks. □

Theorem 5. *The proposed scheme can provide forward security.*

Proof. In our scheme, the user needs to calculate the mask value $r_i^u = H_2\left(r_i^{u-1}\right)$ before each encryption, and then delete r_i^{u-1}. So even if the attacker compromise the user, can only get the current r_i^u and can not get r_i^{u-1} of the previous time period. Therefore, our scheme can guarantee forward security. □

Table 1: Computational overhead of the three schemes

Scheme	**Individual user**	**Cloud server**	**Medical personnel**
PHDA	$6T_{exp} + 3T_{mul}$	$(2n+3)\,T_p + nT_{exp} + (2n+1)\,T_{mul}$	$2T_p + T_{exp}$
MuDA	$2T_{exp} + T_{mul}$	$(n-1)\,T_{mul}$	$2T_{exp} + 2T_{plm}$
SEDA	$1T_h + 2T_{mad} + dT_{mul}$	$d\,(n-1)\,T_{mad}$	$nT_h + 2T_{mad}$

Table 2: Simulation parameters

Parameters	**Values**
Size of simulation area/m^2	100×100
Number of mobile nodes	$20, 40, 60, \cdots, 200$
Number of base stations	*1*
Mobile node communication range /m	*50*
Mobile node average velocity /m/s	*1, 2*
Initial energy /mJ	*1000*
MAC layer protocol	*802.11*
Channel bandwidth /Mbs	*11*
Simulation time /s	*100*

5 Performance Evaluation

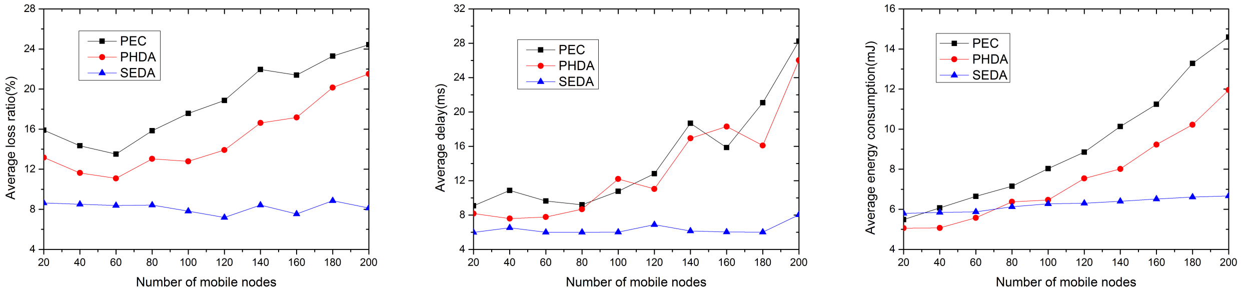
5.1 Computing Complexity

We compare the computational complexity of SEDA with the typical MuDA [1] and PHDA [14] for privacy preserving data aggregation schemes. The computational overhead of each scheme is considered from the following three aspects: the computational cost of a single mobile user, the computational overhead of the cloud service, and the computational overhead of the medical personnel. For SEDA, each mobile user u_i needs to perform 1 hash operation, 2 modulo addition operations and d modular multiplication operations for its privacy data encryption and signature. In the data aggregation phase, the cloud server calculates an encrypted aggregation operation to obtain $C_{12\cdots n}$ need to perform $d(n-1)$ modular addition operation. The medical personnel to verify the signature and decryption data need to n hash operations and 2 modular operations. For PHDA, each mobile user u_i encrypts and signs its health data with 6 exponential modular operations and 3 modular multiplication operations. The cloud service verifies that this received health data signature and aggregation health data requires $(2n+3)$ bilinear pair operations, n exponential exponentiation operations and $(2n+1)$ modular multiplication operations. The medical personnel verify that signature and decryption data requires 2 bilinear pair operations and 1 exponential modular operation. For MuDA, each mobile user u_i encrypts its privacy data requires 2 exponential operations and 1 modular operation. In the data aggregation phase, the cloud server computes an encrypted aggregation operation requires $(n-1)$ modular multiplication operations. The medical personnel decrypts the aggregated health data with 2 bilinear pair operations and 1 discrete logarithmic operation.

The computational complexity of the three schemes is shown in Table 2. Where T_{exp} represents the computational overhead required for exponential modular operation in Z_{N^2}, T_{mul} represents the computational overhead required for modular multiplication operation in $\mathbb{G}$, T_{mad} represents the computational overhead required for modular addition operation in $\mathbb{G}$, T_h represents the computational overhead required for the hash operation, T_p represents the computational overhead required for bilinear pairing operations, T_{plm} represents the computational overhead required to calculate discrete logarithms using Pollard's Lambda method, and n represents the total number of mobile users within the network. As can be seen from Table 1, our scheme is significantly less than the other two schemes, because the bilinear pairing operation and exponential modular operation need to spend much more than the modular operation.

5.2 Simulation Settings

Our simulations are performed in NS-2 [11]. Two main experiments are performed to evaluate the performance of the proposed scheme. In the first experiment, the moving speed of the mobile node was set to 1 m/s. In the second experiment, the moving speed of the mobile node is set to 2 m/s. In all simulation experiment, the mobile nodes are randomly deployed in a $100 \times 100m^2$ monitoring area, the base station node is located in the center of the area. Table 2 shows some basic parameter settings in the simulation. In order to evaluate the transmission efficiency of SEDA health data, there are three different



(a) Loss ratio vs. Number of mobile nodes (b) Delay vs. Number of mobile nodes (c) Energy consumption vs. Number of mobile nodes

Figure 3: The user movement speed is 1m/s

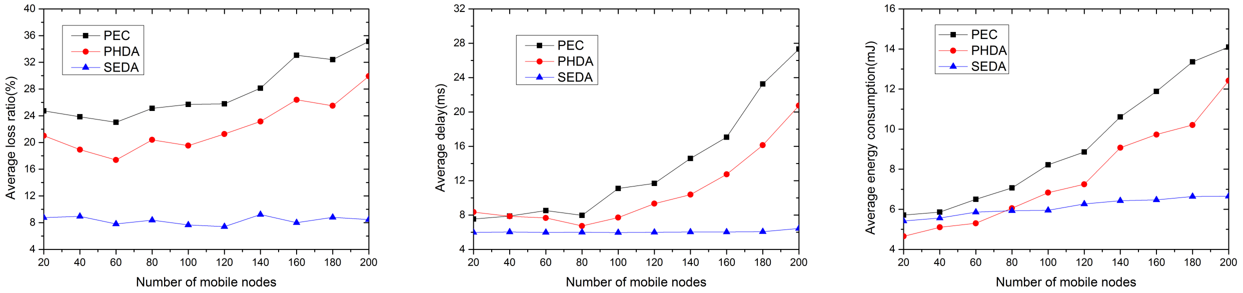
data transmission solutions are considered. The first is PEC [8], a traditional solution of relying on all neighbor nodes to forward the user data. The second is PHDA [14], a method that relies on base stations and any neighbor nodes to forward the user data. The third is our proposed SEDA solution, which relies on the base station and the optimal neighbor nodes to forward the user data.

The performance metrics that we use in simulation experiments are the packet loss ratio, transmission delay and energy consumption. The average packet loss ratio (LR) is defined as $LR = \frac{1}{n-1}\sum_{i=0}^{n-1}\left(M_{AGTs}^{i} - M_{AGTr}^{i}\right)/M_{AGTs}^{i}$, where n represents the number of mobile nodes,M_{AGTs}^{i} represents the total number of the mobile node u_i sends packets of *cbr* (constants bit rate) data flow in the application layer (AGT), and M_{AGTr}^{i} represents the total number of u_i received packets of *cbr* data flow in AGT. The average packet delay (PD) is defined as $PD = \frac{1}{N+1}\sum_{i=0}^{N}\left(T_r^i - T_s^i\right)$, where N represents BS in the AGT layer to receive the total number of packets of *cbr* data flow, T_r^i represents BS receives the $i-th$ packet time, and T_s^i represents u_i sends the $i-th$ packet time. The average energy consumption (EC) is defined as $EC = \frac{1}{n}\sum_{i=0}^{n}\left(E_{init}^{i} - E_{res}^{i}\right)$, where n represents the number of mobile nodes, E_{init}^{i} represents the initial energy value of u_i, and E_{res}^{i} represents the residual energy value of u_i at the end of simulation.

5.3 Simulation Results

In Figure 3, the moving speed of the mobile node is set to 1 m/s. Figure 3(a) shows the relationship between the average packet loss ratio and the number of mobile nodes. As can be seen from the figure, with the increase of the number of mobile nodes, PEC and PHDA packet loss first decreases then increases gradually, because congestion occurs when the area of the coverage area of the mobile node is expanded to a certain extent. SEDA has a low packet loss ratio compared to PEC and PHDA, when the number of mobile nodes is 200, the average loss ratio of SEDA is 66.74% less than PEC and 62.25% less than PHDA. This is because SEDA chooses the best node as the next hop forwarding node, which is not affected by the number of mobile nodes. Figure 3(b) shows the relationship between the average delay and the number of mobile nodes. It can be seen that the average delay of PEC and PHDA is increasing with the number of mobile nodes increasing, while the average delay of SEDA is almost constant. Specifically, when the number of mobile nodes is 200, the average delay of SEDA is 76.4% and 68.93% less than that of PEC and PHDA respectively. This is because the number of forwarding hops for PEC and PHDA is increasing as the number of mobile nodes increases, and SEDA selects the best node as the next hop node, and the hop count does not increase with the number of mobile nodes. Figure 3(c) shows the relationship between the number of mobile nodes and the average energy consumption. As can be seen from the figure, SEDA has a lower energy consumption compared to the other two schemes. Specifically, when the number of mobile nodes is 200, the average energy consumption of SEDA is 54.34% less than PEC and 46.43% less than PHDA. This is because SEDA selects the nearest node to the the target node for forwarding, thereby reducing energy consumption.

In Figure 4, the moving speed of the mobile node is set to 2 m/s. Figure 4(a) shows the relationship between the average loss ratio and the number of mobile nodes. It can be seen from the figure that SCDA has a low packet loss ratio compared to PEC and PHDA, when the number of mobile nodes is 200, the average loss ratio of SEDA is 75.9% less than PEC and 71.71% less than PHDA. As mentioned earlier, this is because SEDA chooses the best node as the next hop node, thereby reducing the number of forwarding hops. Figure 4(b) shows the relationship between the average delay and the number of mobile nodes. This graph shows that SEDA has a low transmission delay compared to the other two schemes. When the number of mobile nodes is 200, the average delay of SEDA is 70.55% and 68.06% less than that of PEC and PHDA. Because SEDA chooses the optimal node as the next hop node, thereby reducing the transmission delay.



(a) Loss ratio vs. Number of mobile nodes

(b) Delay vs. Number of mobile nodes

(c) Energy consumption vs. Number of mobile nodes

Figure 4: The user movement speed is 2m/s

Table 3: To-be tested audio files

Simulation	Avg. packets loss ratio	Avg. Delay	Avg. Energy consumption
Experiment 1	66.74% and 62.25%	76.4% and 68.93%	54.34% and 46.43%
Experiment 2	75.9% and 71.71%	70.55% and 68.06%	52.8% and 44.25%

Figure 4(c) shows the relationship between the average energy consumption and the number of mobile nodes. It can be seen that SEDA has a lower energy consumption compared to the other two schemes. Specifically, when the number of mobile nodes is 200, the average energy consumption of SEDA is 52.8% and 44.25% less than that of PEC and PHDA, respectively. As mentioned earlier, this is because the energy consumption is also decreasing as the number of forwarding hops decreases.

A comparison between experiment 1 and experiment 2 is shown in Table 3. This table shows how much of the packet loss ratio, delay and energy consumption of SEDA is less than that of PEC and PHDA. When the mobile user's moving speed is 2m/s, experiment 2 has a high packet loss ratio compared to experiment 1. Because as the mobile user's mobile speed becomes faster, the network topology becomes faster, thus affecting the packet delivery ratio. Due to the higher number of packets lost in experiment 2, energy consumption and delay are reduced.

6 Conclusion

In this paper, we propose a secure and efficient data aggregation scheme for cloud-assisted WBAN. The scheme uses privacy homomorphism to encrypt user data so that it does not need to be decrypted when aggregating data, ensuring data confidentiality and resisting compromise attacks. At the same time, the user data is forwarded by using the fixed base station node and the best relay node between the user and the base station, thus improving the transmission efficiency of the user data. The experimental results show that our scheme has lower packet loss ratio, smaller delay and less energy consumption.

Acknowledgments

Above work is supported by National Natural Science Foundation (NSF) of China under grant Nos. 61370007, 61572206, U1405254, Huaqiao University graduate research innovation ability cultivation project of China under grant No. 1511314006, Fujian Provincial Natural Science Foundation of China under grant No. 2013J01241, and Program for New Century Excellent Talents of Fujian Provincial under grant No. 2014FJ-NCET-ZR06.

References

[1] L. Chen, R. X. Lu, Z. F. Cao, K. AlHarbi, and X. D. Lin, "Muda: Multifunctional data aggregation in privacy-preserving smart grid communications," *Peer-to-Peer Networking and Applications*, vol. 8, no. 5, pp. 777–792, 2015.

[2] C. Q. Hu, H. J. Li, Y. Huo, T. Xiang, and X. F. Liao, "Secure and efficient data communication protocol for wireless body area networks," *IEEE Transactions on Multi-Scale Computing Systems*, vol. 2, no. 2, pp. 94–107, 2016.

[3] N. Jammali and L. C. Fourati, "Pfka: A physiological feature based key agreement for wireless body area network," in *International Conference on Wireless Networks and Mobile Communications (WINCOM'15*, pp. 1–8, Oct. 2015.

[4] F. A. Khan, A. Ali, H. Abbas, and N. A. H. Haldar, "A cloud-based healthcare framework for security and patients' data privacy using wireless body area networks," *Procedia Computer Science*, vol. 34, pp. 511–517, 2014.

[5] F. A. Khan, A. Ali, H. Abbas, and N. A. H. Haldar, "A secure and efficient one-time password authentication scheme for WSN," *International Journal Network Security*, vol. 19, no. 2, pp. 177–181, 2017.

[6] C. T. Li, C. C. Lee, and C. Y. Weng, "A secure cloud-assisted wireless body area network in mobile emergency medical care system," *Journal of Medical Systems*, vol. 40, no. 5, p. 117, 2016.

[7] C. T. Li, C. W Lee, and J. J. Shen, "An extended chaotic maps-based keyword search scheme over encrypted data resist outside and inside keyword guessing attacks in cloud storage services," *Nonlinear Dynamics*, vol. 80, no. 3, pp. 1601–1611, 2015.

[8] X. H Liang, R. X. Lu, L. Chen, X. D. Lin, and X. M. Shen, "Pec: A privacy-preserving emergency call scheme for mobile healthcare social networks," *Journal of Communications and Networks*, vol. 13, no. 2, pp. 102–112, 2011.

[9] X. D. Lin, R. X. Lu, X. M. Shen, Y. Nemoto, and N Kato, "Sage: A strong privacy-preserving scheme against global eavesdropping for ehealth systems," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 4, pp. 365–378, 2009.

[10] S. Ozdemir, M. Peng, and Y. Xiao, "Prda: polynomial regression-based privacy-preserving data aggregation for wireless sensor networks," *Wireless Communications and Mobile Computing*, vol. 15, no. 4, pp. 615–628, 2015.

[11] R. F. S. Pearlin and G. Rekha, "Performance comparison of AODV, DSDV and DSR protocols in mobile networks using NS-2," *Indian Journal of Science and Technology*, vol. 9, no. 8, pp. 130–141, 2016.

[12] S. N. Ramli, R. Ahmad, M. F. Abdollah, and E. Dutkiewicz, "A biometric-based security for data authentication in wireless body area network (wban)," in *15th International Conference on Advanced Communication Technology (ICACT'13)*, pp. 998–1001, Jan. 2013.

[13] Q. Wang, C. Wang, K. Ren, W. J. 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.

[14] K. Zhang, X. H. Liang, M. Baura, R. X. Lu, and X. M. Shen, "Phda: A priority based health data aggregation with privacy preservation for cloud assisted wbans," *Information Sciences*, vol. 284, pp. 130–141, 2014.

[15] J. Zhou, Z. F. Cao, X. L. Dong, and X. D. Lin, "Security and privacy in cloud-assisted wireless wearable communications: Challenges, solutions, and future directions," *IEEE Wireless Communications*, vol. 22, no. 2, pp. 136–144, 2015.

[16] J. Zhou, Z. F. Cao, X. L. Dong, N. X. Xiong, and A. V. Vasilakos, "4s: A secure and privacy-preserving key management scheme for cloud-assisted wireless body area network in m-healthcare social networks," *Information Sciences*, vol. 314, pp. 255–276, 2015.

Biography

Huaijin Liu received the B.S. degree from Huaqiao University, China, in 2015, where he is currently pursuing the master's degree. His current research interest includes wireless sensor network security, wireless body area network security and privacy protection, wireless vehicle network security.

Yonghong Chen received the B.S. degrees from Hubei National University, and M.Eng. and Ph.D. degree degrees from Chognqing University, Chongqing, China, in 2000 and 2005 respectively. He is currently the professor of of College of Computer Science and Technology, Huaqiao University, Xiamen, China. His research interests include network security, watermarking and nonlinear processing.

Hui Tian received his BSc and MSc degrees in Wuhan Institute of Technology,Wuhan, China in 2004 and 2007, respectively. He received his PhD degree in Huazhong University of Science and Technology, Wuhan, China. He is now an associate professor in the National Huaqiao University of China. His research interests include network and multimedia information security, digital forensics and information hiding.

Tian Wang received his BSc and MSc degrees in Computer Science from the Central South University in 2004 and 2007, respectively. He received his PhD degree in City University of Hong Kong in 2011. Currently, he is a professor in the Huaqiao University of China. His research interests include wireless sensor networks, fog computing and mobile computing.

Yiqiao Cai received the B.S. degree from Hunan University, Changsha, China, in 2007, and the Ph.D. degree from Sun Yat-sen University, Guangzhou, China, in 2012. In 2012, he joined Huaqiao University, Xiamen, China, where he is currently a lecturer with the College of Computer Science and Technology. He is interested in differential evolution, multiobjective optimization, and other evolutionary computation techniques.

Construction and Analysis of Key Generation Algorithms Based on Modified Fibonacci and Scrambling Factors for Privacy Preservation

Amiruddin Amiruddin[1,2], Anak Agung Putri Ratna[1], and Riri Fitri Sari[1]
(Corresponding author: Amiruddin Amiruddin)

Department of Electrical Engineering, Universitas Indonesia[1]
Jl. Margonda Raya, Pondok Cina, Beji, Kota Depok, Jawa Barat 16424, Indonesia
Sekolah Tinggi Sandi Negara, Bogor, Jawa Barat, Indonesia[2]
(Email: amir@stsn-nci.ac.id)

(Received Sept. 24, 2017; revised and accepted Apr. 12, 2018)

Abstract

Cryptographic key is the most important factor for supporting encryption of confidential data before it is transmitted in a communication network. A good cryptographic key has properties of random sequence and long period. For these purposes, a randomness capable and lightweight computing algorithm is required. The randomness capability and computation time of such an algorithm can be measured by using randomness test and algorithmic complexity analysis, respectively. In this paper, two models of key generation algorithm using the modified Fibonacci and scrambling factor were constructed. Such modification and scrambling factor are intended to support the randomness capability and low algorithmic complexity. The proposed key generation algorithms have been simulated and analyzed. The key generation algorithm Model 2 (called hereinafter "Scrambled Fibonacci-based") is better than Model 1 in term of randomness, despite both having similar linear algorithmic complexity, denoted by $\mathcal{O}(n)$.

Keywords: Cryptography; Key Generation; Randomness; Scrambled Fibonacci; Scrambling Factor

1 Introduction

Wireless communication system and its services have become an important component of modern life and society. An example of such a wireless network is the Internet of Things (IoT) that grows rapidly, nowadays, to support human beings need on information. However, due to the nature of the Radio Frequency (RF) spectrum used as shared transmission medium, wireless communications are essentially vulnerable and prone to interception [5]. The next generation of wireless communication systems should support applications with very low communication latency, availability, high reliability and security [27]. To protect from interception and to ensure the data confidentiality, many wireless systems use cryptographic systems with secret keys that are only available to the legitimate senders and recipients.

Various methods or approaches have been proposed to generate long and random encryption keys [36]. Each method or approach has advantages and disadvantages and cannot be applied to all different kinds of applications. Therefore, a key generation function should be tailored and adjusted to the characteristics of the applications that will use it. One important consideration in designing a key generation algorithm is its algorithmic complexity [24]. For applications in low-capacity devices for IoT, low complexity algorithms are required. Unfortunately, the existing key generation algorithms lack the measurement of their complexity.

Fibonacci sequence [10] which is a very famous series function in the field of mathematics can be used to generate encryption keys. It is a sequence of numbers where a number is found by adding up the two preceding numbers. Beginning with 0 and 1, the sequence goes as 0, 1, 1, 2, 3, 5, 8, 13, 21, 34, and so forth. Written as a rule, the expression is $x_i = x_{i-1} + x_{i-2}$. Its lightweight operation and ability to save computing time are of the reasons for using it in the key generation function. Several cryptographic methods used Fibonacci sequence or its behavior for encryption application [9, 12]. Applying Fibonacci is suitable for common areas that do not involve data privacy. However, for confidential data-based applications, it is necessary to make improvement on the Fibonacci function. Moreover, the operation used in Fibonacci produces a regular pattern (ascending or descending) that can be used as an entrance point to analyze the resulted key sequence. Therefore, it is necessary to modify the Fibonacci operation so that the resulted pattern becomes random and hardens the efforts to analyze it.

In this paper, a key generation algorithm based on the modified Fibonacci and scrambling factor is constructed and analysed. The key generation function will be applied to the Internet of Things network with constrained devices that have limited storage, low computing power and energy. The contribution of this work is a construction and analysis of key generation algorithm based on Fibonacci and scrambling factor which satisfies the requirement for random and long period of key sequences.

The remaining of the paper is organized as follows. Section 2 gives a brief overview of the previous related works regarding cryptographic key generation methods. Section 3 describes the key generation definition and its performance measurement. The proposed method is described in Section 4. Section 5 discusses the simulation and the result and Section 6 closes the paper with the conclusion.

2 Related Works

As technology grows, research on cryptographic key management [17,20,22,25,32] continues to be done in various aspects including key generation [14,18,28,29,34,35,37], agreement [3, 6, 7, 13, 15, 33] or exchange [4], distribution [8], assignment [19], authentication [16], and update. However, we focus on efforts for key generation to be used on symmetric cryptosystems. Verma *et al.* [31] proposed a method for generating cryptographic key using biometrics with the help of fingerprint pattern. The algorithm generates key by extracting minutiae points and core point and the final key is obtained from the fingerprint image. Turakulovich *et al.* Turakulovich *et al.* [30] discussed comparative factors of the key generation techniques include randomness, key space, key space of biometric, entropy, measured entropy of biometric, convenience and cost, secure saving, update. However, they mostly discussed the factor for biometric-based key generation technique which is different from our proposed method.

Hossain *et al.* [11] proposed a One Time Key (OTK) generation technique based on User ID (UID) and password. The key generation method involves a server-side process to check for possible collisions between a new UID and a given UID to the previous client. This process takes a long time so it is not applicable for devices with limited storage and energy. Torre *et al.* [29] studied the practical performance of an enhanced channel-based key generation system with a very short roundtrip delay, allowing reciprocal channel assessment with increased accuracy. The reciprocal channel measurements performed by body-worn sensor nodes is used to extract encryption keys. Although the performance is slightly increased due to the shorter round-trip delay, further apparent non-reciprocity in the channel measurements can probably be attributed to inaccuracy of the received signal strength indication in the transceiver chip.

Tavangaran *et al.* [27] studied the secret key generation protocol for a compound Discrete Memoryless Multiple Sources (DMMS) with one-way communication in presence of an eavesdropper. The key generation protocol uses a two phase approach to achieve secret key. In the first step, the sender estimates his state and sends this along with other information which is obtained from his observation to the recipient. In the second step, the recipient uses this information including the estimated state of the sender to generate the secret key. However, the protocol has not been reported whether it has been implemented or not.

Al-Moliki *et al.* [2] enhanced the confidentiality of Visible Light Communication (VLC) networks by suggesting a new key generation protocol for optical Orthogonal Frequency Division Multiplexing (OFDM) schemes in an indoor environment. The keys are extracted from the bipolar OFDM samples produced from optical OFDM schemes. This approach which emphasizes the source of key generation differs from our proposed approach which emphasizes the process of the key generation.

Karimian *et al.* [14] proposed a novel approach of key generation that extracts keys from real-valued ECG features. However, this approach is only suitable for ECG-based applications although it can also be modified for other field applications.

In this research, we proposed new key generation algorithm based on modified Fibonacci and scrambling factor to support long periodicity and randomness of the generated key sequence. The research position of our proposed key generation algorithm among other algorithms is summarized in Table 1.

3 Key Generation and Performance Measurement

3.1 Key Generation

In the field of cryptography, key is the most urgent parameter for data encryption or decryption. By definition, a key is a sequence of a random string of bits created explicitly for scrambling and unscrambling data. Instances of cryptographic processes demanding the usage of keys include, inter alia, the transformation of plain text data into cipher text data (encryption) and vice versa (decryption), the computation and verification of a digital signature, the computation and verification of an authentication code from data, the computation of a shared secret that is used to obtain keying material, and the derivation of additional keying material from a key-derivation key.

There are two types of key, *i.e.* symmetric and asymmetric key. Symmetric key is a key used in a symmetric-key cryptographic algorithm which requires that the key must be kept secret. Asymmetric key is a key used with a public-key algorithm. In asymmetric key cryptography, there are two corresponding keys, *i.e.* private and public keys. A private key is a cryptographic key used with a public-key algorithm that must be kept secret and is uniquely associated with an entity that is authorized to use it. Public key is a key used with a public-key algorithm that may be made public and is associated with a

Table 1: Research position of key generation

Author	**Method**	**Source**	**Implementation**
Verma *et al.*, 2016	extraction	Biometric with finger-print pattern	Simulated on Matlab
Hossain *et al.*, 2016	generation	UID and password	Simulated on mobile phone
Torre *et al.*, 2017	extraction	Reciprocal channel measurements	Not reported
Tanvangaran *et al.*, 2017	generation	Compound: information, estimated state	Not yet
Al-Moliki *et al.*, 2017	extraction	Bipolar OFDM samples	Simulated on Monte Carlo
Karimian *et al.*, 2017	extraction	ECG value	-
Amiruddin *et al.*, 2017	generation	User input	Simulated on Matlab

private key and an entity that is authorized to use that private key.

Key generation is the process of generating keys for cryptographic purpose such as encryption [1,21]. A cryptographic key can be generated through a function in software or hardware with input parameters that can be obtained from various sources, *e.g.* channels used by each pair of users [28], phase fluctuations in fiber links, and values entered by the user.

3.2 Measurement

To measure the performance of the proposed key generation algorithm, several tests were used, *i.e.* key generation speed, key randomness, key periodicity, and algorithmic complexity analysis. The key generation speed was measured by recording the start time and finish time of the key generation process and then subtracting the start time from the finish time. The key randomness was measured by using autocorrelation function, while the key periodicity was manually analyzed. All of these kinds of key performance measurement were also used in [26]. The analysis of algorithmic complexity was measured by following the description and calculation example presented in [24].

4 Proposed Methods

We have constructed two models of new key generation algorithm based on the modified Fibonacci, described in detail as follows.

4.1 Model 1

In Model 1, modification made on Fibonacci series is the application of modulo number (annotated with c). In this model, the first key, K_1, is obtained by the result of $a\%c$. K_2 is obtained by the result of $b\%c$. K_i is obtained by the result of the addition of $(K_{i-1} + K_{i-2})\%c$. By applying a modulus number in this model, the ascending or descending pattern of the generated key sequences is reduced in periodicity. The key generation function of Model 1 is given in pseudocode form in Algorithm 1.

Algorithm 1 Key Generation (Model 1)

```
Begin
Initialize the parameters: $a$, $b$, $n$, $c$
Derive the 1st element of the key
$K(1) \leftarrow mod(a, c)$
Derive the 2nd element of the key
$K(2) \leftarrow mod(b, c)$
Derive the 3rd to n-th element of the key
for $i = 3$ to $n$ do
    $K(i) \leftarrow mod(K(i-1) + K(i-2), c)$
end for
Output the key sequence, K
End
```

The simulation result showed that this model produces key sequences that have a better randomness level than those produced by the original Fibonacci. However, the randomness of the key sequences is not yet satisfied the randomness test. Therefore, we constructed another model of key generation algorithm, Model 2, by adding a scramble factor to the previous model, Model 1. This addition of scrambling factor was expected to make the generated key sequences more random to satisfy the randomness test.

4.2 Model 2

In Model 2, Fibonacci is modified to generate random and long period key sequences. Input parameters for the function are a, b, n and d as the first number, second number, key length, and modulus number, respectively. Through this model, the first key,K_1, is generated from the formula $(a*b-a)\%d$. K_2 is generated similarly from $(a*b-b)\%d$. For $i = 3 : n$, K_i is obtained from the addition of the two previous key, $K_{i-1} + K_{i-2}$, and a scrambling factor, $3*i$ in formula of $K_{i-1} + K_{i-2} + 3*i\%d$. Actually, the original Fibonacci is presented in the addition marked with a + symbol. However, such an addition can cause the resulted key sequences to have a low randomness and easy to be guessed (as described in Model 1). Therefore, in Model 2, a scrambling factor using a multiplication, $3*i$, is added to improve the modified Fibonacci in generating random key sequences. The scrambling factor with the use of multiplication (*) involving an ever-changing

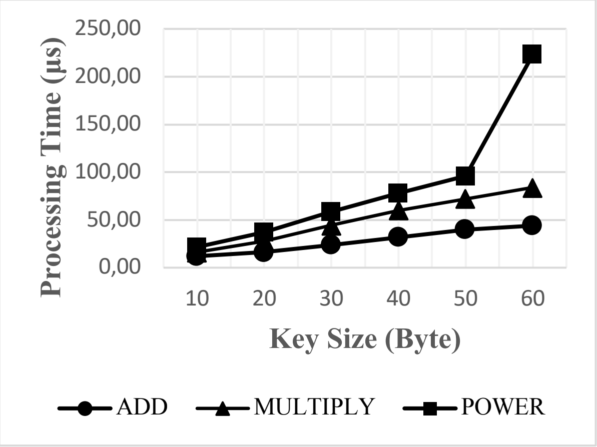


Figure 1: Processing time comparison of several operations

number, i, produces a non-patterned or random number.

The scrambling factor with the use of power (^ operation can also generate random numbers, but has a high computational overhead. The scrambling factor of the addition (+) and subtraction (−) produces a number that is ascending or descending pattern, while the division (:) can cause infinite numbers if the divisor is a zero. In Figure 1 we show the experiment result of processing time comparison among the addition, multiplication, and power operations. Based on this result and the previous explanation, we use the multiplication in our scrambling factor in modifying the Fibonacci sequence.

The proposed generation algorithm of Model 2 is given in pseudocode form in Algorithm 2.

Algorithm 2 Key Generation (Model 2)

```
Begin
Initialize the parameters: $a, b, n, c$
Derive the 1st element of the key
$K(1) \leftarrow mod(a * b - a, c)$
Derive the 2nd element of the key
$K(2) \leftarrow mod(a * b - b, c)$
Derive the 3rd to n-th element of the key
for $i = 3$ to $n$ do
    $K(i) \leftarrow mod(K(i-1) + K(i-2) + 3 * i, c)$
end for
Output the key sequence, K
End
```

The use of $*b - a$, $*b - b$, and $3 * i$ in Algorithm 2 is intended to satisfy the randomness test and long periodicity of the generated key sequences as the requirements of good key as stated by Shannon [34].

The difference of operations among the original Fibonacci, modified Fibonacci Model 1, and Model 2 is presented in Table 2.

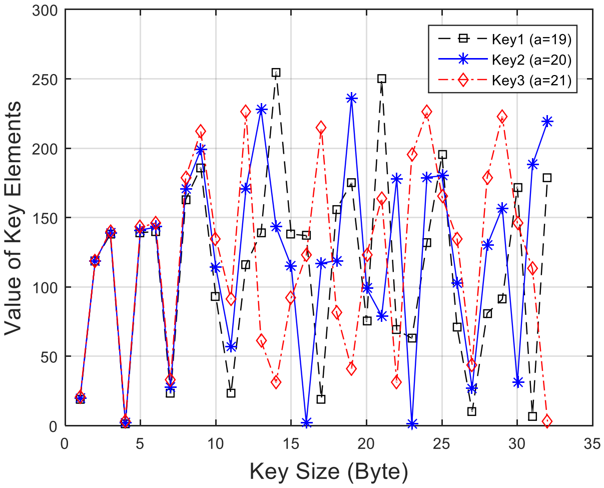


Figure 2: Key element plot of three key sequences with different initial value of parameter a of proposed algorithm Model 1

5 Results and Discussion

5.1 Model 1

The performance of the proposed key generation algorithms Model 1 and Model 2 were evaluated by simulation using Matlab software. The key generation algorithm of Model 1 does not meet the randomness test as in Figure 1, indicated by similar pattern of the three generated key sequences. By varying the value of parameter a with 19 (Key1), 20 (Key2), and 21(Key3), the three key sequences for up to 10 Bytes length have similar plot of key elements and this means that the key sequences are not random. Due to the unsatisfactory result of randomness of proposed Model 1 algorithm, we then proposed Model 2 algorithm.

5.2 Model 2

The key generation simulation used the Algorithm 2 by varying the parameters *i.e.* the first number (a), second number (b), and the key length (key size) (n) and let the modulus number (c) be remains in 256, as the maximum value of alphabet characters. In this simulation, we measured the key generation speed and randomness level, and compared with other algorithm.

5.3 Key Generation Speed

For measuring the speed of key generation, 1000 (one thousand) key sequences were generated. Each key sequence was then recorded at the start and the end time to get the processing time (finish time - start time), then looked for the average processing time by summing up all processing time and then divided by 1000. We can see on Figure 2 the time it took to generate key sequences of

Table 2: Different operations among fibonacci-based algorithms

Original Fibonacci	**Modified Fibonacci Model 1**	**Modified Fibonacci Model 2**
$U_1 = a$	$U_1 = \mod(a, c)$	$U_1 = \mod(a * b - a, d)$
$U_2 = b$	$U_2 = \mod(b, c)$	$U_2 = \mod(a * b - b, d)$
$U_i = U_{i-1} + U_{i-2}$	$U_i = \mod(U_{i-1} + U_{i-2}, c)$	$U_i = \mod(U_{i-1} + U_{i-2} + 3 * i, d)$

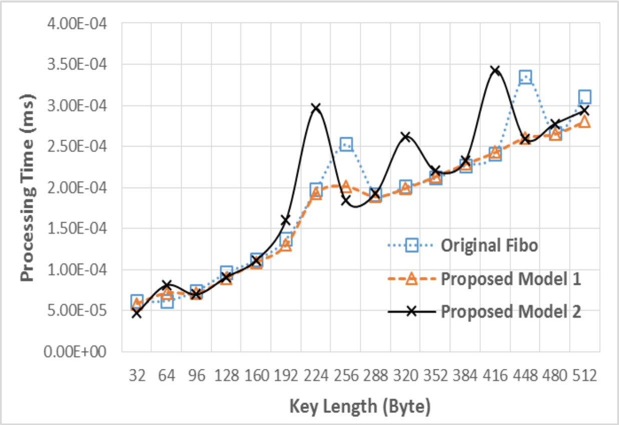


Figure 3: Comparison of key generation processing time (ms) for varying key length (byte)

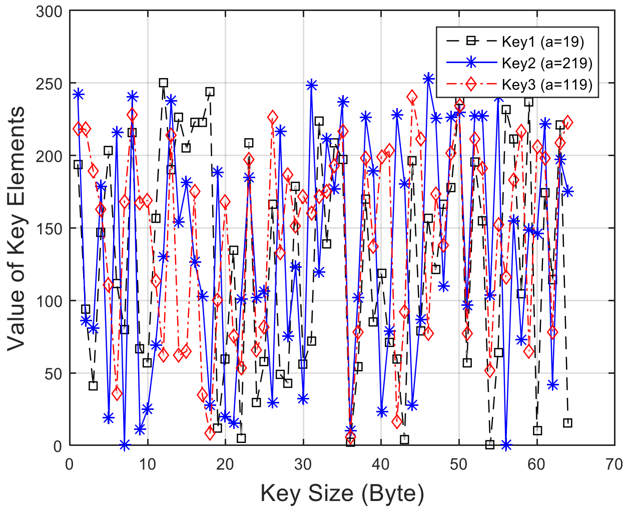


Figure 4: Plot of key elements of three key sequences of 64 Bytes with different value of parameter a: 19 (Key1), 219 (Key2), and 119 (Key3) of the proposed algorithm Model 2

varying key length sizes with the same initial parameter.

It showed that the key generation time increases along with the increase of the key length. However, the proposed Model 1 algorithm has the fastest processing time among all, while the proposed Model 2 has the lowest processing time in certain time but still has a relatively same processing time to the original Fibonacci algorithm. However, this can be explained as a consequence of using more functions in the proposed algorithm Model 2 than the number of functions in the Model 1 and the original Fibonacci. The use of more functions is intended to support the randomness of the generated key sequences. The increase in key length size is not linear with increasing generation time, since the ratio between the two is relatively decreasing as the key length increases.

5.4 Randomness of Key Sequence

Randomness tests are involved in a measuremnet to analyze the distribution of a set of data to see if it is uncorrelated or random. To test the randomness of the generated key sequences, we have simulated key generation by varying the value of variable a, *i.e.* 19, 219, 119 for Key 1, Key 2, and Key 3, as given in Figure 3. It appears that all the key sequences have different and irregular patterns indicating that all of the key sequences are uncorrelated or random.

By varying only the parameter values b from 119 to 124, we generated six key sequences and obtained their plot of key autocorrelation values as shown in Figure 4. The autocorrelation function (ACF), Rk, is calculated using the Equation (1) described in [52], when given a measurement of the variables $Y_1, Y_2, ..., Y_n$ on $X_1, X_2, ..., X_n$, by shifting (lag) of k.

$$R_k = \frac{\sum_{i=1}^{N-k} \left(Y_i - \overline{Y}\right)\left(Y_{i+k} - \overline{Y}\right)}{\sum_{i=1}^{N} \left(Y_i - \overline{Y}\right)^2} \quad (1)$$

As shown in Figure 4, the autocorrelation value of all 64-Byte key sequences with the defined lags of 1 to 20 is in the range between the upper boundary (0.2) and the lower boundary (-0.2) indicating that all of the key sequences are uncorrelated or random. The autocorrelation value for lags of 0 is 1 which means that the two compared key sequences are not random, since there is no key element shift (lag = 0) so there is no difference between the two key sequences. By using Pearson's correlation test, and varying the parameter of b, we have simulated the key generation for 200 key sequences and yielded the correlation coefficient between two key sequences as in Table 3.

The Pearson correlation is calculated using Equation (2),

$$rP = \frac{\sum_{i=1}^{N-k} \left(X_i - \overline{X}\right)\left(Y_{i+k} - \overline{Y}\right)}{\sum_{i=1}^{N} \left(X_i - \overline{X}\right)^2 \left(Y_i - \overline{Y}\right)^2} \quad (2)$$

Table 3: Correlation test (Pearson coefficient and significant test) between two key sequences generated by the proposed Model 2 algorithm

	Correlation between 2 key sequences				
	K_1, K_2	K_2, K_3	K_3, K_4	K_4, K_5	K_5, K_6
Prson Coef	-0.117	-0.085	-0.080	0.078	-0.222
Sig. test	0.535	0.653	0.673	0.681	0.237
Result	Pass	Pass	Pass	Pass	Pass
	Correlation between 2 key sequences				
	K_6, K_7	K_7, K_8	K_8, K_9	K_9, K_{10}	K_{10}, K_1
Prson Coef	-0.062	-0.157	0.056	-0.418	-0.050
Sig. test	0.744	0.406	0.765	0.022	0.790
Result	Pass	Pass	Pass	Pass	Pass

where rP, X, Y, $\overline{(X)}$, $\overline{Y}$, i, n are Pearson correlation, value of variable x, value of variable y, mean value of variable x, mean value of variable y, nth iteration, and number of elements.

A comparison of the randomness of several key sequences between the proposed algorithm and the Raphael algorithm is given in Figure 5, Figure 6, and Table 4. In Figure 5, it can be seen that the plot of key elements between the three key sets generated by the proposed algorithm is more random than that generated by Raphael's algorithm which yields relatively similar plots of key elements.

As shown in Figure 6, by varying only the value of parameter b, the correlation coefficient plot and the P-value of the key sequence generated by the proposed algorithm indicate a random sequence of keys because the value of P-value in average is greater than the correlation coefficient. While in contrast, Raphael's algorithm produces a key sequence that has P-value and a competing correlation coefficient, indicating that the key sequence has a correlation or not random. Further, as shown in Table 4, the comparison of two adjacent key sequences using the Spearman correlation test, indicating that the proposed algorithm passed all randomness tests, while the Raphael algorithm failed on all random tests. All of the above comparisons show the advantages of proposed Model 2 key generation algorithm over Raphael's algorithm.

Table 4: Correlation comparison of two adjacent key sequences of the proposed algorithm

Adj.Key	**Proposed algorithm**			**Raphael's algorithm**		
	Rho	**Pval**	**Test**	**Rho**	**Pval**	**Test**
K_1, K_2	-0.118	0.5356	**Pass**	0.253	0.178	Fail
K_2, K_3	-0.085	0.6533	**Pass**	0.453	0.012	Fail
K_3, K_4	-0.080	0.6739	**Pass**	0.375	0.041	Fail
K_4, K_5	0.078	0.6816	**Pass**	0.005	0.980	**Pass**
K_5, K_6	-0.222	0.2375	**Pass**	0.484	0.007	Fail
K_6, K_7	-0.062	0.7445	**Pass**	0.252	0.180	Fail
K_7, K_8	-0.157	0.4064	**Pass**	0.234	0.214	Fail
K_8, K_9	0.057	0.7659	**Pass**	0.390	0.033	Fail
K_9, K_{10}	-0.419	0.0222	**Pass**	0.103	0.588	**Pass**
K_{10}, K_1	0.051	0.7901	**Pass**	0.002	0.002	Fail

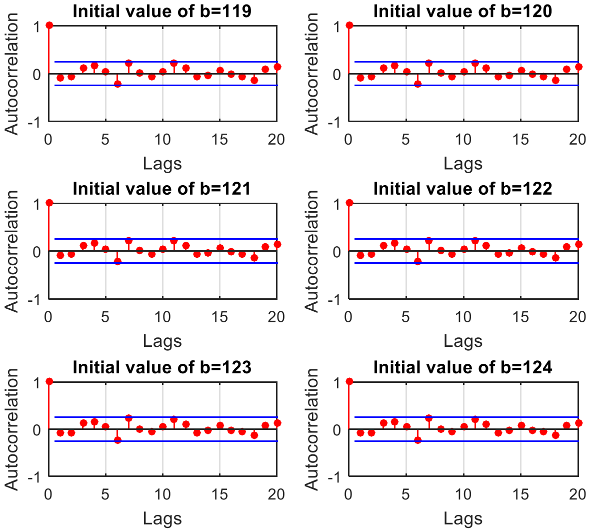


Figure 5: Plot of autocorrelation value of 64 Byte key sequences with different value of parameter b generated by the proposed Model 2 algorithm

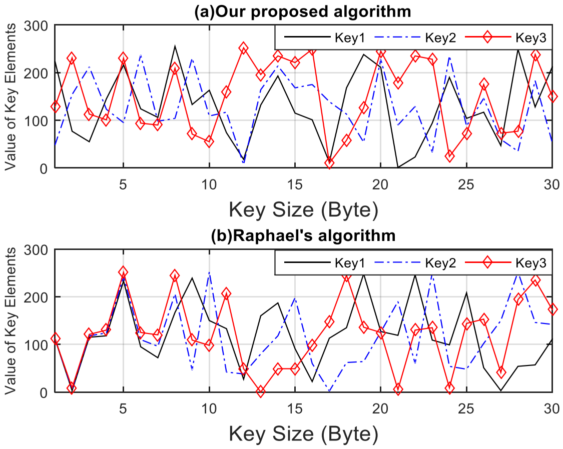


Figure 6: Comparison of key element plot among three key sequences generated by (a) proposed algorithm (b) Raphael's algorithm

5.5 The Periodicity of Key Sequences

Suppose a given key sequence A is 1 3 4 6 5 8 1 3 4 6 5 8 1 3 4 6 5 8 . . . Then the period of key sequence A is 6 (the number of elements from 1 3 4 6 5 8 before experiencing the exact same loop). The longer the period, the better a key sequence. Since the proposed algorithm uses four parameters a, b, d, and n, and there is a scrambling factor that depends on the process iteration, the key sequences generated by the proposed algorithm has a long period equal to n (the key length to be generated, determined by the user). This character is called One Time Key or One Time Pad. Meanwhile, the algorithm used by Raphael *et al.* [23] also satisfies a long period, but since it does not use the modulus function, the resulting key sequence has a regular pattern of ascending or descending. So it does not meet the randomness test and it requires a larger memory to sum the number that will enlarge as the key length increases.

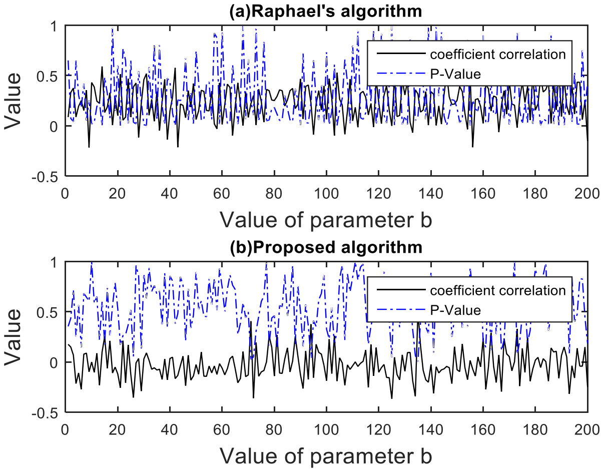


Figure 7: Comparison of correlation coefficient with different value of parameter b

5.6 Complexity of Algorithm

In the Key Generation algorithm Model 1 (see Algorithm 1), there are 10 Line of Codes (LOCs). However, LOCs beginning with a double forward slash (//) are only descriptions and not executed, and hence these parts are ignored and not counted in the analysis of the algorithmic complexity. LOC 3, 5, 7-9 are the processing part of the algorithm that will be calculated on its complexity. In LOC 3, there are 2 instructions *i.e.* mod (a, d) and assignment (=) of the variable K(1). In LOC 5, similar to LOC 3, there are 2 instructions, *i.e.* mod (b, c) and assignment (=) of the variable $K(2)$. LOC 7-9 is an incremental loop as many as $(n + 1 - 3)$ or $(n - 2)$ times with variable i as the counter. The assignment instruction $(i = 3)$ is executed before the loop.

Inside the loop, there are 1 comparison instruction $(i < n+1)$ and 1 incremental counter $(i{+}{+})$. In addition, there are also an assignment operation $(K(i) = \mod (K(i-1)+K(i-2), c))$ involving 1 sum operation (+), 1 modulus operation (mod) and 1 assignment (=) of the variable $K(i)$. Thus, in LOC 7-9 there are 5 instructions repeated $(n - 2)$ times and 1 non-repeated instruction. Thus, in detail, the number of instructions in LOC 3 = 2, LOC 5 = 2, and LOC 7-9 = $5(n - 2) + 1$. Thus the complexity of the Algorithm 1 is $5(n - 2) + 1 + 2 + 2$ or $5(n - 2) + 5$ or $5n - 5$ or $\mathcal{O}(n)$.

In the Key Generation algorithm Model 2 (see Algorithm 2), there are 10 LOCs. LOCs beginning with double forward slash are ignored and not counted in the analysis of the algorithmic complexity. LOC 3, 5, 7-9 is the processing part of the algorithm that will be calculated on its complexity. In LOC 3, there are 4 instructions *i.e.* multiplication (*), subtraction (-), modulus (mod) and assignment (=) of the variable K(1). In LOC 5, similar to LOC 3, there are 4 instructions, *i.e.* multiplication (*), subtraction (-), modulus (mod) and assignment (=) of the variable K(2). LOC 7-9 is an incremental loop as many as $(n + 1 - 3)$ or $(n - 2)$ times with the variable i as the counter. The assignment instruction $(i = 3)$ is executed before the loop.

Inside the loop, there are 1 comparison instruction $(i < n + 1)$ and 1 incremental counter $(i + +)$. In addition, there are also an assignment operation of $(K(i) = \mod (K(i-1) + K(i-2) + 3 * i, d))$ involving 2 addition operations (+), 1 multiplication operation (*), 1 modulus operation (mod) and 1 assignment operation (=) of the variable $K(i)$. Thus, in LOC 5-7, there are 7 instructions repeated (n-2) times and 1 non-repeated instruction. Thus, in detail, the number of instructions in $LOC3 = 4$, $LOC5 = 4$, and LOC 7-9 = $7(n - 2) + 1$. Thus the complexity of the Model 2 algorithm is $7(n - 2) + 1 + 4 + 4$ or $7(n - 2) + 9$ or $7n - 5$ or $\mathcal{O}(n)$.

From the above analysis, it appears that both models of the proposed key generation algorithms have the same algorithmic complexity that is linear complexity expressed by $\mathcal{O}(n)$. However, only the Model 2 satisfies the randomness test.

6 Conclusions

We have utilized the Fibonacci sequence in constructing two proposed models of key generation algorithm, Model 1 (without scrambling factor) and Model 2 (with scrambling factor). The modification by adding a scrambling factor is intended to generate key sequences that satisfy randomness tests and long periodicity which have not been used in measuring the existing key algorithms found in recent literature. Simulation results of the two models indicate that the key generation time increases along with the increase of the key length, but the ratio between key generation time and key length relatively decreases. The randomness test results indicate that the key sequences generated by Model 1 do not meet the randomness test despite having relatively fast computation time.

Model 2 produces key sequences with accepted autocorrelation value (accepted random value) since it is always in the range between the upper boundary (0.2) and the lower boundary (-0.2). The result of algorithmic complexity analysis showed that both of key generation algorithms have a similar linear algorithmic complexity, expressed by $\mathcal{O}(n)$. However, considering all the performance measurement used in the present work, the proposed key generation algorithm Model 2 (Called hereinafter scrambled fibonacci) is the best compared to the proposed Model 1 and original Fibonacci-based Raphael algorithm. The future work of this research would be to implement the proposed Scrambled Fibonacci-based key algorithm in an encryption/decryption application in constrained devices to support security and privacy preservation in the Internet of Things.

References

[1] D. S. AbdElminaam, “Improving the security of cloud computing by building new hybrid cryptography algorithms,” *International Journal of Electronics and Information Engineering*, vol. 8, no. 1, pp. 40–48, 2018.

[2] Y. M. Al-Moliki, M. T. Alresheedi, and Y. Al-Harthi, “Secret key generation protocol for optical OFDM systems in indoor VLC networks,” *IEEE Photonics Journal*, vol. 9, no. 2, pp. 1–15, 2017.

[3] S. Arasteh, S. F. Aghili, and H. Mala, “A new lightweight authentication and key agreement protocol for internet of things,” in *13th International Iranian Society of Cryptology Conference on Information Security and Cryptology (ISCISC'16)*, pp. 52–59, 2016.

[4] M. Azrour, Y. Farhaoui, and M. Ouanan, “A new secure authentication and key exchange protocol for session initiation protocol using smart card,” *International Journal of Network Security*, vol. 19, no. 6, pp. 870–879, 2017.

[5] S. Baksi, J. Snoap, and D. C. Popescu, “Secret key generation using one-bit quantized channel state information,” in *IEEE Wireless Communications and Networking Conference (WCNC'17)*, pp. 1–6, 2017.

[6] M. Bayat, M. Aref, “An attribute based key agreement protocol resilient to KCI attack," *International Journal of Electronics and Information Engineering*, vol. 2, no. 1, pp. 10–20, 2015.

[7] T. Y. Chang, W. P. Yang, M. S. Hwang, “Simple authenticated key agreement and protected password change protocol", *Computers & Mathematics with Applications*, vol. 49, pp. 703–714, 2005.

[8] S. F. Chiou, M. S. Hwang, and S. K. Chong, “A simple and secure key agreement protocol to integrate a key distribution procedure into the DSS,” *International Journal of Advancements in Computing Technology (IJACT'12)*, vol. 4, no. 19, pp. 529–535, 2012.

[9] K. Eguchi, K. Abe, M. Fujimoto, D. Yan, and I. Oota, “The development of a negative single-input/multi-output driver using a fibonacci-like converter,” in *13th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON'16)*, pp. 1–4, 2016.

[10] S. M. Farooq and S. H. S. Basha, “A study on fibonacci series generation algorithms,” in *3rd International Conference on Advanced Computing and Communication Systems (ICACCS'16)*, pp. 1–5, 2016.

[11] S. Hossain, A. Goh, C. H. Sin, and L. K. Win, “Generation of one-time keys for single line authentication,” in *14th Annual Conference on Privacy, Security and Trust (PST'16)*, pp. 686–689, 2016.

[12] C. H. Hsu, H. S. Dang, and T. A. T. Nguyen, “The application of fibonacci sequence and taguchi method for investigating the design parameters on spiral micro-channel,” in *International Conference on Applied System Innovation (ICASI'16)*, pp. 1–4, 2016.

[13] Q. Jiang, S. Zeadally, J. Ma, and D. He, “Lightweight three-factor authentication and key agreement protocol for internet-integrated wireless sensor networks,” *IEEE Access*, vol. 5, pp. 3376–3392, 2017.

[14] N. Karimian, Z. Guo, M. Tehranipoor, and D. Forte, “Highly reliable key generation from electrocardiogram,” *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 6, pp. 1400–1411, 2017.

[15] M. Lavanya and V. Natarajan, “Lightweight key agreement protocol for iot based on IKEv2,” *Computers & Electrical Engineering*, vol. 64, pp. 580-594, 2017.

[16] C. C. Lee, M. S. Hwang, L. H. Li, “A new key authentication scheme based on discrete logarithms", *Applied Mathematics and Computation*, vol. 139, no. 2, pp. 343-349, July 2003.

[17] A. Lei, H. Cruickshank, Y. Cao, P. Asuquo, C. P. A. Ogah, and Z. Sun, “Blockchain-based dynamic key management for heterogeneous intelligent transportation systems,” *IEEE Internet of Things Journal*, vol. 4, no. 6, pp. 1832–1843, 2017.

[18] C. T. Li, M. S. Hwang and Y. P. 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, Aug. 2009.

[19] I. C. Lin, M. S. Hwang, C. C. Chang, “A new key assignment scheme for enforcing complicated access control policies in hierarchy", *Future Generation Computer Systems*, vol. 19, no. 4, pp. 457–462, May 2003.

[20] I. C. Lin, H. H. Ou, M. S. Hwang, “Efficient access control and key management schemes for mobile agents", *Computer Standards & Interfaces*, vol. 26, no. 5, pp. 423–433, 2004.

[21] L. Liu and Z. Cao, "Analysis of two confidentiality-preserving image search schemes based on additive homomorphic encryption," *International Journal of Electronics and Information Engineering*, vol. 5, no. 1, pp. 1–5, 2016.

[22] Z. Mahmood, H. Ning, and A. Ghafoor, "Lightweight two-level session key management for end user authentication in internet of things," in *IEEE International Conference on Internet of Things and IEEE Green Computing and Communications and IEEE Cyber, Physical and Social Computing and IEEE Smart Data*, pp. 323–327, 2016.

[23] A. J. Raphael and Dr. V. Sundaram, "Secured communication through fibonacci numbers and unicode symbols," *International Journal of Scientific and Engineering Research*, vol. 3 no. 4, no. 4, pp. 1–5, 2012.

[24] L. Rosyidi and R. F. Sari, "A practical approach for complexity analysis of autonomic internet of things protocol algorithm," in *19th International Symposium on Wireless Personal Multimedia Communications (WPMC'16)*, pp. 256–261, 2016.

[25] T. H. Sun and M. S. Hwang, "A hierarchical data access and key management in cloud computing," *ICIC Express Letters*, vol. 6, no. 2, pp. 569–574, 2012.

[26] Y. Suryanto, Suryadi, and K. Ramli, "A secure and robust image encryption based on chaotic permutation multiple circular shrinking and expanding," *Journal of Information Hiding and Multimedia Signal Processing-Ubiquitous International*, vol. 7 no. 4, pp. 697–713, 2016.

[27] N. Tavangaran, H. Boche, and R. F. Schaefer, "Secret-key generation using compound sources and one-way public communication," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 1, pp. 227–241, 2017.

[28] C. D. T. Thai, J. Lee, and T. Q. S. Quek, "Physical-layer secret key generation with colluding untrusted relays," *IEEE Transactions on Wireless Communications*, vol. 15, no. 2, pp. 1517–1530, 2016.

[29] P. Van Torre, Q. Van den Brande, J. Verhaevert, J. Vanfleteren, and H. Rogier, "Key generation based on fast reciprocal channel estimation for body-worn sensor nodes," in *11th European Conference on Antennas and Propagation (EUCAP'17)*, pp. 293–297, 2017.

[30] K. Z. Turakulovich and Y. B. Karamatovich, "Comparative factors of key generation techniques," in *International Conference on Information Science and Communications Technologies (ICISCT'16)*, pp. 1–3, 2016.

[31] I. Verma and S. Jain, "Biometric based key-generation system for multimedia data security," in *3rd International Conference on Computing for Sustainable Global Development (INDIACom'16)*, pp. 864–869, 2016.

[32] P. Vijayakumar, M. Azees, A. Kannan, and L. Jegatha Deborah, "Dual authentication and key management techniques for secure data transmission in vehicular ad hoc networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 4, no. 4, pp. 1015–1028, 2016.

[33] F. Wu, L. Xu, S. Kumari, X. Li, J. Shen, K. K. R. Choo, M. Wazid, and A. K. Das, "An efficient authentication and key agreement scheme for multi-gateway wireless sensor networks in iot deployment," *Journal of Network and Computer Applications*, vol. 89, no. Supplement C, pp. 72–85, 2017.

[34] S. Xiao, Y. Guo, K. Huang, and L. Jin, "High-rate secret key generation aided by multiple relays for internet of things," *Electronics Letters*, vol. 53, no. 17, pp. 1198–1200, 2017.

[35] H. Zhang, Y. Liang, L. Lai, and S. Shamai Shitz, "Multi-key generation over a cellular model with a helper," *IEEE Transactions on Information Theory*, vol. 63, no. 6, pp. 3804–3822, 2017.

[36] J. Zhang, T. Q. Duong, A. Marshall, and R. Woods, "Key generation from wireless channels: A review," *IEEE Access*, vol. 4, pp. 614–626, 2016.

[37] J. Zhang, B. He, T. Q. Duong, and R. Woods, "On the key generation from correlated wireless channels," *IEEE Communications Letters*, vol. 21, no. 4, pp. 961–964, 2017.

Biography

Amiruddin Amiruddin, a lecturer at Sekolah Tinggi Sandi Negara (STSN), Indonesia. He received Bachelor?s degree in Informatics from Universitas Budi Luhur. He received his Master?s in Information Technology from the Faculty of Computer Science, Universitas Indonesia (UI). He just received a Ph.D degree from the Electrical Engineering Department, Faculty of Engineering, Universitas Indonesia (UI).

Anak Agung Putri Ratna, a Senior Lecturer of Computer Engineering at Electrical Engineering Department, Faculty of Engineering, Universitas Indonesia (UI). She graduated with a BSc in Electrical Engineering from UI, a Master in Engineering from the Waseda University Japan, and a Ph.D in Electrical Engineering from UI.

Riri Fitri Sari, a Professor of Computer Engineering at the Electrical Engineering Department, Faculty of Engineering, Universitas Indonesia (UI). She graduated with a BSc in Electrical Engineering from UI, a Master in Human Resources Management from the Atmajaya University Jakarta and an MSc in Software Systems and Parallel Processing from the Department of Computer Science, University of Sheffield, UK, funded by British Council Chevening Award, and a Ph.D in Computer Networks from the School of Computing, University of Leeds, UK. Her current main teaching and research area includes Computer Network, Internet of Things (IoT), Cloud Computing, Vehicle Ad Hoc Networks, and ICT implementation.

Spectrogram-based Efficient Perceptual Hashing Scheme for Speech Identification

Qiu-Yu Zhang, Tao Zhang, Si-Bin Qiao, and Dong-Fang Wu
(*Corresponding author: Qiu-Yu Zhang*)

School of Computer and Communication, Lanzhou University of Technology
No.287, Lan-Gong-Ping Road, Lanzhou 730050, China
(Email: zhangqylz@163.com)

(*Received Sept. 19, 2017; revised and accepted Mar. 27, 2018*)

Abstract

In order to meet the requirements of discrimination, robustness and high-efficiency identification of existing speech identification algorithms in mobile speech communication, an efficient perceptual hashing scheme based on spectrogram for speech identification was proposed in this paper. Firstly, a fraction of spectrogram is cut, which represents low frequency information of input speech signal and is less susceptible to common content-preserving manipulations such as MP3 compression, noise addition and volume adjustment etc. Secondly, the local binary pattern (LBP) algorithm is applied to produce a LBP feature image. Finally, the perceptual hashing sequence is obtained by employing an image perceptual hashing algorithm to the LBP feature image. Experimental results show that the proposed approach has a good discrimination, robustness and identification efficiency. It can satisfy the real-time identifying requirements in mobile speech communication.

Keywords: Entropy Rate; LBP; Perceptual Hashing; Spectrogram; Speech Identification

1 Introduction

With the development of multimedia technology and network communication technology, the transmission and storage of speech information become more and more convenient. However, some speech information contains much private information, such as court testimony and military order. Therefore, validating their authenticity becomes a critical issue to multimedia identification techniques [5].

Traditional cryptography hashing algorithms are very sensitive to the changes of speech content because of introducing some distortions while processing speech signal, such as resample and compression, which have an adverse effect on speech content identification. Speech perceptual hashing identification technologies can protect speech information by verifying its authenticity, which can guarantee speech information services more safe and reliable [2, 4]. So, this technology is recently receiving big attention in the area of research.

A lot of spectrogram-based audio fingerprinting algorithms have been proposed in recent year. Rafii *et al.* [12] proposed an audio fingerprinting system to handle different kinds of live version audio, and the fingerprinting is extracted from a binary image, which is obtained from a log-frequency spectrogram by using an adaptive threshold method. Though the system shows a good identification precision to different genres of live music, its robustness is not illustrated in detail. To satisfy robustness of audio fingerprinting system, Zhang *et al.* [15] proposed a feature extracting method based on spectrogram through utilizing scale invariant feature transform (SIFT) local descriptor and the locality sensitive hashing (LSH). Due to the stability of SIFT, the proposed algorithm achieves a high discrimination and robustness, but its time complexity is high. Being similar to [15], SIFT is employed to extract 128 features of spectrogram in [16]. Experimental results show that the system has a good identification rates when the audio lengths are stretched from 65% to 150%. However, due to use Euclidean distance to match the features of 128-dimension descriptors, it is still time-consuming.

Besides there are a few audio fingerprinting algorithms based on the feature of spectrogram, some audio perceptual hashing algorithms with respect to other features have been proposed. Chen *et al.* [3] introduced an audio perceptual hashing algorithm based on Zernike moment. Experiment results show that the algorithm achieves a good discrimination and perceptual robustness. However, generating hashing process in his paper takes too much time. Huang *et al.* [6] proposed a speech perceptual hashing algorithm based on linear prediction analysis, which has a high running efficiency but not a good robustness. Li *et al.* [9] introduced a hashing generating approach by utilizing the correlation of Mel-frequency cepstrum coefficients (MFCC). Instead of traditional hamming distance, it takes advantage of similarity metric function to implement hashing matching and shows a good robustness to re-sampling and MP3 compression. However, the algorithm is computationally expensive. In [8], the combination of modified discrete coefficients transform (MDCT) and non-negative matrix factorization (NMF) is considered as a hashing yielding scheme. It exhibits a good robustness but a poor discrimination. To develop an efficient speech identification system, Zhang *et al.* [14] proposed a speech perceptual hashing algorithm in terms of discrete wavelet packet decomposition (WPD). Although the system can discriminate different speech files, it lacks robustness in the noisy environment. By exploiting linear prediction coefficients (LPC) of speech signal to obtain local features, Chen *et al.* [1] proposed a robust hash function, which gains a better discrimination but has poor effect to resist some speech content-preserving distortions, such as filtering and noise addition.

Aiming at the problems mentioned above, obtaining a compromise between discrimination and robustness of algorithm, and meeting the requirement to enhance identification efficiency, we proposed an efficient speech perceptual hashing identification algorithm based on spectrogram. Firstly, a spectrogram produced by a 4 s original speech is obtained like Figure 1(a). Figure 1(b) is a spectrogram of the original speech signal contaminated by 30 dB white Gaussian noise. Through comparing Figure 1(a) with Figure 1(b), it is obvious that noise has little influence on low frequency portion (namely the bottom half of the spectrogram), then, which is cut to get Figure 1(c). In addition, Figure 1(c) is converted to acquire a feature image (Figure 1(d)) by using LBP algorithm. As can be seen from Figure 1(d), most of textural features are extracted. Lastly, an image perceptual hashing algorithm proposed in [7] is applied to get hashing sequences of LBP feature image, and which are matched to finish speech identification. The experimental results demonstrate that the proposed algorithm can satisfy the real-time need of mobile speech communication.

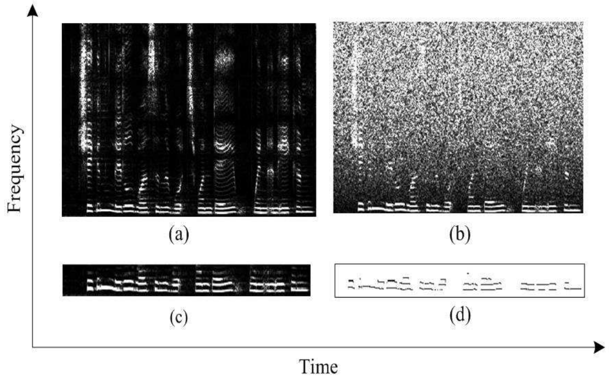


Figure 1: Spectrogram analysis: (a) Spectrogram of an original 4 s speech clip; (b) Spectrogram of adding 30 dB noise to original speech; (c) An image of being cut out from (a); (d) LBP feature image of (c)

The remaining part of this paper is organized as follows. Section 2 does several preliminaries, which are mainly introducing three theories, including LBP descriptor, 2D-DCT and SVD, which will be exploited in this paper. The detailed proposed algorithm is described in Section 3. Subsequently, Section 4 gives the experimental results and performance analysis as compared with other related methods. Finally, we conclude our paper in Section 5.

2 Problem Statement and Preliminaries

2.1 LBP Descriptor

LBP [13] is an effective image texture description method. Owing to having some characteristics, such as simple calculation, rotation and gray invariance, LBP is applicable to real-time system. The method is described briefly as follows: a central pixel point i_c is defined in a local 3×3 neighborhood of a monochrome texture image. Next, i_c is in comparison with one of the joint 8 pixel values i_n (i=1, 2, ..., 8), if i_c is less than in, b_n is set to 1, otherwise to 0 (see Equation(1)). Equation (2) is utilized to get final LBP code values.

$$b_n = \begin{cases} 1 & i_n - i_c > 0 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

$$LBP(x_c, y_c) = \sum_{n=0}^{7} 2^n b_n. \quad (2)$$

In this paper, LBP descriptor is used to extract texture information of spectrogram.

2.2 Two-Dimensional Discrete Cosine Transform (2D-DCT)

Discrete Cosine Transform (DCT) [11] is used to approximate to an image via different amplitude and frequency. Two-dimensional discrete cosine transform (2D-DCT) can be obtained by computing twice one-dimensional DCT in the two directions of row and column. 2D-DCT is frequently applied in image processing since it is characterized by lossless compression and energy concentration. Generally, after 2D-DCT, the main energy of an image is concentrated in the part of low frequency, which is located in top left corner of a 2D-DCT coefficient matrix and representing the stable features of an image. The definition of 2D-DCT is as follows:

$$D(u,v) = \frac{2}{\sqrt{MN}} c(u)c(v) \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y) \quad (3)$$
$$\cos\frac{(2x+1)u\pi}{2M} \times \cos\frac{(2y+1)v\pi}{2N}$$

The 2D-DCT inverse transform is given by:

$$f(x,y) = \frac{2}{\sqrt{MN}}\sum_{u=0}^{M-1}\sum_{v=0}^{N-1} c(u)c(v)D(u,v) \cos\frac{(2x+1)u\pi}{2M} \times \cos\frac{(2y+1)v\pi}{2N} \quad (4)$$

where $D(u,v)$ is a 2D-DCT transform coefficient matrix, the size of input digital image $f(x,y)$ is $M \times N$, $0 \leq x \leq M-1$, $0 \leq y \leq N-1$, $0 \leq u \leq M-1$, $0 \leq v \leq N-1$, $c(u)$ and $c(v)$ are transform parameters and their values are defined as follows:

$$c(u), c(v) = \begin{cases} 1/\sqrt{2} & u,v=0 \\ 1 & \text{otherwise} \end{cases} \quad (5)$$

In this paper, 2D-DCT is utilized to extract the low frequency energy of LBP feature image.

2.3 Singular Value Decomposition

Singular value decomposition (SVD) [10] is a method of algebraic feature extraction and numerical analysis. Thanks to the stability of singular value, it is widely applied in the field of image compression and digital watermark. Supposing A is a $M \times N$ gray image. The SVD transform of A is given by:

$$A = USV^T = \sum_{i=1}^{r} \lambda_i u_i v_i^T \quad (6)$$

where U and V are $M \times M$ and $N \times N$ orthogonal matrices respectively, S is a $M \times N$ matrix. λ is singular value of A and satisfies Equation (7). r is the number of non-zero singular value. $u_i$and v_i are left singular vector and right singular vector corresponded by λ_i.

$$\lambda_1 \geq \lambda_2 \geq \cdots \lambda_r = \cdots = \lambda_M \quad (7)$$

In this paper, SVD is performed to the 2D-DCT low frequency coefficient matrix for obtaining the left singular vector and right singular vector corresponded by the biggest singular value.

3 The Proposed Scheme

The two principal components of speech identification system are hashing generation and hashing matching, and the procedures of proposed perceptual hashing algorithm are shown in Figure 2. The whole steps of the algorithm are depicted as follows: an input speech signal yields a hashing sequence, which is matched with other hashing sequences that are stored in a reference hashing database. And matching results are analyzed to identify input speech content. The specific hashing generation process can be seen from Figure 3. Firstly, one-dimensional speech signal is converted to a two-dimensional spectrogram, due to the high frequency part of speech signal is vulnerable to some distortions, so an image block representing low frequency portion of input speech signal is cut out. Next, by using LBP method, the image block is extracted texture features to gain a LBP feature image. Subsequently DCT coefficient matrix is obtained by utilizing 2D-DCT to the LBP feature image. It is divided into many smaller matrices with the same size, and some of them representing low frequency part of LBP feature image are recombined into a new matrix. Finally, after doing SVD to it, a hash sequence is derived.

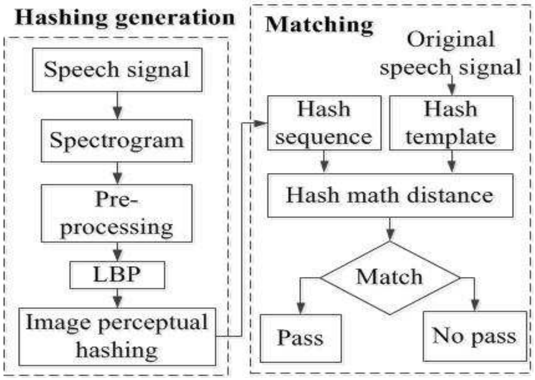


Figure 2: Block diagram of speech perceptual hashing identification algorithm

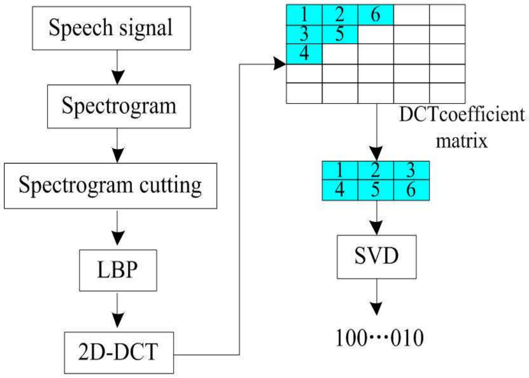


Figure 3: Block diagram of hashing generation of input speech signal

3.1 The Process of Hashing Generation

Assuming the original speech signal is s, the steps of hashing generation are depicted as follows:

Step 1: Short Time Fourier Transform (STFT) is used to obtained the spectrogram of s, which in matrix form is $S_i = \{S_i(k) \mid i = 1,2,\ldots,M, k = 1,2,\ldots,L\}$, where M and L represent the number of rows and columns of the matrix, respectively.

Step 2: Cutting the spectrogram which represents the low frequency part of s, expressed as $Cs_i = S_i$, where, $i = 1, 2, \ldots, N$, and N is the number of rows after S_i is cut.

Step 3: The LBP is performed on Cs to obtain a $M_1 \times N_1$ LBP feature image $L(m, n)$. where M_1 and N_1 are the number of rows and columns of $L(m, n)$ and $m = 1, 2, \ldots, M_1, n = 1, 2, \ldots, N_1$.

Step 4: The 2D-DCT is performed on $L(m, n)$ to obtain a 2D-DCT coefficient matrix $D(m, n)$. In order to extract its top-left-corner low frequency energy conveniently, it is divided into 25 same-sized $p \times q$ matrix blocks $\Psi_{i,j}(p, q)$, and some of them representing low frequency energy of $L(m, n)$ are recombined to derive a new matrix C, which is shown as follows:

$$C = \begin{bmatrix} \Psi_{1,1} & \Psi_{1,2} & \Psi_{2,1} \\ \Psi_{3,1} & \Psi_{2,2} & \Psi_{1,3} \end{bmatrix} \tag{8}$$

where, i and j are block position indices of 2D-DCT block matrix, and each block has p rows and q columns. In this paper, $i = 1, 2, \ldots, 5, j = 1, 2, \ldots, 5, p = 1, 2, \ldots, 6, q = 1, 2, \ldots, 74$.

Step 5: The SVD is performed on C to get a left singular value vector u_1 and a right singular value vector v_1 corresponded by the biggest singular value. Next they are transposed respectively and combine into a feature vector F, which is shown as follows:

$$F(k) = [u_1^T \ v_1^T] \quad 1 \le k \le l \tag{9}$$

where T and k are referred to matrix transpose and feature index respectively, total number of feature is l.

Step 6: F is quantified to obtain a perceptual hashing sequence by using Equation (10).

$$h(k) = \begin{cases} 1 & F(k) \ge 0 \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

where k is hashing index, l is the length of hashing codes.

3.2 Hashing Match

After yielding hashing sequences, the normalized hamming distance, shown in Equation (11), is utilized to match them. The bit error rate (BER) is the ratio between the length of mismatches and the total length of hashing vector, and it is equal to normalized hamming distance in numerical value.

$$BER(h_1, h_2) = D(h_1, h_2) = \frac{1}{l} \sum_{k=1}^{l} \mid h_1(k) - h_2(k) \mid \tag{11}$$

where h_1 and h_2 are two hashing sequences randomly extracted from two speech clips s_1 and s_2, k is hashing index, l is the length of hashing sequence.

In order to estimate the performance of whole perceptual hashing system, a statistical hypothesis testing method is defined as follows:

Given two randomly selecting speech clips s_1 and s_2

H_0**:** if s_1 and s_2 are same two perceptual contents,

$$BER \le \tau$$

H_1**:** if s_1 and s_2 are two different perceptual contents,

$$BER > \tau$$

where τ is perceptual threshold. By setting a reasonable τ and computing BER of two clips s_1 and s_2, if BER$\le \tau$, the two clips can be treated as same two perceptual contents, identification is passed, otherwise not passed.

4 Experimental Results and Analysis

In this section, the performance of the proposed algorithm will be evaluated. The experimental speech data comes from the Texas Instruments and Massachusetts Institute of Technology (TIMIT) speech database and the Text to Speech (TTS) speech database. There are different 1280 speech clips in experimental database recorded by 640 men and 640 women. The format of each speech clip is wav with the length 4 s, which is of the form of 16 bits PCM, mono and sampled at 16 kHz. Experimental hardware environment is Intel(R) Core(TM) i5-3230, 4-core processor, 8 G and 2.6 GHz, software environment is the MATLAB 2013a under Win7 operating system. Next, the proposed method compares with three algorithms in [8, 14, 1], for convenience, which are abbreviated as MDCT-NMF [8], WPD-QT [14] and LPC-NMF [1] respectively. Some parameters involved in this experiment are shown in Table 1.

Table 1: The parameters used in this experiment

Hashing algorithm	Parameters
Proposed	M=257, L=372, N=32, M_1=255, N_1=30, l=234
MDCT-NMF	M=360, L=177, N=100, r=1
WPD-QT	M=64000, N=256, n=16
LPC-NMF	M=360, N=12, r=1

As shown in Table 1, in MDCT-NMF algorithm, speech signal is divided into M frames with L samples, N is the number of lower MDCT coefficients of each frame, and r is the dimension-reduction number of NMF. In WPD-QT algorithm, the length of speech signal is M, wavelet packet coefficients matrix is split into N identical $n \times n$ square blocks. In LPC-NMF algorithm, framing number is M, and N is the order of LPC, and r is the dimension-reduction number of NMF.

4.1 Discrimination Analysis

In this phase, 1, 280 different speech clips are used to calculate BERs in pairs, therefore a total of 818,560 BERs can be obtained and follow the distribution shown in Figure 4. Supposing the generation of binary sequence is random (independent and identical distributed), consequently, these BERs follow binomial distribution (l, μ), where l represents the length of hashing sequence and μ is the probability of that a 0 or 1 is extracted. According to central limit theorem, if l is large, the BERs obey the normal distribution with a mean μ of 0.5 and the standard deviation $\sigma = \sqrt{\mu(1-\mu)/l} = \sqrt{1/4l}$. In our experiment, l=234. After substituting l into above equation, it is found that theoretical standard deviation σ value is 0.0327 (experimental mean and standard deviation value are 0.4904 and 0.0332, respectively), which shows that experimental results are pretty close to theoretical values.

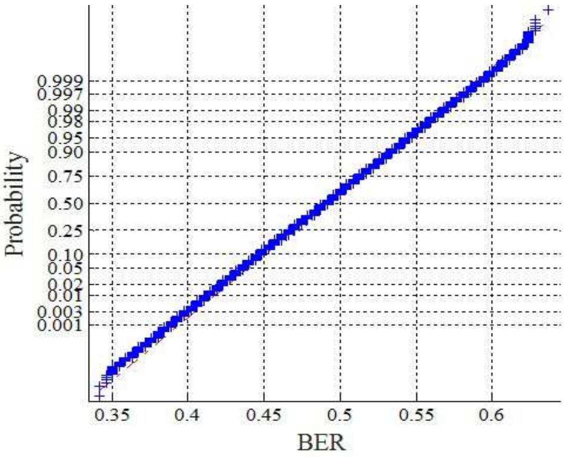


Figure 4: BER normal distribution diagram

The false accept rate (FAR) is commonly used to evaluate discrimination of a speech perceptual hashing system. It refers to the probability that the BER of two different perceptual contents is less than τ, and it is given by Equation (12).

$$FAR(\tau) = \int_{-\infty}^{\tau} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)}{2\sigma^2}} dx \qquad (12)$$

where τ is BER threshold, μ and σ are mean and standard deviation of BERs.

Through comparing proposed algorithm with three other algorithms, the Table 2 shows the FARs under different thresholds, and the conclusion may be drawn: when $\tau < 0.3$, with respect to the discrimination, the proposed method is better than MDCT-NMF and LPC-NMF and close to WPD-QT. This is primarily due to the following reasons, MDCT-NMF method is sensitive to the change of frame size, when sampling rate and frame number are set to 16 kHz and 360 in our experiment respectively, its FARs decreased dramatically; In WPD-QT method, the wavelet packet transform reflects frequency variation of speech signal well, so it has smaller FARs under different thresholds; In LPC-NMF method, there is a certain amount of error when LPC is used to describe vocal tract character. Therefore, the method shows a lower discrimination to different speech signal; for proposed method, different speech signals have obviously distinct spectrograms, so it gains a better discrimination.

Entropy rate (ER) is a comprehensive evaluation criterion on discrimination of perceptual hashing algorithm. It principally overcomes the disadvantages where the discrimination of algorithm is susceptible to hashing size. It ranges from 0 to 1, and the larger its value indicates the higher capacity of discrimination. It can be calculated from following Equation (13) and Equation (14).

$$\begin{aligned} ER &= -[q\log_2 q + (1-q)\log_2(1-q)] && (13) \\ q &= \frac{1}{2}(\sqrt{\frac{\mid \sigma^2 - \sigma_1^2 \mid}{\sigma^2 + \sigma_1^2}} + 1) && (14) \end{aligned}$$

where σ and σ_1 are theoretical and experimental standard deviation of BERs respectively, q is experimental mean value.

As can be observed in Table 3, the entropy rate of proposed method is larger than MDCT-NMF and LPC-NMF except WPD-QT. Thus, by above analyses, proposed algorithm displays a better discrimination.

4.2 Robustness Analysis

Unlike the discrimination analysis, the robustness analysis phase requires to compare BERs yielded from original speech clips with their content-preserving manipulating speech. There are 10 types of content preserving operations shown in Table 4. In order to vividly demonstrate the influence of different content preserving distortions to the 4 s original speech spectrogram (it is shown on Figure 1(a)), six spectrograms are manifested in Figure 5. Compared with Figure 1(a), it can be clearly seen from Figure 5(b), Figure 5(c), Figure 5(d) and Figure 5(e) that the upper part of the spectrogram that reflects high frequency information of the speech signal is subjected to MP3 compression, filtering, noise and echo addition while their lower portions are seldom swayed by these distortions. Moreover, in Figure 5(a), because increasing volume causes the energy of speech fingerprint to rise, some new textures appear, which tends to be useless and interferes with extracting useful textual features. As also can be observed, there is no significant difference between Figure 5(f) and Figure 1(a), this indicates that resampling operation has less impact on Figure 1(a).

The mean and maximum of BERs of the proposed algorithm and three other methods in different content-keeping manipulations are presented in Table 5. In MDCT-NMF approach, when the sampling rate (its value in the original paper is 44.1 kHz while it is 16 kHz in our experiment) is reduced, the length of frame is decreasing on the condition of having same frame number, which results in containing less information in each frame and

Table 2: FAR under different threshold

τ	**MDCT-NMF**	**WPD-QT**	**LPC-NMF**	**Proposed**
0.10	2.94×10^{-21}	5.47×10^{-31}	1.48×10^{-22}	3.17×10^{-32}
0.15	1.14×10^{-16}	4.60×10^{-24}	1.05×10^{-17}	5.74×10^{-25}
0.20	1.11×10^{-12}	4.60×10^{-18}	1.73×10^{-13}	1.09×10^{-18}
0.25	2.75×10^{-09}	5.50×10^{-13}	6.75×10^{-10}	2.23×10^{-13}
0.30	1.68×10^{-06}	7.97×10^{-09}	6.25×10^{-07}	4.88×10^{-09}

Table 3: The comparison of entropy rate

Algorithm	**MDCT-NMF**	**WPD-QT**	**LPC-NMF**	**Proposed**
ER	0.5449	0.9510	0.6730	0.8308

Table 4: Content preserving operations

Type	**Parameters**	**Abbreviation**
Volume adjustment 1	-50%	V1
Volume adjustment 2	+50%	V2
Resampling 1	16-8-16 (kHz)	R1
Resampling 2	16-32-16 (kHz)	R1
Echo addition	100 ms, 0.5	E
Narrowband noise	AWGN,40 dB	NN
Low-pass filter 1	Butterworth filter, 3.4(kHz)	LP1
Low-pass filter 2	FIR filter, 3.4(kHz)	LP2
MP3 compression 1	32 kbps	M1
MP3 compression 2	128 kbps	M1

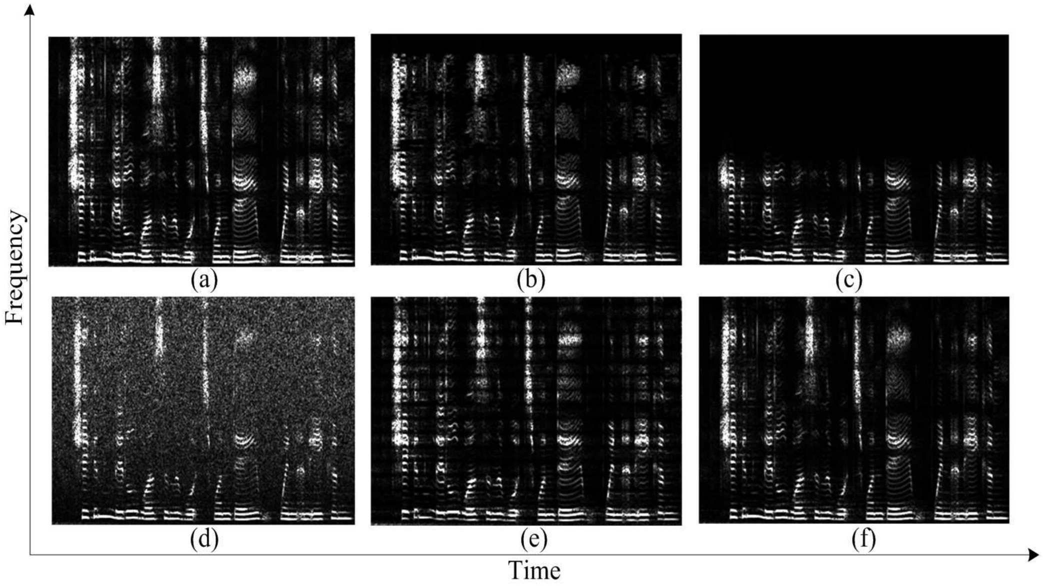


Figure 5: Spectrogram of (a) the 50%-volume-adding version of the 4 s original speech clip; (b) the 32 kbps-MP3-compressing version of the 4 s original speech clip; (c) the Butterworth-filtering version of the 4 s original speech clip; (d) 40 dB-noise-adding version of 4 s original speech clip; (e) echo-adding version of 4 s original speech clip; (f) resampling (16-8-16 kHz) version of the 4 s original speech clip

its degrading robustness. Although the wavelet packet transform can offer a more precise decomposition to signal frequency, the decomposing coefficients of signal high frequency are susceptible to noise. So the WPD-QT method is poor on noise resistance. Since the linear prediction analysis applies several past speech sampling values to approximate to current ones, therefore the change of amplitude has bad influence on LPC-NMF method. Because noise makes the spectrogram blurry, echo makes its textures overlapping. Therefore, the influence of noise and echo to proposed method is obvious. By the analysis above and combining the data in Table 5, the following information is obtained: in terms of robustness, the proposed algorithm outperforms LPC-NMF except the operation of volume addition, and it is also better than MDCT-NMF and WPD-QT on resisting the distortions caused by noise addition and filtering, but it is a little weaker than MDCT-NMF on resisting echo distortion.

In contrast to discrimination analysis, the false reject rate (FRR) is employed to estimate the robustness of a perceptual hashing system. It is the probability that the BERs of same two perceptual contents are more than τ. And its formula can be got from Equation (15).

$$FRR(\tau) = 1 - \int_{-\infty}^{\tau} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)}{2\sigma^2}} dx \qquad (15)$$

where τ is BER threshold, μ and σ are mean and standard deviation of BERs.

In order to describe the discrimination and robustness of the proposed algorithm more adequately, two kinds of BERs are utilized for probability analysis and drawing FAR-FRR curve in a coordinate system, one from the discrimination analysis in Section 4.1 and another from the robustness analysis in Section 4.2. Then, the robustness and discrimination of system can be evaluated by observing whether FAR curve and FRR curve cross, because if they have an intersection, the system cannot judge whether two speech clips are same perceptual contents in intersection area. Figure 6 shows the comparison of FAR-FRR curve between proposed approach and three other methods.

As can be seen in Figure 6(d), there is no one intersection on FAR-FRR curve of the proposed algorithm. Therefore, assuming the matching threshold τ is set to 0.3, when BER<0.3, the identification system can judge that two speech contents are perceptually same, otherwise different, which indicates the proposed algorithm has a better overall discrimination and robustness. For the discrimination analysis in Section 4.1, MDCT-NMF algorithm has a poor discrimination, and on the robustness analysis in Section 4.2, WPD-QT algorithm has a poor robustness against noise. Furthermore, there are bad effects on resisting the operations of noise adding and filtering in LPC-NMF algorithm. All these drawbacks of the three methods result in appearing an intersection on FAR-FRR curve in Figure 6(a), Figure 6(b) and Figure 6(c).

Through the analysis above, it is proved that the proposed algorithm shows satisfactory results on overall robustness and discrimination.

4.3 Efficiency Analysis

For illustrating the complexity and running efficiency of proposed algorithm, the running time is used to evaluate them with 100 speech clips selected randomly from the original speech database, which is took in the process of hashing generation and matching.

As can be seen from Table 6, compared with three other algorithms, the proposed algorithm takes a less time to generate and match hashing sequences. And its running efficiency is 4 times than LPC-NMF, 13 times than WPD-QT and 48 times than MDCT-NMF, which indicates that the proposed method obtains higher running efficiency. Furthermore, the hashing size in proposed algorithm is 234, 360 in LPC-NMF and MDCT-NMF, 250 in WPD-QT, which shows that the proposed algorithm has a stronger compaction.

From the analyses given above, the proposed algorithm has the advantages of high speed and few data, therefore it can meet the efficiency requirement of real-time speech communication.

5 Conclusions

In this paper, we proposed an efficient perceptual hashing scheme based on spectrogram for speech identification. By leveraging computer-vision method, the proposed algorithm adopts LBP to make texture information of a sub-spectrogram block more salient. As well as an image perceptual hashing method is utilized to generate hashing sequences from the sub-spectrogram block, which represents low frequency information of speech signal and is not sensitive to common content keeping distortions. Experimental results show that the proposed scheme achieves better discrimination to different speech clips and good robustness against some routine speech operations, such as noise addition, MP3 compression and filtering. Furthermore, the proposed scheme shows a high running efficiency and stronger compaction. This enables the proposed approach to be used in mobile speech real-time environment well.

There exist some issues to be handled in the proposed algorithm. For example, the robustness resisting echo and volume adjustment need to be further improved. And the performance analysis of speech fragment tampering attack has not yet been taken into account.

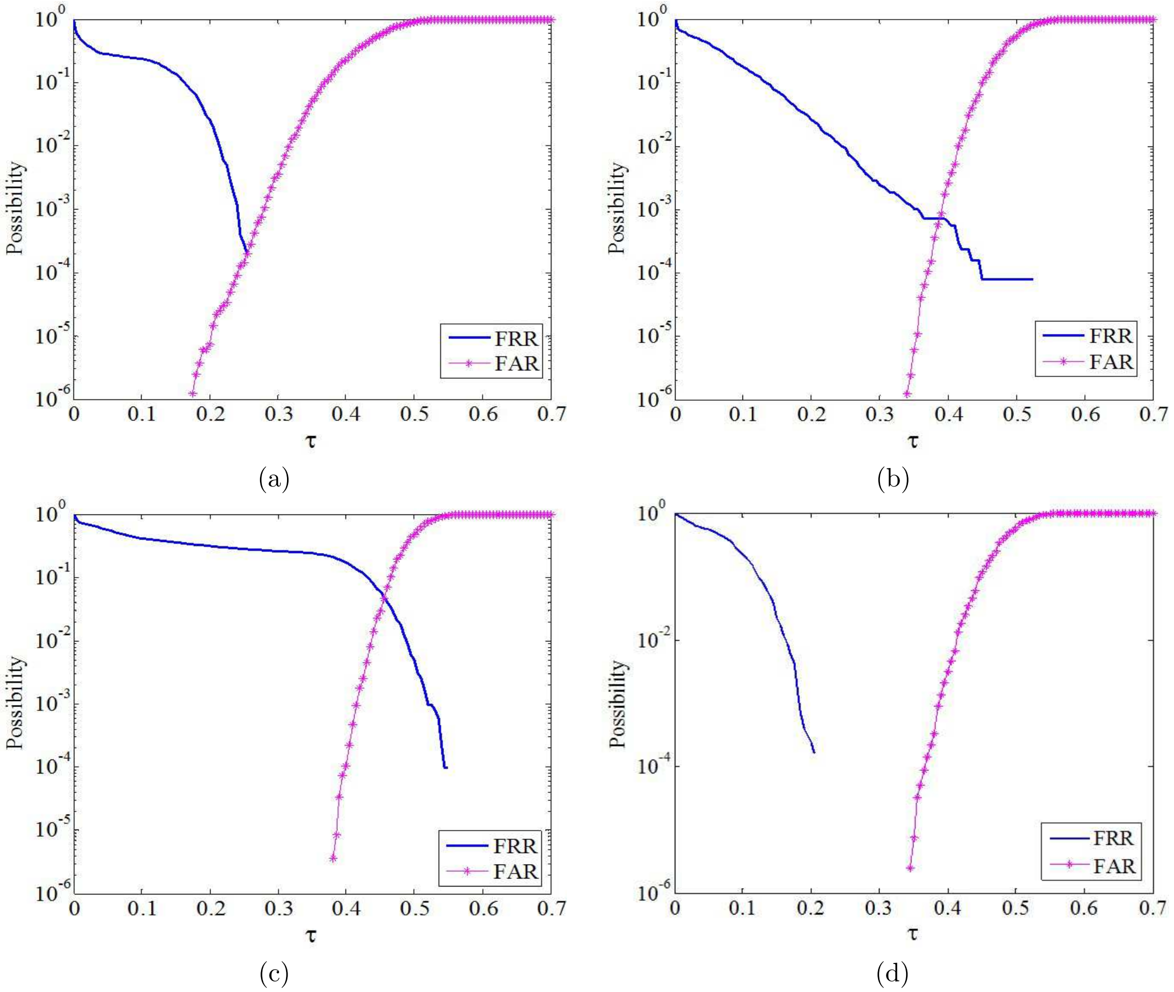
In future work, we will focus on extracting less and more salient features to overcome various degradations in spectrogram and research the tampering attack performance of proposed algorithm.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No.61363078), the Natu-

Table 5: Robustness test

Type	**MDCT-NMF**		**WPD-QT**		**LPC-NMF**		**Proposed**	
	Mean	Max	Mean	Max	Mean	Max	Mean	Max
V1	0.0040	0.0722	0.0008	0.0022	0.0016	0.0330	0.0179	0.0940
V2	0.0256	0.1056	0.0082	0.0100	0.0415	0.0917	0.0479	0.1624
R1	0.0012	0.0194	0.0036	0.0352	0.0260	0.1250	0.0094	0.0470
R2	0.0098	0.0750	0.0489	0.2266	0.1219	0.4028	0.0263	0.0149
E	0.0923	0.1827	0.1066	0.2305	0.2015	0.3000	0.1260	0.2051
NN	0.1357	0.2086	0.1452	0.5273	0.3464	0.5250	0.0918	0.2094
LP1	0.1422	0.2500	0.0864	0.2617	0.4098	0.5389	0.0784	0.1667
LP2	0.1615	0.2583	0.0924	0.2695	0.4303	0.5500	0.0813	0.1851
M1	0.0218	0.0722	-	-	0.1147	0.2920	0.1097	0.1838
M2	0.0035	0.0389	-	-	0.0727	0.2810	0.0248	0.0855



(a)

(b)

(c)

(d)

Figure 6: BER normal distribution diagram. (a) FAR-FRR curve of MDCT-NMF algorithm; (b) FAR-FRR curve of WPD-QT algorithm; (c) FAR-FRR curve of LPC-NMF algorithm; (d) FAR-FRR curve of proposed algorithm

ral Science Foundation of Gansu Province of China (No.1310RJYA004). The authors would like to thank the anonymous reviewers for their helpful comments and suggestions.

References

[1] N. Chen and W. Wan, "Robust speech hash function," *ETRI Journal*, vol. 32, no. 2, pp. 345–347, 2010.

[2] N. Chen, H. D. Xiao, J. Zhu, J. J. Lin, Y. Wang, and

Table 6: Running time

Algorithm	**MDCT-NMF**	**WPD-QT**	**LPC-NMF**	**Proposed**
File length(s)	4	4	4	4
Dominant frequency (GHz))	2.5	2.5	3.3	2.6
Total (s)	130.4	36.81	12.47	2.67

W. H. Yuan, "Robust audio hashing scheme based on cochleagram and cross recurrence analysis," *Electron Letters*, vol. 49, no. 1, pp. 7–8, 2013.

[3] N. Chen and H. D. Xiao, "Perceptual audio hashing algorithm based on zernike moment and maximum-likelihood watermark detection," *Digital Signal Processing*, vol. 23, no. 4, pp. 1216–1227, 2013.

[4] W. R. Ghanem, M. Shokir, and M. Dessoky, "Defense Against Selfish PUEA in Cognitive Radio Networks Based on Hash Message Authentication Code," *International Journal of Electronics and Information Engineering*, vol. 4, no. 1, pp. 12–21, 2016.

[5] Y. B. Huang, Q. Y. Zhang, and W. J. Hu, "Robust speech perception hashing authentication algorithm based on spectral subtraction and multi-feature tensor," *International Journal of Network Security*, vol. 20, no. 2, pp. 206–216, 2018.

[6] Y. B. Huang, Q. Y. Zhang, and Z. T. Yuan, "Perceptual speech hashing identification algorithm based on linear prediction analysis," *Telkomnika Indonesian Journal of Electrical Engineering*, vol. 12, no. 4, pp. 3214–3223, 2014.

[7] S. S. Kozat, R. Venkatesan, and M. K. Mihcak, "Robust perceptual image hashing via matrix invariants," in *International Conference on Image Processing (ICIP '04)*, pp. 3443–3446, Oct. 2004.

[8] J. F. Li, H. X. Wang, and J. Yi, "Audio perceptual hashing based on NMF and MDCT coefficients," *Chinese Journal of Electronics*, vol. 24, no. 3, pp. 579–588, 2015.

[9] J. F. Li, T. Wu, and H. X. Wang, "Perceptual hashing based on correlation coefficient of MFCC for speech authentication," *Journal of Beijing University of Posts and Telecommunications*, vol. 38, no. 2, pp. 89–93, 2015.

[10] N. M. Makbol, B. E. Khoo, and T. H. Rassem, "Block-based discrete wavelet transform-singular value decomposition image watermarking scheme using human visual system characteristics," *IET Image Process*, vol. 10, no. 1, pp. 34–52, 2016.

[11] S. S. Nassar, N. M. Ayad, H. M. Hamdy, H. M. Kelash, H. S. El-sayed, M. A. M. El-Bendary, F. E. A. El-Samie, and O. S. Faragallah, "Efficient audio integrity verification algorithm using discrete cosine transform," *International Journal of Speech Technology*, vol. 19, no. 1, pp. 1–8, 2016.

[12] Z. Rafii, B. Coover, and J. Y. Han, "Robust audio hashing scheme based on cochleagram and cross recurrence analysis," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 644–648, May 2014.

[13] C. S. Rao and S. B. G. T. Babu, *Image Authentication Using Local Binary Pattern on the Low Frequency Components*, India, 2016.

[14] Q. Y. Zhang, P. F. Xing, Y. B. Huang, R. H. Dong, and Z. P. Yang, "An efficient speech perceptual hashing authentication algorithm based on wavelet packet decomposition," *Journal Information Hiding Multimedia Signal Process*, vol. 6, no. 2, pp. 311–322, 2015.

[15] X. Zhang, B. L. Zhu, L. W. Li, W. Li, X. Q. Li, W. Wang, P. Z. Lu, and W. Q. Zhang, "Sift-based local spectrogram image descriptor: A novel feature for robust music identification," *EURASIP Journal on Audio Speech & Music Processing*, vol. 2015, no. 6, pp. 1–15, 2015.

[16] B. L. Zhu, W. Li, Z. R. Wang, and X. Y. Xue, "A novel audio fingerprinting method robust to time scale modification and pitch shifting," in *In Proceedings of the 18th ACM international conference on Multimedia*, pp. 987–990, Oct. 2010.

Biography

Zhang Qiu-yu. Researcher/Ph.D. supervisor, graduated from Gansu University of Technology in 1986, and then worked at school of computer and communication in Lanzhou University of Technology. He is vice dean of Gansu manufacturing information engineering research center, a CCF senior member, a member of IEEE and ACM. His research interests include network and information security, information hiding and steganalysis, multimedia communication technology.

Zhang Tao. received the BS degrees in communication engineering from Lanzhou University of Technology, Gansu, China, in 2015. His research interests include audio signal processing and application, multimedia authentication techniques.

Qiao Si-bin. received the BS degrees in communication engineering from Lanzhou University of Technology, Gansu, China, in 2014. His research interests include audio signal processing and application, multimedia authentication techniques.

Wu Dong-fang. In 2015, Wu Dongfang obtained his bachelor of engineering degree from Northwest University for Nationalities. Currently, he is studying for his master's degree at Lanzhou University of Technology. His research focuses on the industrial control network security.

A Note on the Construction of Lightweight Cyclic MDS Matrices

Akbar Mahmoodi Rishakani[1], Yousef Fekri Dabanloo[1], Seyed Mojtaba Dehnavi[2],
Mohammad Reza Mirzaee Shamsabad[3], Nasour Bagheri[4]
(Corresponding author: Akbar Mahmoodi Rishakani)

Department of Mathematics, Shahid Rajaee Teacher Training University, Tehran, Iran[1]
Shabanlou, Lavizan, Tehran, Iran
Department of Mathematical and Computer Sciences, University of Kharazmi, Tehran, Iran[2]
Department of Mathematics, Shahid Beheshti University, Tehran, Iran[3]
Department of Electrical Engineering, Shahid Rajaee Teachers Training University[4]
(Email: am.rishakani@sru.ac.ir)

(Received Aug. 22, 2017; revised and accepted Jan. 12, 2018)

Abstract

Modern lightweight block ciphers and hash functions apply linear layers for the diffusion purpose. In this paper, we characterize a class of lightweight MDS matrices decomposed into two cyclic matrices. As the main contribution, we presents a class of lightweight 4×4 cyclic MDS matrices lighter than the state-of-the-art which reduces the implementation cost (in terms of number of XOR gates required) of linear diffusion layers for hardware-oriented cryptographic primitives.

Keywords: Branch Number; Cyclic Matrix; Diffusion Layer; Lightweight Cryptographic Primitive; MDS Matrix

1 Introduction

Many modern lightweight block ciphers and hash functions apply MDS or almost MDS matrices as diffusion layers. For example, Midori [3] and QARMA [1, 9] families of block ciphers use almost MDS matrices and LED block cipher [10] and PHOTON hash function [11] use MDS matrices as diffusion layers. The performance of a diffusion layer depends on its branch number and implementation cost which is usually measured by the number of XORs required. Since the branch number of an MDS matrix is already maximum, for constrained applications like RFID and IoT [7, 8, 19, 21, 22], the implementation cost remains the main concern. For this purpose, we provide a hardware-efficient class of lightweight 4×4 cyclic MDS matrices.

1.1 Related Works

Providing MDS matrices that can be implemented with as few XOR operations as possible is one of the essentials in the design of lightweight symmetric primitives.

The XOR metric for measuring the efficiency of hardware implementations was first presented in [13] and later improved in [4] and [12]. Based on results from [18], many publications tried to find as efficient MDS matrices as possible.

In [4], the authors present lightweight cyclic MDS matrices by the use of lightweight multiplication in $\mathbb{F}_{2^m}$ (the field with 2^m elements). The cost of their presented 4×4 MDS matrices is $12m + 12$ XORs, $4 \leq m \leq 8$. Here, m is the size of input words. The authors of [15] construct lightweight 4×4 cyclic MDS matrices with implementation cost of 60 and 108 XORs for 4-bit and 8-bit input words, respectively.

Bai and Wang [2] characterize lightweight 4×4 MDS matrices with 4-bit input words for which the entries implementation needs 10 XORs and overall, the entire matrix requires $4 \times 12 + 10 = 58$ XORs for implementation. Then, a class of 4×4 MDS matrices proposed with the help of Toeplitz matrices with 58 XORs for 4-bit and 123 XORs for 8-bit input words by Sarkar *et al.* in [17]. Later, in [6] Cauchois *et al.* constructed quasi-involutory recursive-like MDS matrices from 2-cyclic codes for which the implementation cost of 4×4 MDS matrices with 4-bit input words is 72 XORs. Zhang *et al.* in [23] provide cyclic 4×4 MDS matrices with 4-bit input words and 12 XORs for entries which overall requires $4 \times 12 + 12 = 60$ XORs for implementation. Recently, Zhou *et al.* [20] proposed two kinds of lightweight 4×4 MDS matrices over 4-bit and 8-bit input words which require $4 \times 12 + 10 = 58$ and $8 \times 12 + 10 = 106$ XORs, respectively.

1.2 Our Contribution

In most of recently presented lightweight primitives, *e.g.* QARMA [1] and Midori [3] block ciphers, almost MDS matrices are used due to the low implementation cost, *i.e.* 24 XORs for 4-bit input words; while the lightest MDS ones take 58 XORs (before this paper). Hence, there is a significant gap between the implementation cost of almost MDS matrices and MDS matrices. On the other hand, employing an almost MDS matrix as diffusion layer, in general, provides lower security bounds for the same number of rounds. Thus, in this paper, we took a step forward to reduce the gap between the implementation cost of almost MDS matrices and MDS ones, to motivate designers to use MDS matrices.

Our concern in this paper is to construct lightweight 4×4 cyclic MDS matrices with efficient implementation in hardware, measured by the number of XOR gates required. We construct 4×4 lightweight MDS matrices by the multiplication of two cyclic matrices. More precisely, one of its multiplicands is a 4×4 cyclic matrix whose entries are binary permutation matrices (which have no implementation cost in hardware) and the other is a cyclic matrix with two non-zero entries per row. We characterize the MDS property of this type of matrices. As a result, we provide lightweight 4×4 cyclic MDS matrices on m-bit input words with the implementation cost of $10m + 4$ XORs for $4 \leq m \leq 8$.

Note that our results would be infeasible without our new approach of representing the MDS matrix as a product. This is because of the fact that an MDS matrix cannot have zero entries. So, a 4×4 MDS matrix over m-bit input words would need already $12m$ XORs only for the additions within the matrix multiplication, which exceeds our results. That is, we benefit from being able to use matrices with many zero entries.

We believe that, it is irrelevant to compare the implementation cost of cyclic MDS matrices with non-cyclic ones, but since cyclic MDS matrices are less studied, we compare our results with cyclic and non-cyclic matrices in Table 1 for m-bit input words, $m = 4, 8$ (details are given in sections 3 and 4). Note that, in Table 1, by $\#A = \#A^{-1}$ we mean the matrices A for which the implementation cost of A and A^{-1} are equal.

1.3 Outline of the Paper

In Section 2, we give the preliminary notations and definitions. Section 3 presents new criteria for constructing cyclic MDS matrices. In Section 4, we verify the implementation cost of our constructions and their inverses. Section 5 concludes the paper.

2 Preliminaries

In this paper, n and m are natural numbers. By $|A|$ we denote the number of elements of a finite set A. We denote the set of all $n \times n$ matrices with entries in R by $\mathcal{M}_n(R)$ and the determinant of a matrix A in $\mathcal{M}_n(R)$ by $det_R(A)$. The XOR of two binary vectors or matrices v and w is denoted by $v \oplus w$, a zero vector or matrix by $\mathbf{0}$ and an identity matrix by I. We represent the finite field with 2 elements by $\mathbb{F}_2$ and use $\mathbb{F}_2^m$ to represent the set of all m-bit vectors. We denote by $\#A$, the number of XORs needed to implement the binary matrix $A \in \mathcal{M}_n(\mathbb{F}_2)$.

By the notation $A = cycl(a_1, a_2, a_3, \ldots, a_n)$ we mean the cyclic matrix

$$A = \begin{pmatrix} a_1 & a_2 & a_3 & \ldots & a_n \\ a_n & a_1 & a_2 & \ldots & a_{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_2 & a_3 & a_4 & \ldots & a_1 \end{pmatrix}$$

For a vector $v = (v_3, v_2, v_1, v_0) \in \mathbb{F}_2^4$ we correspond a number $\bar{v} = \sum_{v_i \neq 0} 2^i$ in hexadecimal representation. So, a matrix $M \in \mathcal{M}_4(\mathbb{F}_2)$ could be represented by four numbers (in hexadecimal representation) corresponding to its rows. For instance, the following matrix is represented by *7bde*:

$$M = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \qquad (1)$$

The set of all invertible binary matrices of order m that have exactly one non-zero entry in each row is denoted by $\mathcal{PM}_m(\mathbb{F}_2)$ and called binary permutation matrices. For example, by our notations, the matrix $2814 \in \mathcal{PM}_4(\mathbb{F}_2)$ maps the vector $x = (x_3, x_2, x_1, x_0) \in \mathbb{F}_2^4$ to (x_2, x_0, x_3, x_1). So, for any $A \in \mathcal{PM}_m(\mathbb{F}_2)$ and $x \in \mathbb{F}_2^m$, $y = xA$ is a vector whose components are a permutation of the components of x and $\#A = 0$.

The i-th component of a vector $x \in (\mathbb{F}_2^m)^n$ is denoted by x_i, *i.e.* $x = (x_{n-1}, ..., x_0)$. The weight of a vector $x \in (\mathbb{F}_2^m)^n$ with respect to m-bit input words is denoted by $wt_m(x)$ and defined as

$$wt_m(x) = |\{\ x_i : x_i \neq 0, 0 \leq i \leq n-1\}|.$$

For example, let

$$x = 1001110000101110,$$

we have, $wt_1(x) = 8$, $wt_2(x) = 6$ and $wt_4(x) = 4$.

Definition 1. *[5] Let $M \in \mathcal{M}_n(\mathcal{M}_m(\mathbb{F}_2))$. The (differential) branch number of M with respect to m-bit input words is defined as*

$$\mathcal{B}_m(M) = \min_{x \neq \mathbf{0}} \{wt_m(x) + wt_m(xM) : x \in (\mathbb{F}_2^m)^n\}.$$

For the matrix M defined in Equation (1), we have $\mathcal{B}_1(M) = 4$, i. e. $wt_1(x) + wt_1(xM) \geq 4$ for any nonzero $x \in \mathbb{F}_2^4$. On the other hand, the matrix M could be considered as a 2×2 matrix with entries in $\mathcal{M}_2(\mathbb{F}_2)$, i. e.

$$M = \left(\begin{array}{cc|cc} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ \hline 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{array}\right) \in \mathcal{M}_2(\mathcal{M}_2(\mathbb{F}_2))$$

Table 1: Comparison between the implementation cost of 4 × 4 MDS matrices, for m-bit input words

Source	Cyclic	$\#A = \#A^{-1}$	XOR count (m=4 / m=8)
[4]	$\surd$	X	60 / 108
[2]	X	X	58 / -
[15]	$\surd$	X	60 / 108
[15]	$\surd$	$\surd$	68 / -
[17]	X	X	58 / 123
[12]	X	X	58 / 116
[16]	$\surd$	$\surd$	60 / 128
[20]	$\surd$	X	58 / 106
[23]	$\surd$	X	60 / -
This paper	$\surd$	X	44 / 84

In this case, one can check that $\mathcal{B}_2(M) = 2$.

It is straightforward to verify that for a matrix $M \in \mathcal{M}_n(\mathcal{M}_m(\mathbb{F}_2))$ we have $\mathcal{B}_m(M) \leq n + 1$.

Definition 2. *[5] A matrix $M \in \mathcal{M}_n(\mathcal{M}_m(\mathbb{F}_2))$ is called MDS with respect to m-bit input words if*

$$\mathcal{B}_m(M) = n + 1.$$

It is worth noting that, with the help of an $n \times n$ MDS matrix with respect to m-bit input words, we can construct an MDS code of length $2n$ over m-bit alphabets [5].

3 Constructing Lightweight 4 × 4 Cyclic MDS Matrices

In this section we verify a class of cyclic 4 × 4 matrices to give sufficient conditions when they are MDS. The proposed matrices are the product of two cyclic matrices such that the non-zero entries of the first factor are in $\mathcal{PM}_m(\mathbb{F}_2)$ and the non-zero entries of the second factor belong to $\mathcal{M}_m(\mathbb{F}_2)$.

For the mentioned verification, we need the following theorems.

Theorem 1. *[5] For $M \in \mathcal{M}_n(\mathcal{M}_m(\mathbb{F}_2))$, M is MDS with respect to m-bit input words if and only if every square submatrix of M of order t, $1 \leq t \leq n$, is invertible.*

Theorem 2. *[14] For $M \in \mathcal{M}_n(\mathcal{M}_m(\mathbb{F}_2))$, if the entries of M in $R = \mathcal{M}_m(\mathbb{F}_2)$ are pairwise commuting, then*

$$det_{\mathbb{F}_2}(M) = det_{\mathbb{F}_2}(det_R(M)).$$

The following lemma is a straightforward result of Theorem 1.

Lemma 1. *Let $A \in \mathcal{M}_m(\mathbb{F}_2)$ and $P_1, P_2, P_3, P_4 \in \mathcal{PM}_m(\mathbb{F}_2) \bigcup \{\mathbf{0}\}$. If the matrix*

$$M = cycl(P_1, P_2, P_3, P_4) \times cycl(I, A, \mathbf{0}, \mathbf{0}),$$

is MDS with respect to m-bit input words, then at most one of P_i's, $1 \leq i \leq 4$, could be zero.

According to Lemma 1, we verify the following class of matrices in more details:

$$M = cycl(P_1, P_2, P_3, \mathbf{0}) \times cycl(I, A, \mathbf{0}, \mathbf{0}). \qquad (2)$$

It is easy to verify that, the matrix defined in Equation (2) is MDS if and only if the matrix

$$M' = cycl(I, P_1^{-1}P_2, P_1^{-1}P_3, \mathbf{0}) \times cycl(I, A, \mathbf{0}, \mathbf{0}),$$

is MDS. As $\mathcal{PM}_m(\mathbb{F}_2)$ is closed under multiplication and inversion, we have $p_1 = P_1^{-1}P_2, p_2 = P_1^{-1}P_3 \in \mathcal{PM}_m(\mathbb{F}_2)$. Now if we assume that p_1, p_2 and A are pairwise commuting, then the determinants of the square submatrices of

$$\begin{aligned} M' &= cycl(I, p_1, p_2, \mathbf{0}) \times cycl(I, A, \mathbf{0}, \mathbf{0}) \\ &= cycl(I, A \oplus p_1, p_1 A \oplus p_2, p_2 A), \end{aligned}$$

are as following.

(a) 1 × 1 submatrices:

$$I, A \oplus p_1, p_1 A \oplus p_2, p_2 A.$$

(b) 2 × 2 submatrices:

$$\begin{aligned} &p_1 p_2 A \oplus I, p_1(A^2 \oplus p_1 A \oplus p_2), \\ &A^2 \oplus p_1 A \oplus p_1^2 \oplus p_2, p_2^2 A^2 \oplus p_1 A \oplus p_2, \\ &p_1 p_2 A^2 \oplus (p_2^2 \oplus I)A \oplus p_1, p_1^2 A^2 \oplus p_2^2 \oplus I, \\ &(p_1^2 \oplus p_2)A^2 \oplus p_1 p_2 A \oplus p_2^2, (p_2^2 \oplus I)A^2 \oplus p_1^2, \\ &p_2 A^2 \oplus p_1 p_2 A \oplus I. \end{aligned}$$

(c) 3 × 3 submatrices:

$$\begin{aligned} &(p_1 \oplus p_1 p_2^2)A^3 \oplus (p_2 \oplus p_1^2 \oplus p_2^3)A^2 \oplus p_1^3 A \\ &\qquad \oplus p_2^2 \oplus p_1^2 p_2 \oplus I, \\ &(p_2 \oplus p_1^2 \oplus p_2^3)A^3 \oplus p_1^3 A^2 \oplus (p_1^2 p_2 \oplus p_2^2 \oplus I)A \\ &\qquad \oplus p_1 \oplus p_1 p_2^2, \\ &p_1^3 A^3 \oplus (p_1^2 p_2 \oplus p_2^2 \oplus I)A^2 \oplus (p_1 \oplus p_1^2 p_2)A \\ &\qquad \oplus p_1^2 \oplus p_2^3 \oplus p_2, \\ &(p_2^2 \oplus p_1^2 p_2 \oplus I)A^3 \oplus (p_1 \oplus p_1 p_2^2)A^2 \\ &\qquad \oplus (p_1^2 \oplus p_2 \oplus p_2^3)A \oplus p_1^3. \end{aligned}$$

(d) 4×4 submatrices:

$$(I \oplus p_1^2 p_2 \oplus p_2 p_1^2 \oplus p_1^4 \oplus p_2^4)(I \oplus A^4).$$

According to Theorem 1 and Theorem 2, M' is MDS with respect to m-bit input words if and only if the aforementioned submatrices are invertible.

In the special case of $p_1 = p_2 = I$, we have the following theorem.

Theorem 3. *Let $A \in \mathcal{M}_m(\mathbb{F}_2)$. The matrix*

$$M = cycl(I, I, I, \mathbf{0}) \times cycl(I, A, \mathbf{0}, \mathbf{0}) \tag{3}$$

is MDS with respect to m-bit input words if and only if $A, A^3 \oplus I, A^7 \oplus I$ are invertible.

Proof. By replacing p_1 and p_2 with I in matrices of (a),(b),(c) and (d), it results that M is MDS if and only if the following matrices are invertible.

$$\begin{gathered} I,\ A,\ I \oplus A,\ A^2,\ (I \oplus A)^2, \\ A(I \oplus A),\ I \oplus A \oplus A^2,\ I \oplus A \oplus A^3, \\ I \oplus A^2 \oplus A^3,\ A(I \oplus A \oplus A^2), \\ A^3 \oplus A^2 \oplus A \oplus I,\ (I \oplus A)^4. \end{gathered} \tag{4}$$

Given that $I \oplus A^3 = (I \oplus A)(I \oplus A \oplus A^2)$, $I \oplus A^7 = (I \oplus A)(I \oplus A \oplus A^3)(I \oplus A^2 \oplus A^3)$, $(I \oplus A)^4 = (I \oplus A)(A^3 \oplus A^2 \oplus A \oplus I)$ and regarding (4), all submatrices of M are invertible if and only if $A, I \oplus A^3, I \oplus A^7$ are invertible, which completes the proof. □

Similar to Theorem 3, the next theorem could be proved.

Theorem 4. *Let $A \in \mathcal{M}_m(\mathbb{F}_2)$. The matrix*

$$M = cycl(I, I, I, \mathbf{0}) \times cycl(I, \mathbf{0}, \mathbf{0}, A)$$

is MDS with respect to m-bit input words if and only if $A, A^3 \oplus I, A^7 \oplus I$ are invertible.

Similarly, we verified the MDS property of matrices

$$M = cycl(I, I, I, \mathbf{0}) \times cycl(I, \mathbf{0}, A, \mathbf{0}).$$

We found out that, there is no matrix $A \in \mathcal{M}_m(\mathbb{F}_2)$ such that M is MDS.

4 Implementation and Experimental Results

In this section, we discuss the implementation cost of the 4×4 cyclic MDS matrices given in Theorem 3 and Theorem 4 and their corresponding inverses. For the matrix M in Equation (3), we have

$$\#M = 10m + 4\#A. \tag{5}$$

This is because the implementation cost of $C = cycl(I, A, \mathbf{0}, \mathbf{0})$ would be $4m + 4\#A$ XORs; since, for the action of C on input words, we should apply A four times plus extra $4m$ XORs for the additions within matrix multiplication. On the other hand, to implement $B = cycl(I, I, I, \mathbf{0})$, we use the following procedure:

$$\begin{aligned} (X_3, X_2, X_1, X_0)B &= (Y_3, Y_2, Y_1, Y_0), \\ Z_0 = X_1 \oplus X_2, Z_1 &= X_0 \oplus X_3, \\ Y_0 = X_0 \oplus Z_0, Y_1 &= X_3 \oplus Z_0, \\ Y_2 = X_2 \oplus Z_1, Y_3 &= X_1 \oplus Z_1, \end{aligned}$$

which shows that B needs $6m$ XORs. By the same calculations, the implementation cost of matrices M verified in Theorem 4 equals to $10m + 4\#A$ XORs.

Now according to Equation (5), the construction of lightweight 4×4 MDS matrices with respect to m-bit input words, $4 \leq m \leq 8$, given in Theorem 3 and Theorem 4, would be reduced to finding invertible matrices $A \in \mathcal{M}_m(\mathbb{F}_2)$ with as low implementation cost as possible such that $I \oplus A^3$ and $I \oplus A^7$ are invertible. Every invertible matrix A with $\#A = 0$ belongs to $\mathcal{PM}_m(\mathbb{F}_2)$; so, at least one of the non-zero entries of A would be on its principal diagonal, i. e. one of the rows of A equals to the corresponding row of I. This means that $I \oplus A$ could not be invertible. Thus, we should search for matrices A with $\#A = 1$.

For this purpose, we have exhaustively searched the proposed matrices A in $\mathcal{S}_m$, $4 \leq m \leq 8$, where, $\mathcal{S}_m$ is the set of all binary matrices $A \in \mathcal{M}_m(\mathbb{F}_2)$, for which just one of the rows has two non-zero entries and the other rows have only one non-zero entry. Clearly, $|\mathcal{S}_m| = \binom{m}{2} m^m$ and for every $A \in \mathcal{S}_m$, we have $\#A = 1$. It takes few hours to find all matrices $A \in \mathcal{S}_8$ ($|\mathcal{S}_8| = 7 \times 2^{26}$) such that A, $I \oplus A^3$ and $I \oplus A^7$ are invertible, by programming. Note that the case of $m = 8$ is the most time consuming case. As a result, we found 48, 240, 960, 480 and 25920 such matrices for $m = 4, 5, 6, 7, 8$, respectively. We present all 48 matrices for $m = 4$ as follows and give a list of five matrices for each of the other cases in Appendix.

$$\begin{gathered} 1286, 1294, 1846, 18c2, 1942, 1a84, 214a, 2158, \\ 281c, 2854, 2948, 2a14, 3814, 3842, 418a, 41c2, \\ 421c, 4298, 4318, 4382, 5182, 5284, 6148, 6218. \\ 1285, 12a4, 1684, 1843, 1862, 1c42, 2149, 2168, \\ 2548, 2815, 2834, 2c14, 4183, 41a2, 4219, 4238, \\ 4582, 4618, 9284, 9842, a148, a814, c182, c218. \end{gathered} \tag{6}$$

According to Equation (5), by choosing A from the list of the matrices (6) or from the Appendix, the implementation cost of the proposed MDS matrices with respect to m-bit input words, $4 \leq m \leq 8$, derived from Theorem 3 and Theorem 4 are $10m + 4$ XORs.

Often, when aiming for the most efficient MDS matrix, the inverse of the matrix is not considered and might have much higher implementation cost. This is because of the fact that, in many applications in symmetric cryptography, we do not need to implement the inverse of components. Examples of such applications include stream ciphers, hash functions, block ciphers in CTR and OFB modes or block ciphers with Feistel or Lai-Massy structures. Accordingly, most of the papers we are comparing with, have not verified the implementation cost of the inverse of their proposed MDS matrices [2,4,12,17,20,23]. However, we give an upper bound for the implementation cost of the inverse of our proposed MDS matrices as following.

For $B = cycl(I, I, I, \mathbf{0})$ we have, $B^{-1} = cycl(I, \mathbf{0}, I, I)$. Therefore, $\#B^{-1} = \#B = 6m$. On the other hand, if $C = cycl(I, A, \mathbf{0}, \mathbf{0})$, then $C^{-1} = \alpha C^3$, $\alpha = (I \oplus A^4)^{-1}$. As $C^2 = cycl(I, \mathbf{0}, A^2, \mathbf{0})$, we have

$$C^{-1} = \alpha C \times cycl(I, \mathbf{0}, A^2, \mathbf{0}).$$

By this decomposition, an upper bound for the implementation cost of C^{-1} is

$$(4m + 4\#A^2) + (4m + 4\#A) + 4\#\alpha.$$

So, the implementation cost of the inverse of the matrices derived from Theorem 3 and Theorem 4 would be bounded by

$$14m + 4\#A + 4\#A^2 + 4\#\alpha.$$

For $m = 4$, we presented 48 candidates of matrix A with 1 XOR implementation cost (listed in (6)), for which the corresponding matrices M in Theorem 3 and Theorem 4 are MDS. Among them, the first 24 matrices have the property that $I \oplus A^4 = A$. In this case, $\alpha C^2 = cycl(A^{-1}, \mathbf{0}, A, \mathbf{0})$. So,

$$C^{-1} = cycl(I, A, \mathbf{0}, \mathbf{0}) \times cycl(A^{-1}, \mathbf{0}, A, \mathbf{0}).$$

This means that, if we select A from the first 24 matrices of the list (6), then the implementation cost of corresponding cyclic MDS matrix M would be 44 XORs and the implementation cost of M^{-1} would be 68 XORs. Note that, for every invertible matrix $A \in \mathcal{S}_m$, we have $A^{-1} \in \mathcal{S}_m$; i. e. $\#A^{-1} = 1$.

To verify whether the matrices presented in Equation (2) are MDS or not, in the case of $m = 4$, we exhaustively checked P_1, P_2, P_3 and A. The total number of such matrices is $(4!)^3 2^{16} = 81 \times 2^{25}$. The result of our programming shows that for all of the resultant MDS matrices, we have $P_1 = P_2 = P_3$.

5 Conclusion

In this paper, we proposed a new class of lightweight 4×4 cyclic MDS matrices with respect to m-bit input words based on the product of cyclic matrices. The resultant MDS matrices need $10m + 4$ XORs for implementation. In comparison to the state-of-the-art, to the best of our knowledge, our proposed cyclic MDS matrices outperform the previous known cyclic MDS matrices.

References

[1] R. Avanzi, "The QARMA block cipher family. Almost MDS matrices over rings with zero divisors, nearly symmetric even-mansour constructions with non-involutory central rounds, and search heuristics for low-latency s-boxes," *IACR Transactions on Symmetric Cryptology*, vol. 2017, pp. 4–44, 2017.

[2] J. Bai and D. Wang, "The lightest 4x4 MDS matrices over GL(4, $\mathbb{F}_2$)," *IACR Cryptology ePrint Archive*, vol. 2016, pp. 686, 2016.

[3] S. Banik, A. Bogdanov, T. Isobe, K. Shibutani, H. Hiwatari, T. Akishita, and F. Regazzoni, "Midori: A block cipher for low energy (Extended version)," *Cryptology ePrint Archive*, Report 2015/1142, 2015.

[4] C. Beierle, T. Kranz, and G. Leander, "Lightweight multiplication in GF(2ˆn) with applications to MDS matrices," in *Advances in Cryptology (CRYPTO'16))*, vol. 9814 of *Lecture Notes in Computer Science*, pp. 625–653, 2016.

[5] M. Blaum and R. M. Roth, "On lowest density MDS codes," *IEEE Translation Information Theory*, vol. 45, no. 1, pp. 46–59, 1999.

[6] V. Cauchois, P. Loidreau, and N. Merkiche, "Direct construction of quasi-involutory recursive-like MDS matrices from 2-cyclic codes," *Cryptology ePrint Archive*, Report 2016/1112, 2016.

[7] Y. C. Chen, W. L. Wang, M. S. Hwang, "RFID authentication protocol for anti-counterfeiting and privacy protection", in *The 9th International Conference on Advanced Communication Technology*, pp. 255–259, 2007.

[8] N. Chikouche, C. Foudil, P. L. Cayrel, and M. Benmohammed, "Improved RFID authentication protocol based on randomized mceliece cryptosystem," *International Journal Network Security*, vol. 17, no. 4, pp. 413–422, 2015.

[9] W. R. Ghanem, M. Shokir, and M. Dessoky, "Defense Against Selfish PUEA in Cognitive Radio Networks Based on Hash Message Authentication Code," *International Journal of Electronics and Information Engineering*, vol. 4, no. 1, pp. 12–21, 2016.

[10] A. P. J. Guo, T. Peyrin and M. Robshaw, "The led block cipher," *Cryptology ePrint Archive*, Report 2012/600, 2012.

[11] J. Guo, T. Peyrin, and A. Poschmann, "The PHOTON family of lightweight hash functions," in *Advances in Cryptology (CRYPTO'11)*, vol. 6841 of *Lecture Notes in Computer Science*, pp. 222–239, 2011.

[12] J. Jean, T. Peyrin, and S. M. Sim, "Optimizing implementations of lightweight building blocks," *Cryptology ePrint Archive*, Report 2017/101, 2017.

[13] K. Khoo, T. Peyrin, A. Y. Poschmann, and H. Yap, "Foam: Searching for hardware-optimal spn structures and components with a fair comparison," in *Cryptographic Hardware and Embedded Systems (CHES'14)*, pp. 433–450, 2014.

[14] I. Kovacs, D. S. Silver, and S. G. Williams, "Determinants of commuting-block matrices," *The American Mathematical Monthly*, vol. 106, no. 10, pp. 950–952, 1999.

[15] Y. Li and M. Wang, "On the construction of lightweight circulant involutory MDS matrices," in *Fast Software Encryption (FSE'16)*, pp. 121–139, 2016.

[16] M. Liu and S. M. Sim, "Lightweight MDS generalized circulant matrices," in *Fast Software Encryption (FSE'16)*, pp. 101–120, 2016.

[17] S. Sarkar and H. Syed, "Lightweight diffusion layer: Importance of toeplitz matrices," *Cryptology ePrint Archive*, Report 2016/835, 2016.

[18] S. M. Sim, K. Khoo, F. Oggier, and T. Peyrin, "Lightweight mds involution matrices," in *Fast Software Encryption (FSE'15)* , pp. 471–493, 2015.

[19] W. L. Tai and Y. F. Chang, "Comments on a secure authentication scheme for iot and cloud servers," *International Journal Network Security*, vol. 19, no. 4, pp. 648–651, 2017.

[20] L. Wang, L. Zhou and Y. Sun, "Construction of lightweight MDS matrices over matrix polynomial residue ring," Cryptology ePrint Archive, Report 2016/1173, 2016.

[21] C. H. Wei, M. S. Hwang, A. Y. H. Chin, "A mutual authentication protocol for RFID", *IEEE IT Professional*, vol. 13, no. 2, pp. 20–24, Mar. 2011.

[22] C. H. Wei, M. S. Hwang, A. Y. H. Chin, "A secure privacy and authentication protocol for passive RFID tags," *International Journal of Mobile Communications*, vol. 15, no. 3, pp. 266–277, 2017.

[23] S. Zhang, Y. Wang, Y. Gao, and T. Wang, "On the construction of the 4 x 4 lightest circulant MDS matrices," in *Proceedings of the International Conference on Cryptography, Security and Privacy, (ICCSP'17)*, pp. 1–6, 2017.

Appendix

Here, for simplicity, a matrix $A \in \mathcal{M}_m(\mathbb{F}_2)$, $5 \leq m \leq 8$, is represented by a sequence of m decimal numbers corresponding to its rows. For example, $(1, 2, 8, 16, 6)$ represents the matrix

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \end{pmatrix}.$$

List of some matrices for $m = 5$:

$$(1, 2, 8, 16, 6), (2, 1, 10, 4, 16), (3, 4, 1, 8, 16),$$
$$(4, 2, 1, 17, 8), (12, 2, 16, 1, 4).$$

List of some matrices for $m = 6$:

$$(1, 2, 4, 16, 32, 10), (2, 1, 32, 20, 4, 8), (4, 2, 1, 16, 8, 34),$$
$$(12, 2, 1, 16, 8, 32), (33, 32, 16, 2, 8, 4).$$

List of some matrices for $m = 7$:

$$(1, 2, 4, 16, 32, 64, 40), (2, 8, 1, 4, 64, 16, 96),$$
$$(9, 1, 32, 16, 2, 64, 4), (17, 1, 4, 2, 32, 64, 8),$$
$$(34, 16, 8, 2, 1, 4, 64).$$

List of some matrices for $m = 8$:

$$(1, 2, 4, 8, 32, 64, 144, 16), (2, 1, 4, 64, 128, 8, 16, 40),$$
$$(20, 32, 1, 128, 16, 2, 64, 8), (80, 128, 8, 1, 4, 64, 32, 2),$$
$$(96, 128, 4, 8, 2, 16, 1, 64).$$

Biography

Akbar Mahmoodi Rishakani received his B.S. and M.S. degrees in pure mathematics from Shahid Beheshti University in 2005 and 2008 respectively. He is now PHD student of mathematical cryptography in Shahid Rajaee Teacher Training University under the supervision of Prof. Hamid Reza Maimani and Prof. Nasour Bagheri. His current research interests include information security, cryptology and combinatorics.

Yousef Fekri Dabanloo received the B.S. degree in pure mathematics from Semnan University in 2011, Semnan, Iran. He received the M.S. degree in 2013 from Sharif University of Technology, Tehran, Iran. Now he is as a PhD student in Shahid Rajaei University, Tehran, Iran. His current research interests include information security and cryptology.

Seyed Mojtaba Dehnavi was born in 1975 in Iran. He received his BSc in applied mathematics and hardware engineering in 2001 from Iranian University of Science and Technology, his MSc in pure mathematics in 2004 from Amir Kabir University of Technology, and his PhD in mathematical cryptography in 2015 from Kharazmi University under supervision of Prof. Hamid Reza Maimani.

Mohammad Reza Mirzaee Shamsabad was born in 1983 in Iran. He received his BSc in applied mathematics in 2006 from Azad University, his MSc in pure mathematics in 2010 from Shahid Bahonar University. He is now a candidate of PhD in mathematical cryptography in Shahid Beheshti University under supervision of Prof. Hossein Hajiabolhassan.

Nasour Bagheri is an assistant professor at electrical engineering department, Shahid Rajaee Teacher Training University, Tehran, Iran. He is the author of more than 60 articles on information security and cryptology. Homepage of the author is available at: https://www.srttu.edu/english-cv-dr-bagheri/.

Exploiting Incremental Classifiers for the Training of an Adaptive Intrusion Detection Model

Marwa R. Mohamed[1], Abdurrahman A. Nasr[2], I. F. Tarrad[3] and Salah R. Abdulmageed[2]

(Corresponding author: Marwa R. Mohamed)

Department of Communication Engineering, University of Helwan[1]
Cairo, Egypt
Department of Computers and Systems Engineering, University of Al Azhar[2]
Department of Communication Engineering, University of Al Azhar[3]
(Email: marwa.rizk@chi.edu.eg)

(Received Dec. 6, 2017; revised and accepted June 18, 2018)

Abstract

Due to the fact that network data is dynamic in nature, the demand for adaptive Intrusion Detection System (IDS) has increased for smart analysis of network data stream. An intrusion detection system is a component of the information security and its main aim is to detect abnormal activities of the network and tries to prevent suspicious data streams that might lead to network security breach. However, most IDS poverty to the capability to detect zero-day or previously unknown attacks. As such, two types of IDS have been contemplated for detecting network threats, namely, signature-based IDS and anomaly detection system. The former depends on stored signatures in a database (thus, its name) to detect intrusions, whereas the latter develops a model based on normal system or network behavior, with the aim of detecting both recognized and novel attacks. The two types of intrusion detection systems confront many problems comprehensive; continuous learning, scalability, a high rate of false alarm, and inability to work in the online model. Here, an Adaptive Intrusion Detection Model (AIDM) is proposed. Such model is an intelligent and learnable anomaly detection model that overcomes the problems of traditional anomaly detection systems namely, high false alarm, real-time learning, and scalability. In this paper, AIDM exploited and studied a set of different incremental machine learning classifiers for intelligent detection and analysis of network data streams is carried out. Such incremental classifiers are Non-Nearest Generalized Exemplar (NNGE), Incremental Naï̈ve Bayes (INB), Hoffeding Trees (HT), Instance-Based K- Nearest Neighbor (IBK) and Radial Basis Function Neural Network (RBFNN). Besides that, we utilized Deep Learning 4 Java Multilayer Perceptron (DL4JMLP) classifier for a deep learning approach. Furthermore, a comparison of results between seven machine learning classifiers has been performed to choose the best classifier result capable of recognizing the incoming unknown attacks from the network traffic. These classifiers are incremental in nature such that it can learn network data streams in real-time without the need for redeployment of network infrastructure. AIDM is evaluated using three different datasets collected from Defense Advanced Research Projects Agency (DARPA), Kyoto University and the Cyber Range Lab of the Australian Centre for Cyber Security (ACCS), the training model was obtained by the aforementioned data mining techniques. The evaluation of AIDM indicated promising and improved results with the deep learning classifier (DL4JMLP) when compared with above-mentioned Incremental classifiers and the recent best known related work for detecting anomalous network traffic.

Keywords: Anomaly Detection; Deep Learning; Incremental Classifiers; Intrusion Detection System; Machine Learning

1 Introduction

Intrusion Detection System (IDS) is one of the most popular technology for securing dynamic network environments which used for monitoring, detecting and responding to any potential presence of abnormal activities by both internal and external intruders. IDS composed of methods and techniques that automate the process of detecting attacks occurring over the network, alert the system administrators to any abnormal events and provide a report for any loss of confidentiality [34].

An IDS is categorized according to detection of attacks into two techniques; anomaly and signature-based detection. The signature (misuse) detection technique [32] used to detect attacks by comparing current activities to a list of observed signatures. It is more successful in detecting known attacks but it produces high false negative alarm with novel attacks. Anomaly detection technique [14] depends on building a dynamic normal behavior model for a user or a network by using machine learning methods wherefore this technique is effective for detecting unknown attacks where any deviation from normal behavior model indicates as an intrusion. The Host-based and Network-based techniques are two primary types of detection sources for IDS [22]. Host-Based Intrusion Detection System (HBIDS) analyzes a certain user profile or system events to detect attacks whereas Network-Based Intrusion Detection System (NBIDS) analyzes the flow of packets on the network.

Intelligent intrusion detection systems based on machine learning classifiers must be fast in monitoring and analyzing such evolved streams in real time. Streaming data generated from computer networks are changing continuously over time and infinite in incoming with a high speed. These characteristics need a number of challenges associated with mining data streams [20] thus these systems should correctly detect unlabeled instances by training them with some labeled samples from each category in dataset to build a learnable model able to classify incrementally unseen samples without any need to previous data and should update itself to accommodate new data arrive over time.

Incremental learning (IL) [37] is a method of machine learning and a classification task which used to build a learnable model for effective classification during training from the dynamic network. Incremental learning technique processes the new instances without any need to train the model where the model updated automatically after receiving each new instance. Incremental learning plays a vital role in anomaly detection system where normal profiles are updated automatically according to any changes occur without a need to train the model using previously instances. The structure of incremental learning classifiers has the ability to detect novel intrusions and handle concept drift in a dynamic network that is changed over time.

Latterly, deep learning approach has achieved revolutionize in the field of artificial intelligence (AI) and become a step toward building independent systems with a higher-level [35].

Deep learning approach has the ability to learn valuable features from a huge amount of unlabelled data stream by emulating the used methods to process data and create patterns help in decision making. We are seeking by using this approach overcome the challenges of traditional intrusion detection systems which uses typical classifiers [6].

Some of the large-scale attacks can be considered acts of cyber war [28], where attacks against any telecommunications infrastructure in every government could cause damage and palpable effects *e.g.* in Turkey 2008, how cyber attacks able to hide the blast in an oil pipeline from the control room thus it was unable to respond immediately by hacking security cameras.

Despite that, IDSs become widespread as a security measure thus it is substantial to know the drawbacks of traditional systems [33] to enhance the performance of IDS. The most limitations of traditional IDS [23] are, they usually produce a large numbers of false reports thus many real attacks are ignored due to the unpredictable activities for both user profile and computer networks, need training examples for system events to describe normal behaviors patterns and inability to analysis the encrypted packets in the network to detect any malicious activities.

In this paper, we are focused on the second type of IDS namely, the anomaly detection which uses different machine learning algorithms for demarcating the region between normal and anomaly behaviors. The incremental method is broadly classified as a single shot and has the ability to evolve the structure of classifiers to process new instances and classes without having to be completely retrained.

To this end, we proposed an Adaptive Intrusion Detection Model (AIDM) based on based on machine-learning techniques [36] and feature extraction which competent to detect zero-time attacks and distinguish between normal and anomaly instances in the computer networks with keeps the knowledge that had been learned previously from old examples.

The rest of paper is structured as follows: Section 2 highlights the related works done. Section 3 discusses the incremental and deep learning classifiers. Section 4 describes our Adaptive Intrusion Detection Model (AIDM) while Section 5 demonstrates the implementation of AIDM with a description of the datasets used and discusses the experimental results. Section 6 presents a comparative study with recent related works. Section 7 summarizes the proposed model.

2 Related Work

Many research papers are published regarding the incremental classifiers for detecting the network attacks in a huge amount of data streams because of they able to build a learnable model for new and unknown attacks [10]. Recently, deep learning approach has become a renowned technique in machine learning due to the developments are done in the field of artificial intelligence. Therefore, most of the researchers utilize the deep learning in the field of intrusion detection systems to improve the performance of the detection and overcome the shortages in the traditional systems where they focus on various offline algorithms that conducted on the most popular KDD-Cup 99 benchmark dataset to survey and evaluate the performance of IDS. KDD-Cup 99 dataset has a lot number of attributes which supportive for testing feature selection techniques. KDD-Cup 99 dataset is prepared and managed by MIT, Lincoln Labs by DARPA Intrusion Detection Evaluation Program. Presently, literature works use some public datasets such as NSL-KDD, Kyoto 2006+, and UNSW-NB15 to evaluate IDS.

Popoola *et al.* [24] proposed an effective feature selection technique for NID using discredited differential evolution (DDE). The results of the proposed technique presented that the model is able to identify 16 features capable of classifying the connections in the training and testing NSL-KDD dataset where achieved 99.92% classification accuracy for the training set and 88.73% accuracy for new attacks moreover it helped in reducing the training and testing time by using (C4.5) classifier.

Amrita *et al.* [25] proposed a hybrid feature selection approach - Heterogeneous Ensemble of Intelligent Classifiers (HYFSA-HEIC) to maximize the accuracy with low false alarm for intelligent lightweight network intrusion detection system which classifies input traffic into intrusion or normal. The ensemble method was built using 6 features and has been evaluated using both datasets KDD-Cup99 and Corrected Test where HYFSA-HEIC combined the hybrid feature selection approach (HyFSA) and a heterogeneous ensemble of intelligent classifiers (HEIC). The Heterogeneous ensemble employed five diverse accurate intelligent classifiers are, NB, NN (SGD), RIPPER, C4.5 and RF where their decisions were combined by utilizing majority voting of elementary combiner based on algebraic combination rule. The results showed that the proposed approach outperformed other methods with a reduction in the training and testing time by 50.79% and 55.30% respectively.

Boujnouni *et al.* [8] explored a new intrusion detection system (IDS) based on information gain criterion to estimate the quality of the attributes in the dataset. The proposed model works sequentially: Firstly preprocessing the intrusion dataset NSL-KDD to exclude varying resolution and ranges and then the novelty model takes a decision whether network traffic is an attack or normal based on SSPV-SVDD as classifier and SMO as a solver. The new IDS with the improved version of Support Vector Domain Description (SVDD) called SSPV-SVDD achieved 77.5% novelty detection rate.

Jayaswal *et al.* in [13] presented an effective and flexible NIDS based on deep learning approach by using self-taught learning (STL). The proposed model has two stages: In the first stage, a sparse auto-encoder used for the feature learning and in the second stage a soft-max regression used for classification NIDS. The evaluation of the model is conducted on the processed NSL-KDD dataset. The implementation used two types of classification 2-class and 5-class on the tested data. The model achieved 88.39% accuracy rate for the 2-class classification whereas SMR achieved 78.06%. For the 5-class classification, the f-measure values are 75.76% and 72.14% for STL and SMR respectively.

Belouch *et al.* in [4] introduced a proposed model based on two-stage and used a classification Reduced Error Pruning Tree (RepTree) algorithm for network intrusion detection system. In the first stage, the incoming traffic is firstly classified according to its protocol TCP, UDP and other, to normal or attack then a preprocessing applied for each subset. In the second stage, in case the traffic was an attack, a pre-trained multiclass classifier identifies the type of attack to provide the best response. The proposed model is evaluated using two datasets are NSL-KDD and UNSW-NB15. Feature selection technique was performed for each protocol to reduce the size of datasets by select features. The proposed model achieved a quite same accuracy of 88.85% with the combination of NBTree + RandomTree classifier with an advantage in training and testing time performance for the NSL-KDD dataset. For UNSW-NB15 dataset the model achieved the best accuracy of 89.95% with Decision tree classifier.

Jabbar *et al.* [12] demonstrated that the novel ensemble classifier (RFAODE) for intrusion detection system which built using two well -known algorithms RF and AODE is efficient for classifying the traffic to normal or malicious. The model outperforms classifiers namely, na?ve Bayes, J48, and PART with 90.51% accuracy using Kyoto dataset.

Shone *et al.* [29] presented a novel deep learning technique for intrusion detection system where they proposed a non-symmetric deep auto-encoder (NDAE) for unsupervised feature learning. The model was implemented using tensor Flow and evaluated using two datasets KDD Cup-99 and NSL-KDD. For KDD Cup '99 dataset the evaluation using 5-class proved that the proposed model was able to offer an average accuracy of 97.85% and a 3.8% improvement on its own accuracy for NSL-KDD dataset using 13-class.

Pattern recognition and increasing the time of learning became one of the main problems for most traditional intrusion detection systems that based on typical machine learning and signature-based approach. These traditional systems have limitations *e.g.* low detection rate for zero-day attacks, the rapid change in attacks behavior and arriving huge amount of data. For all of these issues, we proposed an Adaptive Intrusion Detection Model (AIDM) for network traffic classification to continue the work done from the previous related works. In addition, comparing the accuracy between seven classifiers has been done with a report on results.

3 Learning Classifiers

Different of effective IDS focus on incremental and deep learning approaches due to the ability to increase the model's performance and adapt to new samples without missing its existing knowledge. The intrusion network detection or the network flow classification can be formalized as shown: a set of instances in the dataset $D = \{S1, S2, \cdots, Sn\}$, where each instance has a labeled category class. For building a trained model, some labeled instances are used to learn the model how to link each instance to one category. For this purpose, we evaluated AIDM using seven different classification techniques able to handle data streams with high dimensional features for picking up the most effective classifier for building the proposed model.

3.1 NNGE (Non-Nested Generalized Exemplars)

NNGE is a nearest-neighbor learner [15] that expands generalized exemplars in memory depending on the distance between a set of exemplars using Euclidean distance function. The generalized exemplars are a group of samples explained as rules which reduce the classification time without any effect on the accuracy. NNGE able to classify sequence, multiclass data, and data with a different format. NNGE learns incrementally by classifying then generalizing a new example where an example is generalized and combined with the nearest exemplars having the same class. The generalization process of an exemplar is aborted if NNGE classifier checks that exemplar is similar to any exemplar in the area of feature space and may conflict with a new exemplar. Learning process of NNGE classifier builds a group of generalized exemplars where a hyper-rectangle covers a group of exemplars $\{H^1, H^2, \cdots, H^k\}$. In case of a hyper-rectangle covers only one exemplar, it is considered to be a non-generalized exemplar. Each example in training instances passing through three steps: the first step is the classification, where finding the hyper-rectangle H^k that is nearest to training instances E^j by calculating the distance $D(E, H)$ between an example and the hyper-rectangle H as in Equation (1):

$$D(E,H) = \sqrt{\sum_{i=1}^{n}(w_i \frac{(d(E_i, H_i))}{E^{max} - E^{min}})^2} \qquad (1)$$

Where $E^{max} - E^{min}$ are the ranges of values of attribute I over the training set, H_i is the interval $[H_i^{min}, H_i^{max}]$ which based on the attribute type. The second step is the model adjustment, where the hyper-rectangle H^k splits when covers conflicting example. This step applied when the hyper-rectangle covers an example of a different class. The third step is a generalization, where extend H^k to cover E^j where it extends the closest hyper-rectangle H^k that has the same class. This extension can be done only if the new hyper-rectangle doesn't overlap with other hyper-rectangles having a different class (See Algorithm 3.1).

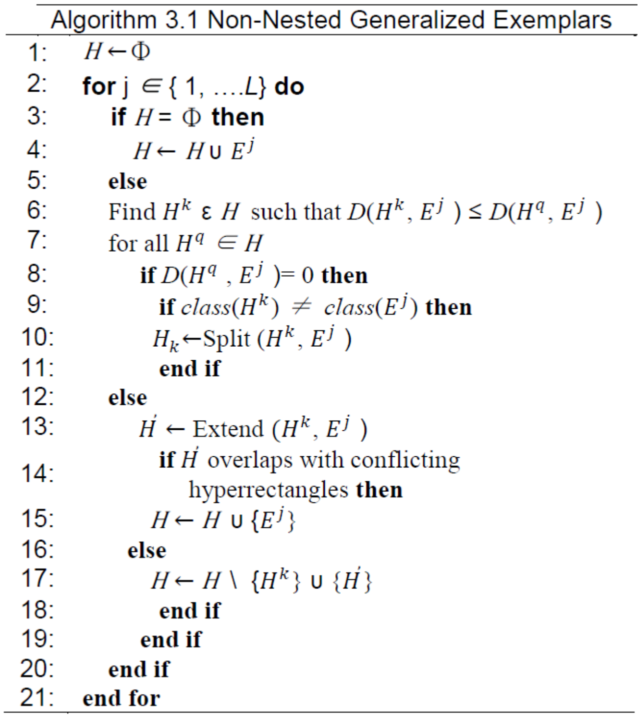
3.2 Naïve Bayes (NB)

NB [21] uses Bayesian classification method which useful in understanding a lot of learning classification algorithms. NB based on Bayes theorem considers one of the simplest probabilistic classifiers. This classifier has the ability to handle high streaming data, use data streaming as supervised classification and effective for various real applications due to its incremental nature. NB classifies new data streams by finding the highest posterior probability where it can predict whether X belongs to the class C_i.

The posterior probability $P(C_i \ x)$ of class C_i defined in Equation (2):

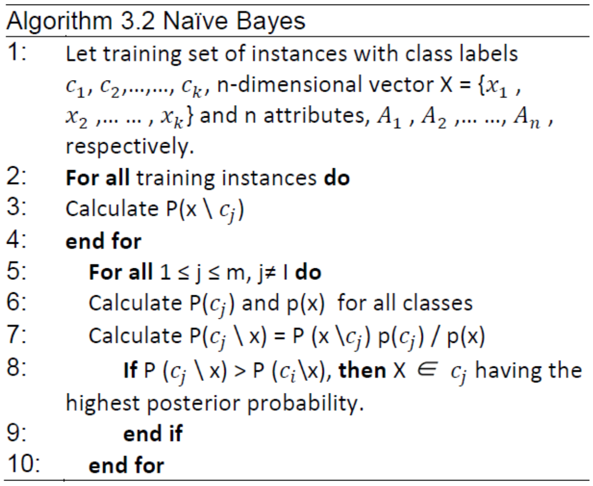
$$P(c_i \ x) = P(x \ C_i)p(C_i)/p(x) \qquad (2)$$

Where c is the predicted class and x is the instance. NB classifier is an independence assumption where the probability of one feature is unrelated to affect the probability of the other. It has two phases namely, a training phase and testing phase. In training phase, NB classifier calculates the probabilities of each attribute and stores this probability then, it calculates all time taken for the probabilities for each attribute. In the testing phase, the time taken to calculate the probabilities for each attribute is proportional to a number of attributes (See Algorithm 3.2).



Algorithm 3.1 Non-Nested Generalized Exemplars

```
$H \leftarrow \Phi$
for j $\in$ { 1, ….$L$} do
  if $H = \Phi$ then
    $H \leftarrow H \cup E^j$
  else
  Find $H^k$ ε $H$ such that $D(H^k, E^j) \leq D(H^q, E^j)$
  for all $H^q \in H$
    if $D(H^q, E^j) = 0$ then
      if $class(H^k) \neq class(E^j)$ then
      $H_k \leftarrow$ Split $(H^k, E^j)$
      end if
  else
    $H' \leftarrow$ Extend $(H^k, E^j)$
      if $H'$ overlaps with conflicting
        hyperrectangles then
      $H \leftarrow H \cup \{E^j\}$
    else
      $H \leftarrow H \setminus \{H^k\} \cup \{H'\}$
      end if
    end if
  end if
end for
```



Algorithm 3.2 Naïve Bayes

```
Let training set of instances with class labels
$c_1, c_2, \ldots, \ldots, c_k$, n-dimensional vector X = {$x_1$ ,
$x_2$ ,… … , $x_k$} and n attributes, $A_1$ , $A_2$ ,… …, $A_n$ ,
respectively.
For all training instances do
Calculate P(x \ $c_j$)
end for
  For all 1 ≤ j ≤ m, j≠ I do
  Calculate P($c_j$) and p(x)  for all classes
  Calculate P($c_j$ \ x) = P (x \$c_j$) p($c_j$) / p(x)
    If P ($c_j$ \ x) > P ($c_i$\x), then X $\in$ $c_j$ having the
highest posterior probability.
    end if
  end for
```

3.3 HT (Hoeffding Tree)

HT is a decision tree method [19] used for classifying data stream and this method considers the most efficient classification technique in data mining. Most existing classification techniques are very sensitive to stream ordering and take time-consuming with a huge amount of data streams. HT is one of the most suitable classifiers for data streaming which it gives results the same extent to batch classifiers and the learning process has constant time per instance. HT uses Hoeffding bound metrics which solved one of the most critical problems in the process of choosing a node which determined how many examples in a subset of the training dataset require for each node.

If r is an actual random variable (r is the information gain of an attribute) with range R and n is the observation of this variable Hoeffding Bound states that with probability $1-\delta$, the actual mean r is at least $\bar{r}-\epsilon$, where δ defined by a user and ϵ can be expressed as in Equation (3):

$$\epsilon = \sqrt{\frac{R^2 \ln(1/\delta)}{2n_l}} \quad (3)$$

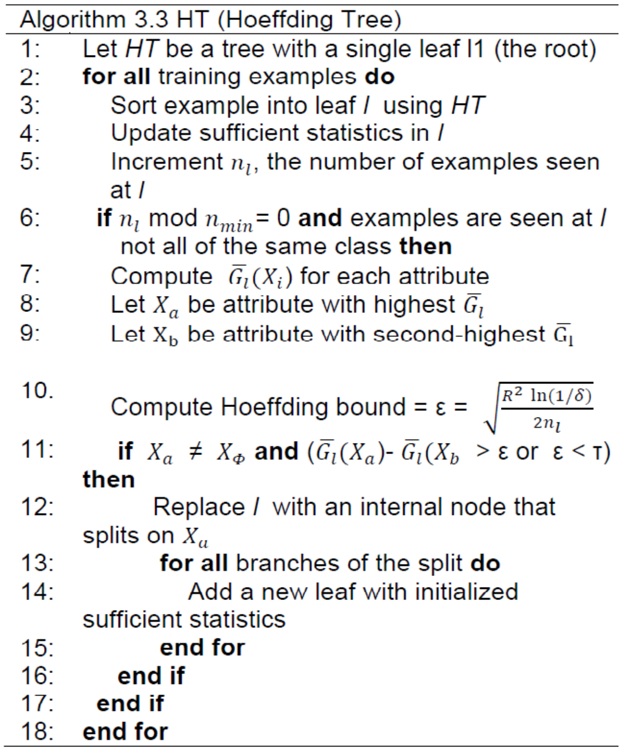
Let $G(X_i)$ be the heuristic measure used to choose test attributes, X_a be the attribute with highest observed G after seeing n examples and X_b the second-best attribute. Let the difference between two observed heuristic values $\Delta G = \bar{G}(X_a) - \bar{G}(X_b) \geq 0$ with a given desired δ. Hoeffding bound states that X_a is the best choice with probability $(1-\delta)$ for the number of examples (n) that have been seen at the node. In another expression we can say that X_a is the best attribute with probability $1-\delta$, if the observed difference of information gain (heuristic measure) $\Delta\bar{G} > \epsilon$ (See Algorithm 3.3).

3.4 Instance-based k-nearest neighbor Ibk (KNN)

IBK [26] is a k-nearest neighbor learner and kind of a simple instance-based learning algorithm It classifies new instance-based by determining the K-nearest neighbors on Euclidean distance metric which defined in Equation (4):

$$d(x_i, x_j) = \sqrt{\sum_{r-1}^{r=n} [a_r(x_i) - a_r(x_j)]^2} \quad (4)$$

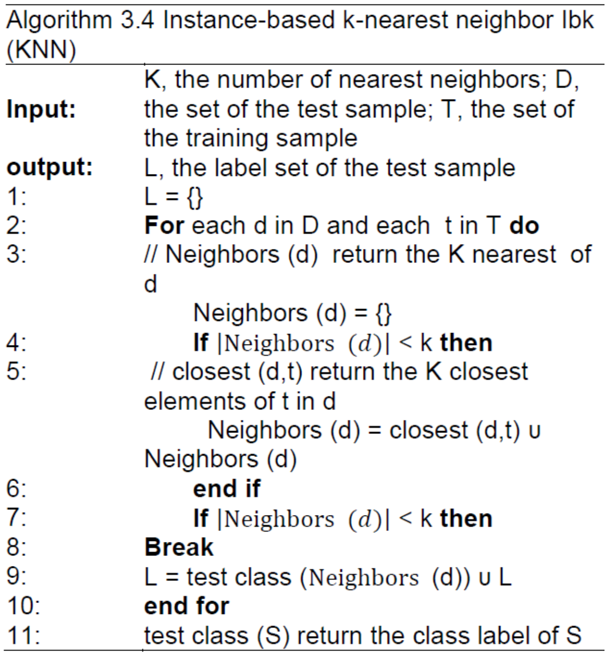
Where $d(x_i, x_j)$ is the distance between the two instances x_i and x_j, $a_r(x)$ is the attribute value of instance x. The nearest neighbors used for classifying new instance and the relation between the time taken in classify new instance is linearly increased with the number of training instances saved in the classifier thus it called Lazy learning classifier. The new classification for new instance can be found easily by comparing the new instance with the stored trained dataset. IBK classifier has limitation *e.g.* computational complexities for finding K-nearest neighbor instances and the processing time for no longer optimal with a huge number of instances because we need to calculate all similarities between all training instances. We can solve this problem by reducing the dimensions of the features or using small datasets and dependency on the training set where building KNN (See Algorithm 3.4).



Algorithm 3.3 HT (Hoeffding Tree)

```
Let HT be a tree with a single leaf l1 (the root)
for all training examples do
    Sort example into leaf l using HT
    Update sufficient statistics in l
    Increment $n_l$, the number of examples seen
    at l
  if $n_l$ mod $n_{min}$= 0 and examples are seen at l
    not all of the same class then
    Compute $\bar{G}_l(X_i)$ for each attribute
    Let $X_a$ be attribute with highest $\bar{G}_l$
    Let $X_b$ be attribute with second-highest $\bar{G}_l$

    Compute Hoeffding bound = ε = $\sqrt{\frac{R^2 \ln(1/\delta)}{2n_l}}$
    if $X_a \neq X_\Phi$ and ($\bar{G}_l(X_a)$- $\bar{G}_l(X_b$ > ε or ε < τ)
then
        Replace l with an internal node that
splits on $X_a$
        for all branches of the split do
            Add a new leaf with initialized
sufficient statistics
        end for
     end if
   end if
end for
```



Algorithm 3.4 Instance-based k-nearest neighbor Ibk (KNN)

```
Input:    K, the number of nearest neighbors; D,
          the set of the test sample; T, the set of
          the training sample
output:   L, the label set of the test sample
          L = {}
          For each d in D and each t in T do
          // Neighbors (d) return the K nearest of
          d
              Neighbors (d) = {}
              If |Neighbors (d)| < k then
           // closest (d,t) return the K closest
          elements of t in d
                Neighbors (d) = closest (d,t) ∪
          Neighbors (d)
              end if
              If |Neighbors (d)| < k then
          Break
          L = test class (Neighbors (d)) ∪ L
          end for
          test class (S) return the class label of S
```

3.5 Radial Basis Function Neural Network (RBFNN)

RBF learner [18] is a function based classifier and the feed-forward artificial neural network. RBF is an alternative to a public MLP classifier but it is superior to MLP in the fast training process and it simple in its structure.

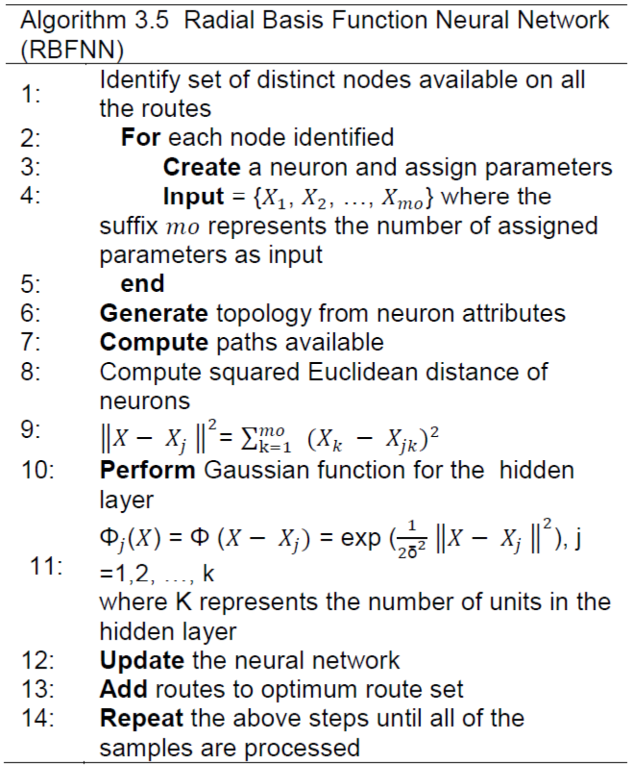
RBF uses Gaussian transfer functions instead of sigmoidal functions which used by MLP learner. RBFNN classifies instances by measuring the distances between centers of hidden and inputs neurons and it consists of three layers: an input layer, which feeds the feature into the network. The number of input neuron is determined by the number of input vector; hidden layer calculates the outcome of the basis functions and used the normalized Gaussian function as the radial basis function which expressed as in Equation (5):

$$\phi_j(X) = \phi(X - X_j) = \exp(\frac{1}{2\delta^2}|X - X_j|^2) \qquad (5)$$

Where δ is the width of the Gaussian and K represents the number of units in the hidden layer, X_j is denoted the center of the ith node; and the output layer, which calculates the linear combination of basis functions as in Equation (6):

$$Y = f_j(x) = [\sum_{i=1}^{k} w_i\phi_i(x)] \qquad (6)$$

Where $f(x)$ is the final output, $\phi_j(X)$ denotes the radial function of the jth hidden node and w_i denotes the hidden-to-output weight corresponding to the ith hidden node (See Algorithm 3.5).

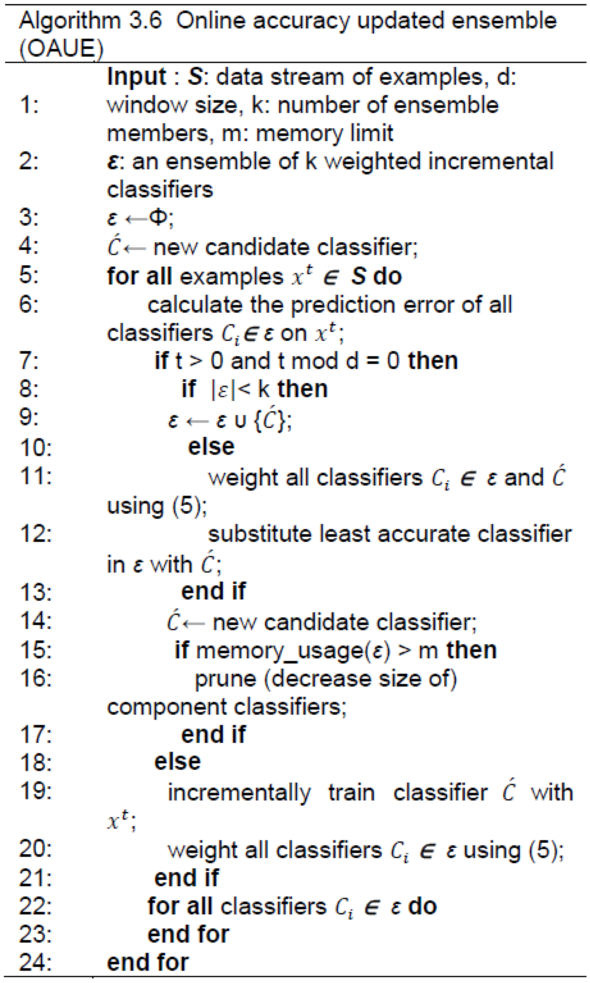


Algorithm 3.5 Radial Basis Function Neural Network (RBFNN)

```
Identify set of distinct nodes available on all the routes
  For each node identified
      Create a neuron and assign parameters
      Input = {$X_1, X_2, \ldots, X_{mo}$} where the suffix $mo$ represents the number of assigned parameters as input
  end
Generate topology from neuron attributes
Compute paths available
Compute squared Euclidean distance of neurons
$\|X - X_j\|^2 = \sum_{k=1}^{mo} (X_k - X_{jk})^2$
Perform Gaussian function for the hidden layer
$\Phi_j(X) = \Phi(X - X_j) = \exp(\frac{1}{2\delta^2}\|X - X_j\|^2)$, j =1,2, …, k
where K represents the number of units in the hidden layer
Update the neural network
Add routes to optimum route set
Repeat the above steps until all of the samples are processed
```

3.6 Online Accuracy Updated Ensemble (OAUE)

Online accuracy updated ensemble is an ensemble algorithm [30] where it is an improved version of accuracy weighted ensemble. OAUE classifier has a weighted pool of classifiers to predict the class of incoming sample where it predicts the new instances by aggregating all predictions using a rule called weighted voting. The new classifier is added to the ensemble after a segment of instances is classified and it replaced by the weakest ensemble member in performance. OAUE is updated incrementally when the performance of each classifier is evaluated using expected prediction error for the most recent instances. The classifier with the poorest performance is replaced from the pool of classifiers and then the weights adjusted according to accuracy (See Algorithm 3.6).



Algorithm 3.6 Online accuracy updated ensemble (OAUE)

```
Input : S: data stream of examples, d: window size, k: number of ensemble members, m: memory limit
ε: an ensemble of k weighted incremental classifiers
ε ←Φ;
Ć← new candidate classifier;
for all examples $x^t$ ∈ S do
    calculate the prediction error of all classifiers $C_i$ ∈ ε on $x^t$;
      if t > 0 and t mod d = 0 then
        if |ε|< k then
          ε ← ε u {Ć};
          else
            weight all classifiers $C_i$ ∈ ε and Ć using (5);
            substitute least accurate classifier in ε with Ć;
        end if
          Ć← new candidate classifier;
          if memory_usage(ε) > m then
            prune (decrease size of) component classifiers;
        end if
      else
          incrementally train classifier Ć with $x^t$;
          weight all classifiers $C_i$ ∈ ε using (5);
      end if
    for all classifiers $C_i$ ∈ ε do
    end for
end for
```

3.7 Deep Learning 4 Java Multilayer Perceptron (DL4JMLP)

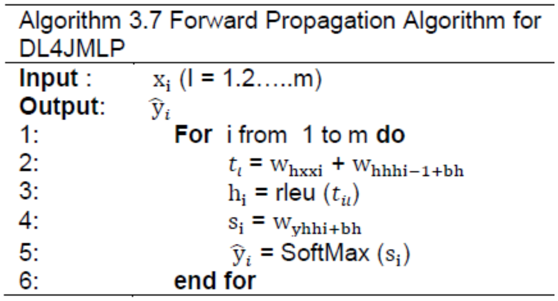
Deeplearning4J (DL4J) is a Java-based toolkit for configuring deep neural networks which consist of multiple layers. These layers are organized by using hyper-parameters which utilize to learn the neural networks. DL4J treats the tasks of loading data and training the algorithms as separate processes and works with a lot of different data types such as images, CSV, ARFF, plain text and Apache Camel Integration. DL4J uses for retrieving the data in a format suited for training a neural net model which uses multiple passes and each pass is called an epoch. DL4J gives data to the scientists and tools developers to build a deep neural network on a high level by using concepts like layer. DL4J employs a builder pattern to build the neural net declaratively. The neural networks are typically trained by using batches of the training data where the updates to the weights and biases of the neural network effect on the outputs of each node of the network. Deep Learning 4 Java Multilayer Perceptron (DL4Jmlp) is a straightforward wrapper which consists of three layers: the Input layer, the hidden layer which uses ReLU (Rectified Linear Unit) activation function and it defined as in Equation (7):

$$F(x) = \max(x, 0). \tag{7}$$

Where $f(x)$ has output 0 if the input is less than 0 and the output is raw if the input is greater than 0. The output layer uses the softmax activation function which squashes the outputs of each unit between 0 and 1and it represented in Equation (8):

$$\sigma(x)_j = \frac{e^{x_j}}{\sum_{k=1}^{k} e^{x_k}} \tag{8}$$

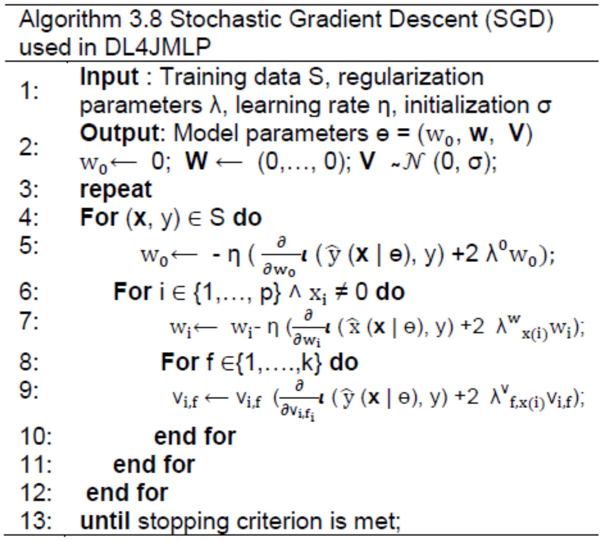
Where x is a vector of the inputs to the output layer, j indexes the output units, and $j = 1, 2, \cdots, K$ (See Algorithm 3.7).



Algorithm 3.7 Forward Propagation Algorithm for DL4JMLP

```
Input :      $x_i$ (l = 1.2.....m)
Output:      $\widehat{y}_i$
    For i from 1 to m do
        $t_\iota = w_{hxxi} + w_{hhhi-1+bh}$
        $h_i$ = rleu ($t_{i\iota}$)
        $s_i = w_{yhhi+bh}$
        $\widehat{y}_i$ = SoftMax ($s_i$)
    end for
```

DL4Jmlp uses the most popular Stochastic Gradient Descent (SGD), as an optimization to the neural network [3]. Which known as incremental gradient descent, is a stochastic approximation of the gradient descent optimization and iterative method for minimizing an objective function that is written as a sum of differentiable functions (See Algorithm 3.8).



Algorithm 3.8 Stochastic Gradient Descent (SGD) used in DL4JMLP

```
Input : Training data S, regularization
parameters λ, learning rate η, initialization σ
Output: Model parameters ө = ($w_0$, w, V)
$w_0 \leftarrow$ 0; W ← (0,..., 0); V ~$\mathcal{N}$ (0, σ);
repeat
For (x, y) ∈ S do
        $w_0 \leftarrow - \eta ( \frac{\partial}{\partial w_0} \iota (\widehat{y} (\mathbf{x} \mid ө), y) + 2 \lambda^0 w_0)$;
    For i ∈ {1,..., p} ∧ $x_i \neq 0$ do
            $w_i \leftarrow w_i - \eta ( \frac{\partial}{\partial w_i} \iota (\widehat{x} (\mathbf{x} \mid ө), y) + 2 \lambda^w{}_{x(i)} w_i)$;
        For f ∈{1,....,k} do
            $v_{i,f} \leftarrow v_{i,f} ( \frac{\partial}{\partial v_{i,f_i}} \iota (\widehat{y} (\mathbf{x} \mid ө), y) + 2 \lambda^v{}_{f,x(i)} v_{i,f})$;
        end for
    end for
 end for
until stopping criterion is met;
```

4 The Proposed Model

The main strategy of the proposed AIDM, Figure 1, is to utilize the study of the incremental and deep learning methodologies for classifying the stream of the computer networks. This methodology depends on selecting the best machine learning classifier which optimizes the efficiency of anomalous detection and we compared the model with the traditional intrusion detection systems. AIDM examines all events circulating in the network by observing abnormal activities and it consists of two phases: the off-line and the on-line phases. In the off-line (training dataset) phase, the model is trained by machine learning classifiers to be more familiarize and learnable for activities exist in the network flow where labeled processed training records are fed into AIDM to construct learnable model able to be tested.

In online (testing dataset) phase, new unlabelled and labeled samples are fed into the model to predict each unlabelled samples based on the extracted trained model from the off-line phase. AIDM encompass three stages as shown in Figure 1 are, dataset collection stage, preprocessing engine stage and classification stage. In the dataset collection stage, datasets samples are collected with known labels are fed into off-line phase whereas a mixed collection of known and unknown labels are fed into the on-line phase. AIDM evaluated using three datasets are, KDD Cup 99 and UNSW-NB15 and real Kyoto 2006+. In the preprocessing engine stage, offline and online network datasets are processed using a pre-processing engine where transformation, feature extraction, and normalization techniques applied for the datasets in order to correctly identify records over which the attacks resemble. In classification stage, AIDM trained using different classifiers on labeled datasets to build a trained model able to use for new datasets. For choosing the best results for AIDM a comparison between the classifiers results is carried out.

4.1 Pre-processing Engine Stage

Data pre-processing [9] considers the first and important step for both the online and the offline phases which used to analyze and transform the datasets to a suitable form to go through learning phase and data analysis. The flow of different network traffic sources is processed by using a sub model in the pre-processing engine by using a Detector for Number of Features in Dataset (DNFD) where a proper feature extraction and transformation (coding) techniques are applied based on the number of features in each dataset. The following section provides a concise definition of the used three techniques of the preprocessing engine.

4.1.1 Data Transformation

Some machine learning algorithms handle numeric features to guarantee the best performance thus we converted the symbolic and text values to numerical values

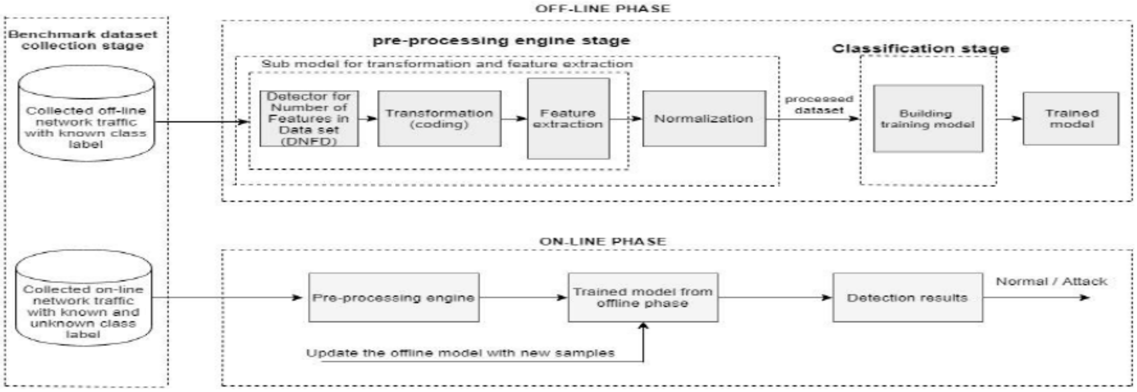
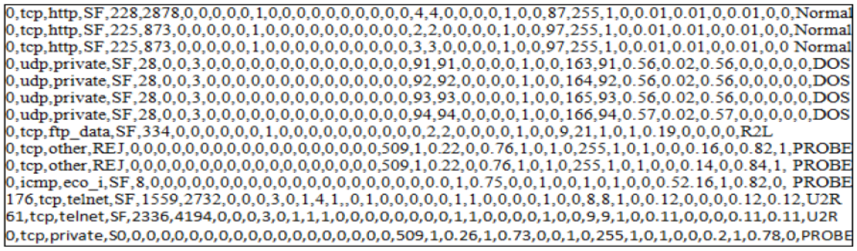

OFF-LINE PHASE
Benchmark dataset collection stage
Collected off-line network traffic with known class label
pre-processing engine stage
Sub model for transformation and feature extraction
Detector for Number of Features in Data set (DNFD)
Transformation (coding)
Feature extraction
Normalization
processed dataset
Classification stage
Building training model
Trained model
ON-LINE PHASE
Collected on-line network traffic with known and unknown class label
Pre-processing engine
Trained model from offline phase
Detection results
Normal / Attack
Update the offline model with new samples

Figure 1: The structure of an Adaptive Intrusion Detection Model (AIDM)



```
0,tcp,http,SF,228,2878,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,4,4,0,0,0,0,1,0,0,87,255,1,0,0.01,0.01,0,0.01,0,0,Normal
0,tcp,http,SF,225,873,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,2,2,0,0,0,0,1,0,0,97,255,1,0,0.01,0.01,0,0.01,0,0,Normal
0,tcp,http,SF,225,873,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,3,3,0,0,0,0,1,0,0,97,255,1,0,0.01,0.01,0,0.01,0,0 Normal
0,udp,private,SF,28,0,0,3,0,0,0,0,0,0,0,0,0,0,0,0,0,0,91,91,0,0,0,0,1,0,0,163,91,0.56,0.02,0.56,0,0,0,0,0,DOS
0,udp,private,SF,28,0,0,3,0,0,0,0,0,0,0,0,0,0,0,0,0,0,92,92,0,0,0,0,1,0,0,164,92,0.56,0.02,0.56,0,0,0,0,0,DOS
0,udp,private,SF,28,0,0,3,0,0,0,0,0,0,0,0,0,0,0,0,0,0,93,93,0,0,0,0,1,0,0,165,93,0.56,0.02,0.56,0,0,0,0,0,DOS
0,udp,private,SF,28,0,0,3,0,0,0,0,0,0,0,0,0,0,0,0,0,0,94,94,0,0,0,0,1,0,0,166,94,0.57,0.02,0.57,0,0,0,0,0,DOS
0,tcp,ftp_data,SF,334,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,2,2,0,0,0,0,1,0,0,9,21,1,0,1,0.19,0,0,0,0,R2L
0,tcp,other,REJ,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,509,1,0.22,0,0.76,1,0,1,0,255,1,0,1,0,0,0.16,0,0.82,1,PROBE
0,tcp,other,REJ,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,509,1,0.22,0,0.76,1,0,1,0,255,1,0,1,0,0,0.14,0,0.84,1, PROBE
0,icmp,eco_i,SF,8,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.0,1,0.75,0,0,1,0,0,1,0,1,0,0,0.52.16,1,0.82,0, PROBE
176,tcp,telnet,SF,1559,2732,0,0,0,3,0,1,4,1,,0,1,0,0,0,0,0,1,1,0,0,0,0,1,0,0,8,8,1,0,0.12,0,0,0,0.12,0.12,U2R
61,tcp,telnet,SF,2336,4194,0,0,0,3,0,1,1,1,0,0,0,0,0,0,0,1,1,0,0,0,0,1,0,0,9,9,1,0,0.11,0,0,0,0.11,0.11,U2R
0,tcp,private,S0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,509,1,0.26,1,0.73,0,0,1,0,255,1,0,1,0,0,0.2,1,0.78,0,PROBE
```

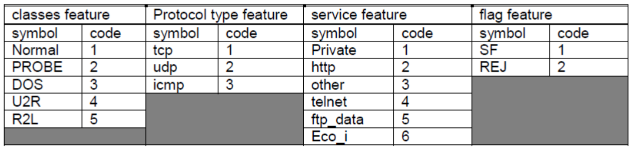
Figure 2: Random records taken from the KDD-CUP99 dataset

where we gave a number to each symbol repeated in the feature, for example, 1 gave to the feature that has a greater number of repeatedly then, less repeatedly takes 2, *etc.* Table 1 shows an example for symbols converted to numerical values based on the previous random set of the KDD-CUP99 dataset in Figure 2. KDD CUP 99, UNSW-NB15 and Kyoto 2006+ datasets have symbolic values; where KDD CUP 99 dataset contains three symbols features are, Protocol type, Service, and Flag features; UNSW-NB15 dataset has three symbols features are protocol, service and state features; Kyoto 2006+ dataset has two symbols features are service and flag. The transformation (coding) technique depends on the DFND used in the sub model which distinguished between the numbers of features in the datasets and the coding table is applied according to the number of features for each dataset.

4.1.2 Feature Extraction

Feature extraction technique [2] is an amelioration process for detecting the abnormal events by selecting eclecticism set of features in the datasets, which are represented using different features and then, removing the features with less influence for achieving a high performance for IDSs. In data analysis for network flow classification we don't need all these features wherefore we need to pick up the most informative features which optimize the performance of machine learning algorithms. In AIDM we have been performed feature extraction technique [11] for each dataset by using information gain (IG) criteria according to the number of detected features by DNFD.

Table 1: Coding of KDD-CUP99 dataset features



classes feature		Protocol type feature		service feature		flag feature	
symbol	code	symbol	code	symbol	code	symbol	code
Normal	1	tcp	1	Private	1	SF	1
PROBE	2	udp	2	http	2	REJ	2
DOS	3	icmp	3	other	3		
U2R	4			telnet	4		
R2L	5			ftp_data	5		
				Eco_i	6		

Figure 3 shows a simple flow chart for the sub model in the preprocessing engine where the collected datasets pass through this sub model where DNFD detect the number of features in the datasets and according to the number of features in the datasets, the implementation of transformation (coding) and feature extraction techniques in a suitable manner for each dataset. Information gain (IG) criteria [1] considers one of the simplest feature selection methods used to rank all features in the datasets based on the entropy which measures the reduction in purity in an arbitrary collection of instances.

The entropy H(X) of variable Y can define as in Equation (9):

$$H(Y) = -\sum_{i} P(y_i)\log_2(p(y_i)) \qquad (9)$$

Where $P(y_i)$ is the prior probability for variable Y. The entropy of variable X after observing value of another

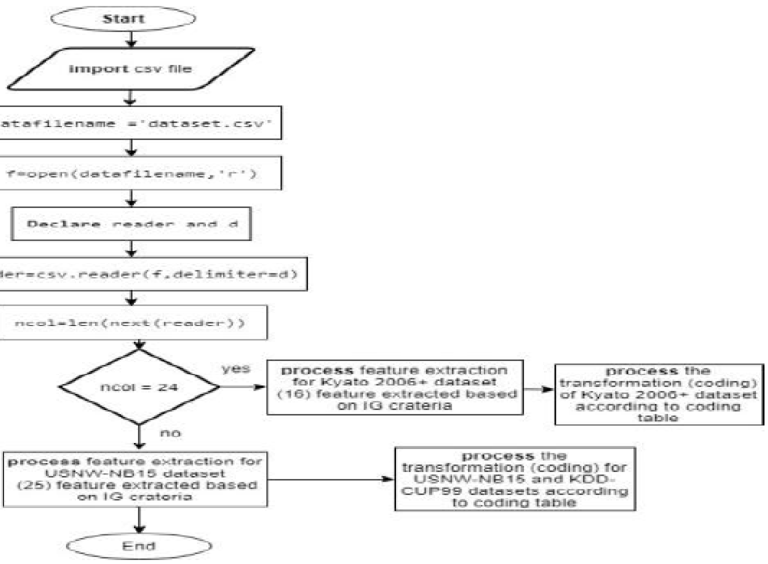

Start
import csv file
datafilename ='dataset.csv'
f=open(datafilename,'r')
Declare reader and d
reader=csv.reader(f,delimiter=d)
ncol=len(next(reader))
ncol = 24
yes
process feature extraction for Kyato 2006+ dataset (16) feature extracted based on IG crateria
process the transformation (coding) of Kyato 2006+ dataset according to coding table
no
process feature extraction for USNW-NB15 dataset (25) feature extracted based on IG crateria
process the transformation (coding) for USNW-NB15 and KDD-CUP99 datasets according to coding table
End

Figure 3: Flowchart of the sub model

variable Y can define as in Equation (10):

$$H(Y\backslash X) = -\sum_i P(x_i) \sum_i P(y_i \backslash x_i) \log_2(p(y_i \backslash x_i)) \quad (10)$$

Where $P(y_i \backslash x_i)$ is the posterior probability of X given Y. The information gain (feature ranking) about X provided by Y to pick up the most important features can be defined as in Equation (11):

$$IG(X/Y) = H(Y) - H(Y\backslash X). \quad (11)$$

We have been extracted 25 features from KDD-CUP99 and USNW-NB15 datasets and 18 features from Kyoto 2006+ dataset. Table 2 shows the optimal features that selected from the three datasets for AIDM.

Table 2: Selected features for the three datasets by IG

datasets	Selected features by using IG criteria
KDD-CUP99	F5, f3, f6, f33, f23, f37, f35, f40, f12,f34,f4,f41,f30,f24,f29,f38,f25, f36,f10,f27,f28,f32,f39,f31 and f13
USNW-NB15	F1,f8,f28,f13,f9,f29,f33,f12, f11,f10,f14,f18,f7,f2,f17,f26, f25,f27,f3,f6,f16,f20,f36,f5 and f19
Kyoto 2006+	F1,f2,f3,f4,f5,f6,f7,f8,f9,f10,f11,f12,f13, f14, f18,f20,,f22 and f23

The intrusion detection accuracy for the selected features by IG criteria [31] outperforms the all features for the three datasets as shown in Figure 4 which increased the effectiveness of all classifiers.

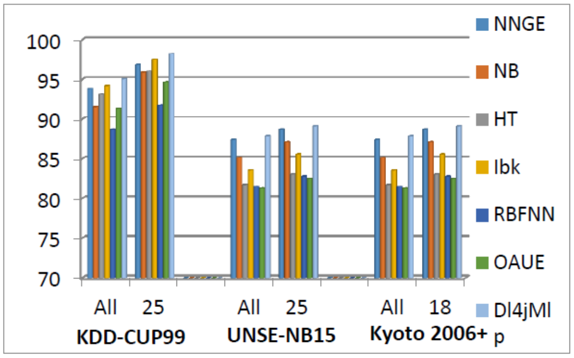


Figure 4: The accuracy of three datasets for all features and selected features

4.2 Data Normalization

The normalization technique [17] is the process of scaling each value of the attributes in the dataset into the specified new range [0, 1] or [-1, 1] which makes dataset acquires a particular property helpful in prediction purpose. Min-max normalization method converts v_i in the dataset to new value x_i using Equation (12):

$$x_i = \frac{v_i - \min(v_i)}{\max(v_i) - \min(v_i)} \quad (12)$$

Where v_i is the existing value of the attribute. The maximum and minimum values are taken over all values of the attribute normally, x_i is set to zero if the maximum is equal to the minimum.

4.3 Data Classification

The classification stage used two experiments to illustrate the importance of the feature extraction technique. The first experiment, AIDM classified the datasets without apply feature extraction technique and the results are recorded of each classifier respectively. The second experiment, AIDM classified the datasets with applying the feature extraction technique on the datasets where the feature extraction has been applied many times to reach the maximum accuracy for each dataset. we reached to the best number of features for each dataset as shown in Table 2. Figure 4 shows a comparison between the three datasets before and after applying the feature extraction technique which proved the importance of this technique in increasing the detection accuracy.

5 AIDM Implementation

AIDM implemented using: Java programming language with aid of an open source framework Massive Online Analysis (MOA), which used for data stream mining and big data processing, and the open source WEKA tool, which have a collection of machine learning algorithms for data mining and preprocessing tasks [16]. For training the proposed model we used 80,000 of processed datasets having known attacks and 10,000 of processed datasets having unknown and known attacks for testing the model.

The performance evaluation has been taken from the final prediction of the datasets classes. The experiments were performed on a laptop having configuration Intel© Core^{TM} i3 CPU M 380 2.53 GHz, 2.53GHz, 4.00GB of RAM and the operating system is windows 7 professional.

5.1 Description of KDD CUP 99, UNSW-NB15 and Kyoto 2006+ Datasets

The evaluation of AIDM conducted on three datasets namely, KDD CUP 99, USNW-NB15 and Kyoto 2006+. The following section presents a brief description of the datasets.

5.1.1 KDD-CUP99 Dataset

KDD-cup 99 benchmark dataset [5] was built based on DARPA'98 evaluation program which is a compressed TCP dump data of network traffic during 7 weeks. Each instance in KDD-cup 99 have 41 features with a labeled class indicates one of five values are, Normal, DOS (denial of service), PROBE (surveillance), U2R (user to root) and R2l (root to local). The KDD-CUP99 dataset may not be a good choice of the neoteric computer networks due to the scarcity in computer networking datasets but it is one of the most criterion realistic benchmark datasets for training and testing intrusion detection system. The KDD-CUP99 dataset has been divided into training dataset with 4,898,431 records and 311,027 records for the testing dataset.

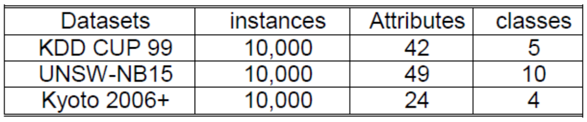
5.1.2 UNSW-NB15 Dataset

The UNSW-NB15 benchmark dataset [7] was developed based on an IXIA tool which used to produce nine modernistic attacks beside to, assortment broad of normal activities. The UNSW-NB15 dataset has been divided into training dataset with 82,332 records and 175, 341 records for the testing dataset. The UNSW-NB15 dataset has 45 features with a label contains nine attacks namely, Reconnaissance, Shellcode, Exploit, Fuzzers, Worm, DoS, Backdoor, Analysis and Generic beside, a normal class. UNSW-NB15 and KDD-CUP99 datasets have six common features are, protocol type, service, duration, source, and destination.

5.1.3 Kyoto 2006+ Dataset

Kyoto 2006 + dataset [27] have been collected at Kyoto University from honey spot systems. Kyoto 2006 + dataset is a real traffic traces devoid of any manual labels which each connection has 23 features and a label contains three values where, '1' value means that the session was normal, '-1' means that known attack was observed in the session and '-2' means that unknown attack was observed in the session. Kyoto 2006+ and KDD-CUP99 datasets have 14 common features namely, Duration, Service, Source bytes, Destination bytes, Count, Same srv rate, Serror rate, Srv serror rate, Dst host count, Dst host srv count, Dst host same src port rate, Dst host serror rate, Dst host srv serror rate and flag beside to, 10 new features namely, IDS detection, Malware detection, Ashula detection, Label, Source IP Address, Source Port Number, Destination IP Address, Destination Port Number, Start Time and Duration. A statistics description for the datasets based on the number of instances, attributes, and classes shown in Table 3.

Table 3: Statistics description of three benchmarks



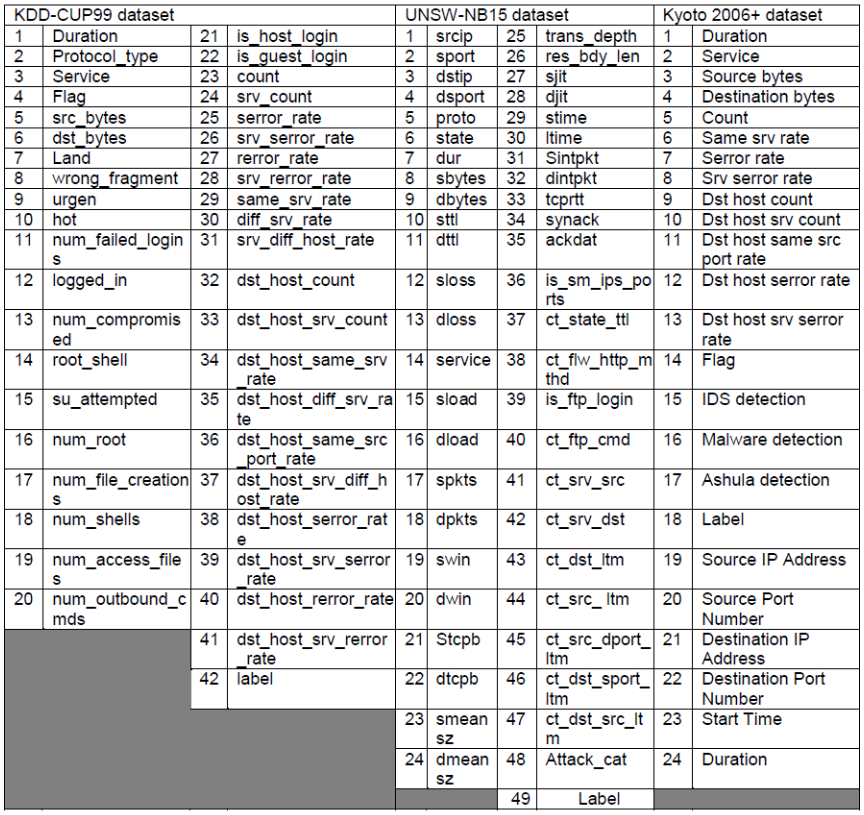
Datasets	instances	Attributes	classes
KDD CUP 99	10,000	42	5
UNSW-NB15	10,000	49	10
Kyoto 2006+	10,000	24	4

Table 4 represented the features for each dataset used in AIDM. We randomized selected the data of 27, 28, 29, 30 and 31 December 2015 where it has the latest updated data. The statistical description of classes used in the testing (on-line) phase for the three benchmark datasets represented in Table 5.

5.2 Performance Evaluation Metrics

AIDM evaluation has been done based on (1) The cumulative accuracy, which refers to the accuracy after classification of all chunks and it is the percentage of corrected instances classified to the total number of instances), (2) kappa statistics it measures the extent of convergence of

Table 4: The Features of the KDD-CUP99, UNSW-NB15 and Kyoto 2006+ datasets



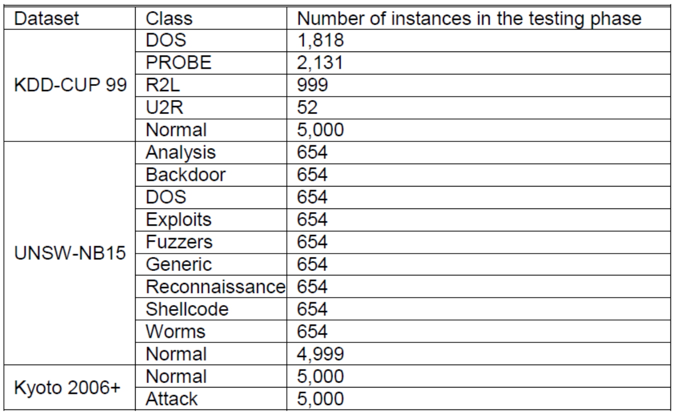
KDD-CUP99 dataset				UNSW-NB15 dataset				Kyoto 2006+ dataset	
1	Duration	21	is_host_login	1	srcip	25	trans_depth	1	Duration
2	Protocol_type	22	is_guest_login	2	sport	26	res_bdy_len	2	Service
3	Service	23	count	3	dstip	27	sjit	3	Source bytes
4	Flag	24	srv_count	4	dsport	28	djit	4	Destination bytes
5	src_bytes	25	serror_rate	5	proto	29	stime	5	Count
6	dst_bytes	26	srv_serror_rate	6	state	30	ltime	6	Same srv rate
7	Land	27	rerror_rate	7	dur	31	Sintpkt	7	Serror rate
8	wrong_fragment	28	srv_rerror_rate	8	sbytes	32	dintpkt	8	Srv serror rate
9	urgen	29	same_srv_rate	9	dbytes	33	tcprtt	9	Dst host count
10	hot	30	diff_srv_rate	10	sttl	34	synack	10	Dst host srv count
11	num_failed_logins	31	srv_diff_host_rate	11	dttl	35	ackdat	11	Dst host same src port rate
12	logged_in	32	dst_host_count	12	sloss	36	is_sm_ips_ports	12	Dst host serror rate
13	num_compromised	33	dst_host_srv_count	13	dloss	37	ct_state_ttl	13	Dst host srv serror rate
14	root_shell	34	dst_host_same_srv_rate	14	service	38	ct_flw_http_mthd	14	Flag
15	su_attempted	35	dst_host_diff_srv_rate	15	sload	39	is_ftp_login	15	IDS detection
16	num_root	36	dst_host_same_src_port_rate	16	dload	40	ct_ftp_cmd	16	Malware detection
17	num_file_creations	37	dst_host_srv_diff_host_rate	17	spkts	41	ct_srv_src	17	Ashula detection
18	num_shells	38	dst_host_serror_rate	18	dpkts	42	ct_srv_dst	18	Label
19	num_access_files	39	dst_host_srv_serror_rate	19	swin	43	ct_dst_ltm	19	Source IP Address
20	num_outbound_cmds	40	dst_host_rerror_rate	20	dwin	44	ct_src_ ltm	20	Source Port Number
		41	dst_host_srv_rerror_rate	21	Stcpb	45	ct_src_dport_ltm	21	Destination IP Address
		42	label	22	dtcpb	46	ct_dst_sport_ltm	22	Destination Port Number
				23	smeansz	47	ct_dst_src_ltm	23	Start Time
				24	dmeansz	48	Attack_cat	24	Duration
						49	Label		

prediction with correct class using Equation (13):

$$K = (Po - Pe)/(1 - Pe). \tag{13}$$

Where Po indicate the relative observed convergence, Pe indicates the hypothetical probability of chance convergence and (3) running time defined as the time taken for execution the classification of each classifier in the proposed model.

Table 5: Classes statistical distribution for 10,000 samples



Dataset	Class	Number of instances in the testing phase
KDD-CUP 99	DOS	1,818
	PROBE	2,131
	R2L	999
	U2R	52
	Normal	5,000
UNSW-NB15	Analysis	654
	Backdoor	654
	DOS	654
	Exploits	654
	Fuzzers	654
	Generic	654
	Reconnaissance	654
	Shellcode	654
	Worms	654
	Normal	4,999
Kyoto 2006+	Normal	5,000
	Attack	5,000

5.3 Experimental Results and Discussion

In this paper, seven classifiers have been used to evaluate the effectiveness and performance of the proposed AIDM to identify unknown diverse attacks in the network traffic and overcome the problems in traditional IDS.

AIDM evaluation is based on the incremental and deep learning techniques for multiclass classification for the datasets. To guarantee the best performance for AIDM we utilized six incremental classifiers and DL4JMLP deep learning classifier consecutively for training the model. The evaluation of AIDM has conducted using the three benchmark datasets compare the results among each trained model by each classifier and learning classifiers proposed by beforehand researchers has been carried out to determine the efficiency AIDM which helpful for testing the model in on-line phase.

Through our results in the proposed model, it is shown that AIDM achieved high accuracy with using Dl4jMlp deep learning classifier and its effectiveness is superior to that of traditional machine learning classifiers in differentiate attacks for multiclass classification.

Final accuracy of AIDM conducted on 10,000 samples of known and unknown attacks from the three datasets the performance was measured for every 1000 instances between samples. AIDM evaluated using a prequential evaluation (Interleaved Test-Then-Train) which used for the real-time evaluation.

Such evaluation technique it is considered an evaluation for streaming data where each sample tests first before using in training.

Table 6 summarizes the classification results of the seven selection classifiers in regard to accuracy (%), kappa statistics and running time.

It clearly shown from Table 6 that the proposed model based on deep learning approach using the Dl4jMlp Classifier and feature selection technique using IG criterion, has achieved the best accuracy of 98.2%, 92.5% and 82.12% for the three datasets, KDD-Cup 99, UNSW-NB15 and Kyoto 2006+ respectively than other incremental classifiers used by AIDM.

It is notable from Table 6 that the classification accuracies for KDD-CUP4 99 and real Kyoto 2006+ datasets achieved higher evaluation than UNSW-NB15 dataset because of the complexity of the UNSW-NB15 dataset that has an assortment of the contemporary anomaly and normal behaviors which has been developed for evaluating NIDSs. Figures 5, 6 and 7 showed the prequential accuracy for the seven classifiers for KDD Cup 99, UNSW-NB15 and Kyoto 2006+ datasets respectively. The three previous figures demonstrated that the AIDM which combined with deep learning Dl4jMlp Classifier and based on feature selection performs the best efficiency with smooth curve compared to the other curves of the classifiers.

The proposed model results as shown in Table 2 indicated promising results in terms of low computational time and high classification results for Kyoto 2006+ dataset where it achieved 92.39% accuracy with a low time of 2.85 sec by using DL4JMLP. It is noticed that the time is somewhat close for both datasets KDD-Cup 99 and UNSW-NB15 which they have the same number of selected features as mentioned in the pre-processing stage.

For the KDD-CUP99 dataset, the DL4JMLP classifier achieved a slight increase in efficiency with 97.9% than Ibk classifier which achieved 97.52%. And also, DL4JMLP classifier outperformed with a slight increase of 89.12% over NNGE which achieved 88.76% for the UNSW-NB15 dataset.

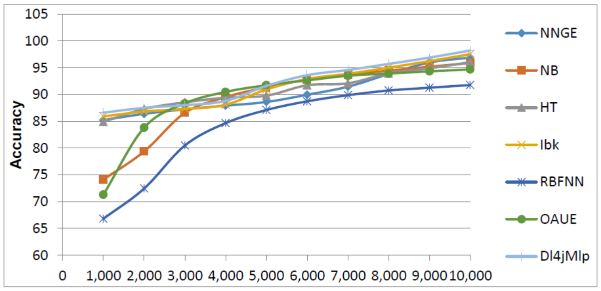


Figure 5: Prequential accuracy (%) for the KDD-CUP4 99 dataset

The accuracy of the seven classifiers has been recorded and illustrated as shown in the three previous figures for every 1000 instances between samples for the three datasets used, and indicated that AIDM with DL4JMLP classifier is a learnable model over time and with increase of the samples.

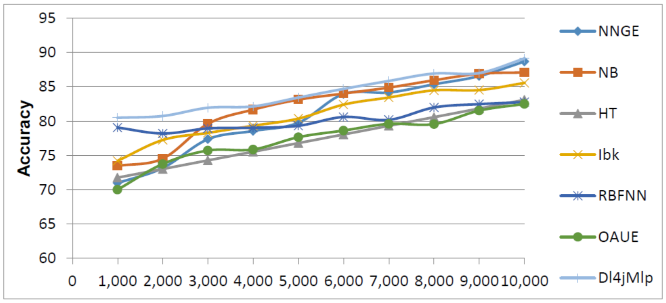


Figure 6: Prequential accuracy (%) for the UNSW-NB15 dataset

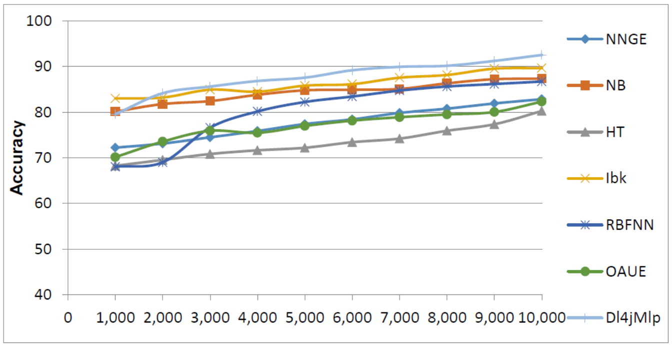


Figure 7: Prequential accuracy (%) for the Kyoto 2006+ dataset

6 Comparative Study

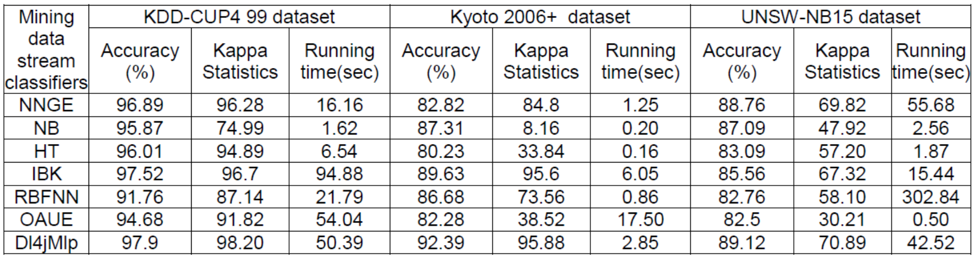
In order to evaluate the performance of our proposed model, we have compared AIDM with some other state-of-the-art models. All evidence proved that AIDM which based on feature selection technique and the Dl4JMlp classifier is a promising model to compare with other state-of-art models. We have made a comparison between AIDM and the proposed model in [13] which runs on the NSL-KDD dataset due to that the two papers based on the deep learning technique.

AIDM helped to increase the detection rate of abnormal attacks by 97.9% accuracy for the KDD-CUP99 dataset and the other model has achieved 78.06% accuracy for the NSL-KDD dataset.

Another comparison has been done between the proposed model in [4] which based on feature extraction technique and runs on UNSW-NB15 and KDD-CUP99 datasets.

AIDM still achieved a high accuracy of 89.12% and 97.9% for UNSW-NB15 and KDD-CUP99 datasets respectively while the proposed model in [4] achieved an accuracy of 88.95% and 89.85% for UNSW-NB15 and KDD-

Table 6: Accuracy, Kappa statistic and running time (sec) for three datasets using six incremental classifiers



Mining data stream classifiers	KDD-CUP4 99 dataset			Kyoto 2006+ dataset			UNSW-NB15 dataset		
	Accuracy (%)	Kappa Statistics	Running time(sec)	Accuracy (%)	Kappa Statistics	Running time(sec)	Accuracy (%)	Kappa Statistics	Running time(sec)
NNGE	96.89	96.28	16.16	82.82	84.8	1.25	88.76	69.82	55.68
NB	95.87	74.99	1.62	87.31	8.16	0.20	87.09	47.92	2.56
HT	96.01	94.89	6.54	80.23	33.84	0.16	83.09	57.20	1.87
IBK	97.52	96.7	94.88	89.63	95.6	6.05	85.56	67.32	15.44
RBFNN	91.76	87.14	21.79	86.68	73.56	0.86	82.76	58.10	302.84
OAUE	94.68	91.82	54.04	82.28	38.52	17.50	82.5	30.21	0.50
Dl4jMlp	97.9	98.20	50.39	92.39	95.88	2.85	89.12	70.89	42.52

CUP99 datasets respectively.

Our AIDM enjoyed with higher performance compared with the other proposed models and this is demonstrated that our model is a learnable and predictive model. Depending on the comparisons with the other two models in [13] and [4], AIDM model based on feature selection technique achieved the best accuracy by the DL4JMLP classifier than other traditional IDS and state-of-the-art models for intrusion detection systems.

7 Conclusion

Most recent researches trend towards the creation of a proposed model for detecting the anomalous connections by using an efficient classification technique and supported by feature extraction technique from datasets. In this paper, an intelligent, effective and learnable model has been built based on machine learning techniques with the ability to reduce feature extracted from the three datasets by selecting the most relevant attributes using information gain method. The proposed model outperformed typical machine intrusion detection systems and incremental learning by using DL4JMLP deep learning classifier in addition feature extraction technique. The evaluation has been conducted on the processed three datasets, namely, KDD CUP 99, UNSW-NB15 and real Kyoto 2006+ datasets in the preprocessing engine, which has a skillful sub model which has ability to detect the number of features in the datasets and apply the proper techniques of feature selection and coding beside, a normalization process for the input datasets. AIDM was tested using 10,000 random sets of the three processed datasets using a classification stage which based on 6 incremental classifiers and one deep learning classifier, namely, Non-Nearest Generalized Exemplar (NNGE), Incremental Naïve Bayes (NB), Hoffeding Trees (HF), Instance-based k-nearest neighbor IBK (KNN), and Radial Basis Function Neural Network (RBFNN), Online Update Accuracy Ensemble (OUAE) and DL4JMLP deep learning classifier. AIDM has achieved a 97.9% high accuracy with DL4JMLP deep learning classifier compared with other incremental learning classifiers as mentioned in the experimental results and discussion section furthermore, it has fulfilled better results compared with the state-of-the-art models as shown in the comparative study section all of these positives indicate that AIDM is powerful, learnable and a good real-time model for classifying abnormal and normal traffics in the network flow.

References

[1] M. S. I. Ahmed, A. M. Riyad, R. L. R. Khan, M. J. K, and E. Shamsudeen, "Information based feature selection for intrusion detection systems," *International Journal of Scientific & Engineering Research*, vol. 8, no. 7, pp. 2362-2366, 2017.

[2] V. R. Balasaraswathi, M. Sugumaran, and Y. Hamid, "Feature selection techniques for intrusion detection using non-bio-inspired and bio-inspired optimization algorithms," *Journal of Communications and Information Networks*, vol. 2, no. 4, pp. 107-119, 2017.

[3] J. E. Ball, D. T. Anderson, and C. S. Chan, "A comprehensive survey of deep learning in remote sensing: theories, tools and challenges for the community," *Journal of Applied Remote Sensing*, vol. 11, no. 4, 2017.

[4] M. Belouch, "A two-stage classifier approach using reptree algorithm for network intrusion detection," *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 6, pp. 389-394, 2017.

[5] B. Chakrabarty, O. Chanda, and M. Saiful, "Anomaly based intrusion detection system using genetic algorithm and K-centroid clustering," *International Journal of Computer Applications*, vol. 163, no. 11, pp. 13-17, 2017.

[6] M. M. U. Chowdhury, F. Hammond, G. Konowicz, C. Xin, H. Wu, and J. Li, "A few-shot deep learning approach for improved intrusion detection," in *8th IEEE Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON'17)*, pp. 456-462, 2017.

[7] K. Datasets, D. G. Mogal, S. R. Ghungrad, and B. B. Bhusare, "NIDS using machine learning classifiers on UNSW-NB15 and KDDCUP99 datasets," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 6, no. 4, pp. 533-537, 2017.

[8] M. El-Boujnouni and M. Jedra, "New intrusion detection system based on support vector domain description with information gain metric," *Interna-*

tional Journal of Network Security, vol. 20, no. 1, pp. 25-34, 2018.

[9] L. Gnanaprasanambikai and N. Munnusamy, "Data preprocessing and classification for traffic anomaly intrusion detection using NSLKDD dataset," *International Journal of Pure and Applied Mathematics*, vol. 119, no. 10, pp. 847-858, 2018.

[10] J. Henrydoss, S. Cruz, E. M. Rudd, M. Gunther, and T. E. Boult, "Incremental open set intrusion recognition using extreme value machine," in *16th IEEE International Conference on Machine Learning and Applications (ICMLA'17)*, pp. 1089-1093, 2017.

[11] V. K. Jabali, "Taxonomy of feature selection in intrusion detection system," *IJCSNS International Journal of Computer Science and Network Security*, vol. 17, no. 6, pp. 88-102, 2017.

[12] M. A. Jabbar, R. Aluvalu, and S. S. Reddy, "RFAODE: A novel ensemble intrusion detection system," in *7th International Conference on Advances in Computing & Communications*, vol. 115, pp. 226-234, 2017.

[13] A. Jayaswal, "Detecting network intrusion through a deep learning approach," *International Journal of Computer Applications*, vol. 180, no. 14, pp. 15-19, 2018.

[14] J. Josemila Baby and J. Jeba, "Survey paper on various hybrid and anomaly based network intrusion detection system," *Research Journal of Applied Sciences*, vol. 12. pp. 304-310, 2017.

[15] H. A. Kholidy, "Attacks detection in SCADA systems using an improved non-nested generalized exemplars algorithm," in *12th International Conference on Computer Engineering and Systems (ICCES'17)*, pp. 607-612, 2017.

[16] A. A. Kolpyakwar, M. G. Ingle, and R. V. Deshmukh, "A survey on data mining approaches for network intrusion detection system," *International Journal of Computer Applications*, vol. 159, no. 1, pp. 20-23, 2017.

[17] D. A. Kumar, S. Venugopalan, "The effect of normalization on intusion detection classifiers (Naïve Bayes and J48)," *International Journal on Future Revolution in Computer Science & Communication Engineering*, vol. 3, pp. 60-64, 2017.

[18] A. Mansourkhaki, M. Berangi, and M. Haghiri, "Comparative application of radial basis function and multilayer perceptron neural networks to predict traffic noise pollution in tehran roads," *Journal of Ecological Engineering*, vol. 19, no. 1, pp. 113-121, 2018.

[19] A. Muallem, S. Shetty, J. W. Pan, J. Zhao, and B. Biswal, "Hoeffding tree algorithms for anomaly detection in streaming datasets: A survey," *Journal of Information Security*, vol. 8, no. 4, pp. 339-361, 2017.

[20] L. R. Nair, S. D. Shetty, and S. D. Shetty, "Streaming big data analysis for real-time sentiment based targeted advertising," *International Journal of Electrical and Computer Engineering*, vol. 7, no. 1, pp. 402-407, 2017.

[21] A. R. Onik and T. Samad, "A network intrusion detection framework based on bayesian network using wrapper approach," *International Journal of Computer Applications*, vol. 166, no. 4, pp. 13-17, 2017.

[22] M. Panda and M. R. Patra, "A comparative study of data mining algorithms for network intrusion detection," *Annals of Emerging Technologies in Computing*, vol. 2, no. 1, pp. 49-57, 2018.

[23] P. Parrend, J. Navarro, F. Guigou, A. Deruyver, and P. Collet, "Foundations and applications of artificial Intelligence for zero-day and multi-step attack detection," *EURASIP Journal on Information Security*, 2018.

[24] E. Popoola and A. Adewumi, "Efficient feature selection technique for network intrusion detection system using discrete differential evolution and decision tree," *International Journal of Network Security*, vol. 19, no. 5, pp. 660-669, 2017.

[25] K. K. Ravulakollu, "A hybrid intrusion detection system: integrating hybrid feature selection approach with heterogeneous ensemble of intelligent classifiers," *International Journal of Network Security*, vol. 20, no. 1, pp. 41-55, 2018.

[26] H. Shapoorifard, "Intrusion detection using a novel hybrid method incorporating an improved KNN," *International Journal of Computer Applications*, vol. 173, no. 1, pp. 5-9, 2017.

[27] I. Sharafaldin, A. Gharib, A. H. Lashkari, and A. A. Ghorbani, "Towards a reliable intrusion detection benchmark dataset," *Software Networking*, vol. 6, no. 1, pp. 177-200, 2017.

[28] S. Sharma, J. Thakkar, and J. Patel, "A survey on supply chain cyber security," in *National Conference on Latest Trends in Networking and Cyber Security*, pp. 77-79, 2017.

[29] N. Shone, T. N. Ngoc, V. D. Phai, and Q. Shi, "A deep learning approach to network intrusion detection," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 1, pp. 1-10, 2017.

[30] V. T. Slavko Gajin, "Ensemble classifiers for supervised anomaly based network intrusion detection," in *13th IEEE International Conference on Intelligent Computer Communication and Processing (ICCP'17)*, pp. 13-19, 2017.

[31] P. Sudha, "Feature selection techniques for the classification of leaf diseases in turmeric," *International Journal of Computer Trends and Technology*, vol. 43, no. 3, pp. 138-142, 2017.

[32] J. Surana, J. Sharma, I. Saraf, N. Puri, and B. Navin, "A survey on intrusion detection system," *International Journal of Engineering Development and Research*, vol. 5, no. 2, pp. 960-965, 2017.

[33] H. Tanaka, "Effectiveness and weakness of quantified/automated anomaly based IDs," *International Journal of Network Security & Its Applications*, vol. 9, no. 6, pp. 1-11, 2017.

[34] C. J. Ugochukwu, E. O. Bennett, and P. Harcourt, "An intrusion detection system using machine learning algorithm," *International Journal of Computer*

Science and Mathematical Theory, vol. 4, no. 1, pp. 39-47, 2018.

[35] R. Vargas, A. Mosavi, R. Ruiz, "Deep learning: A review," *Advances in Intelligent Systems and Computing*, vol. 5, July, 2017.

[36] L. Wang and R. Jones, "Big data analytics for network intrusion detection: A survey," *International Journal of Networks and Communications*, vol. 7, no. 1, pp. 24-31, 2017.

[37] J. Zhong, Z. Liu, Y. Zeng, L. Cui, and Z. Ji, "A survey on incremental learning," in *5th International Conference on Computer, Automation and Power Electronics (CAPE'17)*, no. Cape, pp. 166-174, 2017.

Biography

Marwa R. Mohamed was born in 1982 in Egypt, she received her B.S.C degree in Electrical Engineering in 2005 from the Faculty of Engineering, Helwan University, Egypt. She had worked as a system administrator at Gulf English School (GES)-Cairo. She has successfully completed the Networking security and Management track granted through MCIT Scholarship at Raya Academy for a total of 1200 hours. She was ranked within the top 20% of trainers, accordingly, she has taken part from 04/02/2007 to 04/05/2007 and she has a certificate from Raya with this period.

Abdurrahman A. Nasr is a lecturer of software engineering, Computer, and System Engineering Department, Faculty of Engineering, Al-Azhar University at Cairo. He received his M.Sc. and Ph.D. degrees in electrical engineering from Al-Azhar University in 2012, and 2014 respectively. His fields of interest include artificial intelligence, stochastic process, machine learning, data mining, mathematics and operating systems.

I. F. Tarrad was born in 1959 in Egypt, He received his B.S.c degree in Electrical Engineering in 1984 from the Faculty of Engineering, Al-Azhar University, Egypt. He had worked as a demonstrator (teaching and research assistant) at the Al-Azhar University. He received his Master of Science degree in communication engineering 1989 and Ph.D. from Hungarian Academy of the Sciences Technical University of Budapest in 1996. In 1996 he was appointed as Lecturer at Department of Communication Engineering, Al-Azhar University, his research activities are within mobile, Computer Network, and digital communication.

Salah R. Abdulmageed received his M.S. and Ph.D. in Systems and Computers Engineering from Al-Azhar University in 2002 and 2005, respectively. Since 2017, he is a professor and head of Systems and Computers Engineering Department at Al-Azhar University. He performed his postdoctoral research in 2007 and 2008 in the Computer Science and Engineering Department, School of Engineering, the Southern Methodist University at Dallas, TX, USA. He was a software development manager of TEMPO (Tool for Extensive Management and Performance Optimization) project in Cairo University and Vodafone Egypt asan industrial partner in 2014 and 2015. His research interests include Mobile Computing, Cellular Networks, Sensor Networks, Cognitive Radio Networks, Vehicular Ad-hoc Networks, Big Data and Data Analysis, Internet Services and Applications.

Tight Proofs of Identity-based Signatures without Random Oracle

Huiyan Chen[1], Yanshuo Zhang[1], Zongjie Wan[1], Chenchen Zhang[1,2]
(Corresponding author: Huiyan Chen)

Beijing Electronic Science and Technology Institute[1]
No. 7, Fufeng road, Fengtai district, Beijing 100070, China
Xidian University, Xi'an 710126, China[2]
(Email: chenhy03@126.com)

(Received Nov. 10, 2017; revised and accepted May 5, 2018)

Abstract

It is a very desirable property of an identity-based signature to have a tight security reduction. According to our known knowledge, there are few results on designing identity-based signature schemes with tight security reduction. Inspired by the work of David Galindo *et al.* [13] and based on the signatures proposed by Sven Schäge [36, 37], we construct identity-based signatures which are existentially unforgeable under adaptively chosen message and identity attacks and whose security is also tightly related to Strong Diffie-Hellman assumption in the standard model.

Keywords: Existential Unforgeability; Identity-based Signature; q-Strong Diffie-Hellman Problem; Standard Model

1 Introduction

One focus of modern cryptography has been the construction of identity-based signature scheme that can be rigorously proven secure based on specific computational assumptions.

A number of identity-based signature (IBS) schemes [7–9, 16, 17, 21, 26, 28, 30, 32, 34, 40–43] have been devised since the concept of identity-based cryptography was proposed by Shamir [39] in 1984. At present, there are two known generic constructions of IBS. The first is due to Bellare *et al.* [29]. They show that a large number of previously proposed schemes are instances of their generic construction. The other generic construction is due to Kurosawa and Heng [15]. The construction of Kurosawa and Heng requires an efficient zero-knowledge protocol for proof of knowledge of a signature, which makes their construction applicable to only a few schemes such as RSA-FDH and BLS [22].

1.1 Our Contribution

In this work, we ask the following question: how does one construct identity-based signature with tight security proof in the standard model? The security of an IBS scheme is generally confirmed by a security proof which typically describes a reduction from some hard computational problem to breaking a defined security property of the IBS scheme. The reduction for the IBS scheme is considered as tight when this success probability of an adversary breaking the IBS is roughly equal to the probability of solving the underlying hard problem in roughly the same amount of time. Tightness of security reduction gives explicit bound on the probability that adversary successfully forges a signature for an IBS scheme as a function of its expended resources, and affects the efficiency of the IBS scheme when instantiated in practice: A tighter reduction allows to securely use smaller parameters, *e.g.*, shorter moduli, a smaller group size. Therefore it is a very desirable property of an IBS to have a tight security reduction. According to our known knowledge, there are few results on designing IBS schemes with tight security reductions. In this paper, we study the problem above and our work stems from the results of Sven Schäge [36, 37] and Galindo *et al.* [13]. In [36, 37], Schäge presented combing function based signature and chameleon hash function based signature which are strongly existential unforgeability under adaptively chosen message attack in standard model and which have tight security proof. In [13], Galindo *et al.* gave a Schnorr-like identity-based signature which is existentially unforgeable under adaptively chosen message and identity attack in random oracle model. Galindo *et al.*'s work is different from that of Bellare *et al.* [29], is also different from that of Kurosawa and Heng [15]. Inspired by the work of Galindo *et al.*, we construct four identity-based schemes with tight security reduction from combing function based signature and chameleon hash function based signature [36,37]. According to the type of parings used in our four schemes, we have divided them into two types and denoted them as

TYPE I and TYPE II, respectively. TYPE I is based the fact that there is efficiently computable homomorphism on the bilinear groups $(\mathbb{G}_1, \mathbb{G}_2)$, and TYPE II is just the opposite. According to the efficiency and the security, we compare our IBS scheme with the known IBS schemes in Table 1.

1.2 Related Works

It is a very desirable property of an IBS scheme to have a tight security reduction. Therefore, providing new security proofs for cryptosystems that were already well known to be secure in the random oracle model or for some of their variants (*e.g.*, [2, 3, 10, 11, 27, 36, 37]) and constructing new schemes (*e.g.*, [5, 6, 14, 18, 19, 23–25]) that provide tight security reductions have been a new research focus in in the area of provable security. In addition, to verify whether there is a tight security proof for the Schnorr signature scheme, cryptographers have given considerable research efforts, *e.g.*, [4, 31, 38].

However, the research on tight security reduction for IBS schemes has made little progress. In fact, Hess and Barreto *et al.* gave proofs under the Diffie-Hellman assumption for their respective scheme through Pointcheval and Stern's forking lemma [35] which does not yield tight security reductions. Chen *et al.* [7,8,21] gave proofs under the Diffie-Hellman assumption for their schemes by "ID reduction technique" from [1] which does not yield tight security reductions. Bellare *et al.* [29] defined a framework to provide security proofs for a large family of IBS schemes. Unfortunately, their framework does not provide tight security bounds for the resulting family of IBS. Kurosawa and Heng [15] showed a transformation from any digital signature scheme satisfying certain condition to an IBS scheme and gave security proof for the resulting IBS scheme. Although their security proof avoids the use of the forking technique, their reduction is still quite loose. Until today, there have few results on IBS schemes with tight security reductions except that the scheme was constructed by Libert *et al.* [20].

2 Preliminaries

2.1 Security Notion of Signature Scheme

A signature scheme is made up of three algorithms, KeyGen, Sign, and Verify, for generating keys, signing, and verifying signatures, respectively.

The standard notion of security for a signature scheme is called existential unforgeability under a chosen message attack, which is defined using the following game between a challenger $\mathcal{C}$ and an adversary $\mathcal{A}$:

Setup. $\mathcal{C}$ runs the algorithm `KeyGen` of the signature scheme and obtains both the public key PK and the private key SK. The adversary $\mathcal{A}$ is given PK but the private key SK is kept by the challenger.

Queries. Proceeding adaptively, $\mathcal{A}$ requests signatures on at most q_S messages of his choice $m_1, \ldots, m_{q_S} \in \{0,1\}^*$ under PK. $\mathcal{C}$ responds to each query with a signature $\sigma_i = \mathtt{Sign}(SK, m_i)$.

Forgery. The adversary outputs a pair (m^*, σ^*). The adversary succeeds if the following hold true:

1) $\mathtt{Verify}(PK, m^*, \sigma^*) = \mathtt{accept}$.
2) m^* is not any of $m_1, , \ldots, m_{q_S}$.

We define $\mathtt{AdvSig}_{\mathcal{A}}$ to be the probability that $\mathcal{A}$ wins in the above game, taken over the coin tosses made by $\mathcal{A}$ and the challenger.

Definition 1. *An adversary $\mathcal{A}$ (t, q_S, ε)-breaks a signature scheme if $\mathcal{A}$ runs in time at most t and makes at most q_S signature queries in the above game, and $\mathtt{AdvSig}_{\mathcal{A}}$ is at least ε. A signature scheme is (t, q_S, ε)- existentially unforgeable under adaptively chosen message attacks if no adversary (t, q_S, ε)-breaks it.*

We also consider a slightly stronger notion of security, called strong existential unforgeability. The above game can easily be extended to cover strongly existential unforgeability by changing the second requirement in the forgery stage as follows.

Forgery. The adversary outputs a pair (m^*, σ^*). The adversary succeeds if the following hold true:

1) $\mathtt{Verify}(PK, m^*, \sigma^*) = \mathtt{accept}$.
2) (m^*, σ^*) is not any of $(m_1, \sigma_1), \ldots (m_{q_S}, \sigma_{q_S})$.

Definition 2. *An adversary $\mathcal{A}$ (t, q_S, ε)-breaks a signature scheme if $\mathcal{A}$ runs in time at most t and makes at most q_S signature queries in the modified game above, and $\mathtt{AdvSig}_{\mathcal{A}}$ is at least ε. A signature scheme is (t, q_S, ε)-strongly existentially unforgeable under adaptively chosen message attacks if no adversary (t, q_S, ε)-breaks it.*

2.2 Security Notion of Identity-Based Signature Scheme

An identity-based signature scheme can be described as a collection of the following four algorithms:

Setup. This algorithm is run by the "Private Key Generator" (PKG) on input a security parameter, and generates the public parameters *params* of the scheme and a master secret. PKG publishes *params* and keeps the master secret to itself.

Extract. Given an identity ID, the master secret and *params*, this algorithm generates the private key d_{ID} of ID. PGK will use this algorithm to generate private keys for all entities participating in the scheme and distribute the private keys to their respective owners through a secure channel.

Table 1: Scheme comparison

Scheme	Reduction	Type of Pairing	Pairing Operation	Security Assumption	Random Oracles
KJ [32]	Loose	Type 1	3	CDH	NO
BJ [20]	Tight	Type 1	2	one more CDH	YES
RG [41]	Tight	Type 1	2	SDH	NO
TYPE I	Tight	Type 1,2	2	SDH	NO
TYPE II	Tight	Type 3	4	SDH	NO

Sign. Given a message m, an identity ID, a private key d_{ID} and *params*, this algorithm generates the signature σ of ID on m. The entity with identity ID will use this algorithm for signing.

Verify. Given a signature σ, a message m, an identity ID and *params*, this algorithm outputs `accept` if σ is a valid signature on m for identity ID, and outputs `reject` otherwise.

We recall here the security notion [20] for identity-based signatures which is an extension of the usual notion of existential unforgeability under chosen-message attacks for signature and which is defined security for identity-based signature schemes by the following game between a challenger $\mathcal{C}$ and an adversary $\mathcal{A}$:

Setup. $\mathcal{C}$ runs the algorithm `Setup` of the signature scheme and obtains both the public parameters *params* and the master secret SK. $\mathcal{A}$ is given *params* but the master secret SK is kept by the challenger.

Queries. The adversary $\mathcal{A}$ adaptively makes a number of different queries to the challenger.

1) **Extraction query**.Proceeding adaptively, $\mathcal{A}$ requests extractions on at most q_E identities of his choice $ID_1, \ldots, ID_{q_E} \in \{0,1\}^*$. $\mathcal{C}$ responds to each query with $d_{ID_i} = \texttt{Extract}(param, SK, ID_i)$.

2) **Signature query**.Proceeding adaptively, $\mathcal{A}$ requests signatures on at most q_S messages of his choice $(ID_{i_1}, m_1), \ldots, (ID_{i_{q_S}}, m_{q_S}) \in \{0,1\}^* \times \{0,1\}^*$. $\mathcal{C}$ responds to each query by running $\texttt{Extract}(params,\ SK, ID_{i_j})$ to obtain the private key $d_{ID_{i_j}}$ of ID_{i_j}, then running σ_j=$\texttt{Sign}(params,\ d_{ID_{i_j}},\ ID_{i_j}, m_j)$, last forwarding σ_j to the adversary $\mathcal{A}$.

Forgery. The adversary outputs a tuple $(ID^*,\ m^*,\sigma^*)$. The adversary succeeds if the following hold true:

1) $\texttt{Verify}(params, ID^*, m^*, \sigma^*)$=`accept`.
2) ID^* was not any of $ID_1, \ldots, ID_{q_E}$.
3) (ID^*, m^*) was not any of $(ID_{i_1},\ m_1), \ldots, (ID_{i_{q_S}},\ m_{q_S})$.

We define $\texttt{AdvSig}_{\mathcal{A}}$ to be the probability that $\mathcal{A}$ wins in the above game, taken over the coin tosses made by $\mathcal{A}$ and the challenger.

Definition 3. *An adversary $\mathcal{A}$ $(t, q_E, q_S, \varepsilon)$-breaks an IBS signature scheme if $\mathcal{A}$ runs in time at most t and makes at most q_S signature queries, q_E extraction queries in the above game, and* $\texttt{AdvSig}_{\mathcal{A}}$ *is at least ε. A signature scheme is $(t, q_E, q_S, \varepsilon)$- existentially unforgeable under adaptively chosen message and identity attacks if no adversary $(t, q_E, q_S, \varepsilon)$-breaks it.*

2.3 Bilinear Parings and Complexity Assumptions

We consider the mathematical preliminaries for constructing and proving our signature schemes.

Let us consider three cyclic multiplicative group $\mathbb{G}_1$, $\mathbb{G}_2$ and $\mathbb{G}_T$ of the same prime order p. Let g_1 be a generator of $\mathbb{G}_1$, g_2 be a generator of $\mathbb{G}_2$. Let $\widehat{e} : \mathbb{G}_1 \times \mathbb{G}_2 \to \mathbb{G}_T$ be a bilinear pairing with the following properties:

Bilinearity: $\widehat{e}(u^a, v^b) = \widehat{e}(u, v)^{ab}$ for all $u \in \mathbb{G}_1$, $v \in \mathbb{G}_2$, $a, b \in Z_p$.

Non-degeneracy: There exists $u \in \mathbb{G}_1$, $v \in \mathbb{G}_2$ such that $\widehat{e}(u, v) \neq 1$.

Computability: There is an efficient algorithm to compute $\widehat{e}(u, v)$ for all $u \in \mathbb{G}_1$, $v \in \mathbb{G}_2$.

Definition 4. ***Bilinear Groups.*** *We say that $(\mathbb{G}_1, \mathbb{G}_2)$ are bilinear groups if there exists a group $\mathbb{G}_T$ and a non-degenerate bilinear pairing $\widehat{e} : \mathbb{G}_1 \times \mathbb{G}_2 \to \mathbb{G}_T$, such that the group order of $\mathbb{G}_1, \mathbb{G}_2$, and $\mathbb{G}_T$ is a prime p, and the bilinear map $\widehat{e}$ and the group operations in $\mathbb{G}_1, \mathbb{G}_2$, and $\mathbb{G}_T$ are all efficiently computable.*

Galbraith, Paterson, and Smart [33] defined three types of pairings:

- In Type 1, $\mathbb{G}_1 = \mathbb{G}_2$.
- In Type 2, $\mathbb{G}_1 \neq \mathbb{G}_2$ but there exists an efficient homomorphism ψ: $\mathbb{G}_2 \to \mathbb{G}_1$, while no efficient one exists in the other direction.
- In Type 3, $\mathbb{G}_1 \neq \mathbb{G}_2$ and no efficiently computable homomorphism exist between $\mathbb{G}_1$ and $\mathbb{G}_2$, in either direction.

Although Type 1 pairings were mostly used in the early-age of pairing-based cryptography, they have been gradually discarded in favor of Type 3 pairings.

Definition 5. ***q-Strong Diffie-Hellman Problem (q-SDH).*** *Over bilinear groups $(\mathbb{G}_1, \mathbb{G}_2)$, given as input a $q+3$ tuple of elements $(g_1, g_1^x, g_1^{x^2}, \ldots, g_1^{x^q}, g_2, g_2^x)$ output a pair $(c, g_1^{1/(x+c)})$ for some value $c \in \mathbb{Z}_p \setminus \{-x\}$, where g_1 is a generator of $\mathbb{G}_1$ and g_2 is a generator of $\mathbb{G}_2$.*

An algorithm $\mathcal{A}$ solves the q-SDH problem over bilinear groups $(\mathbb{G}_1, \mathbb{G}_2)$ with advantage ε if

$$\texttt{SDHAdv}_{q,\mathcal{A}} = Pr[\mathcal{A}(g_1, g_1^x, g_1^{x^2}, \ldots, g_1^{x^q}, g_2, g_2^x) = (c, g_1^{1/(x+c)})] \geq \varepsilon$$

where the probability is over the random choice of generators $g_1 \in \mathbb{G}_1$ and $g_2 \in \mathbb{G}_2$, the random choice of $x \in \mathbb{Z}_p^*$, and the random bits consumed by $\mathcal{A}$.

Definition 6. ***Strong Diffie-Hellman Assumption (SDH).*** *We say that the (q, t, ε)-SDH assumption holds over bilinear groups $(\mathbb{G}_1, \mathbb{G}_2)$ if no t-time algorithm has advantage at least ε in solving the q-SDH problem over the bilinear groups $(\mathbb{G}_1, \mathbb{G}_2)$.*

2.4 Chameleon Hash Function and Combining Function

In this section, we review the notions of chameleon hash function and combining function from [36, 37].

A chameleon hash function CH = (CHGen, CHEval, CHColl) consists of three algorithms. The probabilistic polynomial-time algorithm CHGen takes as input the security parameter k and outputs a secret key $\texttt{SK}_{\texttt{CH}}$ and a public key $\texttt{PK}_{\texttt{CH}}$. Given $\texttt{PK}_{\texttt{CH}}$, a random r from a randomization space $\mathcal{R}$ and a message m from a message space $\mathcal{M}$, the algorithm CHEval outputs a chameleon hash value c in the hash space $\mathcal{C}$. Analogously, CHColl deterministically outputs, on input $\texttt{SK}_{\texttt{CH}}$ and $(r, m, m') \in \mathcal{R} \times \mathcal{M} \times \mathcal{M}$, $r' \in \mathcal{R}$ such that $\texttt{CHEval}(\texttt{PK}_{\texttt{CH}}, m, r) = \texttt{CHEval}(\texttt{PK}_{\texttt{CH}}, m', r')$.

Definition 7. ***Collision-resistant chameleon hash function.*** *We say that CH is (ε, t)-collision-resistant if no t-time algorithm, only given PK_{CH}, outputs (r, r', m, m') such that $m \neq m'$ and $CHEval(PK_{CH}, m, r) = CHEval(PK_{CH}, m', r')$ with probability at least ε, where the probability is over the random choices of PK_{CH} and the coin tosses of algorithm.*

For the convenience of writing, we write $\texttt{CH}(r, m)$ to denote $\texttt{CHEval}(\texttt{PK}_{\texttt{CH}}, r, m)$ and $\texttt{CH}^{-1}(r, m, m')$ for $\texttt{CHColl}(\texttt{SK}_{\texttt{CH}}, r, m, m')$ if the keys are obvious from the context.

Definition 8. ***Combining Functions.*** *Let $\mathcal{V}_k$ for $k \in N$ be a collection of functions of the form $z : \mathcal{R} \times \mathcal{M} \rightarrow \mathcal{Z}$ with $|\mathcal{Z}| \leq 2^k$. Let $\mathcal{V} = \{\mathcal{V}_k\}_{k \in N}$. We say that $\mathcal{V}$ is (t, ε, δ)-combining if for all attackers $\mathcal{A}$ there exist negligible functions ε and δ and the following properties hold for randomly picked z from $\mathcal{V}_k$.*

1) *for all $m \in \mathcal{M}$ it holds that $|\mathcal{R}| = |\mathcal{Z}_m|$ where $\mathcal{Z}_m$ is defined as $\mathcal{Z}_m = z(\mathcal{R}, m)$. For all $m \in \mathcal{M}$ and all $t \in \mathcal{Z}$ there exists an efficient algorithm $z^{-1}(t, m)$ that, if $t \in \mathcal{Z}_m$, outputs the unique value $r \in \mathcal{R}$ such that $z(r, m) = t$, and $\perp$ otherwise.*

2) *for randomly picked $t \in \mathcal{Z}$ and $r' \in \mathcal{R}$, we have for the maximal (over all $m \in \mathcal{M}$) statistical distance between r' and $z^{-1}(t, m)$ that*

$$\underset{m \in \mathcal{M}}{\mathrm{MAX}}\{\frac{1}{2}\sum_{r \in \mathcal{R}} |Pr[r' = r] - Pr[z^{-1}(t, m) = r]|\} \leq \delta$$

3) *for all $r \in \mathcal{R}$, it holds for all t-time attackers $\mathcal{A}$ that output (m, m') such $m \neq m'$ and $z(r, m) = z(r, m')$ with probability at most ε.*

2.5 The SDH Signatures

The Boneh-Boyen (BB) signature [5, 6] is proven tightly secure under a new flexible assumption, the q-Strong Diffie Hellman (SDH) assumption and without random oracle. Based on this work, Sven Schäge [36, 37] gives combing function based signature (denoted as $\text{S}_{\texttt{CMB, SDH}}$, where CMB is the abbreviation of combing function) and chameleon hash function based signature (denoted as $\text{S}_{\texttt{CH, SDH}}$), respectively.

For the combining signature $\text{S}_{\texttt{CMB, SDH}}$ and the chameleon signature $\text{S}_{\texttt{CH, SDH}}$, if the combing function is $(t_{comb}, \varepsilon_{comb}, \delta_{comb})$-combining, where functions ε_{comb} and δ_{comb} are negligible, and chameleon hash function is collision-resistant, Sven Schäge [36, 37] gave the following result.

Proposition 1. *The combining signature $S_{\texttt{CMB, SDH}}$, and the chameleon signature $S_{\texttt{CH, SDH}}$ are tightly secure against strong existential forgeries under adaptively chosen message attacks.*

3 $\text{S}_{\texttt{CMB, SDH}}$ Based IBS

3.1 $\text{S}_{\texttt{CMB, SDH}}$ Based IBS over Bilinear Groups with Efficiently Computable Isomorphism

Let $(\mathbb{G}_1, \mathbb{G}_2)$ be bilinear groups with group order $|\mathbb{G}_1| = |\mathbb{G}_2| = p$ for some prime p,ψ be an efficiently computable isomorphism from $\mathbb{G}_2$ to $\mathbb{G}_1$, if $\mathbb{G}_1 = \mathbb{G}_2$, one could take ψ to be the identity map. For the moment we assume that the messages m to be signed are elements in $\mathbb{Z}_p$, but the domain can be extended to all of $\{0, 1\}^*$ using (target) collision resistant hashing.

Setup: Select five random generators a,b,c,$g_1 \in \mathbb{G}_1$ and $g_2 \in \mathbb{G}_2$, and random integers x,y,$z \in \mathbb{Z}_p^*$. Then, the bilinear map over bilinear groups $(\mathbb{G}_1, \mathbb{G}_2)$ is taken as $\widehat{e} : \mathbb{G}_1 \times \mathbb{G}_2 \rightarrow \mathbb{G}_T$. Compute $u = g_2^x \in \mathbb{G}_2$, $h_1 = g_1^{xy} \in \mathbb{G}_1$, $h_2 = g_1^y \in \mathbb{G}_1$, $f_1 = g_1^{xz} \in \mathbb{G}_1$,

$f_2 = g_1^z \in \mathbb{G}_1$. $\pi : \mathcal{R} \times \mathcal{ID} \to \mathcal{Z}$ is a combining function, where we assume that $\mathcal{Z} \subseteq \mathbb{Z}_p$, $\mathcal{R} \subseteq \mathbb{Z}_p$, $\mathcal{ID}$ is an identity space. Also compute $\gamma_a = \widehat{e}(a, g_2) \in \mathbb{G}_T, \gamma_b = \widehat{e}(b, g_2) \in \mathbb{G}_T, \gamma_c = \widehat{e}(c, g_2) \in \mathbb{G}_T$. The public system parameters are the tuple $(a, b, c, g_1, g_2, h_1, h_2, f_1, f_2, u, \gamma_a, \gamma_b, \gamma_c, \pi, \widehat{e})$. The master secret key is the triple (x, y, z).

Extraction: Given the secret key (x, y, z) and an identity $ID \in \mathcal{ID}$, pick chooses a random value $r \in \mathcal{R}$, a random value $r_0 \in \mathbb{Z}_p \backslash \{-x\}$ and compute $\tau = (ab^r c^{\pi(r,ID)})^{1/(x+r_0)} \in \mathbb{G}_1$. Here, the inverse $1/(x + r_0)$ is computed modulo p. The private key corresponding ID is the pair (τ, r, r_0).

Signature: Given a private key (τ, r, r_0) corresponding identity $ID \in \mathcal{ID}$ and a message $m \in \mathbb{Z}_p$. Pick a random value $k \in \mathbb{Z}_p$ and compute $\sigma_1 = \tau$ $((h_1 h_2^{r_0})^m (f_1 f_2^{r_0}))^k$, $\sigma_2 = (u g_2^{r_0})^k$. The signature is $\sigma = (\sigma_1, \sigma_2, r, r_0)$.

Verification: Given the public system parameters $(a, b, c, g_1, g_2, h_1, h_2, f_1, f_2, u, \gamma_a, \gamma_b, \gamma_c, \pi, \widehat{e})$, an identity $ID \in \mathcal{ID}$, a message $m \in \mathbb{Z}_p$, and a signature $\sigma = (\sigma_1, \sigma_2, r, r_0)$, verify that

$$\widehat{e}(\sigma_1, u g_2^{r_0}) = \gamma_a \gamma_b^r \gamma_c^{\pi(r,ID)} \widehat{e}((h_1 h_2^{r_0})^m (f_1 f_2^{r_0}), \sigma_2)$$

If the equality holds the result is valid; otherwise the result is invalid.

On the IBS scheme above, we have the following result.

Theorem 1. *Suppose the $S_{\mathtt{CMB},\ \mathtt{SDH}}$ signature is $(t', q_E + q_S, \varepsilon')$-secure against strongly existential forgery under an adaptively chosen message attack. Then the identity-based signature scheme above is $(t, q_E, q_S, \varepsilon)$-secure against existential forgery under an adaptively chosen massage and identity attack provided that $q_S + q_E \leq q$, $\varepsilon' \geq \varepsilon - 2q_S/p$, and $t' = t + O((6q_S + 4)T)$, where T is the maximum time for an exponentiation in $\mathbb{G}_1$, $\mathbb{G}_2$, and $\mathbb{Z}p$.*

Due to limited space, we omit the proof of Theorem 1.

According to Theorem 1 and Proposition 1, for the $\mathrm{S}_{\mathtt{CMB},\ \mathtt{SDH}}$ based identity-based signature, we get the following result.

Corollary 1. *Suppose SDH assumption holds in bilinear groups $(\mathbb{G}_1, \mathbb{G}_2)$. Then the identity-based signature above is tightly secure against existential forgeries under adaptively chosen massage and identity attacks in standard model.*

3.2 $\mathrm{S}_{\mathtt{CMB},\ \mathtt{SDH}}$ Based IBS over Bilinear Groups without Efficiently Computable Isomorphism

Let $(\mathbb{G}_1, \mathbb{G}_2)$ be bilinear groups where $|\mathbb{G}_1| = |\mathbb{G}_2| = p$ for some prime p, and there are no efficiently computable homomorphisms between $\mathbb{G}_1$ and $\mathbb{G}_2$. For the moment we assume that the messages m to be signed are elements in $\mathbb{Z}_p$, but the domain can be extended to all of $\{0, 1\}^*$ using (target) collision resistant hashing.

1) **Setup:** Select five random generators $a, b, c, g_1 \in \mathbb{G}_1$ and $g_2 \in \mathbb{G}_2$, and random integers $x, y, z \in \mathbb{Z}_p^*$. Then, the bilinear map over bilinear groups $(\mathbb{G}_1, \mathbb{G}_2)$ is taken as $\widehat{e} : \mathbb{G}_1 \times \mathbb{G}_2 \to \mathbb{G}_T$. Compute $u = g_2^x \in \mathbb{G}_2$, $h_1 = g_1^y \in \mathbb{G}_1$, $f_1 = g_1^z \in \mathbb{G}_1$. $\pi : \mathcal{R} \times \mathcal{ID} \to \mathcal{Z}$ is a combining function, where we assume that $\mathcal{Z} \subseteq \mathbb{Z}_p$, $\mathcal{R} \subseteq \mathbb{Z}_p$, $\mathcal{ID}$ is an identity space. Also compute $\gamma_a = \widehat{e}(a, g_2) \in \mathbb{G}_T, \gamma_b = \widehat{e}(b, g_2) \in \mathbb{G}_T, \gamma_c = \widehat{e}(c, g_2) \in \mathbb{G}_T$. The public system parameters are the tuple $(a, b, c, g_1, g_2, h_1, f_1, u, \gamma_a, \gamma_b, \gamma_c, \pi, \widehat{e})$. The master secret key is the triple (x, y, z).

2) **Extraction:** Given the secret key (x, y, z) and an identity $ID \in \mathcal{ID}$, pick chooses a random value $r \in \mathcal{R}$, a random value $r_0 \in \mathbb{Z}_p \backslash \{-x\}$ and compute $\tau = (ab^r c^{\pi(r,ID)})^{1/(x+r_0)} \in \mathbb{G}_1$. Here, the inverse $1/(x + r_0)$ is computed modulo p. The private key corresponding ID is the pair (τ, r, r_0).

3) **Signature:** Given a private key (τ, r, r_0) corresponding identity $ID \in \mathcal{ID}$ and a message $m \in \mathbb{Z}_p$, pick a random value $k \in \mathbb{Z}_p$ and compute $\sigma_1 = \tau (h_1^m f_1)^k$, $\sigma_2 = (u g_2^{r_0})^k$, $\sigma_3 = g_1^k$. The signature is $\sigma = (\sigma_1, \sigma_2, \sigma_3, r, r_0)$.

4) **Verification:** Given the public system parameters $(a, b, c, g_1, g_2, h_1, f_1, u, \gamma_a, \gamma_b, \gamma_c, \pi, \widehat{e})$, an identity $ID \in \mathcal{ID}$, a message $m \in \mathbb{Z}_p$, and a signature $\sigma = (\sigma_1, \sigma_2, \sigma_3, r, r_0)$, verify that

$$\widehat{e}(\sigma_1, u g_2^{r_0}) = \gamma_a \gamma_b^r \gamma_c^{\pi(r,ID)} \widehat{e}(h_1^m f_1, \sigma_2)$$
$$\widehat{e}(\sigma_3, u g_2^{r_0}) = \widehat{e}(g_1, \sigma_2).$$

If the equality holds the result is valid; otherwise the result is invalid.

On the IBS scheme above, we have the following result.

Theorem 2. *Suppose the $S_{\mathtt{CMB},\ \mathtt{SDH}}$ signature is $(t', q_E + q_S, \varepsilon')$-secure against strongly existential forgery under an adaptively chosen message attack. Then the identity-based signature scheme above is $(t, q_E, q_S, \varepsilon)$-secure against existential forgery under an adaptively chosen massage and identity attack provided that*

$$q_S + q_E \leq q,\ \varepsilon' \geq \varepsilon - 2q_S/p,\ \text{and}\ t' = t + O((5q_S + 4)T),$$

where T is the maximum time for an exponentiation in $\mathbb{G}_1$, $\mathbb{G}_2$, and $\mathbb{Z}p$.

Due to limited space, we omit the proof of Theorem 2.

According to Theorem 2 and Proposition 1, for the $\mathrm{S}_{\mathtt{CMB},\ \mathtt{SDH}}$ based identity-based signature, we get the following result.

Corollary 2. *Suppose SDH assumption holds in bilinear groups $(\mathbb{G}_1, \mathbb{G}_2)$. Then the identity-based signature above is tightly secure against existential forgeries under adaptively chosen massage and identity attacks in standard model.*

4 $S_{\texttt{CH, SDH}}$ Based IBS

4.1 Construction over Bilinear Groups with Efficiently Computable Isomorphism

Let $(\mathbb{G}_1, \mathbb{G}_2)$ be bilinear groups where $|\mathbb{G}_1| = |\mathbb{G}_2| = p$ for some prime p, ψ be an efficiently computable isomorphism from $\mathbb{G}_2$ to $\mathbb{G}_1$, if $\mathbb{G}_1 = \mathbb{G}_2$, one could take ψ to be the identity map. For the moment we assume that the messages m to be signed are elements in $\mathbb{Z}_p$, but the domain can be extended to all of $\{0,1\}^*$ using (target) collision resistant hashing.

1) **Setup:** Select random generators $a, b, g_1 \in \mathbb{G}_1$ and $g_2 \in \mathbb{G}_2$, and random integers $x, y, z \in \mathbb{Z}_p^*$. Then, the bilinear map over bilinear groups $(\mathbb{G}_1, \mathbb{G}_2)$ is taken as $\widehat{e} : \mathbb{G}_1 \times \mathbb{G}_2 \to \mathbb{G}_T$. Compute $u = g_2^x \in \mathbb{G}_2$, $h_1 = g_1^{xy} \in \mathbb{G}_1$, $h_2 = g_1^y \in \mathbb{G}_1$, $f_1 = g_1^{xz} \in \mathbb{G}_1$, $f_2 = g_1^z \in \mathbb{G}_1$. $\texttt{CH}$ is a chameleon hash function and its public key is $\texttt{PK}_{\texttt{CH}}$. Also compute $\gamma_a = \widehat{e}(a, g_2) \in \mathbb{G}_T$,$\gamma_b = \widehat{e}(b, g_2) \in \mathbb{G}_T$. The public system parameters are the tuple $(a, b, g_2, h_1, h_2, f_1, f_2, u, \gamma_a, \gamma_b, \texttt{CH}, \texttt{PK}_{\texttt{CH}}, \widehat{e})$. The master secret key is the triple (x, y, z).

2) **Extraction:** Given the master secret key (x, y, z) and an identity $ID \in \mathcal{ID}$, pick chooses a random $r \in \mathcal{R}$, a random $t \in \mathbb{Z}_p \backslash \{-x\}$ and compute $\tau = (ab^{\texttt{CH}(r,ID)})^{1/(x+t)} \in \mathbb{G}_1$. Here, the inverse $1/(x+t)$ is computed modulo p. The private key corresponding ID is the pair (τ, r, t).

3) **Signature:** Given a private key (τ, r, t) corresponding identity ID and a message $m \in \mathbb{Z}_p$, pick a random $k \in \mathbb{Z}_p$ and compute $\sigma_1 = \tau((h_1 h_2^t)^m (f_1 f_2^t))^k$, $\sigma_2 = (u g_2^t)^k$. The signature is $\sigma = (\sigma_1, \sigma_2, r, t)$.

4) **Verification:** Given the public system parameters $(a, b, c, g_1, g_2, h_1, h_2, f_1, f_2, u, \gamma_a, \gamma_b, \gamma_c, \pi, \widehat{e})$, an identity ID, a message $m \in \mathbb{Z}_p$, and a signature $\sigma = (\sigma_1, \sigma_2, r, t)$, verify that

$$\widehat{e}(\sigma_1, u g_2^t) = \gamma_a \gamma_b^{\texttt{CH}(r,ID)} \widehat{e}((h_1 h_2^t)^m (f_1 f_2^t), \sigma_2) \quad (1)$$

If the equality holds the result is valid; otherwise the result is invalid.

On the IBS scheme above, we have the following result.

Theorem 3. *Suppose the $S_{\texttt{CH, SDH}}$ signature is $(t', q_E + q_S, \varepsilon')$-secure against strongly existential forgery under an adaptively chosen message attack. Then the identity-based signature scheme above is $(t, q_E, q_S, \varepsilon)$-secure against existential forgery under an adaptively chosen massage and identity attack provided that $q_S + q_E \le q$, $\varepsilon' \ge \varepsilon - 2q_S/(p-1)$, and $t' = t + O((6q_S + 6)T)$, where T is the maximum time for an exponentiation in $\mathbb{G}_1$, $\mathbb{G}_2$, and $\mathbb{Z}p$.*

Due to limited space, we omit the proof of Theorem 3.

According to Theorem 3 and Proposition 1, for the $S_{\texttt{CH, SDH}}$ based identity-based signature, we get the following result.

Corollary 3. *Suppose SDH assumption holds in bilinear groups $(\mathbb{G}_1, \mathbb{G}_2)$. Then the identity-based signature above is tightly secure against existential forgeries under adaptively chosen massage and identity attacks in standard model.*

4.2 Construction over Bilinear Groups without Efficiently Computable Isomorphism

Let $(\mathbb{G}_1, \mathbb{G}_2)$ be bilinear groups where $|\mathbb{G}_1| = |\mathbb{G}_2| = p$ for some prime p, and there are no efficiently computable homomorphisms between $\mathbb{G}_1$ and $\mathbb{G}_2$. For the moment we assume that the messages m to be signed are elements in $\mathbb{Z}_p$, but the domain can be extended to all of $\{0,1\}^*$ using (target) collision resistant hashing.

1) **Setup:** Select random generators $a, b, g_1 \in \mathbb{G}_1$ and $g_2 \in \mathbb{G}_2$, and random integers $x, y, z \in \mathbb{Z}_p^*$. Then, the bilinear map over bilinear groups $(\mathbb{G}_1, \mathbb{G}_2)$ is taken as $\widehat{e} : \mathbb{G}_1 \times \mathbb{G}_2 \to \mathbb{G}_T$. Compute $u = g_2^x \in \mathbb{G}_2$, $h_1 = g_1^y \in \mathbb{G}_1$, $f_1 = g_1^z \in \mathbb{G}_1$. $\texttt{CH}$ is a chameleon hash function and its public key is $\texttt{PK}_{\texttt{CH}}$. Also compute $\gamma_a = \widehat{e}(a, g_2) \in \mathbb{G}_T$,$\gamma_b = \widehat{e}(b, g_2) \in \mathbb{G}_T$. The public system parameters are the tuple $(a, b, g_2, h_1, f_1, u, \gamma_a, \gamma_b, \texttt{CH}, \texttt{PK}_{\texttt{CH}}, \widehat{e})$. The master secret key is the triple (x, y, z).

2) **Extraction:** Given the master secret key (x, y, z) and an identity $ID \in \mathcal{ID}$, pick chooses a random $r \in \mathcal{R}$, a random $t \in \mathbb{Z}_p \backslash \{-x\}$ and compute $\tau = (ab^{\texttt{CH}(r,ID)})^{1/(x+t)} \in \mathbb{G}_1$. Here, the inverse $1/(x+t)$ is computed modulo p. The private key corresponding ID is the pair (τ, r, t).

3) **Signature:** Given a private key (τ, r, t) corresponding identity ID and a message $m \in \mathbb{Z}_p$, pick a random $k \in \mathbb{Z}_p$ and compute $\sigma_1 = \tau(h_1^m f_1)^k$, $\sigma_2 = (u g_2^t)^k$,$\sigma_3 = g_1^k$. The signature is $\sigma = (\sigma_1, \sigma_2, \sigma_3, r, t)$.

4) **Verification:** Given the public system parameters $(a, b, c, g_1, g_2, h_1, f_1, u, \gamma_a, \gamma_b, \gamma_c, \pi, \widehat{e})$, an identity ID, a message $m \in \mathbb{Z}_p$, and a signature $\sigma = (\sigma_1, \sigma_2, \sigma_3, r, t)$, verify that

$$\widehat{e}(\sigma_1, u g_2^t) = \gamma_a \gamma_b^{\texttt{CH}(r,ID)} \widehat{e}(h_1^m f_1, \sigma_2)$$
$$\widehat{e}(\sigma_3, u g_2^t) = \widehat{e}(g_1, \sigma_2)$$

If the equality holds the result is valid; otherwise the result is invalid.

On the IBS scheme above, we have the following result.

Theorem 4. *Suppose the $S_{\texttt{CH, SDH}}$ signature is $(t', q_E + q_S, \varepsilon')$-secure against strongly existential forgery under an adaptively chosen message attack. Then the identity-based signature scheme above is $(t, q_E, q_S, \varepsilon)$-secure against existential forgery under an adaptively chosen massage and identity attack provided that $q_S + q_E \le q$, $\varepsilon' \ge \varepsilon - 2q_S/(p-1)$, and $t' = t + O((5q_S + 4)T)$, where T is the maximum time for an exponentiation in $\mathbb{G}_1$, $\mathbb{G}_2$, and $\mathbb{Z}p$.*

Due to limited space, we omit the proof of Theorem 4.

According to Theorem 4 and Proposition 1, for the $\mathrm{S}_{\texttt{CH, SDH}}$ based identity-based signature, we get the following result.

Corollary 4. *Suppose SDH assumption holds in bilinear groups* $(\mathbb{G}_1, \mathbb{G}_2)$*. Then the identity-based signature above is tightly secure against existential forgeries under adaptively chosen massage and identity attacks in standard model.*

5 Conclusion

In this paper, according to the fact whether bilinear groups $(\mathbb{G}_1, \mathbb{G}_2)$ have an efficiently computable homomorphism, we give two IBS schemes, which are existentially unforgeable under adaptively chosen message and identity attacks and whose security is tightly related to q-SDH in the standard model, based on $\mathrm{S}_{\texttt{CMB, SDH}}$ proposed by Sven Schäge [36,37]. And then, we apply the idea constructing IBS schemes above to the $\mathrm{S}_{\texttt{CH, SDH}}$ by Sven Schäge [36,37], we also get IBS schemes which are existentially unforgeable under adaptively chosen message and identity attacks and whose security is also tightly related to q-SDH in the standard model.

Acknowledgments

The authors gratefully acknowledge the anonymous reviewers for their valuable comments. This study is supported by the Fundamental Research Funds for the Central Universities.

References

[1] M. Abe and T. Okamoto, "A signature scheme with message recovery as secure as discrete logarithm," in *Advances in Cryptology (ASIACRYPT'99)*, pp. 378–389, 1999.

[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)*, pp. 399–416, 1996.

[3] D. J. Bernstein, "Proving tight security for rabin-williams signatures," in *Advances in Cryptology (EUROCRYPT'08)*, pp. 70–87, 2008.

[4] R. Bhaskar, S. Garg and S. V. Lokam, "Improved bounds on security reductions for discrete log based signatures," in *Advances in Cryptology (CRYPTO'08)*, pp. 93–107, 2008.

[5] D. Boneh and X. Boyen, "Short signatures without random oracles and the sdh assumption in bilinear groups," in *Advances in Cryptology (EUROCRYPT'04)*, pp. 56–73, 2004.

[6] D. Boneh and X. Boyen, "Short signatures without random oracles and the sdh assumption in bilinear groups," *Journal of Cryptology*, vol. 21, no. 2, pp. 149–177, 2008.

[7] H. Chen, Y. Li, "Efficient identity-based signature scheme with partial message recovery," in *Proceedings of the Eighth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing (SNPD'07)*, pp. 883–888, 2007.

[8] H. Chen, Z. Wang, "A practical identity-based signature scheme from bilinear map," in *Emerging Directions in Embedded and Ubiquitous Computing (EUC'07)*, pp. 704–715, 2007.

[9] J. H. Cheon J. C. Cha, "An identity-based signature from gap diffie-hellman groups," in *Public Key Cryptography (PKC'03)*, pp. 18–30, 2003.

[10] J. S. Coron, "Optimal security proofs for PSS and other signature schemes," in *Advances in Cryptology (EUROCRYPT'02)*, pp. 272–287, 2002.

[11] Y. Dodis and L. Reyzin, "On the power of claw-free permutations," in *Third International Conference on Security in Communication Networks*, pp. 55–73, 2002.

[12] N. Fleischhacker, T. Jager, D. Schröder, "On tight security proofs for schnorr signatures," in *Advances in Cryptology (ASIACRYPT'14)*, pp. 512–531, 2014.

[13] F. D. Garcia, D. Galindo, "A schnorr-like lightweight identity-based signature scheme," in *International Conference on Cryptology in Africa*, pp. 135–148, 2009.

[14] E. J. Goh and S. Jarecki, "A signature scheme as secure as the diffie-hellman problem," in *Advances in Cryptology (EUROCRYPT'03)*, pp. 401–415, 2003.

[15] S. H. Heng, K. Kurosawa, "From digital signature to id-based identification/signature," in *Public Key Cryptography (PKC'04)*, pp. 248–261, 2004.

[16] F. Hess, "Efficient identity based signature schemes based on pairings," in *9th Annual International Workshop on Selected Areas in Cryptography*, pp. 310–324, 2002.

[17] M. S. Hwang, J. W. Lo, S. C. Lin, "An efficient user identification scheme based on ID-based cryptosystem", *Computer Standards & Interfaces*, vol. 26, no. 6, pp. 565–569, 2004.

[18] J. Katz and N. Wang, "Efficiency improvements for signature schemes with tight security reductions," in *Proceedings of the 10th ACM Conference on Computer and Communications Security (CCS'03)*, pp. 155–164, 2003.

[19] J. Katz, N. Wang, E. J. Goh, S. Jarecki, "Efficient signature schemes with tight reductions to the diffie-hellman problems," *Journal of Cryptology*, vol. 20, no. 4, pp. 493–514, 2007.

[20] B. Libert and J. J. Quisquater, "The exact security of an identity based signature and its applications," in *Iacr Cryptology Eprint Archive*, 2004. (`http://eprint.iacr.org/2004/102`)

[21] Z. H. Liu, H. Y. Chen, S. W. Lu, "Identity-based signature scheme with partial message recovery," *Chinese Journal of Computers*, vol. 29, no. 9, pp. 1622–1627, 2006.

[22] B. Lynn, D. Boneh and H. Shacham, "Short signatures from the weil pairing," *Journal of Cryptology*, vol. 17, no. 4, pp. 297–319, 2004.

[23] V. Lyubashevsky, M. Abdalla, P. A. Fouque and M. Tibouchi, "Tightly-secure signatures from lossy identification schemes," in *Advances in Cryptology (EUROCRYPT'12)*, pp. 572–590, 2012.

[24] B. C. Mames, "An efficient CDH-based signature scheme with a tight security reduction," in *Advances in Cryptology (CRYPTO'05)*, pp. 511–526, 2005.

[25] B. C. Mames and M. Joye, "A practical and tightly secure signature scheme without hash function," in *Cryptographers' Track at the RSA Conference (CT-RSA'07)*, pp. 339–356, 2007.

[26] J. Mao, J. Zhang, "A novel id-based designated verifier signature scheme," *Information Sciences*, vol. 178, no. 3, pp. 766–773, 2008.

[27] S. Micali and L. Reyzin, "Improving the exact security of digital signature schemes," *Journal of Cryptology*, vol. 15, no. 1, pp. 1–18, 2002.

[28] Y. Mu, F. Zhang, W. Susilo, "Identity-based partial message recovery signatures (or how to shorten id-based signatures)," in *The 9th International Conference on Financial Cryptography and Data Security (FC'05)*, pp. 45–56, 2005.

[29] C. Namprempre, M. Bellare and G. Neven, "Security proofs for identity-based identification and signature schemes," in *Advances in Cryptology (EUROCRYPT'04)*, pp. 268–286, 2004.

[30] T. Okamoto, R. Tso, C. Gu and E. Okamoto, "Efficient id-based digital signatures with message recovery," in *Cryptology and Network Security (CANS'07)*, pp. 47–59, 2007.

[31] P. Paillier and D. Vergnaud, "Discrete-log-based signatures may not be equivalent to discrete log," in *Advances in Cryptology (ASIACRYPT'05)*, pp. 1–20, 2005.

[32] K. G. Paterson and J. C. N. Schuldt, "Efficient identity-based signatures secure in the standard model," in *Australasian Conference on Information Security and Privacy*, pp. 207–222, 2006.

[33] K. G. Paterson, N. P. Smart, S. D. Galbraith, "Pairings for cryptographers," *Discrete Applied Mathematics*, vol. 156, pp. 3113–3121, 2008.

[34] N. M. Paulo, S. L. M. Barreto, B. Libert and J. J. Quisquater, "Efficient and provably-secure identity-based signatures and signcryption from bilinear maps," in *International Conference on the Theory and Application of Cryptology and Information Security*, pp. 515–532, 2005.

[35] D. Pointcheval and J. Stern, "Security arguments for digital signatures and blind signatures," *Journal of Cryptology*, vol. 13, no. 3, pp. 361–396, 2000.

[36] S. Schäge, "Tight proofs for signature schemes without random oracles," in *Advances in Cryptology (EUROCRYPT'11)*, pp. 189–206, 2011.

[37] S. Schäge, "Tight security for signature schemes without random oracles," *Journal of Cryptology*, vol. 28, no. 3, pp. 641–670, 2015.

[38] Y. Seurin, "On the exact security of schnorr-type signatures in the random oracle model," in *Advances in Cryptology (EUROCRYPT'12)*, pp. 554–571, 2013.

[39] A. Shamir, "Identity-based cryptosystems and signature schemes," in *Cryptology (CRYPTO'84)*, pp. 176–180, June 2008.

[40] W. Susilo, F. Zhang, R. S. Naini, "An efficient signature scheme from bilinear pairings and its applications," in *Public Key Cryptography (PKC'04)*, pp. 277–290, 2004.

[41] R. Yanli and G. Dawu, "Efficient identity based signature/signcryption scheme in the standard model," in *The First International Symposium on Data, Privacy and E-Commerce (ISDPE'07)*, pp. 133–137, 2007.

[42] X. Yi, "An identity-based signature scheme from the weil pairing," *IEEE Communications Letters*, vol. 7, no. 2, pp. 76–78, 2003.

[43] S. Zheng, Y. Yang, Z. Wang, L. Wang and Z. Hu, "Provably secure and efficient identity-based signature scheme based on cubic residues," *International Journal of Network Security*, vol. 14, no. 1, pp. 33–38, 2012.

Biography

Chen Huiyan received his PhD degree from Graduate University of Chinese Academy of Sciences in 2007. His research interests include cryptography, information security, and cloud computing.

Zhang Yanshuo received his PhD degree from Academy of Mathematics and Systems Science of the Chinese Academy of Sciences in 2009. His research interests include cryptography, information security.

Wan Zongjie received his M.S. degree from Beijing University of Post and Telecommunications in 2008. His research interests include cryptography, network security, and cloud computing.

Zhang Chenchen is currently a master degree candidate in the School of information and Communication Engineering,Xidian University. His research interests are information security and cryptography.

Cryptanalysis of an ID-based Deniable Threshold Ring Authentication

Tzu-Chun Lin[1], Ting-Yi Yeh[1], Min-Shiang Hwang[2,3]

(Corresponding author: Min-Shiang Hwang)

Department of Applied Mathematics, Feng Chia University[1]
100, Wenhwa Road, Taichung 40724, Taiwan, R.O.C.
Department of Computer Science & Information Engineering, Asia University[2]
500, Lioufeng Rd., Wufeng, Taichung 41354, Taiwan
Department of Medical Research, China Medical University Hospital, China Medical University, Taiwan[3]
(Email: mshwang@asia.edu.tw)

(Received Aug. 21, 2018; revised and accepted Dec. 15, 2018)

Abstract

In this paper, we offer analysis of a deniable threshold ring authentication protocol proposed by Jin *et al.* [5]. The Authors in [5] combined the two concepts, namely, threshold ring signature and deniable authentication, to propose a non-interactive deniable (t, n)- threshold ring authentication protocol. The protocol is the first design of this type. We will point out that this protocol cannot guarantee that at least t legal participates generate a valid signature and cannot withstand message modification attacks under certain restrictions.

Keywords: Bilinear Pairing; Deniable Authentication; Ring Signature; Threshold Signature

1 Introduction

The development of digital signature schemes for different needs is one of most important research topics in cryptography. A (t, n)-threshold signature allows any subset of t or more than t out of n participants to generate a valid signature [4]. Since 1991, Desmedt and Frankel [1] presented the concept of a (t, n)-threshold signature, which has been studied widely. A ring signature allows members of a signature group to anonymously sign messages on behalf of the group, and there is no group manager, group setup process and undo mechanism.

A deniable authentication protocol only allows not the message receiver to verify that the received message is indeed the one sent by the sender, but also ensures the sender's privacy so that this means that the following two special characteristics have to satisfy:

1) It enables an intended receiver to identify the source of a given message.

2) The intended receiver cannot prove the identity of the sender to any third party, even if he/she fully cooperates with the third party.

Due to the above two characteristics, the deniable authentication protocol can be broadly used for online shopping, electronic voting systems [9], e-learning systems and negotiation over Internet, *etc.* The concept of interactive deniable authentication protocol was initially introduced by Dwork *et al.* [2]. Since then, many interactive and non-interactive deniable authentication protocols based on various mathematical theories have been developed [6, 8, 10–15, 17, 18, 20, 24, 25]. The concept of deniable ring authentication was first introduced by Noar. This scheme can be extended to generate threshold authentication [17]. However, Noar's scheme requires an interactive zero knowledge protocol. In 2004, Susilo and Mu [22] proposed a non-interactive deniable ring authentication protocol based on Chamelon hash function and bilinear pairing.

In 2015, Jin *et al.* [5] combined the two concepts to propose a non-interactive deniable threshold ring authentication protocol (called IBDTRA) generated by elliptic curves and bilinear pairing [23]. This is the first design of this type. On the surface, all legal signers are involved in the signature. Actually, only t legal signers have used their own private keys to sign. However, the verifier can only verify that the signature was calculated by the group consisting of legitimate signers. The verifier would not know which participants have jointly calculated the signature.

In this article, we will point out that the IBDTRA cannot guarantee that at least t legal participates generate a valid signature and cannot withstand message modification attacks under certain restrictions.

2 Review of IBDTRA Protocol

2.1 Preliminary

Let G_1 be an additive group of prime order q, G_2 be a multiplicative group with the same order q. In the paper, a bilinear map $e : G_1 \times G_1 \longrightarrow G_2$ has to satisfy the following properties:

1) Bilinearity: $e(aP, bQ) = e(P, Q)^{ab}$, where $P, Q \in G_1$, $a, b \in \mathbb{Z}_q^*$.

2) Non-degeneracy: If $e(P, Q) = 1_{G_2}$ for all $Q \in G_1$, then $P = 1_{G_1}$.

3) Computability: There is an efficient algorithm to compute $e(P, Q)$ for any $P, Q \in G_1$.

Note that if a bilinear pairing is used on elliptic curves, then, usually, the Weil pairing or Tate pairing is recommended.

2.2 The IBDTRA Protocol

In this subsection, we will review Jin *et al.*'s IBDTRA protocol [5] . Let $G_1 =< P >$ be an additive group of prime order q, G_2 be a multiplicative group with the same order q and $e : G_1 \times G_1 \longrightarrow G_2$ be a bilinear map. Let $H_1 : \{0,1\}^* \longrightarrow G_1$ and $H_2 : \{0,1\}^* \longrightarrow \mathbb{Z}_q^*$ be two one-way hash functions. Suppose that there are n legitimate signers and one verifier (receiver). A message m requires at least $t(< n)$ participants to sign. Assumed that each legitimate signer $Sig(i)$ and the verifier have their own identity ID_i, $i = 1, \cdots, n$ and ID_r, respectively.

Setup. PKG constructs a pair of keys: (private key, public key) = (s, P_{pub}), where $s \in \mathbb{Z}_q$ and $P_{pub} := sP$. PKG publishes system parameters are $\{G_1, G_2, q, P, P_{pub}, e, H_1, H_2\}$.

Key Generation. The PKG computes private key $S_i = sQ_i$ for each user, where $Q_i = H_1(ID_i)$, $i = 1, \cdots, n, r$.

Signature. Without loss of generality, suppose the $Sig(1), \cdots, Sig(t)$ are the participating signers and $Sig(1)$ prepares the authenticate on behalf of other participants $Sig(t + 1), \cdots, Sig(n)$.

1) $Sig(1)$ chooses $x_i, h_i \in \mathbb{Z}_q$ as so-called private key for each $Sig(i)$, $i = t + 1, \cdots, n$, and computes the public keys

$$U_i = x_iP - h_iQ_i,$$
$$V_i = x_iP_{pub}.$$

2) For $j = 1, \cdots, t$, each $Sig(j)$ chooses the private key $r_j \in \mathbb{Z}_q^*$ and computes the public key

$$U_j = r_jP.$$

3) $Sig(1)$ computes the hash value

$$h_0 = H_2(\{ID_i\}_{i=1}^n, t, m, ID_r, \{U_k\}_{k=1}^n)$$

and then constructs a polynomial $f(x) \in \mathbb{Z}_q[x]$ of degree $n - t$ with the constant term h_0.

4) For $j = 1, \cdots, t$, each $Sig(j)$ uses the value $h_j := f(j)$ to compute another key

$$V_j = r_jP_{pub} + h_jS_j.$$

5) Anyone can calculate the values

$$V = \sum_{k=1}^n V_k \quad \text{and} \quad W = e(V, Q_r).$$

6) Also, $\sigma = \{\{U_k\}_{k=1}^n, W, f(x)\}$ is the deniable authentication for the message m.

Verification. After receiving (σ, m), the verifier checks whether

$$h_0 \overset{?}{=} H_2(\{ID_i\}_{i=1}^n, t, m, ID_r, \{U_k\}_{k=1}^n).$$

If this is the case, then computes $h_k = f(k)$ for each $k = 1, \cdots, n$, and

$$e(\sum_{k=1}^n (U_k + h_kQ_k), S_r).$$

If

$$e(\sum_{k=1}^n (U_k + h_kQ_k), S_r) = W,$$

then the massage m is accepted.

Figure 1 shows that the participating signers have different algorithms than the other ring participants in making the authentication of a message.

3 Cryptanalysis of IBDTRA Protocol

In this section, we show that there are two weaknesses in Jin *et al.*'s IBDTRA protocol [5].

Theorem 1. *In IBDTRA protocol, the verifier cannot confirm if at least t members are participating in this signature.*

Proof. Assume that only $t - 1$ members, say $Sig(1), \cdots, Sig(t - 1)$, agree to sign. $Sig(1)$ can construct (U_t, V_t) for the absent $Sig(t)$: First, $Sig(1)$ chooses an integer $r_t \in \mathbb{Z}_q$ and computes

$$U_t = r_tP.$$

$Sig(1)$ computes the hash value

$$h_0 = H_2(\{ID_i\}_{i=1}^n, t, m, ID_r, \{U_k\}_{k=1}^n)$$

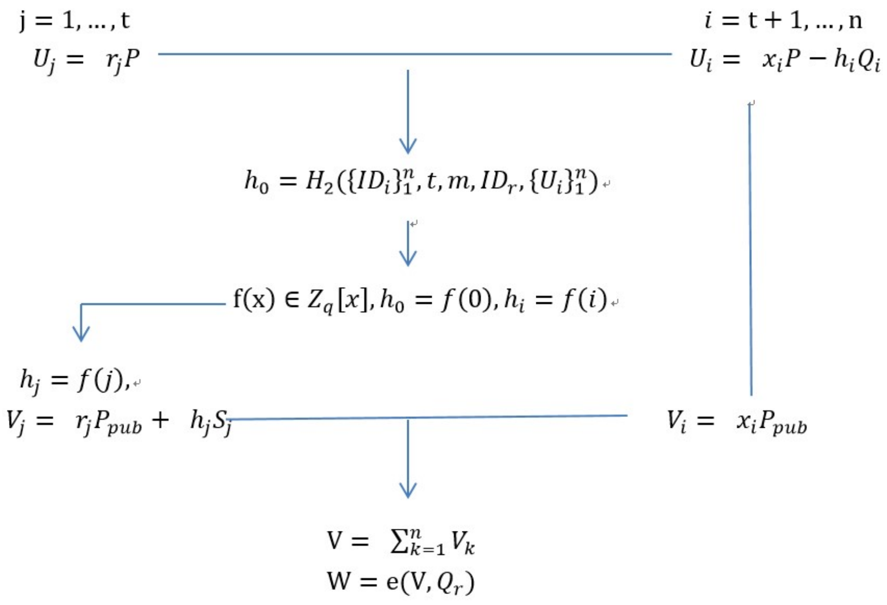


Figure 1: The phase of (t, n)-threshold authentication

and then construces a polynomial $f(x) \in \mathbb{Z}_q[x]$ of degree $n - t$ with the constant term h_0. In order to obtain V_t, $Sig(1)$ needs to find a pair of integers (x_t, y_t) such that

$$r_t \equiv x_t - h_t y_t \bmod q \tag{1}$$
$$Q_t = y_t P. \tag{2}$$

Thus, $V_t := x_t P_{pub}$. □

One of the ways to solve the equation

$$r_t = x_t - h_t y_t$$

in $\mathbb{Z}_q$ relies on the following idea: we choose randomly a pair (x', y') and let $s = x' - h_t y'$. Then $1 = s^{-1}x' - h_t s^{-1} y'$, thus $r_t = s^{-1}x' r_t - h_t s^{-1} r_t y'$. Next, we check whether $Q_t = (s^{-1} r_t y')P$. If it is not the case, then go back to the choice of (x', y'). If G_1 is a subgroup of an elliptic curve over $\mathbb{Z}_q$. To determine an integer y such that $Q = yP$ is just like to solve the ECDLP (Elliptic Curve Discrete Logarithm Problem).

Another way to find the solution (x_t, y_t): assume that (x, y) is a solution of the equation $U_t = xP - h_t Q^*$, where $Q^* = yP$. If $Q^* \neq Q_t$, we consider the point $U^* := xP - h_t Q_t$. Let $aP = U^* - U_t$. Solving the ECDLP is one of the most frequently studied topics in cryptanalysis. So far, there is no known attack of polynomial time for the 160-bit ECDLP [3, 16, 19]. In fact, the difficulty of the ECDLP depends not only on the bit-length of the order $o(P) = q$ but also on the bit-length of the scalar a. In order to invalidate attacks, the bit-length of the selected scalar approximates the bit-length of the prime order q. For this reason, we choose a solution y whose bit-length is approximately the bit-length of the prime order q. If the bit-length of the integer of the subtraction $a \equiv y_t - y \bmod q$ is short enough, then the value of a can be calculated in polynomial time by using exhaustive search. If the value of the integer a is found, then the private key $y_t = y - ah_t^{-1}$ is solved, where $Q_t = y_t P$. Next question, is it easy to find a pair of solution (x, y) such that the bit-length of the scalar a is short enough? In other word, how likely is it to prevent IBDTRA protocol from such attack? Such questions are still open. The following attack uses a similar approach.

Theorem 2. *If $t \leq n/2$, then the IBDTRA protocol is unable to resist the message modify attack.*

Proof. Assume that an eavesdropper intercepts the authenticated message (σ, m) and wants to use the fake message m^* to replace the real message m.

Since $H_2(\cdots, m, \cdots) \neq H_2(\cdots, m^*, \cdots)$, in order for the fake message m^* to pass verification, the eavesdropper must be able to create a new polynomial $g(x) \in \mathbb{Z}_q[x]$ such that the following conditions are satisfied:

1) $h_0^* = g(0)$, where the constant term $h_0^* = H_2(\{ID_i\}_{i=1}^n, t, m^*, ID_r, \{U_k\}_{k=1}^n)$.

2) $\deg g(x) = n - t$.

3) $h_k = g(k)$, $k = 1, \cdots, t$.

Such polynomial $g(x)$ is defined by

$$g(x) := h_0^* + (h_1 - h_0^*)x^l \prod_{i=1, i\neq 1}^{t} \frac{x-i}{1-i}$$
$$+ (h_2 - h_0^*)\frac{x}{2} \prod_{i=1, i\neq 2}^{t} \frac{x-i}{2-i} + \cdots$$
$$+ (h_t - h_0^*)\frac{x}{t} \prod_{i=1}^{t-1} \frac{x-i}{t-i},$$

where $l \geq 0$ is an integer with $t + l = n - t$.

Then, for $i = t + 1, \cdots, n$, set $h_i^* := g(i)$. To find $x_i^* \in \mathbb{Z}_q$ satisfying $U_i = x_i^* P - h_i^* Q_i$, the eavesdropper computes $U_i^* = x_i P - h_i^* Q_i$ and $U_i - U_i^*$. If $U_i - U_i^* = a_i P$ and a_i is a small integer, then the integer a_i can be effectively calculated, and also $x_i^* = a_i + x_i$ and $V_i^* = x_i^* P_{pub}$. □

4 Conclusion

The verifier in IBDTRA protocol uses his/her own private key and public keys of all legal signers together with a bilinear pairing to verify an authenticated message, so he/she can only prove whether a given message is from a legitimate group. The verifier could not know the list of signers who have actually participated in the signature. In this paper we provide a possible way to accomplish the challenge of sender spoofing and modifying messages: First, we give a solution (x, y) of Equation (1) and let $U^* - U_t = aP$ be an element of the cyclic group $< P >$; under the premise that the integer a is small, then the integer a can be effectively calculated by using exhaustive search, so the true solution (x_t, y_t) that satisfies both Equations (1) and (2) is found. Although IBDTRA protocol is not immediately dangerous, due to increased computing power and ongoing research on ECDLP [16,19,21], such attacks still seem to pose a threat to it.

Acknowledgment

This research was partially supported by the Ministry of Science and Technology, Taiwan (ROC), under contract no.: MOST 104-2221-E-468-004 and MOST 105-2410-H-468-009. MOST 106-2221-E-468-002.

References

[1] Y. Desmedt, Y. Frankel, "Shared generation of authentications and Signatures", in *Advances in Cryptology (CRYPTO'91)*, pp. 457–469, 1991.

[2] C. Dwork, M. Naor, A. Sahsi, "Concurrent zero-knowledge", in *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing (STOC'98)*, pp. 409–418, 1998.

[3] J. Hoffstein, J. Pipher, J. H. Silverman, *An Introduction to Mathematical Cryptography*, Springer-Verlag, New York, 2008.

[4] M. S. Hwang and T. Y. Chang, "Threshold signatures: Current status and key issues", *International Journal of Network Security*, vol. 1, no. 3, pp. 123–137, 2005.

[5] C. Jin, C. Xu, L. Jiang, "ID-based deniable threshold ring authentication," in *IEEE 17th International Conference on High Performance Computing and Communications (HPCC'15), IEEE 7th International Symposium on Cyberspace Safety and Security (CSS'15), and IEEE 12th International Conf on Embedded Software and Systems (ICESS'15)*, pp. 1779–1784, 2015.

[6] J. Kar, "ID-based deniable authentication protocol based Diffie-Hellman problem on elliptic curve," *International Journal of Network Security*, vol. 15, no. 5, pp. 357–364, 2013,

[7] H. Krawczyk, T. Rabin, *Chameleon Hashing and Signatures*, 1997.

[8] W. B. Lee, C. C. Wu, W. J. Tsaur, "A novel deniable authentication protocol using generalized ElGamal signature scheme", *Information Sciences*, vol. 177, no. 6, 1376–1381, 2007.

[9] C. T. Li, M. S. Hwang, C. Y. Liu, "An electronic voting protocol with deniable authentication for mobile Ad hoc networks," *Computer Communications*, vol. 31, no. 10, pp. 2534–2540, June 2008.

[10] F. Li, P. Xiong, C. Jin, "ID-based deniable authentication for Ad hoc networks", *Computing*, vol. 96, no. 9, pp. 843–853, 2014.

[11] F. Li, D. Zhong, T. Takagi, "Efficient deniable authenticationed encryption and its application to E-mail," *IEEE Transactions on Inforensics and Security*, vol. 11, no. 11, pp. 2477–2486, 2016.

[12] C. C. Lin, C. C. Chang, "An improved deniable authentication protocol," *International Journal of Computer Science and Network Secruity*, vol. 6, no. 11, pp. 240–242, 2006.

[13] T. C. Lin, "Improvement of an ID-based deniable authentication protocol," *Journal of Electronic Science and Technology*, vol. 16, no. 2, pp. 139–144, 2018.

[14] C. Y. Liu, C. C. Lee, T. C. Lin, "Cryptanalysis of a deniable authentication protocol based on generalized ElGamal signature scheme," *International Journal of Newtork Security*, vol. 12, no. 1, pp. 58–60, 2011.

[15] R. Lu, X. Lin, Z. Cao, L. Qin, X. Liang, "A simple deniable authentication protocol based on the Diffie-Hellman algorithm," *International Journal of Computer Mathematics*, vol. 85, no. 9, pp. 1315–1323, 2008.

[16] S. Miyoshi, Y. Nogami, T. Kusaka, N. Yamai, "Solving 94-bit ECDLP with 70 computers in parallel," *International Journal of Computer and Information Engineering*, vol. 9, no. 8, pp. 1966–1969, 2015.

[17] M. Naor, "Deniable ring authentication," *Advances in Cryptology (Crypto'01)*, LNCS, vol. 2442, pp. 481–498, 2002.

[18] M. D. Raimondo, R. Gennaro, "New approaches for deniable authentication," *Journal of Cryptology*, vol. 22, pp. 572–615, 2009.

[19] O. Schwarz, *General Attacks on Elliptic Curve Based Cryptosystems*, Project Report in Winter 2012-2013, Project Advisor: Barukh Ziv.

[20] Z. Shao, "Efficient deniable authentication protocol based on generalized ElGamal signature scheme," *Computer Standards & Interfaces*, vol. 26, no. 5, pp. 449–454, 2004.

[21] K. Somsuk, C. Sanemueang, "The new modified methodology to solve ECDLP based on brute force attack," *Recent Advances in Information and Communication Technology*, pp. 255–264, Springer, 2018.

[22] W. Susilo, Y. Mu, "Non-interactive deniable ring authentication," in *International Conference on Applied Cryptography and Network Security (ACNS'04)*, LNCS, vol. 3089, pp. 149–163, Springer, 2004.

[23] S. F. Tzeng, M. S. Hwang, "Digital signature with message recovery and its variants based on elliptic curve discrete logarithm problem", *Computer Standards & Interfaces*, vol. 26, no. 2, pp. 61–71, Mar. 2004.

[24] E. J. Yoon, "Security analaysis of Kar's ID-based deniable authentication protocol," *Contemporary Engineering Sciences*, vol. 8, no. 17, pp. 765–771, 2015.

[25] E. J. Yoon, E. K. Ryu, K. Y. Yoo, "Improvement of Fan *et al.*'s deniable authentication protocol based on Diffie-Hellman algorithm," *Applied Mathematics and Computation*, vol. 167, pp. 274–280, 2005.

Biography

Tzu-Chun Lin received the PhD in Mathematics from the Faculties for Mathematics and Science of the Georg-August-University at Göttingen in Germany. She is an associate professor in the Department of Applied Mathematics of Feng Chia University in Taiwan, R.O.C.. Her current research interests include commutative algebras, invariant theory of finite groups and public-key cryptography.

Ting-Yi Yeh received his B.S. degree from the Department of Applied Mathematics at Feng Chia University, Taiwan, ROC. His research interests on information security and public key cryptography.

Min-Shiang Hwang received M.S. in industrial engineering from National Tsing Hua University, Taiwan in 1988; and Ph.D. degree in computer and information science from National Chiao Tung University, Taiwan in 1995. He was a professor and Chairman of the Department of Management Information Systems, NCHU, during 2003-2009. He was also a visiting professor with University of California (UC), Riverside and UC. Davis (USA) during 2009-2010. He was a distinguished professor of Department of Management Information Systems, NCHU, during 2007-2011. He obtained the 1997, 1998, 1999, 2000, and 2001 Excellent Research Award of National Science Council (Taiwan). Dr. Hwang was a dean of College of Computer Science, Asia University (AU), Taichung, Taiwan. He is currently a chair professor with Department of Computer Science and Information Engineering, AU. His current research interests include information security, electronic commerce, database and data security, cryptography, image compression, and mobile computing. Dr. Hwang has published over 300+ articles on the above research fields in international journals.

A Secure and Efficient Computation Outsourcing Scheme for Multi-Users

V. Sudarsan Rao[1] and N. Satyanarayana[2]

(Corresponding author: V. Sudarsan Rao)

Department of CSE, Khammam Institute of Technology and Sciences (KITS)[1]
Telangana, India
Department of CSE, Nagole Institute of Technology and Sciences (NITS)[2]
Telangana, India
(Email: sudharshan.cse2008@gmail.com)

(Received Oct. 9, 2017; revised and accepted Apr. 13, 2018)

Abstract

The outsourcing process is computationally secure if it is performed without unveiling to the other external agent or cloud, either the original data or the actual solution to the computations. Secure multiparty computation computes a certain function without revealing their private secret information. In this paper, a new kind of outsourcing computing protocol is proposed which utilizes multi cloud servers view framework. This paper mainly adopts the Fully Homomorphic Encryption technique (FHE). In our proposed protocol, encrypted data by different users is transformed to cloud. The protocol being non-interactive between users, gives the comparatively lesser computational and communication complexity. The analysis of our proposed protocol is also presented at the end of the paper.

Keywords: Access Control; Circuit Computation; Cloud Computing; Privacy; Secret Information Parameters; Secure Outsourcing

1 Introduction

Beside the tremendous advantages of outsourcing, client faces some challenges by outsourcing the computational task to cloud [2, 3]. These are security, input-output privacy and verification of result. Consider a scenario where some mutually distrusted members are present, and they want to compute a complex function, which involves their own private inputs [6]. This scenario may be termed as secure multi-party computation. Suppose, $U_1, U_2, \cdots, U_m$ are m users, and each posses a private number $n_1, n_2, \cdots, n_m$. Consider function is,

$$\mathbb{FUNC} = f(n_1, n_2, \cdots, n_m),$$

which they want to co-operatively compute, but they don't want to expose n_i of corresponding U_i to other users U_j, $i \neq j$ & $i, j \in (1, 2, \cdots, m)$. Also they should guarantee that $\mathbb{FUNC}$ should not be known by any of the unauthorized user. Its observable that the computation and communication complexities are mostly dependant on the complex nature of computation function. The scenario is shown as Figure 1.

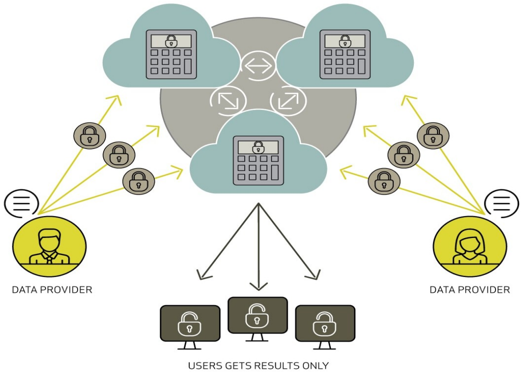


Figure 1: General computational outsourcing scenario

Recently, as the development of cloud computing [33], users' concerns about data security are the main obstacles that impedes cloud computing from wide adoption. These concerns are originated from the fact that sensitive data resides in public cloud [31], which is maintained and operated by untrusted cloud service provider (CSP) [21, 29]. The expectation of users is that the cloud should compute the function having the inputs as private parameters of users in the encrypted/transformed form.

Remaining paper organized as - Section 2 provides a general nomenclature for various secure outsourcing algorithms. Significant state-of-the-art protocols along with the motivation towards the problem and our contribution in this paper is summarized. Preliminaries are given in Section 3. Secure outsourcing using FHE scheme is given in Section 4. Experimental analysis are presented in Section 5. Section 6 presents our proposed scheme along with correctness, security analysis and our experimental simulation results. Section 7 concludes the paper.

2 Secure Outsourcing Algorithms Classification

Increasing no. of smart equipments and their growing need to execute computationally large task resulting the outsourcing of any scientific computation to the cloud server an encouraging solution. The general nomenclature is represented as Figure 2.

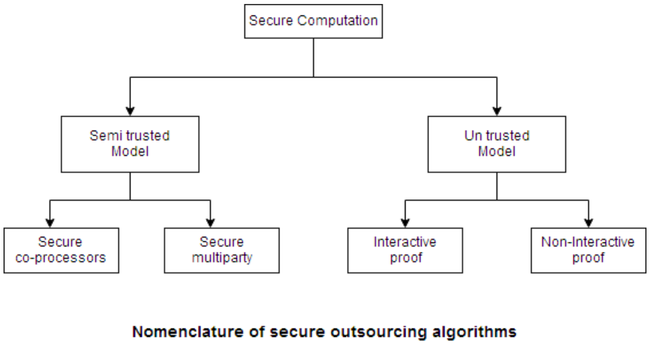


Figure 2: Secure outsourcing algorithms nomenclature

2.1 Related Work

While outsourcing the private data functions to the cloud, there exist many problems and challenges. In past years, much research have been carried out to come up with various solutions for secure computational outsourcing. One solution was proposed by Gentry [16], in 2009 where a joint public key is used to encrypt their private input data and accordingly the notion was termed as Homomorphic encryption, which successively used in the secure outsourcing of practical complex problems. In the work by [1], authors have given a scheme where for encryption purpose, users' public keys are utilized, and cloud will be able to compute the function having their private inputs. A more secure outsourcing was given by Halevi *et al.* [18] in 2011, that was a non-interactive method for secure outsourcing [8]. [23] given a new fully homomorphic scheme, multikey FHE, which applied bootstrapping concept for secure outsourcing of computations. ABE, introduced as fuzzy identity-based encryption in [25], was firstly dealt with by Goyal *et al.* [15]. Two distinct and interrelated notions of ABE were determined in [15]. Accordingly, several constructions supporting for any kinds of access structures were provided [13,24] for practical applications [19,32]. Atallah *et al.* [2] offered an structure for secure outsourcing of scientific computations e.g. multiplication of matrices. Although, the solution used the disguise technique and thus leaded to leakage of private information. Atallah and Li [3] given an efficient protocol to outsource sequence comparison with two servers in secure manner. Furthermore, Benjamin and Atallah [6] addressed the problem of secure outsourcing for widely applicable linear algebraic computations. Atallah and Frikken [4] further studied this problem and gave improved protocols based on the so-called weak secret hiding assumption. Recently, Wang *et al.* [28] presented efficient mechanisms for secure outsourcing of linear programming computation.

In [17], a novel paradigm for outsourcing the decryption of ABE is given. Compared with our work, the two lack of the consideration on the eliminating the overhead computation at attribute authority. In 2014, V.Sudarshan *et al.* [27] proposed an Attribute-Based Encryption mechanism, applied for cloud security. Recently Lai *et al.* [20] given a construction with verifiable decryption, which achieves both security and verifiability without random oracles. Their task supplements a redundancy with ciphertext and uses this redundancy for correctness checking. A.K.Chattopadhyay *et al.* [12] proposed the scheme which uses simple Boolean based encryption and decryption of the data files, which is low in computational cost. Yongjian L. *et al.* [22] given the ABE based construction of scheme where some invalid ciphertexts checking is performed in the decryption algorithm. Kai F. *et al.* [14] designed an efficient user and attribute revocation method. Along with this, analysis and simulation results for their scheme showed that it is both secure and highly efficient. Z. Cao *et al.* [11] remarked that Yu *et al.*'s scheme [30] has two inherent weaknesses:

1) It does not truly mitigate the client's computational burden for key updates.

2) It does not ensure confidentiality since the files uploaded to the cloud by the client are eventually not encrypted at all.

2.2 Motivation and Contribution

In the scenario of outsourcing private inputs or computational function to cloud, There exist hurdles in following two aspects - One is in the users' or customers point of view, where they want to ensure the privacy of its input parameters and results. Another is to cloud servers point of view, where cloud entity is worried about feasibleness of encrypted/transformed inputs and operating on them. In computational outsourcing, users are not participating in the computational function, rather than they outsource the private problem along with parameters to the cloud, but users and cloud servers are not mutually trusted entities. Thus, users would not like to submit their private problem data inputs to the cloud. Thus, encrypting/transforming the private data prior to submission to cloud is a usual solution.

Our contribution in this paper is as -

- We have proposed protocol for secure and an efficient computational outsourcing to cloud. The protocol is completely non-interactive between users.

- We have performed the computational security analysis for our proposed system.

3 Preliminaries

This section discusses some of the significant preliminaries required for secure computational outsourcing.

3.1 Lattice-Based Encryption

As we know that the computational complexity as well as the input parameters' privacy is mostly dependant on the encryption procedure adopted by user. Lattice-Based Encryption [9,10] is considered as secure against quantum computer attacks and much efficient as well as potent than RSA and Elliptic curve cryptosystems.

Lattice based cryptosystem, whose security is based on core lattice theory problems, was introduced by Miklos Ajtai, in 1996. In the same year, first lattice based public key encryption scheme (NTRU) was proposed. Later, much work and improvement [16] was carried out towards this direction involving some additional cryptographic primitives LWE (learning with errors).

3.2 Computational Verifiability

Various different solutions exist for secure computational outsourcing. Homomorphic encryption (HE) can be assumed as a better solution to secure outsourcing of scientific computations, but it is useful when the returned result can be trusted.

Lemma 1. *It is infeasible to factorizing the N in polynomial time if integer factorization in large scale in infeasible.*

Proof. Assume x is an adversary who is able to factorize a number N into primes p and q of probable same bit length in polynomial time. Suppose this operations probability as p'. Each factor $fact_i$ of a number N will at least posses two prime factors. So the probability p''_r that the attacker can factorize it is almost lesser than p'. Thus the resultant probability that attacker can factorize N is $\prod_{i=1}^{m} p''_r \leq (p')^m$. Now if p' is negligible, the resultant probability is also negligible. □

Definition 1. *A matrix $M \in R^{n,n}$ can be called as orthogonal if it is satisfying one of the equivalent conditions*

1) $M.M^T = M^T.M = I_n$;

2) M is invertible and $M^{-1} = M^T$.

4 Secure Outsourcing Using FHE

This section summarizes Sudarshan *et al.* scheme [26] for secure outsourcing of large matrix multiplication computations on cloud. First, the key space is being generated at client side, which will be utilized in further steps. Here, we have considered the scientific computation as 'large matrix multiplication', which the client needs to outsource to cloud server. Here the assumption taken is that the third party or cloud server is untrusted. Client needs to perform problem transformation step for secure outsourcing. Further, the computation inside cloud is performed. After getting the computed result, client will retransform it and get the original result for matrix multiplication problem. The complete description and steps involved in this scheme are summarized as below:

Algorithm 1 Secure Outsourcing using FHE

```
Begin
Generate secret key pair: ${H, Y}$
  where, $H$: is a Hadamard matrix [34] and
  $Y$: is a diagonal matrix selected randomly.
Consider, $M_1$ and $M_2$ are two large matrices, for which the multiplication needs to be computed, thus client will outsource this computation problem to cloud side.
Client computes,
  $M'_1 = H \times M_1 \times Y$
  $M'_2 = Y^{-1} \times M_2$
Client sends $M'_1$ and $M'_2$ to cloud server.
$Result' \leftarrow M'_1 \times M'_2$
The cloud server sends back the computed result to client side.
After getting the computed result, client will retransform it and get the original result for MM problem. The procedure is given as below Algorithm
$Result \leftarrow H^{-1} \times Result'$
End
```

5 Experimental Analysis

This section presents our experimental analysis.

5.1 System Specifications

Our system specifications are as below:

- *Software specifications*
  OS Ubuntu 16.04 LTS, 64 bit; Python version 'Python 3.6.0'.

- *Hardware specifications*
  RAM size 4 GB;
  Processor Intel core i3 4030U CPU @1.90$GHz \times 4$.

5.2 Our Results

We performed the experiments on varying sized secret key pair, arbitrary large sized matrices M_1 and M_2 as input. Problem parameters transformation/encryption, decryption and entire average execution time for executing the protocol is analysed. The graph for encryption phase for various sized input parameters is Figure 1.

The graph for decryption phase for various sized input parameters is in Figure 2.

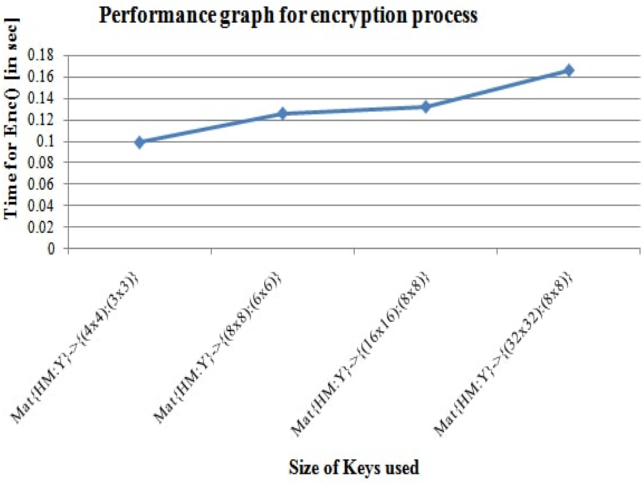


Figure 1: The performance for encryption phase

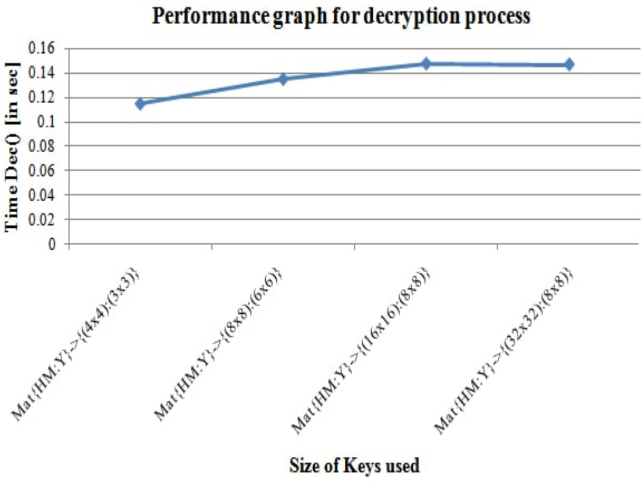


Figure 2: The performance for decryption phase

The graph for overall algorithm execution for various sized input parameters is in Figure 3.

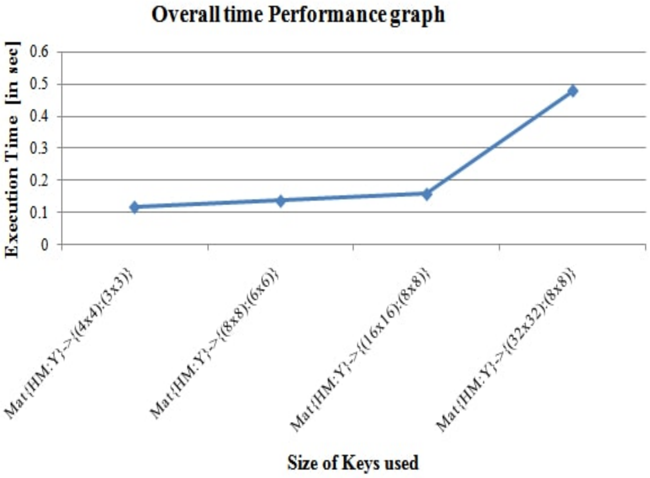
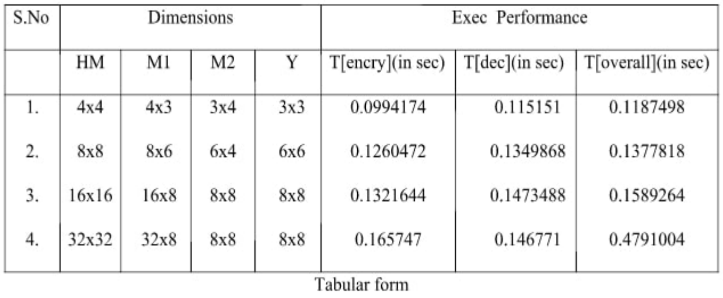


Figure 3: The performance for overall algorithm

In above Performance graphs 1-3, the encryption, decryption and overall execution time (in seconds) for varying experimental instances of secret key matrix pair size dimensions is shown. The end results of execution performance for varying key sizes is presented as Table 1.

Table 1: Execution performance



S.No	Dimensions				Exec Performance		
	HM	M1	M2	Y	T[encry](in sec)	T[dec](in sec)	T[overall](in sec)
1.	4x4	4x3	3x4	3x3	0.0994174	0.115151	0.1187498
2.	8x8	8x6	6x4	6x6	0.1260472	0.1349868	0.1377818
3.	16x16	16x8	8x8	8x8	0.1321644	0.1473488	0.1589264
4.	32x32	32x8	8x8	8x8	0.165747	0.146771	0.4791004

Tabular form

6 Proposed Scheme

In this section, we have proposed an efficient secure computational outsourcing mechanism applicable for multi-users. The system model and proposed mechanism steps are given in subsections below.

6.1 System Model

The proposed system model is represented as in Figure 4.

Notations used are given in Table 2.

Table 2: Notations used in proposed system

$CS1$:	First cloud server
$CS2$:	Second cloud server
c_i:	Ciphertexts (encrypted data of each customer/user U_i)
n:	No. of users
α_i:	Private input corresponding to U_i
ψ:	Probability density function
q:	Prime order
$RAND_i$:	Random number for i^{th} user
$\mathscr{C}_{FUN}$:	Function circuit
R:	Ring structure space
β:	Final computed result

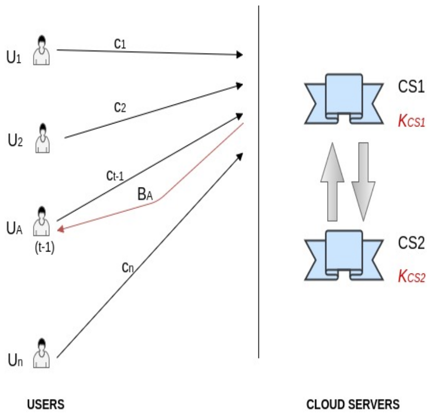


Figure 4: Proposed model

6.2 Protocol Steps

The proposed secure computational outsourcing protocol executes in the below phases

Algorithm 2 Key Gen() and Set up

```
Begin
Perform sampling for ring element space vector $a_i \leftarrow R_q^N$, $\forall\ i = (1, 2, \cdots, n)$; Ring element $SK_i \leftarrow \psi$; $\alpha_i \leftarrow \psi^N$ ($\psi$ represents: probability density function), where
    $$\psi = \int_{-\infty}^{x} P(\xi) d\xi$$
Key pairs of $U_i$: Public key $(a_i.SK_i + 2\alpha_i) \in R_q^N$; Private key $SK_i$.
$CS1$ has its private no. as $K_{CS1}$ & $CS2$ has its private no. as $K_{CS2}$.
$U_i$ shares a random no. $RAND_i$ with $CS1$.
Each user $U_i$ initiates protocol and sends $RAND_i.SK_i$ to $CS2$.
$CS2$ reckons $K_{CS2}.RAND_i.SK_i$ and sends back to $CS1$.
$CS1$ can get $K_{CS2}.SK_i$ by extracting $RAND_i$.
End
```

$\forall i \in (1, 2, \cdots, n)$, U_i uses Lattice based encryption method to encrypt its own problem input α_i. The sub-steps involved in this are as Algorithm-3.

Algorithm 3 Lattice based Encryption

```
Begin
First $U_i$ perform sampling as: $e_i \leftarrow \psi^N$. where,
    $\psi$ is: probability density function(PDF), defined as
    $$\psi = \int_{-\infty}^{x} P(\xi) d\xi$$
Next, each user $U_i$ computes
    $$c_0^i \leftarrow < u_i, e_i > + \alpha_i \in R_q$$
    $$c_1^i \leftarrow < a_i, e_i > \in R_q$$
Further, it gives output as ciphertext,
    $$c_i = (c_0^i, c_1^i) \in R_q^N; (N = 2)$$
End
```

$CS1$ stores all ciphertexts coming from user $U_i (1 \leq i \leq n)$, then further steps are given in Algorithm-4.

Production of the result by cloud servers will follow as steps of Algorithm-5.

6.3 Analysis of Proposed Scheme

Here, we have presented the correctness and security analysis of our proposed scheme.

- Correctness analysis:
  The correctness analysis of given scheme is as follows.

  Theorem 1. *Due to Homomorphic properties of the transformed ciphertexts, the given scheme is correct.*

  Let, P and Q are rings a function $f : P \to Q$ will be ring homomorphism if $\forall x_1, x_2 \in P$.

  – $f(x_1 + x_2) = f(x_1) + f(x_2)$;

  – $f(x_1 * x_2) = f(x_1) * f(x_2)$.

- Security analysis:
  The security analysis of proposed scheme can be analysed as below.

  Theorem 2. *As long as Lattice based encryption is secure and cloud servers CS1 and CS2 are noncolluding, the given protocol is secure enough.*

Algorithm 4 Circuit Computation on Outsourcing

```
Begin
First, $CS1$ transforms the ciphertexts as $c_i \rightarrow c_i^{TR_1}$ where, $c_i^{TR_1} = (c_0^{i_{TR_1}}, c_1^{i_{TR_1}}) = (K_{CS1}.c_0^i, K_{CS1}.(K_{CS2}.SK_i).c_1^i)$.
$CS1$ sends above $c_i^{TR_1}$ to $CS2$.
After receiving $c_i^{TR_1}$, $CS2$ again transforms $c_i^{TR_1}$ into
    $c_i^{TR_2} = (K_{CS2}.K_{CS1}.c_0^i, K_{CS1}.(K_{CS2}.SK_i).c_1^i)$
    take, $K = K_{CS1}.K_{CS2}$
    then, $c_i^{TR_2} = (c_0^{i_{TR_2}}, c_1^{i_{TR_2}}) = (K.c_0^i, K.SK_i.c_1^i)$
$CS2$ then reckons the ciphertext of result by transformed ciphertext of every user's private i/p.
Additive oprn. for each add. gate
    $\Rightarrow c_i^{TR_2} \bigoplus c_j^{TR_2}$
    $\Rightarrow (c_1^{i_{TR_2}} - c_0^{i_{TR_2}}) \bigoplus (c_1^{j_{TR_2}} - c_0^{j_{TR_2}})$
    $\Rightarrow (K.SK_i.c_1^i - K.c_0^i) \bigoplus (K.SK_j.c_1^j - K.c_0^j)$
    $\Rightarrow (K.(SK_i.c_1^i - c_0^i) \bigoplus K.(SK_j.c_1^j - c_0^j))$
    $\Rightarrow K.[(SK_i.c_1^i - c_0^i) \bigoplus (SK_j.c_1^j - c_0^j)]$
    $\Rightarrow K.[\alpha_i + \alpha_j]$
Multiplicative oprn. for every mul. gate -
    $\Rightarrow c_i^{TR_2} \bigotimes c_j^{TR_2}$
    $\Rightarrow (c_1^{i_{TR_2}} - c_0^{i_{TR_2}}) \bigotimes (c_1^{j_{TR_2}} - c_0^{j_{TR_2}})$
    $\Rightarrow (K.SK_i.c_1^i - K.c_0^i) \bigotimes (K.SK_j.c_1^j - K.c_0^j)$
    $\Rightarrow (K.(SK_i.c_1^i - c_0^i) \bigotimes K.(SK_j.c_1^j - c_0^j))$
    $\Rightarrow K^2.[(SK_i.c_1^i - c_0^i) \bigotimes (SK_j.c_1^j - c_0^j)]$
    $\Rightarrow K^2.[\alpha_i \times \alpha_j]$
End
```

Algorithm 5 Production of Result

```
Begin
When $CS2$ performed gate by gate computation on circuit $\mathscr{C}_{FUN}$, it gets some intermediate meta result, which is encrypted by $K_{CS1}$ and $K_{CS2}$ of the cloud servers $CS1$ and $CS2$.
    If $\beta = FUN(\alpha_1, \alpha_2, \cdots, \alpha_n)$ and let's $\theta$ is the no. of multiplicative gates of $\mathscr{C}_{FUN}$.
    then, $\beta' = K^{\theta+1}.\beta = (K_{CS1}^{\theta+1}.K_{CS2}^{\theta+1}).\beta$
To provide results for each user, and ensure that only authorized user set must get final result [*Assume,* $U_{\mathbb{A}}$, $\mathbb{A} \in (1, 2, 3, \cdots, n)$ *is authorized user set to access result*], $CS2$ first sends $\beta'$ to $CS1$.
$CS1$ removes $K_{CS1}^{\theta+1}$ and ties $RAND_{\mathbb{A}}$ to compute $\beta'_{\mathbb{A}} = RAND_{\mathbb{A}}.K_{CS2}^{\theta+1}.\beta$
Then $CS1$ sends $\beta'_{\mathbb{A}}$ to $CS2$.
$CS2$ finally removes $K_{CS2}^{\theta+1}$ and gets $\beta_{\mathbb{A}} = RAND_{\mathbb{A}}.\beta$
Further $CS2$ sends it to authorized users set $U_{\mathbb{A}}$, $\mathbb{A} \in (1, 2, 3, \cdots, n)$.
End
```

In proposed protocol, each user U_i encrypts its private input α_i with the help of its own public key, which is being produced by triggering lattice based encryption scheme. Further, U_i sends $RAND_i.SK_i$ to $CS2$. Then, $CS2$ reckons $K_{CS2}.RAND_i.SK_i$ and sends back to $CS1$. Here, $U_i's$ private key is SK_i, which is protected by $RAND_i$. In the entire process, the user's private keys are not being revealed.

Algorithm 6 Secure Results Reconstruction at Users' side

```
Begin
For each $U_{\mathbb{A}}$, $\mathbb{A} \in (1, 2, 3, \cdots, n)$, it successfully gets the final result $\beta$ by deposing $RAND_{\mathbb{A}}$.
End
```

After transfering computed results, cloud ensures in the protocol that only authorized user set must get final result; (Assume, $U_{\mathbb{A}}$, $\mathbb{A} \in (1, 2, 3, \cdots, n)$ is authorized user set to access result.)

6.4 Experimental Simulation Results

This section presents the simulation results on virtual cloud server. We used *Microsoft Azure cloud server* along with *Android Studio 3.0.1 software platform SDK* to perform simulation on client-cloud outsourcing model. *JDK 9.0.1* is used for backend interface development. The machine used for simulation has specifications as *OS: Windows 10*; *RAM size: 16 GB*; *NVIDIA GeForce GTX 1080 GPU* with *octa core i7 CPU.*

The experimental simulation results obtained are given in Table 3.

In Table 3, we tested our scheme through various client's private data size and key size parameters etc. The client's private data can be in any form like integer, image, text, pdf file, video file etc. We obtained the time taken for encrypt private data, the time taken for upload operation on Azure cloud server and time taken to download and decrypt data on client side. Inside cloud scientific computational operations are being performed on transformed parameters. The overall representation is given in Figure 5.

6.5 Comparative Analysis

This section presents the comparison of our scheme with existing schemes on several factors/parameters. The representation is given in Table 4.

7 Conclusion and Future Work

When users have to compute some complex function, which involves their private inputs then to perform outsourcing is the possible scenario from user side. There exist hurdles in following two aspects: One is in the users' or customers point of view, where they want to ensure the privacy of its input parameters and results. Another is to cloud servers point of view, where cloud entity is worried about feasibleness of encrypted/transformed inputs and operating on them.

Table 3: Experimental simulation

Client data size	Key size	$T.T._{Encry()}$	$T.T._{Upload-on-Cloud-Server}$	$T.T._{Download-res-and-Decry()}$
500 kB	16 bits	2 sec.	2 sec.	2.4 sec.
1 mB	32 bits	2.6 sec.	2.4 sec.	2.8 sec.
2 mB	32 bits	3 sec.	3.2 sec.	3.2 sec.
50 mB	64 bits	6 sec.	6.4 sec.	5.8 sec.
100 mB	64 bits	9.4 sec.	8.8 sec.	9 sec.
500 mB	128 bits	19 sec.	19.2 sec.	19 sec.
1 GB	128 bits	32.2 sec.	32 sec.	32.4 sec.

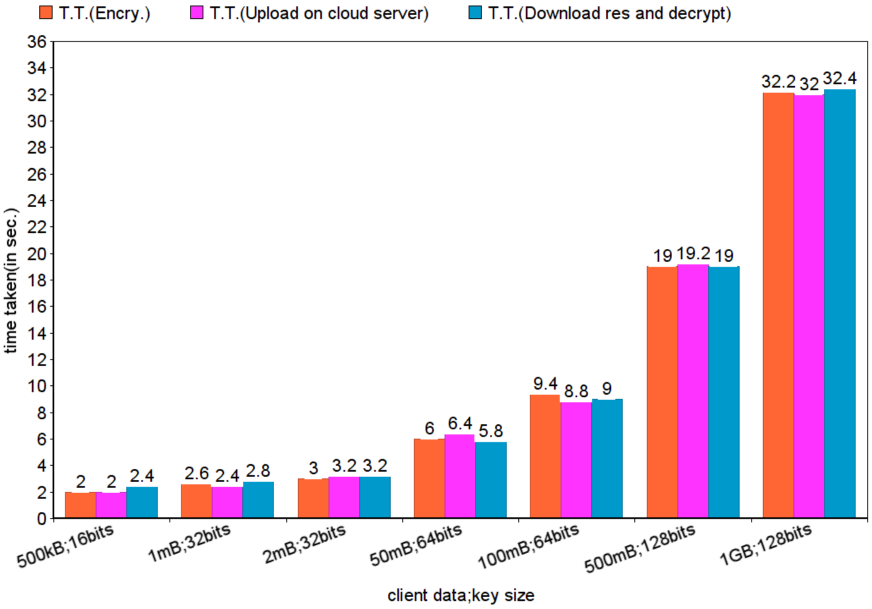
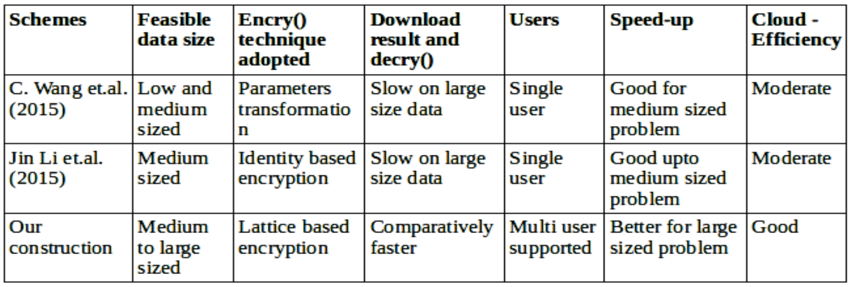


Figure 5: The overall performance

Table 4: Comparison with related work



Schemes	Feasible data size	Encry() technique adopted	Download result and decry()	Users	Speed-up	Cloud - Efficiency
C. Wang et.al. (2015)	Low and medium sized	Parameters transformatio n	Slow on large size data	Single user	Good for medium sized problem	Moderate
Jin Li et.al. (2015)	Medium sized	Identity based encryption	Slow on large size data	Single user	Good upto medium sized problem	Moderate
Our construction	Medium to large sized	Lattice based encryption	Comparatively faster	Multi user supported	Better for large sized problem	Good

In this paper, we have constructed a scheme for secure outsourcing based on multi cloud servers. The computational complexity and security analysis is also given for our proposed system. Finding an efficient, practical and computationally secure outsourcing solution for various specific scientific problems will be our further research work.

References

[1] G. Asharov, A. Jain, A. Lopez-Alt, E. Tromer, V. Vaikuntanathan, and D. Wichs, ”Multiparty computation with low communication, computation and interaction via threshold fhe,” in *Advances in Cryptology (EUROCRYPT'12)*, pp. 483-501, 2012.

[2] M. J. Atallah, K. Pantazopoulos, J. R. Rice, and E. E. Spafford, ”Secure outsourcing of scientific computations,” in *Trends in Software Engineering*, vol. 54, pp. 215-272, 2002.

[3] M. J. Atallah and J. Li, ”Secure outsourcing of sequence comparisons,” *Internatinal Journal Informations Security*, vol. 4, no. 4, pp. 277-287, Oct. 2005.

[4] M. J. Atallah and K. B. Frikken, ”Securely outsourcing linear algebra computations,” in *Proceedings 5th ACM Symposium*, pp. 48-59, 2010.

[5] M. J. Atallah and K. B. Frikken, ”Securely outsourcing linear algebra computations”, in *Proceedings 5th ACM Symposium*, pp. 48-59, 2010.

[6] D. Benjamin and M. J. Atallah, ”Private and cheating-free outsourcing of algebraic computations,” in *Proceedings 6th Annual Conference (PST'08)*, pp. 240-245, 2008.

[7] D. Benjamin, M. J. Atallah, ”Private and cheating-free outsourcing of algebraic computations,” in *Proceedings 6th Annual Conference (PST'08)*, pp. 240-245, 2008.

[8] Z. Brakerski and V. Vaikuntanathan, ”Efficient fully homomorphic encryption from (standard) LWE,” in *Proceedings of the IEEE 52nd Annual Symposium on Foundations of Computer Science (FOCS'11)*, pp. 97-106, 2011.

[9] Z. Brakerski and V. Vaikuntanathan, ”Fully homomorphic encryption from ring-lwe and security for key dependent messages,” in *Advances in Cryptology (CRYPTO'11)*, vol. 6841, pp. 505-524, 2011.

[10] Z. Brakerski and V. Vaikuntanathan, ”Fully homomorphic encryption from ring-lwe and security for key dependent messages,” in Advances in Cryptology (CRYPTO'11), vol. 6841, pp. 505-524, 2011.

[11] Z. Cao, L. Liu, O. Markowitch, ”Analysis of one scheme for enabling cloud storage auditing with verifiable outsourcing of key updates”, *International Journal of Network Security*, vol. 19, no. 6, pp. 950-954, Nov. 2017.

[12] A. K. Chattopadhyay, A. Nag and K. Majumder, ”Secure data outsourcing on cloud using secret sharing scheme”, *International Journal of Network Security*, vol.19, no.6, pp. 912-921, Nov. 2017.

[13] L. Cheung and C. Newport, ”Provably secure ciphertext policy ABE,” in *Proceedings 14th ACM Conference (CCS'07)*, pp. 456-465, 2007.

[14] K. Fan, J. Wang, X. Wang, H. Li and Y. Yang, ”A secure and verifiable outsourced access control scheme in fog-cloud computing”, *Sensors*, vol. 17, no. 7, 2017.

[15] V. Goyal, O. Pandey, A. Sahai, and B. Waters, ”Attribute-based encryption for fine-grained access control of encrypted data,” in *Proceedings 13th ACM Conference Computer Communications Security*, pp. 89-98, 2006.

[16] C. Gentry, ”A fully homomorphic encryption scheme,” Stanford University, 2009. (https://crypto.stanford.edu/craig/craig-thesis.pdf)

[17] M. Green, S. Hohenberger, and B. Waters, ”Outsourcing the decryption of ABE ciphertexts,” in *Proceedings 20th USENIX Conference (SEC'11)*, pp. 34, 2011.

[18] S. Halevi, Y. Lindell, and B. Pinkas, ”Secure computation on the web: computing without simultaneous interaction,” in *Advances in Cryptology*, pp. 132-150, 2011.

[19] F. Han, J. Qin, H. Zhao, and J. Hu, ”A general transformation from KP-ABE to searchable encryption,” *Future Generation Computer Systems*, vol. 30, pp. 107-115, Jan. 2014.

[20] J. Lai, R. Deng, C. Guan, and J. Weng, ”Attribute-based encryption with verifiable outsourced decryption,” *IEEE Transactions on Information Forensics Security*, vol. 8, no. 8, pp. 1343-1354, Aug. 2013.

[21] X. Lei, X. Liao, T. Huang, H. Li, and C. Hu, ”Outsourcing large matrix inversion computation to a public cloud”, *IEEE Transactions on Cloud Computing*, vol. 1, no. 1, 2013.

[22] Y. Liao, Y. He, F. Li, S. Jiang, S. Zhou, ”Analysis of an ABE scheme with verifiable outsourced decryption”, *Sensors*, vol. 18, no. 176, 2018.

[23] A. Lopez-Alt, E. Tromer, and V. Vaikuntanathan, ”Cloud-assisted multiparty computation from fully homomorphic encryption,” *IACR Cryptology ePrint Archive*, vol. 2011, no. 663, 2011.

[24] T. Nishide, K. Yoneyama, and K. Ohta, ”Attribute-based encryption with partially hidden encryptor-specified access structures,” in *Proceedings Applications Cryptography Networks Security*, pp. 111-129, 2008.

[25] A. Sahai and B. Waters, ”Fuzzy identity-based encryption,” in *Proceedings Advanced Cryptology*, pp. 457-473, 2005.

[26] V. Sudarshan, N. Satyanarayana. ”An efficient protocol for secure outsourcing of scientific computations to an untrusted Cloud”, *International Conference on Intelligent Computing and Control (I2C2)*, 2017.

[27] V. Sudarshan, N. Satyanarayana, A. Dileep Kumar. ”Lock-In to the meta cloud with attribute based encryption without outsourced decryption”, *IJCST*, vol. 5, no. 4, 2014.

[28] C. Wang, K. Ren, and J. Wang, "Secure and practical outsourcing of linear programming in cloud computing," in *Proceedings of IEEE Infoaom*, pp. 820-828, 2011.

[29] C. Xiang and C. Tang, "Securely verifiable outsourcing schemes of matrix calculation", *International Journal High Performance Computing and Networking*, vol. 8, no. 2, 2015.

[30] J. Yu, K. Ren, and C. Wang, "Enabling cloud storage auditing with verifiable outsourcing of key updates," *IEEE Transactions on Information Forensics Security*, vol. 11, no. 6, pp. 1362-1375, 2016.

[31] Y. Zhang and M. Blanton, "Efficient secure and verifiable outsourcing of matrix multiplications", *Department of Computer Science and Engineering*, pp. 158-178, 2014.

[32] H. Zhao, J. Qin, and J. Hu, "Energy efficient key management scheme for body sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 11, pp. 2202-2210, Nov. 2013.

[33] D. Zeng, S. Guo, and J. Hu, "Reliable bulk-data dissemination in delay tolerant networks," *IEEE Transactions Parallel Distributed Systems*, vol. 25, no. 8, pp. 2180-2189, 2014.

[34] J. S. Leon, *Hadamard Matrices and Hadamard Codes*, Dec. 27, 2018. (http://homepages.math.uic.edu/~leon/mcs425-s08/handouts/Hadamard_codes.pdf)

Biography

V. Sudarsan Rao received his M.Tech in Computer Science and Engineering from JNTUH,Hyderabad , India, in 2010. Currently, he is a Researcher in the Department of Computer Science and Engineering of JNTUH, Hyderabad. His research interests include Cloud Computing and Security of provably secure symmetric encryption schemes, efficient software implementations of cryptographic primitives, pattern recognition, and Network Security.

Dr. N. Satyanarayana, M. Tech (CS), Ph.D (CSE): The Nagole Institute of Science and Technologies headed by Dr. N.Satyanarayana who possessed the highest qualification in engineering. His qualifications are M.Phil, AMIE(ET), M.Tech (CS), Ph.D(CSE), MISTE, MCSI. He obtained his Ph.D in Computer Science and had a remarkable record of merit in pursuit his engineering studies. He is young and dynamic. He has a disposition to inspire both students and staff to achieve their right goals. With his extensive and rich research experience he is able to run the institution on sound academic ground.

Privacy Preservation for eHealth Big Data in Cloud Accessed Using Resource-Constrained Devices: Survey

Kittur Philemon Kibiwott[1], Zhang Fengli[1], Kimeli Victor K.[3], Omala A. Anyembe[2], Eugene Opoku-Mensah[1]
(Corresponding author: Kittur Philemon Kibiwott)

School of Information and Software Engineering, University of Electronic Science and Technology of China[1]
No.4 Section 2 Jianshe Rd North, Chengdu 610054, China
School of Computer Science and Engineering, University of Electronic Science and Technology of China[2]
Chengdu 611731, China
University of Eldoret[3]
Address: P. O. Box 1125-30100 Eldoret, Kenya
(Email: phkibiwott@gmail.com)

(Received Sept. 18, 2017; revised and accepted Jan. 10, 2018)

Abstract

Mobile technology is proving to offer unprecedented advantage to health professionals by providing a more efficient transmission and access to health services. However, mobile devices are resource-constrained. This is setback whenever storage and computation are required on ehealth big data. To mitigate this drawback, mobile computing is integrated with scalable cloud computing. While this is an advantage on mobile user's side by enlarging the limited resources of the device, it also gives rise to security and privacy concerns. In order to overcome these challenges associated with security and privacy, the data owner (hospital) encrypts data using Attribute Based Encryption (ABE) primitive due to the fine-grained access control advantage it offers then sends ciphertext to the cloud. To realize fast data access, the resource-constrained device securely outsources heavy computations to resource abundant cloud server on its behalf with the guarantee that the server cannot learn anything about plaintext. In this paper, a survey of ABE with outsourced decryption of the existing works that is applicable to resource-constrained device for accessing eHealth big data is provided.

Keywords: ABE; Big Data; Cloud Computing; eHealth; Mobile; Outsource; Resource-Constrained

1 Introduction

Advances in technology have led to the generation of variety of massive data from diverse sources. Consequently, traditional techniques of storage and sharing is difficult to implement. This is because the data is enormous, complex and some in unstructured format. Cloud computing is used to fill this gap. This is due to the scalability advantage it has over its traditional storage counterpart. Therefore, this means a third party will be the custodian of the data, for example a hospital can outsource its eHealth big data to the cloud to be shared with the users like government, insurance companies, patients, doctors and other hospitals. Data access control is an effective way to protect and preserve the privacy of eHealth big data and achieve confidentiality. In this case the owner of the data for instance has to encrypt the data using public key encryption (PKE), outsources it to the cloud and the intended receiver with valid decryption key corresponding to the encryption key recovers the data. But there is a limitation when encrypted data is to be decrypted by many categories of users especially in eHealth big data setting.

Identity Based Encryption (IBE) [54] scheme which regards identities as string of characters was proposed as an alternative to PKE to simplify certificate management process, hence decrease communication overhead. Drawback to this scheme is that when the volume of data is large and complex for example in the case of healthcare big data, the computation cost is high and also time consuming [42]. In the year 2005, another new kind of IBE known as Fuzzy Identity-Based Encryption (FIBE) was proposed by Sahai and Waters [53]. In FIBE, identities is regarded as set of descriptive attributes where a user that has a private key for a given set of attributes can recover the message. This was Attribute-Based Encryption (ABE) at its cradle stage. Attribute-Based Encryption one of the public-key encryption flavors, proves to be applicable in securely accessing eHealth big data. It has emerged to be a promising access control primitive for cloud computing in the recent years. To access data in cloud [60], the data owner has to provide expressive fine-grained access control on how data is to be exchanged with the users.

In ABE scheme, the receiver's private key can decrypt a certain ciphertext only if the associated attributes and access policy correlate. There are two kinds of ABE schemes: Key-Policy ABE (KP-ABE) [4, 21, 31, 47] and Ciphertext-Policy ABE (CP-ABE) [6,20,62]. In KP-ABE scheme, ciphetexts are labeled with sets of attributes and access policies of this attributes are associated with end user's private keys. While in CP-ABE scheme, every ciphertext is associated with an access policy, and every end user's private key is associated with a set of attributes. CP-ABE is regarded as the suitable technology for data access control in cloud storage system because the the data owners defines the access policy [33]. ABE is one of the powerful and most important technology for realizing fine-grained access control of data in the cloud. However, in the majority of ABE schemes, the major drawback is the inefficiency as the size of ciphertext and decryption overhead grows with the complexity of the access policy. This is a setback to users using resource-constrained devices. To overcome this problem and therefore accommodate these devices, secured partial decryption should be carried out by the cloud.

1.1 Motivation

The emergence of smartphones and social media have further extended the usage of mobile devices that people carry these devices everywhere they go. Users *e.g.* patients can communicate with the physicians and obtain the information they require anywhere anytime without being in hospital physically. This not only saves time but also serves well during emergencies. While this is an advantage, the device lack in abundant resources. Big data is voluminous and therefore cannot be accommodated by mobile device which calls for assistance from untrusted or semi trusted unlimited resource cloud server platform for efficient storage and processing of the data. This poses security and privacy challenges as the data is stored by a third party which is outside the data owner's view.

To protect the data from leakage, the owner has to encrypt and define expressively how the data is to be accessed and shared by various users before he offloads to the cloud. When the mobile user with the required credentials wants to recover the message from the cloud, he has to borrow power from the cloud server to perform computational-intensive tasks on his behalf without the server jeopardizing the privacy of data. The overhead on end user's side is reduced significantly [22, 41]. In this paper, extensive review of secure ABE technologies with outsourced decryption suitable for resource-constrained devices to preserve the privacy of eHealth big data in cloud is provided. We will first discuss the characteristics that forms eHealth big data, then define eHealth and security issues, finally we will give comprehensive outsourced decryption technologies, proposed future work and conclusion. This work can serve as a guide to the beginners in understanding the fundamental issues in security and performance of ABE primitive with outsourced decryption.

2 eHealth

eHealth paradigm envisages the transfer of health resources and health care by electronic means. It includes information and data sharing between patients and health service providers, hospitals, health professionals and health information networks, electronic health records, telemedicine services, portable patient-monitoring devices and operating room scheduling softwares [50]. eHealth an implementation of information communication technology is currently one of the major sectors where big data explosion is experienced [38].

2.1 eHealth Big Data

Big Data is defined using 5V's: Volume, Velocity, Variety, Value and Veracity. This data originates from different multiple sources such as networked sensors, mobile devices, web logs, call centers, smartphones and social media sites such as facebook. A forecast by IDC Digital Universe for 2012-2020 reveals that digital data will swell by almost half, that is from 0.8 zettabytes to 40 zettabytes [24]. According to [26], it anticipates that by the year 2020 80% of the US healthcare service providers will have implemented Electronic Health Record (EHR) systems, 80% of the general population will have adopted Personal Health Record (PHR) systems, while 80% of PHR and EHR systems will be connected using Health Information Exchange (HIE) systems hence voluminous amount of the data will be generated. Health big data according to [37] stem from pre-hospital, in-hospital and post-hospital.

The healthcare system data volume in USA hit 150 exabytes in 2011 [1] and its projected to increase more and the drift is due to real-imaging, wearable computing devices etc. Genomic-driven study, probe-driven treatment and health management are the two major sources that generates massive amount of ehealth big data. Due to huge volume of healthcare big data (terabytes to petabytes) and its complexity, it becomes hard to store them in traditional storage. An effective alternative is to store them in cloud owing to the elastic scalability and computation advantage provided by cloud. This means a third party will be the custodian of the data. According to Arora *et al.* [2], since the cloud servers and data owners are not within the same trusted domain, then the biggest concern when big data is stored in third party is security and privacy as cloud storage is untrusted or semi-trusted.

2.2 Big Data Characteristics in eHealth

According to Gartner [17], big data is high-volume, high-velocity and high-variety information assets demanding cost effective, innovative forms of information processing to enhance insight and decision making. IBM added the fourth “V“ (veracity), while Oracle added the fifth “V“ (value).

1) *Volume*: Refers to massive amount of data generated from different sources in healthcare industry [46]. For example the ehealth big data can be generated by medical sensors, doctors, other hospitals, insurance companies, government etc. Utilizing Electronic Health Records (EHR) and its significant growth of the correlated healthcare related data generate increasing volume of health information. In 2012, the digital healthcare data in the entire universe was estimated to be equivalent to 500 Petabytes and its projected to attain 25,000 Petabytes by 2020 [28].

2) *Velocity*: Velocity is needed by healthcare providers and consumers for timely and proper decision making. It refers to the rate at which ehealth big data is generated, stored, analyzed and apportioned to different healthcare providers and consumers. The system should be efficient and secure as patient's data is critical.

3) *Variety*: It refers to different forms of ehealth big data. This can be in structured form *e.g.* EHR which can be easily stored by machine, unstructured, or semi-strucured can be inform of prescription, doctor's notes, images, x-ray etc.

4) *Value*: Refers to extracting meaningful data from eHealth big data. This takes place during processing of healthcare data. Extracting desired data can be used for example in research to curb the future outbreak of diseases.

5) *Veracity*: Refers to ehealth data with different quality, relevance and meaning. Since we have different forms of eHealth big data it follows that we will also have different quality of data. The quality of data has direct implication on the life of the patient. For that matter, eHealth big data quality should be reliable.

2.3 Sources of Big Data in eHealth

According to Iroju *et al.* [46] healthcare data is generated from:

- Biometric Generated Data: Biometric data is the record of data that uniquely identifies people. It originates from individuals‘ bodily characteristics such as facial scans, genetics, retinal scans, heart rate, blood pressure.

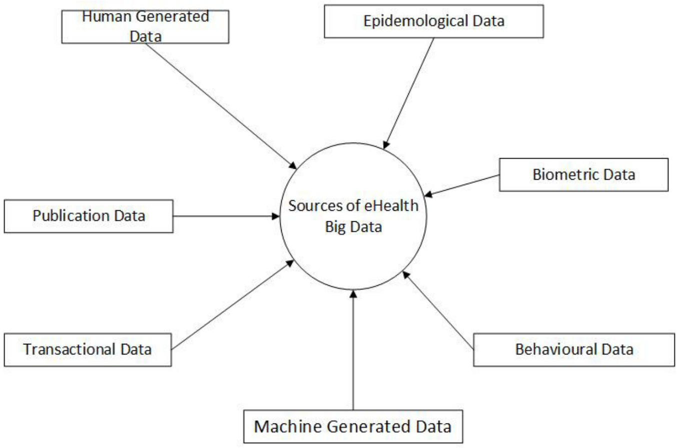


Figure 1: Sources of eHealth big data

- Transactional Generated Data:These include data emanating from healthcare individual claims and billing records. Examples include charges levied records on patients. They can be in semi structured or structured format.
- Publication Generated Data: Refers to data from health researches, medical science reference materials and government proposals data. Health research include exploratory research, descriptive research, explanatory research, and emancipatory research.
- Machine Generated Data: Refers to data that are generated by machines used in the healthcare system. Examples include data coming from remote sensors, wearable devices, x-ray machines, ECG machine, anesthesia machines etc.
- Human Generated Data: Refers to data produced by human beings in the healthcare system. It comprises unstructured and semi structured clinical data such as case prescription notes, hospital admission records, laboratory results, discharge summaries and electronic mails. Digitization of health records such as the use of structured Electronic Health Record (EHR) has also resulted to voluminous data.
- Epidemological Generated Data: These data include vital statistical data, causes of diseases, impacts, identify disease risk factors, health surveys and targets for preventive healthcare, patterns such as disease distribution in population and probe these disease causes.
- Behavioural Generated Data: Refers to data generated from social intercourses and communication tools such as websites and social media sites such as Twitter, Facebook and WhatsApp.

Figure 1 provides a summary of the sources of eHealth big data.

2.4 Information Security and Privacy for eHealth Big Data

2.4.1 eHealth Privacy

Safeguarding personal health information from disclosure, loss, unauthorized access, modification or used without patient's consent [49]. eHealth systems should be build with privacy as a priority [42]. Normally privacy is judged from the harm it causes if an individual information goes public [43]. eHealth data that demands privacy is divided into:

1) Highly risk data: Data that can cause harm to an individual.

2) Restricted data: Data covered by state or federal legislation.

3) Confidential Data: Example is patient's ID.

2.4.2 Confidentiality

Confidentiality makes certain that healthcare information provided to healthcare professions cannot leak to the third party without owner's consent. Confidentiality seeks "non inteference of information and protective actions of information such as security measures" [13].

2.4.3 Integrity

This ensures that the data/information remains unchanged. Data should only be modified by those who are authorized to do so [48]. Loss of data, data breach and the correctness of the data are the concerns as far as integrity of data is concerned when data is outsourced to the cloud.

2.4.4 Availability

The presence of eHealth information to authorized users when it's needed [45]. This permits health professionals to obtain accurate and timely health information that will add value to the patient treatment [48].

2.4.5 Authentication

Sources of eHealth data needs to be determined before its used to confirm its true originality [14]. The moment authentication process is established, it should be clearly stated what data is permitted to be accessed and the requirements for one to access them.

2.5 Access Structures [5]

These are set of qualified families that can construct a secret. Let $\mathbb{A}$ be the universe of attributes. A collection $\mathbb{P} \subseteq 2^{\mathbb{A}} \setminus \{\emptyset\}$ is monotone if $\forall \mathbb{B}, \mathbb{C}$: if $\mathbb{B} \in \mathbb{P}$ and $\mathbb{B} \subseteq \mathbb{C}$,then $\mathbb{C} \in \mathbb{P}$. An access structure is a collection $\mathbb{P}$ of non-empty subsets $\mathbb{P} \subseteq 2^{\mathbb{A}} \setminus \{\emptyset\}$. The sets in $\mathbb{P}$ are called authorized sets, and the sets not in $\mathbb{P}$ are unauthorized sets.

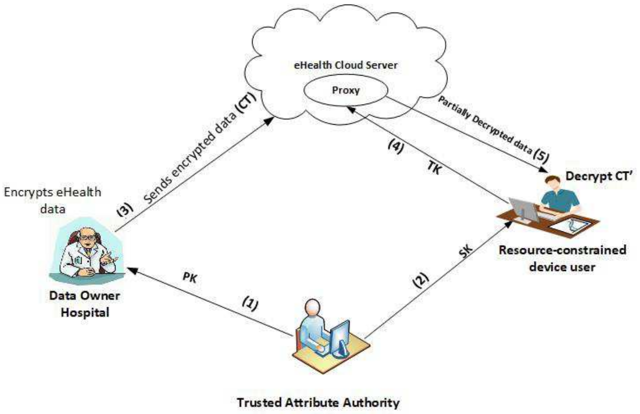
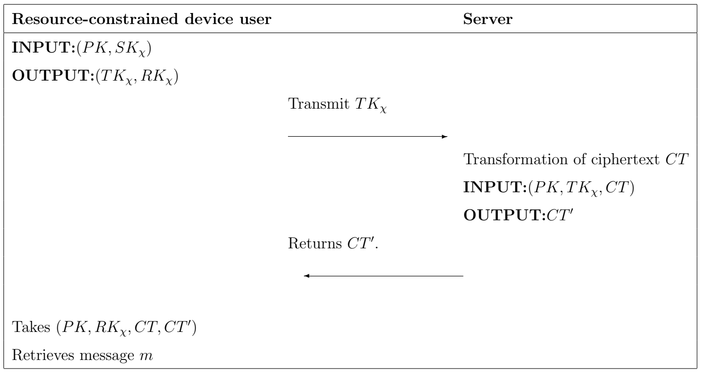

eHealth Cloud Server
Proxy
Sends encrypted data (CT)
(3)
Encrypts eHealth data
Partially Decrypted data (5)
(4)
TK
Decrypt CT'
Resource-constrained device user
SK
(2)
PK
(1)
Data Owner Hospital
Trusted Attribute Authority

Figure 2: Secure eHealth big data access for resource-constrained devices in cloud computing



Resource-constrained device user | **Server**

INPUT:(PK, SK_{χ})

OUTPUT:(TK_{χ}, RK_{χ})

Transmit TK_{χ} →

Transformation of ciphertext CT

INPUT:(PK, TK_{χ}, CT)

OUTPUT:CT'

Returns CT'. ←

Takes (PK, RK_{χ}, CT, CT')

Retrieves message m

Figure 3: Outsourced decryption primitive

2.6 Bilinear Maps

Let $\mathbb{G}$,$\mathbb{G}_{\tau}$ be two multiplicative cyclic groups of prime order p. Let g, g_1 be $\mathbb{G}$ generators and e be a bilinear map; e: $\mathbb{G} \times \mathbb{G} \to \mathbb{G}_T$. Bilinear map e has the following properties:

1) *Bilinearity* : $\forall g, g_1 \in \mathbb{G}$ and $a, b \in \mathbb{Z}_p^*$ we have $e(g^a, g_1^b) = e(g, g_1)^{ab} = e(g^b, g_1^a)$.

2) $Non-degeneracy : e(g, g_1) \neq 1$.

3) *Computability*: There is an efficient algorithm to compute $e(g, g_1)$.

2.7 Complexity Assumptions

The Decisional Bilinear Diffie-Hellman (DBDH) assumption: Let $x, y, z, c \in \mathbb{Z}_p$ be randomly chosen and $g \in \mathbb{G}$ be a generator. The DBDH assumption [53] holds in $\mathbb{G}$ if no probabilistic polynomial-time algorithm can distinguish the tuples $(g, g^x, g^y, g^z, e(g, g)^{xyz})$ from the tuple (g, g^x, g^y, g^z, g^c) with non-negligible advantage. *Discrete Logarithm (DL) Assumption*: Let $(p, \mathbb{G}_1, \mathbb{G}_{\tau}, e)$ be a prime order bilinear group system. Given $(p, \mathbb{G}_1, \mathbb{G}_{\tau}, e,\ g, g^x)$, where $g \in \mathbb{G}$ and $x \in \mathbb{Z}_p^*$ are uniformly chosen randomly, the Discrete Logarithm problem is to compute x. The Discrete Logarithm assumption [14] in the prime order bilinear group system $(p, \mathbb{G}_1, \mathbb{G}_\tau, e)$ is that no probabilistic polynomial-time (PPT) algorithm $\mathcal{A}$ can solve the DL problem with non-negligible advantage. The advantage of $\mathcal{A}$ is defined as

$$\Pr[\mathcal{A}(p, \mathbb{G}_1, \mathbb{G}_\tau, e, g, g^x) = x],$$

where probability space is over g, x chosen randomly and random bits consumed by $\mathcal{A}$.

Decisional Bilinear Diffie-Hellman Exponent (DBDHE) assumption: Let α, t be randomly chosen and $g \in \mathbb{G}$ be a generator. The decisional q-DBDHE assumption [8] is that no probabilistic polynomial-time algorithm $\mathbb{AD}$ can distinguish the tuple $e(g,g)^{\alpha^{q+1}} \in \mathbb{G}_\tau$ from a random element $C \in \mathbb{G}_\tau$ with more than a non-negligible advantage provided ϵ=(g, g_q, g_{q+2},...,g_{2q}, g^t), where g_I is denoting g^{α^I}. Advantage of algorithm $\mathbb{BD}$ solving decisional q-BDHE is:

$$|Pr[\mathbb{BD}(\epsilon, V = e(g,g)^{\alpha^{q+1}}) = 0] - Pr[\mathbb{BD}(\epsilon, V = C) = 0]| \geq \in .$$

Decisional modifed Bilinear Diffie-Hellman (MBDH) assumption: Suppose a challenger $\mathbb{BD}$ randomly selects $\alpha, \beta,\ \gamma,\ z\ \in\ \mathbb{Z}_p$. Decisional modified bilinear Diffie-Hellman (MBDH) [65] is that no probabilistic polynomial-time algorithm $\mathbb{AD}$ can distinguish the tuple $(g, g^\alpha, g^\beta, g^\gamma, (g,g)^{\alpha\beta/\gamma})$ from $(g, g^\alpha, g^\beta,\ g^\gamma,\ (g,g)^z)$ with non-negligible advantage.

2.8 Formal Structure for Generic ABE

The intuition to this basic ABE is that the intended user also referred to as legitimate user with the given set of attributes specified in the access policy at the time of encryption can access the data.

An ABE consists of four basic algorithms [53] as follows:

- *Setup*: This is a probabilistic algorithm, executed by trusted attribute authority. Takes as input security parameter γ and outputs a pair (PK, MSK). Where PK is public parameter while MSK is master secret key.
$$(PK, MSK) \leftarrow Setup(1^\gamma).$$
- *KeyGen*: This algorithm is executed by trusted attribute authority to produce secret/private key. The input to the key generation algorithm is a set of attributes χ, master secret key MSK and public key PK. It yields private/decryption key SK_χ.
$$SK_\chi \leftarrow KeyGen(\chi, MSK, PK).$$
- *Encrypt*: This is a probabilistic algorithm, executed by data owner (sender). Takes as input the message m to be encrypted, the set of attributes χ and public key PK. It yields as output ciphertext CT.
$$CT \leftarrow Encrypt(PK, \chi, m).$$
- *Decrypt*: This algorithm is deterministic and its is executed by the intended user/decryptor. Takes as input ciphertext CT, public key PK, and private key SK satisfied by the set of attributes. It returns as an output a message m.
$$m \leftarrow Decrypt(PK, SK_\chi, CT).$$

Correctness.

$$m \leftarrow Decrypt(Encrypt(PK, \chi, m), PK, SK_\chi).$$

3 ABE with Outsourced Decryption for eHealth Big Data

3.1 Formal Structure for ABE with Outsourced Decryption

The intuition of the ABE with outsourced decryption primitive is that an authorized mobile user that possesses a given set of attributes satisfying the access structure can securely access the data. The hospital encrypts the data and specifies the access policy then sends it to the cloud. When a mobile device user having required set of attributes wants to access the data it will first sends a transformation key to the proxy to perform heavy computation overhead such as compute pairings on his/her behalf. Transformed ciphertext will then be sent to the end user. Limited-resource device user will perform final decryption. Using this primitive improves performance of resource-constrained device. The complete system is as shown in Figure 2.

Attribute Based Encryption with outsourcing decryption for eHealth big data system has the following five entities:

1) *Trusted Attribute Authority*: This is the only trusted entity. It generates public and private keys and parameters for Attribute Based Encryption scheme. Public key is used for encryption of data and private key is used for decryption to recover the original message. Resource-constrained device users also receive from trusted attribute authorities their attributes that corresponds to decryption keys.

2) *Hospital*: It's is the owner of the data. Prior to outsourcing the data to the cloud, it defines how data is to be accessed by authorized users then encrypt it under given access policy.

3) *eHealth cloud*: It is semi-trusted or untrusted entity. it has unlimited resources and therefore provides storage facilities, high computation power and access for eHealth big data.

4) *Proxy*: It interacts with resource-constrained device users. It transforms efficiently using blinding key the encrypted data into a simple ciphertext without learning the plaintext of the data.

5) *Resource-constrained device users*: Authorized entities possessing a set of attributes that satisfies the access policy of the encrypted data can decrypt and recover the message. They receive decryption keys that corresponds to their attributes from Trusted Attribute Authorities. In our setting users can be government, medical research organizations, insurance companies, other hospitals etc.

An ABE with outsourced decryption has the following basic algorithms [22, 36]:

- *Setup*: This is a randomized algorithm, executed by trusted attribute authority. Takes security parameter η as input and produces as output a pair (PK, MSK). Where PK is public parameter while MSK is master secret key.

$$(PK, MSK) \leftarrow Setup(1^{\eta}).$$

- *KeyGen*: This is randomized algorithm executed by trusted attribute authority to yield private key. The input to the key generation algorithm is a set of attributes χ, master secret key MSK and public parameter PK. It produces as output private/decryption key SK_{χ}.

$$SK_{\chi} \leftarrow KeyGen(\chi, MSK, PK).$$

- *Encrypt*: This is a probabilistic algorithm, executed by the owner of the data (sender). Takes a message m to be encrypted, the set of attributes χ and public key PK as inputs. It yields as ciphertext CT output.

$$CT \leftarrow Encrypt(PK, \chi, m).$$

- *Decrypt*: This algorithm is deterministic and its is executed by the intended data user/decryptor. Takes ciphertext CT, public key PK, and private key SK satisfied by the set of attributes as inputs. Returns as output message m.

$$m \leftarrow Decrypt(PK, SK_{\chi}, CT).$$

- *Transform$_{out}$*: This algorithm is executed by resource-constraint user as shown in figure 3. It takes transformation key TK_{χ} and ciphertext CT as input. It returns CT'.

$$CT' \leftarrow Transform_{out}(TK_{\chi}, CT).$$

- *Decrypt$_{out}$*: The algorithm is executed by intended resource-constrained device user. It takes retrieving key RK_{χ} and transformed ciphertext CT′ as input. Returns message m.

$$m \leftarrow Decrypt_{out}(RK_{\chi}, CT').$$

Correctness.

1) $Decrypt(PK, SK_{\chi}, Encrypt(PK, m, \mathbb{A})) = m$
2) $Decrypt_{out}(PK, RK_{\chi}, Transform_{out}(Encrypt(PK, M, \mathbb{A}), PK, \text{TK}_{\chi})) = m$

3.2 Adversary Models

While proposing any system, security is the primary consideration. Users of the system should be convinced that it is secure enough to trust their data into it. Parties involved in the system, rarely trust each other but all have one common thing, they all trust the protocol proposed. Accessing the contents to which they are not authorized to is the main objective of any adversary. They may collude with others or work independently [15]. Some of the system threats in ABE with outsourced decryption originate from [15, 22, 29, 32];

1) *Semi-trusted/untrusted cloud servers colluding with unauthorized users*: It is assumed that the servers provides their services smoothly but may at times be curious of leaking sensitive information such as a patient data to illegitimate users.

2) *Attribute authorities (AA)*: AA may willingly violate data owners by conspiring with the outsiders where they issue them with keys to enable them access unauthorized data.

3) *Legitimate users colluding with each other*: Authorized users can combine their attributes to access unauthorized data which individually could not access.

4) *Replay attacks*: Legitimate users can re-submit the previous tokens to the servers to obtain unauthorized data.

5) *Active attacks*: Unauthorized users may introduce malicious data into the cloud to harvest some data or corrupt them.

6) *Servers colluding with authorized users*: Cloud server may collude with authorized users to obtain unauthorized data for the purpose of using them for their malicious end.

3.3 Security Models

Definition of security of any cryptosystem is based on what is to be achieved and a particular attack. There are three known security models common to all cryptosystems. In order of security strength they are adaptive chosen ciphertext attack (CCA2) [52] (stronger), non-adaptive chosen ciphertext attack (CCA1) [44] and chosen plaintext attack (CPA) [19]. Where CCA2 security model is more secure.

Security goals: To realize fully the benefits of ABE with outsourced decryption for eHealth big data, the following security goals should be met [30, 41, 55]:

1) *Fine-grained access control*: Hospital (owner of data) should be in a position to safeguard its sensitive information using strong security measures. Only legitimate users with the set of attributes defined in the

access policy, manages to retrieve the stored data in the cloud.

2) *Efficient encryption/decryption*: Legitimate users should be able to access the stored information without delay as the health big data is so sensitive as it deals with people's lives.

3) *Collusion resistance*: Since data custodian is the third party which is semi-trusted / untrusted in nature, collusion among different unauthorized parties should be thwarted for example collusion between cloud and illegitimate users to acquire private keys to access unauthorized data which individually could not access should be prevented.

4) *Confidentiality of data*: eHealth big data should not get leaked to the third party without owner's approval.

5) *Convenience*: With the proliferation of mobile devices which is resource-constrained in nature, and which are incapable to finish decryption independently or consumes much time to finish decryption, utilizing outsourcing decryption enables the legitimate users to access a bulk of eHealth data anywhere anytime since heavy computations is offloaded to cloud.

6) *Unidirectional*: The main property of outsourced decryption is unidirectionality. This means the server has capability only to transform original (*e.g.* alice's) ciphertext to another (*e.g.* Bob's) ciphertext and not in reverse direction.

7) *Public verifability*: Information should be available that enables involved parties i.e users to confirm/verify the genuineness of original ciphertext and the transformed ciphertext.

8) *Immediate revocation*: Mischievous/malicious users should easily and completely be revoked from all future data access.

9) *User/ciphertext anonymity*: Disclosure of user's identity or key privacy also referred to as ciphertext anonymity should not be revealed. It should hide users/ciphertext identities.

10) *Scalability*: With the increase of legitimate users, the system efficiency is still guaranteed. The performance of the system cannot be affected by number of legitimate users.

4 State-of-the-art of ABE with Outsourced Decryption for Resource-Constrained Devices

The leading efficiency drawback in the vast majority of ABE is the increase in size of the ciphertext and the decryption cost (computational cost) with the increase in complexity of the access policy. The applications executing in mobile devices which are resource-constrained in nature in terms of battery life, computational resources, storage, and bandwidth may have to hold on for a long time or even abort before execution to finality. Using these devices to access eHealth big data is not faster enough as it is costly due to bilinear pairing operations involved. To curb this limitations, the remedy is to adopt mobile cloud computing where heavy computations are offloaded to the cloud [18, 64].

The first ciphertext-policy attribute-based PRE (CP-ABPRE) scheme, in which a cloud server is authorized to transform a ciphertext under a specified access structure (represented only as AND gates on positive and negative attributes) into the one under another access structure was proposed in 2009 [35]. In this scheme, the user successfully decrypts the ciphertext if and only if the set of positive and negative attributes are embedded in the access structure.

To minimize the number of pairing operations on end users side and hence reduce decryption overhead, ABE with outsourced decryption schemes is proposed where intensive computational tasks is outsourced to cloud service providers [33, 34, 61]. The schemes proposed by Green *et al.* [22] and De *et al.* [12] provides fine-grained access control solution to the lightweight devices such as mobile phones with constrained computing resources which independently cannot successfully execute basic encryption/decryption while protecting sensitive data outsourced to the public cloud [32]. Scheme of [22] achieved CPA-security which was later extended to achieve the stronger RCCA-security in random oracle model. With the provision of outsourced decryption, heavy computations and storage can be offloaded and the light computations be performed by resource-constrained devices effectively and efficiently. Generally in [22], the key generating algorithm is designed to output two key pair to the data user as follows:

1) A short El Gamal kind private key known as retrieving key *rk*.

2) Its paired key known as transformation key *tk*, which is send to the server and its publicly known.

In this scheme a key for blinding (ie transformation key) *tk* is sent to the third-party (server) for translation of any ciphertext CT satisfied by end user's attributes or access policy into a simpler ciphertext CT'. The end user incurs minimal overhead to recover plaintext from transformed ciphertext. The major drawback to this technique is that a ciphertext can be mutated on the transit therefore making the user unable to realize and detect the change. For the case of eHealth it can lead to wrong diagnosis hence cause fatal consequences.

In addressing the same ABE computational problem, where cost grows linearly with respect to complexity of the number of attributes or ciphertext policy and which

Table 1: Comparison of the schemes characteristics with outsourced decryption

Scheme	**Characteristics**							
	Fine-grained access	**Efficiency**	**Collusion resistance**	**Confidentiality**	**Unidirectionality**	**Verifiability**	**Immediate revocation**	**Scalability**
Zhou [64]	√	√	√	√	√	×	×	√
Li [34]	√	√	√	√	√	×	√	√
Lai [29]	√	√	√	√	√	√	×	√
Green [22]	√	√	×	√	√	×	×	√
Li [32]	√	√	√	√	√	√	×	√
Jiguo 16 [33]	√	√	√	√	√	×	√	√
Lin [36]	√	√	×	√	√	√	×	√
Mao [41]	√	√	√	√	√	√	×	√
Jiguo 17 [27]	√	√	×	√	√	√	×	√
Zechao [39]	√	√	×	√	√	√	×	√

Abbreviations: √: Scheme supports the corresponding characteristic, ×: Scheme does not support the corresponding characteristic.

Table 2: Comparison of the security models of ABE schemes with outsourced decryption

Scheme	**Security**	**Model**	**Complexity assumption**
Zhou [64]	IND-CPA	adaptive	Co-DBDH
Li [34]	RCCA	selective	DBDH
Lai [29]	RCCA	selective	DL
Green [22]	CPA	selective	Decisional q-BDHE
Li [32]	RCCA	selective	DBDH
Jiguo 16 [33]	CPA	selective	DCDH
Lin [36]	IND-CPA	adaptive	DL
Mao [41]	IND-CPA	selective	Generic group
Jiguo 17 [27]	IND-CPA	selective	DL
Zechao [39]	IND-CPA	selective	DON'T EXIST

Abbreviations: $DBDH$: Decisional Bilinear Diffie-Hellman, DL: Discrete Logarithm, $q-BDHE$: Decisional Bilinear Diffie-Hellman Exponent, $Co-DBDH$: Co-Decisional Bilinear Diffie-Hellman, $DCDH$-Divisible Computation Diffie-Hellman, $IND-CPA$:Indistinguishable Chosen Plaintext Attack, $RCCA$:Replayable Chosen Ciphertext Attack.

is a bottleneck to resource-constrained devices such as mobile devices, Zechao *et al.* [39] proposed a new CP-ABE scheme known as Offline/online attribute-based encryption with verifiable outsourced decryption protocol. In this scheme, offline/online technique is combined with the outsourced verifiable computation technique using bilinear groups, which supports both offline/online generation of key and encryption, as well as the verifying outsourced decryption. Heavy computation during key generation is executed offline, and encryption can be split into two phases offline and online where heavy tasks is executed offline and lightweight tasks are executed online efficiently. On the other hand, decryption workload is offloaded to the server. To overcome the disadvantage of complexity, Song *et al.* [40] extended the scheme of Emura *et al.* [16] scheme. In [40], an alternative technique is proposed where decryption process is made faster by making use of only a constant number of bilinear operations. The decryption cost and ciphertext length are decreased significantly in comparison with previous protocols. While addressing the same issue of defining the access structure policy and allowing the owner to outsource intensive computation tasks to cloud server providers, Yu *et al.* [61] integrated techniques of Attribute-Based Encryption (ABE), proxy re-encryption and lazy re-encryption. In this scheme, the data owner enforces a unique access structure on each user. While executing this protocol, the cloud servers is prevented from learning about the underlying plaintext.

The bilinear pairings computation is considered to be the most costly operation experienced in pairing-based cryptographic protocols construction. In order to execute the protocol with pairing and hence accommodate devices with limited resources, Benot *et al.* [11] proposed secure delegation of elliptic-curve pairing by resource-constrained device to a more powerful device. Pairing $e(X, Y)$ for example is delegated to a more powerful device (for instance a PC). Delegation is done in such a way that a powerful device cannot learn about the points X and Y. To verify the output and confirm whether the terminal is cheating, the resource-constrained device either yields the correct output or nothing with overwhelming accuracy. However, the drawback to this scheme is that the resource-constrained device restricts itself to a simple

Table 3: Comparison of the storage overhead of schemes with outsourced decryption

Scheme	Length of the Key					Ciphertext	
	Public Key	Master Key	Private Key	Transform Key	Retrieval Key	Original ciphertext	Transformed ciphertext
Zhou [64]	$2\|\mathbb{G}\| + \|\mathbb{G}_\tau\|$	$\|\mathbb{G}\| + \|\mathbb{Z}_p\|$	$(2N+3)\|\mathbb{G}\|$	$(2N+3)\|\mathbb{G}\|$	$\|\mathbb{Z}_p\|$	$(2\mathrm{N}+1)\|\mathbb{G}\| + \|\mathbb{G}_\tau\|$	$(4\mathrm{N}+2)\|\mathbb{G}\| + \|\mathbb{G}_\tau\|$
Li [34]	$(N+4)\|\mathbb{G}\|$	$\|\mathbb{Z}_p\|$	$(2\mathrm{N}+5)\|\mathbb{G}\|$	$2\mathrm{N}\|\mathbb{G}\|$	$2\|\mathbb{G}\|$	$(\mathrm{N}+2)\|\mathbb{G}\| + \|\mathbb{G}_\tau\|$	$2\|\mathbb{G}\| + 2\|\mathbb{G}_\tau\|$
Lai [29]	$(N+5)\|\mathbb{G}\| + \|\mathbb{G}_\tau\|$	$\|\mathbb{Z}_p\|$	$(N+2)\|\mathbb{G}\|$	$(N+2)\|\mathbb{G}\|$	$\|\mathbb{Z}_p\|$	$(4\mathrm{N}+3)\|\mathbb{G}\| + 2\|\mathbb{G}_\tau\|$	$\|\mathbb{G}\| + 4\|\mathbb{G}_\tau\|$
Green [22]	$2\|\mathbb{G}\| + \|\mathbb{G}_\tau\|$	$\|\mathbb{Z}_p\|$	$(N+2)\|\mathbb{G}\|$	$(N+2)\|\mathbb{G}\|$	$\|\mathbb{Z}_p\|$	$(2\mathrm{N}+1)\|\mathbb{G}\| + \|\mathbb{G}_\tau\|$	$2\|\mathbb{G}_\tau\|$
Li [32]	$(N+4)\|\mathbb{G}\|$	$\|\mathbb{Z}_p\|$	$(2\mathrm{N}+3)\|\mathbb{G}\|$	$(2\mathrm{N}+3)\|\mathbb{G}\|$	$\|\mathbb{Z}_p\|$	$(\mathrm{N}+2)\|\mathbb{G}\| + \|\mathbb{G}_\tau\|$	$\|\mathbb{G}_\tau\|$
Jiguo 16 [33]	$2\|\mathbb{G}\| + \|\mathbb{G}_\tau\|$	$\|\mathbb{G}\| + \|\mathbb{Z}_p\|$	$(2N+3)\|\mathbb{G}\|$	$(2N+3)\|\mathbb{G}\|$	$\|\mathbb{Z}_p\|$	$(2\mathrm{N}+4)\|\mathbb{G}\| + \|\mathbb{G}_\tau\|$	$5\|\mathbb{G}_\tau\|$
Lin [36]	$(N+4)\|\mathbb{G}\| + \|\mathbb{G}_\tau\|$	$\|\mathbb{Z}_p\|$	$(N+2)\|\mathbb{G}\|$	$(N+2)\|\mathbb{G}\|$	$\|\mathbb{Z}_p\|$	$(2\mathrm{N}+2)\|\mathbb{G}\|$	$\|\mathbb{G}_\tau\|$
Mao [41]	$(N+4)\|\mathbb{G}\| + \|\mathbb{G}_\tau\|$	$\|\mathbb{Z}_p\|$	$(N+2)\|\mathbb{G}\|$	$(N+2)\|\mathbb{G}\|$	$\|\mathbb{Z}_p\|$	$(2\mathrm{N}+2)\|\mathbb{G}\|$	$\|\mathbb{G}_\tau\|$
Jiguo 17 [27]	$(N+5)\|\mathbb{G}\| + \|\mathbb{G}_\tau\|$	$(N+1)\|\mathbb{Z}_p\|$	$2\|\mathbb{G}\|$	$2\|\mathbb{G}\|$	$\|\mathbb{Z}_p\|$	$5\|\mathbb{G}\| + 2\|\mathbb{G}_\tau\|$	$\|\mathbb{G}\| + 4\|\mathbb{G}_\tau\|$
Zechao [39]	$5\|\mathbb{G}\| + \|\mathbb{G}_\tau\|$	$5\|\mathbb{G}\| + \|\mathbb{G}_\tau\| + \|\mathbb{Z}_p\|$	$(2k+5)\|\mathbb{G}\| + 3k\|\mathbb{Z}_p\|$	$(2k+5)\|\mathbb{G}\| + 3k\|\mathbb{Z}_p\|$	$\|\mathbb{Z}_p\|$	$(1+3l)\|\mathbb{G}\| + 3l\|\mathbb{Z}_p\|$	$3\|\mathbb{G}\| + \|\mathbb{G}_\tau\|$

Abbreviations: N: Attribute size, $|\mathbb{G}|$: bit length of an element in $\mathbb{G}$, $|\mathbb{G}_\tau|$: bit length of an element in $\mathbb{G}_\tau$, $|\mathbb{Z}_p|$: bit length of an element in $\mathbb{Z}_p$.

curve or provided field operations. Not flexible enough to support complex curves.

Muhammad *et al.* [3], proposed attribute-based encryption with encryption and decryption outsourcing that reduces the computational load on both the host and the users using devices that are computationally resource-constrained (*e.g.*. mobile devices). The scheme is comprised of two proxies which are independent and cannot collude, one on the host side and the other one on the user's side. In the former, data owner is allowed to outsource cryptographic creation policy to semi-trusted proxy. The proxy is unable to learn about encrypted messages and is enforced to encrypt the messages based on the policy specified on the attributes. While in the latter, the heavy computation overhead during decryption is reduced by allowing a user to offload the verification policy onto another semi-trusted proxy where it borrows power from the proxy to verify the policy using the user's key transformation attributes. This scheme is provable secure under the generic group model.

To ensure the server legitimately executes outsourced decryption, a number of schemes have been proposed [27, 29, 36, 41, 59]. Lai *et al.* [29] and Mao *et al.* [41] separately introduced verifiability primitive in the outsourced decryption. To accomplish this, an extra instance is added to the existing ABE in encryption/decryption algorithm phases. A drawback to the scheme is that owner of data has to perform an extra work of encrypting the random message then compute checksum value corresponding to two messages. As a result, computation and communication overhead are duplicated. To overcome this drawback, Lin *et al.* [36] proposed a more efficient ABE with verifiable outsourced decryption based on an attribute-based key encapsulation mechanism, a symmetric-key encryption scheme and a commitment scheme in generic model. The scheme in [36] can be considered both in Key-Policy Attribute-Based Encryption (KP-ABE) and also in Ciphertext-Policy Attribute-Based encryption (CP-ABE) settings. Solution to checking the integrity of outsourced data while maintaining privacy and secrecy of the stored data was proposed by Yadav *et al.* [59]. In this scheme, secure operations in data storage can be augmented to provide remote integrity checking. It is carried out by computing just once the hash of data, and therefore mobile user does not need to posses outsourced data. However, from all this primitives the number of the attributes grows linearly with the length of the ciphertext and the size of costly pairing computations. This greatly affects outsourced CP-ABE scheme by limiting verifiability. To avoid this, Jiguo *et al.* [27] proposed Verifiable Outsourced Decryption of Attribute-Based Encryption with Constant Ciphertext Length that saves the communication cost.

Avoiding the distruption of the medical information system while simultaneously achieving fine-grained, privacy and confidentiality properties, Junbeom *et al.* [23] proposed a scheme that is key escrow resilient and which allows the partial decryption of the encrypted medical data by device controller without leaking any private information to the controller. This improves computational efficiency of the medical devices where most of the laborious decryption tasks is delegated to the device controller. However, the schemes do not achieve checkability on the output returned, therefore there is no guarantee of the accuracy of the partial decrypted ciphertext. To provide a solution to this, a fine-grained, multiparty access control with outsourcing decryption, was proposed by Qinlong *et al.* [51] where Cloud Service provider (CSP) can transform original ciphertext defined under access policy to another

Table 4: Comparison of the computation cost of schemes with outsourced decryption

Scheme	Computation cost				Access structure
	Encrypt	Transform$_{out}$	Decrypt	Decrypt$_{out}$	
Zhou [64]	$(2\|A_{ct}\|+1)\mathbb{G}+2\mathbb{G}_\tau$	$(2\|A_{ct}\|+1)\mathbb{G}$	DONT EXIST	$2\|A_{ct'}\|C_p+4\mathbb{G}_\tau$	Threshold
Li [34]	$C_p+(2\|A_{ct}\|+3)\mathbb{G}+2\mathbb{G}_\tau$	$2(\|A_{ct'}\|-1)C_p+2\|A_{ct'}\|\mathbb{G}_\tau$	DONT EXIST	$2C_p+3\mathbb{G}_\tau$	(t,n)-Threshold
Lai [29]	$(8\|A_{ct}\|+10)\mathbb{G}+4\mathbb{G}_\tau+2H$	$4(\|A_{ct'}\|-2)C_p+(4\|A_{ct'}\|-2)\mathbb{G}_\tau$	$4(\|A_{ct'}\|-1)C_p+4A_{ct'}\mathbb{G}_\tau$	$4\mathbb{G}_\tau$	LSSS
Green [22]	$(4\|A_{ct}\|+1)\mathbb{G}+3\mathbb{G}_\tau+\|A_{ct}\|H$	$(\|A_{ct'}\|+2)C_p+3(\|A_{ct'}\|-1)\mathbb{G}+(\|A_{ct'}\|+1)\mathbb{G}_\tau$	DONT EXIST	$2\mathbb{G}_\tau$	LSSS
Li [32]	$C_p+(2\|A_{ct}\|+3)\mathbb{G}+2\mathbb{G}_\tau$	$2(\|A_{ct'}\|-1)C_p+2\|A_{ct'}\|\mathbb{G}_\tau$	DONT EXIST	$2\mathbb{G}_\tau$	Threshold
Jiguo 16 [33]	$(2\|2A_{ct'}\|+6)\mathbb{G}+3\mathbb{G}_\tau+2H$	$2C_p+2\mathbb{G}_\tau$	DONT EXIST	$6\mathbb{G}_\tau$	Tree
Lin [36]	$4(\|A_{ct'}\|+6)\mathbb{G}+\mathbb{G}_\tau+2H$	$2(\|A_{ct'}\|-1)C_p+2(\|A_{ct'}\|-1)\mathbb{G}_\tau$	$2(\|A_{ct'}\|-1)C_p+2A_{ct'}\mathbb{G}_\tau$	$\mathbb{G}_\tau$	LSSS
Mao [41]	$(2\|A_{ct'}\|+8)\mathbb{G}+3H$	$2(\|A_{ct'}\|-1)C_p+2(\|A_{ct'}\|-1)\mathbb{G}_\tau$	$2(\|A_{ct'}\|-1)C_p+2A_{ct'}\mathbb{G}_\tau$	$\mathbb{G}_\tau$	LSSS
Jiguo 17 [27]	$(2\|A_{ct'}\|+6)\mathbb{G}+4\mathbb{G}_\tau+2H$	$4C_p+2\mathbb{G}_\tau$	$4C_p+4\mathbb{G}_\tau$	$4\mathbb{G}_\tau$	non-monotonic AND gate
Zechao [39]	$(3\|A_{ct'}\|+1)\mathbb{G}+2H$	$4C_p+\mathbb{G}_\tau$	$4C_p$	$\mathbb{G}_\tau$	LSSS

Abbreviations: $|A_{ct}|$: Attributes size that belongs to original ciphertext, $|A_{ct'}|$: Attributes size that belongs to transformed ciphertext, C_p: Bilinear pairing operation, $\mathbb{G}$: Group, $\mathbb{G}_\tau$: Target group, H: Hash function.

simpler ciphertext by making use of attribute-based proxy re-encryption. According to Qinlong *et al.* [51], the user utilizing symmetric encryption algorithm encrypts data with random data encryption key. The data encryption key is then encrypted by employing access policy. To reduce computation cost, most of decryption operations are delegated to the CSP. To ensure the output returned from the service provider is correct, checkability is provided to guarantee accurateness of the outsourced/partial decrypted ciphertext. A major drawback to this scheme is inefficiency and its inflexibility when the owner of data for example hospital needs to select some but not all of the data are to be published by particular users. To overcome this, Weng *et al.* [57] proposed a scheme that insures that only ciphertexts which satisfies a stated condition can be re-encrypted. Keywords are the only conditions utilized in this scheme, on the other hand it is not practical in real life applications.

Realizing data privacy is the primary focus when designing any cryptographic scheme. To achieve both forward security and backward security, Xiao *et al.* [58] proposed a scheme that supports efficient outsourced decryption, user revocation and dynamic entry/exit of attribute authorities. In their schemes, user revocation is only related to revoked user. To improve efficiency of accessing remote data and preserve the privacy of the user's identity, Wang *et al.* [56] proposed the first anonymous distributed fine-grained access control protocol with verifiable outsourced decryption in the public cloud which is multi-authority in nature. Later, Camenisch *et al.* [10] proposed scheme is employed which in addition to anonymity employs pseudonym technique where legitimate user's public/private key pair corresponding to pseudonym is generated.

Generation of private key for user's policy can also increase computation overhead. To reduce such local overhead, Li *et al.* in [32] proposed a scheme where attribute authority can outsource partial private key generation to a key generation service provider (KGSP). From this scheme, constant efficiency is achieved at both attribute authority and user's end side. In order to avoid incorrect output, checkability is performed on the outcome returned from KGSP.

ABE computational overhead from exponentiation at user's end side can be relieved by adopting the traditional approach to utilize server-aided techniques [7,25]. But the notable common drawback to these schemes is that by directly utilizing this schemes in ABE, it may not work efficiently [32], and to mitigate this challenge, Zhou *et al.* [64] proposed the ABE scheme which allows secure outsourcing of both the encryption and decryption to cloud service providers.

To bring data close to the user, fog computing [9] was proposed which is an extension of cloud computing. The work related to CP-ABE in fog computing with outsourced decryption have been proposed [63,65]. A system with both outsourced encryption and decryption capabilities in fog computing using CP-ABE was proposed in [63]. In this scheme the workload operations of encryption and decryption are offloaded to the fog nodes.Therefore the computation operations on the data owner's side during encryption and also on users side during decryption are not relevant to the attributes size in the access structure and private keys respectively. Since the update concentrates only on the ciphertext associated with the corresponding updated attribute, the cost incurred by attribute update is minimal and hence efficient.

5 Performance and Security Analysis Comparisons

In this section, comparisons is made of the existing works to analyse the goals and the efficiency costs of the schemes. Table 1 features the comparisons of goals achieved by respective schemes, whereas Table 2, Table 3 and Table 4 highlights comparison of security models with their complexities, efficiency cost from storage perspective and computation efficiency cost for the corresponding schemes against one another respectively. We have used the following notations: N: Attribute size; $\mathbb{G}$: Group; $\mathbb{G}_\tau$: Target group; $|\mathbb{G}|$: bit length of an element in $\mathbb{G}$; $|\mathbb{G}_\tau|$: bit length of an element in $\mathbb{G}_\tau$; $|\mathbb{Z}_p|$: bit length of an element in $\mathbb{Z}_p$; H: Hash function; $|A_{ct}|$: Attributes size that belongs to original ciphertext; $|A_{ct'}|$: Attributes size that belongs to transformed ciphertext; C_p: Bilinear pairing operation.

From the output of Table 1 it can be seen that since each scheme is constructed to realize a given security goal, therefore none of the schemes can achieve all the goals concurrently. However, it shows that all the schemes supports fine-grained access, efficiency, unidirectionality, confidentiality and scalability characteristics while schemes [29, 32, 34, 41, 64] supports collusion resistance. Verifiability is supported by schemes [27,29,32,36,39,41] whereas schemes [33, 34] supports immediate revocation.

Regarding the comparisons of security models as depicted by Table 2, it shows that more than half of the models are selective except schemes [36, 64] which are adaptive. Selective security means the initialization phase comes prior to setup algorithm. In this case, the adversary initially provides the challenger with access structure.

Consequently, in Table 3, compared to other schemes Jiguo *et al.*'s scheme [27] is ideal in terms of key length, while Lin *et al.*'s scheme [36] and Mao *et al.*'s scheme [41] are ideal in terms of ciphertext length. While in Table 4, half of the schemes do not have decrypt algorithm and among their counterparts which have decrypt algorithms, Zechao *et al.*'s scheme [39] has an ideal computation cost.

6 Proposed Future Work

6.1 Accelerating the Efficiency of Attribute-Based Encryption Schemes without Using Outsourced Decryption

Nearly all the existing ABE schemes utilize bilinear pairings as a building block for a useful algorithm construction. However, bilinear pairing has high computational overhead, which makes algorithms complex, costly and therefore inefficient. Building pairing free algorithms or reducing the bilinear pairing size operations improves efficiency by simplifying computation complexity in resource-constrained devices. In addition, employing technologies like lattice to build an ABE scheme can also improve the computational efficiency for resource-constrained device users.

6.2 Reducing Communication Cost

In ABE with outsourced decryption, the resource-constrained users usually sends a transformation key to the unlimited-resource server (proxy) to simplify the ciphertext. The original message is then recovered from the simple ciphertext by the user. This increases the communication overhead between user and proxy. To minimize the overhead, ABE schemes that does not require the user to send the blinding (transformation) key to the proxy before performing final decryption should be built.

7 Conclusion

With the proliferation of mobile devices and development of easy to use application softwares, ABE schemes with outsourced decryption is gaining popularity due to advantages it has that supports devices with limited resource capabilities. By utilizing this primitive, encrypted eHealth big data stored in the unlimited resource cloud can now be accessed by resource-constrained devices where the user has to request a cloud server to perform heavy computations overhead on his/her behalf without learning about the plaintext of the data stored. This paper provides a survey of ABE schemes with outsourced decryption by reviewing the characteristics of eHealth big data, sources of eHealth big data, design structure, investigating adversary and security models and finally comparing efficiency costs for various existing schemes.

Lastly, based on this survey, the future work was proposed to provide roadmap to the solution of the problem encountered during outsourcing decryption.

Acknowledgments

The authors acknowledges the support by National Natural Science Foundation of China under Grant No. 61602097, Sichuan Science-Technology Support Plan Program under Grant Nos. 2016GZ0065, 2015JY0178, 2016ZC2575 and 17ZA0322, Fundamental Research Funds for the Central Universities under Grant No. ZYGX2015J072.

References

[1] J. Andreu-Perez, C. C. Y. Poon, R. D. Merrifield, S. T. C. Wong, and G. Z. Yang, "Big data for health," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, pp. 1193–1208, July 2015.

[2] D. Arora and P. Malik, "Analytics: Key to go from generating big data to deriving business value," in *IEEE First International Conference on Big Data Computing Service and Applications*, pp. 446–452, Mar. 2015.

[3] M. Asim, M. Petkovic, and T. Ignatenko, "Attribute-based encryption with encryption and decryption outsourcing," in *Department of Mathematics and Computer Science*, pp. 21-28, 2014.

[4] N. Attrapadung, B. Libert, and E. de Panafieu, *Expressive Key-Policy Attribute-Based Encryption with Constant-Size Ciphertexts*, pp. 90–108, 2011.

[5] A. Beimel, *Secure Schemes for Secret Sharing and Key Distribution*, Ph.D. dissertation,Technion-Israel Institute of technology, Faculty of computer science, 1996.

[6] J. Bethencourt, A. Sahai, and B. Waters, "Ciphertext-policy attribute-based encryption," in *IEEE Symposium on Security and Privacy (SP'07)*, pp. 321–334, May 2007.

[7] K. Bicakci and N. Baykal, *Server Assisted Signatures Revisited*, pp. 143–156, Springer Berlin Heidelberg, Berlin, Heidelberg, 2004.

[8] D. Boneh, X. Boyen, and E. J. Goh, *Hierarchical Identity Based Encryption with Constant Size Ciphertext*, pp. 440–456. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.

[9] F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, "Fog computing and its role in the internet of things," in *Proceedings of the First Edition of the MCC Workshop on Mobile Cloud Computing (MCC'12)*, pp. 13–16, 2012.

[10] J. Camenisch and A. Lysyanskaya, "An efficient system for non-transferable anonymous credentials with optional anonymity revocation," in *International Conference on the Theory and Applications of Cryptographic Techniques*, pp. 93-118, 2001.

[11] B. Chevallier-Mames, J. S. Coron, N. McCullagh, D. Naccache, and M. Scott, "Secure Delegation of Elliptic-Curve Pairing," in *Smart Card Research and Advanced Application*, vol. 6035, pp. 24–35, 2010.

[12] S. J. De and S. Ruj, "Decentralized access control on data in the cloud with fast encryption and outsourced decryption," in *IEEE Global Communications Conference (GLOBECOM'15)*, pp. 1–6, Dec. 2015.

[13] P. Lobato, de Faria and J. V. Cordeiro, "Health data privacy and confidentiality rights: Crisis or redemption?," *Revista Portuguesa de Saude Publica*, vol. 32, no. 2, pp. 123–133, 2014.

[14] H. Delfs, H. Knebl, and H. Knebl, "Introduction to cryptography," *Principles and Applications*, vol. 2, 2002.

[15] X. Dong, J. Yu, Y. Luo, Y. Chen, G. Xue, and M. Li, "Achieving an effective, scalable and privacy-preserving data sharing service in cloud computing," *Computers & Security*, vol. 42, pp. 151–164, 2014.

[16] K. Emura, A. Miyaji, A. Nomura, K. Omote, and M. Soshi, "A ciphertext-policy attribute-based encryption scheme with constant ciphertext length," in *Proceedings of the 5th International Conference on Information Security Practice and Experience (ISPEC'09)*, pp. 13–23, 2009.

[17] Gartner, "Gartner's three-part definition of big data," *Data Education for Business and IT Professionals*, 2013. (`http://www.dataversity.net/gartners-three-part-definition-of-big-data/`)

[18] I. Giurgiu, O. Riva, D. Juric, I. Krivulev, and G. Alonso, *Calling the Cloud: Enabling Mobile Phones as Interfaces to Cloud Applications*, pp. 83–102, Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.

[19] S. Goldwasser and S. Micali, "Probabilistic encryption and how to play mental poker keeping secret all partial information," in *Proceedings of the Fourteenth Annual ACM Symposium on Theory of Computing (STOC'82)*, pp. 365–377, 1982.

[20] V. Goyal, A. Jain, O. Pandey, and A. Sahai, *Bounded Ciphertext Policy Attribute Based Encryption*, Springer Berlin Heidelberg, Berlin, Heidelberg", pp. 579–591, 2008.

[21] V. Goyal, O. Pandey, A. Sahai, and B. Waters, "Attribute-based encryption for fine-grained access control of encrypted data," in *Proceedings of the 13th ACM Conference on Computer and Communications Security (CCS'06)*, pp. 89–98, 2006.

[22] M. Green, S. Hohenberger, and B. Waters, "Outsourcing the decryption of abe ciphertexts," in *Proceedings of the 20th USENIX Conference on Security (SEC'11)*, pp. 34–34, 2011.

[23] J. Hur and K. Kang, "Dependable and secure computing in medical information systems," *Computer Communications*, vol. 36, no. 1, pp. 20-28, 2012.

[24] IDC, "Big data meets big data analytics," *SAS*, 2012. (`https://www.datanami.com/whitepaper/big_data_meets_big_data_analytics/`)

[25] M. Jakobsson and S. Wetzel, *Secure Server-Aided Signature Generation*, pp. 383–401. Springer Berlin Heidelberg, Berlin, Heidelberg, 2001.

[26] J. G. Peter, "Medical informatics 20/20," *Medicine & Health Sciences*, (`http://www.hoise.com/vmw/08/articles/vmw/LV-VM-01-08-1.html`)

[27] L. Jiguo, S. Fengjie, Z. Yichen, X. Huang, and S. Jian, "Verifiable outsourced decryption of attribute-based encryption with constant ciphertext length," *Security and Communication Networks*, vol. 2017, 2017.

[28] J. Sun and C. Reddy, "Big data analytics for healthcare," *Big Data Analytics in Healthcare*, (`https://www.siam.org/meetings/sdm13/sun.pdf/`)

[29] J. Lai, R. H. Deng, C. Guan, and J. Weng, "Attribute-based encryption with verifiable outsourced decryption," *IEEE Transactions on Information Forensics and Security*, vol. 8, pp. 1343–1354, Aug. 2013.

[30] C. C. Lee, P. S. Chung, and M. S. Hwang, "A survey on attribute-based encryption schemes of access control in cloud environments," *International Journal Network Security*, vol. 15, pp. 231–240, 2013.

[31] A. Lewko, A. Sahai, and B. Waters, "Revocation systems with very small private keys," in *IEEE Symposium on Security and Privacy*, pp. 273–285, May 2010.

[32] J. Li, X. Huang, J. Li, X. Chen, and Y. Xiang, "Securely outsourcing attribute-based encryption with checkability," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, pp. 2201–2210, Aug. 2014.

[33] J. Li, W. Yao, Y. Zhang, H. Qian, and J. Han, "Flexible and fine-grained attribute-based data storage in cloud computing," *IEEE Transactions on Services Computing*, vol. 10, no. 5, pp. 785-796, 2016.

[34] J. Li, X. Chen, J. Li, C. Jia, J. Ma, and W. Lou, *Fine-Grained Access Control System Based on Outsourced Attribute-Based Encryption*, pp. 592–609, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.

[35] X. Liang, Z. Cao, H. Lin, and J. Shao, "Attribute based proxy re-encryption with delegating capabilities," in *Proceedings of the 4th International Symposium on Information, Computer, and Communications Security (ASIACCS'09)*, pp. 276–286, 2009.

[36] S. Lin, R. Zhang, H. Ma, and M. Wang, "Revisiting attribute-based encryption with verifiable outsourced decryption," *IEEE Transactions on Information Forensics and Security*, vol. 10, pp. 2119–2130, Oct. 2015.

[37] J. Liu, Y. Zhang, and C. Xing, "Medical big data web service management platform," in *IEEE 11th International Conference on Semantic Computing (ICSC'17)*, pp. 316–321, Jan. 2017.

[38] W. Liu and E. K. Park, "Big data as an e-health service," in *International Conference on Computing, Networking and Communications (ICNC'14)*, pp. 982–988, Feb. 2014.

[39] Z. Liu, Z. L. Jiang, X. Wang, X. Huang, S. M. Yiu, and K. Sadakane, "Offline/online attribute-based encryption with verifiable outsourced decryption," *Concurrency and Computation: Practice and Experience*, vol. 29, no. 7, pp. e3915, 2017.

[40] S. Luo, J. Hu, and Z. Chen, *Ciphertext Policy Attribute-Based Proxy Re-encryption*, pp. 401–415, Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.

[41] X. Mao, J. Lai, Q. Mei, K. Chen, and J. Weng, "Generic and efficient constructions of attribute-based encryption with verifiable outsourced decryption," *IEEE Transactions on Dependable and Secure Computing*, vol. 13, pp. 533–546, Sep. 2016.

[42] A. Mehmood, I. Natgunanathan, Y. Xiang, G. Hua, and S. Guo, "Protection of big data privacy," *IEEE Access*, vol. 4, pp. 1821–1834, 2016.

[43] B. Mounia and C. Habiba, "Big data privacy in healthcare moroccan context," *Procedia Computer Science*, vol. 63, pp. 575–580, 2015.

[44] M. Naor and M. Yung, "Public-key cryptosystems provably secure against chosen ciphertext attacks," in *Proceedings of the Twenty-Second Annual ACM Symposium on Theory of Computing (STOC'90)*, pp. 427–437, 1990.

[45] HealthIT.gov, *Guide to Privacy and Security of Electronic Health Information*, ver. 2, 2015. (`https://www.healthit.gov/sites/default/files/pdf/privacy/privacy-and-security-guide.pdf/`)

[46] I. Olaronke and O. Oluwaseun, "Big data in healthcare: Prospects, challenges and resolutions," in *Future Technologies Conference (FTC'16)*, pp. 1152–1157, Dec. 2016.

[47] R. Ostrovsky, A. Sahai, and B. Waters, "Attribute-based encryption with non-monotonic access structures," in *Proceedings of the 14th ACM Conference on Computer and Communications Security (CCS'07)*, pp. 195–203, 2007.

[48] R. Pankomera and D. V. Greunen, "Privacy and security issues for a patient-centric approach in public healthcare in a resource constrained setting," in *IST-Africa Week Conference*, pp. 1–10, May 2016.

[49] H. K. Patil and R. Seshadri, "Big data security and privacy issues in healthcare," in *IEEE International Congress on Big Data*, pp. 762–765, June 2014.

[50] J. Pavolotsky, "Demystifying big data," *Telecommunications Policy*, vol. 40, no. 9, pp. 837-854, 2016. (`http://breakinggov.sites.breakingmedia.com/wp-content/uploads/sites/4/2012/10/TechAmericaBigDataReport.pdf/`)

[51] H. Qinlong, M. Zhaofeng, Y. Yixian, N. Xinxin, and F. Jingyi, "Improving security and efciency for encrypted data sharing in online social networks," *China Communications*, vol. 11, pp. 104–117, Mar. 2014.

[52] C. Rackoff and D. R. Simon, "Non-interactive zero-knowledge proof of knowledge and chosen ciphertext attack," in *Proceedings of the 11th Annual International Cryptology Conference on Advances in Cryptology (CRYPTO'91)*, pp. 433–444, 1992.

[53] A. Sahai and B. Waters, *Fuzzy Identity-Based Encryption*, pp. 457–473, Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.

[54] Adi Shamir, *Identity-Based Cryptosystems and Signature Schemes*, pp. 47–53, Springer Berlin Heidelberg, Berlin, Heidelberg, 1985.

[55] J. Shao, Z. Cao, X. Liang, and H. Lin, "Proxy re-encryption with keyword search," *Information Sciences*, vol. 180, no. 13, pp. 2576–2587, 2010.

[56] H. Wang, D. He, and J. Han, "VOD-ADAC: Anonymous distributed fine-grained access control protocol with verifiable outsourced decryption in public cloud," *IEEE Transactions on Services Computing*, vol. 1, no. 99, pp. 1–1, 2017.

[57] J. Weng, R. H. Deng, X. Ding, C. K. Chu, and J. Lai, "Conditional proxy re-encryption secure against chosen-ciphertext attack," in *Proceedings of the 4th International Symposium on Information, Computer, and Communications Security (ASIACCS'09)*, pp. 322–332, 2009.

[58] M. Xiao, M. Wang, X. Liu, and J. Sun, "Efficient distributed access control for big data in clouds," in *IEEE Conference on Computer Communications Workshops*, pp. 202–207, Apr. 2015.

[59] H. Yadav and M. Dave, "Secure data storage operations with verifiable outsourced decryption for mobile cloud computing," in *International Conference on Recent Advances and Innovations in Engineering (ICRAIE'14)*, pp. 1–5, May 2014.

[60] K. Yang, X. Jia, K. Ren, B. Zhang, and R. Xie, "Dacmacs: Effective data access control for multiauthority cloud storage systems," *IEEE Transactions on Information Forensics and Security*, vol. 8, pp. 1790–1801, Nov. 2013.

[61] S. Yu, C. Wang, K. Ren, and W. Lou, "Achieving secure, scalable, and fine-grained data access control in cloud computing," in *Proceedings of IEEE*, pp. 1–9, Mar. 2010.

[62] J. Zhang and Z. Zhang, "A ciphertext policy attribute-based encryption scheme without pairings," in *Proceedings of the 7th International Conference on Information Security and Cryptology (INSCRYPT'11)*, pp. 324–340, 2012.

[63] P. Zhang, Z. Chen, J. K. Liu, K. Liang, and H. Liu, "An efficient access control scheme with outsourcing capability and attribute update for fog computing," *Future Generation Computer Systems*, vol. 78, no. 2, pp. 753-762, 2016.

[64] Z. Zhou and D. Huang, "Efficient and secure data storage operations for mobile cloud computing," in *Proceedings of the 8th International Conference on Network and Service Management (CNSM'12)*, pp. 37–45, 2013.

[65] C. Zuo, J. Shao, G. Wei, M. Xie, and M. Ji, "CCA-secure ABE with outsourced decryption for fog computing," *Future Generation Computer Systems*, vol. 78, no. 2, pp. 730-738, 2016.

Biography

Kittur Philemon Kibiwott is currently a Ph.D candidate at the University of Electronic Science and Technology of China (UESTC). He received Bsc degree in Computer Science from Periyar University (India) and Msc degree Computer Science from Bharathiar University (India). His research area include cloud computing, cryptography and information security.

Zhang Fengli received her Ph.D degree from the University of Electronic Science and Technology of China (UESTC) in 2007 and M.S. degree in 1986. She is currently a Professor at the University of Electronic Science and Technology of China (UESTC). She has published more than eighty papers in refereed international journals and conferences which more than 50 are indexed by SCI and EI. Her research area of interest include mobile data management and application, network security, database.

Kimeli Victor K. is currently Dean of the School of Science at the University of Eldoret (UoE). He obtained his Ph.D from IHIT, China and M.Sc degree from Essex. His research area of interest include Databases, E-commerce, information security and Sensor Networks.

Omala A. Anyembe is currently a Ph.D candidate in the school of Computer Science and Engineering at the University of Electronic Science and Technology of China (UESTC). His research area of interest include cryptography, IoT and information security.

Eugene Opoku-Mensah is currently a Ph.D candidate in the school of Information and Software Engineering at the University of Electronic Science and Technology of China (UESTC). His research area include privacy preservation in cloud computing, Web page Ranking, and data mining.

Achieving an Effective, Confidentiality and Integrity of Data in Cloud Computing

Ghassan Sabeeh Mahmood[1,2], Dong Jun Huang[1], Baidaa Abdulrahman Jaleel[2]

(Corresponding author: Ghassan Sabeeh Mahmood)

School of Information Science and Engineering, Central South University[1]
Changsha 410083, Hunan, China
Computer Science Department, College of Science, University of Diyala, Iraq[2]
(Email: ghassan.programer@gmail.com)

(Received Nov. 3, 2017; revised and accepted Apr. 18, 2018)

Abstract

Cloud computing allows users to store their data remotely. Users can enjoy cloud applications on-demand without the burden of maintaining personal hardware and managing software. Although its advantages are clear, cloud storage requires users to relinquish physical possession of data, and thus, it poses security risks with regard to the correctness of data. In this paper, we propose a new cloud scheme to enhance data security, thereby addressing the aforementioned issue whilst achieving a secure cloud storage service and dependability. A secret image is encrypted by using the Advanced Encryption Standard (AES) algorithm. Then, the encrypted image is embedded into the host image via a steganography technique, which combines Discrete Wavelet Transform (DWT) and Singular Value Decomposition (SVD) to obtain the stego image. To preserve data integrity, a hash value is generated for the stego image using the Secure Hash Algorithm 2 (SHA-2) prior to storing the image in the cloud. After the image is retrieved from the cloud, its hash value is generated using the same algorithm (i.e. SHA-2). Both hash values are then compared to verify whether the data stored in the cloud are changing and to obtain the secret image. The proposed scheme is proven to be secure and highly efficient through an extensive security and performance analysis.

Keywords: Cloud Computing; Cryptography; Steganography; Hash Function

1 Introduction

The cloud computing paradigm allows on-demand network access to a shared set of computing resources (e.g. storage, servers, networks, services and applications) that can be provided immediately [3]. Cloud computing is characterized by five important features, three service models and four deployment models [6]. Its important features are wide network access, location-independent resource pooling, on-demand service, measured service and rapid resource elasticity. Meanwhile, the service models are software as a service, infrastructure as a service and platform as a service, whereas the deployment models include a public cloud, private cloud, hybrid cloud and community cloud [4].

Enterprises and individuals can use the data centre of the cloud for storage without additional burden. Data can be stored and accessed remotely anywhere and anytime. Users can be relieved of the burden of storing and maintaining local information through data outsourcing [17]. However, security issues are key concerns in the cloud, which limit its adoption among organizations. Traditional mechanisms for handling security issues are unsuitable for cloud storage due to its virtual nature [2].

Therefore, the privacy, integrity, security and confidentiality of stored data should be considered in cloud computing. Novel methods should be developed and applied to fulfil all the aforementioned requirements. The best approach is to encrypt data before outsourcing them to cloud computing. For example, the owner allows outsiders to see the outline of his/her data, but only authorized users can recover these data. Such robust demands necessitate the search for encryption solutions for multimedia [19].

Steganography is used with cryptography to verify the confidentiality of data. In this special branch of data hiding, a message is embedded into a cover image based on a shared key, thereby producing a stego image [8]. Steganography methods can be grouped into spatial domain and transform domain methods. In spatial domain methods, the original image levels are modified to encode the secret information. Although these methods achieve a higher payload, they are weak to image processing manipulations and statistical attacks, including image compression, image cropping and noise attacks. In transform domain methods, the image is first changed from the spatial domain to the frequency domain. Then, the image coefficients are altered to hide secret data. Transform domain methods have a lower payload than spatial domain methods, but are robust against statistical attacks. Examples of these methods are discrete wavelet transform (DWT), discrete Fourier transform and discrete cosine transform [11].

Cryptography and steganography work hand-in-hand. A message is scrambled via cryptography, such that it cannot be understood. Then, steganography is performed to hide the message and make it invisible. For example, an encrypted message may arouse the suspicion of the receiver, whereas an imperceptible message will not. Steganography can be useful when using cryptography is illegal. Under such condition, steganography can enable secretly sending a message. However, the manner in which cryptography and steganography are evaluated varies. Cryptography fails when the 'enemy' notices that a message exists in the steganography medium; by contrast, steganography is considered a failure when the 'enemy' is able to reveal the content of the encrypted message [10].

In addition to data confidentiality, integrity is also a key issue in cloud computing. Data can either be manipulated or lost due to accidental or intentional malicious activities, which can be terrifying for the user and embarrassing for the cloud service provider. The cloud provides 'multi-tenancy'; that is, cloud resources will be shared and utilized by multiple users. Consequently, adversaries can take advantage of the vulnerabilities in the cloud. Administration errors, such as failures in data migration or backup/restore process, can also damage data. Accordingly, data integrity is a core issue in outsourcing data over cloud storage [21].

In the current paper, a novel secure cloud storage system is proposed to ensure high data confidentiality and integrity levels. The Advanced Encryption Standard (AES) method is used to encrypt a secret image. Then, the encrypted image is embedded into the cover image using the hybrid steganography scheme DWT - singular value decomposition (SVD) to get the stego image and verify the confidentiality of the data. Thereafter, a hash value for the stego image is generated using the Secure Hash Algorithm 2 (SHA-2) before the stego image is stored in the cloud to maintain data integrity. After the image is retrieved from the cloud, the same algorithm (i.e. SHA-2) is used to generate its hash value. Both hash values are then compared via a verification process to validate whether the data stored in the cloud are altered and to obtain the secret image. The novel contributions of this paper are as follows.

1) An image is decomposed into four frequency subbands (LL, LH, HL and HH) using DWT in information hiding. The HL frequency sub-band, which represents mid-frequencies, is selected. This sub-band is robust against various geometric and filtering noises. Therefore, inserting the secret image into the HL subband does not change the original image data and the appearance of the image is maintained at a high level.

2) The SVD of an image provides three singular matrices (U, S and V). S is a diagonal matrix, whereas U and V are orthogonal matrices. The secret image information will be inserted into the singular values in the S matrix of the original image. The original image will not be misrepresented, even if the singular values are altered. Consequently, the secret image is inserted into the original image using SVD.

3) A secure and efficient scheme that can achieve data confidentiality is developed using the AES algorithm.

4) An efficient data integrity verification process is proposed for this scheme using the SHA-2 hash function.

The remaining parts of the paper are organized as follows. Preliminaries regarding the study are provided in Section 2. Related works are presented in Section 3. The proposed scheme is described in detail in Section 4. The experimental results are discussed in Section 5. Finally, concluding remarks for the paper are given in Section 6.

2 Preliminaries

2.1 SVD

SVD is used in various image-processing applications, including stenography, image watermarking and data compression. It is also adopted to solve various mathematical problems [13]. Matrix SVD is decomposed into three matrices (U, S and V). S is a diagonal matrix, whereas U and V are right and left singular matrices. The singular S matrix includes intensity-related image information. The orthogonal U and V matrices comprise geometric image information. The equation for the decomposition of matrix SVD is as follows:

$$SVD = s_1 U_I V_I^T + s_2 U_2 V_2^T + \cdots + s_r U_r, \qquad (1)$$

where the rank of matrix SVD is indicated by r. U_I, U_2, $\cdots$, U_r and V_I, V_2, $\cdots$, V_r are the columns of the left and right singular values, respectively; whilst s_1, s_2, $\cdots$, s_r are the scalar singular values of the diagonal matrix [7].

2.2 DWT

DWT has recently received considerable attention in various signal-processing applications, including image steganography, because of its capability to provide the necessary data for the analysis and synthesis of signals and to reduce computation time. DWT can detect portions of the host image where secret data are successfully hidden. DWT exhibits an advantage over other approaches because it allows the signal to be reconstructed by applying inverse DWT to frequency bands [16]. DWT is a frequency domain technique. In this approach, the cover image is first transformed into the frequency domain. Then, its frequency coefficients are modified according to the transformed coefficients. DWT hierarchically decomposes an image in single-level decomposition, thereby providing the spatial and frequency descriptions of the image. The image is decomposed into three directions: diagonal, vertical and horizontal. DWT then decomposes the image into four frequency bands: LL, HL, LH and HH. LL represents low-frequency bands, HL and LH represent mid-frequency bands and HH represents high-frequency bands. The LL band presents approximate details, the HL band gives horizontal details, the LH band provides vertical details and the HH band highlights the diagonal details of an image [12].

2.3 Cryptography Algorithms

2.3.1 AES

A user can upload his/her personal data and share them with others in cloud computing. However, if privacy is not secure, then users may not use this cloud service even if the demand is strong [14]. To guarantee data protection in cloud computing, cryptography techniques are adopted as common solutions [15]. Among these techniques, AES is considered the block encryption standard. An AES encryption system is symmetric. This algorithm has different key lengths, i.e. 128, 196 and 256 bits. Packet size is 128 bits. The AES algorithm exhibits good flexibility, and thus, it is extensively used in various hardware and software.

2.3.2 Cryptographic Hash Functions

Cryptographic hash functions are fundamental tools in modern cryptography. These tools are used to ensure data integrity when information is transferred over insecure networks. The Secure Hash Algorithm (SHA) is considered one of the best cryptography hash functions [9]. SHA, which was developed by the National Security Agency, is typically divided into three sub-families: SHA-0, SHA-1 and SHA-2. Data are organized into blocks of bits during hashing with SHA. The number of bits is locked for a specific algorithm. In particular, the SHA-0 and SHA-1 families divide data into 512 bit blocks for processing. By contrast, the algorithm used by the SHA-2 family has varying digest sizes, which are distinguished as SHA-224, SHA-256, SHA-384 and SHA-512. The processing block bit size is variable for the SHA-2 family. In particular, the processing block size of the SHA-224 and SHA-256 sub-families are 512 bits, whereas that of the SHA-384 and SHA-512 sub-families are 1024 bits [18]. This paper uses SHA-512 to guarantee data integrity when transferring information over unprotected networks.

3 Related Works

El-Makkaoui *et al.* [5] presented an enhanced encryption scheme, called Cloud (RSA), based on the Rivest-Shamir-Adleman (RSA) algorithm. Cloud (RSA) uses two discrete keys: evaluation and private keys. The evaluation key $ev = (M)$ is used to implement operations on encrypted data through a third party. The private key $pr = (M, e, k)$, which is known only to the data owner, is used to encrypt and decrypt data. The safety of the private key is based on two factors:

1) The problem of determining the prime factorization of (M);

2) The e^{th} root problem of Cloud (RSA).

Even if the factorization of (M) is given, decrypting the ciphertext encrypted using the Cloud (RSA) encryption scheme is extremely difficult because (e and k) are private. Mandal *et al.* [10] proposed a crypto-stego method, in which the steganography technique embedded private data by using a pixel-mapping method. The encryption and decryption process uses a genetic algorithm, which features crossover and mutation operations. Cryptography and steganography also use a secret key, which is generated by combining certain features of the cover image and the secret key of the user. Bhandari *et al.* [1] proposed a scheme called hybrid encryption (RSA) along with AES by enhancing the security standard of the RSA algorithm. Wang *et al.* [19] presented degradation and encryption techniques for Portable Network Graphics (PNG). In particular, the prefix and noise generation techniques were improved for PNG degradation. In addition, a modified generalized Feistel scheme was developed for encrypting PNG.

Although existing systems have achieved confidentiality, they remain unsuccessful in preserving data integrity. Consequently, a secure system should be developed to achieve effective performance by maintaining confidentiality with data integrity.

4 Proposed Scheme

The basic concept of the proposed scheme is described in Section 4.1. The encrypted secret image is presented in Section 4.2. The steganography method is discussed in Section 4.3. Finally, integrity check using the SHA-512 hash function is presented in Section 4.4.

4.1 Basic Concept

Once log in is successful, the data owner will select the secret image and store it on the cloud server. The secret image selected by the owner will be encrypted using the AES algorithm. Then, the encrypted image will be embedded into the cover image using the hybrid steganography scheme DWT-SVD to get the stego image. Thereafter, SHA-2 is used to generate the hash value of the stego image before it is stored in the cloud to maintain data integrity. The hash value of the image is also generated using SHA-2 after the image is retrieved from the cloud. Both hash values are compared using the verification process to validate whether the data stored in the cloud are altered; then, the secret image is obtained. The proposed system is illustrated in Figures 1 and 2.

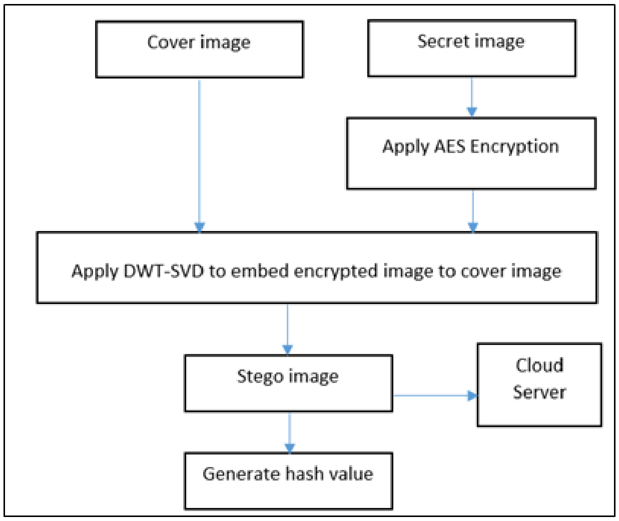


Figure 1: Process of the encrypting and embedding algorithm

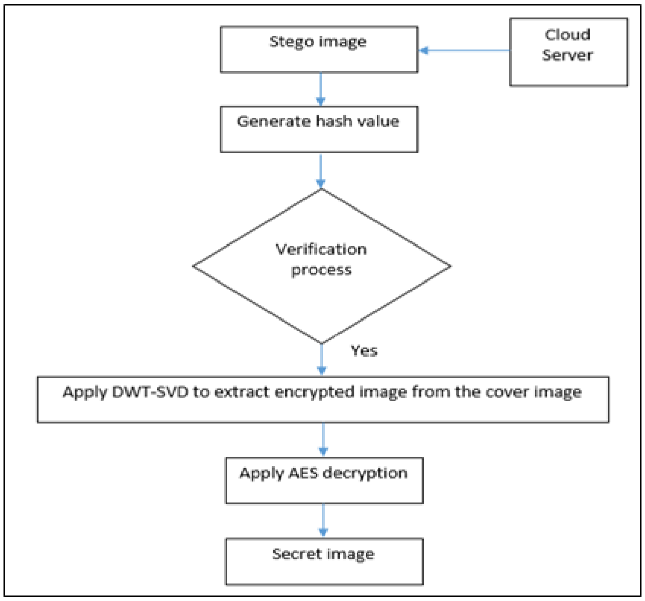


Figure 2: Process of retrieving the secret image algorithm

4.2 Secret Image Encryption

The colour image comprises a set of pixels. Each pixel has three main components: red (R), green (G) and blue (B). Each component is represented by 8 bits. The colour components of the secret image are individually encrypted. All the RGB components are mixed to produce the colour image. The encryption and decryption algorithms of the secret image are presented as in Algorithms 1 and 2.

Algorithm 1 Encryption process

1: The colour components (R, G and B) of the secret image are extracted.
2: The AES algorithm and different keys are used to encrypt each colour component.
3: All the components are combined to obtain the final encrypted image.

Algorithm 2 Decryption process

1: To extract the encrypted image, the stego image is retrieved from the cloud and further decomposed into different colour components.
2: The AES algorithm and the respective keys are used to decrypt the colour components.
3: All the components are combined to obtain the decrypted image.

4.3 DWT-SVD-based Image Steganography

The algorithms used for the DWT-SVD-based image steganography scheme are presented as in Algorithms 3 and 4.

Algorithm 3 Embedding Algorithm

1: The cover and encrypted images are decomposed into sub-bands using DWT.
2: SVD is performed on the HL sub-band to transform the cover and encrypted images.
3: The encrypted image is embedded into the host image.
4: Inverse SVD is performed on the embedded image.
5: Finally, inverse DWT is applied to get the stego image.

Algorithm 4 Extraction Algorithm

1: The stego image using DWT is decomposed into subbands.
2: SVD is performed on the HL sub-band of the decomposed stego image.
3: Extraction is applied to the resultant SVD image.
4: Inverse SVD is performed on the resultant image.
5: Finally, inverse DWT is performed to get the encrypted image.

4.4 Integrity Check Using the SHA-512 Hash Function

The SHA-512 hash function is used to eliminate the clash between two hash values to achieve data integrity. Firstly, the hash value of the stego image is precomputed. Subsequently, the stego image is sent to the cloud and the computed hash value is stored in the local repository. When the clients want to verify data integrity, the file is retrieved from the cloud and the hash value of this file is recomputed. Then, the values are matched. The file is intact if the precomputed and recomputed hash values match. If these values do not match, then the file has been tampered with and its integrity has been compromised. The algorithm of data integrity is described as in Algorithm 5.

```
Algorithm 5 Data Integrity Algorithm
1: The stego image is sent to the cloud after computing its hash value.
2: The computed hash value of the stego image is stored in the secured local repository.
3: After the stego image is downloaded from the cloud, its hash value is recomputed.
4: The hash values are matched to obtain data integrity.
```

5 Experimental Results

The images used in this experiment are offered in Section 5.1. The results of the encryption-based AES algorithm are discussed in Section 5.2. Finally, the robustness test for the proposed scheme is explained in Section 5.3.

5.1 Cover and Secret Images

The sizes of the cover and secret images used in the experiments are 512×512 and 256×256, respectively. The original and secret images are shown in Figures 3(a) and 3(b), respectively.

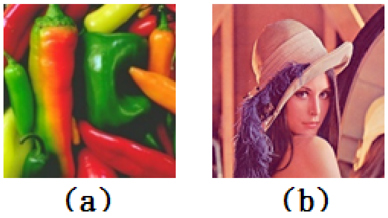


Figure 3: Cover and secret images

Several quality measures, such as peak signal-to-noise ratio (PSNR), mean square error (MSE) and normalized correlation (NC), are used to evaluate the performance of the stego and extracted images [12].

PSNR is a metric used to check the perceptual similarity between the original and stego images. It can be defined as follows:

$$PSNR = 10 \log \frac{255^2}{MSE}, \tag{2}$$

where MSE is calculated between the host image A and the stego image A_s as follows:

$$MSE = \frac{1}{MM} \sum_{i=1}^{M} \sum_{j=1}^{M} (A - A_s)^2. \tag{3}$$

The stego image appears nearly identical to the host image when good imperceptibility is achieved. That is, the host image is unaffected by the embedding process. A PSNR above 40 dB indicates good perceptual fidelity. In the experiment, PSNR is above 40 dB, thereby indicating the effectiveness of the proposed scheme.

NC is used to evaluate the feasibility of the extracted secret image. The similarity between secret images is represented by the number of mismatched data between the inserted and extracted secret images. NC for valid secret images, which represents the characteristics of the extracted secret image, is defined as

$$corr(d, d^*) = \frac{\sum_{i=1}^{N} (d_i - \bar{d})(d_i^* - \bar{d})}{\sqrt{\sum_{i=1}^{N} (d_i - \bar{d})} \sqrt{\sum_{i=1}^{N} (d_i^* - \bar{d})}}, \tag{4}$$

where $(d_i), d_i^*)$ are the original and modified data, whilst d is the mean of the original data.

5.2 Results of the Encryption-based AES Algorithm

The image encryption process using the AES of the secret image obtained as a colour image is offered in Figure 4(a). The encrypted image is produced by combining all the colour components, as shown in Figure 4(b). In Figure 4(c), the decrypted image based on the AES algorithm is shown.

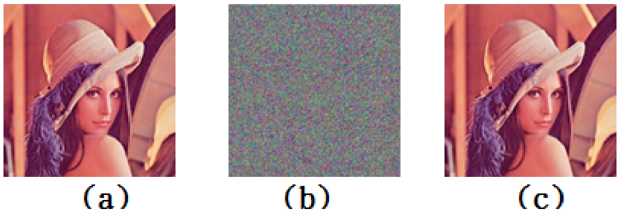


Figure 4: Encrypted and decrypted secret image

The response time of the cryptographic performance in terms of encryption and decryption is highlighted in Table 1.

Table 1: Cryptographic performance

Size (KB)	Response time (s)	
	Encryption	Decryption performance
256	0.4375	0.5227

The preceding experiment showed that the speed of the cryptographic performance depends on the response time of the encryption and decryption processes. In addition, the results demonstrate that the decrypted image is similar to the secret image, and thus, the AES algorithm performs effectively. This algorithm also exhibits good manoeuvrability for image encryption based on this finding.

5.3 Robustness Test of the Proposed Method

The stego and extracted images are shown in Figures 5(a) and 5(b), respectively. The NC of the extracted image is 0.9968.

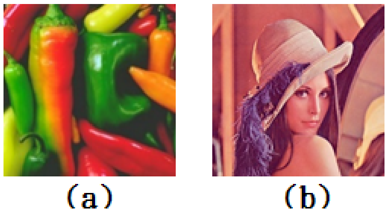


Figure 5: Stego and extracted images

The reliability test for the proposed method is illustrated in Figure 6. The extracted secret image is shown in Figure 6(b) if the fruit image shown in Figure 6(a) is used for detection. Therefore, the secret image cannot be detected using a random reference image.

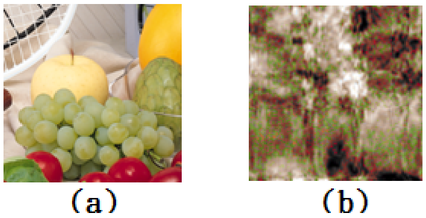


Figure 6: Reliability test

Tables 2 and 3 present the comparison results between the NC and PSNR of the proposed scheme and that of pure SVD. The proposed scheme achieves better result than pure SVD for numerous attacks, including Gaussian noise m = 0, v = 0.001; speckle; compression QF 60%; rotation by 10 (clockwise); shifting attack and average filtering.

Our scheme exhibits stronger anti-interference performance and higher stability than pure SVD when facing various malicious attacks. Efficiency in terms of computation time for embedding and extraction (in seconds) is presented in Table 4.

6 Conclusion

The security of data stored in the cloud is a significant issue. Cryptography techniques have been used in cloud computing to guarantee the confidentiality of private data. However, attackers have numerous chances to break through the security provided by cryptography techniques. In this work, a data security system that combines cryptography and steganography techniques is presented to achieve multi-layer security. Firstly, the AES encryption method is used to encrypt the secret image. Secondly, the hybrid steganography scheme SVD-DWT is applied to hide the encrypted secret image within the cover image to ensure the confidentiality of the data. Thirdly, a hash algorithm is used for the hidden file before and after it is downloaded from the cloud to verify data integrity. As shown in the simulation results, the proposed system provides high-quality image in terms of PSNR. In addition, the system reduces suspicion over the presence of hidden information in an image.

References

[1] A. Bhandari, A. Gupta, and Debasis Das, “Secure algorithm for cloud computing and its applications,” in *6th International Conference Cloud System and Big Data Engineering (Confluence’16)*, IEEE, 2016.

[2] S. Cherillath Sukumaran, M. Mohammed, “DNA cryptography for secure data storage in cloud,” *International Journal of Network Security*, vol. 20, no. 3, pp. 447-454, 2018.

[3] E. F. Coutinho, F. R. de C. Sousa, P. A. L. Rego, D. G. Gomes, J. N. de Souza, “Elasticity in cloud computing: A survey,” *Annals of Telecommunications*, vol. 70, no. 7-8, pp. 289-309, 2015.

[4] S. A. El-Booz, G. Attiya, and N. El-Fishawy, “A secure cloud storage system combining time-based one-time password and automatic blocker protocol,” *EURASIP Journal on Information Security*, 2016.

[5] K. El-Makkaoui, A. Ezzati, and A. Beni-Hssane, “Cloud-RSA: An enhanced homomorphic encryption scheme,” in *Europe and MENA Cooperation Advances in Information and Communication Technologies*, pp. 471-480, Springer, 2017.

[6] S. E. Elgazzar, A. A. Saleh, H. M. El-Bakry, “Overview of using private cloud model with GIS,” *International Journal of Electronics and Information Engineering*, vol. 7, no. 2, pp. 68-78, Dec. 2017.

[7] B. L. Gunjal, S. Mali, “MEO based secured, robust, high capacity and perceptual quality image watermarking in DWT-SVD domain,” *SpringerPlus*, vol. 4, no. 1, Dec. 2015.

[8] L. C. Huang, L. Y. Tseng, 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, Mar. 2013.

[9] S. F. Lu, H. Ali, and O. Farooq, “Proposed approach of digital signature technology for building a web security system based on SHA-2, MRC6 and ECDSA,” in *2nd International Conference on Information Technology and Industrial Automation (ICITIA’17)*, pp. 254-261, 2017.

[10] S. Mandal and S. Bhattacharyya, “Secret data sharing in cloud environment using steganography and encryption using GA,” in *International Conference on Green Computing and Internet of Things*, pp. 1469-1474, 2015.

[11] K. Muhammad, M. Sajjad, I. Mehmood, S. Rho and S. Baik, “A novel magic LSB substitution method

Table 2: Comparison of PSNR values

Techniques	Attacks						
	No attack	Gaussian noise	Speckle	Compression	Rotation	Shifting	Average filtering
Proposed Method	47.6819	39.8514	37.5561	45.7602	42.2842	35.2093	35.9638
Pure SVD	39.4539	29.7291	31.7207	38.9193	26.9063	29.7628	28.8793

Table 3: Comparison of NC values

Techniques	Attacks						
	No attack	Gaussian noise	Speckle	Compression	Rotation	Shifting	Average filtering
Proposed Method	0.9968	0.0176	0.0421	0.0194	0.0135	0.0102	0.0209
Pure SVD	0.9319	0.0217	0.0513	0.0394	0.0329	0.0371	0.0412

Table 4: Embedding and extraction time (in seconds)

Size	Embedding time (s)	Extraction time (s)
256 KB	1.123821	1.456813

(M-LSB-SM) using multi-level encryption and achromatic component of an image," *Multimedia Tools and Applications*, vol. 75, no. 22, pp. 14867-14893, 2016.

[12] N. Narula, D. Sethi, and P. P. Bhattacharya, "Comparative analysis of DWT and DWT-SVD watermarking techniques in RGB images," *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 8, no. 4, pp. 339-348, 2015.

[13] R. Nouri, A. Mansouri, "Digital image steganalysis based on the reciprocal singular value curve," *Multimedia Tools and Applications*, vol. 76, no. 6, pp. 8745-8756, 2017.

[14] M. Ouedraogo, S. Mignon, H. Cholez, S. Furnell, and E. Dubois, "Security transparency: the next frontier for security research in the cloud," *Journal of Cloud Computing*, 2015.

[15] S. Rajput, J. S. Dhobi, and L. Gadhavi, "Enhancing data security using aes encryption algorithm in cloud computing," in *Proceedings of First International Conference on Information and Communication Technology for Intelligent Systems*, vol. 2, Springer, 2016.

[16] P. Ramu, R. Swaminathan, "Imperceptibility-Robustness tradeoff studies for ECG steganography using continuous ant colony optimization," *Expert Systems with Applications*, vol. 49, pp. 123-135, 2016.

[17] M. Y. Shabir, A. Iqbal, Z. Mahmood, and A. Ghafoor, "Analysis of classical encryption techniques in cloud computing," *Tsinghua Science and Technology*, vol. 21, no. 1, pp. 102-113, 2016.

[18] D. W. Walker, C. Mackey, *Secure Hashing Device Using Multiple Different SHA Variants and Related Methods*, U.S. Patent 9, 680, 637, issued June 13, 2017.

[19] Y. Wang, J. Du, X. Cheng, Z. Liu and K. Lin, "Degradation and encryption for outsourced PNG images in cloud storage," *International Journal of Grid and Utility Computing*, vol. 7, no. 1, pp. 22-28, 2016.

[20] Y. Wang, J. Du, X. Cheng, Z. Liu, K. Lin, "Degradation and encryption for outsourced PNG images in cloud storage," *International Journal of Grid and Utility Computing*, vol. 7, no. 1, pp. 22-28, 2016.

[21] F. Zafar, A. Khan, S. Malik, *et al.*, "A survey of cloud computing data integrity schemes: Design challenges, taxonomy and future trends," *Computers & Security*, vol. 65, pp. 29-49, 2017.

Biography

Ghassan Sabeeh Mahmood received his M.S. degree in 2015 from School of Information Science and Engineering, Central South University, China. His current research interests include security of cloud computing.

Dong Jun Huang is Professor in School of Information Science and Engineering, Central South University, China. His current research interests include image processing, communication and content analysis.

Baidaa Abdulrahman Jaleel received her B.S. degree in 2007 from College of Science, Diyala University, Iraq. Her current research interests include image processing and security of cloud computing.

AODVDC: An Improved Protocol Prevents Whirlwind Attacks in Mobile Ad hoc Network

Luong Thai Ngoc[1,2], Vo Thanh Tu[1]

(Corresponding author: Luong Thai Ngoc)

Faculty of Information Technology, Hue University of Sciences, Hue University; Viet Nam[1]
77 Nguyen Hue street, Hue city, Vietnam
Faculty of Mathematics and Informatics Teacher Education, Dong Thap University; Viet Nam[2]
783 Pham Huu Lau street, Ward 6, Cao Lanh city, Dong Thap, Viet Nam
(Email: ltngoc@dthu.edu.vn, vttu@hueuni.edu.vn)

(Received Oct. 28, 2017; revised and accepted Apr. 13, 2018)

Abstract

Ad hoc On-demand Distance Vector routing protocol is one of the most popular reactive protocol used for Mobile Ad hoc Network, is target of many denial-of-Service attack types. Whirlwind attacks uses a malicious node to make one routing-loop on the discovered route. All data packets are dropped due to they over time-life. This article proposes a mechanisms to manage and provide digital certificates (DC) for Mobile Ad hoc Network (MANET) without public key infrastructure. A digital certificates authentication mechanism secure that only "friendly" nodes to collaborate in the route discovery process, goal is to prevent malicious nodes that joined the discovered route, such as Whirlwind. A new routing protocol named AODVDC by integrating our solutions into AODV routing protocol. Using NS2, we evaluate the security performance using scenario where there are nodes move ramdomly and Whirlwind attacks, compared with related protocols. The simulation results showed that our approach has better performance in terms of packet delivery ratio, routing load and route discovery delay compared to related works under attack scenario.

Keywords: AODV; AODVDC; MANET; Network Security; Whirlwind Attacks

1 Introduction

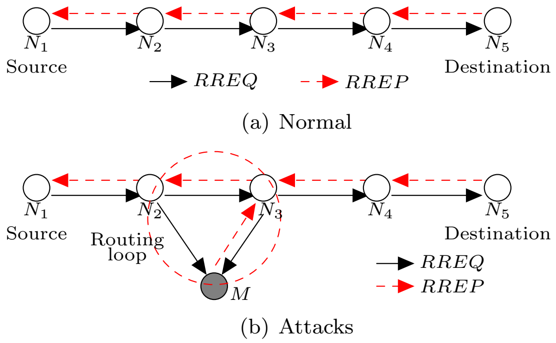
Mobile Ad hoc Network (MANET [8]) is a special wireless, the advantages such as flexibility, mobility, every mobile node acts both as a host and as a router. Routing is the main service provided in network layer, the source node using the route to the destination is discovered and maintained. There are many routing protocols are recommended to MANET, they are classified into proactive, reactive, and hybrid routing [2]. Ad hoc On-demand Distance Vector (AODV [16]) routing protocol is one of the most popular reactive protocol used for Mobile Ad hoc Network, is target of many denial-of-Service attack types [17], such as Blackhole [4, 9], Sinkhole [5], Grayhole [7], Flooding [20], Wormhole [3] and Whirlwind [15].

We focus on Whirlwind attacks type and prevention solution, this attack type target is to make routing-loop which is done with two phases:

Phase 1: Malicious nodes try to set up a routing-loop in the discovered route from source to destination node when receiving route request packet (RREQ) from any source node N_S by using the fake route reply packet (FRREP).

Phase 2: If attacking is successful, all data packets from source to destination node are taken into data whirlwind and automatically dropped due to over time-life.

In Figure 1(a), source node N_1 discovers a new route to destination node N_5 by broadcasting of RREQ to its neighbor nodes named N_2. Intermediate node N_2 is not destination node, it therefore continue broadcasts RREQ packet to its neighbors named N_3 and save reserve route to source N_1, this process repeats at N_3 and N_4 until node N_5 receives the route request packet. When receiving RREQ packet from node N_4, destination node N_5 sends unicast of RREP packet to source on route $\{N_5 \rightarrow N_4 \rightarrow N_3 \rightarrow N_2 \rightarrow N_1\}$. As a result, source node N_1 discovers route to destination in following direction $\{N_1 \rightarrow N_2 \rightarrow N_3 \rightarrow N_4 \rightarrow N_5\}$. Figure 1(b) shows that malicious node M appears in network topology for Whirlwind attack behavior, it is neighbor of both N_2 and N_3 nodes. When receiving the first RREQ packet from node N_2, M adds a entry to destination into its routing table (RT) with minimum cost and next hop (NH) is N_2. When receiving the second RREQ packet from N_3, M adds a entry to source N_1 into its RT with lowest cost and NH is N_3, concurrently sends unicast of FRREP to source N_1 in direction $\{M \rightarrow N_3 \rightarrow N_2 \rightarrow N_1\}$. As a result, N_3's RT has route information to destination via *NH* is M with the lowest cost. The destination node N_5 also sends a RREP packet to source node on direction $\{N_5 \rightarrow N_4 \rightarrow N_3 \rightarrow N_2 \rightarrow N_1\}$. When receiving the RREP packet from node N_4, node N_3 see that the cost to destination is not cheaper than the existing route, the RREP packet is therefore dropped. The results is exist routing-loop on discovered route from N_1 to N_5 including nodes named N_2, N_3, and M. All data packets from N_1 to N_5 node are taken into data whirlwind and automatically dropped due to over time-life.



(a) Normal

(b) Attacks

Figure 1: Description of whirlwind attacks [15]

This article proposes a mechanisms to manage and provide digital certificates for MANET without public key infrastructure (PKI) because MANET is no infrastructure. In addtion, digital certificates authentications mechanism allows only "friendly" node to collaborate in the route discovery process. The remainder of this article is structured as following: In the next Section, we review some related works for security base on digital signature. Section 3 mechanism to manage and provide the digital certificate. Section 4 shows how to authenticate preceding node's DC when an node receiving the control route packets. Section 5 shows the evaluation results by simulation; Finally, conclusions and future works.

2 Related Works

There are some related works to increase security level for AODV routing protocol based on digital signature or one-way hash [12]. Zhou [24] described a solution to distribute the CA role among n nodes of the network using *(n, k + 1)* threshold cryptography scheme. In this scheme the private key is divided into n partial shares $(S_1, S_2, \cdots, S_n)$ where at least*k+1* of n are partial shares which are needed to generate a secret S. The advantage is its increased availability, since any *k+1* among n nodes in the local neighborhood of the requesting node can issue or renew a certificate. And any node, which does not have a private share yet, can obtain a share from any group of at least *k + 1* nodes which has already a share [1].

Zhang described a solution named IKM (id-based key management) as a novel combination of ID-based and threshold cryptography. IKM is a certificateless solution in that public keys of mobile nodes are directly derivable from their known IDs plus some common information. It thus eliminates the need for certificate-based authenticated public-key distribution indispensable in conventional public-key management schemes. IKM features a novel construction method of ID-based public/private keys, which not only ensures high-level tolerance to node compromise, but also enables efficient network-wide key update via a single broadcast message [23].

Zapata in [22] recommended SAODV is improved from AODV to prevent impersonation attacks by changing hop-count (HC) and sequence number (SN) values of route discovery packet. However, SAODV only supports authentication from end-to-end without authenticating hop-by-hop, hence, intermediate node can't certify packet from the preceding node. Addition, because SAODV does not have a mechaism for authentication intermediate node and public key management, malicious nodes can easily join a route by using fake keys.

Sanzgiri [18] recommended ARAN protocol, different from SAODV, route discovery packet (RDP) in ARAN is signed and certified at all nodes. ARAN supplemented the testing member node mechanism, thus, malicious can not pass over security by using fake keys. Structure of RDP and reply route (REP) packets of ARAN is not available with HC to identify routing cost; this means ARAN is unable to recognize transmission expenses to the destination, ARAN argued that the first REP received is the route packet with the best expenses.

Li [10] recommended SEAODV using certification scheme HEAP with symmetric key and one-way hash function to protect route discovery packet. By simulation, the authors has shown that SEAODV is more security with lower communication overhead.

3 Digital Certificates Management and Providing Model

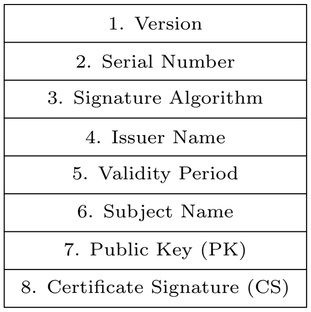
This section describes the digital certificates structure based X.509 and mechanism to manage and provide the Digital Certificate for MANET without PKI. For this approach, we assumptions that each node has a unique identifier and a pair of keys: a private key and a public key. Set of symbols in Table 1 are applied for the presentation.

Table 1: Description of symbols

Variable	Descriptions
N_δ	Node labeled δ
$k_{N_\delta}+$, k_{N_δ}-	Public and private keys of node N_δ
$En(v,k)$	Encrypting v using key k
$De(v,k)$	Decrypting v using key k
$H(v)$	v is hashed by SHA_1 [14] function
IP_{N_δ}	Address of node N_δ
DC_{N_δ}	Digital Certificate of node N_δ

3.1 Digital Certificates

Digital certificates are used to certify the identities of nodes in MANET, it is provided for node automatically from certificate authorities (CA) before nodes collaborate to the discovery route process. We uses a X.509 certificate template, has the structure as Figure 2. Where,



1. Version
2. Serial Number
3. Signature Algorithm
4. Issuer Name
5. Validity Period
6. Subject Name
7. Public Key (PK)
8. Certificate Signature (CS)

Figure 2: DC structure based on X.509 [13]

1) Certificate version;

2) The unique serial number that is assigned by the CA;

3) The public key cryptography and message digest algorithms that are used by CA;

4) The name of the issuing CA;

5) The certificate's start and expiration dates. These define the interval during which the certificate is valid, although the certificates can be revoked before the designated expiration date;

6) The name of the subject of the certificate;

7) The public key and a list of the public key cryptography algorithms;

8) The CA's digital signature, which is created as the last step in generating the certificate by encrypting the hash value of all X.509 certificates attributes with of CA private keys as Formula 1.

$$CS \leftarrow En(H(DC.AllFields\backslash\{CS\}), k_{N_{CA}}-). \qquad (1)$$

Algorithm 1 shows steps to authenticate DC of the packet RREQ (or RREP) if N_i node receiving the packet from preceding node N_j. Node N_i uses the public key ($k_{N_{CA}}+$) of certificate authorities to decrypt the CS field value of packet RREQ (or RREP). If the value after decryption is coincident with the hash value of all fields (excepted CS) for DC then DC is valid, on the contrary then DC is invalid.

Algorithm 1 Checking Digital Certificate

Input: RREQ or RREP packet; **Output:** True if DC is valid; Else return False

```
Boolean IsValidDC(Packet P)
Begin
    $val_1 \leftarrow De(P.DC.CS, k_{N_{CA}}+)$;
    $val_2 \leftarrow H(P.DC.AllFields\backslash\{CS\})$;
    Return ($val_1 == val_2$);
End
```

3.2 Digital Certificate Management

We setup a reliable node named N_{CA} acts as certificate authorities to provide *DC* for all member nodes. In N_{CA} exists a Digital Certificate Database (DCDB) of all nodes as Table 2. Each record in DCDB consists of: Nodes address, OK field controlling the node certificated with *DC* and its Digital Certificates. Where, all attributes (except OK field) are updated by administrators to ensure that only "friendly" nodes are provided with *DC*.

Table 2: Digital certificate database

Nodes	**OK**	**Digital Certificate**
IP_{N_1}	yes	DC_{N_1}
IP_{N_2}	yes	DC_{N_2}
IP_{N_3}	no	DC_{N_3}
...	...	...
IP_{N_n}	yes	DC_{N_n}

3.3 Digital Certificate Providing

We propose a digital certificate providing model which secure that

1) Malicious node can not action as CA node to provide DC to member node;

2) Only the valid member node receives the DC from CA node.

There are two *DCP* and DC_{ACK} packets are used to provide the Digital Certificates for all nodes. They have the structures similar as RREQ and RREP packets, *DCP* packet has a new field named DC to store the digital certificate, DC_{ACK} has two new fields named ACK and KEY, they save acknowledge information and public key from member node. The steps to provide the DC for all nodes following:

- *The first*, administrators update DC of "friendly" nodes to DCDB. Member nodes can not to collaborate in the route discovery process until they have received DC from N_{CA}.

- *The second*, periodically after T_{DC} time interval, node N_{CA} checks all nodes are provided with DC by using the DCDB information. If exist node N_δ that it is not provided with DC (OK = False), N_{CA} broadcasts the *DCP* packet to provide the DC for N_δ.

- *Continuous*, when receiving DC, node N_δ sends DC_{ACK} packet back to N_{CA} to confirm that member node already receives DC if DCP packet is sent by N_{CA} and sent for it.
- *Finally*, when receiving packet DC_{ACK}, N_{CA} checks: If the packet is sent by N_δ, N_{CA} updates OK value is true to DCDB, else this process is fail.

3.3.1 Broadcasting DCP Packet and Saving DC

Node N_{CA} provides a DC for node N_δ by broadcasting DCP packet, is improved from algorithm broadcasting RREQ packet of AODV following:

1) *Generating DCP packet:* Node N_{CA} creates DCP with DC_{N_δ} and broadcasts it to all its neighbors as Formula 2.

$$N_{CA} broadcasts : DCP \leftarrow \{RREQ^* + DC_{N_\delta}\}. \quad (2)$$

Where RREQ is the original RREQ packet of AODV protocol and DC is N_δ 's Digital Certificate. CS field value in DC that it is calculated as Formula 3.*

$$DC.CS \leftarrow En(DC.CS, k_{N_\delta}+). \quad (3)$$

2) *Checking DCP and saving DC:* When node N_δ receives the DCP packet, it tests that DCP is sent by N_{CA} and provided DC for N_δ. If all the conditions are satisfied, N_δ saves DC into its cache and unicasting the DC_{ACK} packet to confirm for N_{CA}. On contrary, the packet is dropped, see in Algorithm 2.

Algorithm 2 Testing and Saving Digital Certificate;

Input: DCP packet; **Output:** True if DC is saved successful; Else return False;

```
1:  Boolean TestAndSaveDC(DCP P)
2:  Begin
3:      $val_1 \leftarrow De(P.DC.CS, k_{N_\delta}-)$;
4:      $val_2 \leftarrow De(val1, k_{N_{CA}}+)$;
5:      If $val_2$ != $H(P.DC.AllFields \backslash \{CS\})$ Then
6:          Dispose(P) and Return False;
7:      Else
8:          $P.DC.CS \leftarrow val1$;
9:          SaveToCache(P.DC);
10:         Sends $DC_{ACK}$ packet back to $N_{CA}$;
11:         Return True;
12: End
```

We clearly see that malicious nodes can easily receive DCP packet come from the N_{CA} node because they are sent in the form of a broadcast. However, the malicious node can not decrypt the contents of the certification in DC of DCP packet because it does not know the secret key of N_δ node. If exists any change in the DC packet resulting in command 5 in Algorithm 2 is true, the DCP packet is canceled, the DC proveding process is fail.

3.3.2 Replying the DC_{ACK} Packet

Member node N_δ sends a DC_{ACK} packet back to confirm for N_{CA}, this algorithm is improved from unicasting RREP packet algorithm of AODV following:

1) *Generating DC_{ACK} packet:* After saving DC successfully, node N_δ unicasts confirmation packet DC_{ACK} to back N_{CA} as Formula 4.

$$N_\delta unicasts : DC_{ACK} \leftarrow \{RREP^* + ACK + KEY\} \quad (4)$$

Where RREP is the original RREP packet of AODV routing protocol and ACK field is calculated by Formula 5, KEY field value is its public key.*

$$DC_{ACK}.ACK \leftarrow En(En(H(IP_{Nca}), k_{N_\delta}-), k_{Nca}+)) \quad (5)$$

2) *Checking DC_{ACK} and updating DCDB:* When node N_{CA} receives the DC_{ACK} packet, it tests DC_{ACK} packet is sent by N_δ and it is target node. If all the conditions are satisfied, N_{CA} updates successfully provided DC to DCDB, on contrary, the packet is dropped, see in Algorithm 3.

Algorithm 3 Testing DC_{ACK} and Updating DCDB

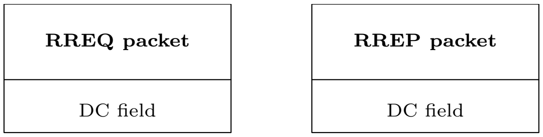
Input: DC_{ACK} packet; **Ouput:** True if DC is provided successful; Else return False

```
Boolean TestDC$_{ACK}$($DC_{ACK}$ P)
Begin
    $val_1 \leftarrow De(P.ACK, k_{N_{CA}}-)$;
    $val_2 \leftarrow De(val1, P.KEY)$;
    If $val_2$ != $H(IP_{Nca})$ Then
        Dispose(P) and Return False;
    If ($IP_{N_\delta}$ exists in DCDB) Then
        DCDBRow row $\leftarrow$ DCDB.Rows[$IP_{N_\delta}$];
        row.OK $\leftarrow$ True;
        Return True;
    Else
        Dispose(P) and Return False;
End
```

We clearly see that a malicious node can hardly receive DC_{ACK} packet because this packet is sent in unincast form. Moreover, malicious nodes can not be act as N_δ to send DC_{ACK} packet to N_{CA}. The reason is because it does not have the secret key of N_δ, and the public key of N_δ was administered by N_{CA}.

4 AODVDC: Improved Protocol Using Digital Certificates

An algorithm has been designed based on reactive routing protocols accepted as standards for routing in MANETs such as AODV. However, the AODV protocol have not any security mechanism for discovery route processing. This is the hole which can be easily exploited by hackers to attack the network by modifying the control packets with fake information. Improved protocol named AODVDC, is proposed by integration of DC authentication algorithm into AODV protocol includes the two phases: Broadcasting route request packet and unicasting route reply packet. The control route packet structures of new protocol is improved from control route packets in AODV, they are supplemented a new field named DC as Figure 3.



(a) New RREQ packet (b) New RREP packet

Figure 3: Control packet structures of AODVDC

4.1 Broadcasting Route Request Packet

Figure 4 describes route request algorithm using DCa method, it is improved from AODV route discovery algorithm as following:

1) *Generating RREQ packet:* If source node (N_S) has not a route to destination node, it starts a new route discovery process by broadcasting the RREQ packet to its all neighbors described as Formula 6.

$$N_S broadcasts : RREQ^* + DC_{N_S} \qquad (6)$$

Where RREQ is the original RREQ packet of AODV routing protocol and DC is its Digital Certificate.*

2) *Processing and forwarding RREQ packet:* When a node receiving a RREQ packet, intermediate or destination node (N_i) processes the packet following:

- If it has not the DC Then N_i drops RREQ packet and The end;
- Else, N_i tests the preceding node 's DC in RREQ packet using *IsValidDC()* function. If DC is invalid Then N_i drops the RREQ packet due to discovered route has malicious node and The end;
- Else, if current node is the destination, it just simply generates and sends back the RREP packet and The end;
- Else, it updates a reverse route toward the source node and updates the RREQ packet using its information and DC before continuous broadcasting the RREQ packet to its all neighbors.

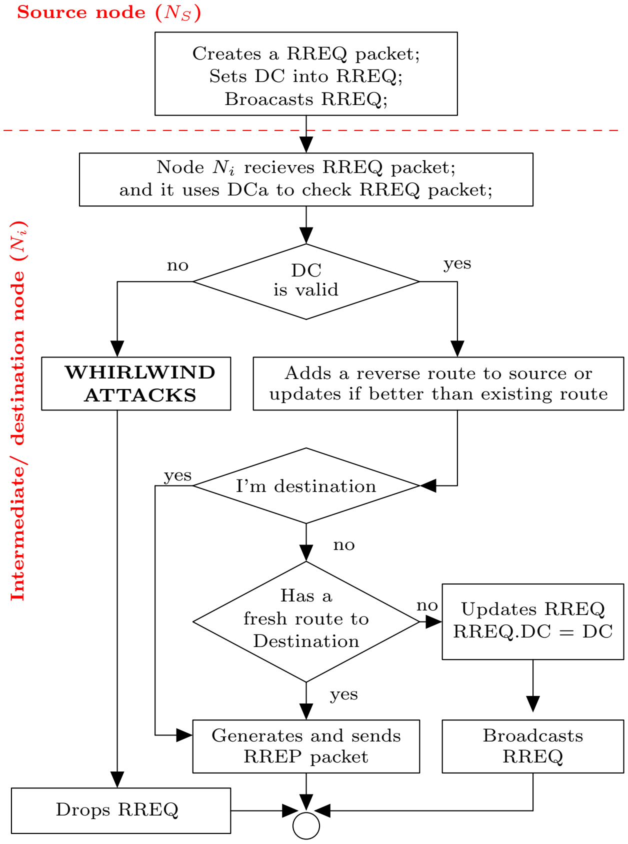

Source node (N_S)
Creates a RREQ packet; Sets DC into RREQ; Broacasts RREQ;
Intermediate/ destination node (N_i)
Node N_i recieves RREQ packet; and it uses DCa to check RREQ packet;
DC is valid
no
yes
WHIRLWIND ATTACKS
Adds a reverse route to source or updates if better than existing route
I'm destination
yes
no
Has a fresh route to Destination
no
Updates RREQ RREQ.DC = DC
yes
Generates and sends RREP packet
Broadcasts RREQ
Drops RREQ

Figure 4: Improved request route algorithm

4.2 Unicasting Route Request Packet

Figure 5 describes route reply algorithm using DCa method, it is improved from AODV route reply algorithm as following:

1) *Generating RREP packet:* A node generates the RREP packet if it is either the destination (N_D) or an intermediate (N_i) which has a "fresh" route to the destination described as Formula 7.

$$N_D unicasts : RREP^* + DC_{N_D} \qquad (7)$$

Where RREP is the original RREP packet of AODV routing protocol and DC is its Digital Certificate.*

2) *Processing and forwarding RREP packet:* When a node receiving a RREP packet, intermediate or destination node (N_i) processes the packet following:

- If it has not the DC Then N_i drops this packet and The end;
- Else, N_i tests the preceding node 's DC in RREP packet using *IsValidDC()* function. If DC is invalid Then N_i drops the RREP packet due to discovered route has malicious node and The end;
- Else, if current node is the source node, it just simply saves a new route or updates if better than existing

route and send data packets from queue to the destination node through discovered route;

- Else, it saves a route to the destination node and updates the RREP packet using its information and DC before continuous unicasting the RREP to back source node.

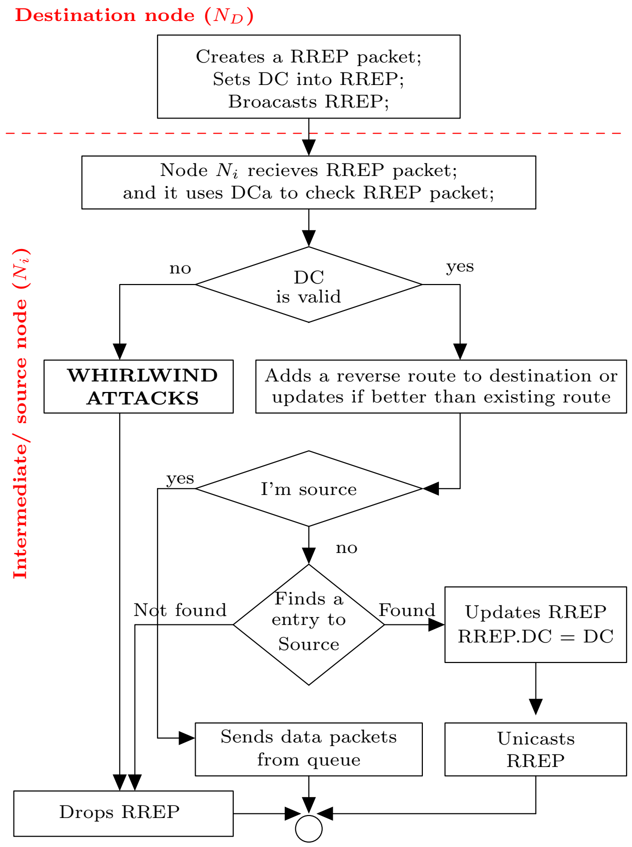


Figure 5: Improved reply route algorithm

5 Simulation Results

We evaluate the Whirlwind attacks prevention performance of AODVDC on simulation system is NS2 - version 2.35 [11]. The simulation area was a rectangular region with a size of 2000 x 2000 m^2, which was chosen to ensure that there existed multiple hops within the network. We use 802.11 MAC layer, 100 normal nodes move with 30m/s maximum speeds under Random Waypoint [21] model, 1 malicious node stays at the center position (red rectangle in Figure 6) and starts to attack at 500s.

Each scenario has 10 pairs of communicating nodes, source sending out constant bit rate (CBR) traffic with packet sizes of 512bytes, rate of 2 packet per second. The first data source is started at second of 0, the following data source is 5 seconds apart from each node. Time 1000 seconds for simulation, FIFO queue type, the detail of simulation parameters are listed in the Table 3.

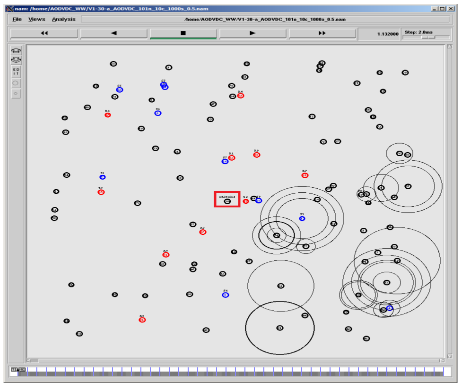


Figure 6: NS2 simulation screen

Table 3: Simulation parameters

Parameters	Setting
Simulation area (m)	2000 x 2000
Simulation time (s)	1000
Number of nodes	101 (1 malicious nodes)
Attack point-time (s)	500^{th}
Wireless standard	IEEE 802.11
Ratio range (m)	250
Mobility model	Random Waypoint
Mobility speed (m/s)	1..30
Number of connection	10 UDPs
Traffic type	CBR
Data rate	2 pkt/s (512 bytes/pkt)
Queue type	FIFO (DropTail)
Routing protocols	AODV, ARAN and AODVDC
Nca	N_{50}

Some used metrics for evaluation following: Packet overhead for providing DC, packet delivery ratio (PDR), routing load (RL) and end-to-end delay (EtE).

5.1 Packet Overhead for Providing DC

We analyse the packet overhead (DCP and DC_{ACK}) for providing the DC in normal network topology. The first scenario simulates for 100 nodes used AODVDC protocol, all DC of normal nodes are setup in DCDB; The second scenario simulates for 100 nodes used AODVDC protocol with 80 normal nodes from 0 to 79 identify in DCDB; The final scenario, we use a scenario similar to the second scenario, and 20 new nodes are installed into DCDB at 300th seconds. The simulation results in Figure 7 shows that AODVDC needs total 57,908 packets DCP and DCACK overhead and 560s to provide DC for all 100 nodes. For the second scenario, there are total 30,096 packets overhead and providing DC has finished during 200s. In the final scenario, there are 36,253 packets overhead and 420s for finished providing DC.

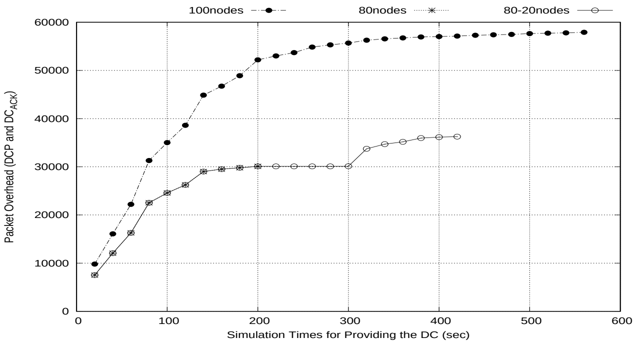


Figure 7: Packet overhead for providing DC

5.2 Whirlwind Attack Prevention Performance

The main purpose for whirlwind attack is to destroy data packets, reduced packet delivery ratio. Figure 8 shows that packet delivery ratio of AODV go down significantly under whirlwind attacks, reduced during simulation times from seconds 500^{th}. The packet delivery ratio of AODVDC increasing from seconds 600^th because it uses first 560 seconds for providing DC for member nodes. After 1000s for simulation with 10UDP connections, the packet delivery ratio of AODV is 71.04% for normal network topology down to 58.02% under whirlwind attacks, reduced 13.02%. The ARAN packet delivery ratio is 59.51% and AODVDC is 65.49%. It is then clear that the AODVDC packet delivery ratio is improved significantly and has better packet delivery ratio compared to ARAN.

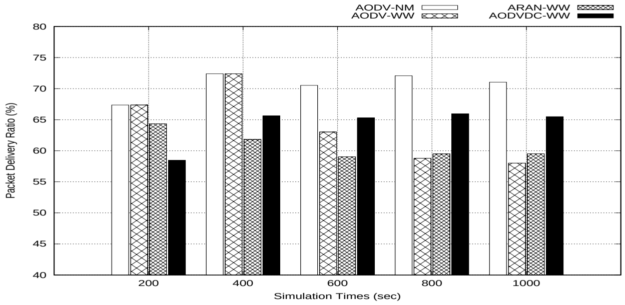


Figure 8: Packet delivery ratio; WW: Whirlwind, NM: Normal

5.3 Routing Load

The AODVDC routing load is larger than original one due to proposed approach used overhead packets DCP and DC_{ACK} for providing the DC for all member nodes. Figure 9 shows that the routing load of AODVDC is larger than AODV, reduced during simulation times due to finished providing DC for member nodes. After 1000s for simulation with 10UDP connections, the routing load of AODVDC is 22.37pkt, AODV is 17.74pkt and ARAN is 20.05pkt.

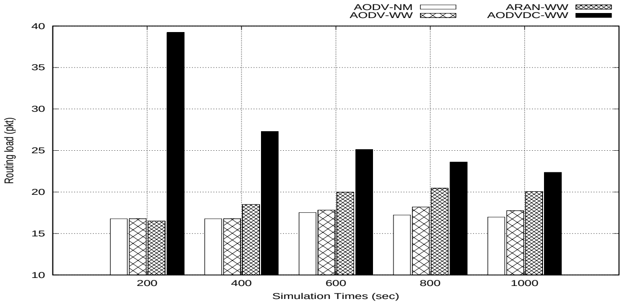


Figure 9: Routing load; WW: Whirlwind, NM: Normal

5.4 End-to-End Delay

Figure 10 shows that all security protocols have end-to-end delay is higher than AODV because of they used RSA public key encryption and hash function SHA_1 for security goal. After 1000s for simulation, end-to-end delay of AODV is 0.867s, ARAN is 1.214s and AODVDC is 1.279s.

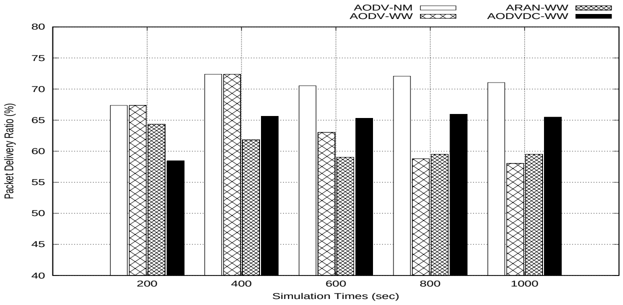


Figure 10: End-to-End delay; WW: Whirlwind, NM: Normal

6 Conclusion

We proposed a mechanisms to manage and provide digital certificates (DC) for Mobile Ad hoc Network (MANET) without public key infrastructure. A new routing protocol named AODVDC by integrating our solutions into the discover route process from AODV protocol. The simulation results showed that our approach has better performance in terms of packet delivery ratio, routing load and route discovery delay compared to related works under attack scenario. However, AODVDC has routing load and end-to-end delay are larger than AODV because it uses new control packets for providing DC for member nodes and uses RSA [6] public key encryption, SHA_1 [14] hashing function.

In the future, we will setup AODVDC with large key to improve the security performance using TLS library [19] and comparison with related works.

References

[1] R. Abderrezak, B. Abderrahim, "A secure architecture for mobile Ad hoc networks," *Mobile Ad-hoc and Sensor Networks*, pp. 424–435, 2006. (https://link.springer.com/chapter/10.1007\%2F11943952_36)

[2] E. Alotaibi and B. Mukherjee, "A survey on routing algorithms for wireless Ad-hoc and mesh networks," *Computer Networks*, vol. 56, no. 2, pp. 940–965, 2012. (http://www.sciencedirect.com/science/article/pii/S138912861100377X)

[3] A. P. Asad and C. McDonald, "Detecting and evading wormholes in mobile Ad-hoc wireless networks," *International Journal of Network Security*, vol. 3, no. 2, pp. 191–202, 2006. (http://ijns.jalaxy.com.tw/contents/ijns-v3-n2/ijns-2006-v3-n2-p191-202.pdf)

[4] C. W. Badenhop, B. W. Ramsey and B. E. Mullins, "An analytical black hole attack model using a stochastic topology approximation technique for reactive Ad-hoc routing protocols," *International Journal Network Security*, vol. 18, no. 4, pp. 667–677, 2016. (http://ijns.jalaxy.com.tw/download_paper.jsp?PaperID=IJNS-2015-06-11-1&PaperName=ijns-v18-n4/ijns-2016-v18-n4-p667-677.pdf)

[5] L. S. Casado, G. M. Fernández, P. G. Teodoro and N. Aschenbruck, "Identification of contamination zones for Sinkhole detection in MANETs," *Journal of Network and Computer Applications*, vol. 54, pp. 62–77, 2015. (http://www.sciencedirect.com/science/article/pii/S1084804515000818)

[6] W. Diffie and M. E. Hellman, "New Directions in Cryptography," *IEEE Transactions on Information Theory*, vol. 22, no. 6, pp. 644-654, 1976. (http://dl.acm.org/citation.cfm?id=2269104)

[7] X. Gao and W. Chen, "A novel gray hole attack detection scheme for mobile Ad-hoc networks," in *IFIP International Conference on Network and Parallel Computing Workshops*, pp. 209–214, 2007. (http://ieeexplore.ieee.org/document/4351486/)

[8] H. Jeroen, M. Ingrid, D. Bart and D. Piet, "An overview of mobile Ad hoc networks: Applications and challenges," *Journal of the Communications Network*, vol. 3, no. 3, pp. 60–66, 2004. (https://biblio.ugent.be/record/317876)

[9] S. Kurosawa, H. Nakayama, N. Kato, A. Jamalipour and Y. Nemoto, "Detecting blackhole attack on AODV-based mobile Ad Hoc networks by dynamic learning method," *International Journal of Network Security*, vol. 5, no. 3, pp. 338–346, 2007. (http://ijns.jalaxy.com.tw/contents/ijns-v5-n3/ijns-2007-v5-n3-p338-346.pdf)

[10] C. Li, Z. Wang and C. Yang, "SEAODV: A security enhanced AODV routing protocol for wireless mesh networks," *Transactions on Computational Science XI*, vol. 6480, pp. 1–16, 2010. (http://link.springer.com/chapter/10.1007\%2F978-3-642-17697-5_1)

[11] S. McCanne and S. Floyd, *The Network Simulator NS2*. (http://www.isi.edu/nsnam/ns/)

[12] J. V. Mulert, I. Welch, and W. K. G. Seah, "Security threats and solutions in MANETs: A case study using AODV and SAODV," *Journal of Network and Computer Applications*, vol. 35, no. 4, pp. 1249–1259, 2012. (http://www.sciencedirect.com/science/article/pii/S1084804512000331)

[13] M. Myers, R. Ankney, A. Malpani, S. Galperin and C. Adams, *X.509 Internet Public Key Infrastructure Online Certificate Status Protocol - OCSP*, RFC 2560, June 1999. (https://tools.ietf.org/html/rfc6960https://tools.ietf.org/html/rfc6960)

[14] National Institute of Standards and Technology, "Secure hash standard," FIPS PUB 180-1, 1995. (https://tools.ietf.org/html/rfc3174)

[15] L. T. Ngoc and V. T. Tu, "Whirlwind: A new method to attack routing protocol in mobile Ad hoc network," vol. 19, no. 5, *International Journal of Network Security*, pp. 832–838, 2017. (http://ijns.jalaxy.com.tw/download_paper.jsp?PaperID=IJNS-2016-06-28-1&PaperName=ijns-v19-n5/ijns-2017-v19-n5-p832-838.pdf)

[16] C. E. Perkins, M. Park and E. M. Royer, "Adhoc on-demand distance vector routing," in *Proceedings of Second IEEE Workshop on Mobile Computing Systems and Applications (WMCSA'99)*, pp. 90–100, 1999. (https://tools.ietf.org/html/rfc3561)

[17] R. D. Pietro, S. Guarino, N. V. Verde and J. Domingo-Ferrer, "Security in wireless Ad-hoc networks - A survey," *Computer Communications*, vol. 51, pp. 1-20, 2014. (http://www.sciencedirect.com/science/article/pii/S0140366414002242)

[18] K. Sanzgiri, B. Dahill, B. N. Levine, C. Shields and E. M. B. Royer, "A secure routing protocol for Ad hoc networks," in *10th IEEE International Conference on Network Protocols (ICNP'02)*, pp. 78–87, 2002. (http://ieeexplore.ieee.org/document/1181388/)

[19] TLS library, RSA source code. (https://tls.mbed.org/rsa-source-code)

[20] V. T. Tu, L. T. Ngoc, "SMA$_2$AODV: Routing protocol reduces the harm of flooding attacks in mobile Ad hoc network," *Journal of Communications*, vol. 12, no. 7, pp. 371–378, 2017. (http://www.jocm.us/index.php?m=content&c=index&a=show&catid=179&id=1122)

[21] J. Yoon, M. Liu and B. Noble, "Random waypoint considered harmful," *Twenty-Second Annual Joint Conference of the IEEE Computer and Communications*, vol. 2, pp. 1312-1321, 2003. (http://ieeexplore.ieee.org/abstract/document/1208967/)

[22] M. G. Zapata, "Secure Ad hoc on-demand distance vector routing," *Mobile Computing and Communications Review*, vol. 6, no. 3, pp. 106–107, 2002.

[23] Y. Zhang, W. Liu, W. Lou and Y. Fang, "Securing mobile Ad hoc networks with certificateless public keys," *IEEE Transactions on Dependable and Secure Computing*, vol. 3, no. 4, pp. 386–399, 2016. (http://ieeexplore.ieee.org/abstract/document/4012650/)

[24] L. Zhou and Z. J. Haas, "Securing Ad hoc networks," *IEEE Network, Special Issue on Network Security*, vol. 13, no. 6, pp. 24-30, 1999.

Luong Thai Ngoc is with the Faculty of Mathematics and Informatics Teacher Education, Dong Thap University. He received B.E. degree in Computer Science from Dong Thap University in 2007 and MSc degree in Computer Science from Hue University of Sciences in 2014. He is a PhD student in Hue University of Sciences now. His fields of interesting are network routing, analysis and evaluation of network performance, security wireless ad hoc network.

Vo Thanh Tu is an associate professor in the Faculty of Information Technology, Hue University of Sciences, Hue University. He received B.E. degree in Physics from Hue University in 1987 and PhD degree in computer science from Institute of Information Technology, Vietnam Academy of Science and Technology in 2005. His fields of interesting are network routing, analysis and evaluation of network performance, security wireless ad hoc network.

Comprehensive Security for Body Area Networks: A Survey

Laura Victoria Morales, David Delgado-Ruiz, and Sandra Julieta Rueda
(Corresponding author: Laura Victoria Morales)

Systems and Computing Engineering Department, Universidad de los Andes
Cra 1 Este No 19A - 40, Mario Laserna Building, Bogotá, Colombia
(Email: l.morales825@uniandes.edu.co)

(Received Sept. 5, 2017; revised and accepted Apr. 13, 2018)

Abstract

Body Area Networks (BANs) are composed of multiple devices that measure, collect, forward and analyze physiological and medical data that may be used for different purposes like activity tracking, health monitoring or medical treatments. Given the type of data BANs manage, several security requirements must be addressed: confidentiality, integrity, privacy, authentication and authorization. This survey studies various proposals that aim to satisfy BAN security requirements, their advances and remaining challenges. We found that the mentioned requirements have not been comprehensively considered; the majority of the studied proposals do not address the entire BAN architecture, they focus on specific components. Although supporting security of individual BAN components is relevant, a comprehensive security view of an entire BAN environment is needed.

Keywords: BAN Security; Body Area Networks; eHealth

1 Introduction

Body Area Networks (BANs) enable wired and wireless communications among different types of devices, such as wearable and implantable sensors, smartphones, tablets and external servers, to collect physiological data for different purposes, particularly to support medical decisions and improve medical care. Data collected by BANs is considered sensitive, thus several security requirements must be addressed. If unauthorized entities gain access to this kind of data, patients may suffer diverse consequences, like job or insurance losses. Modified data may lead to wrong medical decisions; for example, an insulin pump may inject a wrong insulin dose [32,38].

This survey studies different proposals that address security in BAN environments. We classified these proposals using two criteria: addressed security requirements and considered BAN components. The former aspect includes confidentiality, authentication, authorization, integrity and availability. The latter aspect considers BAN components including devices that measure physiological data (sensors and actuators), forward data (personal servers, smartphones or tablets), and store data (external servers and cloud).

We found that the studied projects only secure one or two components of a BAN architecture, sensors and actuators in particular. Although some BAN components are being secured, there is not a comprehensive security proposal for an entire BAN architecture. In order to build this comprehensive view, we must consider other devices like external servers, cloud services that store collected data, and even auxiliary devices, like gateways and access points.

Having a comprehensive view of the entire BAN architecture enables analysts to see security issues that may be hidden otherwise. For instance, some security solutions that work on external servers and cloud services, may not work on sensors and actuators because of their processing restrictions. Key management is particularly challenging, as it must consider different aspects for sensors and actuators, and for external servers. For example, some proposals generate keys using data collected by the sensors; however, a personal or external server cannot automatically compute these keys.

The rest of the paper is organized as follows: Section 2 presents an overview of BAN components and their interactions, Section 3 summarizes BAN security requirements and classifies the studied proposals according to addressed security requirements and BAN components, and Section 5 presents open issues. Section 6 concludes.

2 Body Area Network (BAN) Architecture

This section presents the main BAN components and types of communications. Later, we will use these characteristics to classify the studied security proposals.

2.1 Components

We identified the following categories of BAN components: Sensors and actuators, personal servers, auxiliary network devices, channels, external servers, and cloud services. Figure 1 illustrates BAN components and their interactions.

Sensors and Actuators: Sensors are implanted or wearable devices [10] that measure human physiological functions and environmental features. Actuators are devices that perform specific tasks; for example, the actuator in an insulin pump injects an insulin dose to a patient. Sensors and actuators may be part of a single device; Implantable Medical Devices (IMD) for instance, have sensors, actuators, and even a CPU [46]. Figure 1 shows several sensors: electroencephalography (EEG), electrocardiography (ECG), blood pressure and motion sensors.

Personal servers: These computing devices collect data from the sensors, temporarily store them, and forward them to interested parties, like a patient's medical team or family [8]. Different devices can be used as personal servers depending on a patient's movement restrictions; personal computers or laptops may work for users with mobility restrictions while tablets or smartphones are more adequate for physically active users. Figure 1 shows two devices that may work as personal servers: a tablet and a laptop.

Auxiliary Network Devices and Channels: We consider access points, gateways and cellular towers as auxiliary network devices, as these devices enable communications among components. Most communications are wireless, since the majority of possible personal servers have Wi-Fi antennas or use cellular networks, like smartphones and tablets. A BAN may also have wired communications; for example, when it includes a server deployed in a hospital. Figure 1 shows the following auxiliary devices: an access point, a gateway and a cellular tower.

External servers: External servers are medium and big-sized computing devices that gather and store information sent by several personal servers that belong to different patients. External servers may keep records for a high number of patients and records may have different types of data, like documents, diagnostic images or videos. Therefore, it is desirable to have servers with high storage and processing capacities.

Cloud: Cloud computing services provide additional storage and computing resources that may be needed in several contexts. For example, hospitals that have a high number of patients can use cloud storage. Cloud services may be used to analyze data for different purposes, like studying diseases and their behavior or creating predictive models.

2.2 Communication Types

We organized communication types in tiers, based on distance between communicating devices and the human body: Intra-BAN for close range, Inter-BAN for medium and Beyond-BAN for long range communications [10]. Figure 1 illustrates this classification.

Intra-BAN Communications: This tier covers communications happening within a two-meter radio from the human body, meaning that this tier comprehends sensor-to-sensor and sensor-to-personal server communications. In this tier, communications are typically wireless, using technologies such as Bluetooth and ZigBee, however wired communications are also possible.

Inter-BAN Communications: This tier covers communications between personal servers or sensors and an auxiliary network device to reach external servers. Internet or cellular networks can be used to establish this kind of communications. There are two types of architecture for Inter-BAN: infrastructure-based or ad hoc-based. The first one is used when a patient is confined within a limited space, like a room. In contrast, the ad hoc-based architecture allows a wider coverage, since it uses multiple access points to connect several networks [10].

Beyond-BAN Communications: The third tier covers communications between an access point, external servers and cloud resources, possibly covering metropolitan areas. In most cases a BAN needs a gateway to enable a connection between Inter-BAN and beyond-BAN devices [10].

3 Security Requirements and Solutions in a BAN Architecture

Different governments have different regulations to control management and address security concerns of health related information.

In the United States, the Health Insurance Portability and Accountability Act (HIPAA) provides legislation and security provisions for safeguarding medical information [31,32,62]. The European Union published, in 2016, a new regulation to protect personal data. This regulation 'provides more rights to citizens to be better informed about the use of their personal data, and gives clearer responsibilities to people and entities using personal data. [15]. Australia's Personally Controlled Electronic Health Records and Canada's Health Information Legislation also protect patient's data. Other countries are also working on legislation to protect medical data of their citizens.

Some international standards also address these security concerns. The European Committee for Standardization (CEN/TC 251 - CEN Technical Committee 251) has

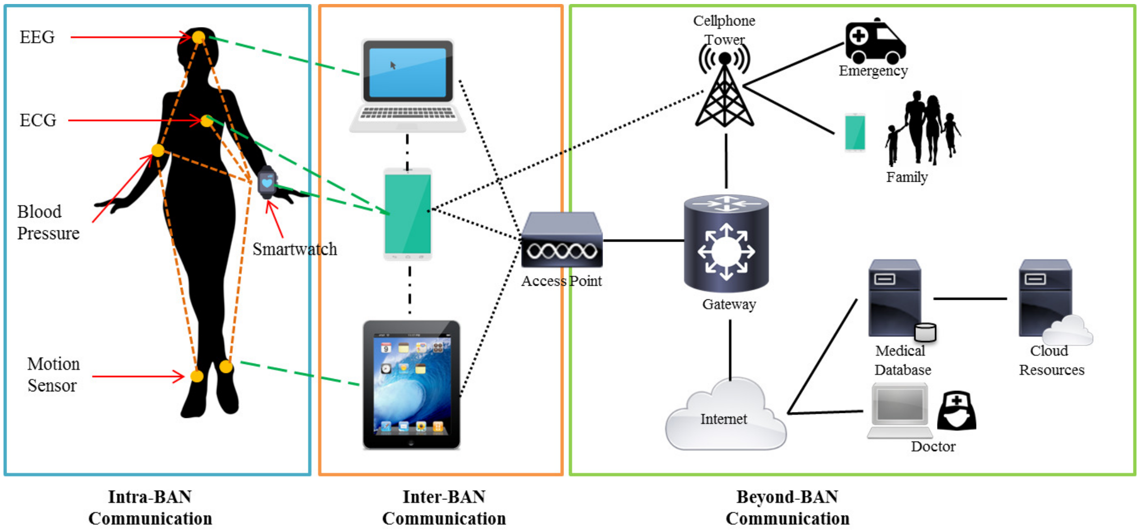


Figure 1: This figure illustrates components and communication types in a BAN architecture. Dashed thin lines show sensor-to-sensor communications and dashed thick lines show sensor-to-personal server communications. A smart watch is used as a sensor and also as a gateway, to send data from sensors to a smartphone, tablet or laptop that serves as the personal server. An access point and a gateway enable communication between personal and external servers. Beyond-BAN communications include communications with elements beyond the access point. Cellphones may use internet or a cellphone network to send data, while laptops and tablets usually do not have access to cellular networks and send data via internet. Extended from [10].

worked to define a standard for data management in the fields of Health Information and Communications Technology in the European Union. The standard establishes requirements for data compatibility and interoperability between systems, as well as data security requirements. The Technical Committee on health informatics of the International Organization for Standardization (ISO/TC 215) has already delivered several standards regarding security of medical records [1].

The IEEE 802.15.6 standard [28] identifies three security levels for BANs.

Level 0. Unsecured communication: No authentication or encryption techniques are used while sending messages.

Level 1. Authentication but not encryption: Authentication and some integrity validation are implemented.

Level 2. Authentication and encryption: Messages are transmitted in authenticated and encrypted frames. The standard also considers integrity, confidentiality and privacy.

Authorization, data freshness and software correctness are security requirement that also appear in a BAN context. In addition, different BAN components, with different features, interact to collect, process and forward data. These components may use different protocols to send sensitive data and may belong to different owners generating a large attack surface and several security concerns.

In this survey we looked for advances and remaining challenges in BAN security. To do so, we selected extended /full papers published in international conferences or journals between 2008 and 2016, that had BAN and security, or variations of these, as keywords. The variations of BAN included Body Area Networks and Body Sensor Networks, and the variations of security included confidentiality, integrity, authentication, authorization, availability and software correctness. We also looked for the first appearances of body sensors security and included one paper from 2003.

3.1 Confidentiality

The goal of this requirement is to guarantee that unauthorized people cannot read protected data. Since sensors generate physiological information that may reveal a disease or disability, data confidentiality is a relevant security requirement in BAN systems [38]. It is important to protect patient's data during transmission between devices, as well as in storage. In addition to medical data, BANs must also protect their device's information such as identification numbers, location, function, configuration and type [47].

3.1.1 Data Encryption and Key Generation

In our set of papers, cryptography is the most studied mechanism to offer confidentiality. However, implementing encryption algorithms for a BAN can be challenging due to power, memory and processing limitations, and low communication ranges of sensors and actu-

Table 1: Papers addressing BAN security requirements per year

	2003	2008	2009	2010	2011	2012	2013	2014	2015	2016	**Total**	**Percentage**
Confidentiality												
Data Encryption												
Secret	0	1	0	0	1	0	1	2	2	0	7	14.8 %
Elliptic Curves	0	0	0	1	1	2	0	1	1	1	7	14.8 %
Total	0	1	0	1	2	2	1	3	3	1	14	29.7 %
Key Generation												
Physiological Signals	1	2	2	3	2	2	3	1	0	1	17	36.1 %
Channel Characterization	0	0	0	0	2	2	3	0	0	0	7	14.8 %
Total	1	2	2	3	4	4	6	1	0	1	24	51%
Key Distribution												
Fuzzy vault	1	1	1	1	0	2	2	0	0	1	9	19.1 %
Diffie-Hellman	0	0	0	1	1	1	1	1	0	1	6	12.7 %
Other	0	2	0	0	1	1	1	2	3	1	11	23.4 %
Total	1	3	1	2	2	4	4	3	3	3	26	55.3%
Access Control												
Authentication	1	3	2	4	6	6	8	3	3	3	39	82.9 %
Authorization	0	0	1	1	1	0	3	0	1	1	8	17 %
Integrity												
Hash Functions	1	0	0	1	2	1	5	1	1	0	12	25.5 %
Session Management	0	1	0	2	3	0	4	1	1	1	13	27.6 %
Availability												
DoS Attack Protection	0	0	1	1	0	2	0	0	0	0	4	8.5 %
Total number of studied papers												**47**

ators [11,12,31,38]. There is another challenge, in some scenarios data must be easily accessed; for example, in a medical emergency. If patient's data is encrypted and the key is not available, then a patient may not receive proper attention [17].

Considering restrictions of sensors and actuators, most of the solutions (38 out of 47 papers, see Table 2) look for efficient encryption methods. The most studied algorithms are secret key encryption and elliptic curve cryptography. Also, physiological values and channel characteristics are used as seeds to generate encryption keys.

Data encryption: Half of the papers that address data encryption use secret key cryptography [6,18–20,34,35,49], while the other half use Elliptic Curve Cryptography [25,30,33,36,37,45,52].

1) Secret Key Cryptography: Secret key algorithms are more suitable for BAN architectures than asymmetric key algorithms, because they use shorter key lengths, thus requiring shorter random numbers and less computational and energy resources [37]. In addition, symmetric encryption and decryption procedures are faster, making this algorithm better for emergency cases, where doctors will need to retrieve data as fast as possible. On the other hand, secret key algorithms must resolve the problem of key-distribution.

2) Elliptic Curves Cryptography (ECC): Since 2010 ECC has gained research interest (see Table 1). ECC is suitable for BAN architectures because it uses small keys; a 160-bit ECC key is as strong as a 1024-bit RSA key [33]. According to the NIST [9], a 2048 RSA key is equivalent to a 224 ECC key [37]. ECC keys can be distributed using protocols such as Diffie-Hellman. Furthermore, these keys can be used to create digital signatures for authentication purposes [25,30,33,36,37,45,52]. However, ECC implementations still must handle unsolved problems, including the creation of a random number generator for private keys, and the distribution of initial parameters [54].

Key Generation:

1) Physiological Values (PVs): PVs are used by around a third of the studied proposals, to generate encryption keys [11,26,40,42,44,46,50,59–61,64,66,67,69]. Some PVs are used as seed for key generation because

   a. They are universal, as the majority of population have them;

   b. Two people do not share the same PVs;

   c. They are easy to collect and to measure;

   d. They are adequate for low computational power devices;

   e. They are difficult to reproduce;

   f. They are random [43,62].

The Heart Rate Variability (HRV) is the most used PV. Several sensors, such as electrocardiograms (ECG or EKG) and photoplethysmograms (PPG), can measure it. An alternative PV is body acceleration where motion sensors measure body movements [41]. Not all PVs are good seeds for key generation because their possible values are not as variable as the HRV [62]. For example, blood glucose, blood pressure and temperature values are expected to be within a predefined small range. Some advantages of using PVs are:

1) Sensors do not need to generate random numbers reducing processing and power consumption;

2) Key security is improved; keys do not have to be distributed as all sensors for the same patient will be able to use the same PV to generate a shared encryption key.

Although PVs have several advantages there are several issues that must be solved before they are more widely accepted:

1) Some PVs can be remotely measured [3, 14], giving unauthorized devices the possibility of generating the shared key;

2) BAN architectures with diverse sensors, measuring different PVs, cannot have a single shared key for all the devices;

3) Personal servers would need access to a given PV to be able to create a key shared with the sensors.

Channel characteristics: A different approach, less used, is to handle channel characteristics for key generation.

One of the used characteristics is the received signal strength indicator (RSSI), a wireless channel feature [4, 5, 56, 58, 70]. However, body movements can affect the strength of the signals produced by implanted sensors as the waves are diffracted and trapped along the skin surface. The environment and involuntary movements such as respiration and heartbeat also affect signal strength, setting the variance of the RSSI values [56, 70]. There is one important advantage, since devices measure RSSI by default, there is no need to use computational resources for key generation.

A different approach, the Body-Coupled Communication (BCC), uses the human body as the communication channel [7]. Some researchers state that BCC may prevent several attacks because attackers would need to be very close to their target to be able to communicate [34]. This approach however, does not consider how to protect communications between sensors or actuators and a personal server.

3.1.2 Key Distribution

More than half of the studied research projects choose an available key-distribution algorithm to deliver secret keys in BAN environments. The most used algorithms are Diffie-Hellman and Fuzzy Vaults (see Table 2) .

1) Diffie-Hellman: A quarter of the papers that address key distribution use Diffie-Hellman for key exchange [25, 30, 33, 35–37]. In particular, 5 of these 6 papers adapt the algorithm to use it with ECC to create a shared secret key using public information derived from the keys generated using an elliptic curve. To use Diffie-Hellman with ECC, two devices need to agree about the curve parameters. With these parameters, each device

   a. Calculates a random number that will work as the private key;

   b. Calculates a point in the curve. This point, multiplied by the private key, will be the public key. The shared key will be a device's private key multiplied by the other device's public key.

2) Fuzzy Vault: Around a fifth of the studied papers use this method for key distribution. In this scheme, a user A hides a secret key (K_a) using a set of values $Set_a = \{a_1, a_2, a_3...a_n\}$. A different user, B, has another set of values $Set_b = \{b_1, b_2, b_3...b_n\}$. User B can obtain the secret key K_a if enough values in Set_b correspond to the values in Set_a [29].

   In BANs, fuzzy vaults are used to distribute secret keys generated with PVs and channel characteristics. In particular, some proposals use PVs to create the fuzzy vault that protects a secret key [11, 40–42, 60, 61, 69]. In [70] and [58], the authors use channel characteristics to create the sets for the fuzzy vault. Additionally, in [27] an enhanced fuzzy vault scheme is used to achieve access control. Fuzzy vaults are adequate because they can handle small errors in the measurements of PVs and channel features; users need to provide some of the values, but not necessarily all of them.

3) Other algorithms: The remaining proposals use other key-distribution algorithms. Among them, Distribution centers, with one node in charge of delivering keys to other devices, is the most used protocol [6, 23, 33, 36, 45]. Some proprietary protocols are also used [52, 59].

   The Internet Security Association and Key Management Protocol (ISAKMP) is also used to implement key exchange procedures and create encrypted connections between two endpoints [39]. Although ISAKMP may be used as a security framework in BAN scenarios, a previous study (where personal servers, in a patient's home, forward medical data, measured by sensors, to a hospital), showed that implementing this protocol increases bandwidth and energy consumption [13].

4) Key Agreement: Some proposals [26, 61] use physiological values and channel features to run a predefined algorithm and generate a shared secret key. Keys are generated, they do not need to be distributed. Some solutions for key agreement also include notifying a patient when a key agreement procedure is occurring in the network; for example, generating a brief vibration [19].

3.2 Access Control

Access control must guarantee that only authorized entities; users, processes or devices; will have access to data collected, forwarded and stored by BAN devices. To guarantee access control, two requirements must be addressed: authentication and authorization.

3.2.1 Authentication

Authentication allows a BAN to establish the identity of a given component, stopping devices that do not belong to a BAN from gaining access to private data. Attackers may pose as a legitimate device, like a sensor or a personal server, to eavesdrop, steal, or send erroneous information, possibly affecting sensors and actuators functionality [12, 32, 38, 48].

Most of the studied proposals (around 83%) present authentication protocols. These protocols may work in conjunction with a key-agreement protocol or may work independently. For example, some authors propose using PVs, channel characteristics or devices' identifications to achieve authentication; if a particular sensor can measure a defined PV, that sensor must be implanted or have physical contact with a patient and should be authenticated as a component of a given BAN [11, 26, 40–44, 46, 50, 59–62, 64, 66, 67, 69].

However, new techniques for measuring a PV without physical contact with the user are emerging. In one example authors implemented two methods for retrieving HRV from videos of human faces [3]. Another example implemented a microwave Doppler for non-contact through-clothing measurement of chest wall movements to obtain heart and respiration rates [14]. Although currently these techniques are not widely used, they suggest that authentication based on proximity may not be enough in the future.

Proposals that use channel characteristics for authentication also assume proximity; only legitimate sensors would be attached to a user and could share the same communication channel in order to have similar RSSI values [5, 34, 55, 56, 58, 63, 70].

Other proposals use a device's identification number for authentication; during an installation phase the id number is registered as part of a group. Later, that device sends its identification and a BAN node, in charge of the authentication, checks if that ID belongs to the group [5, 6, 17, 19, 33, 35, 36, 45, 49, 52]. These approaches may not be enough because identifiers may be faked.

Ho [25] evaluates three authenticated key agreement protocols for Intra-BAN communication: out-of-ban public key exchange, where the devices send their public keys over a secured separate channel. A password to alter the shared key, so only entities with the password can access the key. A numerical display, where a hash is used to guarantee that the other party has the necessary key to obtain the same hash. The implementations of these protocols are found to be resistant against impersonation and man-in-the-middle attacks; additionally, the use of the password protocol is strong to offline dictionary attacks. The author claims that these protocols have been adopted into the IEEE standard on BAN [28].

Previous protocols do not explicitly consider movement. If a person can move, authentication and authorization may be more difficult as sensors and communications would need to switch from one access point to another. In this case the authentication protocol should be able to manage re-authentication to provide the same set of established services at the second access point [65].

3.2.2 Authorization

Authorization requirements restrict access to a patient's medical information according to predefined access rules. For instance, a hospital may have several BANs to monitor several patients storing all data in the same server. However, not all doctors and nurses should have access to information of all patients, only medical personal directly involved with a patient should have access to his or her information. A BAN must implement authorization mechanisms to present data only to authorized entities, like a patient's medical team. In addition, an authorized entity should have access exclusively to needed information; for example, doctors should have access to all the information about the patient, but a pharmacist should only have access to drug prescriptions. A role-based access control is, therefore, necessary in a BAN architecture with multiple users [32].

An approach suggests the creation of behavioral profiles based on access patterns to and from devices in a BAN. Only access requests that are consistent with the profiles are allowed. An authorization mechanism may perform mitigation strategies to control inconsistent requests including passive actions like generating alerts or active actions like jamming the signal to deny access to data [68]. An alternative approach builds behavioral profiles based on places and times. Users, including doctors and nurses, only are allowed to access information from particular locations, such as consulting rooms and hospitals, at specific hours [24].

A different approach uses access policies based on attributes. Every user is assigned a set of attributes (n) and a minimum threshold for authorization (d) is established. If a user has (d) out of (n) attributes, then he or she is authorized to access a piece of information from the BAN personal server [27, 49].

Finally, some proposals use additional devices to perform authentication tasks; an additional device may be used as a proxy for communications among sensors and personal servers, and it allows or denies access requests [66].

Regarding intra-BAN components and communications, one approach is to authenticate and authorize sensors and actuators using proximity. Only devices in close proximity or with physical contact to the human body are authorized to obtain information from sensors [44, 46].

We found that only a few of the studied proposals ad-

Table 2: Security requirements addressed by the studied projects. Encryption, Key Distribution and Authentication are the most studied requirements. (Conventions. Sum: Summary, ECC: Elliptic Curve Cryptography, PV: Physiological Values, CC: Channel Characteristics, Sec: Secret Keys)

Papers	Security Requirements													
	Confidentiality									Access Control		Integrity		Availability
	Encryption					Key Distribution								
	Sum	ECC	PV	CC	Sec	**Sum**	Fuzzy Vault	Diffie-Hellman	Other	Authentication	Authorization	Hash Functions	Session Management	DoS Attack Protection
[43]	●	-	●	-	-	-	-	-	-	●	-	●	-	-
[52]	●	●	-	-	-	●	-	-	●	●	-	-	●	-
[20]	●	-	-	-	●	-	-	-	-	-	-	-	-	-
[25]	●	●	-	-	-	●	-	●	-	●	-	-	-	-
[42]	●	-	●	-	-	●	●	-	-	●	-	-	-	-
[64]	●	-	●	-	-	-	-	-	-	●	-	●	-	-
[22]	-	-	-	-	-	-	-	-	-	●	-	-	●	-
[45]	●	●	-	-	-	●	-	-	●	●	-	-	-	-
[27]	-	-	-	-	-	●	●	-	-	-	●	●	●	-
[23]	-	-	-	-	-	●	-	-	●	-	-	●	-	-
[70]	●	-	-	●	-	●	●	-	-	●	-	●	●	-
[69]	●	-	●	-	-	●	●	-	-	●	-	●	-	-
[13]	-	-	-	-	-	●	-	-	●	●	-	-	-	-
[30]	●	●	-	-	-	●	-	●	-	●	-	-	●	-
[37]	●	●	-	-	-	●	-	●	-	-	-	-	-	-
[63]	●	-	-	●	-	-	-	-	-	●	-	-	-	-
[6]	●	-	-	-	●	●	-	-	●	●	-	-	●	-
[49]	●	-	-	-	●	-	-	-	-	●	●	-	-	-
[36]	●	●	-	-	-	●	-	●	●	●	●	-	-	-
[50]	●	-	●	-	-	-	-	-	-	●	-	-	-	-
[18]	●	-	-	-	●	-	-	-	-	-	-	-	-	-
[58]	●	-	-	●	-	●	-	-	●	●	-	-	-	-
[2]	-	-	-	-	-	-	-	-	-	-	-	-	-	●
[33]	●	●	-	-	-	●	-	●	●	●	-	●	-	-
[5]	●	-	-	●	-	-	-	-	-	●	-	●	●	-
[53]	-	-	-	-	-	-	-	-	-	-	-	-	-	●
[41]	●	-	●	-	-	●	●	-	-	●	-	-	-	-
[4]	●	-	-	●	-	-	-	-	-	-	-	-	-	-
[19]	●	-	-	-	●	●	-	-	●	●	-	-	-	-
[34]	●	-	-	●	●	-	-	-	-	●	-	●	●	-
[46]	●	-	●	-	-	-	-	-	-	●	●	-	-	-
[26]	●	-	●	-	-	●	-	-	●	●	-	-	-	-
[56]	●	-	-	●	-	-	-	-	-	●	-	-	-	-
[68]	-	-	-	-	-	-	-	-	-	●	●	-	●	-
[55]	-	-	-	-	-	-	-	-	-	●	-	-	-	-
[61]	●	-	●	-	-	●	●	-	-	●	-	-	●	-
[17]	●	-	-	-	-	-	-	-	-	●	-	-	-	-
[62]	●	-	●	-	-	-	-	-	-	●	-	●	●	-
[66]	●	-	●	-	-	-	-	-	-	●	●	-	●	-
[60]	●	-	●	-	-	●	●	-	-	●	-	-	●	-
[44]	●	-	●	-	-	-	-	-	-	●	●	-	-	●
[59]	●	-	●	-	-	●	-	-	●	●	-	-	-	-
[11]	●	-	●	-	-	●	●	-	-	●	-	●	-	-
[35]	●	-	-	-	●	●	-	●	-	●	-	●	-	-
[40]	●	-	●	-	-	●	●	-	-	●	-	-	-	-
[67]	●	-	●	-	-	-	-	-	-	●	-	-	-	-
[24]	-	-	-	-	-	-	-	-	-	●	●	-	-	●
Total	38	7	17	7	7	24	9	6	11	39	8	12	13	4

dress authorization requirements (see Tables 1 and 2). However, as previously mentioned, not every agent in a BAN should have access to all data.

3.3 Integrity

Attackers may use several methods to modify a packet; they may capture and edit a packet, and then forward it to a server, or create radio interference to alter bits before a packet reaches a destination [38]. Interference by natural reasons is also possible. There are various consequences; an actuator that receives modified commands will not act according to the actual situation, and an application that receives erroneous data will generate false alarms. In any case, a secure BAN architecture should guarantee that data have not been modified during transmission or storage [12,32,38,48,51].

In addition to unauthorized modifications detection, a BAN also needs to avoid Replay Attacks. In replay attacks adversaries resend/replay old packets trying to make servers believe those packets are valid, possibly generating false alarms or failing to generate warnings. To prevent replay attacks, personal and external servers should evaluate data freshness, a property that indicates if the received information is recent and arrives when expected [32,38,48,51]. To support integrity and avoid replay attacks, the studied proposals use hash functions and session identifiers.

Hash functions. Hash functions are used to verify integrity of a message or stored data by calculating a fixed-size number for a data stream.

Around a quarter of the studied proposals (see Table 2) use hash values to protect messages with medical data. Almost half of these proposals use one of the following hash algorithms: SHA-1, SHA-256, MD5, cyclic redundancy check (CRC) and digests, while the other half use Message Authentication Codes (MAC) [5,27,33–35,69].

A different approach is to use external resources to support integrity checks of stored data. One option is to delegate integrity evaluation, of information stored in external servers, to cloud computing services [23].

Session Management: Replay attacks can be prevented by using session identifiers, such as random numbers [17,30,60,66,70] and timestamps [6,22,27], or by attaching a device ID and a data counter to every message to keep track of arrived messages and avoid repeated ones [5]. It is worth mentioning that only using a device ID and a counter is a technique that may be vulnerable to some attacks, as these values can be guessed.

One of the studied proposals [52] uses databases to store hash values of every received message. If the hash value of an arriving message already is in the database then the message is discarded. Alternatively, in [68], a sequence number is added to every message to track repeated or missing packets in a communication session. Channel characteristics may also be evaluated to find anomalies. If an anomaly is detected, then suspicious packets are jammed.

3.4 Availability

This security requirement, in a BAN, aims to guarantee that data and devices are available whenever they are needed. Physiological and medical information must be available when needed and during emergencies for doctors, nurses and paramedics [31,32,47,48]. Each and every component must be available; if a network does not have enough capacity to transfer all packets, then servers cannot receive data on time [38,47]. If other components, like sensors or servers, are compromised then information cannot be generated or received and warnings cannot be generated.

This requirement is the least studied; only a few of the considered works use protocols to avoid attacks on availability, like Denial-of-Service (DoS) attacks. Two methods were proposed to detect and mitigate these attacks. In the first one, Adaptive Network Profiles, authors create profiles of normal behavior based on different network characteristics like QoS (Quality of Service), traffic patterns and power consumption [2,53]. To detect abnormal behavior, the network is constantly monitored, if an atypical behavior is detected, such as a decrease on QoS or increase in energy consumption, then corrective actions are performed.

The second method works by controlling high energy consumption tasks [24,44]. DoS attacks tend to rapidly drain sensor's resources; data transmission tasks are particularly expensive in energy consumption [44]. To avoid DoS attacks that send high amounts of data, authors created procedures based on proximity; sensors will share information only with devices that are close to the human body. Data transmission only occurs in specific scenarios at controlled environments, thus reducing energy consumption.

While the first approach is designed to protect components and communications in the beyond part of a BAN, the second one is designed to protect the intra-BAN part. Both approaches should be managed within a single framework to guarantee consistency and protect a BAN as a whole.

Authentication procedures also consume more power than other tasks. To avoid battery draining in this case, an approach suggests using profiles; devices only accept communication from predefined devices at specific locations and times [24].

3.5 Software correctness

Sensors and implantable medical devices are controlled by software, and there is always a probability of having software bugs [47]. The Medical Device Recall Report, written by the US Food and Drug Administration (FDA) [16],

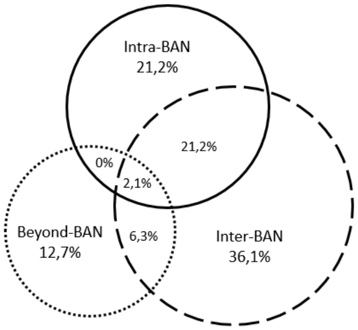


Figure 2: Distribution of works according to addressed BAN communications (Intra-BAN, Inter-BAN and Beyond-BAN communications)

Table 3: Works that secure specific BAN components; Most of the available works focus on sensors and personal servers

Secured BAN Component	***# of Works***	***%***
Sensors	13	27.65%
Sensors and Personal Server	25	53.19%
Personal Server	0	0%
Personal and External Server	4	8.51%
External Server	1	2.12%
External Server and Cloud	1	2.12%
Cloud	2	4.25%
Sensors, Personal and External Server	1	2.12%
Total	47	

states that software design flaws due to lack of proper testing procedures caused the recall of 429 devices. Software design flaws is the main reason to recall devices.

Furthermore, the entire lifecycle of these programs should be managed, not only their design and testing phases. Firmware and software updates must have adequate procedures to prevent the deployment of external firmware or software that may harm or allow unauthorized access to sensors and medical devices [21]. However, we found that none of the studied references consider this aspect.

4 Architecture Analysis

As already mentioned, we classified the studied proposals according to the BAN components they consider and the communications they protect. Table 3 classifies the works based on protected components, while Figure 2 classifies them based on protected communications. Around half of the studied proposals protect two BAN components: sensors and personal servers and their communications.

The interest in Intra-BAN and Inter-BAN communications may be explained because the devices that perform these communications (sensors and actuators) have processing and storage restrictions that have been addressed but are not completely solved. Additionally, these devices use recent technology, presenting new security issues that need to be considered.

In some BAN implementations, sensors may communicate among themselves, not only with the personal server. Around a fifth of the proposals (21.2%) address security of this kind of communications (Intra-BAN communication).

None of the proposals exclusively addressed security of personal servers. This situation may be explained as personal servers are not used for information-gathering or storage tasks, but as gateways between sensors and external servers. However, communications need to be secured and Figure 2 shows that half of the proposals (53.1%) secure communications that involve personal servers.

Few proposals address external servers. This is expected as personal servers usually have good processing and storage capacity. Consequently, traditional security solutions could be used. However, a comprehensive BAN solution must be able to integrate traditional solutions and solutions for devices with restricted resources like sensors and actuators.

We also examined if the proposals considered single or multi-user environments. The proposals that secure a single BAN, one that only involves one patient, are considered as single-user environments. 83% of the authors consider this configuration. On the other hand, a multi-user environment involves multiple patients, their sensors and personal servers send information to a centralized server, typically, a hospital's server. Only 17% of the proposals considered this scenario.

Multi-user BAN environments are relevant because they correspond to real health-care scenarios, like hospitals. Some security issues that have not been explicitly considered emerge in these environments; for instance, a server will need to manage key-generation and key-distribution for different patients.

5 Open Issues

Cloud Computing as a BAN Component: Considering that BANs handle medical information and there are privacy protection standards like The Health Insurance Portability and Accountability Act (HIPAA), BANs must support the requirements standards and legislation have defined. Due to these requirements, the use of cloud computing might be controversial in health related services, as protection and storage of medical information partially depends upon third-party infrastructures and policies.

Currently, cloud computing solutions are not commonly included as part of BANs but are starting to appear. A few of the studied projects considered cloud computing to support medical studies. The focus of these projects is protecting the communication channel between external and cloud servers, and protecting the information stored in the cloud.

Table 4: Percentage of proposals that consider single-user and multi-user environments.

Environment	***# of Works***	***Percentage***
Single-user	39	82.9%
Multi-user	8	17.1%

The proposals presented in [23] and [36] use cloud as a supporting tool to check integrity of a patient's medical information. Other authors secure the communication channel between a BAN and the cloud; for example, in [20] the authors propose a Multi-valued and Ambiguous Scheme to create a cryptographic system, based on secret keys, in order to perform this task.

One proposal addresses the need to support authorization in the cloud; different users should have access to different data according to their particular roles [57]. However, we need more works that study how to support authorization and authentication to grant access to medical records stored in the cloud.

Sensors are the focus of the majority of proposals: As Table 3 shows, most of the projects address security of sensors and personal servers, these are the main topic because the addition of software to control sensors and their role as part of BANs is relatively new. On the other hand, how to comprehensively secure external servers and cloud services that belong to a BAN and store medical data is not a well explored subject.

Multi-user Environment: Most of the studied works protect a single BAN architecture for one patient. They do not consider multi-user BANs, like in the case of a hospital that collects and stores medical information from several patients and must provide different types of access for different physicians and nurses. A few of the authors deal with multi-user environments, and only a small part of these (17.1%, see Table 4) use and try to secure cloud resources [20, 23, 36, 65].

Authorization: Few works consider this subject. One work proposes implementing role-based authorization for personal servers [27], while another one proposes using profiles, based on information like proximity, to allow access to sensor data [44].

Software Correctness: None of the studied works proposes mechanisms to check software correctness. As previously mentioned, according to the US Food and Drug Administration (FDA), software is the main cause for medical devices recalling [16]. This situation happens due to poor procedures to handle software design, update and testing. Therefore, it seems necessary to build methodologies and frameworks to support the development of correct and secure software.

Comprehensive Approach: Most of the analyzed proposals address security of one or two BAN components, they however do not consider the remaining components. None of the studied works addresses the entire environment in a comprehensive way, considering features and requirements across all the components. This view is needed as the security requirements of the collected data do not change depending on the component holding them.

6 Conclusions

In this paper, we made a bibliographic review of security requirements and proposals for BAN architectures. We identified the usual security requirements: Confidentiality, integrity and availability, as well as others like authentication, authorization, data freshness and software correctness.

There are various proposals that address these requirements. We classified them according to the BAN components they protect: sensors, actuators, personal servers, external servers and cloud services; and according to the communications they protect: intra-BAN, inter-BAN and Beyond-BAN. We found that most of the studied proposals only consider one or two BAN components.

We found that approximately 80% of the studied proposals exclusively focus on securing sensors and/or personal servers. The remaining proposals, around 20%, secure external servers and cloud services. Only one proposal considered all components. We argue that a comprehensive view is needed for several reasons. First, the security requirements of medical related data do not change according to the part of the BAN that is holding them. Second, deployed solutions must consider the particular aspects of Intra-BAN, Inter-BAN and Beyond-BAN contexts and communications, but they must be consistent; for example, if a sensor (intra-BAN component) needs to communicate with a personal server (Inter-BAN component), they must establish a protected communication channel. Similarly, a personal server must protect its communications with external servers. Third, a BAN must handle the life cycle of all the algorithms its components run. Finally, when considering a multi-user environment, a BAN external server will need to support security guarantees for several patients.

Acknowledgments

This work was partially supported by the Colombian Administrative Department of Science, Technology, and Innovation (Colciencias).

References

[1] CEN/TC 251, "Business Plan 2015-2018. Health Informatics," 2015. (file:///C:/Users/user/Downloads/

N15-019_Business_Plan_2015-2018_final_clean_20150423.pdf)

[2] H. Abie and I. Balasingham, "Risk-based adaptive security for smart IoT in eHealth," in *7th International Conference on Body Area Networks*, pp. 269–275, 2012.

[3] K. Alghoul, S. Alharthi, H. A. Osman, and A. E. Saddik, "Heart rate variability extraction from videos signals: ICA vs. EVM Comparison," *IEEE Access*, vol. 5, pp. 4711–4719, 2017.

[4] S. T. Ali, V. Sivaraman, and D. Ostry, "Zero reconciliation secret key generation for body-worn health monitoring devices," in *Fifth ACM Conference on Security and Privacy in Wireless and Mobile Networks (WISEC'12)*, pp. 39–50, 2012.

[5] S. T. Ali, V. Sivaraman, D. Ostry, and S. Jha, "Securing data provenance in body area networks using lightweight wireless link fingerprints," in *3rd International Workshop on Trustworthy Embedded Devices (TrustED'13)*, pp. 65–72, 2013.

[6] H. Alyami, J. L. Feng, A. Hilal, and O. Basir, "On-demand key distribution for body area networks for emergency case," in *12th ACM International Symposium on Mobility Management and Wireless Access (MobiWac'14)*, pp. 55–58, 2014.

[7] H. Baldus, S. Corroy, A. Fazzi, K. Klabunde, and T. Schenk, "Human-centric connectivity enabled by body-coupled communications," *IEEE Communications Magazine*, vol. 47, pp. 172–178, June 2009.

[8] D. M. Barakah and M. Ammad-uddin, "A survey of challenges and applications of wireless body area network (WBAN) and role of a virtual doctor server in existing architecture," in *Third International Conference on Intelligent Systems Modelling and Simulation*, pp. 214–219, Feb. 2012.

[9] E. B. Barker, A. L. Roginsky, "Transitions: Recommendation for transitioning the use of cryptographic algorithms and key lengths," *National Intitute of Standards and Technology (NIST'15)*, Nov. 2015.

[10] M. Chen, S. Gonzalez, A. Vasilakos, H. Cao, and V. C. M. Leung, "Body area networks: A survey," *Mobile Networks and Applications*, vol. 16, no. 2, pp. 171–193, 2011.

[11] S. Cherukuri, K. K. Venkatasubramanian, and S. K. S. Gupta, "Biosec: A biometric based approach for securing communication in wireless networks of biosensors implanted in the human body," in *International Conference on Parallel Processing Workshops*, pp. 432–439, Oct. 2003.

[12] J. Chukwunonyerem, A. M. Aibinu, and E. N. Onwuka, "Review on security of wireless body area sensor network," in *11th International Conference on Electronics, Computer and Computation (ICECCO'14)*, pp. 1–10, Sep. 2014.

[13] R. Divya, T. V. P. Sundararajan, and K. Deepak, "Effect of wormhole attack in hierarchical body area network and need for strict security measures," in *6th International Conference on Computing, Communication and Networking Technologies (ICCCNT'15)*, pp. 1–7, July 2015.

[14] A. D. Droitcour, *Non-Contact Measurement of heart and Respiration Rates with a Single-chip Microwave Doppler Radar*, PhD thesis, Stanford University, 2006.

[15] European Patients' Forum, "The new EU Regulation on the protection of personal data: What does it mean for patients?," (http://www.eu-patient.eu/globalassets/policy/data-protection/data-protection-guide-for-patients-organisations.pdf)

[16] A. Ferriter, *Medical Device Recall Report FY 2003 to FY 2012*, US Food and Drug Administration, Technical Report, 2012.

[17] S. Gollakota, H. Hassanieh, B. Ransford, and K. Katabi, D.and Fu, "They can hear your heartbeats: Non-invasive security for implantable medical devices," *SIGCOMM Computer Communication Review*, vol. 41, pp. 2–13, Aug. 2011.

[18] R. M. Gomathi, A. S. Sangari, J. M. L. Manickam, "RC6 based security in wireless body area network," *Journal of Theoretical and Applied Information Technology*, vol. 74, no. 1, Apr. 2015.

[19] D. Halperin, T. S. Heydt-Benjamin, B. Ransford, S. S. Clark, B. Defend, W. Morgan, K. Fu, T. Kohno, and W. H. Maisel, "Pacemakers and implantable cardiac defibrillators: Software radio attacks and zero-power defenses," in *IEEE Symposium on Security and Privacy (SP'08)*, pp. 129–142, May 2008.

[20] N. D. Han, L. Han, D. M. Tuan, H. Peter In, and M. Jo, "A scheme for data confidentiality in cloud-assisted wireless body area networks," *Information Sciences*, vol. 284, pp. 157 – 166, 2014.

[21] S. Hanna, R. Rolles, A. Molina-Markham, P. Poosankam, K. Fu, and D. Song, "Take two software updates and see me in the morning: The case for software security evaluations of medical devices," in *2nd USENIX Conference on Health Security and Privacy (HealthSec'11)*, pp. 6, 2011.

[22] D. He, S. Zeadally, N. Kumar, and J. H. Lee, "Anonymous authentication for wireless body area nNetworks with provable security," *IEEE Systems Journal*, vol. 11, pp. 2590–2601, Dec. 2017.

[23] D. He, S. Zeadally, and L. Wu, "Certificateless public auditing scheme for cloud-assisted wireless body area networks," *IEEE Systems Journal*, vol. 12, no. 1, pp. 64–73, 2018.

[24] X. Hei, X. Du, J. Wu, and F. Hu, "Defending resource depletion attacks on implantable medical devices," in *IEEE Global Telecommunications Conference (GLOBECOM'10)*, pp. 1–5, Dec. 2010.

[25] J. M. Ho, "A versatile suite of strong authenticated key agreement protocols for body area networks," in *8th International Wireless Communications and Mobile Computing Conference (IWCMC'12)*, pp. 683–688, Aug. 2012.

[26] C. Hu, X. Cheng, F. Zhang, D. Wu, X. Liao, and D. Chen, "OPFKA: Secure and efficient ordered-physiological-feature-based key agreement for wireless body area networks," in *IEEE INFOCOM*, pp. 2274–2282, Apr. 2013.

[27] C. Hu, N. Zhang, H. Li, X. Cheng, and X. Liao, "Body area network security: A fuzzy attribute-based signcryption scheme," *IEEE Journal on Selected Areas in Communications*, vol. 31, pp. 37–46, Sep. 2013.

[28] IEEE, "IEEE standard for local and metropolitan area networks," *Wireless Body Area Networks*, Feb. 2012.

[29] A. Juels and M. Sudan, "A fuzzy vault scheme," *Designs, Codes and Cryptography*, vol. 38, no. 2, pp. 237–257, 2006.

[30] S. L. Keoh, "Efficient group key management and authentication for body sensor networks," in *IEEE International Conference on Communications (ICC)*, pp. 1–6, June 2011.

[31] M. Kumar, "Security issues and privacy concerns in the implementation of wireless body area network," in *International Conference on Information Technology*, pp. 58–62, Dec. 2014.

[32] P. Kumar and H. J. Lee, "Review security issues in healthcare applications using wireless medical sensor networks: A survey," *Sensors*, vol. 12, Dec. 2011.

[33] Y. S. Lee, E. Alasaarela, and H. Lee, "Secure key management scheme based on ECC algorithm for patient's medical information in healthcare system," in *The International Conference on Information Networking (ICOIN'14)*, pp. 453–457, Feb. 2014.

[34] C. Li, A. Raghunathan, and N. K. Jha, "Hijacking an insulin pump: Security attacks and defenses for a diabetes therapy system," in *IEEE 13th International Conference on e-Health Networking, Applications and Services*, pp. 150–156, June 2011.

[35] M. Li, S. Yu, J. D. Guttman, W. Lou, and K. Ren, "Secure Ad Hoc trust initialization and key management in wireless body area networks," *ACM Transactions on Sensor Networks*, vol. 9, pp. 18:1–18:35, Apr. 2013.

[36] S. Li, Z. Hong, and C. Jie, "Public auditing scheme for cloud-based wireless body area network," in *IEEE/ACM 9th International Conference on Utility and Cloud Computing (UCC)*, pp. 375–381, Dec. 2016.

[37] J. Liu and K. S. Kwak, "Hybrid security mechanisms for wireless body area networks," in *Second International Conference on Ubiquitous and Future Networks (ICUFN'10)*, pp. 98–103, June 2010.

[38] V. Mainanwal, M. Gupta, and S. K. Upadhayay, "A survey on wireless body area network: Security technology and its design methodology issue," in *International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS'15)*, pp. 1–5, Mar. 2015.

[39] D. Maughan, M. Schertler, M. Schneider, and J. Turner, *Internet Security Association and Key Management Protocol*, RFC 2408 (Proposed Standard), Nov. 1998.

[40] F. Miao, L. Jiang, Y. Li, and Y. T. Zhang, "Biometrics based novel key distribution solution for body sensor networks," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 2458–2461, Sep. 2009.

[41] D. Oberoi, W. Y. Sou, Y. Y. Lui, R. Fisher, L. Dinca, and G. P. Hancke, "Wearable security: Key derivation for body area sensor networks based on host movement," in *IEEE 25th International Symposium on Industrial Electronics (ISIE'16)*, pp. 1116–1121, June 2016.

[42] R. T. Rajasekaran, V. Manjula, V. Kishore, T. M. Sridhar, and C. Jayakumar, "An efficient and secure key agreement scheme using physiological signals in body area networks," in *International Conference on Advances in Computing, Communications and Informatics (ICACCI'12)*, pp. 1143–1147, 2012.

[43] S. N. Ramli, R. Ahmad, M. F. Abdollah, and E. Dutkiewicz, "A biometric-based security for data authentication in wireless body area network (WBAN)," in *15th International Conference on Advanced Communications Technology (ICACT'13)*, pp. 998–1001, Jan. 2013.

[44] K. B. Rasmussen, C. Castelluccia, T. S. Heydt-Benjamin, and S. Capkun, "Proximity-based access control for implantable medical devices," in *16th ACM Conference on Computer and Communications Security (CCS'09)*, pp. 410–419, 2009.

[45] C. Rong and H. Cheng, "Authenticated health monitoring scheme for wireless body sensor networks," in *7th International Conference on Body Area Networks (BodyNets'12)*, pp. 31–35, 2012.

[46] M. Rostami, A. Juels, and F. Koushanfar, "Heart-to-heart (H2H): Authentication for implanted medical devices," in *ACM SIGSAC Conference on Computer and Communications Security (CCS'13)*, pp. 1099–1112, 2013.

[47] M. Rushanan, A. D. Rubin, D. F. Kune, and C. M. Swanson, "SoK: Security and privacy in implantable medical devices and body area networks," in *IEEE Symposium on Security and Privacy*, pp. 524–539, May 2014.

[48] S. Saleem, S. Ullah, and K. S. Kwak, "Towards security issues and solutions in wireless body area networks," in *6th International Conference on Networked Computing (INC'10)*, pp. 1–4, May 2010.

[49] A. S. Sangari and J. M. Leo, "Polynomial based light weight security in wireless body area network," in *IEEE 9th International Conference on Intelligent Systems and Control (ISCO'15)*, pp. 1–5, Jan. 2015.

[50] A. S. Sangari and J. M. L. Manickam, "Public key cryptosystem based security in wireless body area network," in *International Conference on Circuits, Power and Computing Technologies (ICCPCT'14)*, pp. 1609–1612, Mar. 2014.

[51] S. Sangari and M. Manickam, "Security and privacy in wireless body area network," *Indian Streams Research Journal*, vol. 4, no. 8, 2014.

[52] M. Sarvabhatla, M. C. M. Reddy, and C. S. Vorugunti, "A robust biometric-based authentication scheme for wireless body area network using elliptic curve cryptosystem," in *Third International Symposium on Women in Computing and Informatics (WCI'15)*, pp. 582–587, 2015.

[53] R. M. Savola, H. Abie, and M. Sihvonen, "Towards metrics-driven adaptive security management in e-Health IoT applications," in *7th International Conference on Body Area Networks (BodyNets'12)*, pp. 276–281, 2012.

[54] P. G. Shah, X. Huang, and D. Sharma, "Analytical study of implementation issues of elliptical curve cryptography for wireless sensor networks," in *IEEE 24th International Conference on Advanced Information Networking and Applications Workshops*, pp. 589–592, Apr. 2010.

[55] L. Shi, M. Li, S. Yu, and J. Yuan, "BANA: Body area network authentication exploiting channel characteristics," in *Fifth ACM Conference on Security and Privacy in Wireless and Mobile Networks (WISEC'12)*, pp. 27–38, 2012.

[56] L. Shi, J. Yuan, S. Yu, and M. Li, "ASK-BAN: Authenticated secret key extraction utilizing channel characteristics for body area networks," in *Sixth ACM Conference on Security and Privacy in Wireless and Mobile Networks (WISEC'13)*, pp. 155–166, 2013.

[57] Y. Tian, Y. Peng, G. Gao, and X. Peng, "Role-based access control for body area networks using attribute-based encryption in cloud storage," *International Journal of Network Security*, vol. 19, 2017.

[58] G. R. Tsouri and J. Wilczewski, "Reliable symmetric key generation for body area networks using wireless physical layer security in the presence of an On-body eavesdropper," in *4th International Symposium on Applied Sciences in Biomedical and Communication Technologies (ISABEL'11)*, pp. 153:1–153:6, 2011.

[59] K. K. Venkatasubramanian, A. Banerjee, and S. K. S. Gupta, "Ekg-based key agreement in body sensor networks," in *IEEE INFOCOM Workshops*, pp. 1–6, Apr. 2008.

[60] K. K. Venkatasubramanian, A. Banerjee, and S. K. S. Gupta, "Plethysmogram-based secure inter-sensor communication in body area networks," in *IEEE Military Communications Conference (MILCOM'08)*, pp. 1–7, Nov. 2008.

[61] K. K. Venkatasubramanian, A. Banerjee, and S. K. S. Gupta, "PSKA: Usable and secure key agreement scheme for body area networks," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, pp. 60–68, Jan. 2010.

[62] K. K. Venkatasubramanian and S. K. S. Gupta, "Physiological value-based efficient usable security solutions for body sensor networks," *ACM Translations on Sensor Network*, vol. 6, pp. 31:1–36, 2010.

[63] S. Venkatasubramanian and V. Jothi, "Integrated authentication and security check with CDMA modulation technique in physical layer of wireless body area network," in *IEEE International Conference on Computational Intelligence and Computing Research*, pp. 1–6, Dec. 2012.

[64] H. Wang, H. Fang, L. Xing, and M. Chen, "An integrated biometric-based security framework using wavelet-domain HMM in wireless body area networks (WBAN)," in *IEEE International Conference on Communications (ICC'11)*, pp. 1–5, June 2011.

[65] Q. Q. Xie, S. R. Jiang, L. M. Wang, and C. C. Chang, "Composable secure roaming authentication protocol for cloud-assisted body sensor networks," *International Journal of Network Security*, vol. 18, pp. 816–831, Sep. 2016.

[66] F. Xu, Z. Qin, C. C. Tan, B. Wang, and Q. Li, "IMDGuard: Securing implantable medical devices with the external wearable guardian," in *IEEE INFOCOM*, pp. 1862–1870, Apr. 2011.

[67] G. H. Zhang, C. C. Y. Poon, and Y. T. Zhang, "A fast key generation method based on dynamic biometrics to secure wireless body sensor networks for p-health," in *Annual International Conference of the IEEE Engineering in Medicine and Biology*, pp. 2034–2036, Aug. 2010.

[68] M. Zhang, A. Raghunathan, and N. K. Jha, "MedMon: Securing medical devices through wireless monitoring and anomaly detection," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 7, pp. 871–881, Dec. 2013.

[69] Z. Zhang, H. Wang, A. V. Vasilakos, and H. Fang, "Ecg-cryptography and authentication in body area networks," *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, pp. 1070–1078, 2012.

[70] Z. Zhang, H. Wang, A. V. Vasilakos, and H. Fang, "Channel information based cryptography and authentication in wireless body area networks," in *8th International Conference on Body Area Networks (BodyNets'13)*,, pp. 132–135, 2013.

Biography

Laura Victoria Morales is a doctoral student at Universidad de los Andes. Bogotá, Colombia. She holds an M.S. degree on Information Security from Universidad de los Andes. Her research interests include security in IoT devices, Body Area Networks and computer forensics.

David Delgado-Ruiz is a Computing and Systems Engineering student at Universidad de los Andes. Bogotá, Colombia. His research interests include IoT security.

Sandra Julieta Rueda is an Assistant professor at Universidad de los Andes. Bogotá, Colombia. She holds an M.S. degree from Universidad de los Andes and a Ph.D. from The Pennsylvania State University, USA. Her research interests include systems security, access control and policy analysis.

Comments on Privacy-Preserving Yoking Proof with Key Exchange in the Three-Party Setting

Qingfeng Cheng and Xinglong Zhang
(Corresponding author: Qingfeng Cheng)

State Key Laboratory of Mathematical Engineering and Advanced Computing
Zhengzhou, Henan Province 450001, China
(Email: qingfengc2008@sina.com)

(Received Oct. 1, 2017; revised and accepted Mar. 27, 2018)

Abstract

In 2017, Tian, Yang and Mu presented a new three-party key exchange protocol YPKE in radio frequency identification environment, which is based on the HMQV protocol. They claimed that the proposed YPKE protocol in the three-party setting meets user privacy and session key security. In this comment, we point out that the YPKE protocol still has some weaknesses. Our results show that the proposed YPKE protocol cannot provide perfect forward secrecy, and also cannot resist impersonation attack. At the same time, the YPKE protocol is lack of the security of ephemeral private key leakage and unknown key-share, which the original HMQV protocol can achieve.

Keywords: Cryptanalysis; Ephemeral Private Key Leakage Attack; HMQV Protocol; Key Exchange; Key Compromise Impersonation Attack; Perfect Forward Secrecy

1 Introduction

With the rise in technology, radio frequency identification (RFID) protocols [1–3, 10, 13, 14] have become essential components in the Internet of Things (IoT) environment. Usually, in a RFID protocol, the session key to encrypt communication messages among the reader(or server) and the tags (or users) is needed. Key exchange (KE), which can generate the session key, is a fundamental building block in open network. There are many famous KE protocols in the literature, such as MQV protocol [4, 5, 9], HMQV protocol [7] and NAXOS protocol [8].

Recently, Tian, Yang and Mu [12] presented a novel key exchange protocol, called YPKE protocol. The YPKE protocol using yoking proof [6] could generate a common session key among the reader(or server) and the tags (or users). The design of the YPKE protocol was based on the HMQV protocol and Schnorr signature [11]. However, in contrast to the original HMQV protocol, the YPKE protocol needs three round and involves three parties, i.e. a server and two users. In this comment, we will point out that the YPKE protocol exists some weaknesses. We show that the YPKE protocol is lack of perfect forward secrecy, and cannot resist insider impersonation attack, unknown key-share attack and ephemeral private key leakage attack, which the original HMQV protocol can resist.

The remainder of this comment will firstly introduce the original YPKE protocol in Section 2. Then, Section 3 points out the weaknesses of the YPKE protocol. Conclusion will be given in Section 4.

2 Review of the YPKE Protocol

Here, we briefly review the YPKE protocol proposed by Tian *et al.* in 2017. For more details, refer to [12].

Table 1: The notations

Notations	Description
$\mathcal{S}$	the reader/server
TX	a tag/user
τ	security parameter
$\mathcal{G}$	a cyclic additive group of order q, where $\|q\| = \tau$ is a big prime, g is a generator of this group
$SPK_{\mathcal{S}}/SSK_{\mathcal{S}}$	the server's public/secret key, where $SPK_{\mathcal{S}} = (SPK_{\mathcal{S}1}, SPK_{\mathcal{S}2})$ $SSK_{\mathcal{S}} = (SSK_{\mathcal{S}1}, SSK_{\mathcal{S}2})$
EPK_{TX}/ESK_{TX}	TX's ephemeral public/secret key
PK_{TX}/SK_{TX}	TX's public/secret key, where $PK_{TX} = (PK_{TX1}, PK_{TX2})$ $SK_{TX} = (SK_{TX1}, SK_{TX2})$
H_1	a hash function used in the HMQV from $\{0,1\}^*$ to $\{0,1\}^l$
H_1'	a hash function from $\{0,1\}^*$ to $\{0,1\}^\tau$
H_2	a hash function from $\{0,1\}^*$ to Z_q
H_3	a hash function from $\mathcal{G}$ to $\{0,1\}^\tau$
ENC	encryption function

2.1 The Description of YPKE Protocol

In this subsection, we describe the YPKE protocol shown in Figure 1, which needs five step.

1) Server $\mathcal{S}$ sends the key $SPK_{\mathcal{S}}$ to user TA and TB. This step is the same with the original YPKE protocol.

2) Upon receiving the key $SPK_{\mathcal{S}}$, TA and TB respectively send the message (EPK_{TA}, C_{TA}) and (EPK_{TB}, C_{TB}), where $C_{TA} = ENC_{SPK_{\mathcal{S}2}}(PK_{TA})$, $C_{TB} = ENC_{SPK_{\mathcal{S}2}}(PK_{TB})$, $EPK_{TA} = g^{ESK_{TA}}$ and $EPK_{TB} = g^{ESK_{TB}}$.

3) Upon receiving (EPK_{TA}, C_{TA}) and (EPK_{TB}, C_{TB}), $\mathcal{S}$ uses the key $SSK_{\mathcal{S}2}$ to obtain PK_{TA} and PK_{TB}. Then $\mathcal{S}$ sends (c, EPK_{TB}, C'_{TA}) to user TA and (c, EPK_{TA}, C'_{TB}) to user TB, where $C'_{TA} = ENC_{PK_{TA2}}(PK_{TB1})$ and $C'_{TB} = ENC_{PK_{TB2}}(PK_{TA1})$.

4) User TA decrypts $ENC_{PK_{TA2}}(PK_{TB1})$ to obtain PK_{TB1} and uses the HMQV method to compute K_{TATB}. Then TA computes $Y = g^y$, where $y = H_2(K_{TATB}||c)$, and computes $T_{TA} = g^{t_{TA}}$, where $t_{TA} \in_R Z_q$. Further, user TA computes the signature $Sig_{TA} = t_{TA} + e_{TA}SK_{TA1}$, where $e_{TA} = H_2(T_{TA}||EPK_{TA}||PK_{TB1}||C_{TA}||SPK_{\mathcal{S}1}^{ESK_{TA}}||c||Y)$. Finally, TA computes the session key $FSK = H_3(SPK_{\mathcal{S}1}^{y})$ and sends (Sig_{TA}, T_{TA}, Y) to server $\mathcal{S}$. Similarly, user TB also computes the session key $FSK = H_3(SPK_{\mathcal{S}1}^{y})$ and sends (Sig_{TB}, T_{TB}, Y) to server $\mathcal{S}$.

5) Server $\mathcal{S}$ verifies $g^{Sig_{TA}} = T_{TA} \cdot PK_{TA1}^{e_{TA}}$ and $g^{Sig_{TB}} = T_{TB} \cdot PK_{TB1}^{e_{TB}}$. If two equations are right at the same time. Then $\mathcal{S}$ computes the session key $FSK = H_3(Y^{SSK_{\mathcal{S}1}})$. Otherwise, $\mathcal{S}$ aborts the session.

3 Analysis of the YPKE Protocol

In this section, we firstly review some of the security attributes of the KE protocols, and then provide our analysis.

Perfect Forward Secrecy: A user's private key leakage does not compromise the security of session keys generated by this user before the leakage happened.

Insider Impersonation Attack: A user or the server, which involves in the protocol, is malicious, and impersonates another user (or server) to cheat the legal server (or user).

Unknown Key-Share Attack: The adversary M, can corrupt any user, mount the attack between two honest users A and B. At the end of a session, user A convinces that he has shared the session key with user B. However, user B thinks that she has shared the session key with corrupted user C.

Ephemeral Private Key Leakage Attack: The adversary learns the ephemeral private key, and uses it to compute the session key.

3.1 The Lack of Perfect Forward Secrecy

Tian *et al.* claimed that the adversary could not make corrupt queries to the server in their model. However, we think that it is not a reasonable assumption. In the YPKE protocol, there are three parties, a server and two users, whose private key and public key are independent. If the adversary can make queries to two users, he should also make queries to the server.

Since the common session key is $FSK = H_3(Y^{SSK_{S1}})$, the adversary learning the server's private key $SSK_{\mathcal{S}} = (SSK_{\mathcal{S}1}, SSK_{\mathcal{S}2})$ can use the public message Y to achieve $FSK = H_3(Y^{SSK_{S1}})$ easily. It means that the YPKE protocol cannot achieve the property of perfect forward secrecy.

3.2 The Description of Insider Impersonation Attack

In the original YPKE protocol, the server does not verify the identity of two users in the first round communications, so a malicious user can cheat the server successfully. Here, we assume that the user TB is a malicious user. He first fabricates a user TA^* with public key PK_{TA^*} and private key SK_{TA^*}.

1) Server $\mathcal{S}$ sends the key $SPK_{\mathcal{S}}$ to the user TA and the user TB. However, the user TB intercepts the message for the user TA. It means that the user TA even does not know the existence of the session.

2) Upon receiving the key $SPK_{\mathcal{S}}$, TA^*, who is impersonated by TB, and TB respectively send the message (EPK_{TA^*}, C_{TA^*}) and (EPK_{TB}, C_{TB}), where $C_{TA^*} = ENC_{SPK_{\mathcal{S}2}}(PK_{TA^*})$, $C_{TB} = ENC_{SPK_{\mathcal{S}2}}(PK_{TB})$, $EPK_{TA^*} = g^{ESK_{TA^*}}$ and $EPK_{TB} = g^{ESK_{TB}}$.

3) Upon receiving (EPK_{TA^*}, C_{TA^*}) and (EPK_{TB}, C_{TB}), the server $\mathcal{S}$ uses the private key $SSK_{\mathcal{S}2}$ to obtain public key PK_{TA^*} and PK_{TB} respectively. Then the server $\mathcal{S}$ sends the message (c, EPK_{TB}, C'_{TA^*}) to the user TA^* and (c, EPK_{TA^*}, C'_{TB}) to the user TB, where $C'_{TA^*} = ENC_{PK_{TA2^*}}(PK_{TB1})$ and $C'_{TB} = ENC_{PK_{TB2}}(PK_{TA1^*})$.

4) Upon intercepting the message (c, EPK_{TB}, C'_{TA^*}) and receiving the message (c, EPK_{TA^*}, C'_{TB}), user TB randomly chooses a value $y \in_R Z_q$. Then user TB computes $Y = g^y$ and $T_{TA^*} = g^{t_{TA^*}}$, where $t_{TA^*} \in_R Z_q$. Further, user TB computes the signature $Sig_{TA^*} = t_{TA^*} + e_{TA^*}SK_{TA1^*}$, where $e_{TA^*} = H_2(T_{TA^*}||EPK_{TA^*}||PK_{TB1}||C_{TA^*}||SPK_{\mathcal{S}1}^{ESK_{TA^*}}||c||Y)$. Finally, user TB computes the final session

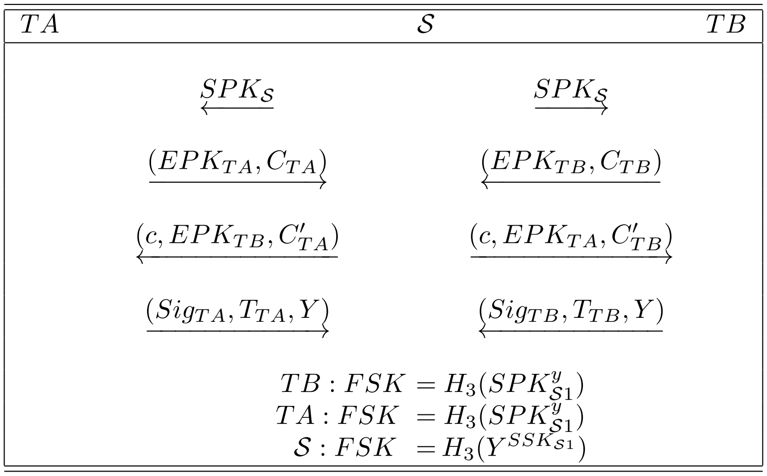

TA
$\mathcal{S}$
TB
$\xleftarrow{SPK_{\mathcal{S}}}$
$\xleftarrow{SPK_{\mathcal{S}}}$
$\xrightarrow{(EPK_{TA}, C_{TA})}$
$\xleftarrow{(EPK_{TB}, C_{TB})}$
$\xleftarrow{(c, EPK_{TB}, C'_{TA})}$
$\xrightarrow{(c, EPK_{TA}, C'_{TB})}$
$\xrightarrow{(Sig_{TA}, T_{TA}, Y)}$
$\xleftarrow{(Sig_{TB}, T_{TB}, Y)}$
$TB : FSK = H_3(SPK^y_{\mathcal{S}1})$
$TA : FSK = H_3(SPK^y_{\mathcal{S}1})$
$\mathcal{S} : FSK = H_3(Y^{SSK_{\mathcal{S}1}})$

Figure 1: The YPKE protocol

key $FSK = H_3(SPK^y_{\mathcal{S}1})$ and impersonates the user TA to send $(Sig_{TA^*}, T_{TA^*}, Y)$ to server $\mathcal{S}$. Similarly, user TB also sends (Sig_{TB}, T_{TB}, Y) to the server $\mathcal{S}$.

5) Server $\mathcal{S}$ verifies $g^{Sig_{TA^*}} = T_{TA^*} \cdot PK^{e_{TA^*}}_{TA1^*}$ and $g^{Sig_{TB}} = T_{TB} \cdot PK^{e_{TB}}_{TB1}$. If two equations are right at the same time. Then $\mathcal{S}$ computes the session key $FSK = H_3(Y^{SSK_{\mathcal{S}1}})$. Otherwise, $\mathcal{S}$ aborts the session.

Now, the session is finished. The server will think that he has shared the common session key with user TA and user TB. However, the user TA does not know the existence of the session completely. So the malicious user TB has successfully cheated the server in the session.

3.3 The Description of Unknown Key-Share Attack

In the original YPKE protocol, the user does not verify the identity of the server, so a malicious user can cheat the other user successfully. Here, we assume that the user TB is a malicious user. He can cheat the user TA, who thinks that she has shared a common session key with the server and the user TB. However, in fact, the server even does not know the existence of the session.

1) The user TB learns the server $\mathcal{S}$'s public key $SPK_{\mathcal{S}}$ and user TA's public key PK_{TA} from other sessions. Then he can impersonate the server to send $\mathcal{S}$'s public key $SPK_{\mathcal{S}}$ to user TA.

2) Upon receiving the key $SPK_{\mathcal{S}}$, TA sends the message (EPK_{TA}, C_{TA}) to the server $\mathcal{S}$, where $C_{TA} = ENC_{SPK_{\mathcal{S}2}}(PK_{TA})$ and $EPK_{TA} = g^{ESK_{TA}}$.

3) The user TB intercepts the message (EPK_{TA}, C_{TA}). Then he impersonates the server $\mathcal{S}$ and sends (c, EPK_{TB}, C'_{TA}) to the user TA, where $C'_{TA} = ENC_{PK_{TA2}}(PK_{TB1})$.

4) User TA decrypts $ENC_{PK_{TA2}}(PK_{TB1})$ to obtain PK_{TB1} and uses the HMQV method to compute K_{TATB}. Then user TA computes $Y = g^y$, where $y = H_2(K_{TATB}||c)$, and computes $T_{TA} = g^{t_{TA}}$, where $t_{TA} \in_R Z_q$. Further, user TA computes the signature $Sig_{TA} = t_{TA} + e_{TA}SK_{TA1}$, where $e_{TA} = H_2(T_{TA}||EPK_{TA}||PK_{TB1}||C_{TA}||SPK^{ESK_{TA}}_{\mathcal{S}1}||c||Y)$. Finally, TA computes the session key $FSK = H_3(SPK^y_{\mathcal{S}1})$ and sends (Sig_{TA}, T_{TA}, Y) to the server $\mathcal{S}$.

5) Upon intercepting the message (Sig_{TA}, T_{TA}, Y), the user TB can compute the session key $FSK = H_3(SPK^y_{\mathcal{S}1})$ and finish the session.

When the session is finished, the user TA will think that he has shared the session key with the server $\mathcal{S}$ and the user TB. In contrast, the server $\mathcal{S}$ does not know the existence of the session. So the malicious user TB has successfully cheated the user TA in the session. It will be dangerous in some situations of IoT environment. The main reason is the lack of authentication, when the user TA communicates with the server $\mathcal{S}$ in the YPKE protocol.

3.4 The Description of Ephemeral Private Key Leakage Attack

Tian *et al.*'s original YPKE protocol was based on the HMQV protocol. However, the HMQV protocol with implicit authentication can resist ephemeral private key leakage attack. However, the adversary, who learns the value of ESK_{TA} and t_{TA} in the YPKE protocol, can use Sig_{TA} to compute the TA's private key SK_{TA1}. It is contradict to the HMQV method. Similarly, if the adversary learns the value of ESK_{TB} and t_{TB}, he also can compute TB's private key SK_{TB1}.

4 Conclusion

In this comment, we analyze the security of the YPKE protocol, and point out that the YPKE protocol still exist some weaknesses. It means that the YPKE protocol is lack of perfect forward secrecy, and cannot resist insider impersonation attack, unknown key-share attack and ephemeral private key leakage attack. In fact, the server and tags in the IoT environment have different compute capability. So it is not an easy task to design an excellent key exchange protocol in such an imbalanced network.

Acknowledgments

The authors would like to thank Prof. Min-Shiang Hwang and the anonymous referees for their helpful comments. This work was supported by the National Natural Science Foundation of China (No. 61872449).

References

[1] M. Chen, S. Chen, Y. Fang, “Lightweight anonymous authentication protocols for RFID systems,” *IEEE/ACM Transactions on Networking*, vol. 25, no. 3, pp. 1475-1488, 2017.

[2] Y. C. Chen, W. L. Wang, M. S. Hwang, “RFID authentication protocol for anti-counterfeiting and privacy protection”, in *The 9th International Conference on Advanced Communication Technology*, pp. 255–259, 2007.

[3] P. Cui, “An improved ownership transfer and mutual authentication for lightweight RFID protocols,” *International Journal of Network Security*, vol. 18, no. 6, pp. 1173-1179, 2016.

[4] L. C. Huang and M. S. Hwang, “An efficient MQV key agreement scheme,” *International Journal of Network Security*, vol. 16, no. 2, pp. 157–160, 2014.

[5] L. C. Huang, C. C. Lee, and M. S. Hwang, “A n2+n MQV key agreement protocol”, *International Arab Journal of Information Technology*, vol. 10, no. 2, pp. 137-142, 2013.

[6] A. Juels, “Yoking-proofs for RFID tags,” in *Proceedings of 2nd IEEE conference on pervasive computing and communications workshops*, pp. 138-143, 2004.

[7] H. Krawczyk, “HMQV: A high-performance secure diffie-hellman protocol,” in *25th Annual International Cryptology Conference*, pp. 546-566, 2005.

[8] B. LaMacchia, K. Lauter, A. Mityagin, “Stronger security of authenticated key exchange,” in *Proceedings of ProvSec, First International Conference on Provable Security*, pp.1-16, 2007.

[9] L. L. Menezes, A. Qu, M. J. Solinas, S. Vanstone, “An efficient protocol for authenticated key agreement,” *Designs, Codes and Cryptography*, vol. 28, no. 2, pp. 119-134, 2003.

[10] Q. Qian, Y. Jia, R. Zhang, “A lightweight RFID security protocol based on elliptic curve cryptography,” *International Journal of Network Security*, vol. 18, no. 2, pp. 354-361, 2016.

[11] C. Schnorr, “Efficient identification and signatures for smart cards,” in *9th Aannual International Cryptology Conference*, pp. 239-252, 1989.

[12] Y. Tian, G. Yang, Y. Mu, “Privacy-preserving Yoking proof with key exchange in the three-party setting,” *Wireless Personal Communications*, vol. 94, no. 3, pp. 1017-1034, 2017.

[13] C. H. Wei, M. S. Hwang, A. Y. H. Chin, “A mutual authentication protocol for RFID”, *IEEE IT Professional*, vol. 13, no. 2, pp. 20–24, Mar. 2011.

[14] C. H. Wei, M. S. Hwang, A. Y. H. Chin, “A secure privacy and authentication protocol for passive RFID tags,” *International Journal of Mobile Communications*, vol. 15, no. 3, pp. 266–277, 2017.

Biography

Qingfeng Cheng received his B.A. degree in 2000 and M.S. degree in 2004 from National University of Defense Technology, and Ph.D. degree in 2011 from Zhengzhou Information Science and Technology Institute. He is now an Associate Professor in the State Key Laboratory of Mathematical Engineering and Advanced Computing. His research interests include cryptography and information security.

Xinglong Zhang born in 1994, is a graduate student in the State Key Laboratory of Mathematical Engineering and Advanced Computing. His main research interests include network protocol and cyber 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.