R-PEKS: RBAC enabled PEKS for secure access of cloud data

R-PEKS: RBAC Enabled PEKS for Secure Access of Cloud Data

K. Rajesh Rao, Indranil Ghosh Ray, Waqar Asif, Ashalatha Nayak and Muttukrishnan Rajarajan, *Senior Member, IEEE*

Abstract—In recent past, few works have been done to achieve multi-user searchable encryption (MUSE) in the context of maintaining personal health record (PHR). This was achieved by combining access control with PEKS, i.e., public key encryption with keyword search. In such applications, changing of permissions are done once in a while. However, till date, no efficient and secure scheme is available in the literature that are suitable for these applications where changing of user permissions are done frequently. In this paper our contributions are twofold. Firstly, we propose a new PEKS scheme for string search, which, unlike the previous constructions, is free from bi-linear pairing and is efficient by 97% compared to PEKS for string search proposed by Ray et.al in TrustCom 2017. Secondly, we introduce role based access control (RBAC) to multi-user PEKS, where an arbitrary group of users can search and access the encrypted files depending upon roles. We termed this integrated scheme as R-PEKS. The efficiency of R-PEKS over PEKS scheme is up to 90%. We provide formal security proofs for the different components of R-PEKS and validate these schemes using commercial data set.

Index Terms—RBAC, PEKS, MUSE, SUSE, Cloud computing

1 Introduction

In cloud, encryption may be a suitable mechanism to protect the data at rest. However, encryption prevents searching within the data which is essential for better usability of the encrypted data. This gives rise to a new area of research, called searchable encryption (SE). One problem of using SE in symmetric setting is the maintenance of index, specially for applications where dataset undergoes frequent update. A variant of SE, called *public key encryption with keyword search* or PEKS is the most popular encryption technique which was introduced in [1] and is free from any index generation. Following this, many researches has been carried out on *single-user searchable encryption* (SUSE) with *access control* mechanisms. However, *multi-user searchable encryption* (MUSE) is becoming more relevant for most of the commercial applications involving large group of users with complex access structure. Some work has also been done on MUSE by delegating the permission of searching among multiple users in an access controlled environment. Most of these works involve attribute based access structure.

SE for keyword search yields huge outputs for most of the commercial applications which deal with large detaset. Most of these outputs are not intended and gives rise to unnecessary network traffic. SE for *string search* is of special interest as this customizes the search. In [2], non-adaptively secure SE for string search was proposed, but the approach was in a symmetric key setting. In [2] authors pointed why it is impossible to design adaptively secure SE in symmetric search. In [3], authors introduced the idea of *string search* using PEKS which is adaptively secure. All existing PEKS schemes are based on pairing based cryptography where the basic component is *bi-linear mapping*. One problem with this technique, specially in the context of string search, is the huge computational cost which makes it less suitable for commercial applications [4]. In this paper, we introduce a new adaptively secure PEKS scheme, which is free from *bi-linear mapping* and compatible with role based access control (RBAC) mechanism for secure search and access of outsourced encrypted data.

Recently, most of the researchers applied SE with *attribute based access control* (ABAC) to provide restricted access based on the attributes for their outsourced personal health record (PHR). In PHR datasets, changing users permissions are minimal [5], [6]. Since ABAC is the most suitable approach for these cases where changing user permissions is minimal, protecting PHR data in cloud by the means of attribute based keyword search over encrypted data is the most preferred choice of researchers [4], [7]–[9].

In big organizations, large number of employees access data under a complex access structure. MUSE finds its application in such cases. If the search needs to be customized using string search, PEKS is the best possible solution for that. Since access structure is defined based on the roles of large number of employees, which are subject to frequent modifications, RBAC is the considered most preferred access control mechanism [6]. For example, securing wireless networks with RBAC in small and medium sized businesses.

- *K. Rajesh Rao is with the Department of Information and Communication Technology, Manipal Institute of Technology, MAHE, Karnataka, INDIA. E-mail: rajesh.kavoor@manipal.edu*
- *Indranil Ghosh Ray is with the Department of Electrical and Computer Engineering, City, University of London, London, ENGLAND. E-mail: Indranil.Ghosh-Ray@city.ac.uk*
- *Waqar Asif is with the Department of Electrical and Computer Engineering, City, University of London, London, ENGLAND. Email: Waqar.Asif@city.ac.uk*
- *Ashalatha Nayak is with the Department of Computer Science and Engineering, Manipal Institute of Technology, MAHE, Karnataka, INDIA. E-mail: asha.nayak@manipal.edu*
- *Muttukrishnan Rajarajan is with the Department of Electrical and Computer Engineering, City, University of London, London, ENGLAND. E-mail: r.muttukrishnan@city.ac.uk*

Here the ad-hoc process of granting and revoking network privileges and access to users becomes extremely difficult to manage, as the number of individuals involved increases beyond a certain point. Thus a combination of RBAC with PEKS is an optimal solution to design access controlled MUSE for frequently changing string search privileges. Additionally, RBAC has a minimal overhead when large number of employees enter and exit the organization. Here we describe an industrial application where our model can provide the most optimal solution, but due to the confidentiality agreement the company's identity is not disclosed.

As per the PCI DSS (Payment Card Industry Data Security Standard), the credit/debit card information should not be stored in plain text. So companies use hashing technique (SHA-256) to store the card information in the database. So during the transaction whenever the user enter his card details, based on the hash value of the card details the information such as name and address are auto filled.

It may be noted that the use of hashing technique is also exposed to the risk of statistical attack. In our model, even if the adversary get the trapdoor he cannot search unless he gets the access of search algorithm. Migrating from hashing to PEKS comes with the additional cost but to make searching process efficient, we introduce role based PEKS, i.e., R-PEKS for such situations. To the best of our knowledge, no secure searching scheme is available using PEKS with RBAC.

In this paper, we consider the following three real world scenarios where PEKS, integrated with RBAC, is a good choice for commercial applications. We coin the term **R-PEKS** for it. We validate and compare our model with existing schemes using the methodologies proposed in [10].

Scenario 1 *An organization wants to outsource its large amount of data to the cloud service provider (CSP). Such data can be accessed by the set of employees who are privileged by their roles to search.*

Scenario 2 *When an employee P wants to access certain data of some employee Q who is working under him, then the privilege is given to P to access Q's data.*

Scenario 3 *When a group G is formed from a subset of employees who are having different roles, then the group level privilege is given to all members for accessing data pertaining to the group.*

Most of the threat models assume that the data owner and data user are trusted. However, cloud service provider (CSP) is considered as *honest-but-curious* [2], [3], [11]. Recently, in MUSE, the threat model was developed by considering that data users are colluding with the CSP where both are *honest-but-curious* [12]. We develop R-PEKS under this threat model.

Main contributions of this paper are as follows:

1) Our proposed model is at the intersection of RBAC and PEKS, which facilitates data owner to provide restricted data search and access based on the RBAC configuration. We implement the access control in three different modes depending on the above scenarios, namely, *single user* (Scenario 1), *multi-user peer to peer* (Scenario 2) and *multi-user group* (Scenario 3).
2) We introduce a new PEKS scheme which, unlike the previous schemes, is free from bi-linear mapping and is more practical and efficient compared to earlier scheme of [3] by 97% but yet providing equivalent level of security.
3) With normal PEKS, the search request of an user u is spread over the whole dataset of the organization, of which a limited portion is meant for the access of the user u. In such PEKS, the access control is handled manually after the search. In this context, R-PEKS enforces the access control during the search and the efficiency of R-PEKS over PEKS scheme is up to 90% when the part of dataset pertaining to user u is 10% of the whole cloud data.

Rest of the paper is organized as follows: In Section 2 we discuss the related work. Section 3 highlights the architecture of our proposed model R-PEKS. Section 4 defines the R-PEKS components and Section 5 describes the R-PEKS design. In Section 6 we discuss the security analysis of different components of R-PEKS. In Section 7 we provide experimental results. In Section 8 we focus on comparison with other schemes and Section 9 concludes the paper and outlines areas for future work.

2 Related work

Access control mechanisms with SUSE scheme provides restricted access to data based on *roles*, *keys* and *attributes*. In role based access, i.e., role based encryption of [13], the security for cloud storage is enabled by the data owner where he encrypts a data using some role and public parameters. In case of key based access control of [14], the accessible file's decryption key is given to the users. Integrating such model with SE increases the complexity of the key management whenever user accesses large number of files. In [8], [9], *multi-field* search query and *fine-grained access control* was implemented in SE with ABAC during file access in the cloud. The file in the cloud is attached with an encrypted index to label the keywords and access policy. So the user can access the file if the keyword is matched and the access policy mechanism grants the permission. Further it allows the user to locally derive the search capability. String search can be viewed upon as ordered multiple keyword search. Few schemes were proposed for string search in symmetric searchable encryption (SSE) [2] and PEKS [3], but both these schemes are without access control mechanism.

Access control mechanisms with MUSE schemes are constructed using *broadcast encryption*, along with *attributes*, *coarse-grained* and *fine-grained* access control. The key problems identified in MUSE are the key management and the access control. Broadcast encryption [15] was the first mechanism to introduce access control in multi-user environment, where the message is encrypted by broadcaster and can be decrypted only by these users who are part of the broadcast channel. Further, based on the broadcast encryption technique, coarser-grained access control were introduced into MUSE in cloud environment in [16]. In [17], Key-Aggregate Searchable Encryption (KASE) was introduced in cloud storage for decreasing the key management during data sharing where each document was encrypted with a different key. This was implemented by generating a single key called key-aggregate in MUSE. In [18] single or multi-keyword search based on attributes and access control were implemented, which is known as *fine-grained access control*.

In [19] multi-keyword search along with authorization in the multiple user setting was studied, where the authorization was meant for a specified period of time.Work has also been done to improve the search efficiency among multiple users using SSE for cloud applications [20]. In recent past, PEKS scheme was combined with ABAC to provide restricted access on *personal health record* (PHR). In [7], authors designed PEKS integrated with ABAC to support multi-keyword search in multi-user settings. In [4], expressive SE scheme in PEKS was studied and developed for outsourced PHRs. In this work authors treated keyword search predicates as access policies and express them as conjunction or any other boolean expression of keywords. Further, in [4], *bi-linear mapping* in prime-order group was used to make it some what secure compared to *bi-linear mapping* in composite-order group. It may be noted that ABAC in PEKS is not a suitable mechanism when access policies are changing quite often. This is because, with every change in the policies, the data owner needs to download, decrypt and re-encrypt the data [5]. In [12], data users colluding with the CSP is considered as a mandatory requirement of MUSE. Moreover in SE, most of the threat model focused only on privacy against honest-but-curious CSP and data user [12], [16].

In the light of above discussions, it is clear that an access control mechanism in SE is very much essential to maintain large number of privileged access. Such privileged access in SE should be conformable for future analysis and change in an efficient way. In this paper, we propose R-PEKS based on RBAC and a new PEKS. Unlike the previous ABAC based schemes, R-PEKS is more suitable for applications where change of user permissions is a frequent activity. Also the underlying PEKS in R-PEKS is free from *bi-linear mapping* and is thus more efficient compared to earlier schemes. In Table 1, we provide a comparison of our scheme with the existing models. It may be noted that our PEKS security relies on CDH assumption (Computational Diffie-Hellman) whereas in [3], [4] the security assumption were based on BDH assumption (Bi-linear Diffie-Hellman). CDH is a very popular hard problem and many state-of-the-art schemes are based on CDH assumptions. For example, authors in [21], [22] also used CDH as a basis of their security proof.

TABLE 1
Properties of different searchable encryption schemes.

Property	KSAC [8]	SSE [2]	PEKS [3]	Expressive keyword search [4]	This paper
Search	multi keyword	string	string	multi keyword	string and multi keyword
SE scheme	SSE	SSE	PEKS	PEKS	PEKS
Mathematical tool	HPE	index using secure hash chain	bi-linear mapping in elliptic curve group	bi-linear pairing in prime order group	exponentiation on prime order multiplicative group
Security assumption	RDSP and IDSP	secured index based on AES and SHA-256	BDH	BDH	CDH
SE setting	single user	single user	single user	multi user	single and multi-user (peer to peer and group)
Access control	ABAC	no	no	ABAC	RBAC
Security	keyword privacy	search pattern privacy	keyword privacy	keyword privacy	keyword privacy and hosted data confidentiality

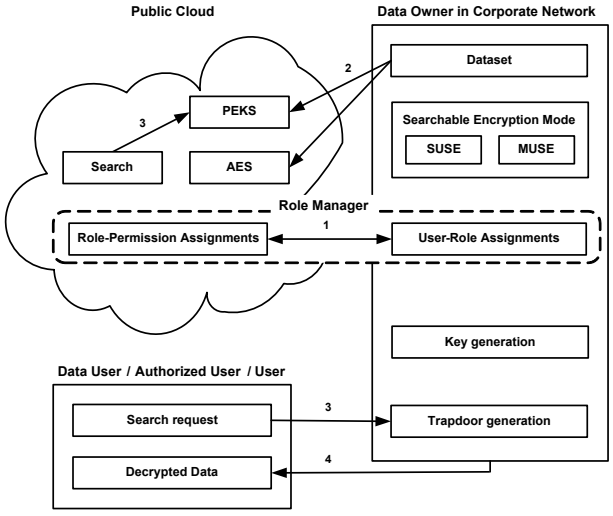

Fig. 1. Design of the R-PEKS settings for secure access of cloud data

3 Architecture of our Proposed Model - R-PEKS

The design of single-user and multi-user R-PEKS settings for secure access of cloud data is given in Figure 1. The **data owner** of Figure 1 is the enterprise which can grant or deny certain set of permissions to its users. Here *encryption* and *searching* are the two operations which constitutes a *permission*. Therefore to maintain the restricted access on such permissions, the data owner maintains the role manager. In Figure 1, the *searchable encryption mode* component enables SUSE and MUSE as two different modes for searching the string request based on the privileges in the PEKS domain. Key pairs are generated for each authorized user in *Key generation* component. A **data user** may be an individual or group with limited data scope who can request for string search. **Public cloud** is responsible to store the encrypted data and search. For the security reason, the roles that are assigned to users as well as permissions are maintained across the **corporate network** and **public cloud** respectively.

The detailed description of the activities are given below:

1) Initially, data owner generates RBAC configuration, key pair for all the authorized users and enables two different modes of searchable encryption, i.e., SUSE and MUSE.
2) Data owner encrypts using PEKS and AES (see Figure 1) and stores the dataset using single-user and multi-user R-PEKS in public cloud.
3) The trapdoor is generated by the data owner on receiving the string search request and *mode* of R-PEKS from the data user. Such trapdoors are used in the public cloud to search and identify the required files.
4) Finally, public cloud return the id's of the files that matches the search to the data owner, which then

decrypts the corresponding AES-encrypted files using AES-decryption for the data user.

Remark 1. It may be noted that PEKS encryption is an one way function as searchable ciphers produced by PEKS encryption can not be decrypted. It is used only for searching using the trapdoor. In the R-PEKS, to retrieve the plain text file, we encrypt the plain text files using PEKS as well as AES and maintain a map between these two encrypted file pointers. Whenever some PEKS encrypted file pointer is detected by the search algorithm, the corresponding AES encrypted file is sent to the data owner for decryption and retrieval of the plain text file.

Remark 2. We maintain a map since AES encryption and decryption components of R-PEKS are straight forward, we skip this component while describing R-PEKS in the following sections.

4 R-PEKS Components

4.1 PEKS: Definitions and Preliminaries

Document collections and Data Structures: Let $\triangle = \{w_1, w_2, \ldots, w_d\}$ be a dictionary of d words and $\mathbb{P}(\triangle)$ be the set of all possible documents which are collections of words. Let $D \subseteq \mathbb{P}(\triangle)$ be the collection of n documents $D = (D_1, D_2, \ldots, D_n)$. Let $id(D_i)$ be the unique identifier for the document D_i. We denote the list of all n document identifiers in D by $id(D)$, i.e., $id(D) = \{id(D_1), \ldots, id(D_n)\}$. Furthermore, let $D(w_j)$ be the collection of all documents in D containing the word w_j. A string s of l words is an ordered tuple $(w_1, w_2, \ldots, w_l)$. Let $D(s)$ denote a collection of documents in D that contains the string s. It is easy to check that $D(s) \subseteq \bigcup_{i=1}^{l} D(w_i)$. We denote by $\delta(D)$, all the distinct keywords connected to the document collection D.

Cryptographic Primitives: Here we define cryptographic primitives that are needed for our PEKS scheme for string search.

We typically denote an arbitrary negligible function by $negl$ such that for any arbitrary polynomial $p(.)$, there exists an integer a such that for all $\lambda > a$, $negl(\lambda) < \frac{1}{p(\lambda)}$ [23].

We also use *pseudo prime number generator* [24], denoted by $PPNG(1^\lambda)$ which outputs a λ-bit probabilistic prime number. For a finite set S, we denote the operation of picking an element uniformly at random from S by $x \leftarrow S$.

Public Key Encryption with Searching:

Definition 1. A non-interactive public key encryption with keyword search scheme consists of the following polynomial time randomized algorithms:

1. KeyGen(λ): This algorithm takes security parameter, λ, and generates a public/private key pair A_{pub}, A_{priv}.
2. PEKS_Enc(A_{pub}, w): For a public key A_{pub}, and a word w, this algorithm produces a searchable encryption of w.
3. Trapdoor($A_{priv,w}$): Given a private key A_{priv} and a word w, this algorithm produces a trapdoor t_w.
4. Search(A_{pub}, C, t_w): Given the public key A_{pub} and searchable encryption $C = PEKS\_Enc(A_{pub}, w')$ and a trapdoor t_w, this algorithm outputs 'yes' if $w = w'$, otherwise 'no'.

Two cryptographic hash functions are used in our scheme Π_{PEKS}, namely H_1 and H_2 which are as follows: $H_1 : \{0,1\}^* \times \{0,1\}^* \mapsto \mathbb{Z}_p$, $H_2 : \{0,1\}^* \times \mathbb{Z}_p \mapsto \{0,1\}^\lambda$. For our implementation we take G as $\mathbb{Z}_p$. Thus, for a word $w \in \{0,1\}^*$, $H_{1\langle k_2, k_1\rangle}(w) = H_{1\langle k_2\rangle}\left(H_{1\langle k_1\rangle}(w)\right) \in \mathbb{Z}_p$.

4.2 Our PEKS scheme

In this section we present our PEKS scheme Π_{PEKS} for string search.

Scheme 1 **(Π_{PEKS}).** The scheme Π_{PEKS} is a collection of four polynomial time algorithms (KeyGen, PEKS_Enc, Trapdoor, Search) such that:

1. KeyGen(1^λ) : KeyGen is a probabilistic key generation algorithm that is run by the data owner to setup the scheme. It takes a security parameter λ, and returns the setup. Since KeyGen is randomized, we write it as $(p, \mathbb{Z}_p, a, k_1, k_2) \leftarrow$ KeyGen(1^λ). The public key is $pk = (k_2)$ and the secret key is $sk = \{a, k_1\}$ where $a \leftarrow \mathbb{Z}_p$.
2. PEKS_Enc(k_1, k_2, a, D_i) : PEKS_Enc is a probabilistic algorithm run by the data owner to generate the cipher text C_i corresponding to document D_i. Since PEKS_Enc is deterministic, we write this as $C_i = PEKS\_Enc(k_1, k_2, a, D_i)$.
3. Trapdoor(a, s) : Trapdoor is a deterministic algorithm run by the data owner to generate a trapdoor for a given string of words $s = (w_1, w_2, \ldots, w_l)$. It takes k_1, k_2, a and s as input and outputs $t = (t_1, t_2, \ldots, t_l)$, where t_i is the trapdoor corresponding to the word w_i. Since trapdoor is deterministic, we write this as $t = Trapdoor(k_1, k_2, a, s)$.
4. Search(C, t, k_2) : Search is run by the server in order to search for the documents in D that contain the string s. It takes a ciphertext collection C corresponding to D and the trapdoor t corresponding to s and returns $D(s)$, the set of identifiers of documents containing the string s.

The Construction: Now we provide the actual algorithms for key generation (Algorithm 1), PEKS encryption (Algorithm 2), trapdoor generation (Algorithm 3) and searching (Algorithm 4).

Algorithm 1: KeyGen

Require: security parameter λ.
Ensure: $\mathbb{Z}_p$, a, k_1 and k_2.
$(p, \mathbb{Z}_p, a, k_1, k_2) \leftarrow KeyGen(1^\lambda)$;
$p \leftarrow PPNG(1^\lambda)$;

Remark 3. To encrypt the document D_i, we read the whole document as stream of words and form a ordered sequence$(w_1, w_2, \ldots, w_l)$. A string is a subsequence of this sequence which is encrypted by applying PEKS_Enc on every word in the in the string. If the same word occurs multiple times, for each instances the encrypted footprint will be different depending on the position (denoted by k in the algorithm) of the word in the string.

Algorithm 2: PEKS_Enc

Require: g, a, k_1, k_2, p and $D = (D_1, \ldots, D_n)$.
Ensure: $C = (c_1, \ldots, c_n)$.
Form a collection $W = \{w_1, \ldots, w_d\}$ of all distinct words occurring in D;
$i \leftarrow 1$;
while $i \leq n$ **do**
open D_i in read mode and C_i in write mode;
$j \leftarrow 1$;
$wp \leftarrow 0$;
while (!EOF) do
read j-th string in $s_j = (w_1, \ldots, w_m)$;
$k \leftarrow 1$;
while $k \leq m$ **do**
$A = H_{1\langle k_2, k_1\rangle}(w_k)$;
$B = H_{2\langle k_2\rangle}(A^{a.(k+wp)})$;
Append B in C_i;
$k \leftarrow k + 1$;
$wp \leftarrow wp + 1$;
end while
$j \leftarrow j + 1$;
end while
$i \leftarrow i + 1$;
end while

Algorithm 3: Trapdoor

Require: $w = (w_1, w_2, \ldots, w_l)$, k_s, a, p.
Ensure: $t = (t_1, \ldots, t_l)$.
$j \leftarrow 1$;
while $j \leq l$ **do**
$A_j = H_{1\langle k_2, k_1\rangle}(w_j)$;
$t_j = A_j^a$;
$j \leftarrow j + 1$;
end while

Algorithm 4: Search

Require: $t = (t_1, \ldots, t_l)$, $\{C_1 \ldots, C_n\}$, k_2.
Ensure: $encrypted\_file\_pointers$, a list of encrypted document pointers;
$exists = false$; $counter \leftarrow 0$;
$i \leftarrow 1$;
while $i \leq n$ **do**
read all blocks of the form $[B]$ from C_i and arrange B's in $listB$;
$start\_pointer = -1$;
$m \leftarrow 1$;
while ($m \leq listB.length$) **do**
compute $chk = H_{2\langle k_2\rangle}(t_1^m)$);
if ($chk == listB[m]$) **then**
$start\_pointer = m$;
$exists = true$;
$counter = 1$;
break;
else
$m = m + 1$;
end if
if ($counter = 1$) **then**
$k \leftarrow 2$;
while ($k \leq l$) **do**
if $(H_{2\langle k_2\rangle}(t_k^{start\_pointer+counter}) == listB[start\_pointer + counter])$ **then**
$counter \leftarrow counter + 1$;
else
$exists = false$;
break;
end if
$k \leftarrow k + 1$;
end while
end if
if ($exists = true$) **then**
add i to $encrypted\_file\_pointers$
end if
end while
end while

Remark 4. To search in a cipher, we first read all the ciphers of the form $[B]$ in a list called $listB$. To find the match of the first trapdoor, i.e., t_1, we check it against each entry of $listB$. This is done in the second while loop. If match is found, then the index of that block is stored in $start\_pointer$ and rest of $l - 1$ trapdoors are checked against next $l - 1$ blocks in $listB$ starting from $start\_pointer + 1$. If match is found in all l successive steps, we add the file pointer in $encrypted\_file\_pointer$ and go for the next file. If the match fails in any step we repeat the matching of t_1 for the remaining blocks in the same way until the file is exhausted.

Remark 5. In the PEKS, we assume that data owner creates the data, encrypts using PEKS and uploads to public cloud for future search. So the KeyGen and PEKS_Enc are run by data owner. To search, the query is converted to trapdoor by data owner and is given to cloud. Cloud runs the search using these trapdoors along with the public key of data owner. In R-PEKS, data owner is the data administrator who is responsible for creating employee accounts and provide employees role based permissions to search on the encrypted data stored in server. So, in R-PEKS, data owner runs KeyGen for each employees/users and provide them with their keys. Permissions are nothing but files which are connected to roles (see Figure 2). Depending on the roles, data owner fetch the keys of the corresponding employee/user and encrypt the files. To search, employees request data owner for the legitimate trapdoor for a given search query. Search is run in server using the trapdoor and the public of the corresponding data and the search results are sent to data owner who further decrypts and redirect the files to the employee based on the permissions.

In the next lemma, we study the correctness of the search algorithm.

Lemma 1 (Correctness). Let $s = (w_1, \ldots, w_l)$ be a string such that s is in the document D_i in the document collection D and $C_i \leftarrow PEKS\_Enc(k_1, k_2, a, D_i)$. Let $[B_1], \ldots, [B_l]$ are the encrypted blocks in C_i for the string s taken in order. Then Search(C, t, k_2) will point out the identifier corresponding to D_i where $t = (t_1 \ldots, t_l)$

is the trapdoor corresponding to s taken in order, i.e., $id(D_i) \in Search(C, t, k_2)$.

Proof: Let $A_1 = H_{1\langle k_2,k_1\rangle}(w_1)$. for some j. Note that, then $B_1 = H_{2\langle k_2\rangle}(A_1^{am})$ for some integer m depending on the position of the word w_1 in the file. Now,

$$\begin{aligned} H_{2\langle k_2\rangle}(t_1^m)) &= H_{2\langle k_2\rangle}(H_{1\langle k_2,k_1\rangle}(w_1)^a)^m) \\ &= H_{2\langle k_2\rangle}(A_1^{am}) \\ &= B_1 \end{aligned}$$

This detects $[B_1]$ for t_1. Two consecutive blocks B_j and B_{j+1} are detected for two consecutive words w_j and w_{j+1} if there are two consecutive numbers, say k and $k+1$, such that $B_{j+1} = H_{2\langle k_2\rangle}(A_{j+1}^{a(k+1)})$, and $B_j = H_{2\langle k_2\rangle}(A_j^{a(k)})$, where $A_{j+1} = H_{1\langle k_2,k_1\rangle}(w_{j+1})$ and $A_j = H_{1\langle k_2,k_1\rangle}(w_j)$. Hence the result follows. □

Remark 6. It may be noted that the second while loop of string search algorithm (i.e., Algorithm 4) is responsible for maintaining the ordering of keywords which is essential to detect string. We have also achieved multi-keyword search by relaxing this condition for ordering. For the experimental results we have used the variant which is dealing with string search only, i.e., Algorithm 4.

Therefore our PEKS scheme Π_{PEKS} can search string and multi keyword based on the user's request. In order to perform a selective search we integrated our PEKS scheme Π_{PEKS} with RBAC model (i.e., R-PEKS). The detailed explanation about RBAC components and function which are crucial for integrating with our PEKS scheme Π_{PEKS} are given in the Section 4.3.

4.3 RBAC model

The NIST RBAC reference model defines different RBAC elements, RBAC assignments and mapping functions [25]. The pictorial representation of NIST RBAC model is shown in Figure 2. Let us define some of the elements, assignments

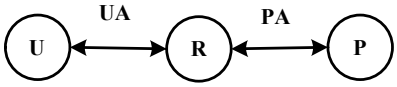

Fig. 2. NIST RBAC [25]

and functions which are crucial for R-PEKS. Let U denote a set of users. Let R denote a set of roles where each role is a job assigned to some user in an organization. Let P denote a set of permissions where each permission is an approval to perform an operation on an object. Formally we express this by $P = 2^{OP \times OB}$, where OP is the set of operations and OB is the set of objects. Let S_u denote a set of users who authorize the user u to search and access data pertaining to them. Let G denote a set of authorized users, i.e., a group. Let **UA** denote a many-to-many mapping which is user-to-role assignment relation, i.e., **UA** $\subseteq U \times R$. Let **PA** denote a many-to-many mapping which is a permission-to-role assignment relation, i.e., **PA** $\subseteq P \times R$. Also from [25], **assigned_permissions:** $R \rightarrow 2^P$ is a mapping function from role to a set of permissions. So, for some $r \in R$, **assigned_permissions(r)** = { p ∈ P | (p, r) ∈ **PA** }.

RBAC Configuration. Let **UP** denote a many-to-many mapping of user-to-permission assignment relation i.e., **UP** $\subseteq U \times P$. Also let us consider the function **RoleMining** which on input **UP**, outputs **UA** and **PA**, i.e., **RoleMining**: $U \times P \rightarrow$ { **UA**, **PA** }. The access control on RBAC can be defined as a *CheckAccess* function which is responsible for authorization process. The function is defined as *CheckAccess*: $U \times P \rightarrow P$. *CheckAccess* takes an user u and the set of all permissions and returns legitimate set of permissions pertaining to user u through roles. Formally, $CheckAccess(u, P) = \{p : p \in P \wedge \forall r, (u, r) \in \mathbf{UA} \wedge (p, r) \in \mathbf{PA}\}$. Similarly, we define *CheckGroupAccess*: $G \times P \rightarrow P$, where G is a group of users who may have different roles. Through *CheckGroupAccess* the permission is granted for all members of the group depending on role assigned to the group.

5 R-PEKS Design

In this section we present our R-PEKS for single-user and multi-user settings. *Those components of Π_{PEKS}, that are reused in R-PEKS schemes as it is, are not described here. For these components of Π_{PEKS}, which are modified with access control functionality, we use the meta character "*" for the same set of parameters used in the corresponding component in Π_{PEKS}. Also, we modify the name of such components by prefixing the original name used in Π_{PEKS} by R.*

5.1 R-PEKS scheme

Single-user R-PEKS scheme is a SUSE in R-PEKS, where user u is authorized to search in the encrypted domain restricted by the assigned role. Further the construction is given below.

Scheme 2. A **single-user R-PEKS scheme**, Π_{suse}, is a collection of five polynomial-time algorithms (KeyGen, R-PEKS_Enc, Trapdoor, R-Search, *CheckAccess*) such that:

1. R-PEKS_Enc($*, CheckAccess(u, .)$): R-PEKS_Enc is a probabilistic algorithm run by the data owner to generate the cipher text C_i for D_i, if the permission set P_u, obtained from *CheckAccess*, allows access to the D_i.
2. R-Search($*, CheckAccess(u, .)$): Search is run by the CSP in order to search only for the privileged documents in D which are accessible under the permissions obtained by *CheckAccess*.

Remark 7. In single user R-PEKS scheme, the data owner performs only the selective encryption and CSP performs only the selective search based on the roles assigned to the user using $CheckAccess$.

Multi-user R-PEKS: peer to peer scheme is a MUSE in R-PEKS. Here user u is authorized to search in the encrypted domain of users in S_u. The construction is given below.

Scheme 3. A **multi-user R-PEKS: peer to peer scheme** Π_{musePP} is a collection of seven polynomial-time algorithms (KeyGen, R-PEKS_Enc, AddUser, RevokeUser, Trapdoor, R-Search, *CheckAccess*) such that:

1. AddUser(u, U): AddUser is a deterministic algorithm run by the data owner to add a new user. It is assigned with a new user $u^{'} \in U$ and updates S_u.

2. RevokeUser(u, U): RevokeUser is a deterministic algorithm run by the data owner to remove a existing user. It is revoked with the existing user $u^{'} \in U$ and updates S_u.
3. R-Search($*, CheckAccess(u, .)$): Search is run by the CSP in order to search only for the privileged documents in D which are accessible under the permissions obtained by *CheckAccess* over S_u.

Remark 8. In multi-user R-PEKS: peer to peer scheme, the requisite permission given by the data owner to search on others data is managed by AddUser and RevokeUser to add and revoke the user's privilege respectively. Data owner performs the selective encryption for users based on the assigned roles using the keys generated for the individual users. Finally, CSP performs the selective search for the authorized user who has requisite permission to search on others data.

Multi-user R-PEKS: group scheme is a MUSE in R-PEKS. Here group G is authorized to search in the encrypted domain of some users (i.e., subset of U) depending on the role assigned to G.

Scheme 4. An **multi-user R-PEKS: group scheme** Π_{museG} is a collection of seven polynomial-time algorithms (R-KeyGen, R-PEKS_Enc, Group_AddUser, Group_RevokeUser, Trapdoor, R-Search, *CheckGroupAccess*) such that:

1. R-KeyGen(1^{λ}) : R-KeyGen is a probabilistic key generation algorithm that is run by the data owner to setup the scheme for the group G.
2. R-PEKS_Enc($*, CheckGroupAccess(G, .)$): R-PEKS_Enc is a probabilistic algorithm run by the data owner to generate the cipher text C_i for D_i, if the permission set P_G, obtained from *CheckGroupAccess* allows access to the D_i.
3. Group_AddUser and Group_RevokeUser are the deterministic algorithms run by the data owner to add and remove a user from the group respectively. Since these functions are very much similar to AddUser and RevokeUser, it is not shown explicitly.
4. R-Search($*, CheckGroupAccess(G, .)$): Search is run by the CSP in order to search only for the privileged documents in D which are accessible under the permissions obtained by *CheckGroupAccess* over group G.

Remark 9. In multi-user R-PEKS: group scheme, the group which is a subset of users are created by the data owner. Key generation, group privileges and managing the group is done by the data owner. Further, data owner performs the selective encryption based on the assigned roles for the group using the generated keys. Finally, based on the request the CSP performs the selective search based on the assigned roles for the group.

Remark 10. We observe that R-PEKS is independent of the underlying PEKS, i.e., Π_{PEKS} in a sense that any PEKS system can be used to install R-PEKS. It may be noted that in R-PEKS for single user, i.e., Π_{suse}, the encryption is done using underlying PEKS encryption whenever there is a permission which is derived from role based structure given by *CheckAccess(.)* (See Scheme 2). Once the permission is given, the encryption is done using encryption algorithm of Π_{PEKS}, i.e. $PEKS\_Enc(.)$. Similar analysis holds for Search(). Thus the idea of R-PEKS can be instantiated with respect to any arbitrary PEKS system. For example, in Subsection 7.2, for the analysis purpose, RBAC is instantiated with PEKS proposed in [3] under the name *r-PEKS*. To the best of our knowledge, prior to this work, PEKS scheme of [3] was the only adaptively secure PEKS scheme for string search and so we select this scheme for comparison.

6 Security Analysis

Threat model. In R-PEKS scheme, we consider both CSP and data users as *honest-but-curious*. Under this assumption, CSP can infer additional privacy information, i.e., role-permission assignments and other informations related to *search pattern* and *access pattern* of the data [15]. Further, the malicious data users may collude with CSP to access unauthorized files by tweaking the roles. Unlike [18], the key and role management is handled by the data owner. So in our model, CSP cannot collude with any revoked malicious users in accessing the unauthorized privileges.

In the next two subsections we discuss security of R-PEKS and its components. In Subsection 6.1, we show that our PEKS is adaptively secure under CDH assumption. In Subsection 6.2, we show that our R-PEKS system ensures data confidentiality by secure access.

6.1 Security of PEKS

In this section we show that Π_{PEKS} is secure against adaptive string attack. The basic idea behind an adaptive string attack is that the adversary $\mathcal{A}$ is allowed to ask for PEKS encryptions of multiple strings chosen adaptively. The definition of security requires that $\mathcal{A}$ should not be able to distinguish the PEKS encryption of two arbitrary strings of same length, even when $\mathcal{A}$ is given access to $PEKS\_Enc()$ and $Trapdoor()$ oracle. We first define an experiment for any PEKS scheme $\pi = (KeyGen, PEKS\_Enc, Trapdoor, Search)$, any adversary $\mathcal{A}$, and any value k of the security parameter.

Definition 2. [$Game\_Adaptive\_Generic_{\mathcal{A},\pi}(1^k)$]

1. A key (pk, sk) is generated by running $KeyGen(1^k)$.
2. The adversary $\mathcal{A}$ is given input 1^k and oracle access to $PEKS\_Enc(.)$ and $Trapdoor(.)$ and outputs a pair of strings s_0 , s_1 of the same length, say m.
3. $b \leftarrow \{0, 1\}$ and then a ciphertext $c \leftarrow PEKS\_Enc(s_b)$ is computed and given to $\mathcal{A}$. We call c the challenge ciphertext.
4. The adversary $\mathcal{A}$ continues to have oracle access to $PEKS\_Enc(.)$ and $Trapdoor(.)$ and outputs a bit b'.
5. The output of the experiment is defined to be 1 if $b' = b$, and 0 otherwise. In case $Game\_Adaptive\_Generic_{\mathcal{A},\pi}(1^k) = 1$, we say that $\mathcal{A}$ succeeded.

Since our scheme deals with two cryptographically strong hash functions, namely, H_1 and H_2, we would like to

extend Definition 2 by giving oracle access of these two hash functions to adversary, which leads to the modified security definition as given below.

Definition 3. [$Game\_Adaptive_{\mathcal{A},\pi}(1^k)$]

1. A key (pk, sk) is generated by running $KeyGen(1^k)$.
2. The adversary $\mathcal{A}$ is given input 1^k and oracle access to $PEKS\_Enc(.)$, $H_1(.)$, $H_2(.)$ and $Trapdoor(.)$ and outputs a pair of strings s_0 , s_1 of the same length, say m.
3. $b \leftarrow \{0,1\}$ and then a ciphertext $c \leftarrow PEKS\_Enc(s_b)$ is computed and given to $\mathcal{A}$. We call c the challenge ciphertext.
4. The adversary $\mathcal{A}$ continues to have oracle access to $PEKS\_Enc(.)$, $H_1(.)$, $H_2(.)$ and $Trapdoor(.)$ and outputs a bit b'.
5. The output of the experiment is defined to be 1 if $b' = b$, and 0 otherwise. In case $Game\_Adaptive_{\mathcal{A},\pi}(1^k) = 1$, we say that $\mathcal{A}$ succeeded.

It is easy to observe that if a scheme is secure under the Definition 3, then it is also secure under Definition 2. This is because from the point of view of adversary, if oracle access of these two hash functions are withdrawn from Definition 3, then this becomes Definition 2. Also reducing these two oracle access amounts to reduce the degree of freedom of adversary in terms of fetching information from oracle and thus adversary can not get more information than what he could with Definition 3. We record this trivial fact in the following theorem without proof.

Theorem 1. If a PEKS scheme is secure under the Definition 3, then it is also secure under the Definition 2.

So it suffices to show that our scheme is secure under Definition 3. Consider the following definition which is based on Definition 3.

Definition 4. A PEKS scheme, denoted by $\pi = (KeyGen, PEKS\_Enc, Trapdoor, Search)$, is said to be adaptively secure under chosen plain text attack if for all probabilistic polynomial time adversaries $\mathcal{A}$, there exists a negligible function $negl$ such that $Pr[Game\_Adaptive_{\mathcal{A},\pi}(1^k) = 1] \leq \frac{1}{2} + negl(k)$, where the probability is taken over the random coins used by $\mathcal{A}$, as well as the random coins used in the game.

In the next theorem, we prove that our scheme Π_{PEKS} is adaptively secure. The proof relies on the hardness of *Computational Diffie-Hellman Problem* (CDH). Now we provide an variant of CDH problem for multiplicative group which is suitable for our scheme.

Computational Diffie-Hellman Problem (CDH): Let $(\mathbb{G}, \circ)$ be a multiplicative abelian group. Also let g be the generator of $\mathbb{G}$ and let $x, y \in \mathbb{Z}$ such that $a = g^x$, $b = g^y$ and $c = g^z$. The CDH problem is as follows:
given a, b, g as input, compute g^{xyz}. CDH is said to be intractable if all polynomial time algorithms have a negligible advantage in solving CDH.

Suppose $\mathcal{A}$ is a polynomial size adversary that has advantage $negl(\lambda)$ in breaking Π_{PEKS}, i,e., $Pr[Game\_Adaptive_{\mathcal{A},\Pi_{PEKS}}(1^k) = 1] \leq \frac{1}{2} + negl(\lambda)$. Suppose $\mathcal{A}$ makes at most n_{H_2} hash function queries to H_2 and at most and at most n_T trapdoor queries. Here we construct a simulator $\mathcal{S}$ and in Theorem 2 we show that $\mathcal{A}$ solves the CDH problem with probability at least $\epsilon' = \frac{negl(\lambda)}{e \times n_T \times n_{H_2}}$, where e is the base of natural logarithm. The construction of simulator and the proof technique of Theorem 2 is similar to that of [1] but in a different setting on *strings* instead of *keywords* and also the proof is done against CDH assumption.

The simulator $\mathcal{S}$. Let g be the generator of $\mathbb{G}$. Let the simulator $\mathcal{S}$ is given g, $p_1 = g^a$, $p_2 = g^b$, $p_3 = g^c$, $p_4 = g^{ac}$. The simulator $\mathcal{S}$ simulates the challenger and interacts with the adversary $\mathcal{A}$ as follows:
$\mathcal{S}$ sets $A_{pub} = [g, p_1]$.

1. (simulating H_1^*): Whenever $\mathcal{A}$ queries for H_1, $\mathcal{S}$ maintains a list $\langle w_j, h_j, a_j, c_j \rangle$ called $list_1$ which is initially empty. When $\mathcal{A}$ queries for w_i, $\mathcal{S}$ responds as follows:
   - I. If w_i already appears in $list_1$, say $\langle w_i, h_i, a_i, c_i \rangle$, then algorithm $\mathcal{S}$ responds with $H_1^*(w_i) = h_i \in \mathbb{G}$.
   - II. Otherwise $\mathcal{S}$ generates a random coin $c_i\{0,1\}$ such that $Pr[c_i = 0] = \frac{1}{n_T+1}$.
   - III. $\mathcal{S}$ picks a random $a_i \in \mathbb{Z}_p$. If $c_i = 0$, $\mathcal{S}$ sets $h_i = p_2 \times g^{a_i} \in \mathbb{G}$, otherwise $h_i = g^{a_i} \in \mathbb{G}$.
   - IV. $\mathcal{S}$ adds $\langle w_i, h_i, a_i, c_i \rangle$ to the list and responds to $\mathcal{A}$'s query by setting $H_1^*(w_i) = h_i \in \mathbb{G}$.
2. (simulating H_2^*): Whenever $\mathcal{A}$ queries for $H_2(t)$ for a new t, $\mathcal{S}$ picks a random value v from $\{0,1\}^\lambda$ and sets $H_2^*(t) = v$ and adds $\langle t, v \rangle$ to $list_2$. If t is already in the list, then $\mathcal{S}$ just sets $H_2^*(t) = v$.
3. (simulating trapdoor) : To get the simulated trapdoor corresponding to w_i, $\mathcal{S}$ sets $h_i = H_1^*(w_i)$. If $c_i = 0$, it stops. Otherwise $\mathcal{S}$ sets the trapdoor $t_i = p_1^{a_i}$.

challenge phase :

$\mathcal{A}$ produces a pair of challenge strings $s_0 = (w_{0,1}, \ldots, w_{0,m})$ and $s_1 = (w_{1,1}, \ldots, w_{1,m})$ which are of same length, say, m.

- I $\mathcal{S}$ computes $H_1^*(w_{i,j}) = h_{i,j}$ for $0 \leq i \leq 1$, $1 \leq j \leq m$. For $i = 0$ and 1, let the corresponding entries in $list_1$ are $\langle w_{i,k}, h_{i,k}, a_{i,k}, c_{i,k} \rangle$, where $1 \leq k \leq m$.
- II $\mathcal{S}$ randomly selects $b \leftarrow \{0,1\}$.
- III $\mathcal{S}$ responds to the challenge $\{[B_1], [B_2], \ldots, [B_m]\}$ where $\mathcal{S}$ choses B_i's randomly from $\{0,1\}^\lambda$.
- IV $\mathcal{A}$ can continue to issue trapdoor queries for strings other than s_0 and s_1.

Before going into the Theorem 2, here we will prove some inequalities which are crucial for the theorem.

Lemma 2. Let us consider the following events :
ε_1 : The event denoting $\mathcal{S}$ does not abort while $\mathcal{A}$ is making trapdoor queries.
ε_2 : The event denoting $\mathcal{S}$ does not abort while $\mathcal{A}$ is making challenge.
ε_3 : The event denoting $\mathcal{A}$ queried against at least one of the strings s_0 and s_1.
Then, (1) $P[\varepsilon_1] \geq \frac{1}{e}$, (2) $P[\varepsilon_1] \geq \frac{1}{n_T}$ and (3) $P[\varepsilon_3] \geq 2 \times negl(\lambda)$.

Proof:

1. Since in $list_1$, the distribution of c_i's are independent of the distribution of h_i's, we have P[one trapdoor query triggering abort] $= \frac{1}{(n_T+1)}$. So,
$P[\varepsilon_1] = (1-P[\text{one trapdoor query triggering abort}])^{n_T}$
$= \left(1 - \frac{1}{(n_T+1)}\right)^{n_T} \geq \frac{1}{e}$.
2. $P[\mathcal{S}$ will abort during challenge$] = P[c_{i,k} = 1 : i \in \{0,1\}, k \in \{1,\ldots,m\}] = \left(1 - \frac{1}{n_T+1}\right)^{2m} \leq 1 - \frac{1}{n_T}$.
So, $P[\varepsilon_2] \geq 1 - p[\mathcal{S}$ will abort during challenge$] = \frac{1}{n_T}$.
3. Let ε_3 be the event denoting $\mathcal{A}$ queried against at least one of the strings s_0 and s_1.
$P[\ \mathcal{A}$ breaks the scheme$] = P[b = b']$
$= P[(b = b')|\varepsilon_3]P[\varepsilon_3] + P[(b = b')|\varepsilon_3']P[\varepsilon_3']$
$\leq \frac{1}{2}P[\varepsilon_3] + \frac{1}{2}$.
From Definition 4, $negl(\lambda) \leq |P[b = b'] - \frac{1}{2}| = \frac{1}{2}P[\varepsilon_3]$. Thus $P[\varepsilon_3] \geq 2\epsilon$.

□

Theorem 2. *Π_{PEKS} is adaptively secure against chosen keyword attack in the random oracle model assuming CDH is intractable.*

Proof: It may be noted that the challenge implicitly defines B_1 as $H_2^*\left(H_1(w_{b,1})^{ac}\right)$. Now,

$$B_1 = H_2^*\left(H_1(w_{b,1})^{ac}\right) = H_2^*\left(g^{(b+a_b)ac}\right)$$

So similar computation for $w_{b,k}$, $k = 2,\ldots,m$ indicates that this is a valid PEKS for the string s_b.

Output phase :

Eventually $\mathcal{A}$ outputs the guess $b' \in \{0,1\}$. Then $\mathcal{S}$ picks a random pair (t,v) from $list_2$ and outputs $\frac{t}{(p_4)^{a_b}}$ as its guess for g^{abc}, where a_b is the value used in the challenge phase. $\mathcal{A}$ must have issued the query for either s_0 or s_1 as otherwise $\mathcal{A}$'s view on the PEKS will be independent of s_0 or s_1 and thus $\mathcal{A}$ cannot have the advantage of ϵ in breaking the scheme. Therefore with probability $\frac{1}{2}$, $list_2$ contains an entry (t,v) such that $t = g^{ac(b+a_b)}$. Let ε_0 be the event denoting that $\mathcal{S}$ selects this pair (t,v). Then $P[\varepsilon_0] = \frac{1}{n_{H_2}}$, and then $\mathcal{S}$ outputs $\frac{t}{(p_4)^{a_b}} = \frac{g^{ac(b+a_b)}}{g^{a_b ac}} = g^{abc}$. It is easy to check that $P[\mathcal{S}$ solves CDH $] = P[\varepsilon_0] \times P[\varepsilon_1] \times P[\varepsilon_2] \times P[\varepsilon_3]$. So, from Lemma 2, $P[\mathcal{S}$ solves CDH $] \geq \frac{negl(\lambda)}{e \times n_T \times n_{H_2}}$. □

Discussion on Adaptive Security of Π_{PEKS}. The basic idea behind the security proof of Theorem 2 is that the adversary $\mathcal{A}$ is allowed to ask for PEKS encryptions and trapdoors of multiple keywords and strings chosen adaptively. This is formalized by allowing $\mathcal{A}$ to interact freely with an encryption and trapdoor oracle. So from $\mathcal{A}$'s point of view, encryption and trapdoor functionality comes as a *black-box* that encrypts keywords and strings of $\mathcal{A}$'s choice using the secret key which is unknown to $\mathcal{A}$. Since keywords are mapped into finite field elements using the hash function H_1 and the final output of PEKS encryption maps the encrypted keyword to $\{1,0\}^*$ using the hash function H_2, $\mathcal{A}$ is also provided with the oracle access of these two hash functions. When $\mathcal{A}$ queries its oracle by providing it with a keyword as input, the PEKS encryption oracle returns a ciphertext as the reply. Since PEKS encryption is randomized, the oracle uses fresh random coins each time it responds to a query. The definition of security requires that $\mathcal{A}$ should not be able to distinguish the encryption of two arbitrary keywords or strings of same lengths, even when $\mathcal{A}$ is given access to $PEKS\_Enc(.), Trapdoor(.), H_1(.)$ and $H_2(.)$. The security of our scheme relies on the CDH assumption, i.e., so long as CDH assumption holds, we need to show that $\mathcal{A}$ can not win the game defined in Definition 3 with a probability much more than $\frac{1}{2}$. In the proof, we constructed a simulator $\mathcal{S}$, who simulates the challenger in such a way so as to solve CDH. We have shown that if adversary $\mathcal{A}$ wins the game defined in Definition 3, then $\mathcal{S}$ solves the CDH in non-negligible probability. Thus if CDH is believed to be insolvable, reasonably we may infer that $\mathcal{A}$ can not win the game of Definition 3.

6.2 Security of R-PEKS: Data Confidentiality by Secure Access

In this section we show that R-PEKS ensures hosted data confidentiality, i.e., protection of access privilege during string search.

Definition 5. R-PEKS is said to have secure access under injection of faulty roles, if R-PEKS access only the files within the scope of the user's privilege during string search.

In the next theorem, we prove that our R-PEKS access only the files based on the roles during string search which ensures hosted data confidentiality by secure access.

Theorem 3. *Even when the data user colludes with a CSP by the injection of faulty roles, they are unable to access the hosted data by string search over any file that is not in the scope of the data user's assigned roles/privileges*

Proof: In R-PEKS, let P be the set of all permissions. Let $r_i \in R_u$ and $r_j \notin R_u$. Also let user u access the permission set $P(r_i)$ through the assigned role set R_u , i.e. *CheckAccess*$(u,P) = P(r_i)$. So, $P(r_i) = \{p : p \in P, (u,r_i) \in \textbf{UA} \wedge (p,r_i) \in \textbf{PA}\}$.

From the role mining algorithm, there is a threshold δ, such that
$|\textbf{assigned\_permissions(r}_i\textbf{)} \cap \textbf{assigned\_permissions(r}_j\textbf{)}| \leq \delta$. Therefore the similarity rate ranges from 0 to δ. According to [26], we carry out the security mutation analysis by focusing on fault injection of roles, done by CSP while colluding with the data user, as given below :

Original search request of user u is served by the data owner by presenting this request as R-Search$(*, CheckAccess(u,P))$. Since $r_i \in R_u$, R-Search$(*, CheckAccess(u,P))$ should be transformed into R-Search$(*, P(r_i))$.

Under the fault injection assumption, let r_i be replaced by CSP to a faulty role, say r_j. Thus original search request, i.e., R-Search$(*, CheckAccess(u,P))$ transforms in to R-Search$(*, P(r_j))$, where $P(r_j) = \{p : p \in P, (u,r_j) \in \textbf{UA} \wedge (p,r_j) \in \textbf{PA}\}$. There are two possibilities:

Case 1: $|\textbf{assigned\_permissions(r}_i\textbf{)} \cap \textbf{assigned\_permissions(r}_j\textbf{)}| > 0$

Case 2: $|\textbf{assigned\_permissions(r}_i\textbf{)} \cap \textbf{assigned\_permissions(r}_j\textbf{)}| = 0$

In Case 1, string search is performed on files common to r_i and r_j. In Case 2, no operation is performed. Suppose the actual number of files identified for a string search on a user without faulty injection is F. Also let us assume that same number of files are identified for successive *k-1* searches on the same string. Finally, let the number of identified files for a same string search caused by the fault injection of roles on k^{th} iteration of search is δ. If files returned after search are the files that are accessible according to role r_i, we call it a *success*. If T is the expected number of files satisfying a particular search string after k iterations, then we define *score* S, on *success* as $S = \frac{(k-1)*\mathrm{F}+\delta}{\mathrm{T}}$. It is easy to check that $S \leq 1$. This is because searches for user u are performed using trapdoors generated from u's private key on data encrypted by u's public key. So from the PEKS privacy, under no situations, the files corresponding to the permissions [**assigned_permissions(**r_j**)** $-$ **assigned_permissions(**r_i**)**] will respond to the search, even if they contain the query string as these files are not encrypted using u's public key. Thus these files never contribute in S.

If $S < 1$, then the system is under attack by the injection of faulty role which affected the string search. If $S = 1$, then $\delta = \mathrm{F}$ and it is a fuzzy state. In such cases the string search is not affected even though the system is under attack. Therefore in all the above cases R-PEKS scheme is secure under the Definition 5, which ensures high level of hosted data confidentiality when data user collude with the CSP by injecting the faulty role during the string search. □

7 Experimental Results

In this section we present an experimental set-up by generating the RBAC model on the PEKS environment. We also provide the performance analysis of our proposed model on TIMIT dataset [27]. Finally, we validate the correctness of our model, i.e. the data confidentiality using *test automation framework* [28]. The implementation is done on Intel Core(TM) $i5-7500$ with 16 GB RAM using Java in Windows platform. For the cryptographic primitives, 'Jpair' library is used.

7.1 Creation of RBAC configuration using RMiner

The creation of RBAC configuration using RMiner [29] is achieved in two phases. Firstly, we generate **UP** (see Subsection 4.3) and secondly we generate roles and its assignments to users as well as permissions, based on the **UP**. The required objects to generate a **UP** are taken from the TIMIT speech corpus [27], which is a phonemically and lexically transcribed speech of American English speakers of different dialects. This dataset is comprised of 42 distinct phonetic symbols giving rise to an average of approximately 200 words in 4607 files in a document of size 19 MB. For the experimentation purpose we considered file level access, i.e., each file is treated as an object and each object is assigned a permission, which is either 1 (grant) or 0 (deny) under two operations, namely encryption and search. It is assumed that if a file has a permission for encryption then it also has a permission for search. Further, we have incorporated 500 users in the implementation. The users and files are given as an input to RMiner [29], which is a role mining tool used for the generation of **UP** and for the creation of RBAC configuration, i.e., **UA** and **PA** (see Subsection 4.3). The generated **UP** for the given number of users and permissions is listed in Table 2. Also, in Table 2, the parameters presented in forth, fifth, sixth and seventh rows are given as input and the rest are generated by RMiner tool. We use

TABLE 2
Statistics of the generated RBAC components for TIMIT speech dataset [27] using RMiner

Input/Output	Values
Total number of users	500
Total number of files	4607
Generated |UP|	11,26,308
Mean of permissions in role	25
Variance of permissions in role	10
Mean of roles in user	10
Variance of roles in user	5
Generated |UA|	313
Generated |PA|	27,857
Generated |R|	285

HPRoleMinimization algirithm of RMiner, which provides the most compact and optimal solution in the creation of RBAC configuration compared other algorithms in the RMiner tool. The total time taken by RMiner algorithm is 46.43 seconds.

7.2 Performance analysis on PEKS and R-PEKS

All performances are analyzed based on the average time taken by 5 different users to perform encryption and search. These are illustrated on the speech dataset by comparing with the existing models. For the analysis purpose RBAC is instantiated with PEKS proposed in [3] and further it is referred as *r-PEKS*.

Performance of PEKS. In the next lemma, we provide complexity analysis of our scheme.

Lemma 3. Using Π_{PEKS}, the number of group operations for searching a query of l-word string in one ciphertext document C is $O\left(\left(\frac{|C|}{\lambda}\right) + (l-1)\right)$.

Proof: It is easy to observe that each block of the form $[B]$ is of size λ bits. Thus in the encrypted document C, the number of blocks is $\frac{|C|}{\lambda}$. To detect the first block requires $O\left(\frac{|C|}{\lambda}\right)$ group operations. Since the blocks are not shuffled, to detect rest of the $l-1$ blocks, $l-1$ group operations are needed. Thus the number of group operations is $O\left(\left(\frac{|C|}{\lambda}\right) + (l-1)\right)$. □

Figure 3 compares the average PEKS encryption time for our PEKS scheme against PEKS of [3]. We plotted the graph by considering the average time in seconds along Y-axis against the number of files needed to encrypt along X-axis. For both schemes, the graph reflects a linear growth with the increase in number of files. However average time taken by our PEKS scheme is 5 seconds to encrypt 250 files which is very much less compare to the existing PEKS [3] where it took 188 seconds for same operation.

Figure 4 represents the performance of searching over encrypted files. We plotted the average time of search (in seconds along Y-axis) against the number of encrypted files (along X-axis). The graph shows a linear growth with the

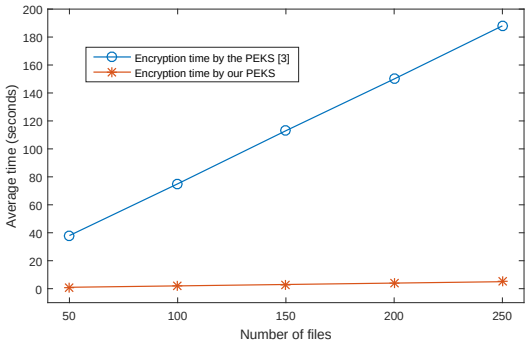

Fig. 3. Comparison of encryption time for our PEKS and PEKS [3]

increase in the number of files. The average time taken by our PEKS scheme to search over 250 encrypted documents is 2.5 seconds which is 97% efficient compared to the performance of PEKS of [3], where they took 88 seconds for the same search operation.

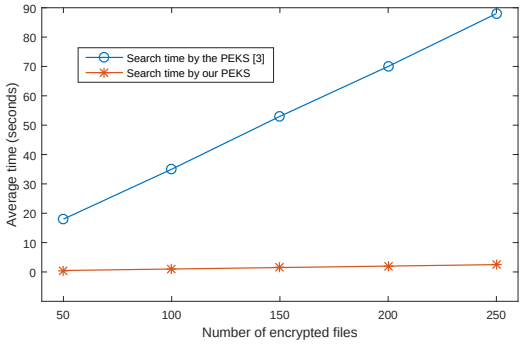

Fig. 4. Comparison of search time for our PEKS and PEKS [3]

Performance of R-PEKS and PEKS. We note that, both the PEKS schemes can have a better performance when integrated with RBAC. Now, we discuss the enhancements of R-PEKS over our PEKS. In Figure 5, we provide comparison of R-PEKS and our PEKS search performance. We plotted the

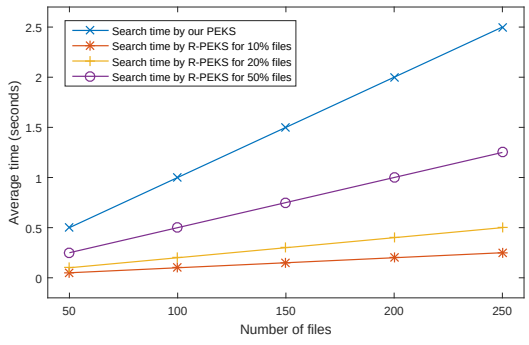

Fig. 5. Comparison of search time for R-PEKS and PEKS

graph by considering the average time of search (in seconds along Y-axis) against the available number of files (along X-axis) using R-PEKS and our PEKS scheme. In all the cases, graph shows a linear growth with the increase in number of files, however the rate of growth reflects the efficiency of R-PEKS against PEKS scheme. In PEKS, since the search is not refined by privileges (i.e., by roles), the search is performed among all the existing files. So the average search time using our PEKS scheme for any percentage of matching files is 2.5 seconds on 250 files. Since R-PEKS is selective search, the string search request is performed only on the privileged files. So the average search time using R-PEKS is 0.25 seconds, 0.5 seconds and 1.25 seconds when the assigned roles contain 10%, 20% and 50% privileged files over 250 files respectively. So effectively from the user's point of view R-PEKS will be faster compare to PEKS. This explains the efficiency of R-PEKS over our PEKS by 90% when the required data is present only in 10% files over 250 files.

Performance of different PEKS schemes with RBAC. In [10], authors proposed idea of generating the synthetic datasets to evaluate the performance of role mining algorithms. Towards this, we have generated the two synthetic datasets presented in Table 3 and Table 4 to compare the performance of R-PEKS. Both the synthetic datasets consist of five parameters, namely, number of users (#U), number of roles (#R), number of permissions for each user (#P), maximum number of permissions for a role and scope of privileges (range of files). In the first synthetic dataset, user's scope of privileges is a varying parameter and other parameters are constant as shown in Table 3 and the corresponding comparison study is presented in Figure 6. In the second synthetic dataset, user's permission is a varying parameter and other parameters are constant as shown in Table 4 and the corresponding comparison study is presented in Figure 7.

TABLE 3
Users with varying scope of privileges for a constant number of permissions

#U	#R	#P	Maximum number of permissions for a role	Scope of privileges
10	20	10	5	50
10	20	10	5	100
10	20	10	5	150
10	20	10	5	200
10	20	10	5	250

In Figure 6 and Figure 7 we plot the average time (along Y-axis) needed by the users to encrypt the documents and to access the data by searching (along X-axis) based on the values tabulated in Table 3 and Table 4 respectively. Figure 6 reflects a linear growth of average time of encryption against scope of privileges (range of files). Figure 6 shows that for R-PEKS, average time varies from 0.5 seconds to 1.7 seconds, whereas for r-PEKS, the average time varies from 8 seconds to 9.2 seconds over scope of privileges varying from 50 to 250 for a constant 10 number of permissions. In case of searching, R-PEKS takes constant average time of 0.1 seconds whereas r-PEKS takes 3.6 seconds respectively under same scope of privileges and permissions as mentioned in Table 3. Therefore, the above comparison revels that the efficiency of R-PEKS over r-PEKS during encryption and search are 81% and 97% respectively.

Figure 7 reflects a linear growth with the increase in the number of permissions for a constant scope of privileges.

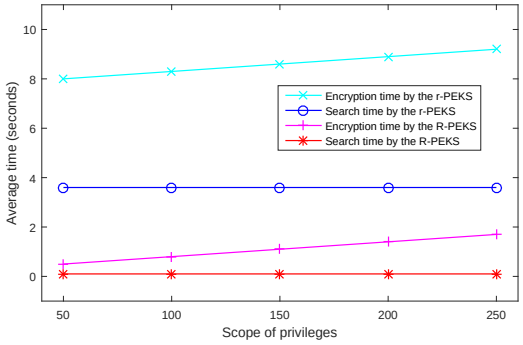

Fig. 6. Comparison of encryption and search time for r-PEKS and R-PEKS on varying scope of privileges for a constant number of permissions

TABLE 4
Users with varying number of permissions for a constant scope of privileges

#U	#R	#P	Maximum number of permissions for a role	Scope of privileges
10	20	10	5	80
10	20	20	5	80
10	20	30	5	80
10	20	40	5	80
10	20	50	5	80

Further, the average time needed for encrypting 50 files in a scope of 80 privileges using R-PEKS and r-PEKS are around 2.5 seconds and 40 seconds respectively. The average time needed to search the data over the 50 encrypted files using R-PEKS and r-PEKS are 0.5 seconds and 17.5 seconds respectively. Above comparison revels that the efficiency of

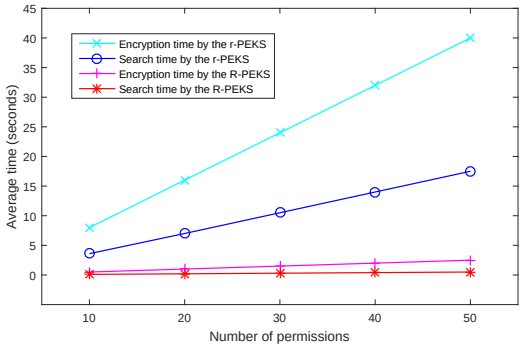

Fig. 7. Comparison of encryption and search time for r-PEKS and R-PEKS on varying number of permissions for a constant scope of privileges

R-PEKS over r-PEKS during encryption and search are 93% and 97% respectively.

Automation of Secure Access by R-PEKS. In this section we determine the functional correctness of our proposed model on accessing only permitted files using R-PEKS in single-user and multi-user environment. To validate this experimentally, we have carried out the testing by developing a test automation framework, which can spawn 100 parallel users having different roles in every 30 seconds as shown in Figure 8. In the *logger* block of Figure 8, the the user behavior patterns are saved for analysis.

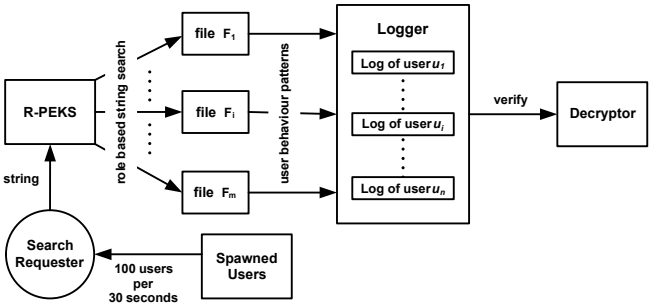

Fig. 8. Test automation framework to determine the correctness of role based file access for string search using R-PEKS

To determine the correctness, 10 different users are retrieved randomly from the log and all the user behavior patterns are monitored. After one hour of experiment, the data that are saved in the logger are studied which reveals not a single instance of unauthorized access.

8 Comparison With Other Schemes

8.1 Comparison with scheme in [30]

In [30], authors proposed MUSE scheme with efficient access control for cloud storage, where the keyword index and trapdoor can be generated with the help of a proxy server. To achieve this authors constructed a new MUSE scheme, where the keyword index and trapdoor can be generated with the help of an additional proxy server. It may be noted that in our scheme, we need not to generate any such index and thus is free from the additional index management and is also free from the risk of leakage from the index.

Also in [30], the core mathematical construct for the searchable encryption is bilinear pairing, which is a computationally intensive module. It may be noted that the scheme proposed in this paper is the PEKS scheme which is free from such bilinear pairing. To the best of our knowledge this is the only such PEKS scheme which is free from bilinear pairing and is much more efficient and easy to implement.

The search complexity of the scheme proposed in this paper is number of finite field operation which is linear in $|C|$ for a cipher file of size $|C|$. Although the complexity analysis of search is not explicitly mentioned, but from Algorithm 4 of [30], it is clear that search complexity is $O(n)$ where n is the number of keywords in a cipher file. Since we are doing sentence search where each word is of length λ, n in [30] is equivalent to $\frac{|C|}{\lambda}$, where $|C|$ is the size of cipher file. So, when expressed in our notation, search complexity of [30] is also liner in $|C|$ but operations are in elliptic curve which is costly compared to finite field operations.

8.2 Comparison with scheme in [31]

In [31] authors proposed template of adaptively secure searchable encryption scheme with access control without any practical instantiation and so the comparison seems difficult. However we observed that their scheme is based on SSE of [15]. One problem SSE is that, adaptive security and string search can not be achieved simultaneously. In this paper, we emphasis on customized search under three application scenarios by enforcing string search without

compromising the security. So the scheme of [31] can not be adapted for the application scenarios which are considered in this paper.

9 Conclusion and Future work

RBAC enabled PEKS is the most suitable and efficient solution for secure search applications in multi-user setting where user permissions are updated frequently. Towards this, we have designed R-PEKS. We have designed a new PEKS, called Π_{PEKS} for this. In the threat model, we have considered honest-but-curious CSP as well as data users and have analyzed the security requirements. Finally we have shown R-PEKS to be secure under the definition of adaptive security and also provides hosted data confidentiality by secure access. The experiment is conducted on TIMIT speech dataset [27] to evaluate the performance of our proposed model. We have shown that with respect to the PEKS of [3] for string identification, our PEKS scheme, i.e., Π_{PEKS} is efficient by 97%. We have also shown that using R-PEKS, an user can gain up to 90% efficiency on searching when the share of data pertaining to him is 10% of the whole cloud data. For comparison study, we also have implemented PEKS of [3] with RBAC and named it r-PEKS. With the generated synthetic dataset, we have shown that the efficiency of R-PEKS against r-PEKS is up to 87% for encryption and 97% for searching. We have developed the *test automation framework* and have determined the functional correctness of R-PEKS by it. It might be of interest to explore how PEKS can be integrated with dynamic user permission assignment using least privilege user-role assignment problem for a better performance and security.

Acknowledgment

This research was partly supported by European Commissions Horizon 2020 research and innovation project FORTIKA under grant agreement No 740690.

References

[1] D. Boneh, G.D. Crescenzo, R. Ostrovsky, and G. Persiano. Public Key Encryption With Keyword Search. In *Proceedings of 23rd International Conference on the Theory and Applications of Cryptographic Techniques*, Switzerland, 2004.

[2] I.G. Ray, Y. Rahulmathavan, and M. Rajarajan. A New Lightweight Symmetric Searchable Encryption Scheme for String Identification. *IEEE Transactions on Cloud Computing*, 2018. doi: 10.1109/TCC.2018.2820014.

[3] I.G. Ray and M. Rajarajan. A Public Key Encryption Scheme for String Identification. In *Proceedings of 16th IEEE International Conference On Trust, Security And Privacy In Computing And Communications*, Sydney, 2017.

[4] H. Cui, Z. Wan, R.H. Deng, G. Wang, and Y. Li. Efficient and Expressive Keyword Search Over Encrypted Data in Cloud. *IEEE Transactions on Dependable and Secure Computing*, 15(3), 2018.

[5] L.Y. Zhang, Y. Zheng, J. Weng, C. Wang, Z. Shan, and K. Ren. You Can Access But You Cannot Leak: Defending against Illegal Content Redistribution in Encrypted Cloud Media Center. *IEEE Transactions on Dependable and Secure Computing*, 2018. doi: 10.1109/TDSC.2018.2864748.

[6] D.R. Kuhn, E.J. Coyne, and T.R. Weil. Adding attributes to Role-Based Access Control. *IEEE Computer*, 43(6):79–81, 2010.

[7] K. He, J. Guo, J. Weng, J. Weng, J.K. Liu, and X. Yi. Attribute-Based Hybrid Boolean Keyword Search over Outsourced Encrypted Data. *IEEE Transactions on Dependable and Secure Computing*, 2018. doi: 10.1109/TDSC.2018.2864186.

[8] Z. Shen, J. Shu, and W. Xue. Keyword Search With Access Control Over Encrypted Cloud Data. *IEEE Sensors Journal*, 17(3):858–868, 2017.

[9] J. Benaloh, M. Chase, E. Horvitz, and K. Lauter. Patient controlled encryption: Ensuring privacy of electronic medical records. In *Proceedings of 16th ACM Conference on Computer and Communications Security*, Chicago, 2009.

[10] J. Vaidya, V. Atluri, J. Warner, and Q. Guo. Role engineering via prioritized subset enumeration. *IEEE Transactions on Dependable and Secure Computing*, 7(3):300–314, 2010.

[11] Q. Wang, M. He, M. Du, S.S.M. Chow, R.W.F. Lai, and Q. Zou. Searchable Encryption over Feature-Rich Data. *IEEE Transactions on Dependable and Secure Computing*, 15(3), 2018.

[12] C.V. Rompay, R. Molva, and M. Onen. Secure and Scalable Multi-User Searchable Encryption. In *Proceedings of 6th International Workshop on Security in Cloud Computing*, Republic of Korea, 2018.

[13] L. Zhou, V. Varadharajan, and H. Hitchens. Achieving Secure Role-Based Access Control on Encrypted Data in Cloud Storage. *IEEE Transactions on Information Forensics and Security*, 8(12), 2013.

[14] E.J. Goh, H. Shacham, N. Modadugu, and D. Boneh. Sirius: Securing remote untrusted storage. In *Proceedings of 10th Annual Network and Distributed System Security Symposium*, California, 2003.

[15] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky. Searchable Symmetric Encryption: Improved Definitions and Efficient Constructions. *Journal of Computer Security*, 19(5):895–934, 2011.

[16] Z. Liu, Z. Wang, X. Cheng, C. Jia, and K. Yuan. Multi-user Searchable Encryption with Coarser-Grained Access Control in Hybrid Cloud. In *Proceedings of 4th International Conference on Emerging Intelligent Data and Web Technologies*, China, 2013.

[17] B. Cui, Z. Liu, and L. Wang. Key-Aggregate Searchable Encryption (KASE) for Group Data Sharing via Cloud Storage. *IEEE Transactions on Computers*, 65(8):2374–2385, 2016.

[18] W. Sun, S. Yu, W. Lou, Y.T. Hou, and H. Li. Protecting your right: Attribute-based keyword search with fine-grained owner-enforced search authorization in the cloud. In *Proceedings of 33rd Annual IEEE International Conference on Computer Communications*, Toronto, 2014.

[19] Y. Yang, X. Liu, and R.H. Deng. Multi-user Multi-Keyword Rank Search over Encrypted Data in Arbitrary Language. *IEEE Transactions on Dependable and Secure Computing*, 2017. doi: 10.1109/TDSC.2017.2787588.

[20] G. Wang, C. Liu, Y. Dong, P. Han, H. Pan, and B. Fang. IDCrypt: A Multi-User Searchable Symmetric Encryption Scheme for Cloud Applications. *IEEE Access*, 6:2908–2921, 2017.

[21] W. Shen, J. Qin, J. Yu, R. Hao, and J. Hu. Enabling identity-based integrity auditing and data sharing with sensitive information hiding for secure cloud storage. *IEEE Transactions on Information Forensics and Security*, 14(2):331–346, 2019.

[22] M. Liu, Y. Zhao, and S. Chen. eCK-security authenticated key agreement protocol based on CDH assumption. In *Proceedings of 2nd IEEE International Conference on Computer and Communications*, China, 2016.

[23] J. Katz and Y. Lindell. *Introduction to Modern Cryptography*. CRC press, 2014.

[24] D.R. Stinson. *Cryptography: Theory and Practice*. CRC press, 2005.

[25] R. Sandhu, D. F. Ferraiolo, and D. R. Kuhn. The NIST Model for Role-Based Access Control: Towards a Unified Standard. Technical report, Laboratory for Information Security Technology, 2000.

[26] D. Xu, M. Tu, M. Sanford, L. Thomas, D. Woodraska, and W. Xu. Automated Security Test Generation with Formal Threat Models. *IEEE Transactions on Dependable and Secure Computing*, 9(4):526–540, 2012.

[27] http://www.fon.hum.uva.nl/david/ma_ssp/2007/timit/train/dr5/fsdc0/. 2007.

[28] Lucile Wybouw-Cognard. Test automation framework, August 22 2002. US Patent App. 09/925,872.

[29] R. Li, H. Li, W. Wang, X. Ma, and X. Gu. RMiner: A Tool Set for Role Mining. In *Proceedings of 18th ACM Symposium on Access Control Models and Technologies*, Netherlands, 2013.

[30] Z. Lv, M. Zhang, and D. Feng. Multi-user Searchable Encryption with Efficient Access Control for Cloud Storage. In *Proceedings of 6th International Conference on Cloud Computing Technology and Science*, Singapore, 2014.

[31] N. Löken. Searchable Encryption with Access Control. In *Proceedings of 12th International Conference on Availability, Reliability and Security*, Italy, 2017.

K. Rajesh Rao received a BE degree in Computer Science and Engineering from the Visvesvaraya Technological University, Belagavi, MTech in Computer Science and Information Security from Manipal Institute of Technology, MAHE, India. He is having one year work experience as Graduate Technical Intern at Intel Corporation, Bangalore. He is currently a Research Assistant at City, University of London, UK and also working towards his PhD. He is also an Assistant Professor-Senior in the Department of Information and Communication Technology at the Manipal Institute of Technology, Manipal. His research interest include but are not limited to cloud security, role mining and applied cryptography.

Indranil Ghosh Ray received a B.Sc. degree (Mathematics Honors) from the Calcutta University, in 2000, and a MCA (Master of Computer Applications) degree in 2003 from University of Kalyani. He worked in software industry for six years then as software engineer and senior software engineer. He joined Indian Statistical Institute, Kolkata, India as Junior Research Fellow in 2010 and was promoted to Senior Research Fellow in 2013. He received a Ph.D degree in computer science from Indian Statistical Institute in 2016. He is a Research Associate with the Information Security Group, School of Engineering and Mathematical Sciences, City University of London, UK since June 2016. His research interests include Homomorphic Encryption and its application in Privacy preserving Cloud Computing, Searchable Symmetric Encryption (SSE), Public Key Encryption with keyword search, MDS codes and its applications in Lightweight Cryptography, algebraic immunity of S-Boxes based on Power Mappings but are not limited to these only.

Waqar Asif received his Ph.D. degree from City, University of London in 2016 and since then he is working as a Research Fellow at the same. He did his MS and BEng in 2012 and 2009 respectively from well reputed institutions in Pakistan. Before moving to the UK, he served as a Lecturer in Bahria University Pakistan for a year. He secured multiple merit based scholarships which includes an Erasmus Mundus STrongTies Scholarship for one and a half year for Cyprus where he worked as a researcher in Frederick University. His research interest include but are not limited to graph theory, sensor networks, network performance metrics, block chain and network privacy.

Ashalatha Nayak is a Professor and currently heading the Department of Computer Science and Engineering at the Manipal Institute of Technology, MAHE, Manipal, India. She obtained her BTech and MTech in Computer Science and Engineering from the Mangalore University, Karnataka, India, PhD from the School of Information Technology, IIT Kharagpur in the area of model based testing. Her research interests include semantic web, software testing, intelligent agents and cloud security.

Muttukrishnan Rajarajan is a Professor of Security Engineering at City, University of London, UK. He currently leads the Information Security Group at City and his research interests are in the areas of privacy preserving data analytics, cloud computing, internet of things security and wireless networks. He has published well over 300 papers and continues to be involved in the editorial boards and technical programme committees of several international security and privacy conferences and journals. He is a visiting Researcher at the British Telecommunications Security Research and Innovation laboratory and is an advisory board member of the Institute of Information Security Professionals, UK and acts as an advisor to the UK Government's Identity Assurance programme (Verify UK). He is a Senior Member of IEEE.