Industrial Engineering Journal
ISSN: 0970-2555

Volume : 53, Issue 5, May : 2024

SECURE SEMEANTIC SEARCH OPTIMAL MATCHING OVER ENCRYPTED DATA

Dr. M . Swapna

Dr. M. Swapna, *Assistant Professor CSE, Vaagdevi College of Engineering(Autonomous), India*

B.Harshitha,UG student, CSE, Vaagdevi College of Engineering(Autonomous),India

R.Nikitha,UG student, CSE, Vaagdevi College of Engineering(Autonomous),India

T.Pavankumar,UG student CSE, Vaagdevi College of Engineering(Autonomous), India

B.Praveen,UG student CSE, Vaagdevi College of Engineering(Autonomous),India

Industrial Engineering Journal
ISSN: 0970-2555

Volume : 53, Issue 5, May : 2024

SECURE SEMEANTIC SEARCH OPTIMAL MATCHING OVER ENCRYPTED DATA

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.

Keywords:

Industrial Engineering Journal
ISSN: 0970-2555

Volume : 53, Issue 5, May : 2024

1.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. [1] 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. [2] 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 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 schemesconsider the semantic relationship among words to performquery expansion on the plaintext, then still use the query wordsand extended semantically related words to perform exactmatching with the specific keywords in outsourced documents.We can roughly divide these schemes into three categories:secure semantic searching based synonym [3], [4], securesemantic 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 elementarysemantic information among words. For example, synonymschemes only use synonym attributes; mutual informationmodels only use the co-occurrences information. AlthoughLiu et al. [9] introduce the Word2vec technique to utilizethe semantic information of word embeddings, their approachdamages the semantic information due to straightly aggre-gating all the word vectors. We think that secure semanticsearching schemes should further utilize a wealth of semanticinformation among words and perform optimal matching ontheciphertext for high search accuracy.

In this paper, we propose a secure verifiable semantic searching scheme that treats matching between queries anddocuments as an optimal matching task. We treat the documentwords as “suppliers,” the query words as “consumers,” and thesemantic information as “product,” and design the minimumword transportation cost (MWTC) as the similarity metricbetween queries and documents. Therefore, we introduce wordembeddings to represent words and compute Euclidean dis-tance as the similarity distance between words, then formulatethe word transportation (WT) problems based on the wordembeddings representation. However, the cloud server couldlearn sensitive information in the WT problems, such as thesimilarity between words. For semantic optimal matching ontheciphertext, we further propose a secure transformationto 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 theencrypted MWTC as measurements without learning sensitiveinformation. Considering the cloud server may be dishonestto return wrong/forged search results, we explore the dualitytheorem of linear programming (LP) and derive a set ofnecessary and sufficient conditions that the intermediate dataproduced in the matching process must satisfy. Thus, we canverify whether the cloud solves correctly RLP problems andfurther confirm the correctness of search results. Our new ideasare summarized as follows:

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

2. For secure semantic optimal matching on the ciphertext,we formulate the word transportation (WT) problem andpropose a secure transformation technique to transformWT problems into random linear programming (LP)problems for obtaining the encrypted minimum wordtransportation cost as measurements between queries anddocuments.

3. For supporting verifiable searching, we explore the dual-ity theorem of LP and present a novel insight that usingthe intermediate data produced in the matching processas proof to verify the correctness of search results.

2. LITERATURE SURVEY

Practical techniques for searches on encrypted data

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. Our 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.

Smart cloud search services verifiable keyword-based semantic search over encrypted cloud data

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.

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

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 multikeyword ranked searching in a cloud environment.

3. PROBLEM STATEMENT

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

Industrial Engineering Journal
ISSN: 0970-2555

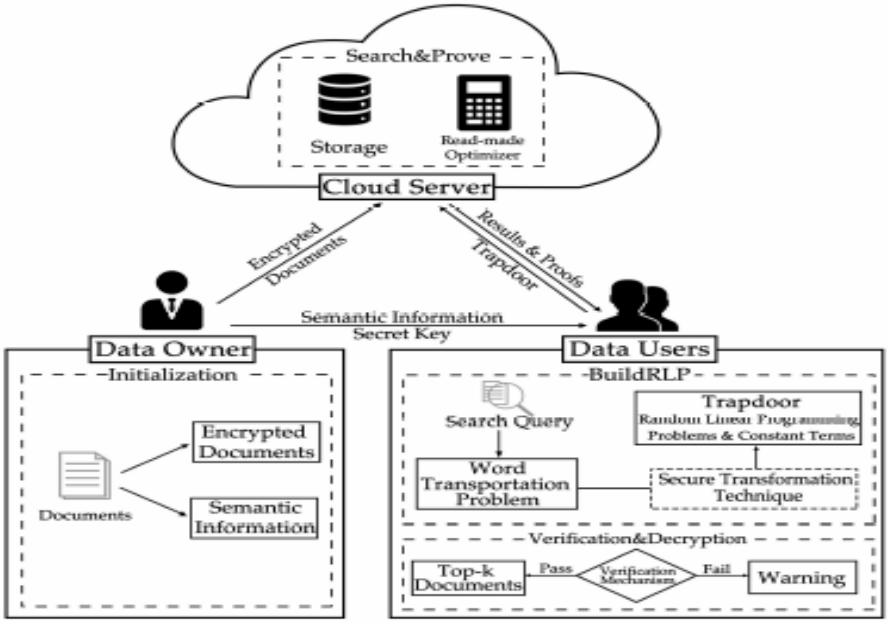
Volume : 53, Issue 5, May : 2024

semantic information among words and perform optimal matching on the ciphertext for high search accuracy.

4. PROPOSED SYSTEM

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.

5. ARCHITECTURE



Industrial Engineering Journal
ISSN: 0970-2555

Volume : 53, Issue 5, May : 2024

6. IMPLEMENTATION

Modules:

1.Cloud Server

View Dataowner

View Datausers

View Files

View Top-K Files

Logout

2.DataOwner

Upload File

View Files

View File Request

View Top-k Files

Logout

3.DataUser

Search File

View File Response

View Top-k Files

Download File

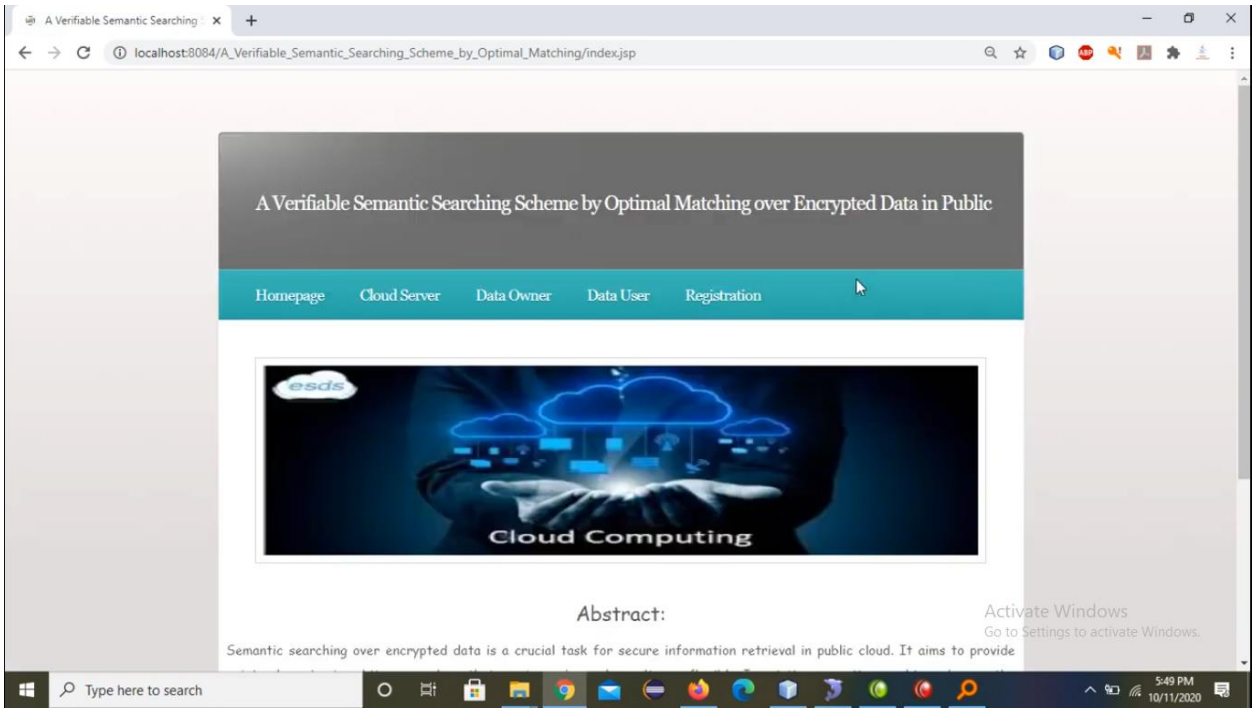
Logout

Industrial Engineering Journal
ISSN: 0970-2555

Volume : 53, Issue 5, May : 2024

7. OUTPUT EXPERIMENTS

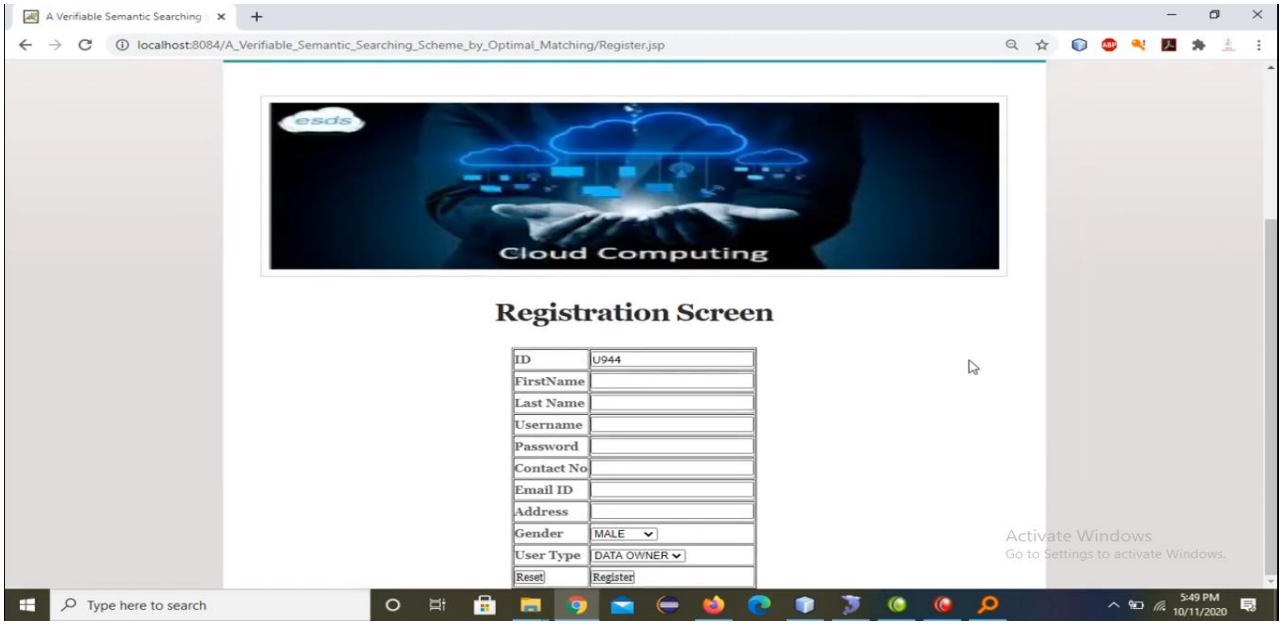
Home Screen



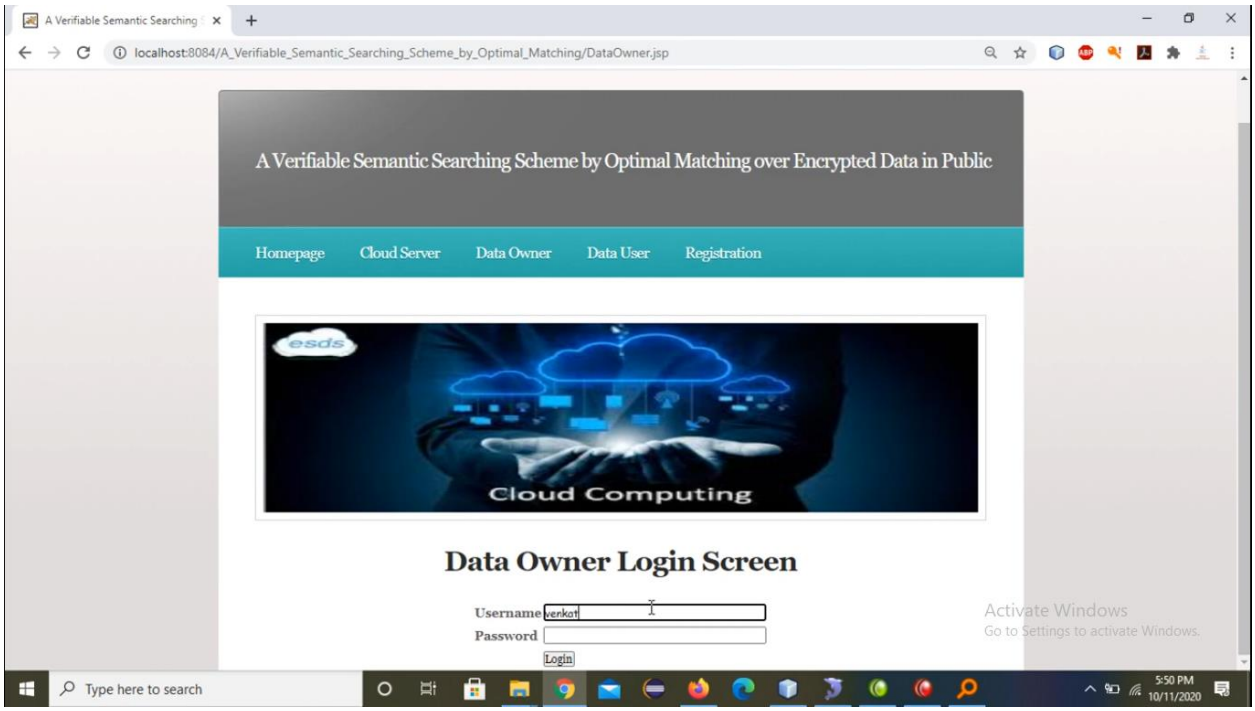
Registration page

Industrial Engineering Journal
ISSN: 0970-2555

Volume : 53, Issue 5, May : 2024



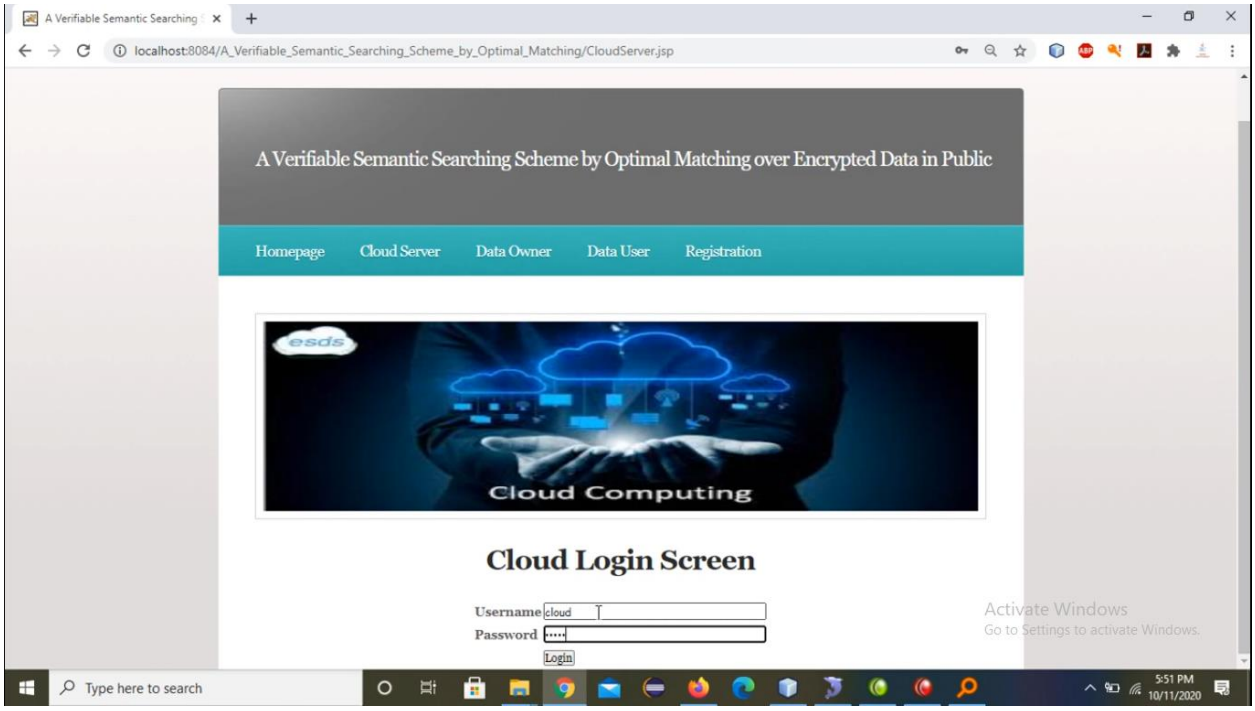
Data Owner Login



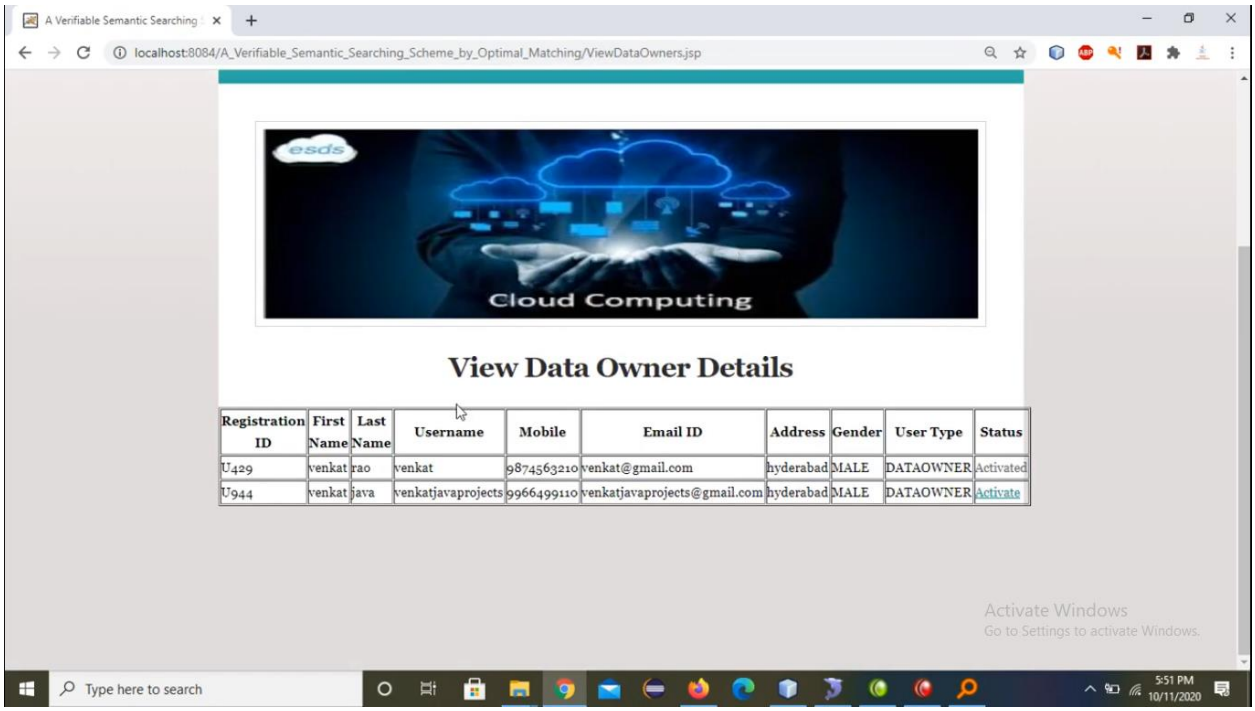
Industrial Engineering Journal
ISSN: 0970-2555

Volume : 53, Issue 5, May : 2024

Cloud Login



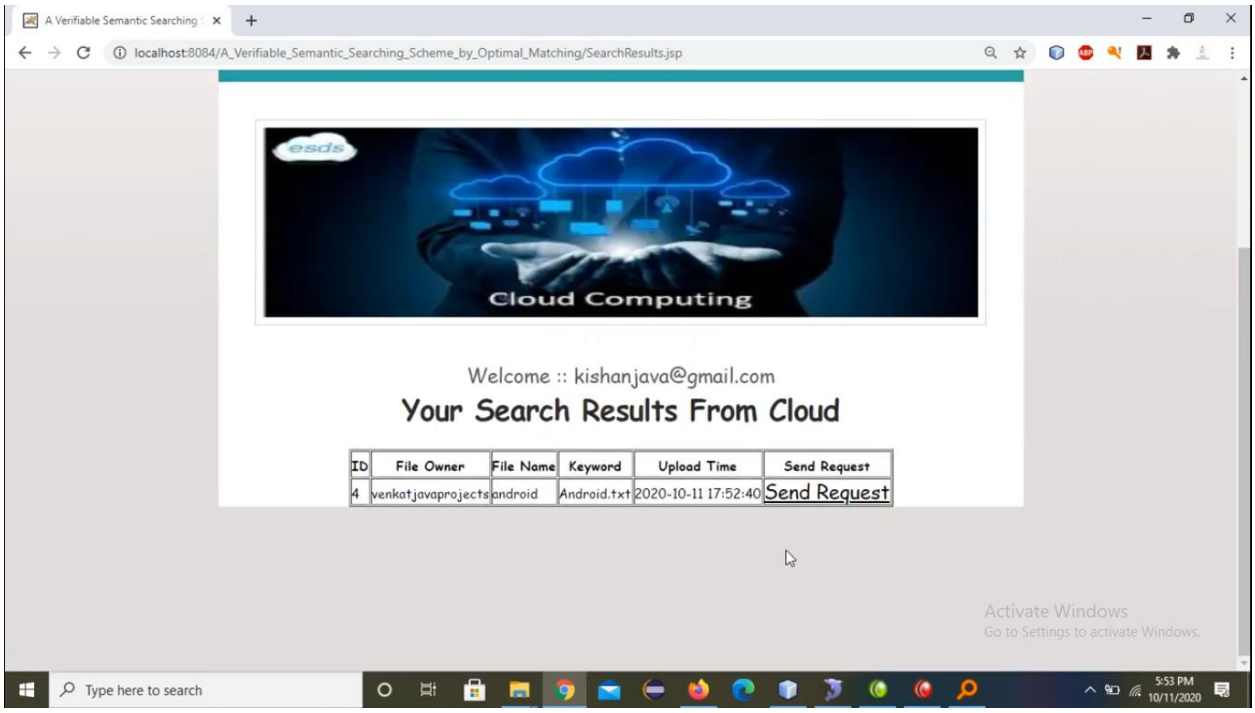
Owner Details



Industrial Engineering Journal
ISSN: 0970-2555

Volume : 53, Issue 5, May : 2024

Search Results



8. 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 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 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.

9. 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: multi-keyword 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