International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 04 | Apr 2020 www.irjet.net p-ISSN: 2395-0072

K-Gram based Composite Secret Sign Search over Encrypted Cloud Information

Aditi Gaur[1], Y.Geetha Sai[2], Shriya G.N [3]

[1,2,3]*Student, Dept. of Computer Science and Engineering, SRM Institute of Science and Technology, Tamil Nadu, India*

Abstract - *A growing number of owners of data has transferred our data to cloud servers. Cloud data owners tend to outsource information in an encrypted form for confidentiality protection purposes. Developing an effective and reliable search method for cipher-text is therefore important. One problem is that the relationship between documents is usually hidden in the encryption process, which can lead to a substantial deterioration of the efficiency of the search accuracy. Using the keyword-based app, all access data from the cloud. The stable multi-keyword ranked cloud to search for encrypted data, a top-k quest for big data encryption against breaches of privacy, and attempt to find an efficient and safe solution to this problem. It provides open operations such as downloading, removing, adding information. Using tree structure and nebulous search method here to retrieve the cloud data. These types of techniques are used to solve the keyword conjecture attack problem. This offers transparent operations such as copying, deleting, and information adding. The cloud data is extracted using a tree structure and nebulous search method here. These types of strategies are used to solve the problem of attack by keyword conjecture.*

Key Words: **Cloud computing, Multi keyword search, Java, Semantic resemblance, Encryption**

1. INTRODUCTION

Cloud computing infrastructure is ground-breaking new technology and it significantly accelerates the production of data storage, processing, and delivery on a wide scale. Nevertheless, as data owners outsource their private data to public cloud providers that are not within their trusted management jurisdictions, security and privacy become major concerns. To avoid leakage of information, sensitive data must be encrypted. Most existing works find only queries of single keywords without sufficient ranking schemes. The keyword dictionary is static with the current multi-graded search strategy, and cannot be easily expanded as the number of keywords grows. It also takes no account of the user behaviour and the frequency of keyword access. The out-of-order the ranking problem can emerge for the query matching result which contains a large number of documents. This makes it difficult for the data used to find the subset which most likely meets their requirements. In this paper, we propose a versatile multi-keyword query scheme, called MKQE to resolve the above disadvantages. MKQE significantly decreases the overhead management during the extension of the dictionary keyword. When generating the query result it takes into account keyword weights and user access history. Hence, records that have higher levels of access and that fit more to the access history of the users get higher rankings in the corresponding result collection. The purpose of the project is to provide users with various types of search options to get the full number of file searches from the cloud's encrypted data.

2. RELATED WORK

Using fuzzy search algorithms to upload the file we can create lots of keywords for each file hereby providing the statics based on different conditions, we can highlight the output report of all these enhanced search mechanisms. By providing the statics dependent on various circumstances, we can highlight the output report of all such enhanced search mechanisms. Keyword Guessing Attack is going to happen in this program, the hackers can easily guess the keyword because they can easily hack our content from the cloud server. The current search method can only provide the result based on the matching framework for Boolean keywords, meaning that the environment would consider the exact file name exactly the same as the keyword that the file would be retrieved from the server, it will not have any search results for misspelled keywords. And the new search system never produces results based on related keywords. Using fuzzy search algorithms to upload the file we can create lots of keywords for each file hereby providing the statics based on a different condition, we can highlight the output report of all these enhanced search mechanisms. By providing the statics dependent on various circumstances, we can highlight the output report of all such enhanced search mechanisms.

3. LITERATURE SURVEY

a. Title: Practical technological searches on encrypted data
Author: D. X. Song, D. Wagner, and A. Perrig
Year: 2000
Description: To that the protection and privacy risks, it is beneficial to store data on data storage servers such as mail servers and file servers in encrypted form. Ciphertext;

provide isolation of the search query, meaning that the untrusted server can't know more about the plaintext than the search result; provide controlled search so that the untrusted server can learn something more about the plaintext. The algorithms we present are simple, fast (the encryption and search algorithms only require O(n) stream cipher and block cipher operations for a document of length n), and add almost no overhead space and communication and are therefore practical to use nowadays. Yet this typically means one has to compromise protection features. For example, if a client needs only documents containing those terms to be retrieved, it was not previously understood how to allow the data storage server to conduct the search and address the question without sacrificing data confidentiality. Within this paper, we define our cryptographic schemes for the search problem on encrypted data and provide security proofs for the resulting cryptosystems. There are a number of important they are known to be secure: they provide known confidentiality for encryption, in the sense that the untrusted server cannot learn more about the plaintext when only provided.

b. Title: privacy synonym based fuzzy multi keyword graded search over encrypted cloud
Author: B.Wang, S.Yu, W. Lou, and Y. T. Hou
Year: 2014

Description: Cloud Storage is one of the most commonly used cloud technologies at this time. As cloud use increases, sensitive and personal data are also outsourced which makes it necessary to protect the confidentiality and integrity of this data. A simple way to secure data is to encrypt it before outsourcing, but the recovery of the necessary files from the encrypted cloud becomes a problem involving searching through the encrypted data. Different schemes have been proposed to fix this issue of searching over encrypted cloud data, but none of the current schemes offer ideal user search experience that Plaintext search. In this paper, we suggest privacy search meaning Fuzzy Multi-Keyword Ranked Search over Encrypted Cloud Data, a plan of action that enhances the user search experience to the greatest level by offering multi-keyword searches based on both fuzzy and synonymous, thus bringing encrypted search experience closer to free search engines for text. In addition, by using a binary tree-based dynamic index, the scheme increases index generation time and search the time relative to existing schemes. Experimental findings reflect the effectiveness of this proposed scheme as it decreases the search time, i.e. the time to locate the required data, by 90 percent and minimizes the overhead of database updates when new files need to be uploaded (database generation time)Compared with current effective indexing schemes for similar datasets in the literature. To this point, search time optimization along with index generation time was not done before.
c.Title: Similarity over outsourced cloud data achieving used and privacy secured
Author: C.Wang, K. Ren, S. Yu, and K. M. R. Urs
Year: 2012

Description: As the data generated by individuals and companies needing to be processed and used is growing rapidly, data owners are encouraged to outsource their complex local data management systems to the cloud for their tremendous versatility and economical savings. Nevertheless, because sensitive cloud data that need to be encrypted before outsourcing, which obsolescence the conventional data management service based on plain text keyword search, how privacy-assured usage mechanisms for outsourced cloud data are therefore of paramount importance. Considering the vast number of on-demand data users and the massive volume of outsourced data files in the cloud, the issue is especially daunting because it is incredibly difficult to satisfy the realistic performance criteria, device accessibility, and user search experiences at the highest level. In this paper, we discuss the question of searching for stable and efficient similarity over outsourced cloud data. Similarity search is a fundamental and powerful method commonly used in the retrieval of plaintext content, but not quite explored in the encrypted data domain. First, our mechanism architecture takes advantage of a suppressing technique to create storage-efficient keyword similarity from a given document set, editing distance as the similarity metric. We then create a private tire-traverse search index based on that, and display it correctly achieves the given search similarity functionality with constant search time complexity. Under stringent security treatment, we formally prove the privacy-preserving assurance of the proposed system. To show the generality of our system and further, expand the scope of applications, we also demonstrate that our new construction naturally supports fuzzy search, a notion previously studied that aims only to tolerate types and representation inconsistencies in the user search data. The comprehensive experiments with real data set on Amazon's cloud platform further show the validity and practicality of the proposed mechanism.

d. Title: Practical and efficient key word based search over cloud data
Author: C. Orencik, M. Kantarcioglu, and E. Savas
Year: 2013

Description: Cloud computing solutions are becoming increasingly popular each year, as many companies tend to outsource their data using reliable and efficient cloud infrastructure, thus reducing hardware ownership costs. Even though the benefits are welcomed, privacy remains a problem awaiting addressing. We are implementing an effective privacy conserving search method whereby uses min hash functions over encrypted cloud data. Most literature work can support only one feature search in queries which reduces performance. One of the main benefits of our proposed approach is multi-keyword search functionality in a single query. The proposed method is proven to fulfill the concept of adaptive mean time

protection. We also incorporate an efficient ranking capability that is based on the values of keyword document pairs in the term frequency-inverse document frequency (TF-IDF). Our research shows that the proposed scheme is known to be reliable, efficient and successful in protecting privacy.

e.Title: Idea conceptual resemblance the methodology focused on generative grammar and concepts characteristics constitution
Author: M. Li, B. Lang, and J. Wang
Year: 2015

Description: In information retrieval, knowledge acquisition and many other fields, calculation of memantine similarity between words are important. The latest studies aim primarily at single definitions consisting of single words. They typically ignore the unique constitutional characteristics of compounds for the compound concepts composed of multiple terms and only process them as single concepts, which can affect the ultimate accuracy. In this paper, we propose a novel ontology-based method for the measurement of the Compound Concept Semantic Similarity called CCSS that exploits the features of the concept constitution. The composite is decomposed into Subject headings and Auxiliary Words (SAA) in CCSS and the connections between such two sets are used to calculate the similarity. Additionally, the defects that can arise from SAA recognition are corrected. In addition, many ontological sources of knowledge such as taxonomic characteristics, regional density, and depth are regarded. Extensive experimental evaluations show that our approach greatly outperforms current approaches.

4. EXISTING SYSTEM

Keyword-based information retrieval, commonly used for searching from cloud storage on plaintext data. The data encryption is a standard way of raising information leakage. However, this will make it a very difficult job to use data on the server-side, such as looking for encrypted data. Researchers have suggested several cipher text search schemes in recent years, by integrating the techniques of cryptography. Such methods require large operations and are very time-complex. Keyword Guessing Attack is going to happen in this program, the hackers can easily guess the keyword because they can easily hack our content from the cloud server. The current search method can only have the result based on the Boolean keyword matching scheme, meaning that the actual name of the file would be the same as the keyword that the file would be retrieved from the server, and no search results would be given for misspelled keywords. And the current search system still never delivers the result based on related keywords. Multi keyword search and fuzzy search were introduced separately, and a combination of the two will not result in a stable and efficient multi-keyword fuzzy search scheme. Fuzzy multi-keyword search over encrypted user data protection problem.

5. PROPOSED SYSTEM

Qualified search scheme for cloud storage information by multi-keyword. Here we use the nebulous keyword collection to build the keywords that are all feasible for misspelling. Search keyword get encrypted and test the file name in the cloud server with the original encrypted collection if the keyword matches then we link the nebulous keyword set for that particular keyword and search the file list based on the nebulous keywords, recover the files from the cloud server and find the search output here as well. Extensive preliminary findings on real-life data sets demonstrate that our proposed solution can greatly boost the ability to protect privacy infringements, the scalability and the time-efficiency of query processing over state-of - the-art methods. The keyword control frequencies are taken into consideration when the program produces a graded return results list. The data owner should monitor the ability to unlink Query rates without losing accuracy and better-protecting data privacy.

6. MODULES

a. Login/New User: Within this section, there is a lot of protection to the login construction itself. Typically, the name of the user account and the correct password of that account are sufficient to explain and login, but here are some more actions to do more.
b. Upload File: We want to load the input data in this module then read the input data file and apply the pre-processing to that input file. And the attached file can be transferred into the next steps.
c. Frequent search: This module, we get the non-stop words as input and measure the number of words and find repeated occurrence from non-stop words of each and every word.
d. Similarity search: From the maximum frequency word we find the weight age of each and every word to measure the similarity between the terms based on the similarity that we will divide the terms into clusters.
e. Linear search: In this module we will create keyword search, each cluster has n number of similar terms as keywords we will find the file for that cluster using the lexical analysis tool.
f. Mail alert process: The user's upload and update process must first get the secret key in the relevant user email I d and then apply the secret key to encrypted data to send the server storage and decrypt it using the secret key to access the corresponding data file to the private key conversion of the server storage solution using the Sharing KeyGen (SKA, t, m)..
g. File Downloading process: The method of uploading files is to get the corresponding secret key to the corresponding file to the user mail I d and decrypt the file data afterward. The decryption key for the file transfer

process to storage servers, so that storage servers perform the decryption procedure. And download the file.

7. ARCHITECTURE DIAGRAM

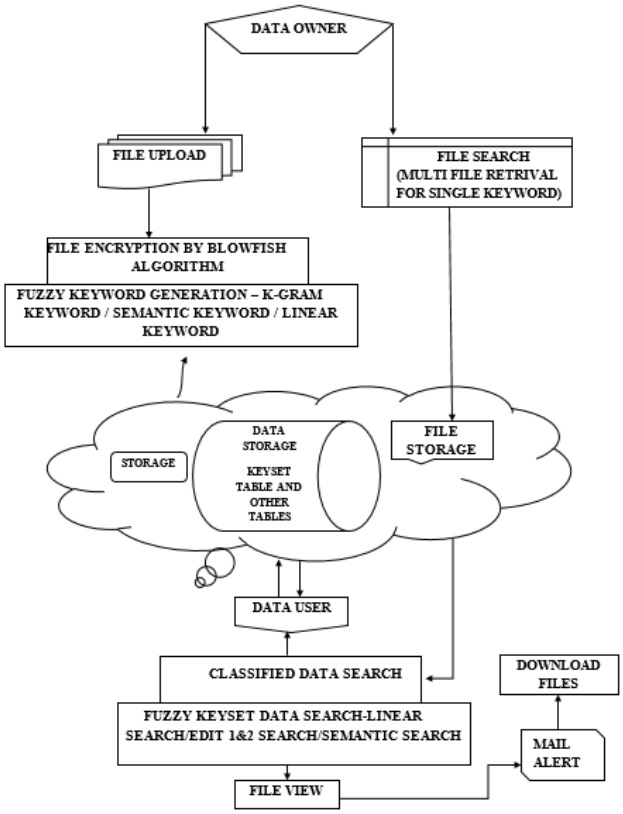

Fig-1: Architecture diagram

8. OUTPUT SCREENSHOTS

Fig-2: Main page

Fig-3: Admin login

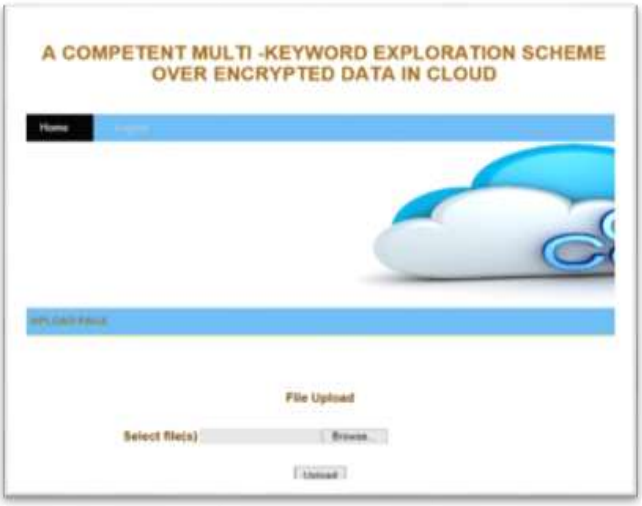

Fig-4: File upload

Fig-5: Search page

Fig-6: Download file

9. FUTURE ENHANCEMENT

It maintains the confidentiality and security of the files stored in the cloud and this keyword-based search helps preserve the above. This can be achieved in a different way to incorporate more functionality at every part of the device to improve the system's security and health.

10. CONCLUSION

A protected multi-keyword search scheme ranked above-encrypted cloud data which simultaneously facilitates dynamic update operations such as removal and document insertion. The cloud server runs through different paths on the index and the data user expects different results but with the same high degree of query accuracy whilst. The keyword-based search is also one commonly used data operator in many applications for database and information retrieval, and its conventional processing methods cannot be applied directly to encrypted data. So how to process these requests over encrypted data while maintaining data privacy at the same time. Then we build the multi-keyword top-k search scheme to boost the search performance, which divides the dictionary into multiple groups and only needs to be stored In the sense that you don't need to send the exact filename to download the file, if you send the maximum number of repeated terms, that time will be downloaded in the decrypted format as well as the original file. This helps keep the files in the cloud secure.

REFERENCES

[1] D. X. Song, D. Wagner, and A. Perrig, “Practical techniques for searches on encrypted data,” in IEEE Symposium on Security and Privacy, 2000, pp. 44–55.

[2] B. Wang, S. Yu, W. Lou, and Y. T. Hou, “Privacy-preserving multi-keyword fuzzy search over encrypted data in the cloud,” in IEEE International Conference on Computer Communications, 2014, pp. 2112–2120.

[3] M. Kuzu, M. S. Islam, and M. Kantarcioglu, “Efficient similarity search over encrypted data,” in IEEE 28th International Conference on Data Engineering (ICDE), 2012, pp. 1156–1167.

[4] C. Wang, K. Ren, S. Yu, and K. M. R. Urs, “Achieving usable and privacy-assured similarity search over outsourced cloud data,” in 2012 Proceedings of IEEE INFOCOM, 2012, pp. 451–459.

[5] C. Orencik, M. Kantarcioglu, and E. Savas, “A practical and secure multi-keyword search method over encrypted cloud data,” in IEEE Sixth International Conference on Cloud Computing (CLOUD), 2013, pp. 390–397.

[6] M. Li, B. Lang, and J. Wang, “Compound concept semantic similarity calculation based on ontology and concept constitution features,” in Tools with Artificial Intelligence (ICTAI), 2015 IEEE 27th International Conference on, 2015, pp. 226–233.

[7] Z. Zhou, Y. Wang, and J. Gu, “A new model of information content for semantic similarity in wordnet,” in Future Generation Communication and Networking Symposia, 2008. FGCNS’08. Second International Conference on, vol. 3. IEEE, 2008, pp. 85–89.

[8] H. Hu, J. Xu, C. Ren, and B. Choi, “Processing private queries over untrusted data cloud through privacy homomorphism,” in 2011 IEEE 27th International Conference on Data Engineering, 2011, pp. 601–612.

[9] Z. Fu, X. Sun, N. Linge, and L. Zhou, “Achieving Effective Cloud Search Services: Multi-keyword Ranked Search over Encrypted Cloud Data Supporting Synonym Query,” IEEE Transaction Consumer Electronics, vol. 60, no.1, pp. 164-172, 2014.

[10] Q. Chai and G. Gong, “Verifiable Symmetric Searchable Encryption for Semi- Honest-but-Curious Cloud Servers,” IEEE International Conference on Communications (ICC’12), pp. 917-922, 2012.