A Secure and Efficient E-Medical Record System via Searchable Encryption in Public Platform

Lei Xu[1], Chungen Xu[1], and Xing Zhang[2]
[1] School of Science, Nanjing University of Science & Technology, Nanjing, China
[2] School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang, China
[E-mail: xuleinjust@yeah.net; xuchung@njust.edu.cn zhangxing8928@gmail.com]
*Corresponding author: Chungen Xu

Received February 12, 2016; revised April 10, 2017; accepted May 25, 2017; published September 30, 2017

Abstract

This paper mainly presents a secure and efficient e-Medical Record System via searchable encryption scheme from asymmetric pairings, which could provide privacy data search and encrypt function for patients and doctors in public platform. The core technique of this system is an extension public key encryption system with keyword search, which the server could test whether or not the files stored in platform contain the keyword without leaking the information about the encrypted file. Compared with former e-medical record systems, the system proposed here has several superior features: (1)Users could search the data stored in cloud server contains some keywords without leaking anything about the origin data. (2) We apply asymmetric pairings to achieve shorter key size scheme in the standard model, and adopt the dual system encryption technique to reduce the scheme's secure problem to the hard Symmetric External Diffie-Hellman assumption, which could against the variety of attacks in the future complex network environment. (3) In the last of paper, we analyze the scheme's efficiency and point out that our scheme is more efficient and secure than some other classical searchable encryption models.

Keywords: keyword search encryption, e-medical record, asymmetric pairings, dual system encryption

This research was supported by research grant from Natural Science Foundations of Jiangsu Province of China [Grant No. BK20141405].

https://doi.org/10.3837/tiis.2017.09.024 **ISSN : 1976-7277**

1. Introduction

Electronic medical record(EMR) is also called computerized medical record systems or patient records based computers. It uses electronic equipment(such as: computers, health cards, etc.) to save, manage, transfer and reproduce the digital medical records of the patient instead of the handwritten paper medical records. It includes all the information of diagnosis and treatment of patients in the hospital. The United States National Institute for medical research defined EMR as a specific system of patient records, which could provide user for the ability of access private data, alert, tip, and clinical decision. According to application of electronic medical record system to in hospital department, although the user can get the relative data from other system directly to complete the whole content of first page, which achieves the goal that data and resource share together, they also need to face the risk of privacy information leakage and loss. Especially in the environment of modern rapid cloud storage technique.

Well known that, cloud storage is a new information storage technology, users can transmit their personal files, photos and videos to cloud through PC client, mobile terminals, such as smart phones, tablet computers at any time. While using cloud technique may reduce the burden of local data management and system maintenance costs, the data stored in the cloud will be out from the physical control of users, that the cloud server administrators and illegal users could obtain the information by accessing the data without limitation. Many companies and individual users try to encrypt the data firstly and then store the ciphertext in the cloud server to protect their data. This method is simple, but brings a lot of problems. For example, in a hospital, if the doctor needs to find the relevant medical record of some attributes or keywords, he should download all the uploaded records, decrypt and then retrieve. This will give rise to two problems: 1) If the user has uploaded a large number of files, download them one by one may cause the server blockage; 2) Decrypting all files downloaded will also take up a lot of local computing resources and result in low efficiency.

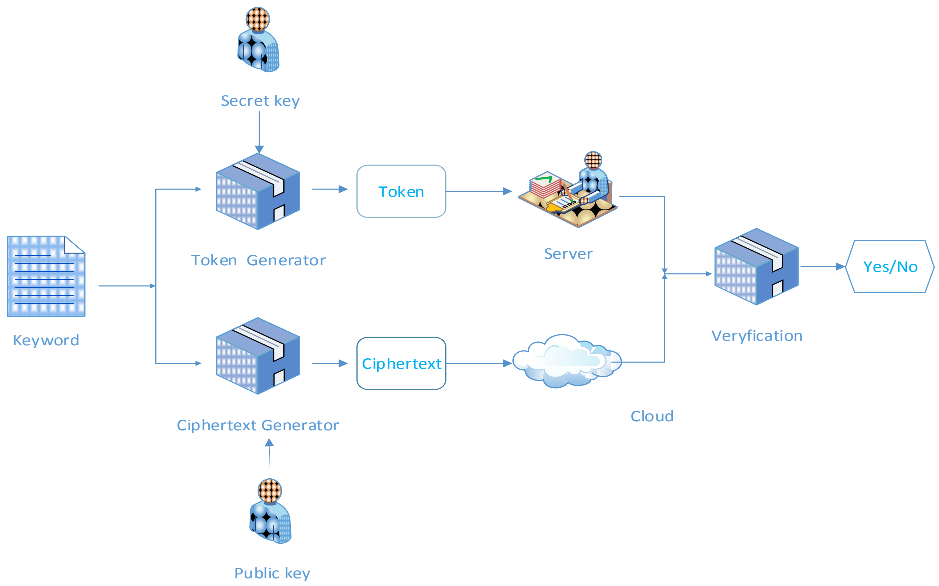


Fig. 1. The flow chart of confidential handling over original data

In order to solve this problem better, searchable encryption came into being, and has been extensively studied and developed in recent years. Searchable encryption initially originated from the private information extraction problem suggested by Chor et al.[1] in 1995. Through this mechanism (as in **Fig. 1**), users can encrypt data first, and then stores the ciphertext in the cloud server. When the user wants to search for the file with keyword $"w"$, he can send the token of the keyword to the cloud server. The cloud will receive the search capability and test for matching with each file, and if the match is successful, it means that the file contains that keyword. Through the above process, the user does not need to waste overhead network and storage space for file that does not contain the keyword; second, the keyword search could perform on the clouds, make full use of the powerful computing ability; finally, users do not have to perform decryption operation to meet the conditions, saving the local computing resources. So it provides us a very good ideal to solve the storage and search problem of privacy information in e-medical record system.

1.1 Related Work

In general, the searchable encryption can be seen as a set of cryptographic protocols with searchable ability. Public key encryption with keyword search(PEKS) was first proposed by Boneh[2] where a sender generates ciphertext associated with keyword under the public key in 2004. In their work, they gave the first concept and used anonymous identity-based encryption scheme to construct the first PEKS scheme which allows the gateway in communication to have the ability to test whether "urgent" is a keyword in the email without learn anything else about the email.

Independent of Boneh's work, Waters et al.[3] presented an approach for constructing searchable encrypted audit logs in the same year, which can be combined with any number of existing approaches for creating tamper resistant logs. In particular, they implemented an audit log for database queries that used hash chains for integrity and identity based encryption with extracted keywords to enable searching on the encrypted log. In addition, Golle and Waters[4] gave a searchable encryption with conjunctive keywords in another paper, i.e their scheme can search the files contained keywords $"w_1", \cdots, "w_n"$. This solves the single keyword problem that has appeared in Boneh's paper.

However, for previous PEKS schemes' adversary did not consider the relationship between the target token and the search results before, Curtmola, Garay, Kamara, Ostrovsky[5] described a stronger adaptive adversary model. In their model, an adversary would decide next query based on previous searching trapdoor and search results as references. They also designed two kinds of schemes, the first one could ensure the security under the none-adaptive case, and use the linked list, array and table data structure to connect the different pieces of the keywords, while in the second scheme, in order to reach the adaptive semantic security, they proposed a broadcast encryption, using the method of sharing, which enables users to make the sharing of the ciphertext data search. And later in 2012, Kurosawa and Ohtaki propose a verifiable searchable encryption scheme[6] that is secure against active adversaries and/or a malicious server. The scheme constructed on a MAC tag inside the index to bind a query to an answer, and was proved to be semantic security against active adversaries, which covers keyword privacy as well as reliability of the search results.

Functional Encryption(FE) is an exciting new paradigm that generalizes public key encryption by Boneh et.al [7]. In functional encryption, each decryption key corresponds to a specific function. When the holder of a decryption key for the function f gets an encryption of a message m, the only thing his key allows him to learn is $f(m)$, but nothing more. Public key encryption scheme with keyword search can be considered as a special type of FE, there are also many other kinds of function encryption corresponding to their different nature, such as: Attribute-based function encryption(ABE) and predicate-based function encryption. Attribute-based function searchable encryption was suggested as a searchable function encryption with unique property by Zheng Q et al. [8] in 2014, where ciphertexts must be accessed by a data owner's access control policy, and predicate encryption[9] was a generalized notion for public key encryption that enables one to encrypt attributes as well as a message. When we set the special function be the ability of searching, that FE scheme can helps us solve many practical problems.

Since the excellent properties of the several function encryptions, the research of FE design has become popular, T.F. Vallent et.al[10] proposed an efficient public key encryption with keyword search protocol which is pairing-free and is resilient against offline keyword guessing attack based on the Diffie-Hellman problem and the ElGamal encryption scheme in 2014. L. Xu et al. used asymmetric pairing to design the first dual form searchable encryption[11] in 2015. Furthermore, there also have been many other schemes[12], [13] with special function from security and practice. In this paper, we plan to use them to design a practical e-medical system.

Medical care is a major event related to people's livelihood, the design of the electronic medical records in the world starts relatively late, and the design of electronic medical records at home and abroad showed a trend of diversification. For example, some confirm the relationship between doctors and patients by using signature system, other ensure the safety of patient information can be stored in a public platform use encryption algorithm. In this paper, our main purpose is to introduce the research status and dynamic development in the world from the point of the development of searchable encryption and the functional development of electronic medical records, and then propose a novel scheme and e-medical record system different from previous ones by using the asymmetric pairings.

1.2 Organization

We organize the rest of the paper as follows. In Section 2, we describe the definition of the PEKS and provide its security model, and then give the related hard problems and complexity assumptions. Section 3, 4 provide a novel PEKS scheme from symmetric pairing and design a practical e-medical record system model based on the proposed scheme with an encryption scheme. Section 5 proves the scheme's security under a statical assumption and follow with a complexity analysis in Section 6. Finally, we end the paper with a brief conclusion.

2. Preliminaries

In this section, we first review the definition of the public key encryption with keyword search, and then present some hard problems with its complexity assumption on pairings related to our security proof.

2.1 Public Key Encryption with Keyword Search

Referring to the Boneh's work[2], we give the Extension-PEKS definition as follows:

Definition 1. A Extension Public Key Encryption with Keyword Search (e-PEKS) scheme[2] for client and server consists of four polynomial-time algorithms, proceeds as follows:

- **Setup**: Take as input a security parameter λ, generate public key pk and secret key sk for client and server respectively . Public the their public key pk, and keep the secret key sk_{client}, sk_{server} to themselves.
- **TokenGen**: Take as input the client's private key sk_{client} and a keyword “w”, generate a token T_w for the keyword “w”.
- **SEncrypt**: Take as input the public key pk and a keyword “w” and a designed server with index “I”, produce a searchable ciphertext of keyword “w”.
- **Verify**: Take as input the public key pk, server's secret key sk_{server}, a valid ciphertext as $S = \textbf{SEncrypt}(pk, w', I)$, and token $T_w = \textbf{TokenGen}(sk_{client}, w)$, output 1 if $w = w'$ and 0 otherwise.

Definition 2. Let λ be the security parameter and $\mathcal{A}$ be the adversary. The security game between $\mathcal{A}$ and the simulator $\mathcal{B}$ simulates as follows:

- **Setup**: The challenger runs the Setup(λ) algorithm to generate $(pk, sk_{client}, sk_{server})$.It gives pk to the attacker $\mathcal{A}$.
- **Phase 1**: The attacker $\mathcal{A}$ can adaptively ask the challenger for the token T_w for any keyword $w \in \mathbb{Z}_q^*$ of his choice.
- **Challenge**: At some point, the attacker $\mathcal{A}$ sends the challenger two words w_0, w_1 on which it wishes to be challenged. The only restriction is that none of w_0 nor w_1 has been queried for token in Phase 1. The challenger picks a random $b \in \{0,1\}$ and gives the attacker $C = PEKS(pk, w_b)$ as the challenge ciphertext.
- **Phase 2**: The attacker can continue to ask for trapdoors T_w for any keyword w of his choice as long as $w \neq w_0, w_1$.
- **Guess**: Eventually, the attacker $\mathcal{A}$ outputs $b' \in \{0,1\}$ and wins the game if $b = b'$. Such an adversary $\mathcal{A}$ is called an IND-CKA adversary. $\mathcal{A}$'s advantage in attacking the scheme is defined as the following function of the security parameter λ:

$$Adv_{\epsilon,A}(\lambda) = |Pr[b = b'] - 1/2|.$$

The probability is over the random bits used by the challenger and the adversary.

Definition 3. We say that a e-PEKS is semantically secure against an adaptive chosen keyword attack if for any polynomial time attacker $\mathcal{A}$ we have that $Adv_{\mathcal{A}}(s)$ is a negligible function.

2.2 Asymmetric Bilinear Pairings and Dual Pairing Vector Spaces

We use the following [14] to describe asymmetric bilinear maps and bilinear map groups:

Definition 4. Let $\mathbb{G}_1, \mathbb{G}_2$ and $\mathbb{G}_T$ be three cyclic multiplicative groups having the same large prime order q and g_1, g_2 are respective generators of $\mathbb{G}_1, \mathbb{G}_2$. A mapping $e: \mathbb{G}_1 \times \mathbb{G}_2 \to \mathbb{G}_T$ is called a cryptographic bilinear map if it satisfies the following properties.

- Bilinearity: $e(u^a, v^b) = e(u,v)^{ab}$ for all $u \in \mathbb{G}_1, v \in \mathbb{G}_2$ and $a, b \in \mathbb{Z}_q$.
- Non-degeneracy: If $\mathbb{G}_1 = \langle g_1 \rangle$ and $\mathbb{G}_2 = \langle g_2 \rangle$, then $\mathbb{G}_T = \langle e(g_1, g_2) \rangle$, namely, $e(g_1, g_2) \neq 1$.

- Computability: There exists an efficient algorithm to compute $e(u,v)$ for all $u \in \mathbb{G}_1, v \in \mathbb{G}_2$. In additional to refer to individual elements of $\mathbb{G}$, we will also consider "vectors" of group element. For $\mathbf{v} = (v_1,\ldots,v_n) \in \mathbb{Z}_q^n$ and $g \in \mathbb{G}$, we write $g^{\mathbf{v}}$ to denote a $n-$tuple of elements of $\mathbb{G}$:

$$g^{\mathbf{v}} := (g^{v_1},\ldots,g^{v_n})$$

we can also perform scalar multiplication and vector addition in the exponent. For any $a \in \mathbb{Z}_q$ and $\mathbf{v},\mathbf{w} \in \mathbb{Z}_q^n$, we have:

$$g^{a\mathbf{v}} := (g^{av_1},\ldots,g^{av_n}) \text{ and } g^{\mathbf{v}+\mathbf{w}} := (g^{v_1+w_1},\ldots,g^{v_n+w_n})$$

Definition 5. For a constant dimension n, we call two random bases $\mathbb{B} := (\mathbf{b_1},\ldots, \mathbf{b_n})$ and $\mathbb{B}^* := (\mathbf{b_1^*},\ldots, \mathbf{b_n^*})$ of $\mathbb{Z}_q^n$ dual orthonormal[15], when

$$\mathbf{b}_i \cdot \mathbf{b}_j^* = 0 (\text{mod } q)$$

where $i \neq j$, and

$$\mathbf{b}_i \cdot \mathbf{b}_i^* = \psi (\text{mod } q)$$

for all i, where ψ is a random element of $\mathbb{Z}_q$.

Then for generators $g_1 \in \mathbb{G}_1, g_2 \in \mathbb{G}_2$, we have

$$e(g_1^{\mathbf{b}_i}, g_2^{\mathbf{b}_j^*}) = 1$$

whenever $i \neq j$, here 1 denotes the unit element in $\mathbb{G}_T$.

Lewko[16] describe a standard algorithm to generate such bases as $\mathbf{Dual}(\cdot)$. We use the notation $(\mathbb{D},\mathbb{D}^*,\mathbb{B},\mathbb{B}^*) \leftarrow \mathbf{Dual}(\mathbb{Z}_q^4, \mathbb{Z}_q^4)$ in the rest of this work.

2.2 Symmetric External Diffie-Hellman Assumptions

Definition 6. [Decisional Diffie-Hellman Assumption in $\mathbb{G}_1$][15]:Give a group generator $\mathcal{G}$, we define the following distribution:

$$\mathbb{G} := (\mathbb{G}_1,\mathbb{G}_2,\mathbb{G}_T, g_1, g_2, e, q) \xleftarrow{R} \mathcal{G}(1^\lambda)$$
$$a,b,c \xleftarrow{R} \mathbb{Z}_q$$
$$D := (\mathbb{G}; g_1, g_2, g^a, g^b)$$

We assume that for any PPT algorithm,

$$Adv_{\mathcal{A}}^{DDH1}(\lambda) := | Pr[\mathcal{A}(D, g_1^{ab})] - Pr[\mathcal{A}(D, g_1^{ab+c})] |$$

is negligible in the security parameter λ .

Notice that the above assumption also applies to $\mathbb{G}_2$.

Definition 7. The Symmetric External Diffie-Hellman assumption(SXDH)[17] holds if DDH problems are intractable over both $\mathbb{G}_1$ and $\mathbb{G}_2$.

2.3 Subspace Assumptions via SXDH

Definition 8. (DS1: Decisional Subspace Assumption in $\mathbb{G}_1$)[16] Given a group generator $\mathcal{G}(\cdot)$, define:

$$\mathbb{G} := (\mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_T, g_1, g_2, e, q) \xleftarrow{R} \mathcal{G}(1^\lambda)$$
$$(\mathbb{B}, \mathbb{B}^*) \xleftarrow{R} \mathbb{Z}_q, \quad \tau_1, \tau_2, \mu_1, \mu_2 \xleftarrow{R} \mathbb{Z}_q$$
$$U_i := g_2^{\mu_1 \mathbf{b}_i^* + \mu_2 \mathbf{b}_{2i}^*}, \quad V_i := g_1^{\tau_1 \mathbf{b}_i}$$
$$Z_i := g_1^{\tau_1 \mathbf{b}_i + \tau_2 \mathbf{b}_{2i1}}, \quad 1 \le i \le k$$
$$D := \mathbb{G}, g_2^{\mathbf{b}_1^*}, g_2^{\mathbf{b}_2^*}, \ldots, g_2^{\mathbf{b}_k^*}, g_2^{\mathbf{b}_{2k+1}^*}, \ldots, g_2^{\mathbf{b}_n^*},$$
$$g_1^{\mathbf{b}_1}, \ldots g_1^{\mathbf{b}_n}, U_1, U_2, \ldots, U_k, \mu_2$$

where k, n are constant positive integers that satisfy $2k \le n$. We assume that for any PPT algorithm $\mathcal{A}$

$$Adv_{\mathcal{A}}^{DS1}(\lambda) := | Pr[\mathcal{A}(D, V_1, \ldots, V_k)] - Pr[\mathcal{A}(D, Z_1, \ldots, Z_k)] |$$

is negligible in the security parameters λ.

Due to the number of keywords is only one in this paper, we set $n = 4, k = 2$ .Moreover, We require the following lemma from[15][19] in our security proof.

Lemma 1[15]**.** Let $C := \{(\mathbf{x}, \mathbf{v}) \mid \mathbf{x} \cdot \mathbf{v} \ne 0, \mathbf{x}, \mathbf{v} \in \mathbb{Z}_q^n\}$. For all $(\mathbf{x}, \mathbf{v}) \in C, (r, w) \in C, \rho, \tau \leftarrow \mathbb{Z}_q$, and $A \xleftarrow{R} \mathbb{Z}_q^{n \times n}$,

$$Pr[x(\rho A^{-1}) = r \wedge v(\tau A^t) = w] = \frac{1}{\#C}$$

In other words, $\rho x A^{-1}$ and $\tau v A^t$ are uniformly and independently distributed when $x \cdot v \ne 0$.

Lemma 2[19]**.** If the DDH assumption holds in $\mathbb{G}_1$, then the Subspace assumption in $\mathbb{G}_1$ stated in Definition 6 also holds. More precisely, for any adversary $\mathcal{A}$ against the Subspace assumption in $\mathbb{G}_1$, there exist probabilistic algorithms $\mathcal{B}$ whose running time are essentially the same as that of $\mathcal{A}$, such that

$$Adv_{\mathcal{A}}^{DS1}(\lambda) \le Adv_{\mathcal{B}}^{DDH1}(\lambda)$$

3 Our Construction

3.1 Basic searchable encryption system

We now construct our SE scheme via asymmetric pairings from extending the shorter IBE scheme suggested by Chen J. et al[15].

Let $\mathbb{G}_1, \mathbb{G}_2$ and $\mathbb{G}_T$ be groups of some large prime order q, and $e: \mathbb{G}_1 \times \mathbb{G}_2 \to \mathbb{G}_T$ be an admissible bilinear map. Our construction works as follows:

Setup(1^λ). The algorithm takes in the security parameter λ and generates a bilinear pairing $\mathbb{G} := (\mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_T, g_1, g_2, e, q)$ where q is a large prime. Then the algorithm samples random dual orthonormal bases $(\mathbb{D}, \mathbb{D}^*, \mathbb{B}, \mathbb{B}^*) \leftarrow$ **Dual**$(\mathbb{Z}_q^4, \mathbb{Z}_q^4)$. Let $\mathbf{d}_1, \ldots, \mathbf{d}_4$ denote the elements of $\mathbb{D}$, $\mathbf{d}_1^*, \ldots, \mathbf{d}_4^*$ denote the elements of $\mathbb{D}^*$, $\mathbf{b}_1, \ldots, \mathbf{b}_4$ denote the elements of $\mathbb{B}$ and $\mathbf{b}_1^*, \ldots, \mathbf{b}_4^*$ denote the elements of $\mathbb{B}^*$. The algorithm also picks $\alpha, \beta \in \mathbb{Z}_q$ randomly, and computes $e(g_1, g_2)^{\alpha \mathbf{d}_1 \mathbf{d}_1^*}$ and $e(g_1, g_2)^{\beta \mathbf{b}_1 \mathbf{b}_1^*}$. Finally, makes the parameters

$$params := \{\mathbb{G}, e(g_1, g_2)^{\alpha \mathbf{d}_1 \mathbf{d}_1^*}, e(g_1, g_2)^{\beta \mathbf{b}_1 \mathbf{b}_1^*}, g_1^{\mathbf{d}_1}, g_1^{\mathbf{d}_2}, g_1^{\mathbf{b}_1}, g_1^{\mathbf{b}_2}\}$$

known to the public and sends the keys

$$sk_{client} := \{\alpha, g_2^{\mathbf{d}_1^*}, g_2^{\mathbf{d}_2^*}\}, sk_{server} := \{\beta, g_2^{\mathbf{b}_1^*}, g_2^{\mathbf{b}_2^*}\}$$

to the client and server as their own secret key separately in a secure channel way.

SEncrypt$(params, w, I)$. Choose $s \in \mathbb{Z}_q$ randomly, the searchable ciphertext of keyword "w" for the designed server "I" is constructed as:

$$S = [g_1^{s\mathbf{d}_1 + sw\mathbf{d}_2 + s\mathbf{b}_1 + sI\mathbf{b}_2}, e(g_1, g_2)^{s(\alpha \mathbf{d}_1\mathbf{d}_1^* + \beta \mathbf{b}_1\mathbf{b}_2^*)}] = [S_1, S_2]$$

TokenGen(sk_{client}, w). Take as input the master key sk_{client} and keyword "w", output the token of the keyword "w":

$$T_w = g_2^{(\alpha + rw)\mathbf{d}_1^* - r\mathbf{d}_2^*}$$

Verify$(params, T_w, S, sk_{server})$. After receive the token of keyword "w", the designed server tests if $e(S_1, g_2^{(\beta+I)\mathbf{b}_1^* - \mathbf{b}_2^*} T_w) = S_2$ with his secret key. If so, output 1; if not, output 0. The scheme's correctness is easy to test.

$$\begin{aligned} e(S_1, \mathrm{g}^{(\beta+I)\mathbf{b}_1^* - \mathbf{b}_2^*} T_w) &= e\left(g_1^{s\mathbf{d}_1 + sw\mathbf{d}_2 + s\mathbf{b}_1 + sI\mathbf{b}_2}, g_2^{(\alpha+rw)\mathbf{d}_1^* - r\mathbf{d}_2^* + (\beta+I)\mathbf{b}_1^* - \mathbf{b}_2^*}\right) \\ &= e(g_1, g_2)^{(s\mathbf{d}_1 + sw\mathbf{d}_2 + s\mathbf{b}_1 + sI\mathbf{b}_2)[(\alpha+rw)\mathbf{d}_1^* - r\mathbf{d}_2^* + (\beta+I)\mathbf{b}_1^* - \mathbf{b}_2^*]} \\ &= e(g_1, g_2)^{s(\alpha+rw)\mathbf{d}_1\mathbf{d}_1^* - srw\mathbf{d}_2\mathbf{d}_2^* + s(\beta+I)\mathbf{b}_1\mathbf{b}_1^* - sI\mathbf{b}_2\mathbf{b}_2^{**}} \\ &= e(g_1, g_2)^{s(\alpha \mathbf{d}_1\mathbf{d}_1^* + \beta \mathbf{b}_1\mathbf{b}_1^*)} = S_2 \end{aligned}$$

3.2 Semi-Function Algorithm

We use the concepts of semi-functional PEKS and semi-functional trapdoors in our proof and provide algorithms that generate them. We notice that these algorithms are only provided for definitional purposes, and are not part of the e-PEKS system.

TokenGen. Algorithm $\mathcal{B}$ picks random values $r, x_3, x_4 \in \mathbb{Z}_q$ and forms a semi-functional token as:

$$T_{W'} = g_2^{(\alpha+rw)\mathbf{d}_1^* - r\mathbf{d}_2^* + x_3\mathbf{d}_3^* + x_4\mathbf{d}_4^*}$$

PEKS. The algorithm picks random values $s, y_3, y_4, z_3, z_4 \in \mathbb{Z}_q$ and forms a semi-functional PEKS as:

$$S' = [g_1^{s\mathbf{d}_1 - sw\mathbf{d}_2 + s\mathbf{b}_1 - sI\mathbf{b}_2 + y_3\mathbf{d}_3 + y_4\mathbf{d}_4 + z_3\mathbf{b}_3 + z_4\mathbf{b}_4}, e(g_1, g_2)^{s(\alpha \mathbf{d}_1\mathbf{d}_1^* + \beta \mathbf{b}_1\mathbf{b}_1^*)}]$$

We observe that if one applies the verification procedure with a semi-functional token and a normal ciphertext, verification will succeed because $\mathbf{d}_3$, $\mathbf{d}_4$ are orthogonal to all of the vectors in exponent of S, and hence have no effect on verification. Similarly, verification of a semi-functional PEKS by a normal trapdoor will also succeed because $\mathbf{d}_3$, $\mathbf{d}_4$ are orthogonal to all of the vectors in the exponent of T_w. When both the PEKS and token are semi-functional, the result of $e(S_1, \mathrm{g}^{(\beta+I)\mathbf{b}_1^* - \mathbf{b}_2^*} T_w) = S_2$ will have an additional term, namely

$$e(g_1, g_2)^{x_3 y_3 \mathbf{d}_3 \mathbf{d}_3^* + x_4 y_4 \mathbf{b}_4 \mathbf{b}_4^*} = e(g_1, g_2)^{(x_3 y_3 + x_4 y_4)\psi}$$

Verification will fail unless $x_3 y_3 + x_4 y_4 \equiv 0 \ (mod\ q)$. If this modular equation holds, we say that the trapdoor and PEKS pair is nominally semi-functional.

4. The proposed secure e-Medical Record System

This section we mainly construct a secure e-Medical Record system in public platform by using a secure encryption system and above basic searchable scheme.

Init: Take as input the security parameter λ, and output a bilinear pairing tuple $\mathbb{G} := (\mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_T, g_1, g_2, e, q)$ where $\mathbb{G}_1, \mathbb{G}_2$ and $\mathbb{G}_T$ be groups of some large prime order q, and $e: \mathbb{G}_1 \times \mathbb{G}_2 \to \mathbb{G}_T$ be an admissible bilinear map. In addition, there also needs a secure encryption algorithm $\mathbf{Enc}(\cdot)$ and decryption algorithm $\mathbf{Dec}(\cdot)$ to ensure the medical record's confidentiality with a key k.

Registration: When one wants to use this system, he needs to apply to become a legitimate user. So he submits his application to KGC. KGC runs the algorithm $\mathbf{Dual}(\cdot)$ firstly to obtain two random dual orthonormal bases, $(\mathbb{D}, \mathbb{D}^*, \mathbb{B}, \mathbb{B}^*) \leftarrow \mathbf{Dual}(\mathbb{Z}_q^4, \ \mathbb{Z}_q^4)$. Let $\mathbf{d}_1, \ldots, \mathbf{d}_4$ denote the elements of $\mathbb{D}$, $\mathbf{d}_1^*, \ldots, \mathbf{d}_4^*$ denote the elements of $\mathbb{D}^*$, $\mathbf{b}_1, \ldots, \mathbf{b}_4$ denote the elements of $\mathbb{B}$ and $\mathbf{b}_1^*, \ldots, \mathbf{b}_4^*$ denote the elements of $\mathbb{B}^*$. Then selects α, β from a uniform distribution on Z_q, computes the secret key $sk_{client} = \{\alpha, g_2^{\mathbf{d}_1^*}, g_2^{\mathbf{d}_2^*}\}$ and $sk_{server} = \{\beta, g_2^{\mathbf{b}_1^*}, g_2^{\mathbf{b}_2^*}\}$ for the patient and doctor respectively. Finally, KGC sends the secret key for searchable encryption with an encryption key k to the user, and output the system parameters

$$params := \{\mathbb{G}, e(g_1, g_2)^{\alpha \mathbf{d}_1 \mathbf{d}_1^*}, e(g_1, g_2)^{\beta \mathbf{b}_1 \mathbf{b}_1^*}, g_1^{\mathbf{d}_1}, g_1^{\mathbf{d}_2}, g_1^{\mathbf{b}_1}, g_1^{\mathbf{b}_2}\} .$$

Storage: The function of data storage will be described as **Fig. 2**. When the patient needs to upload his private data into cloud storage, he should do as follows:

1. Set the data he wants upload be the form of $P = (Q \mid M)$ where M is patient's medical record message and Q is some special strings of the record, such as user's name, ID number or e-mail address.

2. Compute $\mathbf{Enc}(k, M) \to C$ and $\mathbf{SEncrypt}(pk, H(\mathrm{Q}), I) \to SC$ respectively to ensure the confidentiality of the data and the special string of the record, here I denotes the doctor's identity message.

3. Run $\mathbf{TokenGen}(sk_{patient}, H(Q)) \to T$ to generate a special token for the special string of the record.

4. Finally, upload the encrypted data C and the searchable ciphertext SC as the form of $(SC \mid C)$ into cloud, and keep the token T by himself.

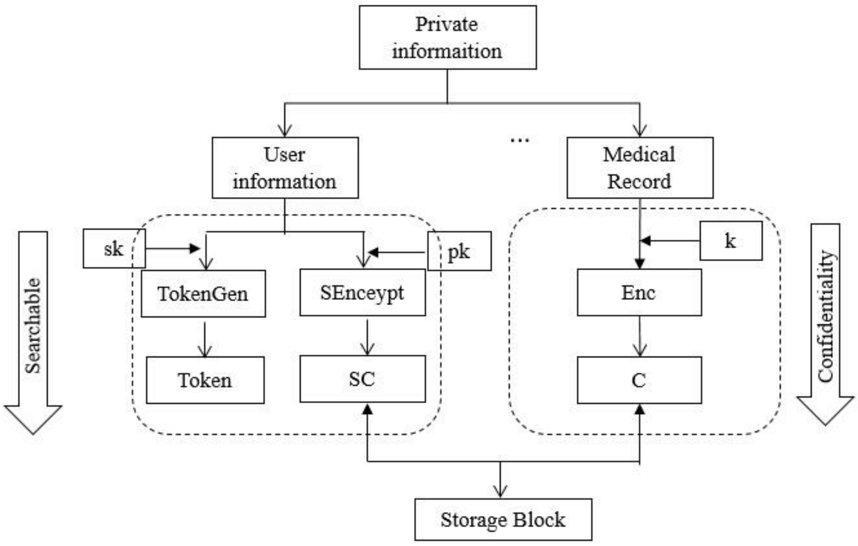


Fig. 2. The flow chart of confidential handling over original data

Retrieval: This processing mainly helps the doctor or user to search and get back the encrypted medical record, which he stored in the public platform of the hospital as **Fig. 3**. The detailed procedure is as follows:

1. This just needs the user to send his token T of the keyword to the cloud server.
2. In response, the server search the searchable ciphertext of the keywords in the head of each data and return the right data with $\mathbf{Verify}\left(params,T,SC,sk_{server}\right)=1$ to the user.
3. Finally, user recovers the initial data with algorithm $\mathbf{Dec}(C,k)$ by the key k .

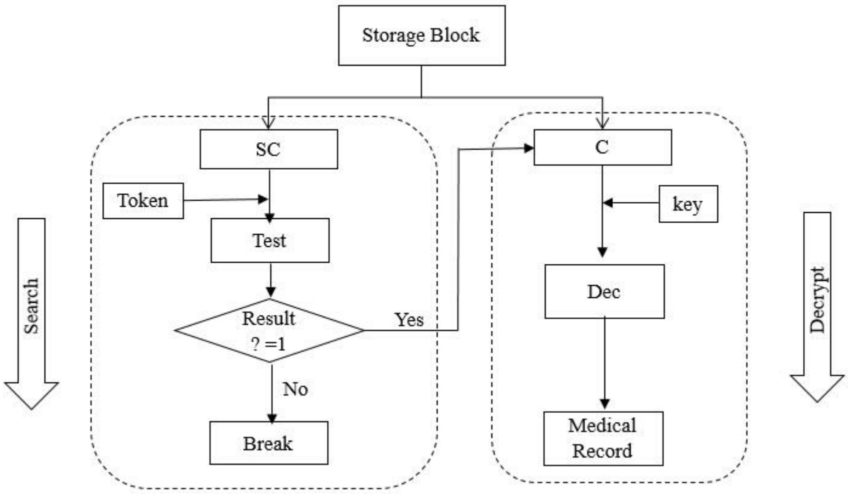


Fig. 3. The flow chart of retrieval handling over encrypted data

4. Security Analysis

Theorem 1. The non-interactive searchable encryption scheme above is semantically secure against a chosen keyword attack in the standard model under the Symmetric External Diffie-Hellman assumption. More precisely, for any adversary $\mathcal{A}$ against the extension PEKS scheme, there exist some probabilistic algorithms $\mathcal{B}_0,\mathcal{B}_1,\ldots,\mathcal{B}_\kappa$ whose running times are essentially the same as that of $\mathcal{A}$, such that

$$Adv_{\mathcal{A}}^{ePEKS}(\lambda) \le Adv_{\mathcal{B}_0}^{DS1}(\lambda) + \sum_{\kappa=1}^{t} Adv_{\mathcal{B}_\kappa}^{DS2}(\lambda) + \frac{t}{q}$$

where t is the maximum number of $\mathcal{A}$'s token queries.

We adopt the dual system encryption methodology by Waters[18] to prove the security of our extension-PEKS scheme.

For a probabilistic polynomial-time adversary $\mathcal{A}$, which makes at most t token queries, we organize the security proof of the scheme by the following sequence of games between $\mathcal{A}$ and a challenger $\mathcal{B}$.

- $\mathbf{Game}_R$: is the real security game.
- $\mathbf{Game_0}$: is the same as $\mathbf{Game_{Real}}$ except that the challenge ciphertext is semi-functional.
- $\mathbf{Game}_\kappa$: for κ from 1 to t, $\mathbf{Game}_\kappa$ is the same as $\mathbf{Game_0}$ except that the first κ tokens are semi-functional and the remaining tokens are normal.
- $\mathbf{Game}_F$: is the same as $\mathbf{Game}_\kappa$, except that the challenge PEKS is a semi-functional ciphertext of a random message in $\mathbb{Z}_q^*$. Denote S_w^F as the challenge ciphertext in $\mathbf{Game}_F$.

Lemma 3. Suppose that there exists an probability polynomial time adversary $\mathcal{A}$ where

$$| Adv_{\mathcal{A}}^{Game_{Real}}(\lambda) - Adv_{\mathcal{A}}^{Game_0}(\lambda) | = \epsilon$$

Then there exists an corresponding algorithm $\mathcal{B}_0$ such that $Adv_{\mathcal{B}_0(\lambda)}^{DS1} = \epsilon$, with $k = 4$ and $n = 8$.

Proof. Assume that $\mathcal{B}_0$ is given $D := (\mathbb{G}; g_2^{\mathbf{f}_1^*}, g_2^{\mathbf{f}_2^*}, g_2^{\mathbf{e}_1^*}, g_2^{\mathbf{e}_2^*}, g_1^{\mathbf{f}_1}, \ldots, g_1^{\mathbf{f}_4}, g_1^{\mathbf{e}_1}, \ldots, g_1^{\mathbf{e}_4}, U_1, U_2, \mu_2)$ along with C_1, C_2, C_3, C_4. We require that $\mathcal{B}_0$ decides whether C_1, C_2, C_3, C_4 are distributed as $g_1^{\tau_1\mathbf{f}_1}, g_1^{\tau_1\mathbf{f}_2}, g_1^{\tau_1\mathbf{e}_1}, g_1^{\tau_1\mathbf{e}_2}$ or $g_1^{\tau_1\mathbf{f}_1+\tau_2\mathbf{f}_3}$, $g_1^{\tau_1\mathbf{f}_2+\tau_2\mathbf{f}_4}, g_1^{\tau_1\mathbf{e}_1+\tau_2\mathbf{e}_3}$, $g_1^{\tau_1\mathbf{e}_2+\tau_2\mathbf{e}_4}$.

Setup. $\mathcal{B}_0$ simulates $\mathbf{Game}_{Real}$ or $\mathbf{Game}_0$ with adversary $\mathcal{A}$, depending on the distribution of C_1, C_2. To compute the public parameters and master secret key, $\mathcal{B}_0$ chooses two random invertible matrix $A, B \in \mathbb{Z}_q^{2\times 2}$ and set dual orthonormal bases $\mathbb{D}, \mathbb{D}^*, \mathbb{B}, \mathbb{B}^*$ to:

$$\mathbf{d}_1 := \mathbf{f}_1, \mathbf{d}_2 := \mathbf{f}_2, (\mathbf{d}_3,\ \mathbf{d}_4) := (\mathbf{f}_3,\ \mathbf{f}_4)A, \mathbf{d}_1^* := \mathbf{f}_1^*, \mathbf{d}_2^* := \mathbf{f}_2^*, (\mathbf{d}_3^*,\ \mathbf{d}_4^*) := (\mathbf{f}_3^*,\ \mathbf{f}_4^*)(A^{-1})'$$

$$\mathbf{b}_1 := \mathbf{e}_1, \mathbf{b}_2 := \mathbf{e}_2, (\mathbf{b}_3,\ \mathbf{b}_4) := (\mathrm{e}_3,\ \mathbf{e}_4)B, \mathbf{b}_1^* := \mathbf{e}_1^*, \mathbf{b}_2^* := \mathbf{e}_2^*, (\mathbf{b}_3^*,\ \mathbf{b}_4^*) := (\mathbf{e}_3^*,\ \mathbf{e}_4^*)(B^{-1})'$$

We note that $\mathbb{D}, \mathbb{D}^*, \mathbb{B}, \mathbb{B}^*$ are properly distributed, and will reveal nothing about $\mathcal{A}$. In addition, $\mathcal{B}$ cannot generate $g_2^{\mathbf{d}_3^*}, g_2^{\mathbf{d}_4^*}, g_2^{\mathbf{b}_3^*}, g_2^{\mathbf{b}_4^*}$, but these will not be needed for creating normal parameters. $\mathcal{B}_0$ chooses random value $\alpha, \beta \in \mathbb{Z}_q$, and computes $e(g_1, g_2)^{\alpha \mathbf{d}_1 \cdot \mathbf{d}_1^*}$, $e(g_1, g_2)^{\beta \mathbf{b}_1 \cdot \mathbf{b}_1^*}$. It then gives $\mathcal{A}$ the public parameters

$$params := \{\mathbb{G}, e(g_1, g_2)^{\alpha \mathbf{d}_1 \cdot \mathbf{d}_1^*}, e(g_1, g_2)^{\beta \mathbf{b}_1 \cdot \mathbf{b}_1^*},\ g_1^{\mathbf{d}_1}, g_1^{\mathbf{d}_2}, g_1^{\mathbf{b}_1}, g_1^{\mathbf{b}_2}\}$$

with $sk_{server} := \{\beta, g_2^{\mathbf{b}_1^*}, g_2^{\mathbf{b}_2^*}\}$ and the keep the secret key $sk_{client} := \{\alpha, g_2^{\mathbf{d}_1^*}, g_2^{\mathbf{d}_2^*}\}$ is known to $\mathcal{B}_0$.

Token Queries. Since $\mathcal{B}_0$ has the msk, it simply responds to all of $\mathcal{A}$'s token queries by running the normal **TokenGen(•)** algorithm. Compute the token of keyword "w" as:

$$T_w = g_2^{(\alpha + rw)\mathbf{d}_1^* - r\mathbf{d}_2^*}$$

Challenge. $\mathcal{A}$ sends $\mathcal{B}_0$ two keywords w_0 and w_1. Then $\mathcal{B}_0$ chooses a random bit $\beta \in \{0,1\}$ and $s \in \mathbb{Z}_q$, constructs the challenge ciphertext as follows:

$$S_1 := C_1(C_2)^{w_\beta} C_3(C_4)^I, S_2 := e(C_1, g_2^{\mathbf{f}_2^*})^\alpha e(C_3, g_2^{\mathbf{e}_2^*})^\beta$$

Here $\mathcal{B}_0$ sets $s := \tau_1$, and gives $[S_1, S_2]$ to $\mathcal{A}$. If C_1, C_2, C_3, C_4 are equal to $g_1^{\tau_1 \mathbf{f_1}}$, $g_1^{\tau_1 \mathbf{f_2}}$, $g_1^{\tau_1 \mathbf{e}_1}$, $g_1^{\tau_1 \mathbf{e_2}}$, then this properly distributed normal trapdoor of w_β. In this case, $\mathcal{B}_0$ has properly simulated $\mathbf{Game}_R$. If C_1, C_2, C_3, C_4 are equal to $g_1^{\tau_1 \mathbf{f_1} + \tau_2 \mathbf{f_3}}, g_1^{\tau_1 \mathbf{f_2} + \tau_2 \mathbf{f_4}}, g_1^{\tau_1 \mathbf{e_1} + \tau_2 \mathbf{e_3}}, g_1^{\tau_1 \mathbf{e_2} + \tau_2 \mathbf{e_4}}$ instead, then the ciphertext element S_1 has an additional term of $w_\beta \tau_2 \mathbf{f_3} - \tau_2 \mathbf{f_4} + I\tau_2 \mathbf{e_3} - \tau_2 \mathbf{e_4}$ as its component in the span of $\mathbf{f_3}, \mathbf{f_4}, \mathbf{e}_3, \mathbf{e}_4$.

The coefficients here in the basis $\mathbf{f_3}, \mathbf{f_4}, \mathbf{e_3}, \mathbf{e_4}$ form the vector $(w_\beta \tau_2, -\tau_2)$ and $(I\tau_2, -\tau_2)$. To compute the coefficient in the basis $\mathbf{d_3^*}, \mathbf{d_4^*}$ and $\mathbf{b_3^*}, \mathbf{b_4^*}$, we multiply the matrix A^{-1}, B^{-1} by the transpose of this vector, obtaining $\tau_2 A^{-1}(w_\beta, -1)'$ and $\tau_2 B^{-1}(I, -1)'$. Since A and B are both random, these coefficients are uniformly random from Lemma 1. Therefore, in this case, B_κ has properly simulated $\mathbf{Game_0}$. This allow B_κ to leverage $\mathcal{A}$'s advantage ϵ between $\mathbf{Game}_R$ and $\mathbf{Game}_\kappa$ to achieve an advantage ϵ against the Subspace assumption in $\mathbb{G}_1$, namely $Adv_{\mathcal{B}_\kappa}^{DS1} = \epsilon$.

Therefore, in this case, $\mathcal{B}_0$ has properly simulated $\mathbf{Game_0}$. This allow $\mathcal{B}_0$ to leverage $\mathcal{A}$'s advantage ϵ between $\mathbf{Game}_{Real}$ and $\mathbf{Game_0}$ to achieve an advantage ϵ against the Subspace assumption in $\mathbb{G}_1$, namely $Adv_{\mathcal{B}_0}^{DS1} = \epsilon$. □

Lemma 4. Suppose that there exists an adversary $\mathcal{A}$ that makes at most t trapdoor queries and $|Adv_{\mathcal{A}}^{Game_{\kappa-1}}(\lambda) - Adv_{\mathcal{A}}^{Game_\kappa}(\lambda)| = \epsilon$ for some κ where $1 \le \kappa \le q$. Then there exists an algorithm $\mathcal{B}_\kappa$ such that $Adv_{\mathcal{B}_0(\lambda)}^{DS1} = \epsilon - 1/q$, with $k = 4$ and $n = 8$.

Proof. $\mathcal{B}_\kappa$ begins by taking in an instance $D := (\mathbb{G}; g_2^{\mathbf{f_1^*}}, g_2^{\mathbf{f_2^*}}, g_2^{\mathbf{e_1^*}}, g_2^{\mathbf{e_2^*}}, g_1^{\mathbf{f_1}}, \ldots, g_1^{\mathbf{f_4}}, U_1, U_2, \mu_2)$ along with C_1, C_2, C_3, C_4 of the Decisional Subspace problem. We now describe how $\mathcal{B}_\kappa$ executes the Setup, Token Queries and Challenge algorithm to decide whether C_1, C_2, C_3, C_4 are distributed as $g_1^{\tau_1 \mathbf{f_1}}$, $g_1^{\tau_1 \mathbf{f_2}}$, $g_1^{\tau_1 \mathbf{e}_1}$, $g_1^{\tau_1 \mathbf{e_2}}$ or $g_1^{\tau_1 \mathbf{f_1} + \tau_2 \mathbf{f_3}}, g_1^{\tau_1 \mathbf{f_2} + \tau_2 \mathbf{f_4}}, g_1^{\tau_1 \mathbf{e_1} + \tau_2 \mathbf{e_3}}, g_1^{\tau_1 \mathbf{e_2} + \tau_2 \mathbf{e_4}}$. The following proof can reference to lemma 3. □

Lemma 5. Suppose that there exists an algorithm $\mathcal{A}$ that makes at most t queries, then we can build an algorithm $\mathcal{B}$ that has $Adv_{\mathcal{A}}^{Game_t} = Adv_{\mathcal{A}}^{Game_{Final}}$.

Proof. Similar with above one, to prove this Lemma, we just need to show the joint distributions of $\left(params, S_{w_\beta}^F, \{T_{w_\kappa^F}\}_{l=1,\ldots,t}\right)$ in $\mathbf{Game}_t$ and that of $\left(params, S_{w_X}^X, \{T_{w_\kappa^F}\}_{l=1,\ldots,t}\right)$ in $\mathbf{Game}_F$ are equivalent to the adversary's view, where $S_{W_X}^X$ is a semi-functional e-PEKS of a random message in $\mathbb{Z}_q$.

For this purpose, we pick $A := (\xi_{i,j}) \xleftarrow{R} \mathbb{Z}_q^{2\times 2}$ and define new dual orthonormal bases $\mathbb{F} := (\mathbf{f_1}, \cdots, \mathbf{f}_4)$, $F^* := (\mathbf{f_1^*}, \cdots, \mathbf{f_4^*})$ and $\mathbb{E} := (\mathbf{e_1}, \cdots, \mathbf{e}_4)$, $\mathbb{E}^* := (\mathbf{e_1^*}, \cdots, \mathbf{e_4^*})$ as follows:

$$\begin{pmatrix}\mathbf{f_1}\\ \mathbf{f_2}\\ \mathbf{f_3}\\ \mathbf{f_4}\end{pmatrix} := \begin{pmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ \xi_{1,1} & \xi_{1,2} & 1 & 0\\ \xi_{2,1} & \xi_{2,2} & 0 & 1\end{pmatrix}\begin{pmatrix}\mathbf{d_1}\\ \mathbf{d_2}\\ \mathbf{d_3}\\ \mathbf{d_4}\end{pmatrix}, \quad \begin{pmatrix}\mathbf{f_1^*}\\ \mathbf{f_2^*}\\ \mathbf{f_3^*}\\ \mathbf{f_4^*}\end{pmatrix} := \begin{pmatrix}1 & 0 & -\xi_{1,1} & -\xi_{2,1}\\ 0 & 1 & -\xi_{1,2} & -\xi_{2,2}\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{pmatrix}\begin{pmatrix}\mathbf{d_1^*}\\ \mathbf{d_2^*}\\ \mathbf{d_3^*}\\ \mathbf{d_4^*}\end{pmatrix}$$

and

$$\begin{pmatrix}\mathbf{e_1}\\ \mathbf{e_2}\\ \mathbf{e_3}\\ \mathbf{e_4}\end{pmatrix} := \begin{pmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ \xi_{1,1} & \xi_{1,2} & 1 & 0\\ \xi_{2,1} & \xi_{2,2} & 0 & 1\end{pmatrix}\begin{pmatrix}\mathbf{b_1}\\ \mathbf{b_2}\\ \mathbf{b_3}\\ \mathbf{b_4}\end{pmatrix}, \quad \begin{pmatrix}\mathbf{e_1^*}\\ \mathbf{e_2^*}\\ \mathbf{e_3^*}\\ \mathbf{e_4^*}\end{pmatrix} := \begin{pmatrix}1 & 0 & -\xi_{1,1} & -\xi_{2,1}\\ 0 & 1 & -\xi_{1,2} & -\xi_{2,2}\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{pmatrix}\begin{pmatrix}\mathbf{b_1^*}\\ \mathbf{b_2^*}\\ \mathbf{b_3^*}\\ \mathbf{b_4^*}\end{pmatrix}$$

It is easy to check that $\mathbb{F}$, $\mathbb{E}$ and $\mathbb{F}^*, \mathbb{E}^*$ are also dual orthonormal, and are distributed the same as $\mathbb{D}$, $\mathbb{B}$ and $\mathbb{D}^*, \mathbb{B}^*$. Then the public parameters, challenge ciphertext, and queried tokens $(pp, S^F_{w_\beta}, \{T_{w^F_\kappa}\}_{l=1,\dots,n})$ in Game $_n$ are expressed over bases $\mathbb{D}, \mathbb{D}^*$ and $\mathbb{B}$, $\mathbb{B}^*$ as

$$pp := \{\mathbb{G}, A, g_1^{\mathbf{d_1}}, g_1^{\mathbf{d_2}}, g_1^{\mathbf{b_1}}, g_1^{\mathbf{b_2}}\}$$

$$C^F_{w_\beta} := [C_1 = g_1^{(s+w_\beta)\mathbf{d_1} - s\mathbf{d_2} + z_3\mathbf{d_3} + z_4\mathbf{d_4} + (s+I)\mathbf{b_1} - s\mathbf{b_2} + z_3\mathbf{b_3} + z_4\mathbf{b_4}}, C_2 = (e(g_1, g_2)^{\alpha\mathbf{d_1}\mathbf{d_1^*} + \beta\mathbf{b_1}\mathbf{b_1^*}})^s]$$

$$\{T^F_{w_l} = g_2^{(\alpha + rw_l)\mathbf{d_1}^* - r\mathbf{d_2}^* + t_{l,3}\mathbf{d_3^*} + t_{l,4}\mathbf{d_4^*}}\}_{l=1,\dots,t}$$

Then we can express them over bases $\mathbb{F}$ and $\mathbb{F}^*$ as

$$pp := \{\mathbb{G}, A, g_1^{\mathbf{f_1}}, g_1^{\mathbf{f_2}}, g_1^{\mathbf{e_1}}, g_1^{\mathbf{e_2}}\}$$

$$C^F_{w_\beta} := [C_1 = g_1^{s'\mathbf{d_1} + s''\mathbf{d_2} + z_3\mathbf{d_3} + z_4\mathbf{d_4}}, C_2 = (e(g_1, g_2)^{\alpha\mathbf{f_1}\mathbf{f_1^*} + \beta\mathbf{e_1}\mathbf{e_1^*}})^s]$$

$$\{T^F_{w_l} = g_2^{(\alpha + rw_l)\mathbf{d_1}^* - r\mathbf{d_2}^* + t_{l,3'}\mathbf{d_3^*} + t_{l,4'}\mathbf{d_4^*}}\}_{l=1,\dots,t}$$

where (s', s'') are the linear combination of some uniformly values which are all uniformly picked from $\mathbb{Z}_q$.

In other words, the coefficients $(s + rw_\beta, -r)$ of $\mathbf{d_1}, \mathbf{d_2}$ in the S_1 term of the challenge searchable ciphertext is changed to random coefficients $(s', s'') \in \mathbb{Z}_q \times \mathbb{Z}_q$ of $\mathbf{f_1}, \mathbf{f_2}$, thus the challenge ciphertext can be viewed as a semi-functional ciphertext of a random message in G_T and under a random keyword in w. Moreover, all coefficients $\{(t_{l,3'}, t_{l,4'})\}_{l=1,\dots,t}$ of $\mathbf{f}_1, \mathbf{f_2}$ in the $\{T^F_{w_l}\}_{l=1,\dots,t}$ are uniformly distributed since $\{(t_{l,3}, t_{l,4})\}_{l=1,\dots,t}$ of $\mathbf{d_3^*}, \mathbf{d_4^*}$ are all independent random values. Thus $(pp, C^F_{w_\beta}, \{T_{w^F_\kappa}\}_{l=1,\dots,t})$ expressed over bases $\mathbb{F}$ and $\mathbb{F}^*$ is distributed as $(pp, S^R_{w_R}, \{T_{w^F_\kappa}\}_{l=1,\dots,t})$ in Game $_F$.

In the adversary's view, both $(\mathbb{D}, \mathbb{D}^*)$ and $(\mathbb{F}, \mathbb{F}^*)$ are consistent with the same public key. Therefore, the challenge searchable ciphertext in the two ways, in Game $_n$ over bases $(\mathbb{D}, \mathbb{D}^*)$ and in $Game_F$ over bases $(\mathbb{F}, \mathbb{F}^*)$. Thus, Game $_t$ and Game $_F$ are statistically indistinguishable.

Through the above three Lemma, we have that the advantage gap between $\mathbf{Game}_R$ and $\mathbf{Game}_0$ is bounded by the advantage of the $DS1$, and the distribution of the challenge PEKS remains same from the adversary's view because of the statical indistinguishability we required. For κ from 1 to t, the gap between $\mathbf{Game}_{\kappa-1}$ and $\mathbf{Game}_{\kappa}$ is bounded by the advantage of $DS2$. Similarly, we require a statical indistinguishability argument to show that the distribution of the κ-th semi-function key remains the same from the adversary's view. The last step shows a statical way to transform $\mathbf{Game}_{\kappa}$ to $\mathbf{Game}_F$ and prove they are equivalent for adversary's view. So we have

$$Adv_{\mathcal{A}}^{PEKS}(\lambda) \le Adv_{\mathcal{B}_0}^{DS1}(\lambda) + \sum_{\kappa=1}^{t} Adv_{\mathcal{B}_\kappa}^{DS2}(\lambda) + \frac{t}{q}$$

These means that: If $DS1$ and $DS2$ assumption holds, then the adversary's advantage of breaking the PEKS scheme is negligible.

5. Complexity and efficiency analysis

In this section, we simply analyze the complexity and efficiency of our scheme by giving its computation and communication cost and comparing with some classical searchable encryption construction. Here we set all the number of keywords be one so to compare easily. Let $|\mathbb{G}_1|,|\mathbb{G}_2|,|\mathbb{G}_T|,|\mathbb{Z}_q|$ respectively denote the size of the element of $\mathbb{G}_1,\mathbb{G}_2,\mathbb{G}_T,\mathbb{Z}_q$, , then the detailed communication cost of the proposed scheme is listed in **Table 1**.

Table 1. The comparison of communication cost in several classical schemes in section 3.1

Scheme	Setup	TokenGen	SEncrypt	ROM
Valent et.al	$2\lvert\mathbb{G}_1\rvert+2\lvert\mathbb{Z}_q\rvert$	$\lvert\mathbb{G}_1\rvert$	$3\lvert\mathbb{G}_1\rvert$	Yes
Fang et.al	$8\lvert\mathbb{G}_1\rvert+3\lvert\mathbb{Z}\rvert$	$\lvert\mathbb{G}_2\rvert+\lvert\mathbb{Z}_q\rvert$	$4\lvert\mathbb{G}_1\rvert+2\lvert\mathbb{G}_T\rvert$	Yes
Xu et.al	$8\lvert\mathbb{G}_1\rvert+\lvert\mathbb{G}_T\rvert$	$\lvert\mathbb{G}_2\rvert$	$4\lvert\mathbb{G}_1\rvert+\lvert\mathbb{G}_T\rvert$	No
Our scheme	$16\lvert\mathbb{G}_1\rvert+2\lvert\mathbb{G}_T\rvert$	$\lvert\mathbb{G}_2\rvert$	$4\lvert\mathbb{G}_1\rvert+\lvert\mathbb{G}_T\rvert$	No

Through the table above, we find that we can achieve a e-PEKS scheme with designed tester and user in standard security model without significantly more communication consumption. Moreover, we make the token in the proposed can be transmitted in an open channel by some special treatment of the ciphertext, which will be able to avoid the problem of information leakage due to the loss of trap door. Additionally, we also elaborate more on these details by listing the running time of every algorithm in several classical searchable encryption scheme that are similar with ours in **Fig. 3**. From the above table, we notice that the SE scheme proposed in section 3 is more efficient than Agrawal's and Park's scheme in paper.

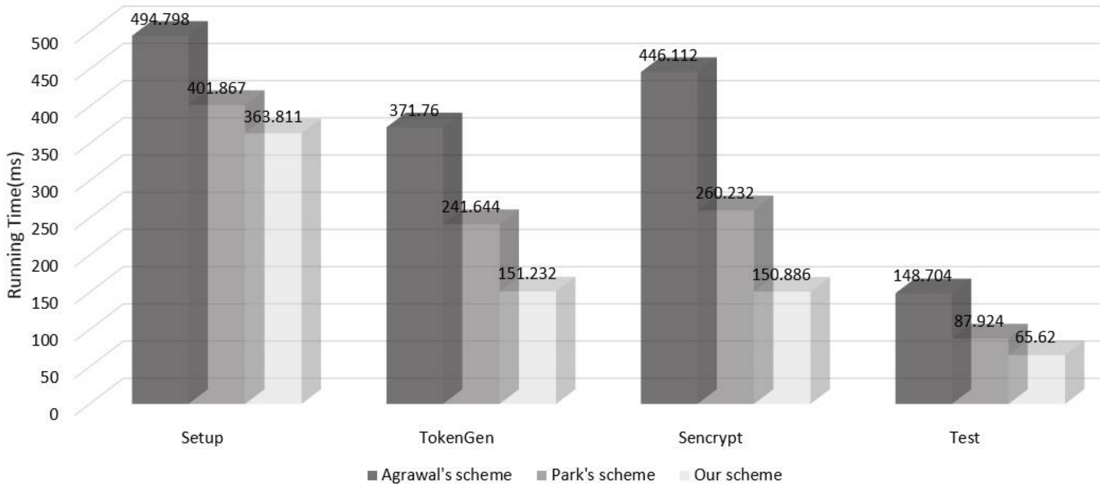


Fig. 3. The running time of several classical searchable encryption scheme

7. Conclusion

We construct an efficient and practical searchable encryption scheme via asymmetric pairing in the standard model, and prove the security of the scheme by using the dual system technique to reduce it to the decisional-Subspace assumption. We also give the detailed communication cost and computation cost of the proposed scheme and point out that our scheme is more efficient than other classical ones by comparing the running time with some classical searchable encryption in each phase.

Acknowledgment

This work is partially supported by The Natural Science Foundation of Jiangsu Province (No. BK20141405). The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

References

[1] B. Chor, O. Goldreich, E. Kushilevitz, et al., "Private information retrieval," in *Proc. of Foundations of Computer Science*, pp. 41-50, 1995. Article (CrossRef Link)

[2] D. Boneh, G. Di Crescenzo, R. Ostrovsky, et al., "Public key encryption with keyword search," in *Proc. of International Conference on the Theory and Applications of Cryptographic Techniques*. Springer Berlin Heidelberg, pp.506-522, May 2-6, 2004. Article (CrossRef Link)

[3] B. R. Waters, D. Balfanz, G. Durfee, et al., "Building an Encrypted and Searchable Audit Log," in *Proc. of NDSS*. vol.4, pp.5-6, February 5-6, 2004. Article (CrossRef Link)

[4] P. Golle, J. Staddon, B. Waters, "Secure conjunctive keyword search over encrypted data," in *Proc. of the 2nd Intl Conf. on Applied Cryptography and Network Security (ACNS)*. pp. 31-45, June 8-11, 2004. Article (CrossRef Link)

[5] R. Curtmola, J. Garay, S. Kamara, R. Ostrovsky, "Searchable symmetric encryption: Improved definitions and efficient constructions," in *Proc. of the 13th ACM Conf. on Computer and Communications Security (CCS).* pp. 79-88, October 30 - November 3, 2006. Article (CrossRef Link)

[6] Kaoru Kurosawa and Yasuhiro Ohtaki, "UC-Secure searchable symmetric encryption," in *Proc. of FC (LNCS)*,vol. 7397, pp. 285–298, February 27–March 2, 2012. Article (CrossRef Link)

[7] B. Dan, A. Sahai, and B. Waters,"Functional encryption: Definitions and challenges," in *Proc. of Theory of Cryptography*,pp.253–273, March 28-30, 2011. Article (CrossRef Link)

[8] Q. Zheng, S. Xu, G. Ateniese. "VABKS: verifiable attribute-based keyword search over outsourced encrypted data," in *Proc. of Infocom*, pp.522-530, 2014. Article (CrossRef Link)

[9] J. Katz, A. Sahai, B. Waters, "Predicate encryption supporting disjunctions, polynomial equations, and inner products," *Journal of cryptology*, vol. 26, no. 2, pp. 191-224, 2013. Article (CrossRef Link)

[10] T. F. Vallent, H. Kim, "A Pairing-Free Public Key Encryption with Keyword Searching for Cloud Storage Services," in *Proc. of e-Infrastructure and e-Services for Developing Countries*, pp.70-78, November 25-27, 2014. Article (CrossRef Link)

[11] L. Xu, C. G. Xu, "Efficient and Secure Data Retrieval Scheme Using Searchable Encryption in Cloud Storage," in *Proc. of International Symposium on Security and Privacy in Social Networks and Big Data*, pp.15-21, November 16-18, 2015. Article (CrossRef Link)

[12] R. Zhang, H. Imai, "Generic combination of public key encryption with keyword search and public key encryption," in *Proc. of Cryptology and Network Security*, pp.159-174, December 12-14, 2007. Article (CrossRef Link)

[13] L. Fang, W. Susilo, C. Ge, et al., "Public key encryption with keyword search secure against keyword guessing attacks without random oracle," *Information Sciences*, vol. 238, no. 7, pp. 221–241, 2013. Article (CrossRef Link)

[14] D. Boneh, X. Boyen, "Efficient Selective Identity-Based Encryption Without Random Oracles," *Journal of Cryptology*, vol. 24, no. 4, pp. 659-693, 2011. Article (CrossRef Link)

[15] J. Chen, H. W. Lim, S. Ling, et al., "Shorter IBE and signatures via asymmetric pairings," in *Proc. of Pairing-Based Cryptography*, pp.122-140, November 22-24, 2013. Article (CrossRef Link)

[16] A. Lewko, "Tools for Simulating Features of Composite Order Bilinear Groups in the Prime Order Setting," in *Proc. of Eurocrypt 2012*, pp.318–335, April 15–19, 2012. Article (CrossRef Link)

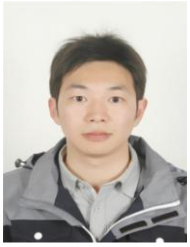
[17] J. Groth, A. Sahai, "Efficient Non-interactive Proof Systems for Bilinear Groups," in *Proc. of EUROCRYPT 2008*, LNCS, vol. 4965, pp.415–432, April 13-17, 2008. Article (CrossRef Link)

[18] B. Waters, "Dual System Encryption: Realizing Fully Secure IBE and HIBE under Simple Assumptions," in *Proc. of CRYPTO 2009*, pp.619–636, August 16-20, 2009. Article (CrossRef Link)

[19] A. Lewko, B. Waters, "New Techniques for Dual System Encryption and Fully Secure HIBE with Short Ciphertexts," in *Proc. of Theory of Cryptography*, pp.455–479, February 9-11, 2010. Article (CrossRef Link)

[20] A. Lewko, T. Okamoto, A. Sahai, et al., "Fully Secure Functional Encryption: Attribute-Based Encryption and (Hierarchical) Inner Product Encryption," in *Proc. of Advances in Cryptology EUROCRYPT 2010*, pp.62–91, May 30- June 3, 2010. Article (CrossRef Link)

[21] J. H. Park, "Inner-product encryption under standard assumptions," *Designs, Codes and Cryptography*, vol. 58, no. 3, pp. 235–257, 2011. Article (CrossRef Link)



Lei Xu received his bachelor degree from Anhui Normal University, China, in 2012. From 2012 to now, he is working his Ph.D. degree in School of Science, Nanjing University of Science and Technology, Nanjing, China. During the period from April 2017 to April 2018, he is also a visiting Ph.D. student at Faculty of Information Technology, Monash University. His main research interests focus on public key cryptography and information security, including searchable encryption mechanism and identity-based encryption.

Chungen Xu received the M.S. degree from East China Normal University, Shanghai, China, in 1996 and the Ph.D. degree from Nanjing University of Science and Technology in 2003. He is a professor in the Department of Mathematics, School of Sciences, Nanjing University of Science and Technology. His current interests are in the areas of computer and network security, cryptography and coding.

Xing Zhang received the Ph.D. degree from Nanjing University of Science & Technology, China, in 2016. From 2016 to now, she is working in School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang, China. During the period from November 2013 to May 2014, she was also a visiting Ph.D. student at School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. Her research interests include information security and cryptography, and the encryption scheme based on cellular automata.