International Journal of Research in Engineering, Science and Management
Volume-3, Issue-2, February-2020
www.ijresm.com | ISSN (Online): 2581-5792

Kgram based Composite Secret Sign Search Over Encrypted Cloud Information

V. M. Sangeetha Priya[1], R. Sangeetha[2], K. Reshma[3], N. Suganthi[4]

[1,2,3]*Student, Department of Information Technology, Jeppiaar SRR Engineering College, Padur, India*

[4]*Assistant Professor, Department of Information Technology, Jeppiaar SRR Engineering College, Padur, India*

***Abstract*: A Large number of data owners have moved our data into cloud servers and prefer to outsource documents in an encrypted form for the function of confidentiality preserving. Therefore, it is essential to develop reliable cipher text search method. The secured multi-keyword ranked search from the encrypted data from the cloud, top-k search problem for big data encryption against privacy breaches is a major problem. By using tree structure and nebulous search method for retrieving data from the cloud and used to solve the problem of keyword guessing attack. The blowfish algorithm is used for the encryption and decryption process. We propose a group of multi-keyword top-k search based on the idea of partition, where a group of indexes are constructed for all documents. The experimental results on real-life data sets demonstrate that this approach can significantly improve the capability of defending the privacy breaches, scalability and time efficiency of query processing and also deal with deletion and insertion of documents flexibly.**

***Keywords*: Multi keyword search, Semantic, Top-k search, Security, Encryption, Secret key.**

1. Introduction

Cloud computing infrastructure is a promising new technology and greatly accelerates the development of large scale data storage and distribution. It depicts on sharing of resources to achieve coherence of scale. However, security and privacy become major concerns when data owners outsource their private data into the public cloud servers that are not within their trusted management domains. To avoid information leakage, sensitive data have to be encrypted before uploading onto the cloud servers, which is major challenge to support keyword-based queries and rank the correct matching results on the encrypted data. To overcome the security drawbacks the proposed system involves the combination of multikeyword search and fuzzy search with encrypting the data. The existing system has lot of security issues such as Keyword Guessing Attack. In this hacker can easily identify the keyword and they can easily hack our content from cloud server. Existing search system will provide the result only based on the Boolean keyword matching system, it means whether it will find the exactly file.

Keyword guessing attack is a major problem in retrieve the data from the cloud. The proposed system can generate lot of keywords for each file using fuzzy search algorithms on uploading the file. While searching files it can retrieve the maximum additional files by matching the corresponding generated fuzzy keywords with the file name of all files which is available in the cloud server. The proposed system combine these methods together into an efficient and secure approach to reduce statistical attacks.

2. Related work

Semantic-Aware Searching Over Encrypted Data for Cloud Computing, Xingmin, et, al., (2018), In recent days with the increasing adoption of cloud computing, a growing number of users outsource their datasets to cloud. To preserve the privacy, the datasets are usually encrypted before outsourcing. However, the common practice of encryption makes the effective utilization of the data difficult. For example, it is difficult to search the given keywords in encrypted datasets. Many schemes are proposed to make encrypted data searchable based on keywords. However, keyword-based search schemes ignore the semantic representation information of users retrieval, and cannot completely meet with users search intention. Therefore, how to design a content-based search scheme and make semantic search more effective and context-aware is a difficult challenge. This paper propose ECSED, a novel semantic search scheme based on the concept hierarchy and the semantic relationship between concepts in the encrypted datasets. ECSED uses two cloud servers. One is used to store the outsourced datasets and return the ranked results to data users. The other one is used to compute the similarity scores between the documents and the query and send the scores to the first server. To further improve the search efficiency, it utilize a tree-based index structure to organize all the document index vectors. It employs the multi keyword ranked search over encrypted cloud data as our basic frame to propose two secure schemes. The experiment results based on the real world datasets show that the scheme is more efficient than previous schemes. In this paper, to address the problem of semantic retrieval, here proposed effective schemes based on concept hierarchy. To improve accuracy, the process extend the concept hierarchy to expand the search conditions. In addition, a tree-based index structure is constructed to organize all the document index vectors, which are built based on the concept hierarchy for the aspect of search efficiency.

A Practical And Secure Multikeyword Search Method Over Encrypted Cloud Data, C. Orencik, et al., (2013), In this paper, Cloud computing technologies become more and more popular every year, as many organizations tend to outsource their data utilizing robust and fast services of clouds while lowering the cost of hardware ownership. Although its benefits are welcomed, privacy is still a remaining concern that needs to be addressed. We propose an efficient privacy-preserving search method over encrypted cloud data that utilizes hash functions. Most of the work in literature can only support a single feature search in queries which reduces the effectiveness. One of the main advantages of our proposed method is the capability of multikeyword search in a single query. The proposed method is proved to satisfy adaptive semantic security definition. We also combine an effective ranking capability that is based on term frequency inverse document frequency values of keyword document pairs. Our analysis demonstrates that the proposed scheme is proved to be privacy-preserving, efficient and effective.

Component Concept Semantic Similarity Calculation Based On Ontology and Concept Constitution Features, J. Wang, et al., (2015), The computation of semantic similarity between words is important in information retrieval, knowledge acquisition and many other fields. The existing studies are mainly aiming at single concepts composed of single terms. For the compound concepts composed of multiple terms, they usually neglect the special constitution features of compounds and only process them as single concepts, which may affect the ultimate accuracy. In this paper, we propose a novel ontology-based Compound Concept Semantic Similarity calculation approach called CCSS which exploits concept constitution features. In CCSS, the compound is decomposed into Subject headings and Auxiliary words (SaA), and the relationships between these two sets are used to measure the similarity. Besides, the errors that may be caused by SaA recognition are corrected. Moreover, several information sources of ontology such as taxonomical features, local density and depth are considered. Extensive experimental evaluations demonstrate that our approach significantly outperforms existing approaches. Processing Private Queries Over Entrusted Data Cloud Through Privacy Homomorphism, B. Choi, et. al., (2014), Query processing that preserves both the data privacy of the owner and the query privacy of the client is a new research problem. It shows increasing importance as cloud computing drives more businesses to outsource their data and querying services. However, most existing studies, including those on data outsourcing, address the data privacy and query privacy separately and cannot be applied to this problem. In this paper, we propose a holistic and efficient solution that comprises a secure traversal framework and an encryption scheme based on privacy homomorphism. The framework is scalable to large datasets by leveraging an index-based approach. Based on this framework, we devise secure protocols for processing typical queries such as k-nearest-neighbor queries (kNN) on R-tree index. Moreover, several optimization techniques are presented to improve the efficiency of the query processing protocols. Our solution is verified by both theoretical analysis and performance study.

Semantic Sequential Query Expansion for Biomedical Article Search, Fang, et al., (2018), The conventional sequential dependence model (SDM)has been proved to perform better than the bag of words model for biomedical article search because it pays attention to the sequence information within queries. However, a few researches have been conducted on combining semantic and sequence information together. Hence, this project proposes the semantic sequential dependence model in this paper, which provides an innovative combination of semantic information and the conventional SDM. Specifically, the synonyms are obtained automatically through the word embeddings which are trained on the domain-specific corpus by selecting an appropriate language model. Then, these synonyms are utilized to generate possible sequences with the same semantics as the original query and these sequences are fed into SDM to obtain the final retrieval results. The proposed approach is evaluated on 2016 and 2017 BioASQ benchmark test sets and the experimental results show that our query expansion approach outperforms the baseline and other participants in the BioASQ competitions. The original query can be re-organized into several new queries by replacing one or more query keywords with their synonyms. Finally, these queries are processed by SDM and used for ranking the final search results. The project uses five comparable neural-network-based language models such as NNLM, LBL, WORD2VEC, ORDER to learn the vector representation of word semantics. In this paper, we propose a new query expansion model SSDM that combines the semantic information of words and the conventional SDM for retrieving biomedical documents. In addition, the word embeddings used to generate the synonyms of query keywords are trained by a specific language model on a domain-specific corpus. For many practical tasks, the word embeddings generated by the Word2vec tool have a sufficiently good effect, which helps us save a lot of time. By comparing to the approaches of other participating teams, our experimental results on different batches of 2016 and 2017 BioASQ competitions demonstrate great effectiveness and robustness of the proposed SSDM approach.

Lightweight searchable public-key encryption for cloud assisted wireless sensor networks, Peng Xu, et al., (2018), The industrial Internet of Things is flourishing, which is unprecedentedly driven by the rapid development of wireless sensor networks (WSNs) with the assistance of cloud computing. The new wave of technology will give rise to new risks to cyber security, particularly the data confidentiality in cloud-assisted WSNs (CWSNs). Searchable public-key encryption (SPE) is a promising method to address this problem. In theory, it allows sensors to upload public-key cipher texts to the cloud, and the owner of these sensors can securely delegate a keyword search to the cloud and retrieve the intended data while maintaining data confidentiality. However, all existing and semantically secure SPE schemes have expensive costs in terms of generating cipher texts and searching keywords. Hence, this paper proposes a lightweight SPE (LSPE) scheme with semantic security for CWSNs. LSPE reduces a large number of the computation-intensive operations that are adopted in previous works; thus, LSPE has search performance close to that of some practical searchable symmetric encryption schemes. In this paper it proposes a lightweight and semantically secure SPE scheme called LSPE for the scenario of CWSNs.

3. Limitations of existing system

- Multi keyword search and fuzzy search have been implemented separately
- Combination of the multi keyword search and fuzzy search does not lead to a secure and efficient multi-keyword fuzzy search scheme
- The multi-keyword fuzzy search over encrypted data problem with user data privacy protection
- It involves high time complexity

4. Proposed system

The competent search scheme to search the documents from the cloud server using multi-keyword. Here we using the nebulous keyword set it will create the all feasible misspell keywords. Search keyword get encrypt and it will check with the collection of original encrypted the file name in the cloud server if the keyword will get matched then we connect the nebulous keyword set for that particular keyword and it search the file list based on that nebulous keywords, it will retrieve the files from the cloud server and here we consider the searching performance also. Extensive experimental results on real-life data sets demonstrate that our proposed approach can significantly improve the capability of defending the privacy breaches, the scalability and the time efficiency of query processing over the state-of-the-art methods.

5. System design

The proposed architectural model of this project is shown in the figure 1.

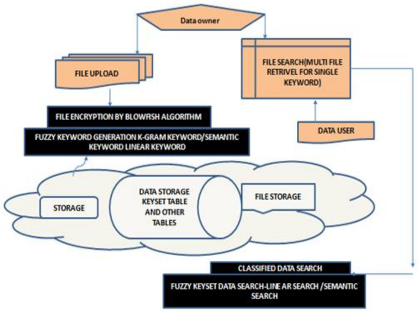

Data owner
FILE UPLOAD
FILE SEARCH(MULTI FILE RETRIVEL FOR SINGLE KEYWORD)
FILE ENCRYPTION BY BLOWFISH ALGORITHM
FUZZY KEYWORD GENERATION K-GRAM KEYWORD/SEMANTIC KEYWORD LINEAR KEYWORD
DATA USER
STORAGE
DATA STORAGE KEYSET TABLE AND OTHER TABLES
FILE STORAGE
CLASSIFIED DATA SEARCH
FUZZY KEYSET DATA SEARCH-LINE AR SEARCH /SEMANTIC SEARCH

Fig 1. Overall architecture

The proposal consists of the following proposals:

A. Authentication
B. File uploading
C. Keyset generation
D. Searching process
E. Alert message
F. File download

A. *Authentication*

Authentication process involves the login credentials of both admin and user It is done so inorder to give access only to the valid user. This process includes collecting the user details such as providing mail id and password for the authentication credentials. In case of a new user, the first process involves registration followed by logging in to the user profile. Every time the user has to login to the profile and must log out after completing the process.

B. *File uploading process*

Transferring data from one remote system to another under the control of a local system is remote uploading. Remote uploading is used by some online services. It is also used when the local computer has a slow connection to the remote systems, but they have a fast connection between them. Without remote uploading functionality, the data would have to first be downloaded to the local host and then uploaded to the remote file hosting server, both times over slow connections. In this module, we want to load the input document then read the input document file and want to implement the pre-processing to that input file. So that the file attached can be processed to the next phases.

C. *Keyset generation*

Keysets are generated for the file based on permutation combination of the words of the file name. The keyset is splitted in kgram1 and kgram2 by splitting it into the combination of four and three letters respectively. This process involves blowfish encryption algorithm for encrypting the file uploaded by the admin.

D. *Search process*

The Searching process includes combinations of multikeyword search and fuzzy search. It involves linear, semantic, kgram1 and kgram2 type of searches. If the exact file name is known then the user can opt for linear search for fetching and downloading the file from the cloud. If the user knows the meaning of the words then semantic search will be a suitable choice. Semantic search involves meaningful synonyms for the keywords specified. Incase of misspelled keywords then the choice of search would be kgram1 and kgram2.The first involves combination of four words and the latter includes combinations of three words. The combinations of words are based on permutation functions. Once the file is uploaded by the admin it is then encrypted and then decrypted by the user by applying the secret key.

E. Alert notification

Alert notification is machine-to-person communication that is important or time sensitive. An alert may be a calendar reminder or a notification of a new message. Alerts are typically delivered through a notification and the most common application of the service is machine-to-person communication. Very basic services provide notification services via email or SMS. The uploading and downloading process of the user is first get the secret key in the corresponding user email id and then apply the secret key to encrypted data to send the server storage and decrypts it by using his secret key to download the corresponding data file in the server storage system's the secret key conversion using the Share Key Generation.

F. File download

Downloading generally transfers entire files for local storage. File downloading process is to get the corresponding secret key to the corresponding file to the user mail id and then decrypt the file data. The file downloading process decryption key to storage servers such that storage servers perform the decryption Operation and the file is downloaded.

6. Data flow diagram

It is a simple graphical formalism that can be used to represent a system in terms of the input data to the system, various processing carried out on these data, and the output data is generated by the system. The data flow diagram (DFD) is one of the most important modeling tools. It is used to model the system components. These components are the system process, the data used by the process, an external entity that interacts with the system and the information flows in the system. DFD shows how the information moves through the system and how it is modified by a series of transformations. It is a graphical technique that depicts information flow and the transformations that are applied as data moves from input to output. DFD is also known as bubble chart. A DFD may be used to represent a system at any level of abstraction. DFD may be partitioned into levels that represent increasing information flow and functional detail.

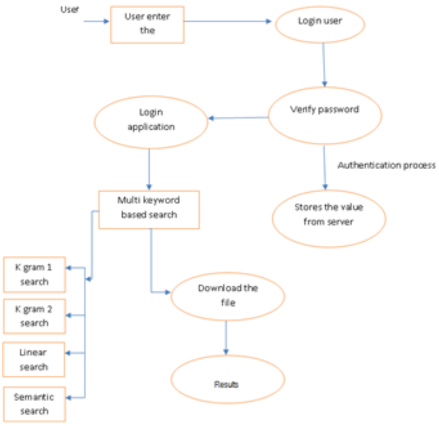


Fig. 2. Data flow diagram

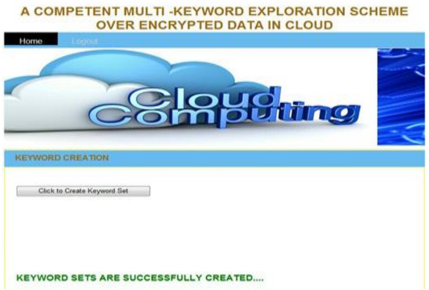
7. Algorithm or methodology

A. Blow fish algorithm

Blowfish is a symmetric block cipher, designed by Bruce Schneider and included in many cipher suites and encryption products. Blowfish provides a good encryption rate in software. Schneider designed Blowfish as a general-purpose algorithm, intended as an alternative to the aging DES and free of the problems and constraints associated with other algorithms Notable features of the design include key-dependent S-boxes and a highly complex key schedule. Blowfish has a 64-bit block size and a variable key length from 32 bits up to 448 bits. It is a 16-round Feistel cipher and uses large key-dependent S-boxes. In structure it resembles CAST-128 which uses fixed S-boxes. Blowfish is a fast block cipher except when changing keys. Each new key requires pre-processing equivalent to encrypting about 4 kilobytes of text, which is very slow compared to other block ciphers. This prevents its use in certain applications, but is not a problem in others. Blowfish has a memory footprint of just over 4 kilobytes of RAM. This constraint is not a problem even for older desktop and laptop computers though it does prevent use in the smallest embedded systems such as early smartcards. A reduced-round variant of Blowfish is known to be susceptible to known-plaintext attacks on reflectively weak keys. Blowfish implementations use 16 rounds of encryption, and are not susceptible to this attack.

8. Results and discussion

A secure multi-keyword ranked search scheme over encrypted cloud data, which simultaneously supports dynamic update operations like deletion and insertion of documents. The cloud server traverses different paths on the index, and the data user receives different results but with the same high level of query accuracies in the meantime. The keyword-based search is such one widely used data operator in many database and information retrieval applications, and its traditional processing methods cannot be directly applied to encrypted data. Therefore, how to process such queries over encrypted data and at the same time guarantee data privacy. Then, in order to improve the search efficiency, we design the group multi-keyword top-k search scheme, which divides the dictionary into multiple groups and only needs to store In the sense no need to give exact filename to download the file, if you are going to give maximum number of time repeated words, that time also original file will be downloaded in decrypted format. This helps to maintain the security of the files in the cloud. Thus the secret and sensitivity of the files stored in the cloud is secured and this keyword based search helps in maintaining the above. This can be made in a better way that more of the features can be added at any part of the system to increase the security and safety of the system. The snapshot describes the keyset generation after uploading the file by the admin.

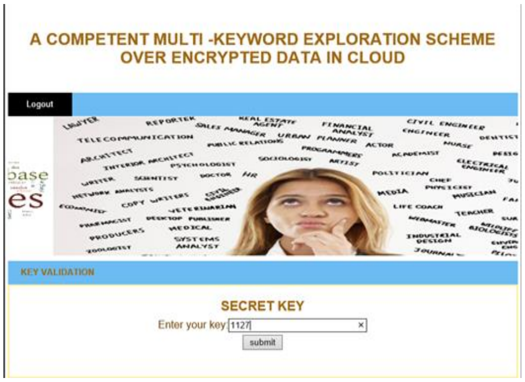

A COMPETENT MULTI -KEYWORD EXPLORATION SCHEME OVER ENCRYPTED DATA IN CLOUD

KEYWORD CREATION

KEYWORD SETS ARE SUCCESSFULLY CREATED....

Fig. 3. Keyset generation

The secret key authentication is made to make the document more confidential to find the authenticate user.


A COMPETENT MULTI -KEYWORD EXPLORATION SCHEME OVER ENCRYPTED DATA IN CLOUD

KEY VALIDATION

SECRET KEY

Enter your key 1127

Fig. 4. Secret key access

The file can be finally downloaded after entering the secret key.

9. Conclusion

Thus the proposed system helps to maintain the security of the files in the cloud. Thus the secret and sensitivity of the files stored in the cloud is secured and this keyword based search helps in maintaining the above. This can be made in a better way that more of the features can be added at any part of the system to increase the security and safety of the system.

References

[1] Y. Zhang, Y. Li, and Y. Wang, "Secure and efficient searchable publickey encryption for resource constrained environment based on pairings under prime order group," Seur. Commun. Network. vol. 2019, pp. 5280806-1-5280806-14, Apr.2019.

[2] J. Yao, Y. Zheng, C. Wang, and X. Gui, "Enabling search over encrypted cloud data with concealed search pattern, "IEEE Access, vol. 6, pp. 11112-11122, 2018.

[3] S. Jiang et. al., "Publicly verifiable Boolean query over outsourced encrypted data," IEEE Trans. Cloud Comput., vol. 10, no. 5, pp .356-362, Aug. 2017.

[4] H. Cui, Z. Wan, R. Debg, G, Wang, and Y. Li, "Efficient and expressive Keyword search over encrypted data in cloud," IEEE Trans. Depend. Sec. Comput., Vol. 15, no. 3, 409-422, June 2016.

[5] Qzheng, S. Xu, and G. Ateniese, "Vabks: Verfiable attribute-based Keyword search over outsourced encrypted data," in Proc. INFOCOM, July 2015, pp. 522-530.

[6] B. Wang, Y. Hou, M. Li, H. Wang, H. Li, and F. Li, "Tree-Based Multi-Dimensional Range Search Encrypted Data with Enhanced Privacy," Beijing, China: Springer, 2014.

[7] Lai, X. Zhou, R. H. Deng, Y. Li, and K. Chen, "Expressive search on encypted data," in proc, 8th ACM SiNSAC Symp. Inf., Comput. Commun. Secur., May 2013, pp. 243-252.

[8] 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, Jan.2013.

[9] G. Pirro, "A semantic similarity metric combining features and intrinsic information content," IEEE Transactions on Data & Knowledge Engineering, vol. 68, no.11, pp. 1289-1308, 2013.