Retrieval of Encrypted Data Using Multi Keyword Top – K Algorithm

Mrs. P.Shanmuga Priya M.E(Ph.d), Preethi.D, Priya.J, shanthini.B

Abstract- Cloud computing is the computing technology that involves a large number of computers connected through a communication network such as internet, similar to utility computing and distributed computing. It has ability to run a program or application on many connected computers at the same time. Searchable symmetric encryption (SSE) allows retrieval of encrypted data over cloud. Basically focus on addressing data privacy issues using searchable symmetric encryption (SSE). To eliminate the leakage, two-round searchable encryption (TRSE) scheme that supports top-k multi-keyword retrieval along with it propose a log file generation module. The efficiency of the system is enhanced by making retrieval of data in more secured and efficient manner. The vector space model helps to provide sufficient search accuracy, and the homomorphic encryption enables users to involve in the ranking while the majority of computing work is done on the server side by operations only on ciphertext. As a result, information leakage can be eliminated and data security is ensured. Thorough security and performance analysis show that the proposed scheme guarantees high security and practical efficiency.

I. INTRODUCTION

Cloud computing has emerging as a promising pattern for data outsourcing and high quality data services. Searchable symmetric encryption (SSE) allows retrieval of encrypted data over cloud.The concepts of similarity relevance and scheme robustness to formulate the privacy issue in searchable encryption schemes, and then solve the insecurity problem by proposing a two-round searchable encryption (TRSE) scheme. Novel technologies in the cryptography community and information retrieval community are employed, including homomorphic encryption and vector space model. In the proposed scheme, the majority of computing work is done on the cloud while the user takes part in ranking, which guarantees top k multi-keyword retrieval over encrypted cloud data with high security and practical efficiency.

II. MOTIVATION

In which a log file will be generated at the server for each action done by any user (both owner and data user) for any file. This information can be used by the cloud admin to know about the is happened while uploading, encrypting or downloading the files. This information can also be used by data owner to know the statistics about his file's downloads. The log file act as the permanent storage medium. It has the history of all retrieval in the cloud server.A log file provide more information about the current updates .

III. EXISTING SYSTEM

In order to improve feasibility and save on the expense in the cloud paradigm, it is preferred to get the retrieval result with the most relevant files that match users' interest instead of all the files, which indicates that the files should be ranked in the order of relevance by users' interest and only the files with the highest relevances are sent back to users. A series of searchable symmetric encryption schemes have been proposed to enable search on ciphertext.

Traditional SSE schemes enable users to securely retrieve the ciphertext, but these schemes support only Boolean keyword search, i.e., whether a keyword exists in a file or not, without considering the difference of relevance with the queried keyword of these files in the result. Preventing the cloud from involving in ranking and entrusting all the work to the user is a natural way to avoid information leakage. However, the limited computational power on the user side and the high computational overhead precludes information security.

IV. DISADVANTAGES

To improve security without sacrificing efficiency, schemes presented in show that they support top-k single keyword retrieval under various scenarios.

Authors of made attempts to solve the problem of top-k multi-keyword over encrypted cloud data.

These schemes, however, suffer from two problems - Boolean representation and how to strike a balance between security and efficiency.

In the former, files are ranked only by the number of retrieved keywords, which impairs search accuracy. In the latter, security is implicitly compromised to tradeoff for efficiency, which is particularly undesirable in security-oriented applications.

V. PROPOSED SYSTEM

The concepts of similarity relevance and scheme robustness to formulate the privacy issue in searchable encryption schemes, and then solve the insecurity problem by proposing a two-round searchable encryption (TRSE) scheme. Novel technologies in the cryptography community and information retrieval community are employed, including homomorphic encryption and vector space model. In the proposed scheme, the majority of computing work is done on the cloud while the user takes part in ranking, which guarantees top k multi-keyword retrieval over encrypted cloud data with high security and practical efficiency.

Proposal of log file generation. In which a log file will be generated at the server for each action done by any user (both

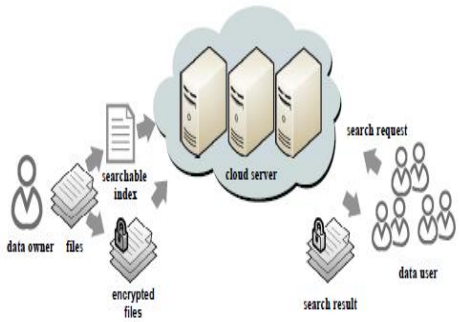
owner and data user) for any file. This information can be used by the cloud admin to know about the issues happened while uploading, encrypting or downloading the files. This information can also be used by data owner to know the statistics about his file's downloads.

VI. ADVANTAGES

The concepts of similarity relevance and scheme robustness. It perform the first attempt to formulate theprivacy issue in searchable encryption, and we show server side ranking based on order-preserving encryption (OPE) inevitably violates data privacy

The two-round searchable encryption (TRSE) scheme, which fulfills the secure multi-keyword top-k retrieval over encrypted cloud data. Specifically, for the first time we employ relevance score to support multi-keyword top-k retrieval. Thorough analysis on security demonstrates the proposed scheme guarantees high data privacy. Furthermore, performance analysis and experimental results show that our scheme is efficient for practical utilization.

VII. SYSTEM ARCIETETURE



VIII. MODULES DESCRIPTION

1 .Data owner module

A cloud computing system hosting data service, as illustrated in Figure in which three different entities are involved: Cloud server, Data owner and Data user. The cloud server hosts third-party data storage and retrieve services. Since data may contain sensitive information, the cloud servers cannot be fully entrusted in protecting data. For this reason, outsourced files must be encrypted. Any kind of information leakage that would affect data privacy is regarded as unacceptable.

2. Encrypt module

To alleviate the computational burden on user side, computing work should be at the server side, need an encryption scheme to guarantee the operability and security at the same time on server side. Homomorphic encryption allows specific types of computations to be carried out on the corresponding ciphertext. The result is the ciphertext of the result of the same operations performed on the plaintext. That is, homomorphic encryption allows computation of ciphertext without knowing anything about the plaintext to get the correct encrypted result.

3. Searchable index module

The data owner has a collection of n files C = {f1, f2, ..., fn} to outsource onto the cloud server in encrypted form and expects the cloud server to provide keyword retrieval service to data owner himself or other authorized users. To achieve this, the data owner needs to build a searchable index I from a collection of l keywords W = 5{w1, w2, ...,wl} extracted out of C, and then outsources both the encrypted index I_ and encrypted files onto the cloud server.

4 Multi-keyword module

This module is used to help the user to get the accurate result based on the multiple keyword concepts. The users can enter the multiple words query, the server is going to split that query into a single word after search that word file in our database. Finally, display the matched word list from the database and the user gets the file from that list. The data user is authorized to process multi-keyword retrieval over the outsourced data. Thus the data user encrypts the query and sends it to the cloud server that returns the relevant files to the data user. Afterwards, the data user can decrypt and make use of the files.

5. Log file generation

A log file will be generated at the server for each action done by any user (both owner and data user) for any file. This information can be used by the cloud admin to know about the issues happened while uploading, encrypting or downloading the files. This information can also be used by data owner to know the statistics about his file's downloads.

IX. CONCLUSION

This motivate and solve the problem of secure multi-keyword top-k retrieval over encrypted cloud data. A two-round searchable encryption (TRSE) scheme employing the fully homomorphic encryption, which fulfills the security requirements of multi-keyword top-k retrieval over the encrypted cloud data. By security analysis,the proposed scheme guarantees data privacy. According to the efficiency evaluation of the proposed scheme over real dataset, extensive experimental results demonstrate that our scheme ensures practical efficiency. A log file will be generated at the server for each action done by any user (both owner and data user) for any file. This information can also be used by data owner to know the statistics about his file's downloads.

X. FUTURE ENHANCEMENT

The end vertices and initial triangle obtained from cover and stegno models are identical in all cases in experiments, it cannot guarantee that both are always exactly identical. One simple solution is to simply project vertices on the x; y; and z-axes. However, this approach cannot withstand similarity transformations. A better approach for determining vertex traverse list is required in the near future.

Another limitation is that this approach cannot withstand certain malicious attacks such as smoothing, additional noise, non uniform scaling, simplification, and vertices re sampling. As

a result, the proposed approach is not suitable for the applications of digital content protection and authentication.

REFERENCES

[1] M. Armbrust, A. Fox, R. Griffith, A. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, and M. Zaharia. "A view of cloud computing," Communication of the ACM 53 (4): 50 58, 2010.

[2] M. Arrington, "Gmail disaster: Reports of mass email deletions," http://www.techcrunch.com/2006/12/28/gmail-disasterreports-of-mass-email-deletions/, December 2006.

[3] Amazon.com, "Amazon s3 availability event: July 20, 2008," http://status.aws.amazon.com/s3-20080720.html, 2008.

[4] RAWA News, "Massive information leak shakes Washington over Afghan war," http://www.rawa.org/temp/runews/2010/08/20/massive-information-leak-shakeswashington-over-afghan-war.html, 2010

[5] AHN, "Romney hits Obama for security information leakage," http://gantdaily.com/2012/07/25/romney-hits-obama-for-security-information-leakage/, 2012

[6] Cloud Security Alliance, "Top threats to cloud computing," http://www.cloudsecurity alliance.org, 2010.

[7] C. Leslie, "NSA has massive database of Americans' phone calls," http://usatoday30.usatoday.com/news/washington/2006-05-10/.

[8] R. Curtmola, J. A. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: improved definitions and efficient constructions," in Proc. of ACM CCS, 2006.

[9] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, "Secure ranked keyword search over encrypted cloud data," in Proc. of ICDCS, 2010.

[10] S. Zerr, D. Olmedilla, W. Nejdl, and W. Siberski, "Zerber+r: Top-k retrieval from a confidential index," in Proc. of EDBT, 2009.

[11] M. van Dijk, C. Gentry, S. Halevi and V. Vaikuntanathan, "Fully Homomorphic Encryption over the Integers," in Gilbert, H. (ed.) EUROCRYPT. LNCS, vol. 6110, pp. 24-43, 2010.

[12] M. Perc, "Evolution of the most common English words and phrases over the centuries," the Journal of the Royal Society Interface, 2012. / mec/2003-2004/.

[13] O. Regev, "New lattice-based cryptographic constructions," JACM 51(6), pp. 899-942, 2004.

[14] N. Howgrave-Graham, "Approximate integer common divisors," in Silverman, J.H. (ed.) CaLC' 01. LNCS, vol. 2146, pp. 51-66, 2001.

[15] NSF Research Awards Abstracts 1990-2003: http://kdd.ics.uci.edu/databases/nsfabs /nsfawards.html.

[16] 20 Newsgroups: http://kdd.ics.uci.edu/databases/20newsgroups/20newsgroups.html.

[17] S. Gries, "Useful statistics for corpus linguistics," in Aquilino Sanchez Moises Almela (eds.), A mosaic of corpus linguistics: selected approaches, 269-291. Frankfurt am Main: Peter Lang, 2010.

[18] C. Gentry, "Fully homomorphic encryption using ideal lattices," in Proc. of STOC, pp.169-178. ACM, New York, 2009.

[19] D. Dubin, "The Most Influential Paper Gerard Salton Never Wrote," LIBRARY TRENDS, Vol. 52, No. 4, pp. 748-764, 2004.

[20] A. Cuyt, V. Brevik Petersen, B. Verdonk, H. Waadeland and W.B. Jones, "Handbook of Continued fractions for Special functions," Springer Verlag, 2008.

[21] T. H. Cormen, C. E. Leiserson, R. L. Rivest, C. Stein, "Introduction to Algorithms," MIT Press and McGraw-Hill. pp. 856-887. 2001.

[22] D. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in Proc. of IEEE Symposium on Security and Privacy, 2000.

[23] D. Boneh, G. Crescenzo, R. Ostrovsky and G. Persiano, "Public-key encryption with keyword Search," in Proc. of Eurocrypt, 2004.

[24] A. Swaminathan, Y. Mao, G.-M. Su, H. Gou, A. L. Varna, S. He, M. Wu, and D. W. Oard, "Confidentiality-preserving rank-ordered search," in Proc. of theWorkshop on Storage Security and Survivability, 2007.

[25] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-preserving multikeyword ranked search over encrypted cloud data," in Proc. of IEEE INFOCOM, 2011.

[26] H. Hu, J. Xu, C. Ren and B. Choi, "Processing private queries over untrusted data cloud through privacy homomorphism," in Proc. of ICDE, 2011.

[27] P. Golle, J. Staddon, and B. Waters, "Secure conjunctive keyword search over encrypted data," in Proc. of ACNS, pp. 31-45, 2004.

[28] L. Ballard, S. Kamara, and F. Monrose, "Achieving efficient conjunctive keyword searches over encrypted data," in Proc. of ICICS, 2005.

[29] J.-S. Coron, A. Mandal, D. Naccache, and M. Tibouchi, "Fully Homomorphic Encryption over the Integers with Shorter Public Keys," in Proc. of CRYPTO, 2011.

[30] N. Smart, F. Vercauteren, "Fully homomorphic encryption with relatively small key and ciphertext sizes," in Proc. of PKC, 2010.

AUTHORS

First Author – Mrs. P.Shanmuga Priya M.E(Ph.d), E mail- priyacse51@gmail.com
Second Author – Preethi.D
Third Author – Priya.J
Fourth Author – Shanthini.B, E mail- shandhini92@gmail.com