JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

A Verifiable Semantic Searching Scheme Over Encrypted Data in Public Cloud

[1] Mrs. Shaiqua Khan, [2]Sakshi Chatap, [3]Mrunal Marotkar, [4] Avanti Tagde , [5] Gaurav Urmale

[1]Assistant Professor, [2]Student, [3]Student, [4]Student, [5]Student

[1]Computer Science Engineering,
[1]G H Raisoni University Saikheda, Chhindwara, Madhya Pradesh, India

Abstract : Semantic search over encrypted data is a key task for secure information retrieval in the public cloud. Queries and search results are flexible as it aims to provide search services for any word. Existing semantic search schemes rely on predictive results from predefined keywords to validate search results from the cloud and do not support verifiable search. Also, the query is expanded to plain text and exact matching is done using expanded semantic words. Predefined keyword with limited precision. This post proposes a secure and verifiable semantic search scheme. For semantically optimal matching of ciphertexts, we formulate a word transfer (WT) problem that computes the minimum word transfer cost (MWTC) as the similarity between a query and a document, and transform the WT problem into a random linear programming (LP) suggests a safe transformation that converts to the problem. . Get encrypted MWTC. For verifiability, we examine the LP duality theorem to design a verification mechanism that verifies the correctness of search results using the intermediate data generated in the matching process. Security analysis shows that our scheme can guarantee verifiability and confidentiality. Experimental results on two data sets show that our scheme has higher accuracy than the other schemes.

Keywords: public cloud, results verifiable searching, secure semantic searching, word transportation.

I. INTRODUCTION

Cloud storage is becoming more and more popular in recent trends as it offers many advantages over traditional storage solutions. Cloud storage allows businesses to maintain their own data storage infrastructure rather than maintain it. , you can purchase the amount of storage you need from a cloud storage provider (CSP) to meet your storage needs. CSP can be used to handle all data maintenance tasks such as backup and restore. It also enables remote access to all data and optimizes operations across different locations. All these benefits enable businesses to significantly reduce operational costs by simply offloading business data to cloud storage. Beside these benefits that provided by the cloud storage, however, many security problems arise in cloud storage that prevent companies from migrating their data to cloud storage [7]. Due to the facts that cloud storage is usually hosted by third party provider other than the data owners and cloud storage infrastructure is usually shared among different users, data stored in cloud storage can be easily targeted by the masquerade attack [1, 8] and the insider data theft attack [9, 10]. These attacks threaten the data security and the data privacy of the stored data, as result, the data owners cannot rely on CSP to secure their confidential data. These attacks also induce the data owners to encrypt all their sensitive data such as the social security numbers (SSN), credit card information, and personal tax information before they can be saved in cloud storage. The encryption approach may have strengthened the data security of cloud data, but it has also degraded the data efficiency because the encryption will reduce the searchability of the data. Especially in cloud computing environments, it is inconvenient to download and decrypt all encrypted data from a remote cloud server before a user searches. Therefore, an efficient scheme to support retrieval of encrypted data in cloud computing will be very important before cloud storage becomes available to many enterprises. Recent research has proposed many schemes to enable keyword searches on encrypted data in cloud computing. The most common approach for these schemes is to index the keywords contained in each uploaded data file. The inherent scalability and flexibility of cloud computing is making cloud services popular and attracting cloud customers to offload storage and computing to the public cloud. Cloud computing

technology has made remarkable progress in both academia and industry, but cloud security is becoming one of the key factors limiting its progress. Incidents of data breaches in cloud computing, such as Apple's Fappening and Uber's data breach, are increasingly in the public eye. Fundamentally, cloud services should be trustworthy and honest, following defined protocols to ensure data confidentiality and integrity. Unfortunately, cloud server providers have complete control over their data and execution logs, which could lead them to engage in real-world fraud such as: B. Spying on sensitive data or performing erroneous calculations. Therefore, cloud customers should encrypt data and set up result verification mechanisms before offloading storage and compute to the cloud.

II. LITERATURE SURVEY

In this section, we review some previously proposed schemes to support keyword searches over encrypted cloud data, review previously proposed Wikipedia similarity matching techniques, and review content-based Solve your ad problem. Keyword Searching Against Encrypted Cloud Data Koletka and Hutchison, in his [2], created a proprietary data structure called a Secure File Object (SFO) to enable keyword searching against encrypted cloud data. I suggested. When the data owner uploads the data file to cloud storage, the client-side application creates her SFO, appends it to the encrypted data file, and then uploads it to cloud storage. Each SFO contains information describing the data file to upload. During SFO creation, the client-side application extracts the unique keywords from the data file you upload, encrypts them, and creates a list of encrypted keywords that is stored in the SFO. If the User wishes to search for a particular keyword, the User will send the Keyword to the Data Owner. Data owners calculate search power by encrypting keywords with the same key that was used to generate the list of encrypted keywords in SFO. The user can send the returned lookup function from the data owner to the cloud her server. If the list of encrypted keywords in SFO contains a search function, the cloud server will return an encrypted data file. Semantic search over encrypted data in this proposed Cloud Computing 5 SFO scheme will be implemented by the authors to provide simple keyword search over encrypted cloud data. The main drawback of the SFO scheme is that the scheme only supports keyword searches using the exact keywords as they appear in the data file. If there are typos in the keywords used to generate the search function, the cloud server will not be able to find the correct encrypted data file. To overcome the shortcomings of [2], Li, Wang et al. In [3], we proposed a “wildcard-based fuzzy set construction (WFSC)” scheme that enables fuzzy keyword searches over encrypted cloud data. The key concept behind WFSC is maintaining an index that covers all possible variations of a keyword within a defined edit distance. Instead of simply scrambling the keywords extracted from the data file, WFSC inserts wildcard characters into the keywords to expand each extracted keyword into a set of modified keywords.

III. RESEARCH METHODOLOGY

1. System Architecture

As shown in Figure 1, our system involves three entities: data owners, data users, and cloud servers. Data owners have a lot of useful documentation, but local machine resources are limited. So the owner is very willing to do her Initialize () to initialize the proposed schema. The owner obtains document F by encrypting it. Regarding cloud servers, our scheme is more efficient than the "semi-honest servers" used in other secure semantic search schemes [3], [4], [5], [6], [7]. Resist sophisticated security models. [8], [9]. In our model, rogue cloud servers return false or bogus search results and attempt to learn sensitive information, but do not maliciously delete or tamper with offloaded documents. Therefore, our secure semantic scheme should guarantee verifiability and confidentiality under such a security model. Regarding verifiability, we first reformulate the definitions of result forgery attacks and evidence forgery attacks in [24], and then apply game-based security definitions to analyze the verifiability of the proposed scheme in Section VII. adopted. Definition 1 (result forgery attack). A result forgery attack consists of a rogue cloud server trying to return incorrect search results to a user for some reason. Formally, q is any search term and C is the encrypted document. Then let T(C,q) be the correct search result and R(C,q) be the search result returned from the cloud server. In this attack, R(C,q) 6 = T(C,q).

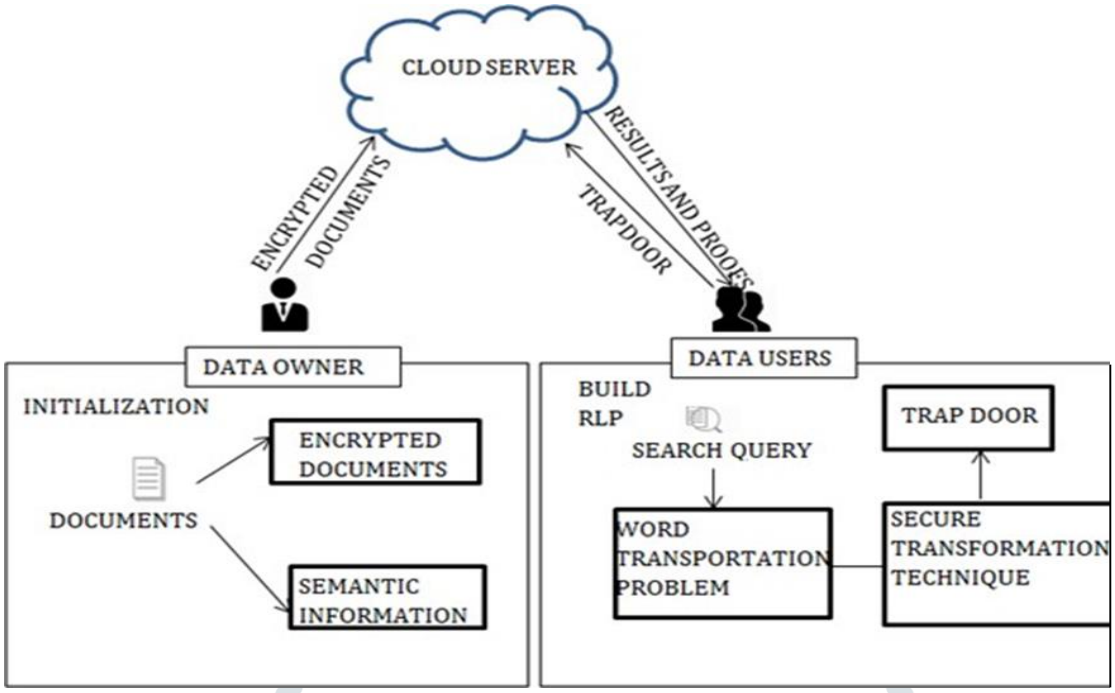

Fig 1 : The system architecture of our secure verifiable semantic searching scheme.

1.1 PROPOSED METHOD

The proposed scheme formulate the Word Transportation (WT) problem and propose a secure transformation technique to transform WT problems into random Linear Programming (LP) problems for obtaining the encrypted minimum word transportation cost as measurements between queries and documents. For supporting verifiable searching, we explore the duality theorem of LP and present a novel insight that using the intermediate data produced in the matching process as proof to verify the correctness of search results.

1.2 Advantages

- The trials results show that the delicacy of our proposed scheme is better than that of other schemes.
- The results demonstrate the effectiveness of our secure verifiable semantic searching scheme grounded on word transportation optimal matching.
- The delicacy of the proposed scheme using the description motifs is still advanced than that of other schemes. A reason is that the word transportation optimal matching is salutary to dissect the semantic relationship between the words and the significance of the distinction among words in long- text queries.
- It's worth spending farther time to get advanced quest delicacy for the operations in practical scripts. Take the medical profile as an illustration, a correct particular medical profile of a case is essential and useful to help the croaker make a precise complaint opinion and health evaluation.

1.3 Disadvantages

- The traditional searchable encryption schemes demand that query words must be the predefined keywords in the outsourced documents, which leads to an egregious limitation of these schemes that similarity dimension solely base on the exact matching between keywords in the queries and documents.

- Scheme only verifies whether all the documents containing the extended keywords are returned to users or not, and needs users to rank all the documents for getting top- k related documents. thus, it's challenging to design a secure semantic searching scheme to support verifiable searching.

- The being scheme is unfit to support semantic searching and introduces multiple rounds of communication between data possessors.

IV. Conclusion

We propose a secure verifiable semantic searching scheme that treats matching between queries and documents as a word transportation optimal matching task. Therefore, we investigate the fundamental theorems of linear programming (LP) to design the word transportation (WT) problem and a result verification mechanism. We formulate the WT problem to calculate the minimum word transportation cost (MWTC) as the similarity metric between queries and documents, and further propose a secure transformation technique to transform WT problems into random LP problems. Therefore, our scheme is simple to deploy in practice as any ready-made optimizer can solve the RLP problems to obtain the encrypted MWTC without learning sensitive information in the WT problems. Meanwhile, we believe that the proposed secure transformation technique can be used to design other privacy-preserving linear programming applications. We bridge the semantic-verifiable searching gap by observing an insight that using the intermediate data produced in the optimal matching process to verify the correctness of search results. Specifically, we investigate the duality theorem of LP and derive a set of necessary and sufficient conditions that the intermediate data must meet. The experimental results on two TREC collections show that our scheme has higher accuracy than other schemes. In the future, we plan to research on applying the principles of secure semantic searching to design secure cross-language searching schemes.

V. References

1. D.X. Song, D. Wagner, and A. Perring, “Practical techniques for searches on encrypted data, "in orc. IEEE Symp. Secure. Privacy, 2000, pp. 44-55.
2. Z. Fu, J. Shu, X. Sun, and N. Linge, “Smart cloud search services: verifiable keyword-based semantic search over encrypted cloud data,” IEEE Trans. Consum. Electron., vol. 60, no. 4, pp. 762–770, 2014.
3. Z. J. Fu, X. M. Sun, N. Linge, and L. Zhou, “Achieving effective cloud search services: multi-keyword ranked search over encrypted cloud data supporting synonym query,” IEEE Trans. Consum. Electron., vol. 60, no. 1, pp. 164–172, 2014.
4. T. S. Moh and K. H. Ho, “Efficient semantic search over encrypted data in cloud computing,” in Proc. IEEE. Int. Conf. High Perform. Comput. Simul., 2014, pp. 382–390.
5. N. Jadhav, J. Nikam, and S. Bahekar, “Semantic search supporting similarity ranking over encrypted private cloud data,” Int. J. Emerging Eng. Res. Technol., vol. 2, no. 7, pp. 215–219, 2014.
6. Y. G. Liu and Z. J. Fu, “Secure search service based on word2vec in the public cloud,” Int. J. Comput. Sci. Eng., vol. 18, no. 3, pp. 305–313, 2019.
7. E. J. Goh, “Secure indexes.” IACR Cryptology ePrint Archive, vol. 2003, pp. 216–234, 2003
8. K. Kurosawa and Y. Ohtaki, “UC-secure searchable symmetric encryp- tion,” in Proc. Int. Conf. Financial Cryptography Data Secur. Springer,2012 pp. 285-298.
9. T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient estimation of word representations in vector space,” in Proc. Int. Conf. Learn. Represent., 2013.,