

Provable Attribute Based Keyword Search Over Outsourced Encrypted Data

LOGANAYAKI.R ¹ MATHESWARAN.V ²

1. PG Student, Dept.of Computer Applications, VSB Engineering College, Karur.
2. Assistant Professor, Dept.of Computer Applications, VSB Engineering College, Karur.

Abstract: Search over encrypted data is a critically important enabling technique in cloud computing, where encryption-before outsourcing is a fundamental solution to protecting user data privacy in the untrusted cloud server environment. Many secure search schemes have been focusing on the single-contributor scenario, where the outsourced dataset or the secure searchable index of the dataset are encrypted and managed by a single owner, typically based on symmetric cryptography. In this paper, we focus on a different yet more challenging scenario where the outsourced dataset can be contributed from multiple owners and are searchable by multiple users, i.e. multi-user multi-contributor case. Inspired by attribute-based encryption (ABE), we present the first attribute-based keyword search scheme with efficient user revocation (ABKS-UR) that enables scalable fine-grained (i.e. file-level) search authorization. Our scheme allows multiple owners to encrypt and outsource their data to the cloud server independently. Users can generate their own search capabilities without relying on an always online trusted authority. Fine-grained search authorization is also implemented by the owner-enforced access policy on the index of each file. Further, by incorporating proxy re-encryption and lazy re-encryption techniques, we are able to delegate heavy system update workload during user revocation to the resourceful semi-trusted cloud server. We formalize the security definition and prove the proposed ABKS-UR scheme selectively secure against chosen-keyword attack. To build confidence of data

user in the proposed secure search system, we also design a search result verification scheme.

1. INTRODUCTION

Cloud computing has emerged as a new enterprise IT architecture. Many companies are moving their applications and databases into the cloud and start to enjoy many unparalleled advantages brought by cloud computing, such as on-demand computing resource configuration, ubiquitous and flexible access, considerable capital expenditure savings, etc. However, privacy concern has remained a primary barrier preventing the adoption of cloud computing by a broader range of users/applications. When sensitive data are outsourced to the cloud, data owners naturally become concerned with the privacy of their data in the cloud and beyond. Encryption-before-outsourcing has been regarded as a fundamental means of protecting user data privacy against the cloud server (CS). However, how the encrypted data can be effectively utilized then becomes another new challenge. Significant attention has been given

and much effort has been made to address this issue, from secure search over encrypted data, secure function evaluation, to fully homomorphic encryption systems that provide generic solution to the problem in theory but are still too far from being practical due to the extremely high complexity.

This paper focuses on the problem of search over encrypted data, which is an important enabling technique for the encryption-before-outsourcing privacy protection paradigm in cloud computing, or in general in any networked information system where servers are not fully trusted. Much work has been done, with majority focusing on the single-contributor scenario, i.e., the dataset to be searched is encrypted and managed by a single entity, which we call owner or contributor in this paper. Under this setting, to enable search over encrypted data, the owner has to either share the secret key with authorized users, or stay online to generate the search trapdoors, i.e., the “encrypted” form of keywords to be searched, for the users upon request. The same symmetric key will be used to encrypt the dataset (or the searchable index of the dataset) and to generate the trapdoors. These schemes seriously limit the users’ search flexibility.

Consider a file sharing system that hosts a large number of files, contributed from multiple owners and to be shared among multiple users. This is a more challenging multi-owner multi-user scenario. How to enable multiple owners to encrypt and add

their data to the system and make it searchable by other users? Moreover, data owners may desire fine-grained search authorization that only allows their authorized users to search their contributed data. By fine-grained, we mean the search authorization is controlled at the granularity of per file level. Symmetric cryptography based schemes are clearly not suitable for this setting due to the high complexity of secret key management. Although authorized keyword search can be realized in single-owner setting by explicitly defining a server-enforced user list that takes the responsibility to control legitimate users’ search capabilities, i.e., search can only be carried out by the server with the assistance of legitimate users’ complementary keys on the user list, these schemes did not realize fine-grained owner-enforced search authorization and thus are unable to provide differentiated access privileges for different users within a dataset. Asymmetric cryptography is better suited to this dynamic setting by encrypting individual contribution with different public keys. For example, Hwang and Lee implicitly defined a user list for each file by encrypting the index of the file with all the public keys of the intended users. However, extending such user list approach to the multi-owner setting and on a per file basis is not trivial as it would impose significant scalability issue considering a potential large number of users and files supported by the system. Additional challenges include how to handle the updates

of the user lists in the case of user enrollment, revocation, etc., under the dynamic cloud environment.

In this paper, we address these open issues and present an authorized keyword search scheme over encrypted cloud data with efficient user revocation in the multi-user multi-data-contributor scenario. We realize fine-grained owner-enforced search authorization by exploiting ciphertext policy attribute-based encryption (CP-ABE) technique. Specifically, the data owner encrypts the index of each file with an access policy created by him, which defines what type of users can search this index. The data user generates the trapdoor independently without relying on an always online trusted authority (TA). The cloud server can search over the encrypted indexes with the trapdoor on a user's behalf, and then returns matching result if and only if the user's attributes associated with the trapdoor satisfy the access policies embedded in the encrypted indexes. We differentiate attributes and keywords in our design. Keywords are actual content of the files while attributes refer to the properties of users. The system only maintains a limited number of attributes for search authorization purpose. Data owners create the index consisting of all keywords in the file but encrypt the index with an access structure only based on the attributes of authorized users, which makes the proposed scheme more scalable and suitable for the large scale file sharing system. In order to further release the

data owner from the burdensome user membership management, we use proxy re-encryption and lazy reencryption techniques to shift the workload as much as possible to the CS, by which our proposed scheme enjoys efficient user revocation. Formal security analysis shows that the proposed scheme is provably secure and meets various search privacy requirements.

Contributions can be summarized as follows:

- 1) We design a novel and scalable authorized keyword search over encrypted data scheme supporting multiple data users and multiple data contributors. Compared with existing works, our scheme supports fine-grained owner-enforced search authorization at the file level with better scalability for large scale system in that the search complexity is linear to the number of attributes in the system, instead of the number of authorized users.
- 2) Data owner can delegate most of computationally intensive tasks to the CS, which makes the user revocation process efficient and is more suitable for cloud outsourcing model.
- 3) We formally prove our proposed scheme selectively secure against chosen-keyword attack.
- 4) We propose a scheme to enable authenticity check over the returned search result in the multi-user multi-data-contributor search scenario.

2. LITERATURE SURVEY

Wei Zhang, Yaping Lin, Sheng Xiao, Jie Wu, and Siwang Zhou [1], explore the problem of secure multi-keyword search in multi-keyword search. PRMSM model in this system searches a keywords without knowing actual data of trapdoors as well as keywords. This system preserves the keywords and files systematically. In this system sum of the relevance scores is used to search result in metric. Authors defined the problem of secure search over encrypted data. Additive Order and Privacy Preserving Function family (AOPPF) is proposed to preserve the privacy of relevant scores of different functions. This system works on Ranked Multi-keyword Search over Multi-owner, Data owner scalability, Data user revocation and Security Goals of system.

M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, [2] provides the simple figure to evaluate the comparison between cloud computing and conventional computing. It also identifies functional and non-functional opportunities of cloud storage.

C. Wang, S. S. Chow, Q. Wang, K. Ren, and W. Lou [3] provide data security in cloud this paper proposed a privacy-preserving public auditing system. This system handles multiple audit session different users for their outsourced data files. The privacy-preserving

public auditing scheme required to design auditing protocol to prevent data from flowing away. Therefore it is not completely solve the problem of privacy preserving in key management. Therefore unauthorized data leaked problem cannot be solved by this system. TPA audit outsourced data when it is required. Authors were utilizes homomorphic linear authenticator and random masking to provide assurance that TPA cannot learn about knowledge of data.

D.Song, D.Wagner, and A.Perrig,[4],describes cryptographic schemes for the problem of searching on encrypted data. It also provides proofs of security for the resulting crypto systems. This scheme is provably secure for remote searching on encrypted data using an untrusted server. This system searches data remotely from untrusted server. This system provides the proofs of security that required for crypto systems. This system worked efficiently for query isolation as they are simple and fast. Only $O(n)$ stream cipher required for encryption and search algorithm.

R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky [5],reviewing existing notions of security and propose new and stronger security definitions called as Searchable symmetric encryption (SSE). This scheme allows outsourcing the data to other party. They proved stronger security level. This system solves the problem of searchable symmetric encryption. This system provides guarantee of security for user which aims to perform search

at once. Two new SSE constructions are proposed for stronger security definitions.

P. Golle, J. Staddon, and B. Waters [6], proposed protocols that allow for conjunctive keyword queries on encrypted data. It solves the problem of secure Boolean search.

This technique is based on simple keyword search method. This system proposed an approach that defines meta-keywords that are associated with documents. Problem with this approach is that It requires, $2m$ keyword search for every keyword m .

C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, [7], proposed schemes in this paper support only boolean keyword search. This scheme solves the problem of supporting efficient ranked keyword search. By doing this effective utilization of remotely stored encrypted data is achieved in Cloud Computing. Authors were mainly concerning on searching effective as well as secure ranked keyword searching for encrypted data. This system uses SSE technique for keyword searching. For ranking function TF x IDF rules are used. For security purpose OPSE crypto primitive is developed in this system.

N. Cao, C. Wang, M. Li, K. Ren, and W. Lou [8] define and solve the problem of multi-keyword ranked search over encrypted cloud data (MRSE) and they also concern with preserving strict system-wise privacy in the cloud computing paradigm. MRSE schemes to achieve various stringent privacy requirements

in two different threat models. Coordinate matching technique is used to capture the relevance of data documents required for query. This system uses “inner product similarity” to search number of keywords in the document. To attempt this purpose authors were proposing MRSE technique. Compare to other multikeyword ranked searching technique this system produces very overheads.

J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou [9], formalizes and solves the problem of effective fuzzy keyword search over encrypted cloud data and maintains keyword privacy. An advance technique is proposed i.e. wildcard-based technique for searching fuzzy keywords. Fuzzy keyword search technique improves the usability of system by returning files with exactly matching keywords that are pre-defined. Proxy-server in this system is used give response for receiver keyword query. PEKS does not have a requirement of coordination between sender and receiver when they are firstly join in opposite. This system requires special methods for sorting the keywords.

2.1 EXISTING SYSTEM

The problem of search over encrypted data, which is an important enabling technique for the encryption-before-outsourcing privacy protection paradigm in cloud computing, or in general in any networked information system where servers are not fully trusted. Much work has been done, with majority focusing on the

single-contributor scenario, i.e. the dataset to be searched is encrypted and managed by a single entity, which we call owner or contributor in this paper. Under this setting, to enable search over encrypted data, the owner has to either share the secret key with authorized users, or stay online to generate the search trapdoors, i.e. the “encrypted” form of keywords to be searched, for the users upon request. The same symmetric key will be used to encrypt the dataset (or the searchable index of the dataset) and to generate the trapdoors. These schemes seriously limit the users’ search flexibility.

2.1.1 Drawbacks of existing system

- ❖ Huge cost in terms of data usability. For example, the existing techniques on keyword-based information retrieval, which are widely used on the plaintext data, cannot be directly applied on the encrypted data. Downloading all the data from the cloud and decrypt locally is obviously impractical.
- ❖ Existing System methods not practical due to their high computational overhead for both the cloud sever and user.

2.2 PROPOSED SYSTEM

The proposed system addresses these open issues and present an authorized keyword search scheme over encrypted cloud data with efficient user revocation in the multi-user multi-data-contributor scenario. We realize fine-grained owner-enforced search authorization by exploiting ciphertext policy

attribute-based encryption (CPABE) technique. Specifically, the data owner encrypts the index of each file with an access policy created by him, which defines what type of users can search this index. The data user generates the trapdoor independently without relying on an always online trusted authority (TA). The cloud server (CS) can search over the encrypted indexes with the trapdoor on a user’s behalf, and then returns matching result if and only if the user’s attributes associated with the trapdoor satisfy the access policies embedded in the encrypted indexes. We differentiate attributes and keywords in our design. Keywords are actual content of the files while attributes refer to the properties of users. The system only maintains a limited number of attributes for search authorization purpose.

2.2.1 Advantages

1. Due to the special structure of the tree-based index, the proposed search scheme can flexibly achieve sub-linear search time and deal with the deletion and insertion of documents.
2. A searchable encryption scheme that supports both the accurate multi-keyword ranked search and flexible dynamic operation on document collection.
3. Due to the special structure of the tree-based index, the search complexity of the proposed scheme is fundamentally kept to logarithmic. And in practice, the proposed scheme can

achieve higher search efficiency by executing our “Greedy Depth-first Search” algorithm. Moreover, parallel search can be flexibly performed to further reduce the time cost of search process.

3. SYSTEM OVERVIEW

This system considers three users, Data owner, data user, and cloud server.

Data owner has a collection of data documents $D = \{d_1, d_2, \dots, d_m\}$. A set of distinct keywords $W = \{w_1, w_2, \dots, w_n\}$ is extracted from the data collection D . The data owner will firstly construct an encrypted searchable index I from the data collection D . All files in D are encrypted and form a new file collection, C . Then, the data owner upload both the encrypted index I and the encrypted data collection C to the cloud server.

Data user provides t keywords for the cloud server. A corresponding trapdoor w T through search control mechanisms is generated. The system assumes that the authorization between the data owner and the data user is approximately done.

Cloud server received w T from the authorized user. Then, the cloud server calculates and returns to the corresponding set of encrypted documents. Moreover, to reduce the communication cost, the data user may send an optional number l along with the trapdoor T so that the cloud server only sends back top- l files that are most relevant to the search query.

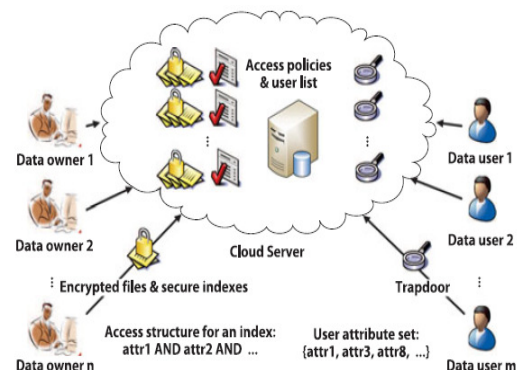


Fig.1 Architecture

Data owner has a collection of data documents to be send to cloud server in the encrypted format. To activate the searching capability over encrypted data, data owner, before sending data, will first build an encrypted searchable manifestation (index), and then outsource both the index and the encrypted document collection to cloud server. To search the document, an authorized user require a corresponding trapdoor through search mechanisms, Upon receiving from data users, cloud server is responsible to search the index and return the corresponding set of encrypted documents. To improve document retrieval accuracy, search result should be ranked by cloud server according to some ranking criteria. Cloud server only sends back top- k documents that are most relevant to the search query. There is one another entity is exist i.e. Unauthorized User. If that Unauthorized user tries to access any data from clod then alert will be generated in the form of mail and message. The alert is given to the authorized person who is owner of that data.

The proposed system presents an authorized keyword search scheme over encrypted cloud data with efficient user revocation in the multi-user multi-data-contributor scenario. It can be realized that fine-grained owner-enforced search authorization by exploiting ciphertext policy attribute-based encryption (CP-ABE) technique. Specifically, the data owner encrypts the index of each file with an access policy created by him, which defines what type of users can search this index. The data user generates the trapdoor independently without relying on an always online trusted authority (TA). The cloud server can search over the encrypted indexes with the trapdoor on a user's behalf, and then returns matching result if and only if the user's attributes associated with the trapdoor satisfy the access policies embedded in the encrypted indexes. We differentiate attributes and keywords in our design. Keywords are actual content of the files while attributes refer to the properties of users. The system only maintains a limited number of attributes for search authorization purpose. Data owners create the index consisting of all keywords in the file but encrypt the index with an access structure only based on the attributes of authorized users, which makes the proposed scheme more scalable and suitable for the large scale file sharing system. In order to further release the data owner from the burdensome user membership management, we use proxy re-

encryption and lazy reencryption techniques to shift the workload as much as possible to the CS, by which our proposed scheme enjoys efficient user revocation. Formal security analysis shows that the proposed scheme is provably secure and meets various search privacy requirements.

The proposed ABKS-UR scheme in the cloud aims to achieve the following functions and security goals:

Authorized keyword search: The secure search system should enable data-owner-enforced search authorization, i.e., only users that meet the owner-defined access policy can obtain the valid search result. Besides achieving finegrained authorization, another challenge is to make the scheme scalable for dynamic cloud environment.

Supporting multiple data contributors and data users: The designed scheme should accommodate many data contributors and data users. Each user is able to search over the encrypted data contributed from multiple data owners.

Efficient user revocation: Another important design goal is to efficiently revoke users from the current system while minimizing the impact on the remaining legitimate users.

Authenticity of search result: To make the proposed authorized keyword search scheme verifiable and enable data user to check the authenticity of the returned search result.

Security goals: Mainly concerned factors with secure search related privacy requirements, and define them as follows.

Keyword semantic security: Since it presents a novel attribute-based keyword search technique, it will formally prove it semantically secure against chosen keyword attack under selective ciphertext policy model.

Trapdoor unlinkability: This security property makes the CS unable to visually distinguish two or more trapdoors even containing the same keyword. Note that the attacker may launch dictionary attack by using public key to generate arbitrary number of indexes with keyword of his choice, and then search these indexes with a particular trapdoor to deduce the underlying keyword in the trapdoor, which is referred to as predicate privacy and it cannot be protected inherently in the PKC-based search scenario.

3.1 MODULE DESCRIPTION

SYSTEM SETUP

At this initial phase, the TA defines the public parameter, and generates PK and MK. The main computation overhead is $3n$ exponentiations in G , one exponentiation in G_1 and one pairing operation on the TA side. The time cost for system setup is very efficient and is linear to the number of attributes in the system.

NEW USER ENROLLMENT

When a new legitimate user wants to join in the system, he has to request the TA to generate the secret key SK, which needs exponentiations in G . The TA also needs one exponentiation in G_1 to generate a new PK component for the user. A data owner may also allow the user to access the dataset by adding him onto the corresponding user list, which incurs one exponentiation in G_1 . It is obvious that the time cost to enroll a new user is proportional to the number of attributes in the system.

SECURE INDEX GENERATION

The size of secure index is constant if the number of attributes is pre-fixed in the system setup phase regardless of the actual number of keywords in a file for both single keyword and conjunctive keyword search scenarios. Moreover, the data owner approximately needs to generate a secure index for a file. Furthermore, we evaluate the practical efficiency of creating secure indexes for 10000 files. It exhibits the expected linearity with the number of attributes in the system. When there exist 30 attributes in the system, the data owner would spend about 8 minutes encrypting the indexes for 10000 files. Note that this computational burden on the data owner is a onetime cost. After all the indexes outsourced to the CS.

TRAPDOOR GENERATION

With the secret key, data user is free to produce the trapdoor of any keyword of

interest, which requires about $2n+1$ group exponentiations in G . Moreover, the experimental result that our proposed authorized keyword search scheme enjoys very efficient trapdoor generation. In accordance with the numerical computation complexity analysis, the trapdoor generation will need more time with the increased number of attributes. Numerical evaluation of ABKS-UR and result verification Operation Computation complexity System Setup New User Enrollment Secure Index Generation Trapdoor Generation Per-index Search Data preparation Search phase Result authentication This is for a new intended keyword search over one authorized dataset. Denotes the per-index search operation.

SEARCH

To search over a single encrypted index, the dominant computation of ABKS-UR is pairing operations, while APKS needs pairing operations. The practical search time of ABKS-UR and APKS on a single secure index with different number of attributes respectively. With the same number of system attributes, ABKS-UR is slightly faster than APKS. Moreover, compared with APKS, ABKS-UR allows users to generate trapdoors independently without resorting to an always online attribute authority, and it has a broader range of applications due to the arbitrarily-structured data search capability. Notice that the search complexity of the scheme will varies a lot for different data users, since the

dataset search authorization only allows users on the user lists to further access the corresponding datasets. Assume that there exist 10000 files and 30 system attributes.

4. CONCLUSION

The first verifiable attribute-based keyword search scheme in the cloud environment is designed, which enables scalable and fine-grained owner-enforced encrypted data search supporting multiple data owners and data users. Compared with existing public key authorized keyword search scheme, this scheme could achieve system scalability and fine-grainedness at the same time. Different from search scheme with predicate encryption, the proposed scheme enables a flexible authorized keyword search over arbitrarily-structured data. In addition, by using proxy reencryption and lazy re-encryption techniques, the proposed scheme is better suited to the cloud outsourcing model and enjoys efficient user revocation. On the other hand, it can make the whole search process verifiable and data user can be assured of the authenticity of the returned search result. The system also formally proves the proposed scheme semantically secure in the selective model.

FUTURE SCOPE

In future, many enhancements can be done like admin option can be enabling to maintain encrypted files over cloud; users can be authorized via OTP and so on. This system is currently work on single cloud, In future is will extended up to sky computing & Provide better security in multi-user systems.

REFERENCES

- [1] W. Sun, S. Yu, W. Lou, Y. T. Hou, and H. Li, "Protecting your right: Attribute-based keyword search with fine-grained ownerenforced search authorization in the cloud," in Proc. IEEE Conf. Comput. Commun., 2014, pp. 226–234.
- [2] S. Yu, C. Wang, K. Ren, and W. Lou, "Achieving secure, scalable, and fine-grained data access control in cloud computing," in Proc. IEEE Conf. Comput. Commun., 2010, pp. 1–9.
- [3] M. Li, S. Yu, Y. Zheng, K. Ren, and W. Lou, "Scalable and secure sharing of personal health records in cloud computing using attribute-based encryption," IEEE Trans. Parallel Distrib. Syst., vol. 24, no. 1, pp. 131–143, Jan. 2013.
- [4] S. Kamara and K. Lauter, "Cryptographic cloud storage," in Proc. 14th Int. Conf. Financial CryptographyData Security, 2010, pp. 136–149.
- [5] D. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in Proc. IEEE Symp. Security Privacy, 2000, pp. 44–55.
- [6] Y. Huang, D. Evans, J. Katz, and L. Malka, "Faster secure twoparty computation using garbled circuits," in Proc. 20th USENIX Conf. Security Symp., 2011, p. 35.
- [7] C. Gentry, "A fully homomorphic encryption scheme," Ph.D. dissertation, Dept. Comput. Sci., Stanford Univ., Stanford, CA, USA, 2009.