International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 04 Issue: 11 | Nov -2017 www.irjet.net p-ISSN: 2395-0072

A Survey on Efficient Privacy-Preserving Ranked Keyword Search Method

Athira Sankar[1], Soumya Murali[2]

PG Scholar[1], Assistant Professor[2]
[1,2] Sree Buddha College of Engineering, Kerala Technological University, India.

Abstract - *As terabytes of data has been increasing, cloud become an important storage to outsource sensitive information for privacy preserving. Cloud computing is an emerging model in the IT infrastructure which organizes huge resource of computing, storage and applications, manipulating, configuring and accessing application online with great efficiency and minimum economic overhead. An efficient method to reduce information leakage is by data encryption. So the data Owners outsource the document in an encrypted format for privacy preserving. In, this paper presents a hierarchical clustering method for fast search semantics. The hierarchical clustering, cluster the document into sub cluster until the constraint on the maximum size of cluster is reached. In search phase, this method can reach a linear computational complexity against an exponential size increase of document collection. In order to verify the search result; a hash sub-tree is designed for authentication.*

1. Introduction

Cloud has become one of the significant fields in the IT enterprise due to the increasing in huge amount of data. Cloud computing is a paradigm, where large pool of system are connected in private/public network for providing infrastructure for application, data and cloud storage.

Cloud server provider (CSP) provides network services; infrastructure application in the cloud. CSP's has different forms of services: Software as a service (SAAS), Platform as a service (PAAS), and Infrastructure as a service (IAAS).

To reduce information leakage, data encryption is a traditional way. Data encryption protects data confidentiality as it is stored on a computer system and transmitted using the internet. Many cryptographic techniques are used for data encryption.

In this paper [1], a vector space model is used. A vector space model is an algebraic model for representing text document as vectors of identifiers. Vector space model is used for ranking, indexing, information filtering, and information retrieval. Here, the documents and queries are represented by vector that is in invertible matrix form. One of the known scheme for representing vector space model is by tf-idf weighting. Tf is the term frequency that the number of words occurs in a document. Idf is the inverse document frequency that measure how much information the word provides, whether the term is common across all documents.

The objective of the work is to investigate the problem of maintaining the close relationship between different plain documents over an encrypted domain. The second objective is to propose the MRSE-HCI architecture to speed up server-side searching phase. The third objective is to design a search strategy to improve the rank privacy. Finally, a Merkle hash tree and cryptographic signature to authenticated tree structure

The rest of the paper is organized as follows, Section II discusses the Related Works on privacy preserving keyword search and. Section III describes the conclusion.

2. Literature Review

Nowadays, more and more people are motivated to outsource their local data to public cloud servers for great convenience and reduced costs in data management. This section describes the privacy preserving schemes on multi keyword search and ranking method.

Slawomir Grzonkowski et al [2] propose a security analysis for authentication protocol in the CE cloud services. The current weakness of the protocol, overcome by Zero Knowledge Proof (ZKP). ZKP technique is used for protecting the user password. Here there is an alternative protocol for ZKP that is SeDici 2.0 is also described. SeDici 2.0 is a third party trusted (TTP) protocol based ZKP. The main aim of the ZKP is to deliver a better anti- phishing solution. ZKP technique addresses the problem of phishing and mutual authentication.

In the case of SeDici 2.0, the client is able to communicate with the consumer services and the authentication services. SeDici 2.0 does not need any physical token. The proposed protocol can provide a suitable solution for CE-based cloud authentication services.

Searchable Symmetric Encryption (SSE) [3] allows a party to outsource the storage of its data to another party in a private manner. SSE is a collection of polynomial time algorithm, such as:

- K ← Gen(I^{Y}): It is is a probabilistic algorithm run by the client to generate a key.
- (I, c) ← Enc(K, D): is a probabilistic algorithm that takes a document collection D and a key K as input
- t ← Trpdr(K, w): is a deterministic algorithm run by the client to generate a trapdoor t for a keyword w, where w ∈ Δ.
- R ← Search(I, t): is a deterministic algorithm run by the server to search for documents in D that contain a keyword w.
- di ← Dec(K, ci): is a deterministic algorithm run by the client to decrypt a single encrypted document ci.

SSE scheme is of three types: SSE-1, SSE-2. MULTI-SSE. SSE-1 is used for building the index. The technical issues of SSE-1 are large address space and small number entries.SSE-2 is similar as SSE-1 .SSE-2 is used for pre-processing and padding the index. MULTI-SSE is used for indexes and trapdoors require same security notions as single-user SSE.

In this paper [4], a hierarchical clustering for cipher text search in a big data environment is proposed. The hierarchical clustering clusters the document based on minimum similarity threshold. Then partitioning the resultant cluster into sub-cluster until a constraint of maximum size of cluster is reached.

During the search phase, this approach can reach a linear computational complexity against exponential size of document collection. The hierarchical clustering is used for better clustering result. Thus large collection of document can be easily clustered thereby, improving efficiency of the search. The proposed system brings improvement in search efficiency, rank security, and the similarity between retrieved documents.

Secure conjunctive keyword ranked search over encrypted cloud data [5] choose the principle of coordinate matching that is used to identify the similarity between query and data document. The algorithm used in the conjunctive keyword search is pailler cryptosystem, Rijndael algorithm, cosine similarity search.

Pailler cryptosystem is an homomorphism algorithm. It is used to encrypt n decrypt file content. Rijndael algorithm is a symmetric AES. It is used to convert secret key into bytes. Cosine similarity search is a nearest neighbor search that is used to identify the top k relevant document of the query.

N Cao et al [6] propose a basic idea for the MRSE based on secure inner product computation. In this paper, MRSE scheme is applied to provide similarity ranking effective.

The proposed system algorithms are:

1) Algorithm to provide multi keyword ranked search.
2) The secure kNN algorithm is used to encrypt the index and query vectors.
3) A Greedy Depth-first Search algorithm based on index tree.
4) The LSH algorithm is suitable for similar search but cannot provide exact ranking.
5) {I′s ; ci} ← GenUpdateInfo (SK; Ts; i; up type)), this algorithm generates the update information {I′s ; ci} which will be sent to the cloud server.

Secure ranked keyword search over cloud data [7] tackles the problems of enabling searchable encryption system. Ranking is calculated using TF*IDF rule.

Efficient ranked searchable symmetric encryption scheme are using order preserving symmetric encryption (OPSE). The OPSE is a deterministic encryption scheme where the numerical ordering of the plaintexts gets preserved by the encryption function.

Here proposed a searching method to improve the efficiency of ranked keyword search algorithms. The efficient one-to many order preserving mapping function, which allows the effective ranking to be designed. This kind of techniques has the ability to categorize, and search large collections of unstructured text on a conceptual basis.

In this paper [8] propose a novel approach, called multi keyword query encrypted data (MKQE). In MKQE, adopt inner product similarity to quantitatively evaluate the coordinate matching. MKQE defines an index vector for each file based on the keywords it contains. Two invertible matrices and a bit vector are also used for the index vector encryption and the trapdoor generation. In MKQE, when a multi keyword query comes, a query vector based on the set of requesting keywords is constructed.

3. Conclusion

There are many privacy preserving keyword search method to provide data confidentiality and security. Among that, MRSE with hierarchical clustering method is efficient privacy preserving keyword search method. Hierarchical clustering method gives better and efficient result. MRSE-HCI architecture is used to adapt data explosion, online information retrieval and semantic search.

At the same time Merkle hash tree is used for authenticating the index tree structure. Merkle hash is data structure used to store large data's.

References

[1] Chi Chen, Xiaojie Zhu, Peisong Shen, Jiankun Hu," An efficient privacy preserving ranked Keyword search method", IEEE transactions on parallel and distributed systems, vol. 27, no. 4, april 2016

[2] S. Grzonkowski, P. M. Corcoran, and T. Coughlin, "Security analysis of authentication protocols for next-generation mobile and CE cloud services," in Proc. IEEE Int. Conf. Consumer Electron., 2011, Berlin, Germany, 2011, pp. 83–87.

[3] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: improved definitions and efficient constructions," in Proc. 13th ACM Conf. Comput. Commun. Security, Alexandria, Virginia, 2006, pp. 79–88.

[4] C. Chen, X. J. Zhu, P. S. Shen, and J. K. Hu, "A hierarchical clustering method For big data oriented ciphertext search," in Proc. IEEE INFOCOM, Workshop on Security and Privacy in Big Data, Toronto, Canada, 2014, pp. 559–564.

[5] P. Golle, J. Staddon, and B. Waters, "Secure conjunctive keyword search over encrypted data," in Proc. Proc. 2nd Int. Conf. Appl. Cryptography Netw. Security, Yellow Mt, China, 2004, pp. 31–45.

[6] N. Cao, C. Wang, M. Li, K. Ren, and W. J. Lou, "Privacy-preserving multi-keyword ranked search over encrypted cloud data," in Proc. IEEE INFOCOM, Shanghai, China, 2011, pp. 829–837.

[7] C. Wang, N. Cao, J. Li, K. Ren, and W. J. Lou, "Secure ranked keyword search over encrypted cloud data," in Proc. IEEE 30th Int. Conf. Distrib. Comput. Syst., Genova, ITALY, 2010, pp. 253–262.

[8] R. X. Li, Z. Y. Xu, W. S. Kang, K. C. Yow, and C. Z. Xu, "Efficient Multi-keyword ranked query over encrypted data in cloud computing, Futur. Gener. Comp. Syst., vol. 30, pp. 179–190, Jan. 2014.