JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 33, 485-497 (2017)
DOI:10.1688/JISE.2017.33.2.12

Post-Quantum Secure Public Key Broadcast Encryption with Keyword Search*

YANG YANG[1,+], SHU-LVE YANG[2], FENG-HE WANG[3] AND JIN SUN[4]
[1]*College of Mathematics and Computer Science*
[2]*College of Physics and Information Engineering*
Fuzhou University
Fuzhou, 350108 China
[3]*Department of Mathematics and Physics*
Shandong Jianzhu University
Jinan, 250101 China
[4]*Department of Application Mathematics*
Xi'an University of Technology
Xi'an, 710048 China
[+]*E-mail: yang.yang.research@gmail.com*

In remote data storage system, the information privacy is the main concern of the users. Data confidentiality, keyword search and access control are important characteristics that should be supported in such system. With the advent of quantum computer, number theory related assumptions are vulnerable to quantum attack. In this context, we propose a post-quantum secure searchable encryption scheme. Data owner's files are stored in encrypted form. The keyword search function can be authorized to multiple users without sharing private key. Based on the learning with errors (LWE) assumption, the scheme is proved secure against chosen-keyword attack. The intractability of LWE problem guarantees its security in quantum age.

Keywords: post-quantum secure, multi-user, public key encryption with keyword search, chosen keyword attack, learning with errors

1. INTRODUCTION

With the mushroom growth of information technology, the demand for remote data storage increased dramatically in recent years. More and more important information are centralized into cloud including enterprise financial documents, government information and personal electronic health files, *etc*. At the same time, the security risk of data management in cloud arouses. Besides external attacks, there is no effective way to prevent users' private data from been watched stealthily by internal personnel or sold to a third party for profit. The privacy concern should be addressed to dispel client misgivings. Data encryption could provide data confidentiality to protect the data privacy but at the same time hinders the traditional data utilization operation, including the most commonly used information retrieval.

In order to provide both data protection and convenient keyword retrieval, the concept of public key encryption scheme with keyword search (PEKS) is proposed by Boneh *et al*. [1] in 2004. In a PEKS scheme, a sender generates an encrypted document

Received December 8, 2015; revised February 18 & March 27, 2016; accepted May 13, 2016.
Communicated by Hung-Min Sun.
* The work is supported by National Natural Science Foundation of China under Grant No. 61402112, 61472307, 61472309, 61303198, 61303223; Science and Technology Project of Fujian Education Department under Grant No. JA12028; Major Science and Technology Project of Fujian Province under Grant No. 2015H6013.

together with an encrypted set of keywords. The original plaintext file can be encrypted with a standard public key encryption algorithm and the keywords should be encrypted with PEKS scheme. User should submit a trapdoor of the searched keyword generated by his private key. In the search phase at server side, the data server tests whether the encrypted set of keywords contains the same keyword encapsulated in trapdoor. In that way, all files contain the searched keyword can be retrieved and returned to user.

So far the overwhelming majority existing searchable encryption schemes are built based on the number theory assumptions such as big integer factorization problem and discrete logarithm problem (DLP). However, both DLP and factorization problems can be solved in polynomial time by quantum algorithm. It means that those algorithms rely on these problems are fragile in post-quantum age. As a result, there is a pressing need to have novel searchable encryption scheme that can resist quantum attack.

In this paper, we construct a novel primitive called public key broadcast encryption scheme with keyword search (BEKS) based on lattice assumption to resist quantum attack. This system enables a group of users (with distinct private key) to outsource encrypted files to the data server. Each user in the group is permitted to search all the encrypted documents that contain the chosen keyword, in which the trapdoor of keyword is generated using their own private key. In another word, the group member do not need to share a common private key in order to do search operation. We have designed a systematic solution with a system model, a security model and detailed security analysis. In the proposed scheme, the basis delegation technique and pre-image sample function (PSF) is utilized. The dimension of lattice is maintained in the process of delegation. The size of public key, private key and ciphertext are constant. This scheme is also proved to be secure against chosen keyword attack (CKA) assuming that learning with errors (LWE) assumption holds.

This paper is organized as follows. Related works are analyzed in Section 2. Section 3 introduces the concept of lattice, the hardness assumption, the lattice delegation. Section 4 presents the system model, the security model and the proposed lattice based searchable broadcast encryption scheme. Section 5 shows the analysis and security proof of the construction. A summarization is given in Section 6.

2. RELATED WORK

Since the seminal work proposed by Boneh *et al*. [1], searchable encryption becomes a hot research topic. Waters [2] showed that the PEKS schemes based on bilinear map could be applied to build encrypted and searchable auditing logs. Crescenzo *et al*. [3] proposed a PEKS scheme based on Jacobi symbols. Hu *et al*. [4] suggested a PEKS scheme that enables multi-keyword search. Attrapadung *et al*. [5] presented the concept of searchable broadcast encryption without any concrete construction. Abdalla *et al*. [6] deals with the transformations among primitives related to PEKS. Later, Han *et al*. [7] gave out a general transformation from attribute based encryption (ABE) to searchable encryption. Some efforts have also been devoted to make PEKS versatile. The work of this kind includes schemes that support fuzzy keyword search [8] and rank search [9]. However, these traditional searchable encryption schemes are fragile to quantum attack.

Broadcast encryption (BE) allows multiple users share encrypted data from a sender,

who selects an arbitrary set *S* of receivers and encrypts data for them. Only the users in *S* is able to decrypt the protected data. BE is very useful in digital right management systems. In order to simplify the certificate management problem, a use's identity information can be employed for public keys. Delerablee proposed the first identity based broadcast encryption (IBBE) scheme in [10]. Later, Sakai *et al*. provided a more efficient IBBE scheme [11], in which the size of the private key is constants. Zhang *et al*. [12] constructed a new scheme and is proved secure in selective-ID model with a long ciphertexts, which means that the secret message broadcasted to the receivers will grow linearly with the number of designated group. The shortage will become serious with the increment of network scale. Moreover, these BE schemes could not support keyword search function.

In 1994, a seminal work was proposed by Shor *et al* [13]. They showed that discrete logarithm and prime factorization problem can be solved in polynomial time by a quantum computer. With the advent of quantum computer, these available public key schemes used for Internet protection will be broken. The result prompt the research of designing crypto schemes that could resist quantum attack. Fortunately, there remain mathematically hard problems that can be used in cryptography for post-quantum age, for instance, Lattice based problems. Till now, there is no polynomial time quantum algorithm to solve lattice based problems. A hardness assumption called learning from errors (LWE) problem is defined by Regev *et al*. [14] in 2005 together with shortest vector problem (SVP). Regev also proposed a concrete scheme relies on LWE and a security proof is provided. There are many lattice based cryptography utilize LWE problem as the foundation, for example, the public key encryption (PKE) scheme [15], hierarchical identity-based encryption (HIBE) schemes [19-21] and other cryptography schemes [16-18].

In 2012, Zhang *et al*. [22] proposed a lattice based searchable encryption scheme. However, serious problem exists in their scheme. In fact, their scheme has not even generated public and private keys for users. The result is that the scheme has a confused logic. The trapdoor of keyword is not protected by user's private key. Even if an attacker has no access permission from a data owner, he is also capable to issue a keyword search on the encrypted documents of the data owners. Moreover, the test equation does not hold. Later on, Gu [23] and Hou [24] *et al*. proposed PEKS schemes from lattice, respectively. These two constructions are quite similar and both of them are designed based on the scheme shown in [21]. Furthermore, these PEKS schemes that based on lattice assumption are all designed for single user and no data sharing can be realized. It cannot provide a remote storage system for a number of users.

3. PRELIMINARIES

3.1 Lattice and Hardness Problem

Let $A=[a_1, \ldots, a_m]$ denote a set of linearly independent vectors. An n-dimensional lattice generated by A is defined as $\Lambda=\{Ac=\sum_{i\in\{1,\cdots,n\}} c_i a_i \in Z\}$, where A plays the role of basis for this lattice. For a prime number q and a vector $y\in Z_q^n$, two kinds of modular lattices are defined as $\Lambda_q^{\perp}(\mathrm{A})=\{ x\in Z_q^m | Ax=0(\mathrm{mod}\ q)\}$ and. $\Lambda_q^y(A)=\{x\in Z_q^m | Ax=y(\mathrm{mod}\ q)\}$.

Definition 1: (Learning with Errors (LWE) Problem) Let n, m, q be positive integers, $s\in Z_q^n$ be a vector and χ be an error distribution over Z_q^m. $A_{s,\chi}$ is a distribution obtained

by computing $\{A, A^T s+x(\text{mod } q)\}$ where $A \in Zq^{n\times m}$ is randomly chosen and errors vector x is chosen in accordance with the error distribution χ. The decisional LWE problem is to distinguish $A_{s,\chi}$ from a uniform distribution.

Lemma 1: [21] Let e be a vector in Z^m and let $y \in \bar{\Psi}_{\alpha}^{m}$. Then the quantity $|e^T y|$ treated as an integer in $[0, q-1]$ satisfies $|e^T y| \le ||e|| q\alpha\omega(\sqrt{\log m})+||e||\sqrt{m}/2$ with all but negligible probability in m.

3.2 Trapdoors for Lattices and Basic Delegation Algorithm

Lemma 2: [25] Let n, q, m be positive integers with $q\ge 3$ (an odd integer) and $m=[6n\log q]$. There exists a PPT algorithm *TrapGen* that outputs a pair $(A\in Z_q^{n\times m}, B\in Z_q^{m\times m})$ such that A is statistically close to uniform on $Z_q^{n\times m}$ and B is a basis of $\Lambda_q^{\perp}(A)$ such that $||B||\le O(n\log q)$ and $||\tilde{B}||\le O(n\log q)$.

Let $L_{TG}=O(\sqrt{n\log q})$ to be the maximum Gram-Schmidt norm of a basis produced by *TrapGen*.

Definition 2: [19] Defines $\sigma_1 = L_{TG}\omega(\sqrt{\log m}) = O(\sqrt{n\log q})\omega(\sqrt{\log m})$. Let $D_{m\times m}$ denote the distribution on matrices in $Z_{m\times m}$ defined as $(D_{Z^m, r})^m$ conditioned on the resulting matrix being Z_q-invertible. An algorithm *SampleR*(1^m) is used to samples matrices in $Z_{m\times m}$ from a distribution that is statistically close to $D_{m\times m}$.

Definition 3: [19] Taken as input a matrix $A\in Z_q^{n\times m}$, a Z_q-invertible matrix R sampled from $D_{m\times m}$, a short basis B for $\Lambda_q^{\perp}(A)$ and $\sigma \ge ||\tilde{B}||(\sigma_1\sqrt{m}\omega(\log^{3/2} m))$, the algorithm *BasisDel*(A, R, B, σ) outputs a short basis B' for $\Lambda_q^{\perp}(A')$, where $A'=AR^{-1}$.

Theorem 1: [19] Let n, q, m, k be positive integers with $q>2$ (prime number) and $m\ge 2n\log q$. There exist a PPT algorithm *SampleRwithBasis* such that on input of $A\in Z_q^{n\times m}$, it outputs a matrix $R\in Z^{m\times m}$ that is within negligible statistical distance from $D_{m\times m}$. and a basis B' for $\Lambda_q^{\perp}(AR^{-1})$ that satisfies $||\tilde{B}'|| \le \sigma/\omega(\sqrt{\log km})$ with overwhelming probability.

Let $A\in Z_q^{n\times km}$ and denote $A=[A_1, \ldots, A_k]$, where $A_i\in Z_q^{n\times km}$. For $S\subseteq\{1, \ldots, k\}$, $S=\{i_1, \ldots, i_j\}$, denote A_s as $[A_{i_1}, \ldots, A_{i_j}]$.

Theorem 2: Let n, q, m, k be positive integers with $q\ge 2$ and $m\ge 2n\log q$. There exist a PPT algorithm *GenSamplePre* such that on input of $A_S\in Z_q^{n\times km}$, a set $R\subseteq\{1, \ldots, k\}$, a basis B_R for $\Lambda_q^{\perp}(A_R)$, a vector $y\in Z_q^n$ and an integer $L\ge ||\tilde{B}_R||\cdot\omega(\sqrt{\log km})$, it outputs e that is within negligible statistical distance from the distribution $D_{\Lambda_q^y(A_S),r}$.

4. SEARCHABLE BROADCAST ENCRYPTION SCHEME FROM LATTICES

4.1 System Model

In the system model shown in Fig. 1, the data service provider (DSP) is responsible

to store the uploaded documents and respond to the search query of user. The DSP is deemed as semi-trusted, who is curious about the private information hidden in the stored files. Thus, the data sender has to encrypt their documents before they are outsourced. The keywords indexes of files are also uploaded to the DSP. This system allows a data user group to upload their encrypted files to DSP. At the same time, each user in the group is capable to query on the encrypted data. Each user in the system possesses a distinct private key. The selected keyword will be encapsulated to a keyword trapdoor using their own private key so that the DSP could not eavesdrop any plaintext information from the keyword trapdoor. In order to respond to the request, the DSP will search for the matched files from the data server and then return the results. Similarly, the DSP is not able to get any private information in the test process.

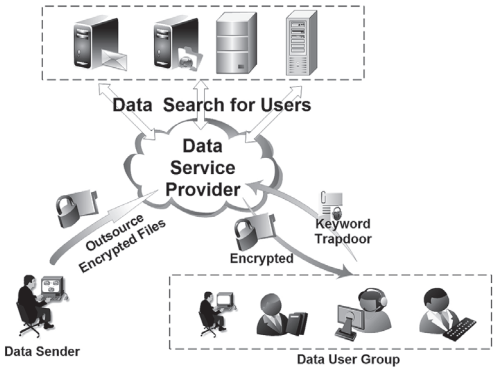

Fig. 1. System model.

4.2 Proposed Scheme

In this subsection, we design a searchable broadcast encryption scheme from lattices as follows. Let $H:\{0, 1\}^* \rightarrow Z_q^{n\times km}$ be hash function such that $H(KW)\sim D_{m\times m}$ for any keyword $KW\in\{0, 1\}^*$.

- ***Setup*(1^λ)→(*PP*, *MSK*):** This algorithm takes the security parameter 1^λ as input. Use algorithm *TrapGen* to select a uniformly random $n\times m$-matrix $A\in Z_q^{n\times m}$ with a basis B for lattice $\Lambda_q^\perp(A)$ such that $\| \tilde{B} \| \le L_{TG} = O(\sqrt{n\log q})$. It randomly chooses a matrix $V\in Z_q^{n\times t}$. Let $V=(v_1, \ldots, v_t)$, where $v_i\in Z_q^n$. It returns the public parameter $PP=(A, V)$ and master secret key $MSK=B$.
- ***KeyGen*(*MSK*, *ID*$_i$)→(*pk*$_i$, *sk*$_i$):** This algorithm takes the master secret key *MSK* and user's identity ID_i as an input. Let the matrix $A_i=A\cdot H(ID_i)^{-1}$ and $\sigma_1 = L_{TG}\omega(\sqrt{\log m})$. Choose a random parameter $\tau_i \ge \| \tilde{B} \| \sigma_1\sqrt{m}\omega(\log^{3/2} m)$. Run *BasisDel* algorithm to generate B_i, in which $B_i\leftarrow BasisDel(A, H(ID_i), B, \tau_i)$. According to definition 3, B_i is a short basis for $\Lambda_q^\perp(A_i)$. It returns $pk_i=A_i$, $sk_i=B_i$.

Note: The user is able to verify the received public key and secret key pair through the following procedure; (1) Verify whether the equation $A_i\cdot B_i=0(\text{mod } q)$ holds. If it holds, it means that B_i is a basis for lattice $\Lambda_q^\perp(A_i)$. If it does not holds, B_i is not a basis for lattice

$\Lambda_q^{\perp}(A_i)$. (2) If the verification in step 1 is true, verify whether $\| \tilde{B}_i \| < \sigma_R / \omega(\sqrt{\log m})$, in which $\sigma_R = \sqrt{n \log q} \cdot \omega(\sqrt{\log m})$ [19]. It is used to guarantee that B_i is a short basis for lattice $\Lambda_q^{\perp}(A_i)$.

- ***BEKS(KW, S, {pk_i}_{ID_i∈S})→CT*:** This algorithm takes a keyword KW, a receiver set $S=\{ID_{\theta_1}, \ldots, ID_{\theta_k}\}$ and their public keys as inputs. Let $A_S=[A_1 \cdot H(ID_{\theta_1} \| KW)^{-1} \| \ldots \| A_k \cdot H(ID_{\theta_k} \| KW)^{-1})] \in Z_q^{n \times km}$. Randomly select a vector $s \in Z_q^n$ and $C_1 \in \{0, 1\}^t$. Then, it computes $p=A_S^T s+x_1$, $C2=V^T s+x_2+C_1 \lfloor q/2 \rfloor$, where $x_1 \leftarrow \chi^{km}$, $x^2 \leftarrow \chi^t$, $\chi = \Psi_\alpha^k$. It outputs $CT=(p, C_1, C_2, A_S)$ as the ciphertext.
- ***Trapdoor(sk_i, KW_j)→T_{KW_j,i}*:** This algorithm takes sk_i and KW_j as inputs. Let $T_{KW_j,i} = A_i \cdot H(ID_i \| KW_j)^{-1}$. Select a random $\sigma_2 \geq \| \tilde{B}_i \| \sigma_1 \sqrt{m} \omega(\log^{3/2} m))$, a short random basis $T_{KW_j,i} \leftarrow BasisDel(A_i, H(ID_i \| KW_j), B_i, \sigma_2)$ for lattice $\Lambda_q^{\perp}(A_{KW_j,i})$ will be generated by running basis delegation algorithm. It outputs $T_{KW_j,i}$ as the trapdoor of KW_j for user ID_i.
- ***Test(CT, T_{KW_i}, S)→1/0*:** This algorithm takes CT, T_{KW_i}, S as inputs. If $ID_i \notin S$, it returns 0. Otherwise, it generates $e_j \leftarrow GenSamplePre(A_S, T_{KW,i}, S=\{ ID_i \}, v_j, \lambda)$, where $\lambda = \sigma \sqrt{km} \omega(\sqrt{\log km}) \geq \| \tilde{B}_i \| \omega(\sqrt{\log km})$ and $j=1, ..., t$. According to theorem 2, e_j is within negligible statistical distance from the distribution $D_{\Lambda_q^{v_j}(A_S), \lambda}$ and $\| e_j \| \leq \lambda \sqrt{km}$. Let $C_2=(C_{2,1}, \ldots, C_{2,t})$ and $c_{2,j} \in Z_q$. Compute $\eta_j = c_{2,j} - e_j^T p \in Z_q$ for $j=1, ..., t$. Let $\mu_j = 0$ if η_j is closer to 0 than to $|q/2| \in Z_q$. Otherwise, $\mu_j=1$. If $[\mu_1, \ldots, \mu_t]$ equals to C_1, the algorithm returns 1 to indicate that the KW is included in CT. It outputs 0, otherwise.

Discussion: In this scheme, all user's private keys are generated by TTP. It is efficient for the system to add a new user. TTP just runs the *KeyGen* algorithm to generate a new public/private key pair for the enrolled user. In order to revoke a user, TTP sends a revocation notification to the data server. The revocation notification contains the revoked user identity, the revocation date together with a signature signed by TTP using the master secret key. Since the data server is responsible for the *test* algorithm, it will not respond on a keyword query from the revoked user.

5. ANALYSIS OF THE PROPOSED SCHEME

5.1 Parameters and Correctness

The correctness of this scheme is inherited from the properties of the trapdoor functions in [21] and the choice of parameters. We have $A_s e_j = v_j$ for $j=1, \ldots, t$ due to $e_j \leftarrow GenSamplePre(A_S, T_{KW,i}, S=\{ID_i\}, v_j, \lambda)$. If CT is a legally constructed ciphertext for keyword KW, then $p=A_S^T s+x_1$, $c_{2,j}=v_j^T s+x_{2,j}+c_{1,j} \lfloor q/2 \rfloor$. It is easy to find $\eta_i = c_{2,j} - e_j^T p = (v_j^T s + x_{2,j} + c_{1,j} \lfloor q/2 \rfloor) - e_j^T (A_S^T s + x_1) = c_{1,j} \lfloor q/2 \rfloor + (x_{2,j} - e_j^T x_1)$ since $v_j^T = (A_S e_j)^T = e_j^T A_S^T$. The formula $x_{2,j} - e_j^T x_1$ is the error term.

Due to the fact that $\| e_j \| \leq \lambda \sqrt{km} = \sigma km \omega(\sqrt{\log km})$ and Lemma 1, we have

$$\begin{aligned} |x_{2,j} - e_j^T x_1| &\leq \sigma km \omega(\sqrt{\log km})(q \alpha \omega(\sqrt{\log km}) + \sqrt{km}/2) \\ &\leq \sigma q \alpha km \omega(\log km) + \sigma (km)^{3/2} \omega(\sqrt{\log km}) \\ &\leq \sigma q \alpha Nm \omega(\log Nm) + \sigma (Nm)^{3/2} \omega(\sqrt{\log Nm}) \end{aligned}$$

In order to make the scheme work correctly, we need the following requirements.

Algorithm *TrapGen* is able to operate, then $m \geq 6n\log q$. The error term $| x_{2,j} - e_j^T x_1 |$ should no more than $q/5$. In trapdoor generation phase, algorithm *BasisDel* can execute, then $\sigma_2 > \| \tilde{B}_i \| \sigma_1 \sqrt{m}\omega(\log^{3/2} m)$, $\sigma_1 = \mathrm{LTG}\omega(\sqrt{\log m})$. Regev's LWE reduction can operate, then $k\alpha q > 2\sqrt{n}$.

We set the parameters (q, m, σ, α) as follows to satisfy the requirements (assume $n^{\delta} > \lceil \log q \rceil = O(k\log n)$). Using the similar techniques in [19], we set $m = 6n^{1+\delta}$, $q = m^{3N/2+2}\omega(\log^{2N+1} n)$, $\sigma = m^{7/2}\omega(\log^4 n)$ and $\alpha = 1/(\sigma N m \omega(\log n))$. Due to m is an integer and q is a prime number, we should round up m to the nearest larger integer and q to the nearest larger prime number.

5.2 Comparison and Performance Analysis

Since lattice based SE scheme is quite different from traditional SE scheme in structure, this scheme is compared with the available lattice based searchable encryption schemes [23, 24] in Table 1. The scheme in [22] is also claimed to be a searchable encryption scheme based on lattice. However, we omit it here due to that it suffers from serious flaws, which is analyzed in introduction. The storage and computation efficiency of the schemes is compared in Table 1, as well as the function and security. Since the schemes in [23] and [24] are constructed based on the IBE scheme in [21], these two constructions are almost the same. Thus, their performance are the same shown in Table 1. The storage efficiency of these schemes is analyzed and compared as following. This proposal needs more storage overhead due to the broadcast feature.

Table 1. Comparison.

	Scheme	[23]	[24]	Ours
Storage Efficiency	Public key size	$mn\cdot\log q$	$mn\cdot\log q$	$mn\cdot\log q$
	Private key	$m^2\cdot\log q$	$m^2\cdot\log q$	$m^2\cdot\log q$
	Trapdoor size	$m^2\cdot\log q$	$m^2\cdot\log q$	$m^2\cdot\log q$
	Ciphertext size	$(m+1)\log q$	$(m+1)\log q$	$(kmn+km+t)\log q$
Computation Efficiency	Setup	–	–	1*sa*
	KeyGen	1*sa*	1*sa*	1*psf*
	PEKS (BEKS)	2*mul*+2*add*	2*mul*+2*add*	2*mul*+3*add*
	Trapdoor	1*psf*	1*psf*	1*psf*
	Test	1*mul*	1*mul*	1*mul*+1*psf*
	Broadcast	No	No	**Yes**
	Lattice-based	Yes	Yes	**Yes**
Security	Security	CKA, ROM	CKA, ROM	**CKA, ROM**

psf: computation cost of running PSF algorithm.
sa: computation cost of running short basis extraction algorithm.
syn: computation cost of extending a synonym set Γ_{KW} of keyword *KW*.
mul: computation cost of multiplication between matrix.
add: computation cost of addition between matrix.
CKA: secure against chosen keyword attack.
ROM: random oracle model.

- **Public Key Size:** The public key of [23, 24] is a matrix $A \in Z_q^{n \times km}$, which has size $mn \cdot \log q$ bits. In our scheme, the public key pk_i is A_i, which has $mn \cdot \log q$ bits.
- **Private Key Size:** The private key sizes are the same in these schemes.
- **Trapdoor Size:** The trapdoor sizes are the same in these schemes.
- **Ciphertext Size:** The ciphertext of [23, 24] consists of p and c, which have size $m \cdot \log q$ bits and $\log q$ bits, respectively. In this scheme, (p, C_1, C_2, A_S) form the ciphertext, which have size $km\log q$, t, $t\log q$, $kmn\log q$ bits, respectively. (t bit is very small compared with other quantities, which is omitted in the table.)

The computation efficiency comparison shows that this construction requires one more *add* operation in encryption phase. In test process, our scheme needs one more *psf* operation. These additional computations are consumed to realize the broadcast function. This proposal supports more useful function compared with the schemes in [23, 24]. It can support multi-user setting and broadcast function while others can not. In that way, each authorized user can query on the encrypted files. Data sharing can be realized in this scheme. If the user of the schemes in [23, 24] desires to authorize the query ability to other users, he has to share his private key with those users. While in our scheme, the data user group is able to make keyword query using their own private key. The security of these three schemes is the same. They are all proved in random oracle model (ROM) and secure against chosen keyword attack (CKA).

To conclude, this scheme has the same security level as the schemes in [23, 24]. Moreover, this scheme realizes "broadcast" function for multiple-user without sharing secret key. A little more storage and computation overhead is required to achieve the broadcast function.

5.3 Security Analysis

Theorem 4: Let $m=\lceil 6n\log q \rceil$ and $q \le 3$ be a prime number, $q > 2\sqrt{n}/\alpha$, $\alpha \in (0,1)$. The proposed scheme is semantically secure against chosen keyword attacks assuming $LWE_{q,\chi}$ is intractable, where $\chi = \bar{\Psi}_\alpha^k$.

***Proof*:** Let l_1 and l_2 be the maximum bit length of identity *ID* and keyword *KW*. Let N be the maximum size of broadcast user set. Suppose there is an adversary $\mathcal{A}$ to break the scheme with advantage ε. Then we can construct a challenger $\mathcal{C}$ to solve the LWE problem with advantage $\varepsilon' \ge \varepsilon(1-2N^2/2^{l_1(l_1+l_2)})$. Adversary $\mathcal{A}$ firstly selects k^* to be the size of the challenge user set S^* and sends it to $\mathcal{C}$.

KeyGen: $\mathcal{C}$ does the following steps to generate public parameter *PP*.

1. Challenger $\mathcal{C}$ firstly obtains (k^*m+1) samples $(u_j, v_j) \in Z_q^n \times Z_q (0 \le j \le k^*m)$ from the LWE oracle, in which all u_j are randomly chosen and either all v_j are also randomly selected or all equal to $u_j^T s + \tau_j$. In the above sampling process, $s \in Z_q^n$ is a uniform secret and τ_j is the independent Gaussian Noise that is chosen in accordance with error distribution χ.
2. Set $A_i^* = (u_{(i-1)\times m+1}, \ldots, u_{i\times m}) \in (Z_q^n)^m$ for $1 \le i \le k^*$. Set $V = u_0$.
3. Select a random $A \in Z_q^{n \times m}$. The public parameter is $PP = (A, V)$, which is sent to adversary $\mathcal{A}$.

Phase 1: Attacker $\mathcal{A}$ adaptively issues the following queries.

- **Hash queries.** $\mathcal{A}$ may adaptively query the random oracle H for Q_H times. Challenger $\mathcal{C}$ maintains a list $\mathcal{H}_{list}$. $\mathcal{C}$ firstly sets $\mathcal{H}_{list}$ to be an empty matrices set and then answers the queries as follows.

–$\mathcal{H}_1$: For the queried identity ID_i, if $ID_i \notin \mathcal{H}_{list}$, challenger $\mathcal{C}$ does the following operations. $\mathcal{C}$ runs *SampleRwithBasis*(A) algorithm to obtain a random $R_{i,0} \in D_{m\times m}$ together with a short basis B_i for lattice $\Lambda_{qdc}^{\perp}(A_i)$, where $A_i=AR_{i,0}^{-1}$. Save the tuple (ID_i, $\perp$, $R_{i,0}$, A_i, B_i) in list $\mathcal{H}_{list}$ and returns $R_{i,0}$ to $\mathcal{A}$. If $ID_i \in \mathcal{H}_{list}$, $\mathcal{C}$ returns $R_{i,0}$ to $\mathcal{A}$ from the tuple directly.

–$\mathcal{H}_2$: For the queried tuple (ID_i, KW_j), challenger $\mathcal{C}$ does the following operations. If (ID_i, $\perp$)$\notin \mathcal{H}_{list}$, $\mathcal{C}$ runs $\mathcal{H}_1$ to obtain the tuple (ID_i, $\perp$, $R_{i,0}$, A_i, B_i). Otherwise, $\mathcal{C}$ gets ($R_{i,0}$, A_i, B_i) from $\mathcal{H}_{list}$ directly. $\mathcal{C}$ chooses a random $R_{i,j} \in D_{m\times m}$ such that $R_{i,0} \neq R_{i,j}$. Save the tuple (ID_i, KW_j, $R_{i,j}$, A_i, B_i) in list $\mathcal{H}_{list}$ and returns $R_{i,j}$ to $\mathcal{A}$. If (ID_i, KW_j)$\in \mathcal{H}_{list}$, $\mathcal{C}$ returns $R_{i,j}$ to $\mathcal{A}$ from the tuple directly.

- **Private key queries.** On the private key query from adversary $\mathcal{A}$ on identity ID_i, it is assumed that $\mathcal{A}$ has already made a hash query on ID_i. Challenger $\mathcal{C}$ looks up the hash list $\mathcal{H}_{list}$ to retrieve the (ID_i, $\perp$, $R_{i,0}$, A_i, B_i) tuple. $\mathcal{C}$ returns Bi as the private key of identity ID_i.
- **Trapdoor queries.** On the trapdoor generation query from adversary $\mathcal{A}$ on identity ID_i and keyword KW_j, it is assumed that $\mathcal{A}$ has already made a hash query on (ID_i, KW_j). $\mathcal{C}$ looks up the hash list $\mathcal{H}_{list}$ to retrieve the (ID_i, KW_j, $R_{i,j}$, A_i, B_i) tuple. Then, $\mathcal{C}$ runs $T_{KW_{j,i}} \leftarrow BasisDel(A_i, R_{i,j}, B_i, \sigma_2)$ in order to get a short basis $T_{KW_{j,i}}$ for lattice $\Lambda_q^{\perp}(A_i \cdot H(ID_i \| KW_j))^{-1}$, in which $\sigma_2 > \|\tilde{B}_i\| \sigma_1 \sqrt{m} \omega(\log^{3/2} m)$. The constructed short basis $T_{KW_{j,i}}$ is a trapdoor for keyword KW_j and identity ID_i, which is returned to adversary $\mathcal{A}$.

Challenge: When $\mathcal{A}$ decides that phase 1 is over, $\mathcal{A}$ sends the challenger $\mathcal{C}$ two keywords KW_0^*, KW_1^* and a receiver set $S^*=(ID_1^*, \ldots, ID_{k^*}^*)$ on which he wants to be challenged, in which $k \leq N$. The restriction is that The secret key of ID_β is not queried in phase 1, in which $ID_\beta \in S^*$. The trapdoor of (KW_α, ID_β) is not queried in phase 1, in which $KW_\alpha \in \{KW_0^*, KW_1^*\}$ and $ID_\beta \in S^*$. If the hash values of ID_β and (ID_β, KW_α) ($KW_\alpha \in \{KW_0^*, KW_1^*\}$, $ID_\beta \in S^*$) are queried in phase 1, challenger $\mathcal{C}$ will abort. If the game is not aborted, the challenger $\mathcal{C}$ will choose a random $\gamma^* \in \{0, 1\}$ and $C_1^* \in \{0, 1\}$. Compute $p^*=(v_1, \ldots, v_{km})^T$, $C_2^*=v_0+C_1^* \lfloor q/2 \rfloor$, $A_S^*=[A_1^* \| \ldots \| A_{k^*}^*]$. Then $\mathcal{C}$ gives the attacker $\mathcal{A}$ a challenge ciphertext $CT^*=(p^*, C_1^*, C_2^*, A_S^*)$

The constructed ciphertext CT^* is a valid challenge one due to the following reason. $A_i^*=(u_{(i-1)\times m+1}, \ldots, u_{i\times m}) \in (Z_q^n)^m$ is used to simulate $A_i \cdot H(ID_{\theta_i}^* \| KW_{\gamma^*})^{-1}$. Set $x_1=(\tau_1, \ldots, \tau_{k^*m}) \in (Z_q^n)^{k^*m}$. Since $A_S^*=[A_1^* \| \ldots \| A_{k^*}^*]$ and $v_j \in Z_q$ are randomly selected or equal to $u_j^T s+\tau_j$ for $1 \leq j \leq 3km$, then we can deduce p^* is either a randomly selected element or $p^*=(A_S^*)^T s+x_1$. Set $x_2=\tau_0$. Since $V=u_0$ and v_0 is randomly selected or all equal to $u_0^T s+\tau_0$, then we have C_2^* is either a randomly selected element or $C_2^*=V^T s+x_2+C_1^* \cdot \lfloor q/2 \rfloor$.

Phase 2: Adversary $\mathcal{A}$ continues to issue the queries as in phase 1 with the constraint that The secret key of ID_β should not be queried, in which $ID_\beta \in S^*$. The trapdoor and hash of (KW_α, ID_β) should not be queried, in which $KW_\alpha \in \{KW_0^*, KW_1^*\}$ and $ID_\beta \in S^*$.

Guess: Finally, the adversary $\mathcal{A}$ outputs a guess $\gamma' \in \{0, 1\}$, $\mathcal{C}$ returns genuine if $\gamma'=\gamma^*$ or random if $\gamma' \neq \gamma^*$ as its answer for the LWE problem.

Probability Analysis: Let E_1 be the event that the hash of ID_β is queried in phase 1, in which $ID_\beta \in S^*$. Let E_2 be the event that the hash of (ID_β, KW_α) is queried in phase 1, in which $KW_\alpha \in \{ KW_0^*, KW_1^* \}$, $ID_\beta \in S^*$. Then,

$$Pr(abort) = Pr(E_1) \cdot Pr(E_2) = \frac{k^*}{2^{l_1}} \cdot \frac{2k^*}{2^{l_1+l_2}} \le \frac{2N^2}{2^{l_1(l_1+l_2)}},$$

$$Pr(\overline{abort}) = 1 - Pr(abort) \ge 1 - \frac{2N^2}{2^{l_1(l_1+l_2)}}.$$

In the guess phase, if the adversary $\mathcal{A}$ is capable to break the scheme with advantage ε, $A_{s,\chi}$ will appear with probability $1/2+\varepsilon$ at least. Then, the advantage of $\mathcal{C}$ against the LWE problemn is at least $\varepsilon' \ge \varepsilon(1-2N^2/2^{l_1(l_1+l_2)})$.

The distribution of the challenge ciphertext is statistically close to the real security scenario since p^*, C_2^*, A_S^* are all constructed using the LWE instances. Then, the challenge ciphertext CT^* will have the same distribution as in the LWE game when the LWE instances are chosen to be genuine. When LWE instances are randomly selected elements, so will be the elements in CT^*. From the proof above, it can be seen that if adversary $\mathcal{A}$ is able to break this scheme, then $\mathcal{A}$ can also break the LWE problem. In this proof, the ciphertext C_1^* is denoted as a single bit. It can be easily extended to the multi-bit scenario, which is omitted here.

6. CONCLUSIONS

In this paper, a novel searchable broadcast encryption scheme is proposed, which is also the first primitive of this kind that is built on lattice assumption. Lattice basis delegation methods and pre-image sample function (PSF) are used in the construction. It achieves constant communication and computation overhead. At the same time, it enables multiple users to search on the encrypted data using their own private key. Based on the hardness of the decisional LWE problem, we prove that the proposed scheme is secure against the chosen keyword attacks (CKA).

REFERENCES

1. D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search," in *Proceedings of Eurocrypt Annual International Conference on the Theory and Applications of Cryptographic Techniques*, 2004, pp. 506-522.
2. B. R. Waters, D. Balfanz, G. Durfee, and D. K. Smetters, "Building an encrypted and searchable audit log," in *Proceedings of the 11th Annual Network and Distributed System Security Symposium*, 2004, pp.1-10.
3. G. Crescenzo and V. Saraswat, "Public key encryption with searchable keywords based on Jacobi symbols," in *Proceedings of International Conference on Cryptology*, 2007, pp. 282-296.

4. C. Hu, P. He, and P. Liu, "Public key encryption with multi-keyword search," in *Proceedings of International Conference on Network Computing and Information Security*, 2012, pp. 568-576.
5. N. Attrapadung, J. Furukawa, and H. Imai, "Forward-secure and searchable broadcast encryption with short ciphertexts and private keys," in *Proceedings of Annual International Conference on the Theory and Application of Cryptology and Information Security*, 2006, pp. 161-177.
6. M. Abdalla, D. Catalano, and D. Fiore, "Verifiable random functions: Relations to identity-based key encapsulation and new constructions," *Journal of Crypto*, Vol. 27, 2013, pp. 544-593.
7. F. Han, J. Qin, H. Zhao, and J. Hu, "A general transformation from KP-ABE to searchable encryption," *Future Generation Computer Systems*, Vol. 30, 2014 pp. 107-115.
8. B. Wang, S. Yu, W. Lou, and T. Hou, "Privacy-preserving multi-keyword fuzzy search over encrypted data in the cloud," in *Proceedings of IEEE Conference on Computer Communications*, 2014, pp. 2112-2120.
9. R. Li, Z. Xu, W. Kang, K. C. Yow, and C. Z. Xu, "Efficient multi-keyword ranked query over encrypted data in cloud computing," *Future Generation Computer Systems*, Vol. 30, 2014, pp. 179-190.
10. C. Delerablee, "Identity-based broadcast encryption with constant size ciphertexts and private keys," in *Proceedings of Asiacrypt*, 2007, pp. 200-215.
11. R. Sakai and J. Furukawa, "Identity-based broadcast encryption, Cryptology ePrint Archive," Report 2007/217, http://eprint.iacr.org.
12. L. Zhang, Y. Hu, and N. Mu, "An identity-based broadcast encryption protocol for ad hoc networks," in *Proceedings of the 9th International Conference for Young Computer Scientists*, 2008, pp. 1619-1623.
13. P. W. Shor, "Polynomial-time algorithm for prime factorizeation and discrete logarithm on a quantum computer," *SIAM Journal on Computing*, 1997, Vol. 26, pp. 1484-1509.
14. O. Regev, "On lattices, learning with errors, random linear codes, and cryptography," in *Proceedings of ACM Symposium on Theory of Computing*, 2005, pp. 84-93.
15. C. Gentry, S. Halevi, and V. Vaikuntanathan, "A simple BGN-type cryptosystem from LWE," in *Proceedings of Eurocrypt Annual International Conference on the Theory and Applications of Cryptographic Techniques*, 2010, pp. 506-522.
16. S. Goldwasser, Y. Kalai, C. Peikert, and V. Vaikuntanathan, "Robustness of the learning with errors assumption," in *Proceedings of Innovations in Computer Science*, 2010, pp. 230-240.
17. S. D. Gordon, J. J. Katz, and V. Vaikuntanathan, "A group signature scheme from lattice assumptions," in *Proceedings of AsiaCrypt Annual International Conference on the Theory and Application of Cryptology and Information Security*, 2010, pp. 395-412.
18. W. Jin and J. Bi, "Lattice-based identity-based broadcast encryption," Cryptology ePrint Archive, Report 2010/288, 2010.
19. S. Agrawal , D. Boneh, and X. Boyen, "Lattice basis delegation in fixed dimension and shorter-ciphtertext hierarchical IBE," in *Proceedings of International Cryptology Conference*, 2010, pp. 98-115.

20. D. Cash, D. Hofheinz, and E. Kiltz, "BonsaiTrees, or How to delegate a lattice basis," in *Proceedings of Eurocrypt Annual International Conference on the Theory and Applications of Cryptographic Techniques*, 2010, pp. 523-552.
21. C. Gentry, C. Peikert, and V. Vaikuntanathan, "Trapdoors for hard lattices and new cryptographic constructions," in *Proceedings of ACM Symposium on Theory of Computing*, 2008, pp. 197-206.
22. J. Zhang, B. Deng, and X. Li, "Learning with error based searchable encryption scheme," *Journal of Electronics* (*China*), Vol. 29, 2012, pp. 473-476.
23. C. Gu, Y. Guang, Y. Zhu, and Y. Zheng, "Public key encryption with keyword search from lattices," *International Journal of Information Technology*, Vol. 19, 2013, pp. 1-10.
24. C. Hou, F. Liu, H. Bai, and L. Ren, "Public key encryption with keyword search from lattice," in *Proceedings of IEEE International Conference on P2P*, *Parallel*, *Grid*, *Cloud and Internet Computing*, 2013, pp. 336-339.
25. J. Alwen and C. Peikert, "Generating shorter bases for hard random lattices," *Theory of Computing Systems*, Vol. 48, 2011, pp. 535-553.

Yang Yang (杨旸**)** is an Associate Professor in the College of Mathematics and Computer Science at Fuzhou University, China. She received her Ph.D. degree from the Xidian University in 2011. Her research interests are in the area of information security and privacy protection.

Shu-Lve Yang (杨书略**)** is an M.S. candidate in the College of Physics and Information Engineering at Fuzhou University. His current research interests include information security and searchable encryption.

Feng-He Wang (王凤和) is an Associate Professor of Shandong Jianzhu University. He received the Ph.D. degree in Xidian University. His research interests are in the areas of public key cryptography, information security.

Jin Sun (孙瑾) is an Associate Professor in the Department of Mathematical Science of Xi'an University of Technology. She received her Ph.D. degree from the Xidian University in 2012. Her current research interests include network security, cryptography and the designs for broadcast encryption scheme.