A Verifiable Semantic Searching Scheme by Optimal Matching over Encrypted Data in Public Cloud

Ch. Rushyendra Mani[1], Sk. Shahida[2]

[1]Assistant Professor, Dept. of MCA, Audisankara College of Engineering and Technology (AUTONOMOUS), Gudur, AP, India.

[2]PG Scholar, Dept. of MCA, Audisankara College of Engineering and Technology (AUTONOMOUS), Gudur, AP, India.

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 plaintext 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.

1.INTRODUCTION

Internet scalability and flexibility of cloud computing make cloud services so popular and attract cloud customersto 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 devel- opment. The events of data breaching in cloud computing, such as the Apple Fappening and the Uber data breaches, are increasingly attracting public attention. In principle, the cloud services are trusted and honest, should ensure data confidentiality and integrity according to predefined protocols. Unfortunately, as the cloud server providers take full control of data and execute protocols, they may conduct dishonest behavior in the real world, such as sniffing sensitive data or performing incorrect calculations. Therefore, cloud customers should encrypt their data and establish a result verification mechanism before outsourcing storage and

computation to the cloud. Since Song et al. [1] proposed the pioneering workabout the searchable encryption scheme, searchable encryption has attracted significant attention. However, the traditional searchable encryption schemes require that query words must be the predefined keywords in the outsourced documents, which leads to an obvious limitation of these schemes that similarity measurement solely base on the exact matching between keywords in the queries and documents. Therefore, some works proposed semantic searching schemes to provide retrieval service to arbitrary words, making the query words and search results flexible and uncertain. However, the verifi- able searching schemes are dependent on forecasting the fixed results of predefined keywords to verify the correctness of the search result returned by the cloud. Therefore, the flexibility of semantic schemes and the fixity of verifiable schemes enlarge the gap between semantic searching and verifiable searching over encrypted data. Although Fu et al. [2] proposed a verifi- able 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. Therefore, it is challenging to design a secure semantic searching scheme to support verifiable searching.

Most of the existing secure semantic searching schemesconsider the semantic relationship among words to perform query expansion on the plaintext, then still use the query words and extended semantically related words to perform exact matching with the specific keywords in outsourced documents. We can roughly divide these schemes into three categories: secure semantic searching based synonym [3], [4], secure semantic searching based mutual information model [5], [6], secure semantic searching based concept hierarchy [2], [7], [8]. We can see that these schemes only use the elementary semantic information among words. For example, synonym schemes only use synonym attributes; mutual information models only use the co-occurrences information. Although Liu et al. [9] introduce the Word2vec technique to utilize the semantic information of word embeddings, their approach damages the semantic information due to straightly aggre- gating all the word vectors. We think that secure semantic searching schemes should further utilize a wealth of semantic information among words and perform optimal matching on the ciphertext for high search accuracy.

2. LITERATURE SURVEY

Since Song et al. [1] proposed the notion of searching over encrypted cloud data, searchable encryption has received significant attention for its practicability in the past 20 years. Therefore, many works have made efforts on the security as well as functionality in the searchable encryption field.

Along the research line about security, many works formulate the definitions of security as well as novel attack pattern against the existing schemes. Goh et al. [10] formulated a security model for document indexes known as semantic security against adaptive chosen keyword attack (IND-CKA), which requires the document indexes not to reveal contents of documents. However, we note that the definition of IND-CKA does not indicate that the queries must be secure. Curtmola et al. [11] further improved security definitions for symmetricsearchable encryption, then put forth chosen-keyword attacks and adaptive chosen-keyword attacks. Besides, Islam et al. [12] first introduced the access pattern disclosure used to learn sensitive information about the encrypted documents, then Liu et al. [13] presented a novel attack based on the search pattern leakage.

Stefanov et al. [14] introduced the notions of forward security and backward security for the dynamic searchable encryption schemes that support data addition and deletion.

Along another research line about functionality, many works introduced practical functions to meet the demand in practice, such as ranked search and semantic searching for improving search accuracy. Additionally, some works proposed verifiable searching schemes to verify the correctness of search results. Ranked Search over Encrypted Data. Ranked search means that the cloud server can calculate the relevance scores be- tween the query and each document, then ranks the documents without leaking sensitive information. The notion of single- keyword ranked search was proposed in [15] that used a modified one-to-many order-preserving encryption (OPE) to encrypt relevance scores and rank the encrypted documents. Cao et al. [16] first proposed a privacy-preserving multi- keyword ranked search scheme (MRSE), which represents documents and queries with binary vectors and uses the secure kNN algorithm (SeckNN) [17] to encrypt the vectors, then use the inner product of the encrypted vectors as the similarity measure. Besides, Yu et al. [18] introduced homomorphic encryption to encrypt relevance scores and realize a multi- keyword ranked search scheme under the vector space model. Recently, Kermanshahi et al. [19] used various homomorphic encryption techniques to propose a generic solution for supporting multi-keyword ranked searching schemes that can resist against several attacks brought by OPE-based schemes. Secure Semantic Searching. A general limitation of traditional searchable encryption schemes is that they fail to utilize semantic information among words to evaluate the relevance between queries and documents. Fu et al. [3] proposed the first synonym searchable encryption scheme under the vector space model to bridge the gap between semantically related words and given keywords. They first extended the keyword set from the synonym keyword thesaurus built on the New American Roget's College Thesaurus (NARCT), then used the extended keyword set to build secure indexes with SeckNN. Using the order-preserving encryption algorithm, [5] and [6] presented secure semantic searching schemes based on the mutual information model. Xia et al. [6] proposed a scheme that requires the cloud to constructs a semantic relationship library based on the mutual information used in [20]. However, any schemes based on the inverted index can calculate the mutual information model. Using the SeckNN algorithm, [7], [8], [2] proposed secure semantic searching schemes based on the concept hierarchy. For example, Fu et al. [8] proposed a central keyword semantic extension searching scheme which calculates weights of query words based on grammatical relations, then extends the central word based on the concept hierarchy tree from WordNet. Inspired by word embedding used in plaintext information retrieval [21], [22], Liu et al. [9] introduced the Word2vec to represent both queries and documents as compact vectors. However, their approach damagesthe semantic information of word embedding due to straightly aggregating all the word vectors of the words.

3. PROPOSED SYSTEM

we propose a secure verifiable semanticsearching scheme that treats matching between queries and documents as an optimal matching task. We treat the document words as "suppliers," the query words as "consumers," and the semantic information as "product," and design the minimum word transportation cost (MWTC) as the similarity metric between queries and documents. 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. For semantic optimal matching on the ciphertext, we further propose a secure transformation to transform WT problems into random linear programming (LP) problems. In this way, the cloud can leverage any ready- made optimizer to solve the RLP problems and obtain the encrypted MWTC as measurements without learning sensitive information. Considering the cloud server may be dishonest to return wrong/forged search results, we explore the duality theorem of linear programming (LP) and derive a set of necessary and sufficient conditions that the intermediate data produced in the matching process must satisfy. Thus, we can verify whether the cloud solves correctly RLP problems and further confirm the correctness of search results. Our new ideas are summarized as follows:

1. Treating the matching between queries and documents as an optimal matching task, we explore the fundamental theorems of linear programming (LP) to propose a se- cure verifiable semantic searching scheme that performs semantic optimal matching on the ciphertext.

2. For secure semantic optimal matching on the ciphertext, we 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.

3. 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.

A. System Architecture

As illustrated in Fig. 1, there are three entities involved in our system: the data owner, data users, and the cloud server.

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 F to get ciphertext documents C with secret key K, then outsources C to the cloud server. The data owner builds forward indexes I, then sends indexes I and K to data users.

Data users are the searching requesters that send the trap-

door of a query to the cloud server for acquiring top-k related documents. Specifically, users input arbitrary query words q, then perform BuildRLP () to generate word transportation problems Ψ, after transform Ψ to random linear programming problems Ω and the corresponding constant terms Δ as a trap- door. Afterward, users receive top-k encrypted documents and proofs Λ returned from the cloud. Users perform VerDec () to decrypt documents when Λ passes our verification mechanism. The cloud server is an intermediate service provider that stores the encrypted document dataset C and performs the retrieval process. Once receiving the trapdoor, the cloud server performs SeaPro () for leveraging any ready-made optimizer to solve the Ω, then obtains the encrypted minimum word transportation cost values with Δ. The cloud ranks the values in ascending order and returns the top-k encrypted documents to users. In the process, the cloud server also provides proofs

Λ for proving the correctness of the search results.

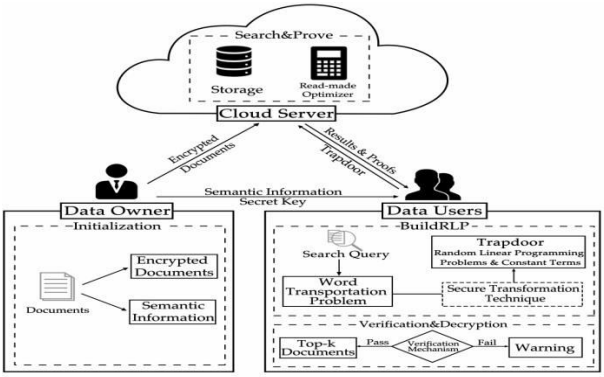

Figure 1. The system architecture of our secure verifiable semantic searching scheme.

B. Security Model

We assume that the data owner is trusted, and the data users are authorized by the data owner. The communication channels between the owner and users are secure on existing security protocols such as SSL, TLS.

With regard to the cloud server, our scheme resists a more challenging security model which is beyond the "semi-honest server" used in other secure semantic searching schemes [3], [4], [5], [6], [7], [8], [9]. In our model, the dishonest cloud server attempts to return wrong/forged search results and learn sensitive information, but would not maliciously delete or tamper with the outsourced documents. Therefore, our secure semantic scheme should guarantee the verifiability, and confidentiality under such a security model.

4. 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.

5. 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] Srikanth veldandi., et al. "Grid Synchronization Failure Detection on Sensing the Frequency and Voltage beyond the Ranges." Journal of Energy Engineering and Thermodynamics, no. 35, Aug. 2023, pp. 1–7. https://doi.org/10.55529/jeet.35.1.7.

[3] Srikanth veldandi, et al. "Intelligents Traffic Light Controller for Ambulance." Journal of Image Processing and Intelligent Remote Sensing, no. 34, July 2023, pp. 19–26. https://doi.org/10.55529/jipirs.34.19.26.

[4] Srikanth veldandi, et al. "Smart Helmet with Alcohol Sensing and Bike Authentication for Riders." Journal of Energy Engineering and Thermodynamics, no. 23, Apr. 2022, pp. 1–7. https://doi.org/10.55529/jeet.23.1.7.

[5] Srikanth veldandi, et al. "An Implementation of Iot Based Electrical Device Surveillance and Control using Sensor System." Journal of Energy

Engineering and Thermodynamics, no. 25, Sept. 2022, pp. 33–41. https://doi.org/10.55529/jeet.25.33.41.
[6] Srikanth veldandi, et al "Design and Implementation of Robotic Arm for Pick and Place by using Bluetooth Technology." Journal of Energy Engineering and Thermodynamics, no. 34, June 2023, pp. 16–21. https://doi.org/10.55529/jeet.34.16.21.
[7] Srikanth, V. "Secret Sharing Algorithm Implementation on Single to Multi Cloud." Srikanth | International Journal of Research, 23 Feb. 2018, journals.pen2print.org/index.php/ijr/article/view/11641/11021.
[8] V. Srikanth. "Managing Mass-Mailing System in Distributed Environment" v srikanth | International Journal & Magazine of Engineering, Technology, Management and Research, 23 August. 2015. http://www.ijmetmr.com/olaugust2015/VSrikanth-119.pdf
[9] V. Srikanth. "SECURITY, CONTROL AND ACCESS ON IOT AND ITS THINGS" v srikanth | INTERNATIONAL JOURNAL OF MERGING TECHNOLOGY AND ADVANCED RESEARCH IN COMPUTING, 15 JUNE. 2017. http://ijmtarc.in/Papers/Current%20Papers/IJMTARC-170605.pdf

[10] E. J. Goh, "Secure indexes." IACR Cryptology ePrint Archive, vol. 2003, pp. 216–234, 2003.

[11] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: improved definitions and efficient constructions," J. Comput. Secur., vol. 19, no. 5, pp. 895–934, 2011.

[12] M. S. Islam, M. Kuzu, and M. Kantarcioglu, "Access pattern disclosure on searchable encryption: Ramification, attack and mitigation." in Proc. ISOC Network Distrib. Syst. Secur. Symp., vol. 20, 2012, pp. 12–26.

[13] C. Liu, L. H. Zhu, M. Z. Wang, and Y. A. Tan, "Search pattern leakage in searchable encryption: Attacks and new construction," Inf. Sci., vol. 265, pp. 176–188, 2014.

[14] E. Stefanov, C. Papamanthou, and E. Shi, "Practical dynamic searchable encryption with small leakage." in Proc. ISOC Network Distrib. Syst. Secur. Symp., vol. 71, 2014, pp. 72–75.

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

[16] N. Cao, C. Wang, M. Li, K. Ren, and W. J. Lou, "Privacy-preserving multi-keyword ranked search over encrypted cloud data," IEEE Trans. Parallel Distrib. Syst., vol. 25, no. 1, pp. 222–233, 2013.

[17] W. K. Wong, D. W. Cheung, B. Kao, and N. Mamoulis, "Secure knn computation on encrypted databases," in Proc. ACM Symp. Int. Conf. Manage. Data, 2009, pp. 139–152.

[18] J. D. Yu, P. Lu, Y. M. Zhu, G. T. Xue, and M. L. Li, "Toward secure multikeyword top-k retrieval over encrypted cloud data," IEEE Trans. Dependable Secure Comput., vol. 10, no. 4, pp. 239–250, 2013.

[19] S. K. Kermanshahi, J. K. Liu, R. Steinfeld, and S. Nepal, "Generic multi- keyword ranked search on encrypted cloud data," in Proc. Springer Eur. Symp. Res. Comput. Secur., 2019, pp. 322–343.

[20] L. F. Lai, C. C. Wu, P. Y. Lin, and L. T. Huang, "Developing a fuzzy search engine based on fuzzy ontology and semantic search," in Proc. IEEE Int. Conf. Fuzzy Syst., 2011, pp. 2684–2689.

[21] A. Imani, A. Vakili, A. Montazer, and A. Shakery, "Deep neural net- works for query expansion using word embeddings," in Proc.

Springer Eur. Conf. Inf. Retrieval, 2019, pp. 203–210.

[22] Y. Long, L. Liu, Y. Shen, and L. Shao, "Towards affordable semantic searching: Zero-shot retrieval via dominant attributes," in Proc. AAAI Conf. Artif. Intell., 2018, pp. 7210–7217.

. Commun., 2012, pp. 917–922.

[23] Q. Chai and G. Gong, "Verifiable symmetric searchable encryption for semi-honest-but-curious cloud servers," in Proc. IEEE Int. Conf

AUTHOR'SPROFILE

CH. RUSHYENDRA MANI iscurrently working as Assistant Professor in Audisankara College of Engineering and Technology (AUTONOMOUS), NH-5, Bypass Road, Gudur, Tirupati (Dt.), Andhra Pradesh, India.

SK.SHAHIDAis pursuing MCA from Audisankara College of Engineering and Technology (AUTONOMOUS), NH-5, Bypass Road, Gudur, Tirupati (Dt.), Andhra Pradesh, India.