IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A PROVABLE CONTEXTUAL RETRIEVAL STRATEGY IN THE PUBLIC CLOUD USING OPTIMAL MATCHING OVER SECURED DATA

Ms.S.Jonisha[1], Mohammed Alameen A[2], Guhan P[3], Mokinder S[4]

Assistant Professor[1],Students[2,3,4]

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING PERI INSTITUTE OF TECHNOLOGY

Abstract

Semantic searching over encrypted data is a crucial task for secure information retrieval in public cloud. It aims to provide retrieval service to arbitrary words so that queries and search results are flexible. In existing semantic searching schemes, the verifiable searching does not be supported since it is dependent on the forecasted results from predefined keywords to verify the search results from cloud, and the queries are expanded on plain text and the exact matching is performed by the extended semantically words with predefined keywords, which limits their accuracy. In this paper, we propose a secure verifiable semantic searching scheme. For semantic optimal matching on ciphertext, we formulate word transportation (WT) problem to calculate the minimum word transportation cost (MWTC) as the similarity between queries and documents, and propose a secure transformation to transform WT problems into random linear programming (LP) problems to obtain the encrypted MWTC. For verifiability, we explore the duality theorem of LP to design a verification mechanism using the intermediate data produced in matching process to verify the correctness of search results. Security analysis demonstrates that our scheme can guarantee verifiability and confidentiality. Experimental results on two datasets show our scheme has higher accuracy than other schemes.

Introduction

Inherent scalability and flexibility of cloud computing make cloud services so popular and attract cloud customers to outsource their storage and computation into the public cloud. Although the cloud computing technique develops magnificently in both academia and industry, cloud security is becoming one of the critical factors restricting its development. We have proposed a verifiable semantic searching scheme that extends the query words to get the predefined keywords related to query words, then they used the

extended keywords to search on a symbol-based trie index. However, their 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.

The aim of this project is to provide semantic searching over encrypted data. It helps to retrieve the information in public cloud with high security. This project is to secure verifiable semantic searching scheme on cipher text. We have used duality theorem of Linear Programming (LP) for the verification mechanism.

Therefore, we introduce word embeddings to represent words and compute Euclidean distance as the similarity distance between words, then formulate the word transportation (WT) problems based on the word embeddings representation. However, the cloud server could learn sensitive information in the WT problems, such as the similarity between words.

Literature Survey

It is desirable to store data on data storage servers such as mail servers and file servers in encrypted form to reduce security and privacy risks. But this usually implies that one has to sacrifice functionality for security. For example, if a client wishes to retrieve only documents containing certain words, it was not previously known how to let the data storage server perform the search and answer the query, without loss of data confidentiality. We describe our cryptographic schemes for the problem of searching on encrypted data and provide proofs of security for the resulting crypto systems. Their techniques have a number of crucial advantages.

They are provably secure: they provide provable secrecy for encryption, in the sense that the untrusted server cannot learn anything about the plaintext when only given the ciphertext; they provide query isolation for searches, meaning that the untrusted server cannot learn anything more about the plaintext than the search result; they provide controlled searching, so that the untrusted server cannot search for an arbitrary word without the user's authorization; they also support hidden queries, so that the user may ask the untrusted server to search for a secret word without revealing the word to the server. The algorithms presented are simple, fast (for a document of length n, the encryption and search algorithms only need O(n) stream cipher and block cipher operations), and introduce almost no space and communication overhead, and hence are practical to use today.

With the increasing popularity of the pay-as-you- consume cloud computing paradigm, a large number of cloud services are pushed to consumers. One hand, it brings great convenience to consumers who use intelligent terminals; on the other hand, consumers are also facing serious difficulties that how to search the most suitable services or products from cloud. So how to enable a smart cloud search scheme is a critical

problem in the consumer-centric cloud computing paradigm. For protecting data privacy, sensitive data are always encrypted before being outsourced. Although the existing searchable encryption schemes enable users to search over encrypted data, these schemes support only exact keyword search, which greatly affects data usability.

Moreover, these schemes do not support verifiability of search result. In order to save computation cost or download bandwidth, cloud server only conducts a fraction of search operation or return a part of result, which is viewed as selfish and semi-honest-but-curious. So, how to enhance flexibility of encrypted cloud data while supporting verifiability of search result is a big challenge. To tackle the challenge, a smart semantic search scheme is proposed in this paper, which returns not only the result of keyword-based exact match, but also the result of keyword-based semantic match. At the same time, the proposed scheme supports the verifiability of search result. The rigorous security analysis and performance analysis show that the proposed scheme is secure under the proposed model and effectively achieves the goal of keyword-based semantic search.

In recent years, consumer-centric cloud computing paradigm has emerged as the development of smart electronic devices combined with the emerging cloud computing technologies. A variety of cloud services are delivered to the consumers with the premise that an effective and efficient cloud search service is achieved. For consumers, they want to find the most relevant products or data, which is highly desirable in the "pay-as-you use" cloud computing paradigm. As sensitive data (such as photo albums, emails, personal health records, financial records, etc.) are encrypted before outsourcing to cloud, traditional keyword search techniques are useless.

Meanwhile, existing search approaches over encrypted cloud data support only exact or fuzzy keyword search, but not semantics-based multi-keyword ranked search. Therefore, how to enable an effective searchable system with support of ranked search remains a very challenging problem. This paper proposes 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.

Cloud storage has become more and more popular as it provides many benefits over traditional storage solutions. Despite the many benefits provided by cloud storage, many security problems have also arisen in cloud storage, which prevents companies from migrating their data to cloud storage. As a result, the owners encrypt their sensitive data before storing it in cloud storage. While encryption increases the security of the data, it also reduces the searchability of the data and thus, the efficiency of the search. Recently, research

has been done on several schemes which enable keyword searching on encrypted data in cloud computing. However, these schemes contain weaknesses which make them impractical when applied to real-life scenarios.

In this paper, we developed a system to support semantic search on encrypted data in cloud computing with three different schemes which are "Synonym-Based Keyword Search (SBKS)", "Wikipedia-Based Keyword Search (WBKS)", and "Wikipedia-Based Synonym Keyword Search (WBSKS)". Our results demonstrated that our schemes are more efficient in terms of performance and storage requirements than the former proposed schemes. Therefore, our developed schemes are more practical than the former proposed schemes

With the appealing features of cloud computing, cloud becomes an important infrastructure of enterprise IT. A large amount of data is being outsourced to the cloud. Before outsourcing the data is being encrypted. Encryption makes simple but important functionalities like search operations over cloud data difficult.

The traditional and efficient plaintext keyword search technique has no effect on encrypted data. The existing searchable encryption schemes support only exact keyword search, not support semantic based search. Hence we propose a search scheme wherein the semantic relationship and synonym of the query keyword are considered with the help of data structures like Semantic Relationship Library (SRL), Inverted Index. The result files are displayed in order according to total relevance score.

SYSTEM DESIGN

SYSTEM ARCHITECTURE

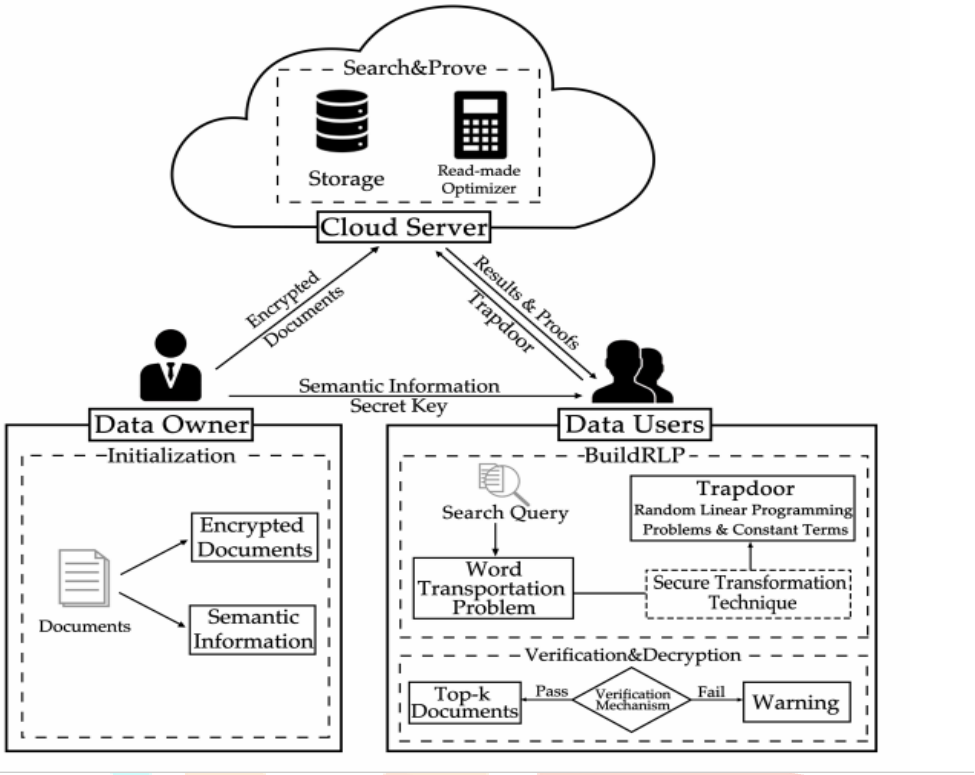

Fig 1:System architecture

MODULES ALGORITHM IMPLEMENTATION

4.2.1 MODULES

- Data owner
- Data User
- Cloud Server

4.2.2 MODULES DESCRIPTION

Data Owner

In the first module, we develop the Data owner module. Firs the new data owner registers and then gets the login credentials. Once after the credentials are authenticated, then the data owner can able to login to the system. The data owner has the option of Uploading the file to the cloud, My files, requested files, download files option, where each has their perspective functionalities. The data owner has a lot of useful documents, but only has limited resources on the local machines. Therefore, the owner is highly motivated to perform Initialize () for initializing the proposed scheme. The owner encrypts documents to get cipher text documents with secret key, then outsources to the cloud server. The data owner builds forward indexes, then sends indexes to data users. The data owner is the initiator who initializes the secure searching scheme. Not like in other schemes, the data owner in our scheme does not need to perform massive cryptographic operations, such as order-preserving encryption and homomorphic encryption. For the owner, the main steps

including (1) generating symmetrical encryption secret key; (2) encrypting documents; (3) building forward indexes.

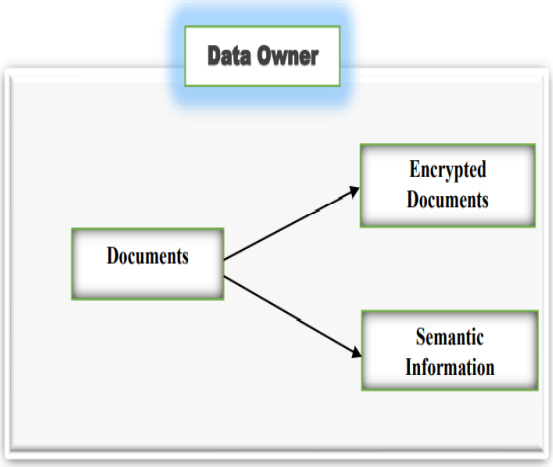

Fig 2:data owner process

Data User

In the next module, we design the Data User Module, where the data user is registered with the details and gets the credentials to login to the system. Data User has the option of Search Key, Search file, Requested files, downloaded files. Each has their perspective functionalities to support the data user operations. Data users are the searching requesters that send the trapdoor of a semantic query to the cloud server for acquiring top-k related documents that are encrypted and stored in cloud server. The secret key generated is valid for 5 minutes only for the enhanced security purpose. Each time the key is received to the data user through the registered email through a secured channel to avoid intruders and attackers.

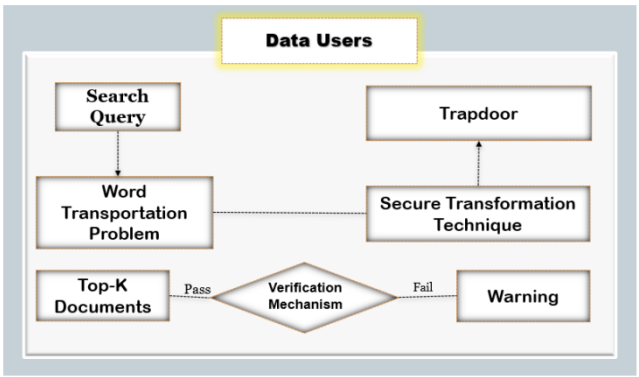

Fig 3:data user process

Cloud Server (CS)

In this module, we develop the Cloud server module, where we design the system to upload the files in a free cloud server named Drive HQ. Is an entity with powerful computation and storage resources. CS stores a massive amount of encrypted data, and receives the semantic queries to look for the required documents on behalf of the data user. The cloud finds the relevant documents, and sends them back to the data user. The cloud server is an intermediate service provider that performs the retrieval process. In our

scheme, the cloud needs to perform the main steps including (1) performing optimal matching on ciphertext and generating proofs in the matching; (2) calculating and ranking the measurements.

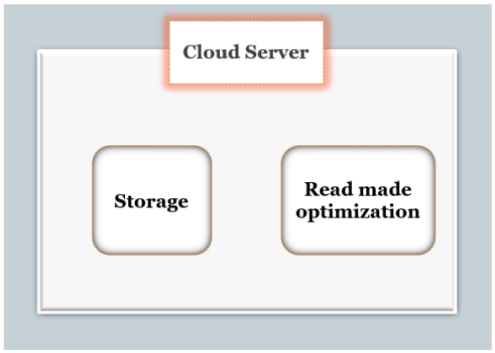

Fig 4:cloud server process

PERFORMANCE ANALYSIS

GRAPH:

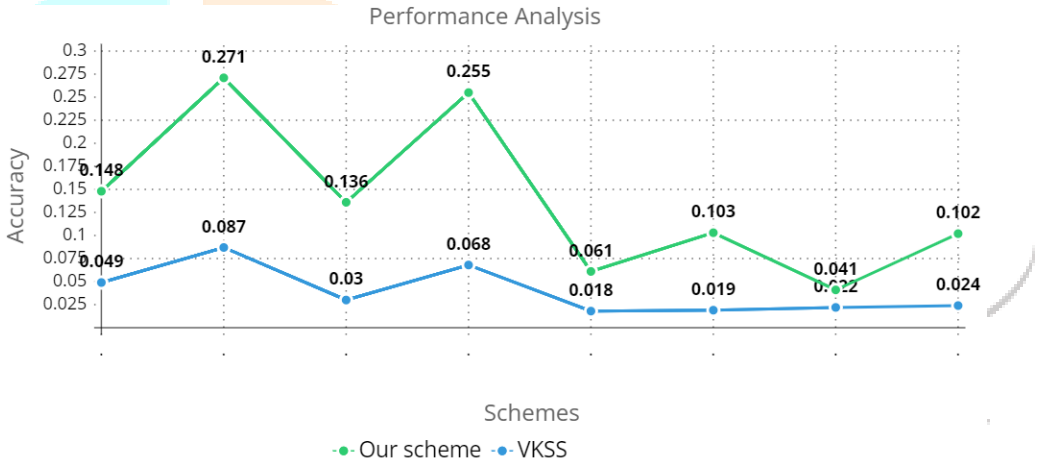

Fig 5: performance graph

RESULTS AND THEORY DESCRIPTION

We conducted extensive experiments to evaluate the performance of the time cost and accuracy between our scheme and the VKSS scheme which is the only one supports secure verifiable semantic searching in prior works found. We report the results of time cost experiments over the Robust04 dataset using title topics as queries due to limited space. We do not consider the time cost of encrypting documents and decrypting documents since the time cost is the same in both schemes. From the performance evaluation of the time cost, we can see that the owner and users in our scheme can ease the computational burden by paying the time cost in cloud compared with the VKSS.

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.

FUTURE SCOPE

The proposed scheme outsources the complex computational of performing proof generation task and semantic matching task to the cloud. Therefore, our scheme is more in line with the outsourcing computation characteristics of the cloud computing paradigm. The main reason why our scheme spends more time in cloud is that the computation of optimal matching on cipher text is related to the size of linear programming problem. Therefore, in the future, we plan to design other schemes to reduce the time cost of optimal matching on cipher text according to find.

REFERENCES

[1] D. X. Song, D. Wagner, and A. Perrig, “Practical techniques for searches on encrypted data,” in Proc. IEEE Symp. Secur. 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] Z. H. Xia, Y. L. Zhu, X. M. Sun, and L. H. Chen, “Secure semantic expansion based search over encrypted cloud data supporting similarity ranking,” J. Cloud Comput., vol. 3, no. 1, pp. 1–11, 2014.

[7] Z. Fu, L. Xia, X. Sun, A. X. Liu, and G. Xie, “Semantic-aware searching over encrypted data for cloud computing,” IEEE Trans. Inf. Forensics Security, vol. 13, no. 9, pp. 2359–2371, Sep. 2018.