Semantic Searching Scheme Over Encrypted Data in Pubic Cloud

Mrs.Shaiqua Khan[1],Sakshi Chatap[2],Mrunal Marotkar[3],Avanti Tagde[4],Gaurav Urmale[5]

[1](Department of Computer Science & Engineering/ G H Raisoni University Saikheda Chhindwara ,India)

[2](Departmen of Computer Science & Engineering/ G H Raisoni University Saikheda Chhindwara , India)

Abstract: The end of this design is give semantic searching over translated data is a pivotal task for secure information reclamation in public pall. We propose a secure empirical 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, So we propose a secure metamorphosis to transfigure WT problems into arbitrary Linear Programming(LP) problems to gain the translated MWTC for verifiability, we explore the duality theorem of LP to design a verification medium using the intermediate data produced in matching process to corroborate the correctness of hunt results. Security analysis demonstrates that our scheme can give the verifiability and confidentiality to give high delicacy than other schemes.

Key Word Secure semantic searching, verifiable searching, word transportation, public cloud.

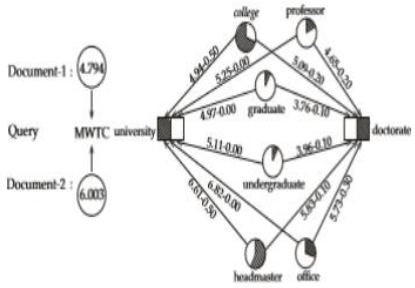
I. INTRODUCTION

Cloud storage is becoming more and more popular in the recent trend as it provides many benefits over the traditional storage solutions. With cloud storage, corporations can purchase only the needed amount of storage from the cloud storage provider (CSP) to fulfill their storage needs instead of maintaining their own data storage infrastructures. They can rely on CSP to handle all data maintenance tasks such as backup and recovery. It also allows all data to be accessed remotely in order to streamline their operation among different locations. With all these benefits, companies can significantly reduce their operation cost by simply outsourcing their business data to cloud storage. Beside these benefits that provided by the cloud storage, however, many security problems arise in cloud storage that prevent companies from migrating their data to cloud storage Due to the facts that cloud storage is usually hosted by third party provider other than the data owners and cloud storage infrastructure is usually shared among different users, data stored in cloud storage can be easily targeted by the masquerade attack and the insider data theft attack .

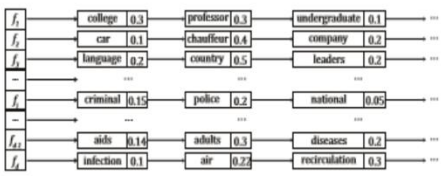
These attacks threaten the data security and the data privacy of the stored data, as result, the data owners cannot rely on CSP to secure their confidential data. These attacks also induce the data owners to encrypt all their sensitive data such as the social security numbers (SSN), credit card information, and personal tax information before they can be saved in cloud storage. The encryption approach may have strengthened the data security of cloud data, but it has also degraded the data efficiency because the encryption will reduce the searchability of the data. Especially in the cloud computing environment, it is impractical for the user to download and decrypt the entire encrypted data from the remote cloud server before a search can occur. Therefore, an efficient scheme that supports search over encrypted data in cloud computing becomes very significant before many enterprises can take advantage of the cloud storage. Many schemes were proposed in recent researches that enable keyword search over encrypted data in cloud computing. The most common approach of these schemes is indexing the keywords contain in each uploading data file

II. Material And Methods

This section presents the main proposed methods, including the secure conversion method shown in Figure 1, the word passing problem, and the verification mechanism .Best Transport Fit is an example of a term.



The word weight is expressed as the relative area of the shadow. The line length represents . This is a rough Euclidean distinction between two related words. Segments numbered M-N represent transmissions containing two concatenated words. In this case, the MWTC between document and query 1 is 4.794, and the MWTC between query and document 2 is 6.003, indicating that the question is more relevant compared to documents 2 and 1.



Problem of Transporting Words for the Best Matching

How to say "transit" (WT) problem as an optimal transport linear programming problem that handles requests Match documents with the ideal matching operation. as pictured. 2 Find the smallest possible word Transfer costs, we use the WT task (MWTC) to compare documents with requests for similarity. Us Enter the index directly into the document's semantic data. Publishes a paper in the WT issue. Figure 3 shows an example. A study on direct indexing. The distribution of keywords for a document is defined as the direct index weight for: all keywords. As a result, you need to select keywords for each page and determine the importance of each keyword all documents. Searches for keywords in the system using term frequency inverse document frequency (TF-IDF). A criterion that does not lose its generality.

Furthermore, we establish weights utilising (1): weight(w,f) =1 |fi|·(1 + lnfi,w)·ln1 + d fw, (1)where w is a specific keyword, f is a specific document, |fi| fi,w to indicate the length of the document

The number of documents containing the phrase w is denoted by the term "frequency TF" of the keyword w in f, and d

$$\begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}$$

The total number of documents in the data set. Query term weights are defined equally and applied equally

How to submit a request. In this study, we standardized the weight of each document or query to 1. Documents are treated as "suppliers", words used in queries as "consumers", and semantic data as "products" provided. Send documents and request indexing.

Therefore, Given the forward index of a document f and the query q, we may state the WT problem as follows.:

WT(f,q) =min

m X i=1

n X j=1

fi,jdi,j

depending on n X j=1

fi,j = efi , i = 1,2,3 ...,m

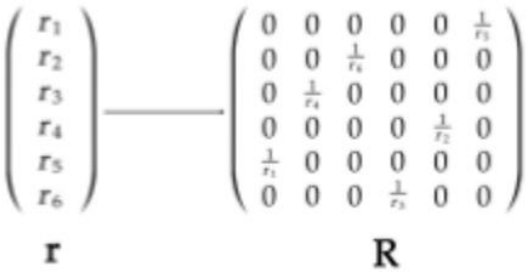
m X i=1

fi,j = eqj,j = 1,2,...,n fi,j ≥ 0 m X i=1 n X j=1 fi,j = 1,

Secure Transformation Technique

Because the basic WT problem potentially reveal sensitive information, word transportation concerns cannot be immediately adapted to the secure semantic searching strategy. As a result, we present a method of safe transformation for realising semantically ideal ciphertext matching, ensuring information Transparency and honesty in word transportation situations. Users in our system use our safe a method for transforming WT issues are transformed into random linear programming (RLP) challenges, which the cloud may solve using any ready-made optimizer, and obtain without learning important information the minimum word transmission cost in encrypted

form (EMWTC). Every WT problem is encrypted specifically via our secure transformation method. ψ = (c,V,W,I) using a unique secret key KT = (A,Q,γ,R,r), in which A is an mn×mn invertible random matrix, Q is an invertible, (m n)(m n) random matrix., γ is a genuine benefit, An mn×1 random vector is r. R is a generalized permutation matrix of size mn×mn. original The objective function cTx is first transformed into an encrypted version cTAy-cTr where x = Ay r. One Possible Solution to Your Problem An RLP problem is represented by the symbol y, representing an mn × 1 solution vector. Each ri must be at least 0. where i=1, 2,..., mn. Ay Modify the original WT release by using x instead of -r. to ψ (4). In (4), the constraint IAy ≥ Ir is defined similarly to the ith component a = IAy of vector T1. Does not fall below the ith component of vector T2 = Ir with i=1,2,...,mn. Minimum cTAy−cTr according to VAy = W VrIAy ≥ Ir



$$\mathbf{r} \qquad \mathbf{R}$$

The most important component of the secret key KT is the matrix R. It has non-speech requirements and is used to hide. Confidential information in the sense of transmission problems Ix ≥ 0. Reciprocal values of random components Vector r - the nonzero element of R. Thanks to the identity matrix I, we see that I IA = A. So we fix Original WT issue. ψ ~ . In the constraint is defined as Ay ≥ r, which means that the ith a No component of the vector T3 = Ay falls below the ith part of vector r. where i=1,2,...,mn.min QVAy = Q(W Vr) γcTAy−γcTr according to Ay ≥ r

Result Verification Mechanism

We built a result verification system that verifies using intermediate data generated through matching. Reliability of search results. Since best ciphertext matching is a linear programming (LP) problem, To We created a verification method, apply the powerful LP problem theorem, and extend the theory of LP duality. double A programming task for each RLP task is initially created by ω. To create a double problem in light of (7) for ω Use the Lagrange multiplier θ. Below is maximum g(s,t) according to V0Ts I0Tt = c0 t ≥ 0 g(s,t) = W0Ts LTt.

Attribute based encryption

Attribute based encryption (ABE) is a relatively new perception of public key encryption for data-centric security solutions. Traditionally, we view encryption as a way for a user to cipher data to a specific target recipient. The user encrypts the data under the recipient's public key such that only the exact recipient holding the matching private key can decrypt it.

However, in various real-world applications, it is essential that we provide access and share data confidentially and in a trusted manner according to a given policy without prior knowledge of who the recipient is.

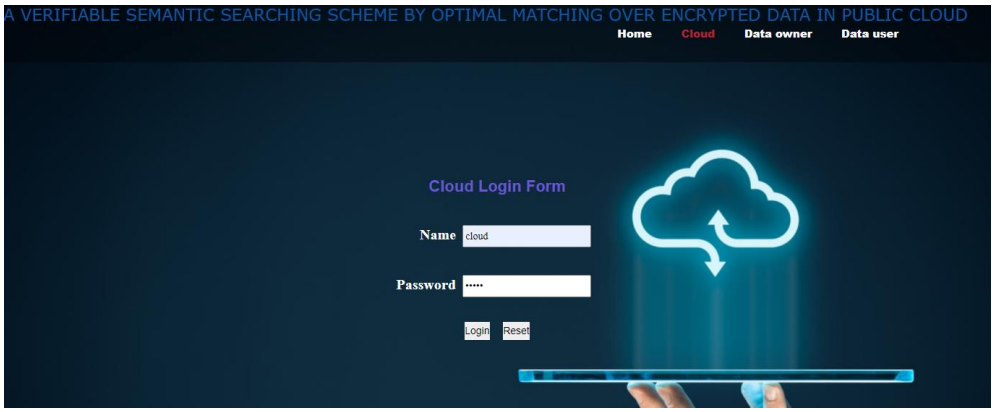
This is where ABE comes into place and offers a more scalable approach than existing public key cryptosystems. ABE can advance trusted sharing and access of data by basing access and decryption on a person's role / privileges within an organization or in other contexts, rather than by a person's specific identity.

The word weight is expressed as the relative area of the shadow. The line length represents . This is a rough Euclidean distinction between two related words. Segments numbered M-N represent transmissions containing two concatenated words. In this case, the MWTC between document and query 1 is 4.794, and the MWTC between query and document 2 is 6.003, indicating that the question is more relevant compared to documents 2 and 1.

III. Result

Home Page:

Cloud home page:



View data owners page:

IV. Discussion

In this study, we provided an arbitrary word search service so that queries and search results are flexible. Therefore, we propose a safe transformation to turn the WT problem into a random linear

Programming Problem (LP) A verification mechanism that examines the LP duality theorem and verifies the correctness of search results using intermediate data produced in the matching process in order to obtain a verifiable encrypted MWTC design.

V. Conclusion

We propose a secure and verifiable semantic search scheme that treats matching between queries and documents as an optimal matching task for word transfer. Therefore, we study basic linear programming (LP) theorems to design word transport (WT) problems and result verification mechanisms. We formulate the WT problem to compute the minimum word transfer cost (MWTC) as a similarity metric between queries and documents, and propose a more secure transformation method. Convert the WT problem to a random LP

VI. References

[1]. 1). D.X. Song, D. Wagner, and A. Perring, “Practical techniques for searches on encrypted data, "in orc. IEEE Symp. Secure. Privacy, 2000, pp. 44-55.

[2].

[3]. 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.

[4]. 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.

[5]. 4).T. S. Moh and K. H. Ho, “Efficient semantic search over encrypted data in cloud computing,” in Proc. IEEE. Int.