International Journal on Future Revolution in Computer Science & Communication Engineering
Volume: 4 Issue: 7
ISSN: 2454-4248
36 – 39

Semantic Search Approach in Cloud

Gurmeet Kaur Saini, Akhil Chaurasia, Rajni Rani
Assistant Professor,
Chandigarh University
Gurmeetsaini02@gmail.com
akhilchaurasia47@gmail.com
rajnigarg92@gmail.com

Abstract: With the approach of cloud computing, more and more information data are distributed to the public cloud for economic savings and ease of access. But, the encryption of privacy information is necessary to guarantee the security. Now a days efficient data utilization, and search over encrypted cloud data has been a great challenge. Solution of existing methods depends only on the keyword of submitted query and didn't examine the semantics of keyword. Thus the search schemes are not intelligent and also omit some semantically related documents. To overcome this problem, we propose a semantic expansion based similar search solution over encrypted cloud data. The solution of this method will return not only the exactly matched files, but also the files including the terms semantically related to the query keyword. In this scheme, a corresponding file metadata is constructed for each file. After this, both the encrypted file metadata set and file collection are uploaded to the cloud server. With the help of metadata set file, the cloud server maintains the inverted index and create semantic relationship library (SRL) for the keywords set. After receiving a query request from user , this server firstly search out the keywords that are related to the query keyword according to SRL. After this, both the query keyword and the extensional words are used to retrieve the files to fulfill the user request. These files are returned in order according to the total relevance score. Our detailed security analysis shows that our method is privacy-preserving and secure than the previous searchable symmetric encryption (SSE) security definition. Experimental evaluation demonstrates the efficiency and effectives of the scheme.

Keywords: *Secure; Semantic expansion, Rank, SRL(Semantic relationship library),Cloud data.*

*****

Introduction

The cloud makes it possible for you to access your information from anywhere at any time. While a traditional computer setup requires you to be in the same location as your data storage device, the cloud takes away that step. The cloud removes the need for you to be in the same physical location as the hardware that stores your data. Your cloud provider can both own and house the hardware and software necessary to run your home or business applications. This is especially helpful for businesses that cannot afford the same amount of hardware and storage space as a bigger company. Small companies can store their information in the cloud, removing the cost of purchasing and storing memory devices.

Additionally, because you only need to buy the amount of storage space you will use, a business can purchase more space or reduce their subscription as their business grows or as they find they need less storage space. Each provider serves a specific function, giving users more or less control over their cloud depending on the type. When you choose a provider, compare your needs to the cloud services available. The information housed on the cloud is often seen as valuable to individual with malicious intent. There is a lot of personal information and potentially secure data that people store on their computers, and this information is now being transferred to the cloud. This makes it critical for you to understand the security measures that your cloud provider has in place, and it is equally important to take personal precautions to secure your data. The multi keyword retrieval over encrypted cloud data achieves high security and privacy.

Cloud makes it possible for us to access our information from anywhere at any time. At any time it removes the need for us to be in the same physical location due to the following features:

SaaS (Software as a Service)

It provides all the functions of a sophisticated traditional application to many customers and often thousands of users, but through a Web browser, not a locally- installed application.

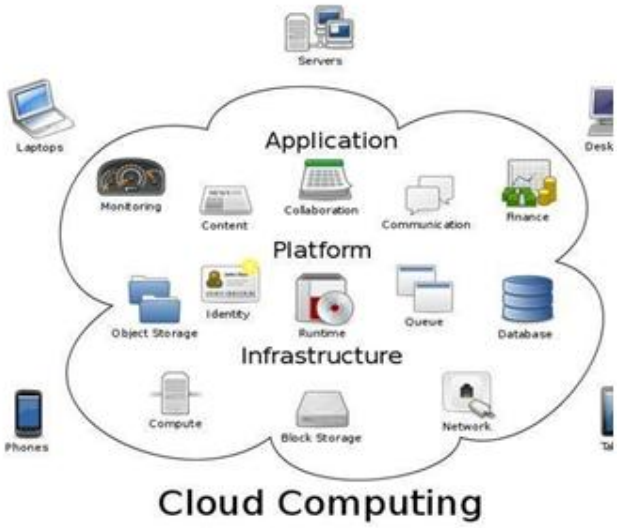
PaaS(Platform as a Service)

Delivers virtualized servers on which customers can run existing applications or develop new ones without having worry about maintaining operating systems, server hardware, load balancing or computing capacity.

IaaS(Infrastructure as a Service)

Delivers utility computing capability, typically as raw virtual servers, on demand that customers configure and manage.

 ISSN: 2454-4248
36 – 39



Cloud Computing enables cloud customers to enjoy the on-demand high quality applications and services from a centralized pool of configurable computing resources. This new computing model can relieve the burden of storage management, allow universal data access with independent geographical locations, and avoid capital expenditure on hardware, software, and personnel maintenances, etc [1]. As cloud computing becomes mature, lots of sensitive data is considered to be centralized into the cloud servers, e.g. personal health records, secret enterprise

data, government documents, etc [1,2]. The straightforward solution to protect data privacy is to encrypt sensitive data before being outsourced. Unfortunately, data encryption, if not done appropriately, may reduce the effectiveness of data utilization. Typically, a user retrieves files of interest to him/her via keyword search instead of retrieving back all the files. Such keyword based search technique has been widely used in our daily life, e.g. Google plaintext keyword search. However, the technologies are invalid after the keywords are encrypted.

LITERATURE SURVEY

Zhangjie Fu et. al., uses an effective approach to solve the problem of multi-keyword ranked search over encrypted cloud data supporting synonym queries. The main contribution of this paper is summarized in two aspects: multi-keyword ranked search to achieve more accurate search results and synonym-based search to support synonym queries. Extensive experiments on real-world dataset were performed to validate the approach, showing that the proposed solution is very effective and efficient for multi-keyword ranked searching in a cloud environment. [1].

J. Li, Q. Wang et. al.,uses the Fuzzy keyword search method that enhances system usability by returning the matching files containing exact match of the predefined keywords or the closest possible matching files based on keyword similarity semantics, when *exact* match fails. They exploit edit distance to quantify keywords similarity and develop an advanced technique on constructing fuzzy keyword sets, which greatly reduces the storage and representation overheads [2].

C. Wang, N. Cao et. al., define and solve the problem of effective yet secure ranked keyword search over encrypted cloud data. Ranked search greatly enhances system usability by returning the matching files in a ranked order regarding to certain relevance criteria (e.g., keyword frequency). To achieve more practical performance, they propose a definition for ranked searchable symmetric encryption, and give an efficient design by properly utilizing the existing cryptographic primitive, order- preserving symmetric encryption (OPSE). Thorough analysis shows that this solution provides security guarantee compared to previous SSE schemes.[3]

N. Cao, C. Wang et. al., propose a basic idea for the MRSE(multi-keyword ranked search over encrypted data) based on secure inner product computation, and then give two significantly improved MRSE schemes to achieve various stringent privacy requirements in two different threat models. To improve search experience of the data search service, we further extend these two schemes to support more search semantics. Thorough analysis investigating privacy and efficiency guarantees of proposed schemes is given.[4]

PROBLEM FORMULATION

Existing search approaches cannot accommodate such requirements like ranked search, semantics- based search etc. The ranked search enables cloud customers to find the most relevant information quickly. Ranked search can also reduce network traffic as the cloud server sends back only the most relevant data. In the real search scenario, it is quite common that cloud customers' searching input might be the synonyms of the predefined keywords, not the exact or fuzzy matching keywords due to the possible synonym substitution (reproduction of information content), such as commodity and goods, and/or her/his lack of exact knowledge about the data. The existing searchable encryption schemes support only exact or fuzzy keyword search. **System model**

 ISSN: 2454-4248
36 – 39

We consider the system model involving three different entities: data owner, data user and cloud server, as illustrated in Figure 1.

Data owner uploads a collection of n text files F ={F1, F2, F3,Fn} in encrypted form C, together with the encrypted metadata set, to the cloud server. Note that, a corresponding file metadata is constructed for each file. Each file in the collection is encrypted with common symmetric encryption algorithm, e.g. AES.

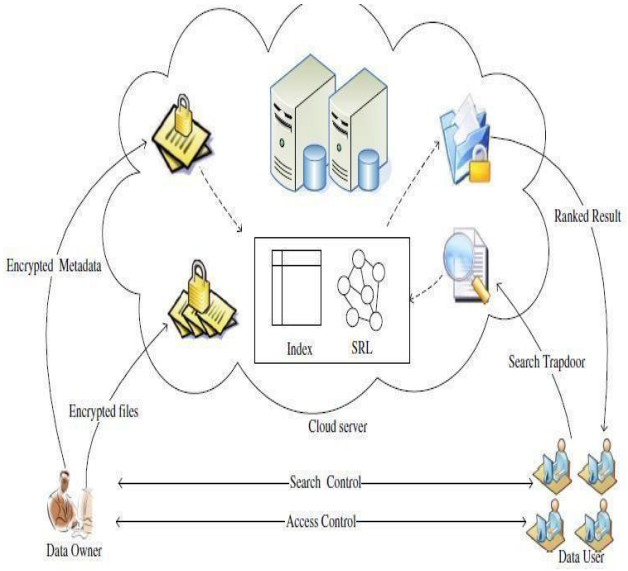


Fig: Framework of the semantic expansion based similar search over encrypted cloud data.

Cloud server first constructs the index and SRL using the metadata set provided by data owner, thus reduce the computing burden on owner, e.g. index creating. Upon receiving the request Tw, the cloud server automatically expands the query keyword based on SRL. Then the server searches the index, and returns the matching files to the user in order. Finally, the access control mechanism, which is out of the scope of this paper, is employed to manage the capability of the user to decrypt the received files.

Design goals

To enable effective and secure ranked semantic expansion search over outsourced cloud data under the aforementioned model, our mechanism should achieve the following design goals.

1) **Ranked semantic expansion search**: To design a similar search scheme that supports semantic search over encrypted cloud data by expanding the query keyword upon semantic relationship of terms, which finally returns the retrieved files in order.

2) **Security guarantee**: To prevent cloud server from learning the plaintext of the data files and keywords. Compared to the existing SSE schemes, the scheme should achieve the as-strong-as possible security strength.

3) **Efficiency:** To achieve the above goals with minimum communication and computation overhead.

Semantic query expansion

In the domain of plaintext retrieval, automatic query extension has been a technique to improve the recall and precision of retrieval for a long time [20]. It uses the semantically related words to expand the particular query, and makes the query request more satisfy the users intent. The key step of semantic query expansion is to find out the semantic relationship between the keywords. Some researchers utilized readily available corpus independent knowledge models [21], e.g. WordNet, EuroWordNet, and some others dynamically constructed the semantic relationship from the document collection by the technologies such as term clustering [22,23], and mutual information model [24-26]. Among these technologies, mutual information model is widely used [24,26-29]. Refer to the formula used in [26], which adopted the mutual information model to implement semantic search in web. The mutual information I(x, y) is defined as $I(x,y) \equiv \log 2 / p(x)\, p(y)$

Here P(x, y) is the probability of observing x and y together. p(x) and p(y) are the probabilities of observing x and y independently in the collection. The higher the semantic relationship between x and y is, the larger the co-occurrence degree is, and consequently the larger the mutual information I(x, y) is. Then normalize the mutual information into a value of relationship in interval [0, 1]. The semantic relationship library will be constructed as a weighted graph structure showed in Figure 3.

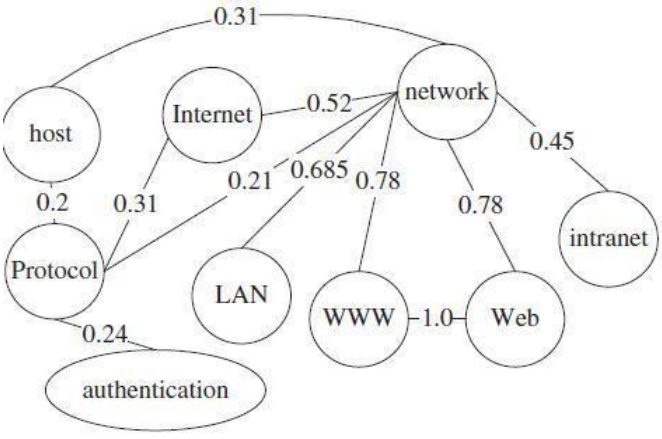


Fig: An example of semantic relationship library.

Order-preserving Encryption (OPE)

The OPE is a deterministic encryption scheme, whose encryption function preserves the numerical ordering in plaintext-space [31,32]. any order- preserving function can be defined as a

 ISSN: 2454-4248
36 – 39

combination of M out of N ordered items, which can be calculated by (N/M).

The adversary has to execute exhaustive enumeration, namely searching over all the possible combination, to break the encryption. So the number of combination, which is maximized when M= N/2, should be large enough to ensure the security.

File metadata

A piece of file-metadata is constructed for each file. The file-metadata consists of the file ID, keywords, and the relevance scores of keywords in response to the file. If file Fi contains keyword wj, a tuple wj, sji is insert into metadata M(Fi), where sji represents the relevance score of keyword wj response to file Fi. All of the file metadata constitute metadata set, which is shown in Figure 4.

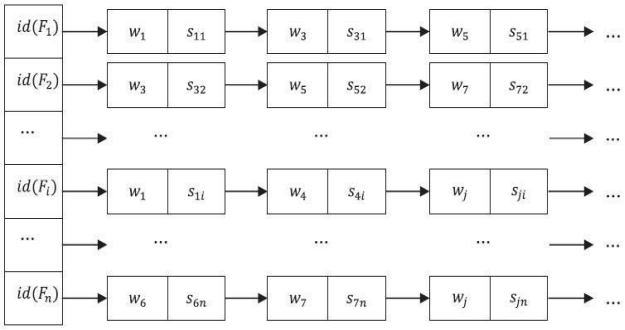


Fig 4: Example of meta data

V. CONCLUSIONS

Retrieving the encrypted cloud data based on the customer needs is the challenging one, and also the retrieved data does not fulfil the customer. In this paper we use the Vector Space to retrieve the encrypted data from the cloud based on the Scoring. Scoring is a natural way to weight the relevance. Based on the relevance score, files can then be ranked in either ascending or descending and it is retrieved accordingly. It has the ability to incorporate term weights, measure similarities between almost anything such as ranking documents according to their possible relevance. So with this model the customer satisfaction and the efficient retrieval are possible without affecting the privacy of the data.

REFERENCES

[1]. Ren K, Wang C, Wang Q (2012) Security challenges for the public cloud. IEEE Internet Comput 16(1):69–73
[2]. Kamara S, Lauter K (2010) Cryptographic cloud storage. In: Financial Cryptography and Data Security. Springer, Berlin/Heidelberg, pp 136–149
[3]. Song DX, Wagner D, Perrig A (2000) Practical techniques for searches on encrypted data. In: Proceedings of IEEE Symposium on Security and Privacy. IEEE, Berkeley, California, pp 44–55
[4]. Goh E-J (2003) Secure indexes. Cryptology ePrint Archive, Report 2003/216
[5]. Boneh D, Di Crescenzo G, Ostrovsky R, Persiano G (2004) Public key encryption with keyword search. In: Advances in Cryptology-Eurocrypt 2004. Springer, Berlin/Heidelberg, pp 506–522
[6]. Chang Y-C, Mitzenmacher M (2005) Privacy preserving keyword searches on remote encrypted data. In: Applied Cryptography and Network Security. Springer, Berlin/Heidelberg, pp 442–455
[7]. Curtmola R, Garay J, Kamara S, Ostrovsky R (2006) Searchable symmetric encryption: improved definitions and efficient constructions. In: Proceedings of the 13th ACM conference on Computer and communications security. ACM, Alexandria, VA, USA, pp 79–88
[8]. Bellare M, Boldyreva A, O'Neill A (2007) Deterministic and efficiently searchable encryption. In: Advances in Cryptology-CRYPTO 2007. Springer, Berlin/Heidelberg, pp 535–552
[9]. Wang C, Cao N, Li J, Ren K, Lou W (2010) Secure ranked keyword search over encrypted cloud data. In: 30th IEEE International Conference on Distributed Computing Systems (ICDCS). IEEE, Genoa, Italy, pp 253–262
[10]. Wang C, Cao N, Ren K, Lou W (2012) Enabling secure and efficient ranked keyword search over outsourced cloud data. IEEE Trans Parallel Distrib Syst 23(8):1467–1479
[11]. Cao N, Wang C, Li M, Ren K, Lou W (2011) Privacy-preserving multi-keyword ranked search over encrypted cloud data. In: Proceedings of IEEE INFOCOM. IEEE, Shanghai, China, pp 829–837
[12]. Yang C, Zhang W, Xu J, Xu J, Yu N (2012) A Fast Privacy-Preserving Multi- keyword Search Scheme on Cloud Data. In: International Conference on Cloud and Service Computing (CSC). IEEE, Shanghai, China, pp 104–110
[13]. Stefanov E, Papamanthou C, Shi E (2014) Practical Dynamic Searchable Encryption with Small Leakage. NDSS '14, San Diego, CA, USA
[14]. Wang C, Ren K, Yu S (2012) Urs KMR Achieving usable and privacy-assured similarity search over outsourced cloud data. In: Proceedings of IEEE INFOCOM. IEEE, Orlando, Florida, USA, pp 451–459
[15]. Li J, Wang Q, Wang C, Cao N, Ren K, Lou W (2010) Fuzzy keyword search over encrypted data in cloud computing. In: Proceedings of IEEE INFOCOM. IEEE, San Diego, CA, USA, pp 1–5
[16]. Chuah M, Hu W (2011) Privacy-aware bedtree based solution for fuzzy multi-keyword search over encrypted data. In: 31st International Conference on Distributed Computing Systems Workshops (ICDCSW). IEEE, Minneapolis, Minnesota, USA, pp 273–281
[17]. Liu C, Zhu L, Li L, Tan Y (2011) Fuzzy keyword search on encrypted cloud storage data with small index. In: IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS). IEEE, Beijing, China, pp 269–273
[18]. Ibrahim A, Jin H, Yassin AA, Zou D (2012) Approximate Keyword-based Search over Encrypted Cloud Data. In: IEEE Ninth International Conference on e-Business Engineering (ICEBE). IEEE, Hangzhou, China, pp 238–245
[19]. Bringer J, Chabanne H (2012) Embedding edit distance to enable private keyword search. Human-centric Comput Inf Sci 2(1):1–12