IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified

Vol. 7, Issue 4, April 2018

Effective and Experienced Keyword Search over Encrypted Cloud Data

T. Tamililakkiya[1], S.R. Boopathybalan[2]

ME Scholar, Mother Teresa College of Engineering and Technology, Mettusalai, Pudukkottai, India[1]

Assistant Professor, Mother Teresa College of Engineering and Technology, Mettusalai, Pudukkottai, India[2]

Abstract: Cloud computing enables the paradigm of data outsourcing. However, to protect data privacy, sensitive cloud data have to be encrypted before outsourced to the commercial public cloud, which makes effective data utilization service is a challenging task. Even though searchable encryption techniques allow users to securely search over encrypted data through keywords, they support only Boolean search and are not yet sufficient to meet the effective data utilization need that is inherently demanded by large number of users and huge amount of data files in cloud. Hence it is necessary to allow multiple keywords in the search request and return documents in the order of their relevance to the keywords. The works on searchable encryption focused on gram based keyword search or special symbol based keyword search or symbol based keyword search and the result produced by them are rarely sorted. An effective method proposed for this challenging problem is efficient search over encrypted cloud data. This method establishes a set of strict privacy requirements for such a secure cloud data utilization system through MRSE. Among various multi-keyword semantics, this method chooses the efficient similarity measure of coordinate matching. Then according to Top K Query method the sorted results are produced. The privacy is preserved by the chunk of data stored in a various server‘s. Then further improvisation is taken to introduce low overhead on computation and communication in future.

Keywords: Encryptions techniques, PEKS, PIR, PHR, HIPAA, MRSE.

I. INTRODUCTION

Consider a cloud-based healthcare information system that hosts outsourced Personal Health Records (PHRs) from various healthcare providers. The PHRs are encrypted in order to comply with privacy regulations like HIPAA. In order to facilitate data use and sharing, it is highly desirable to have a Searchable Encryption (SE) scheme which allows the cloud service provider to search over encrypted PHRs on behalf of the authorized users (such as medical researchers or doctors) without learning information about the underlying plaintext. Note that the context we are considering supports private data sharing among multiple data providers and multiple data users. Therefore, SE schemes in the private-key setting , which assume that a single user who searches and retrieves his/her own data, are not suitable. On the other hand, Private Information Retrieval (PIR) protocols, which allow users to retrieve a certain data-item from a database which publicly stores data without revealing the data-item to the database administrator, are also not suitable, since they require the data to be publicly available. In order to tackle the keyword search problem in the cloud-based healthcare information system scenario, we resort to Public-Key Encryption with Keyword Search (PEKS) schemes, which is firstly propose. In a PEKS scheme, a ciphertext of the keywords called "PEKS ciphertext" is appended to an encrypted PHR. To retrieve all the encrypted PHRs containing a keyword, say "Diabetes", a user sends a "trapdoor" associated with a search query on the keyword "Diabetes" to the cloud service provider, which selects all the encrypted PHRs containing the keyword "Diabetes" and returns them to the user while without learning the underlying PHRs. However, the solution in as well as other existing PEKS schemes extending from only supports equality queries. Set intersection and Meta keywords1 can be used for conjunctive keyword search. However, the approach based on set intersection leaks extra information to the cloud server beyond the results of the conjunctive query, whilst the approach using Meta keywords require 2 m meta keywords to accommodate all the possible conjunctive queries for m keywords. In order to address the above deficiencies in conjunctive keyword search, schemes such as the ones in were put forward in the public-key setting. Ideally, in the practical applications, search predicates (i.e., access policies or structures) should be expressive such that they can be expressed as conjunction, disjunction or any Boolean formulas2 of keywords. In the above cloud-based healthcare system, to find the relationship between diabetes and age or weight, a medical researcher may issue a search query with an access structure.

II. LITERATURE REVIEW

HONGWEI LI et.al, In mobile cloud computing, a fundamental application is to outsource the mobile data to external cloud servers for scalable data storage. The outsourced data, however, need to be encrypted due to the privacy and

DOI 10.17148/IJARCCE.2018.7464

confidentiality concerns of their owner. This results in the distinguished difficulties on the accurate search over the encrypted mobile cloud data. To tackle this issue, in this paper, we develop the searchable encryption for multi-keyword ranked search over the storage data. Specifically, by considering the large number of outsourced documents (data) in the cloud, we utilize the relevance score and k-nearest neighbor techniques to develop an efficient multi-keyword search scheme that can return the ranked search results based on the accuracy.

Zhihua Xia et.al, due to the increasing popularity of cloud computing, more and more data owners are motivated to outsource their data to cloud servers for great convenience and reduced cost in data management. However, sensitive data should be encrypted before outsourcing for privacy requirements, which obsoletes data utilization like keyword-based document retrieval. In this paper, we present a secure multi-keyword ranked search scheme over encrypted cloud data, which simultaneously supports dynamic update operations like deletion and insertion of documents. Specifically, the vector space model and the widely-used TF×IDF model are combined in the index construction and query generation. We construct a special tree-based index structure and propose a “Greedy Depth-first Search” algorithm to provide efficient multi-keyword ranked search.

V. Amudha et.al, In case of sharing the group of documents in cloud storage the document owners uses the encrypted key for secured sharing. For a single owner, the document contains one trapdoor key then the user can download it using the key, but for multiple owners this concept does not works. To overcome this practical problem this paper proposes a solution for KASE in case of federated clouds. This is a practice of interconnecting the cloud computing environment of two or more service providers for the purpose of load balancing traffic.

Zhangjie Fu et.al, Keyword-based search over encrypted outsourced data has become an important tool in the current cloud computing scenario. The majority of the existing techniques are focusing on multi-keyword exact match or single keyword fuzzy search. However, those existing techniques find less practical significance in real-world applications compared with the multi-keyword fuzzy search technique over encrypted data. The first attempt to construct such a multi-keyword fuzzy search scheme used locality-sensitive hashing functions and Bloom filtering to meet the goal of multi-keyword fuzzy search. Nevertheless, effective for a one letter mistake in keyword but was not effective for other common spelling mistakes.

Wei Zhang et.al, with the advent of cloud computing, it has become increasingly popular for data owners to outsource their data to public cloud servers while allowing data users to retrieve this data. For privacy concerns, secure searches over encrypted cloud data have motivated several research works under the single owner model. However, most cloud servers in practice do not just serve one owner; instead, they support multiple owners to share the benefits brought by cloud computing. In this paper, we propose schemes to deal with Privacy preserving Ranked Multi-keyword Search in a Multi-owner model (PRMSM). To enable cloud servers to perform secure search without knowing the actual data of both keywords and trapdoors, we systematically construct a novel secure search protocol.

M. Veerabrahma chary et.al, Observing the view of cloud computing, it has become augmenting popular for data owners to outside supplier their information to public cloud servers while allowing data users to regain this data. To relate to seclusion, safe searches over encrypted cloud data have provoke more research works under the sole owner model. However, most cloud servers in practice do not just Serve unique owner; instead, they support multiple owners to share the benefits brought by cloud computing. In this paper, we suggest -To keep safe the secrecy and several owner model search several keywords and Ranked. To make possible cloud servers to execute safe to look omission knowing the real information of both keywords and trapdoors, To keep alive the privacy of related scores between keywords and files and rank the search result, we suggest a novel Additive Order and Privacy Preserving Function family and dynamic hidden key creation rule and a new data user to establish as genuine rule.

III. SYSTEM ANALYSIS

A. Existing System

This straightforward approach apparently provides fuzzy keyword search over the encrypted files while achieving search privacy using the technique of secure trapdoors. However, this approaches serious efficiency disadvantages. The simple enumeration method in constructing fuzzy key-word sets would introduce large storage complexities, which greatly affect the usability.

For example, the following is the listing variants after a substitution operation on the first character of keyword
CASTLE: {AASTLE, BASTLE, DASTLE, YASTLE, ZASTLE}.

B. Proposed System

- **Wildcard – Based Technique**

DOI 10.17148/IJARCCE.2018.7464

- **Gram - Based Technique**
- **Symbol – Based Tree – traverse Search Scheme**

Wildcard – Based Technique: In the above straightforward approach, all the variants of the keywords have to be listed even if an operation is performed at the same position. Based on the above observation, we proposed to use an wildcard to denote edit operations at the same position. The wildcard-based fuzzy set edits distance to solve the problems.

For example, for the keyword CASTLE with the pre-set edit distance 1, its wildcard based fuzzy keyword set can be constructed as

SCASTLE, 1 = {CASTLE, *CASTLE,*ASTLE, C*ASTLE, C*STLE, CASTL*E, CASTL*, CASTLE*}.

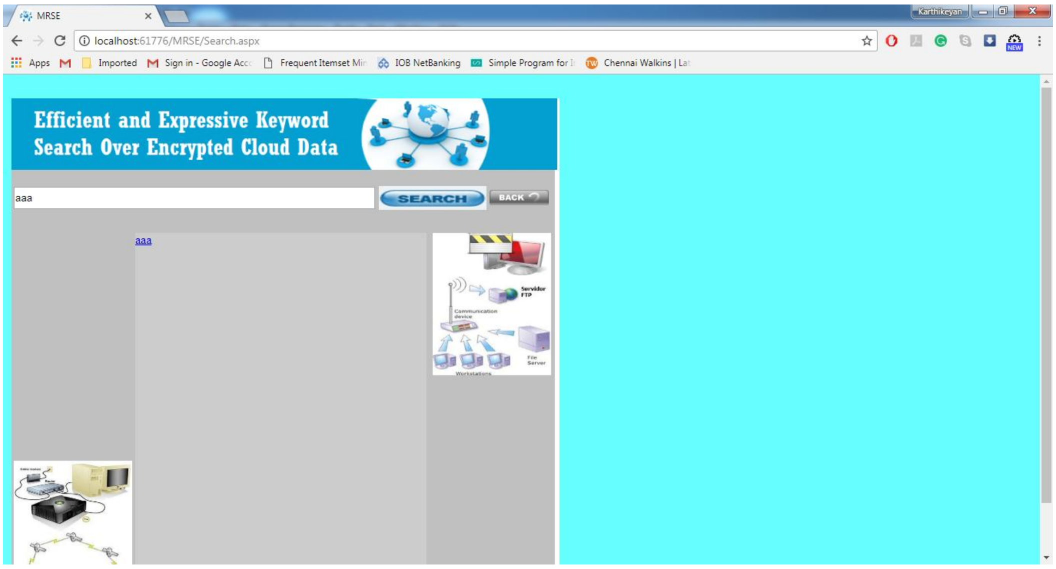
Edit Distance:

- Substitution
- Deletion
- Insertion
- **Substitution** : changing one character to another in a word;
- **Deletion** : deleting one character from a word;
- **Insertion**: inserting a single character into a word.

Gram – Based Technique: Another efficient technique for constructing fuzzy set is based on grams. The gram of a string is a **substring** that can be used as a signature for efficient approximate search. While gram has been widely used for constructing inverted list for approximate string search, we use gram for the matching purpose. We propose to utilize the fact that any primitive edit operation will affect at most one specific character of the keyword, leaving all the remaining characters untouched. In other words, the relative order of the remaining characters after the primitive operations is always kept the same as it is before the operations.
For example, the gram-based fuzzy set SCASTLE, 1 for keyword CASTLE can be constructed as

{CASTLE, CSTLE, CATLE, CASLE, CASTE, CASTL, ASTLE}.

Symbol – Based Tree – traverse Search Scheme: To enhance the search efficiency, we now propose a symbol-based tree-traverse search scheme, where a **multi-way tree** is constructed for storing the fuzzy keyword set over a finite symbol set. The key idea behind this construction is that all trapdoors sharing a common prefix may have common nodes. The root is associated with an empty set and the symbols in a trapdoor can be recovered in a search from the root to the leaf that ends the trapdoor. All fuzzy words in the trie can be found by a depth-first search.
In this section, we consider a natural extension from the previous single-user setting to multi-user setting, where a data owner stores a file collection on the cloud server and allows an arbitrary group of users to search over his file collection.



MRSE

localhost:61776/MRSE/Search.aspx

Efficient and Expressive Keyword Search Over Encrypted Cloud Data

aaa

Fig. 1 MRSE Keyword search

IV. IMPLEMENTATION

A. System Model:

This project considers a cloud data system consisting of data owner, data user and cloud server. Given a collection of n encrypted data files C = (F1, F2, . . . , FN) stored in the cloud server, a predefined set of distinct keywords W = {w1, w2, ...,wp}, the cloud server provides the search service for the authorized users over the encrypted data C. We assume the authorization between the data owner and users is appropriately done. An authorized user types in a request to selectively retrieve data files of his/her interest. The cloud server is responsible for mapping the searching request to a set of data files, where each file is indexed by a file ID and linked to a set of keywords. The fuzzy keyword search scheme returns the search results according to the following rules: 1) if the user's searching input exactly matches the pre-set keyword, the server is expected to return the files containing the keyword1; 2) if there exist typos and/or format inconsistencies in the searching input, the server will return the closest possible results based on pre-specified similarity semantics.

B. Threat Model:

This module consider a semi-trusted server. Even though data files are encrypted, the cloud server may try to derive other sensitive information from users' search requests while performing keyword-based search over *C*. Thus, the search should be conducted in a secure manner that allows data files to be securely retrieved while revealing as little information as possible to the cloud server. In this paper, when designing fuzzy keyword search scheme, we will follow the security definition deployed in the traditional searchable encryption.

More specifically, it is required that nothing should be leaked from the remotely stored files and index beyond the outcome and the pattern of search queries.

C. Design Goals:

In this, we address the problem of supporting efficient yet privacy-preserving fuzzy keyword search services over encrypted cloud data. Specifically, we have the following goals:

- To explore new mechanism for constructing storage efficient fuzzy keyword sets;
- To design efficient and effective fuzzy search scheme based on the constructed fuzzy keyword sets;
- To validate the security of the proposed scheme.

V. CONCLUSION & FUTURE WORK

In order to allow a cloud server to search on encrypted data without learning the underlying plaintexts in the public key setting, Boneh proposed a cryptographic primitive called Public-key Encryption with Keyword Search (PEKS). Since then, considering different requirements in practice, e.g., communication overhead, searching criteria and security enhancement, various kinds of searchable encryption systems have been put forth. However, there exist only a few public-key searchable encryption systems that support expressive keyword search policies, and they are all built from the inefficient composite-order groups. In this work, we focused on the design and analysis of public-key searchable encryption systems in the prime-order groups that can be used to search multiple keywords in expressive access policies. Based on a large universe key-policy attribute-based encryption scheme given, we presented an expressive searchable encryption system in the prime order groups which supports expressive access structures expressed in any monotonic Boolean formulas. Also, we proved its security in the standard model, and analyzed its efficiency using the Charm framework.

As our future work, this project will explore supporting other multi keyword semantics (e.g., weighted query) over encrypted data, integrity check of rank order in search result and privacy guarantees in the stronger threat model. As our future work, is to concentration the encrypted data of semantic keyword search in order that can confront with the more sophisticated search.

REFERENCES

[1] C. Wang, Q. Wang, K. Ren, and W. Lou, ―Privacy-Preserving Public Auditing for Data Storage Security in Cloud Computing,‖ Proc. IEEE INFOCOM, 2010.

[2] C. Wang, Q. Wang, K. Ren, N. Cao, and W. Lou, ―Toward Secure and Dependable Storage Services in Cloud Computing,‖ IEEE Trans. Services Computing, vol. 5, no. 2, pp. 220-232, Apr.-June 2012

[3] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, ―Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data,‖ Proc. IEEE INFOCOM, , jan, 2014.

[4] E.-J. Goh, ―Secure Indexes,‖ Cryptology ePrint Archive, http:// eprint.iacr.org/2003/216. 2003

[5] N. Cao, S. Yu, Z. Yang, W. Lou, and Y. Hou, ―LT Codes-Based Secure and Reliable Cloud Storage Service,‖ Proc. IEEE INFOCOM, pp. 693-701, 2012.

[6] S. Kamara and K. Lauter, ―Cryptographic Cloud Storage,‖ Proc. 14th Int'l Conf. Financial Cryptograpy and Data Security, Jan. 2010.

[7] Singhal, ―Modern Information Retrieval: A Brief Overview,‖ IEEE Data Eng. Bull., vol. 24, no. 4, pp. 35-43, Mar. 2001.

DOI 10.17148/IJARCCE.2018.7464

[8] Y.-C. Chang and M. Mitzenmacher, ―Privacy Preserving Keyword Searches on Remote Encrypted Data,‖ Proc. Third Int‗l Conf. Applied Cryptography and Network Security, 2005.

[9] R. Curtmola, J.A. Garay, S. Kamara, and R. Ostrovsky, ―Searchable Symmetric Encryption: Improved Definitions and Efficient Constructions,‖ Proc. 13th ACM Conf. Computer and Comm. Security (CCS ‗06), 2006.

[10] D. Boneh, G.D. Crescenzo, R. Ostrovsky, and G. Persiano, ―Public Key Encryption with Keyword Search,‖ Proc. Int‗l Conf. Theory and Applications of Cryptographic Techniques (EUROCRYPT), 2004.

[11] M. Bellare, A. Boldyreva, and A. ONeill, ―Deterministic and Efficiently Searchable Encryption,‖ Proc. 27th Ann. Int‗l Cryptology Conf. Advances in Cryptology (CRYPTO ‗07), 2007.

[12] J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou, ―Fuzzy Keyword Search Over Encrypted Data in Cloud Computing,‖ Proc. IEEE INFOCOM, Mar. 2010.

[13] D. Boneh, E. Kushilevitz, R. Ostrovsky, and W.E.S. III, ―Public Key Encryption That Allows PIR Queries,‖ Proc. 27th Ann. Int‗l Cryptology Conf. Advances in Cryptology (CRYPTO ‗07), 2007.

[14] P. Golle, J. Staddon, and B. Waters, ―Secure Conjunctive Keyword Search over Encrypted Data,‖ Proc. Applied Cryptography and Network Security, pp. 31- 45, 2004.

[15] L. Ballard, S. Kamara, and F. Monrose, ―Achieving Efficient Conjunctive Keyword Searches over Encrypted Data,‖ Proc. Seventh Int‗l Conf. Information and Comm. Security (ICICS ‗05), 2005.

DOI 10.17148/IJARCCE.2018.7464