International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2023.1204001 |

A VERIFIABLE SEMANTIC SEARCHING SCHEME BY OPTIMAL MATCHING OVER ENCRYPTED DATA IN PUBLIC CLOUD

Dr. C. Dastagiraiah, Madhiraju Sushmitha, K Jashwanth Roy, M Vamshi Krishna

Associate Professor, Department of CSE, Anurag University, Hyderabad Telangana, India

U.G. Student, Department of CSE, Anurag University, Hyderabad Telangana, India

U.G. Student, Department of CSE, Anurag University, Hyderabad Telangana, India

U.G. Student, Department of CSE, Anurag University, Hyderabad Telangana, India

ABSTRACT: Semantic searching over encrypted data is a critical endeavor for ensuring secure information retrieval in the public cloud, offering flexibility in querying and search result generation. Current semantic searching schemes lack verifiability, as they depend on forecasted results from predefined keywords for verifying search outcomes, expanding queries on plaintext, and performing exact matching with semantically extended words and predefined keywords, thereby limiting accuracy. To address these challenges, this paper proposes a novel secure verifiable semantic searching scheme. It introduces a formulation of the Word Transportation (WT) problem to achieve semantic optimal matching on ciphertext, aiming to calculate the Minimum Word Transportation Cost (MWTC) as a measure of similarity between queries and documents. Additionally, a secure transformation method is proposed to convert WT problems into random Linear Programming (LP) problems, enabling the computation of encrypted MWTC. For ensuring verifiability, the duality theorem of LP is explored to design a verification mechanism leveraging intermediate data generated during the matching process, thereby verifying the correctness of search results. A comprehensive security analysis is conducted, demonstrating that the proposed scheme can guarantee both verifiability and confidentiality. Experimental evaluations are conducted on two datasets, showcasing the superior accuracy of the proposed scheme compared to existing approaches. By enabling flexible retrieval services for arbitrary words and addressing the limitations of existing schemes, the proposed semantic searching framework offers a robust solution for secure information retrieval in the cloud environment. Additionally, its ability to ensure verifiability enhances the trustworthiness of search outcomes, making it suitable for a wide range of applications where data security and accuracy are paramount concerns.

KEYWORDS: Semantic searching, Encrypted data retrieval, Verifiability, Cloud security, Linear programming

I. Introduction

The project focuses on addressing the critical need for secure information retrieval in the public cloud through semantic searching over encrypted data. Existing semantic searching schemes lack verifiability, relying on predefined keywords and plaintext expansion for search accuracy. To overcome this limitation, a novel secure verifiable semantic searching scheme is proposed. The approach involves formulating a Word Transportation (WT) problem to calculate the Minimum Word Transportation Cost (MWTC) as a measure of similarity between queries and documents on ciphertext. A secure transformation method is introduced to convert WT problems into random Linear Programming (LP) problems, ensuring encryption of MWTC. Verifiability is achieved by leveraging the LP duality theorem to design a mechanism that utilizes intermediate data generated during the matching process to verify search results' correctness. Security analysis confirms the scheme's ability to guarantee both verifiability and confidentiality. Experimental validation on two datasets demonstrates the scheme's superior accuracy compared to existing approaches. Overall, the proposed scheme offers a flexible and secure solution for semantic searching over

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2023.1204001 |

encrypted data, enhancing information retrieval capabilities in the public cloud while ensuring data privacy and integrity.

II. Related work

Over the past two decades, searchable encryption has garnered significant attention due to its practicality, facilitating secure information retrieval over encrypted cloud data. Numerous works have concentrated on enhancing both the security and functionality of searchable encryption schemes. Regarding security, scholars have formulated various definitions and attack patterns to evaluate the robustness of existing schemes. Goh et al. introduced a security model termed semantic security against adaptive Chosen Keyword Attack (IND-CKA), ensuring document indexes conceal document contents. Curtmola et al. extended security definitions, including chosen-keyword attacks and adaptive chosen-keyword attacks. Additionally, privacy concerns led to the introduction of access pattern disclosure and novel attacks, such as search pattern leakage.

In parallel, research efforts have focused on enhancing functionality to meet practical demands. This includes ranked search capabilities, where cloud servers evaluate relevance scores between queries and documents without disclosing sensitive information. Cao et al. proposed a privacy-preserving Multi Keyword Ranked Search Scheme (MRSS), utilizing binary vectors and secure kNN algorithms. Furthermore, advancements in homomorphic encryption facilitated multi keyword ranked search schemes, enabling relevance score encryption and calculation under the vector space model.

Furthermore, the incorporation of semantic information has been pivotal in enhancing search accuracy. Traditional schemes often fail to exploit semantic relationships between words, limiting evaluation of query-document relevance. Fu et al. pioneered synonym searchable encryption, extending keyword sets using synonym thesauri and integrating secure indexes. Xia et al. introduced semantic extension searching schemes based on concept hierarchies, improving accuracy by weighting query words based on grammatical relations and extending central words using hierarchy trees.

Finally, the quest for verifiable searching over encrypted data has led to innovative approaches ensuring the correctness of search results. Some schemes verify the presence of encrypted documents containing specific query words, while others focus on verifying ranked search results. Wang et al. proposed a single keyword ranked verification scheme based on hash chains, while Sun et al. introduced a multi-keyword ranked verifiable searching scheme using Merkle Hash trees and cryptographic signatures. However, challenges remain in supporting semantic searching and minimizing communication overhead, especially in multi-data owner scenarios.

III. METHODOLOGY

3.1 Proposed Method

The proposed system aims to address the limitations of existing semantic retrieval schemes by introducing a secure and verifiable semantic retrieval approach. First, the system formulates a word transportation (WT) problem and calculates the minimum word transportation cost (MWTC) as a similarity measure between a query and a document. In this paper, we propose a secure and verifiable semantic retrieval scheme that considers matching between queries and documents as an optimal matching task. We consider document words as "suppliers", query words as "consumers" and semantic information as "products", and design the minimum word transfer cost (MWTC) as a measure of similarity between queries and documents. Therefore, we introduce word embeddings to represent words, calculate the Euclidean distance as the similarity distance between words, and formulate the word transportation (WT) problem based on the word embedding representation. However, the cloud server may learn the sensitive information of the WT problem, such as the similarity between words. We also propose a secure transformation to transform the WT problem into a stochastic linear programming (LP) problem for optimal ciphertext semantic matching. In this way, the cloud can use an off-the-shelf optimizer to solve the RLP problem without learning any sensitive information and obtain the encrypted MWTC as a measurement. Considering that cloud servers may be imprecise and return incorrect/fabricated

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2023.1204001 |

search results, we investigate the linear programming (LP) duality theorem that the intermediate data generated in the matching process must satisfy a set of necessary and sufficient conditions such that, You can check whether the cloud correctly solves the RLP problem and further check the accuracy of the search results.

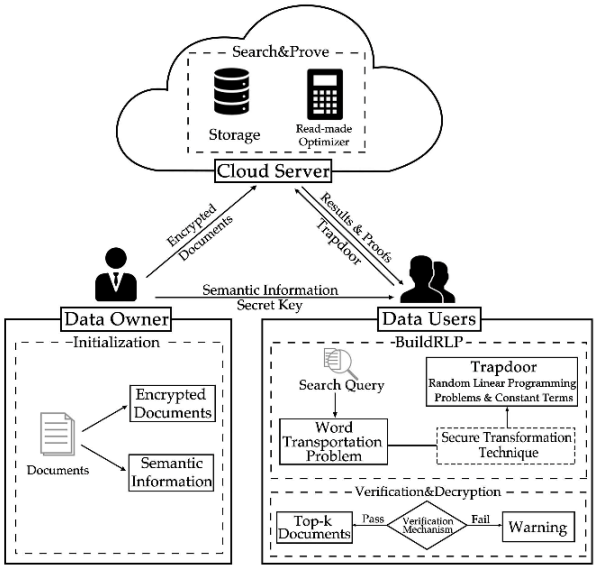

Figure 1 System Architecture

3.2 Example, consider a query "machine learning" and a document containing the phrase "artificial intelligence and machine learning." The system calculates the MWTC by determining the minimum cost to transform the words in the query to those in the document, considering semantic relationships and context.

Next, the system transforms the WT problems into random Linear Programming (LP) problems to obtain encrypted MWTC. This ensures that the matching process is performed securely over encrypted data, preserving confidentiality. To achieve verifiability, the system leverages the duality theorem of LP to design a verification mechanism. This mechanism uses intermediate data produced during the matching process to verify the correctness of search results. For instance, if the search result claims a high similarity score between the query and a document, the verification mechanism checks the consistency of this score with the encrypted MWTC obtained during the matching process. By combining semantic matching, encryption, and verification mechanisms, the proposed system ensures both accuracy and security in semantic searching over encrypted data.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2023.1204001 |

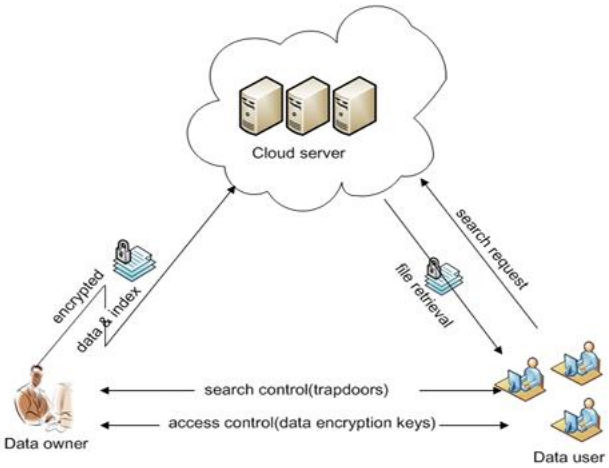

Figure 2 Framework of search encrypted cloud data

3.3 Experiment Modules

The various aspects of the implementation with hypothetical scenarios to illustrate how the project functions.

Here's a breakdown of the software architecture:

3.3.1 Login Page:

This page serves as the entry point for users, allowing them to authenticate and access the system's features securely.

3.1.2 User Registration Page:

Here, new users can create accounts by providing necessary information, such as username, email, and password, enabling them to utilize the system's functionalities.

3.1.3 Data Owner Registration Page:

This page facilitates the registration process for data owners, enabling them to register their data securely within the system for subsequent management and retrieval.

3.1.4 File Upload Page:

Users can utilize this page to upload files securely to the system, ensuring that their data is stored and managed in a protected environment.

3.1.5 Verification Page:

Upon certain actions or transactions, this page prompts users to verify their identity or confirm specific details, enhancing security and trust within the system's operations.

IV. EXPERIMENTAL RESULTS

4.1 The proposed secure verifiable semantic searching scheme addresses limitations of existing approaches by introducing a novel method based on the word transportation problem and random linear programming. Experimental results on two datasets validate the effectiveness of the proposed scheme, exhibiting higher accuracy compared to prior approaches.

The proposed system offers several advantages over existing solutions in the field of searchable encryption and semantic searching:

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2023.1204001 |

- Verifiability: One of the key advantages is the integration of a verification mechanism, ensuring the correctness of search results. By leveraging the LP duality theorem and intermediate data from the matching process, users can verify the integrity of the retrieved documents without compromising security.
- Enhanced Security: The scheme guarantees both verifiability and confidentiality. Security analysis demonstrates the robustness of the proposed system against various attacks, thereby ensuring the privacy of the stored data and search queries.
- Semantic Searching: Unlike traditional searchable encryption schemes, which often fail to utilize semantic information effectively, the proposed system incorporates semantic searching capabilities. By formulating word transportation problems and leveraging random LP transformations, it can evaluate the relevance between queries and documents more accurately, improving search accuracy.
- Flexibility: The system aims to provide retrieval services for arbitrary words, allowing for flexible queries and search results. This flexibility is crucial in real- world applications where users may need to search for a wide range of terms and concepts.
- Optimal Matching: Through the formulation of the word transportation problem, the system calculates the Minimum Word Transportation Cost (MWTC) as the similarity measure between queries and documents. This approach ensures optimal matching on ciphertext, leading to more precise search results.
- Experimental Validation: The proposed system's efficacy is supported by experimental results on two datasets, demonstrating higher accuracy compared to existing schemes. This empirical validation reinforces the practical utility and effectiveness of the proposed approach.

The proposed system offers a comprehensive solution that addresses the limitations of existing searchable encryption schemes by providing verifiability, enhanced security, semantic searching capabilities, flexibility, optimal matching, and empirical validation of its effectiveness. These advantages make it a promising approach for secure information retrieval in public cloud environments.

V. 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.

VI. REFERENCES

[1] Guoxiu Liu, Geng Yang, Shuangjie Bai and Qiang Zhou “Effective Fuzzy Semantic Searchable Encryption Scheme Over Encrypted Cloud Data”, IEEE,2020

[2] S. Tahir, S. Ruj, Y. Rahulamathavan, M. Rajarajan and C. Glackin, "A new secure and lightweight searchable encryption scheme over encrypted cloud data", IEEE Trans. Emerg. Topics Comput., vol. 7, no. 4, pp. 530-544, Oct. 2019.

[3] Xuelong Dai, Hua Dai, Chunming Rong and Fu Xiao,”Enhanced Semantic-Aware Multi-Keyword Ranked Search Scheme Over Encrypted Cloud Data”Oct.-Dec. 2022, pp. 2595-2612, vol. 10

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2023.1204001 |

[4] H. Shen, L. Xue, H. Wang, L. Zhang, and J. Zhang.B+-tree based multi-keyword ranked similarity search scheme over encrypted cloud data.IEEE Access, 9:150865–150877, 2021a.
[5] D. Sharma.Searchable encryption : A survey.Information Security Journal: A Global Perspective, 32(2):76–119, Mar. 2023.
[6] P. Srivani, S. Ramachandram, and R. Sridevi.Multi-key searchable encryption technique for index-based searching.International Journal of Advanced Intelligence Paradigms, 22(1-2):84–98, Jan. 2022.
[7] G. Sucharitha, V. Sitharamulu, S. N. Mohanty, A. Matta, and D. Jose.Enhancing secure communication in the cloud through blockchain assisted-cp-dabe.IEEE Access, 11:99005–99015, 2023.
[8] Y. Tang, Y. Chen, Y. Luo, S. Dong, and T. Li.VR-PEKS: A Verifiable and Resistant to Keyword Guess Attack Public Key Encryption with Keyword Search Scheme, 2023.
[9] M. Ali, H. He, A. Hussain, M. Hussain, and Y. Yuan.Efficient Secure Privacy Preserving Multi Keywords Rank Search over Encrypted Data in Cloud Computing.Journal of Information Security and Applications, 75:103500, 2023.
[10] Lanxiang Chen; Yujie Xue; Yi Mu; Lingfang Zeng; Fatemeh Rezaeibagha," CASE-SSE: Context-Aware Semantically Extensible Searchable Symmetric Encryption for Encrypted Cloud Data" ,IEEE,March-April 2023D. X. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in Proc. IEEE Symp. Secur. Privacy, 2000, pp. 44– 55.
[11] 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.
[12] 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
[13] 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.
[14] 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.
[15] Z. J. Fu, X. L. Wu, Q. Wang, and K. Ren, "Enabling central keywordbased semantic extension search over encrypted outsourced data," IEEE Trans. Inf. Forensics Security., vol. 12, no. 12, pp. 2986–2997, 2017.
[16] 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. [10] E. J. Goh, "Secure indexes." IACR Cryptology ePrint Archive, vol. 2003, pp. 216–234, 2003.
[17] 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.
[18] 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.