ISSN: 2321-7782 (Online)
e-ISJN: A4372-3114
Impact Factor: 6.047

Volume 5, Issue 6, June 2017

International Journal of Advance Research in Computer Science and Management Studies

Research Article / Survey Paper / Case Study

Available online at: www.ijarcsms.com

Secure Multi-keyword Search using OPE over Cloud

Harshali A. Agutale[1]
M.E. (Computer) Pursuing
Savitribai Phule, Pune University
RMD Sinhgad School of Engineering,
Warje, Pune, Maharashtra – India

Prof. Kanchan M. Varpe[2]
Assistant Professor
Department of Computer Engineering
RMD Sinhgad School of Engineering,
Warje, Pune, Maharashtra – India

Abstract: ***Cloud Computing has become a great platform today where the cloud users can store the sensitive files remotely so as to enjoy on-demand high quality applications, to avoid data loss that occur due to server failure/damage and to solve the problem of insufficiency of the storage capacity. Sensitive files are encrypted before outsourcing to the cloud using various encryption techniques. This makes the searching of files through the cipher text more complicated. Order Preserving Encryption (OPE) technique has been used for the ranked search over the encrypted cloud data that encrypts the relevance score of the inverted index. OPE scheme comes with various securities loop holes such as the attacker can easily reveal the distribution of plaintext over cipher text and differential attack. To address this issue the proposed system develops a more secure searchable encryption scheme for multi-keyword ranked search over cloud data. The system maintains the confidentiality of files and their keywords where the cloud server will not be able to penetrate into the sensitive cloud data and the cloud user will be obtain with the fast ranked search. As compared to the existing system the proposed system has improved security, ranked search and the execution time.***

Keywords: Cloud Computing, Ranked search and Relevance Score, Searchable Encryption, Order Preserving Encryption, Differential attack.

I. INTRODUCTION

Outsourcing data to cloud requires a trustworthy cloud service provider and a system which eases the cloud user to upload files and download the required data/files. Traditionally downloading the required files will be the tedious task as the cloud user need to download all the files present over the server, decrypt each file then search into each file the required that data and then decide to whether this file is relevant or not. This problem was the reason as many searching technique were emerged down the line. Searchable encryption gets evolved from single keyword to multi keyword search, fuzzy search pattern, rank search, relevant search and so on. Rank search is also one of the important factors that have to be addressed, as the result of the search should always according to the relevancy. The key factor of the search over encrypted data is to protect the search query i.e. its constants, search query outcome and the retrieved data. For obtaining rank search on cloud, OPE is an efficient technique to encrypt relevance scores of the inverted index.

II. RELATED WORK

In order to search sensitive encrypted data/files over the cloud user need to download all files, encrypt it and then search the data with the keyword. To overcome this searchable encryption was proposed which allows querying in an encrypted domain preserving its privacy. Searchable encryption is a technique to search encrypted data over the cloud. There are two types of searchable encryption one is searchable public key encryption abbreviated as SSE and searchable symmetric encryption abbreviated as SPE.

© 2017, IJARCSMS All Rights Reserved

Song et al. [12] first introduced SSE scheme that successfully search encrypted data but it supports only single keyword, multi keyword search is not possible with SSE. The SSE scheme securely gives the result of the search request as an encrypted searchable index is built whose contents are secured from the server. This method only supports the boolean keyword search in which either keyword is present in the file or not but this scheme does not give the result as per the relevancy of the queried keyword with the files.

Wang et al. [4] proposed a searching technique for multi keyword search in a single query. It also involves fuzzy search technique in which even if the user misspelled the queried word this technique will give the result that is closely matched to the keyword.

In [2] the vector space model and the TF IDF model are combined in the index construction (while uploading the file) and query generation (while searching the file). Tree-based index structure is used to store the index and Greedy Depth-first Search algorithm is used to provide multi-keyword ranked search. This scheme reduces the search time as tree based structure is used for search and it also give the flexibility to deal with the insertion and deletion of the documents over the cloud.

Agrawal et al. [13] introduced a technique in 2004 to retrieve match files in the order of the relevance with the help of indexing technique. Order-preserving symmetric encryption is proposed in [12] for permitting effective range queries upon encrypted data. This will enable the quick search of documents that contain a given keyword. OPE supports ranked search thus giving the result according to the relevancy of the documents. OPE solves the encrypted query problem thus reducing the time for the retrieval of the query result. The OPE property states that if plaintext x1 < x2 then the cipher text E(x1) < E(x2). The relevance score and inverted index are secured with the OPE.

Boldyreva et al. [8] proposed a more secured form of OPE. Reddy et al. [12] proposes a scheme called as Randomized Order Preserving Encryption abbreviated as ROPE, a novel OPE scheme that leaks nothing beyond the order. But the OPE scheme was supposed to have deterministic encryption which means that plain text will always encrypted to a fixed cipher text. Deterministic OPE leaks information about plaintext distribution. OPE schemes cannot meet the standard notion of security called indistinguishability against chosen-plaintext attack (IND-CPA), as OPE scheme is not only deterministic, but also leak the order-relations among the plaintexts.

To overcome the disadvantages of OPE scheme One to Many OPE scheme was proposed by Wang et al. [6] were a probabilistic scheme is proposed that will not conceal the distribution of the plaintext. Even though the probabilistic OPE gives the security expected but a differential attack can occur and adversaries can estimate distribution of plaintext from the differences of cipher texts.

III. Proposed Methodology

Information retrieval becomes difficult in the encrypted domain over the cloud because the amount of outsourced files can be very large and traditional search patterns cannot be deployed to cipher text retrieval directly. Apart from this cloud can be curious to know the data because of which data breach is possible. To overcome this problem, a secure multi-keyword ranked based search scheme is proposed to generate a query in an encrypted domain that will not only give the search result according to the relevant order but also maintain the privacy and the security of the system.

IV. Proposed System

One to many OPE algorithm is use in the proposed system to encrypt the keywords. The purpose of both OPE and One-to-Many OPE is to prevent information leakage to the cloud server.

Below diagram shows the difference between deterministic OPE and Probabilistic or One-to-many OPE scheme.

© 2017, IJARCSMS All Rights Reserved

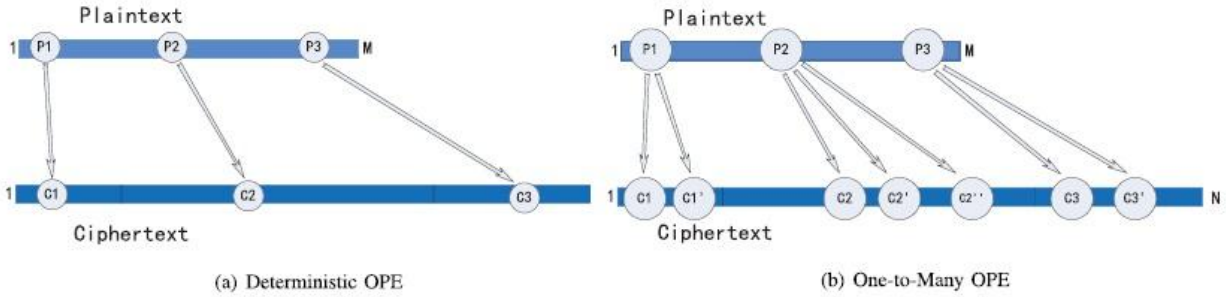


Figure 7.1: Comparison between deterministic and One-to-Many OPE [1]

Below pseudo code represents the One-to-Many OPE Algorithm:

Algorithm 1 Binary Search

1: procedure BINSEARCHOPE(K,D,R,m)

2: M←Length(D); N←Length(R)

3: d←min(D)-1; r←min(R)-1

4: y←r + ceil(N/2)

5: coin←TapeGen(K,(D,R,0||y))

6: x←d + HGD(coin,M,N,y-r)

7: x = d + f

8: if m <= x then

9: D←{d+1,..,x}

10: R←{r+1,..,y}

11: else

12: D←{x+1,..,d+M}

13: R←{y+1,..,r+N} end if

14: return{D,R}

Algorithm 2 Probabilistic OPE

procedure OPM(K,D,R,m,id(F))

While D!=1

do {D,R}= BINSEARCHOPE(K,D,R,m)

end while

coin←TapeGen(K,(D,R,1||m,id(F)))

c←R

c=round(coin)

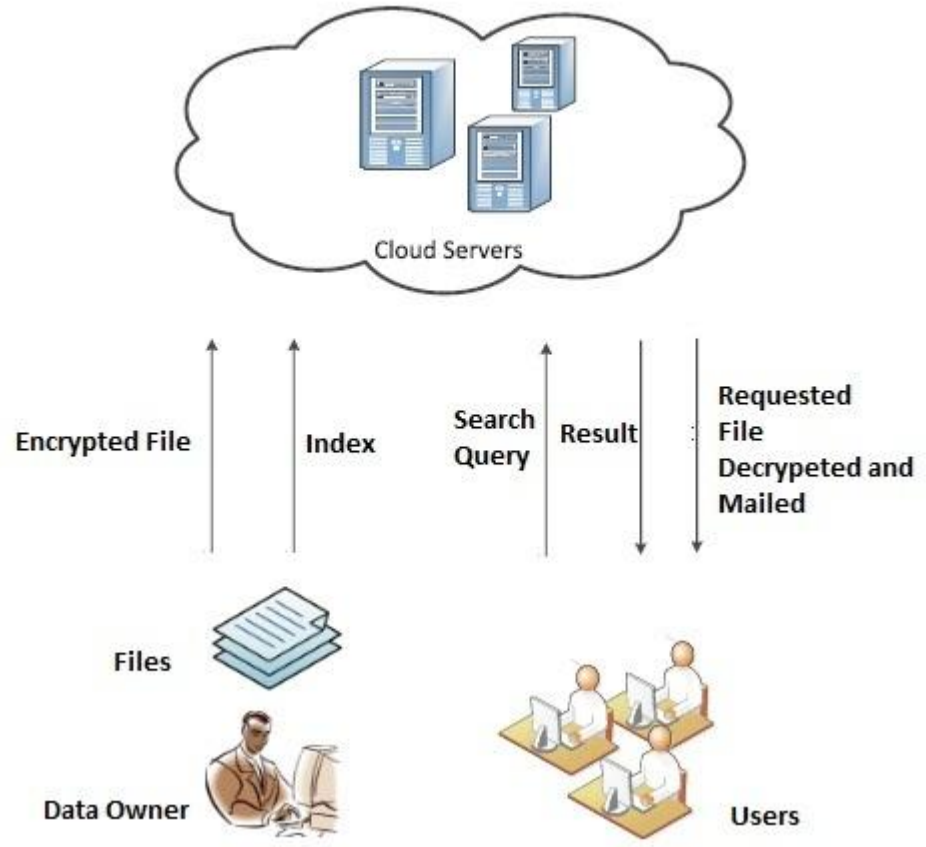
return{c}

The system contains 3 entities as:

1. CSP (Cloud Service Provider): GUI to upload the file and cloud database to store data.

2. *Data* Owner (Registered Cloud User): *Data* Owner will upload the file.

3. Cloud User (Searches the file): Searches files using keywords.



There will be two phases as upload and the download phase. In upload phase the data owner uploads a file. Key generation algorithm is called and random keys are generated as shown below:

P_k = random_gen[set off(character, number), random_user_id]

Uploaded file is encrypted and the file contents are collected as plaintext and stop words are removed. Inverted index are built by using the One to many OPE that uses the generated key in above step. To this inverted index in between noise or dummy key words are added to secure the database from attacks. The trapdoor is generated which will have set of encrypted keywords. The encrypted keywords and file together are encrypted and store on the cloud database as shown below:

$CSP(F) = E'(F) + T_d$

Where T_d is set of encrypted keywords

In the download phase the user that wants to search a file will enter particular keyword or multiple keywords. A SQL query will be fired on cloud database where the keywords will be in encrypted form. The keywords in the query will be matched with the keywords stored in the database. If match found the corresponding files are retrieved and displayed on GUI according to the relevance order. [1] One-to-many OPE is prone to differential attacks. Differentials attacks can be encountered in a way where the cipher texts are arranged in an ascending order. Adding noise between the keywords will give wrong difference between the cipher values thus protecting the system from the attack.

© 2017, IJARCSMS All Rights Reserved

V. RESULT

The Multi-keyword ranked Search system is deployed on MS-Azure cloud. The time to upload the file and search the file is less so the overall execution time is decreased with maintaining the search efficiency.

Below graph shows the comparative study between existing and proposed system.

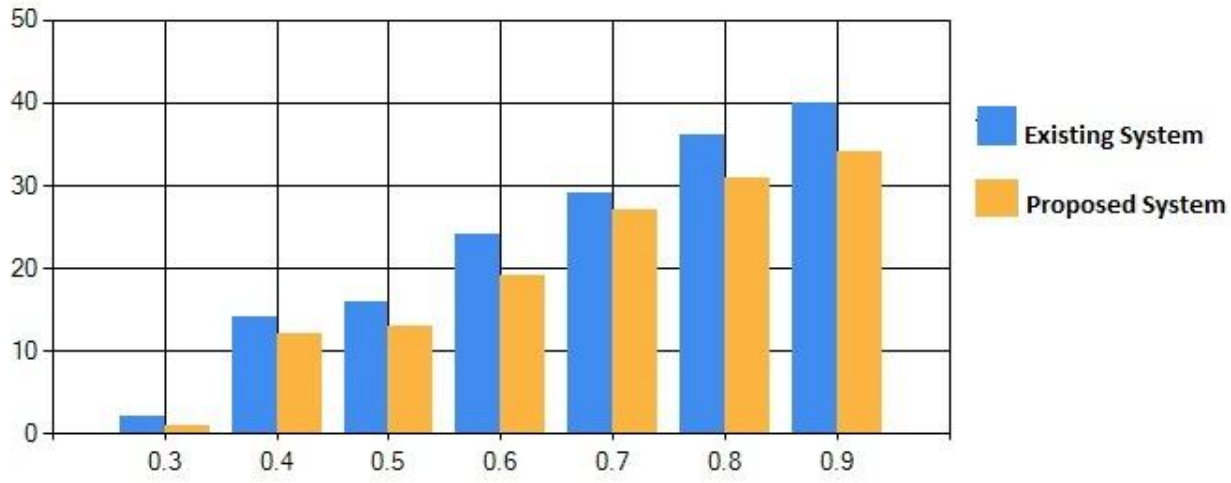
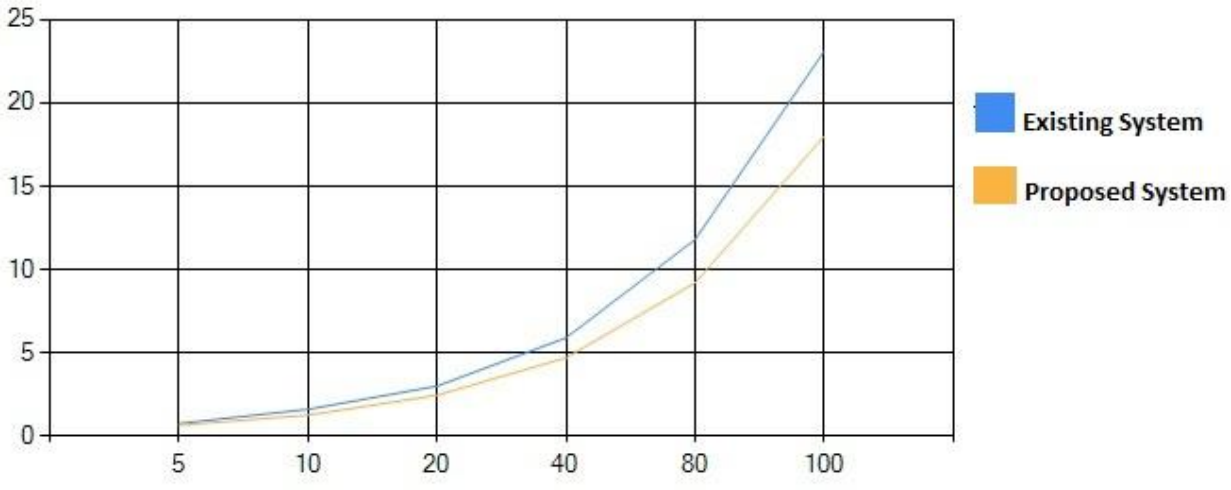


Figure 9.2 shows the comparative study of time taken to search the keyword and number of frequency of the matched keyword in the document. The time taken in the proposed system to search single and multiple keywords is comparatively less than the existing system.



By observing both the graphs, we can conclude that the proposed system gives better results as compared to the existing system in terms of various parameters.

VI. CONCLUSION

The multi-keyword search system gives better ranking of the documents thus providing the user with the search result in the relevant manner of the documents that are stored on the cloud server and for which the user is searching. Apart from this system will also provide security against the differential attack that is discussed in [1] by saving dummy keywords or document ids to the inverted index. As the required files are decrypted and sent to the search user via mail the keys are secure from the end users. In future the one to many OPE can be more improved to save from different attacks by dividing the plain text into several sets and dividing the corresponding bucket into sub-bucket and then mapping each plain text sets to sub-bucket that will cover

© 2017, IJARCSMS All Rights Reserved

up the original distribution of the plain text. Apart from this the keyword search can be more improved by adding fuzzy keyword search where the spelling mistakes are ignored to give efficient search result.

References

1. Ke Li, Weiming Zhang, Ce Yang, and Nenghai Yu "Security Analysis on One-to-Many Order Preserving Encryption-Based Cloud Data Search", In IEEE Transactions on Information Forensics and Security, VOL. 10, NO. 9 SEPTEMBER 2015.
2. "A Secure and Dynamic Multi-keyword Ranked Search Scheme over Encrypted Cloud Data" :Zhihua Xia, Member, IEEE, Xinhui Wang, Xingming Sun, Senior Member, IEEE, and Qian Wang, Member, IEEE, 2015.
3. K. Srinivasa Reddy and S.Ramachandram "A New Randomized Order Preserving Encryption Scheme", In International Journal of Computer Applications (0975 – 8887) Volume 108 – No 12, December 2014.
4. B. Wang, S. Yu, W.Lou, and Y.T.Hou, "Privacy-preserving Multi-keyword Fuzzy Search over Encrypted Data in the Cloud", in Proc. IEEE INFOCOM, Apr./May 2014, pp. 2112-2120.
5. Raluca Ada Popa , Frank H. Li , Nickolai Zeldovich "An Ideal-Security Protocol for Order Preserving Encoding" In Proc. of the 34th IEEE Symposium on Security and Privacy 2013.
6. C. Wang, N. Cao, K. Ren, and W. Lou, "Enabling secure and efficient ranked keyword search over outsourced cloud data", IEEE Trans. Parallel Distrib. Syst., vol. 23, no. 8, pp. 1467–1479, Aug. 2012.
7. L. Xiao and I.-L. Yen, "Security analysis for order preserving encryption schemes", in Proc. 46th Annu. Conf. Inf. Sci. Syst., Mar. 2012, pp. 1–6.
8. Boldyreva, N. Chenette, and A. O'Neill, "Order-preserving encryption revisited: Improved security analysis and alternative solutions", in Advances in Cryptology. Berlin, Germany: Springer-Verlag, 2011, pp. 578–595.
9. R. Fielding,"Architectural styles and the design of network-based software architectures ", Ph.D. thesis, University of California, Irvine , 2000.
10. N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-preserving multi-keyword ranked search over encrypted cloud data ", in Proc. IEEE INFOCOM, Apr. 2011, pp. 829–837.
11. B. Zhang and F. Zhang,"An efficient public key encryption with conjunctive-subset keywords search ", J. Netw. Comput. Appl., vol. 34, no. 1, pp. 262–267, 2011.
12. D. X song, D. Wagner, and A.Perrig, "Practical Techniques for Searches on Encrypted Data ", in proc. IEEE symp. Secur.Privacy, May 2000, pp. 44-55.
13. R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu, "Order preserving encryption for numeric data ", in Proc. ACM SIGMOD Int. Conf. Manage. Data, 2004, pp. 563–574.
14. W.-A. Taylor. (2000). Change-Point Analysis: A Powerful New Tool for Detecting Changes. [Online]. Available: http://www.variation.com/cpa/tech/changepoint.html
15. J. Li, E. R. Omiecinski, Efficiency and security trade-off in supporting range queries on encrypted databases, Data and Applications Security 2005,pp. 69-83.12
16. D. Boneh, G. D. Crescenzo, R. Ostrovsky, and G.Persiano, "Public key encryption with keyword search ",in Proc. of EUROCRYP04,volume 3027 of LNCS. Springer,2004.
17. E.-J. Goh, "Secure indexes", Cryptology ePrint Archive, Report 2003/216.

© 2017, IJARCSMS All Rights Reserved