www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

SECURE SEMANTIC SEARCH IN PUBLIC CLOUDS: OPTIMAL MATCHING OVER ENCRYPTED DATA

[1] Dr.CH.G.V.N.Prasad,[2] V.Aasrith,[3] Y.Shankar,[4] V.Deepthi,[5] R.Sandhya,[6] V.Girish Kumar
[1]Professor,[23456]B.Tech Students
Department Of Computer Science & Engineering
Sri Indu College Of Engineering & Technology,Sheriguda, Ibrahimpatnam

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.

I. INTRODUCTION

Internet 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. 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.Proposed the pioneering work about 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 verifiable 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. 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.

Index in Cosmos
JUNE 2025, Volume 15, ISSUE 2
UGC Approved Journal

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

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 forgetting 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 schemes consider 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, secure semantic searching based mutual information model, secure semantic searching based concept hierarchy .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. 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.

In this paper, we propose a secure verifiable semantic searching 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 dis- tance 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.

II. LITERATURE SURVEY

Title: Achieving effective cloud search services: multi-keyword ranked search over encrypted data supporting synonym query

Author: A. K. Das, M. Wazid, N. Kumar, A. V. Vasilakos, and J. J. P. C. Rodrigues.

Abstract: 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 realworld dataset were performed to validate the approach, showing that the proposed solution is very effective and efficient for multikey word ranked searching in a cloud environment.

Title: Efficient semantic search over encrypted datain cloud computing.

Author: W. Yang, S. Wang, J. Hu, G. Zheng, and C. Valli.

Abstract: 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.

Title: Semantic search supporting similarity ranking over encrypted private cloud data.

Author: C. Yuan, X. Sun, and Q. M. J. Wu

Abstract: 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.

Title: Secure semantic expansion based search over encrypted cloud data supporting similarity ranking.

Author: C.-C. Chang and N.-T. Nguyen

Abstract: With the advent of cloud computing, more and more information data are outsourced to the public cloud for economic savings and ease of access. However, the privacy information has to be encrypted to guarantee the security. To implement efficient data utilization, search over encrypted cloud data has been a great challenge. The existing solutions depended entirely on the submitted query keyword and didn't consider the semantics of keyword. Thus the search schemes are not intelligent and also omit some semantically related documents. In view of the deficiency, as an attempt, we propose a semantic expansion based similar search solution over encrypted cloud data. Our solution could return not only the exactly matched files, but also the files including the terms semantically related to the query keyword. In the proposed scheme, a corresponding file metadata is constructed for each file. Then both the encrypted metadata set and file collection are uploaded to the

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

cloud server. With the metadata set, the cloud server builds the inverted index and constructs semantic relationship library (SRL) for the keywords set. After receiving a query request, the cloud server first finds out the keywords that are semantically related to the query keyword according to SRL. Then both the query keyword and the extensional words are used to retrieve the files. The result files are returned in order according to the total relevance score. Eventually, detailed security analysis shows that our solution is privacy- preserving and secure under the previous searchable symmetric encryption (SSE) security definition. Experimental evaluation demonstrates the efficiency and effectives of the scheme.

III. SYSTEM ANALYSIS & DESIGN

EXISTING SYSTEM

Most of the existing secure semantic searching schemes consider 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 secure semantic searching based mutual information model secure semantic searching based concept hierarchy. We can see that these schemes only use the elementary semantic information among words.

Introduce the Word2vec technique to utilize the semantic information of word embeddings, their approach damages the semantic information due to straightly aggregating 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.

DISADVANTAGES

1. Limited Semantic Understanding: Existing systems often rely on basic word relationships and may not fully grasp the nuances of semantic meanings, leading to less accurate results.
2. Data Privacy Risks: Query expansion on plaintext in many existing systems can expose sensitive information, compromising data privacy.
3. Keyword Dependency: Some systems depend heavily on predefined keywords, which can restrict the flexibility and adaptability of queries.
4. Reduced Search Accuracy: Due to limited semantic understanding, existing systems may not deliver highly accurate search results.

PROPOSED SYSTEM

In this paper, we propose a secure verifiable semantic searching 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 readymade 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:

Treating the matching between queries and

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

documents as an optimal matching task, we explore the fundamental theorems of linear programming (LP) to propose a secure verifiable semantic searching scheme that performs semantic optimal matching on the ciphertext.

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.

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.

ADVANTAGES

1. If the distributor sees “enough evidence” that an agent leaked data, wemay stop doing business with him, or may initiate legal proceedings.
2. In this project we develop a model for assessing the “guilt” of agents.

SYSTEM ARCHITECTURE

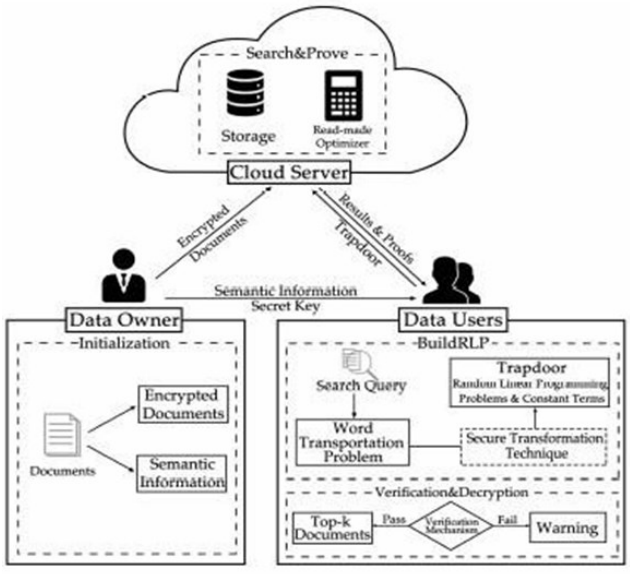


Fig. SYSTEM ARCHITECTURE

IV. IMPLEMENTATION

MODULES

- CLOUD SERVER
- DATAOWNER
- DATAUSER

MODULE DESCRIPTION

CLOUD SERVER:

The cloud server is the infrastructure provided by the cloud service provider where the data is stored and processed. It acts as the intermediary between the data owner and the data user, facilitating the secure retrieval and processing of the data during the semantic search process. The cloud server ensures that the data remains protected and accessible only to authorized users while maintaining the necessary privacy and security measures.

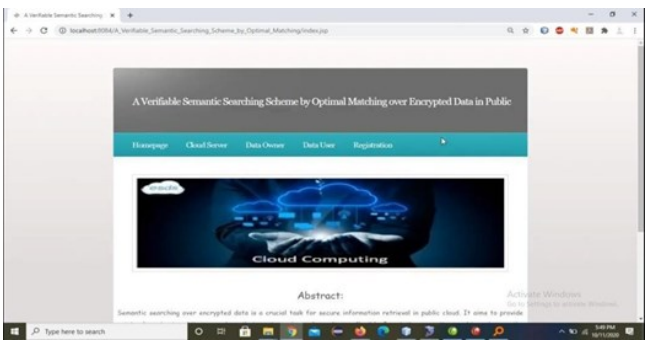
DATA OWNER:

The data owner is the entity that owns and controls the data being searched. They have the authority to decide how the data is stored, accessed, and shared. The data owner's role is crucial in ensuring the privacy and security of the data during the search process. They are responsible for managing and protecting the data, determining who has access to it, and ensuring its privacy and security. The data owner holds the rights and responsibilities associated with the data, including making decisions about its storage, usage, and sharing.

DATA USER:

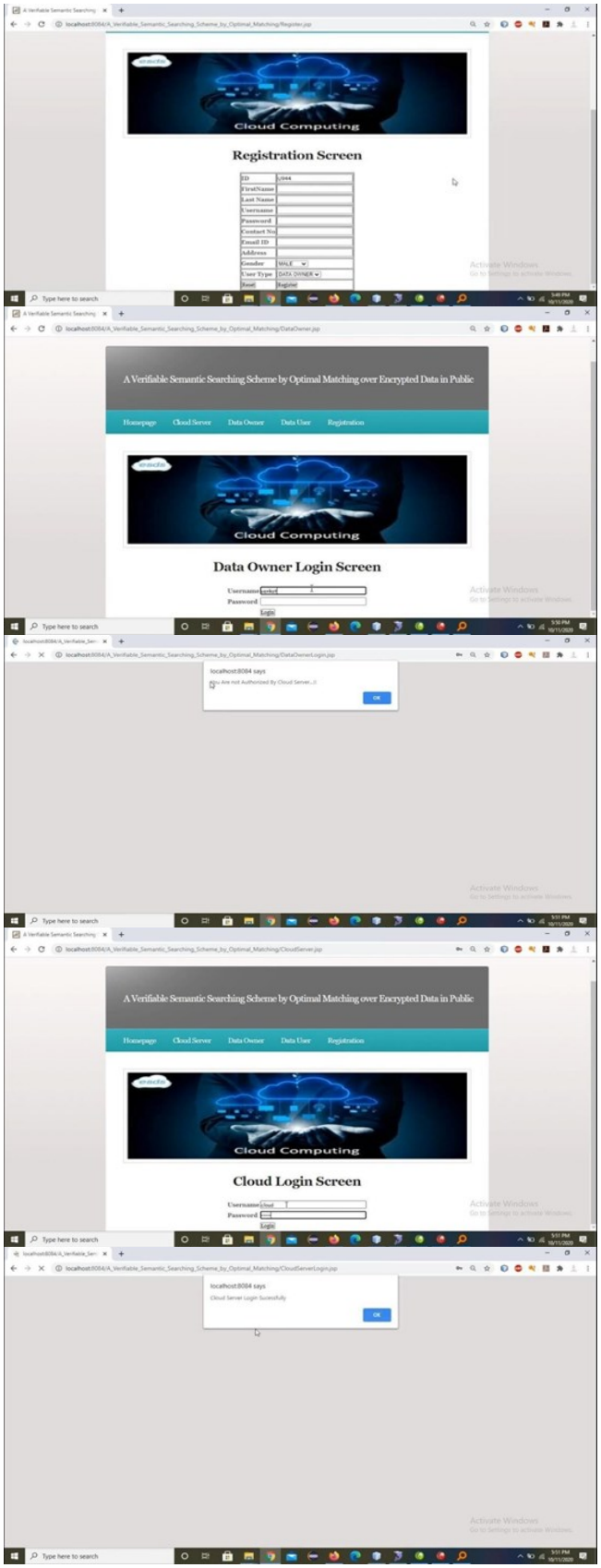
The data user is the individual or entity that performs the search on the data owned by the data owner. The data user is granted access to the data based on permissions and authorization set by the data owner. They utilize the secure semantic search functionality provided by the public cloud to retrieve relevant information from the data while preserving privacy and security.

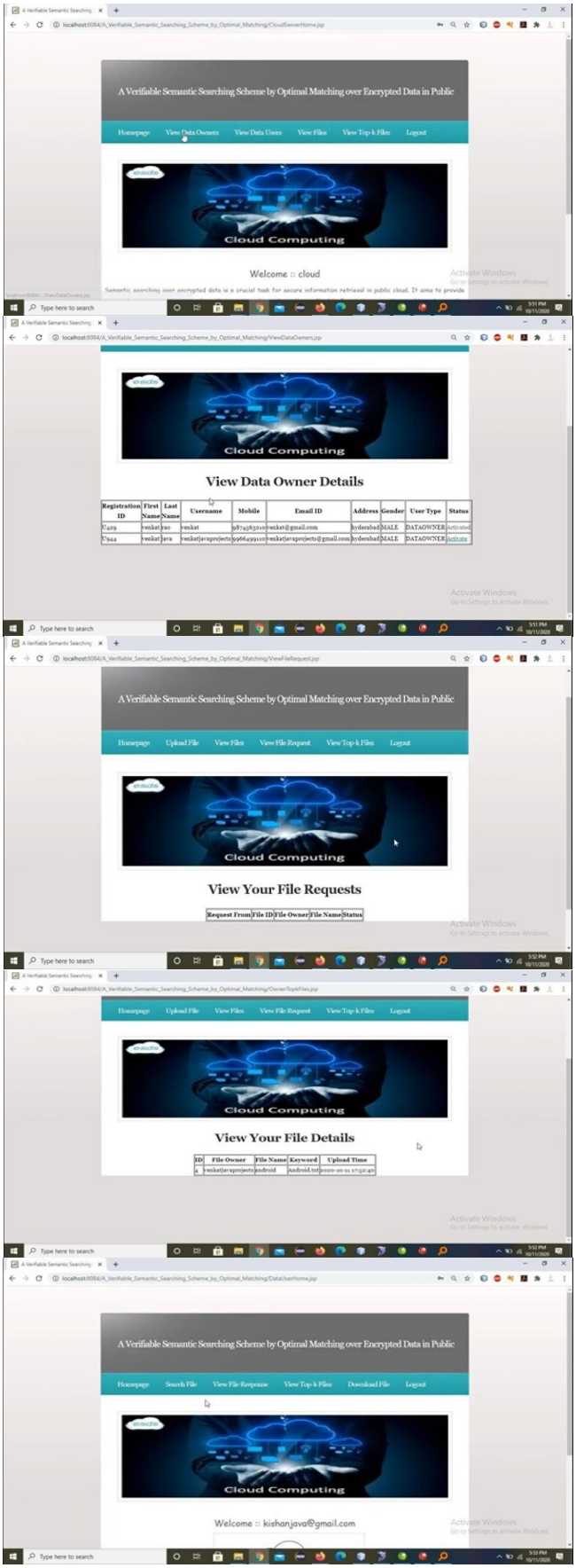
V. SCREENSHOTS:



Index in Cosmos
JUNE 2025, Volume 15, ISSUE 2
UGC Approved Journal

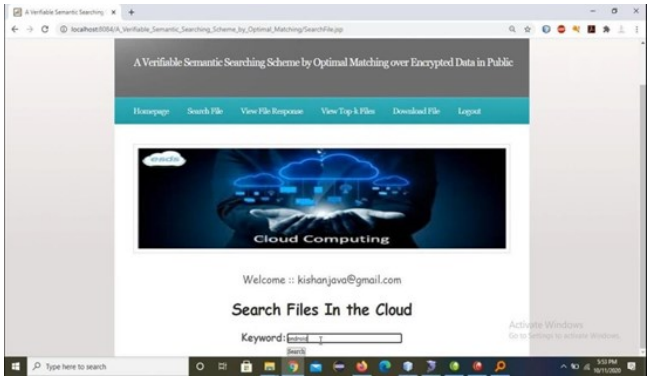
www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86





Index in Cosmos
JUNE 2025, Volume 15, ISSUE 2
UGC Approved Journal

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86



VI. CONCLUSION

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 trans-form WT problems into random LP problems. Therefore, our scheme is simple to deploy in practice as any readymade 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 this mantic-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.

FUTURE SCOPE

The future scope for a secure semantic search in public clouds project holds considerable potential. Advancements in data security, enhanced semantic analysis, AI integration, multimodal search, cross-platform compatibility, and industry-specific solutions are all areas for growth. Additionally, ensuring scalability, compliance with evolving regulations, and user privacy features will be essential for long-term success. Collaboration with industry domains, feedback loops for user input, and ongoing research will help the project remain at the forefront of technological advancements, ensuring its continued relevance and effectiveness in an everchanging digital landscape

REFERENCES

1. D. X. Song, D. Wagner, and A. Perrig, “Practical techniques for searcheson encrypted data,” inProc. 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 cloudsearch services: multikeyword ranked search over encrypted cloud datasupporting 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 datain cloud computing,” inProc. IEEE. Int. Conf. High Perform. Comput.Simul., 2014, pp. 382–390.
5. N. Jadhav, J. Nikam, and S. Bahekar, “Semantic search supportingsimilarity ranking over encrypted private cloud data,”Int. J. EmergingEng. 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 semanticexpansion based search over encrypted cloud data supporting

similarityranking,"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 searchingover encrypted data for cloud computing,"IEEE Trans. Inf. ForensicsSecurity, vol. 13, no. 9, pp. 2359–2371, Sep. 2018.
8. Z. J. Fu, X. L. Wu, Q. Wang, and K. Ren, "Enabling central keyword-based semantic extension search over encrypted outsourced data,"IEEETrans. Inf. Forensics Security., vol. 12, no. 12, pp.2986–2997, 2017.
9. Y. G. Liu and Z. J. Fu, "Secure search service based on word2vec inthe public cloud,"Int. J. Comput. Sci. Eng., vol. 18, no. 3, pp. 305–313,2019.
10. E. J. Goh, "Secure indexes."IACR Cryptology ePrint Archive, vol. 2003,pp. 216–234, 2003.