ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

Exploration Engine for the Cloud with Privacy Fortification

Rohini R. Mahajan, Ganesh N. Dhanokar

M.E.2nd Year, Dept. of Computer Science Engineering and Information Technology, G.H.R.I.E.M., Jalgaon, Maharashtra, India

Assist. Professor, Dept. of Computer Science Engineering and Information Technology, G.H.R.I.E.M., Jalgaon, Maharashtra, India

ABSTRACT: Besides considering the benefits of cloud computing data, proprietors are developing more priority to put their data in the cloud. Data in the cloud should be encrypted before outsourcing. Considering the huge network traffic and size of the data it is necessary to allow multiple keywords in the search request to get appropriate results. Currently, the mechanism of search encryption is mainly focusing on the limited keywords. In the present paper, for the first time, we are introducing multi-keyword ranked search over encrypted data in the cloud computing (MRSE) by considering privacy. For multi-keyword search mechanism on the cloud, we choose the most similar "Coordinate matching" i.e., the maximum similar results to capture the search query. Initially, the idea proposed for the MRSE based on secure inner product computation, it gives two significantly improved MRSE schemes to achieve various stringent privacy requirements in two different threat models.

KEYWORDS: Accuracy, Cloud computing, Multi-keyword, Privacy preserving

I. INTRODUCTION

Distributed computing is a field of computer science that studies variety of distributed systems. A distributed system is a software system in which components located on networked computers communicate and co-ordinate their actions by passing the messages. The components interact with each other in order to achieve a common goal. There are numerous alternatives for the message passing mechanism, including RPC-like connectors and message queues. Three significant characteristics of distributed systems are concurrency of components, lack of a global clock, and independent failure of components. An important objective and challenge of distributed system is location transparency. Examples of distributed systems vary from SOA-based systems to massively multiplayer online games to peer-to-peer applications.

A computer program that runs in a distributed system is called a 'distributed program', and distributed programming is the process of writing such programs. Distributed computing also refers to the use of distributed systems to solve computational problems. In distributed computing, a problem is divided into many tasks, each of which is solved by one or more computers, which communicate with each other by message passing. The word *distributed* in terms such as "distributed system", "distributed programming", and "distributed algorithm" originally referred to computer networks where individual computers were physically distributed within some geographical area. The terms are nowadays used in a much wider sense, even referring to autonomous processes that run on the same physical computer and interact with each other by message passing. While there is no single definition of a distributed system, the following defining properties are commonly used:

- There are several autonomous computational entities, each of which has its own local memory.
- The entities communicate with each other by message passing.

In this article, the computational entities are called '*computers* or *nodes*'.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

A distributed system may have a common goal, such as solving a large computational problem.[1] Alternatively, each computer may have its own user with individual needs, and the purpose of the distributed system is to co-ordinate the use of shared resources or provide communication services to the users.
Other typical properties of distributed systems include the following:

- The system has to tolerate failures in individual computers.
- The structure of the system (network topology, network latency, the number of computers) is not known in advance, the system may consist of different kinds of computers and network links, and the system may change during the execution of a distributed program.
- Each computer has only a limited, incomplete view of the system. Each computer may identify only one part of the input.

Distributed systems are groups of networked computers, which have the same goal for their work. The terms "concurrent computing", "parallel computing", and "distributed computing" have a lot of overlap, and no clear distinction exists between them. The same system may be characterized both as "parallel" and "distributed"; the processors in a typical distributed system run con-currently in parallel. Parallel computing may be seen as a particular tightly coupled form of distributed computing, and distributed computing may be seen as a loosely coupled form of parallel computing. Nevertheless, it is possible to roughly classify concurrent systems as "parallel" or "distributed" using the following criteria:

II. RELATED WORK

N. Cao, C. Wang, M. Li, K. Ren, and W. Lou
With the advent of cloud computing, data owners are motivated to outsource their complex data management systems from local sites to the commercial public cloud for great flexibility and economic savings. But for protecting data privacy, sensitive data have to be encrypted before outsourcing, which obsoletes traditional data utilization based on plaintext keyword search. Thus, enabling an encrypted cloud data search service is of paramount importance. Considering a large number of data users and documents in the cloud, it is necessary to allow multiple keywords in the search request and return documents in the order of their relevance to these keywords. Related works on searchable encryption focus on single keyword search or Boolean keyword search, and rarely categorize the search results. In the present work, for the first time, we define and solve the challenging problem of privacy-preserving multi-keyword ranked search over encrypted data in cloud computing (MRSE).

S. Kamara and K. Lauter
We consider the problem of building a secure cloud storage service on top of a public cloud infrastructure where the service provider is not completely trusted by the customer. We describe, at a high level, several architectures that combine recent and non-standard cryptographic primitives in order to achieve our goal. We survey the benefits such architecture would provide to both customers and service providers and give an overview of recent advances in cryptography motivated specifically by cloud storage.

Y.-C. Chang and M. Mitzenmacher
We consider the following problem: a user $\mathcal{U}$ wants to store his files in an encrypted form on a remote file server $\mathcal{S}$. Later the user $\mathcal{U}$ wants to efficiently retrieve some of the encrypted files containing (or indexed by) specific keywords, keeping the keywords themselves secret and not jeopardizing the security of the remotely stored files. For example, a user may want to store old e-mail messages encrypted on a server managed by Yahoo or another large vendor, and later retrieve certain messages while travelling with a mobile device. In our studies, we offer solutions for this issue under well-defined security requirements. Our schemes are efficient in the sense that no public-key cryptosystem is involved. Indeed, our approach is independent of the encryption method chosen for the remote files. They are also incremental, in that $\mathcal{U}$ can submit new files which are secure against previous queries but still searchable against future queries.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

A. EXISTING SYSTEM

The effective data retrieval needs the large amount of documents demand the cloud server to perform result relevance ranking, instead of returning undifferentiated results. Such ranked search system enables data users to find the most relevant information quickly, rather than burdensomely sorting through every match in the content collection. Ranked search can also elegantly eliminate unnecessary network traffic by sending back only the most relevant data, which is highly desirable in the "pay-as-you-use" cloud paradigm. For privacy protection, such ranking operation, however should not leak any keyword related information. On the other hand, to improve the search result accuracy as well as to enhance the user searching experience, it is also necessary for such ranking system to support multiple keywords search, as a single keyword search often yields far too coarse results.

B. DRAWBACKS OF EXISTING SYSTEM

- The encrypted cloud data search system remains a very challenging task because of inherent security and privacy obstacles, including various strict requirement.
- On enrich the search flexibility, they are still not adequate to provide users with acceptable result ranking functionality.

III.PROPOSED METHODOLOGY AND DISCUSSION

For the first time, we exemplify and resolve the problem of multi-keyword ranked search over encrypted cloud data (MRSE) while preserving strict system-wise privacy in the cloud computing paradigm. Among various multi-keyword semantics, we choose the efficient similarity measure of "co-ordinate matching," i.e., as many matches as possible, to capture the relevance of data documents to the search query. Specifically, we use "inner product similarity", i.e., the number of query keywords appearing in a documents, to quantitatively evaluate such similarity measure of that document to the search query. During the index construction, each document is associated with a binary vector as a sub-index where each bit represents whether corresponding keyword is contained in the document. Thesearch query is also described as a binary vector where each bit means whether the corresponding keyword appears in this search request, so the similarity could be exactly measured by the inner product of the query vector with the data vector. However, directly outsourcing the data vector or the query vector will violate the index privacy or the search privacy. To meet the challenge of supporting such multi-keyword semantic without privacy breaches, we propose a basic idea for the MRSE using secure inner product computation, which is adapted from a secure k-nearest neighbour (kNN) technique, and then give two significantly improved MRSE schemes in a step-by-step manner to achieve various stringent privacy requirements.

A. MODULES

- Data Owner Module
- Data User Module
- Encryption Module
- Rank Search Module

B. MODULES DESCRIPTION

Data Owner Module
This module helps the owner to register those details and also include login details. This module helps the owner to upload his file with encryption using RSA algorithm. This ensures the files to be protected from an unauthorized user.

Data User Module
This module includes the user registration login details. This module is used to help the client to search the file using the multiple keywords concept and get the accurate result list based on the user query. The user is going to select the

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

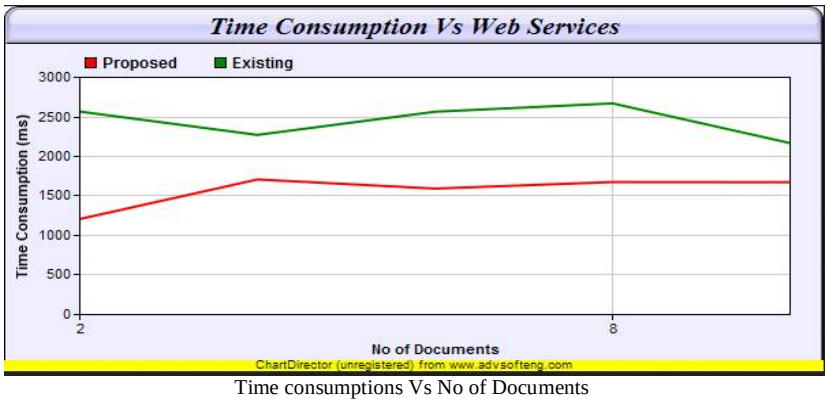
required file and register the user details and get activation code in mail email before entering the activation code. After user can download the Zip file and extract that file.

Encryption Module:
This module is used to help the server to encrypt the document using RSA Algorithm and to convert the encrypted document to the Zip file with activation code and then an activation code sends to the user for download.

Rank Search Module
These modules ensure the user to search the files that are searched frequently using rank search.This module allows the user to download the file using his secret key to decrypt the downloaded data. This module allows the Owner to view the uploaded files and downloaded files

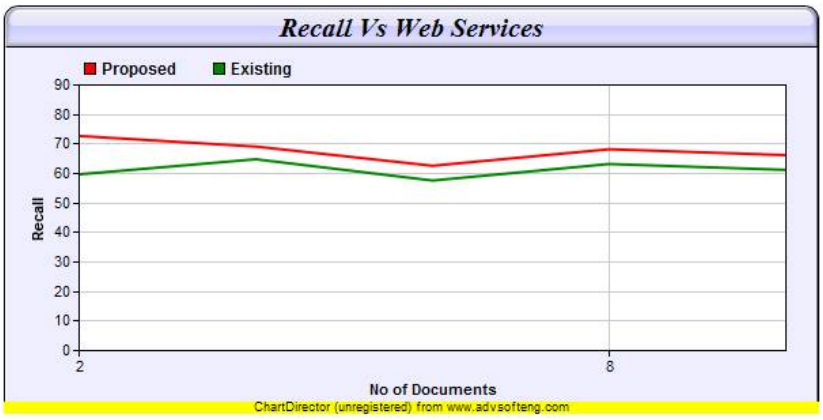
IV. EXPERIMENTAL RESULTS WITH GRAPHS

For the corroboration of the accuracy of the algorithm theresults obtained from the algorithm were compared in the ground truth dataset. It is evident that the proposed algorithm was able tosuccessfully details with more than95% accuracy. The algorithm's accuracy of searching data over encrypted cloud was higher (98.1%). The results showed that our algorithm could maintainthe high accuracy. The graphical representations of the results are shown in the following figures and the results are discussed in the legends of respective figure.



Time consumptions Vs No of Documents

Fig.1 describes the total time consumption of the previous over the existing systems, with comparison to number of documents. As per figure above it can be seen that proposed system takes less time than previous system. However, the outcome of the results may slightly vary depending on the change of environment and configurations



Recall Vs No of Documents

Fig.2 illustrates recall of the previous and current systems generated results, with comparison to number of documents. As per figure above it can be seen that proposed system has higher recall percentage than previous system.

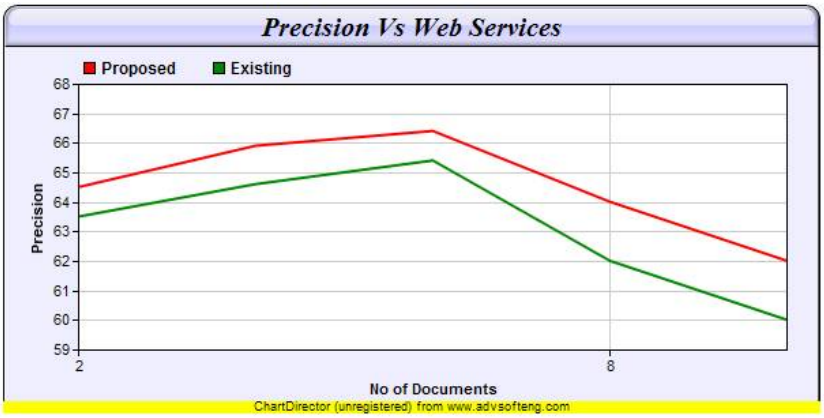
ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016



Precision Vs No. of Documents

Fig.3 demonstrates precision of the previous and existing systems generated results, with comparison to number of documents. The graphical representation shows that proposed system has higher percentage of precision than previous system.

V. CONCLUSION

Collectively, we define,confirmed and solve the problem of multi-keyword ranked search over encrypted cloud data, and established a variety of privacy requirements. Among various multi-keyword semantics, we choose the efficient similarity measure of "coordinate matching," i.e., as many matches as possible, to effectively capture the relevance of outsourced documents to the query keywords, and use "inner product similarity" to quantitatively evaluate such similarity measure. For meeting the challenge of supporting multi-keyword semantic without privacy breaches, we proposed a basic idea of MRSE using secure inner product computation. Additionally, we provided two improved MRSE schemes to achieve various stringent privacy requirements in two distinctive threat models. Further we also investigated some auxiliary enhancements of our ranked search mechanisms, including supporting more search semantics, i.e., TF_IDF, and dynamic data operations. Comprehensive analysis investigating privacy and efficiency securities of proposed schemes is given, and experiments on the real-world data set show our proposed schemes introduces low overhead on both computation and communication. In future investigation, we will explore examining the integrity of the rank order in the search result assuming the cloud server is untrusted.

REFERENCES

1. Cao, N. C., Wang, M. Li., Ren, K. and Lou, W., "Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data," Proc. IEEE INFOCOM, pp. Ren, 829-837, 2011.
2. Vaquero, L.M., Rodero-Merino, L. Caceres, J., and Lindner, M., "A Break in the Clouds: Towards a Cloud Definition," ACM SIGCOMM J., Comput. Commun. Rev., Vol. 39, no. 1, pp. 50-55, 2009.
3. Cao, N., Yu, S., Yang, Z., Lou, W., and Hou, Y., "LT Codes-Based Secure and Reliable Cloud Storage Service," Proc. IEEE INFOCOM, pp. 693-701, 2012.
4. Kamara,S., and Lauter K., "Cryptographic Cloud Storage," Proc. 14th Int'l Conf. Financial Cryptograpy and Data Security, 2010.
5. Singhal, A. "Modern Information Retrieval: A Brief Overview," IEEE Data Eng. Bull., Vol. 24, no. 4, pp. 35-43, 2001.
6. Witten I.H., Moffat, and Bell, T.C., "Managing Gigabytes: Compressing and Indexing Documents and Images". A. Morgan Kaufmann Publishing, 1999.
7. Song, D., Wagner, D., and Perrig, A., "Practical Techniques for Searches on Encrypted Data," Proc. IEEE Symp. Security and Privacy, 2000.
8. Goh E. J., "Secure Indexes," Cryptology ePrint Archive, http:// eprint.iacr.org/2003/216, 2003.
9. Chang,Y. C. and Mitzenmacher, M.,"Privacy Preserving Keyword Searches on Remote Encrypted Data," Proc. Third Int'l Conf. Applied Cryptography and Network Security, 2005.
10. Curtmola, J., Garay, A., Kamara, S., and Ostrovsky, R., "Searchable Symmetric Encryption: R. Improved Definitions and Efficient Constructions," Proc. 13th ACM Conf. Computer and Comm. Security (CCS '06), 2006.
11. Boneh, D., Crescenzo, C.D., Ostrovsky, R., and Persiano, G., "Public Key Encryption with Keyword Search," Proc. Int'l Conf. Theory and Applications of Cryptographic Techniques (EUROCRYPT), 2004.
12. Bellare, M., Boldyreva, A., and ONeill, A., "Deterministic and Efficiently Searchable Encryption," Proc. 27th Ann. Int'l Cryptology Conf. Advances in Cryptology (CRYPTO '07), 2007.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

13. Abdalla, M., Bellare, M., Catalano D., Kiltz, E., Kohno, T., Lange, T., Malone-Lee, J., Neven, G., Paillier, P., and Shi, H., "Searchable Encryption Revisited: Consistency Properties, Relation to Anonymous Ibe, and Extensions," J. Cryptology, vol. 21, no. 3, pp. 350- 391, 2008.
14. Li, J., Wang, Q., Wang, C., Cao, N., Ren, K., and Lou, W., "Fuzzy Keyword Search Over Encrypted Data in Cloud Computing," Proc. IEEE INFOCOM, 2010.
15. Boneh, D., Kushilevitz, E.,and Ostrovsky, R., W.E.S. III, "Public Key Encryption That Allows PIR Queries," Proc. 27th Ann. Int'l Cryptology Conf. Advances in Cryptology (CRYPTO '07), 2007.
16. Golle, P., Staddon, J., and Waters, B., "Secure Conjunctive Keyword Search over Encrypted Data," Proc. Applied Cryptography and Network Security, pp. 31-45, 2004.
17. Ballard, L., Kamara, S., and Monrose, F., "Achieving Efficient Conjunctive Keyword Searches over Encrypted Data," Proc.Seventh Int'l Conf. Information and Comm. Security (ICICS '05), 2005.
18. Boneh,D., and Waters, B., "Conjunctive, Subset, and Range Queries on Encrypted Data," Proc. Fourth Conf. Theory Cryptography (TCC), pp. 535-554, 2007.
19. Brinkman, R., "Searching in Encrypted Data," PhD thesis, Univ. of Twente, 2007.
20. Hwang,Y., and Lee, P., "Public Key Encryption with Conjunctive Keyword Search and Its Extension to a Multi-User System," Pairing, Vol. 4575, pp. 2-22, 2007.
21. Katz, J., Sahai, A., and Waters, B., "Predicate Encryption Supporting Disjunctions, Polynomial Equations, and Inner Products," Proc. 27th Ann. Int'l Conf. Theory and Applications of Cryptographic Techniques (EUROCRYPT), 2008.
22. Lewko, T. Okamoto, A. Sahai, K. Takashima, and B. Waters, "Fully Secure Functional Encryption: Attribute-Based Encryption and (Hierarchical) Inner Product Encryption," Proc. 29th Ann. Int'l Conf. Theory and Applications of Cryptographic Techniques (EUROCRYPT '10), 2010.
23. Shen, E., Shi, E., and Waters, B., "Predicate Privacy in Encryption Systems," Proc. Sixth Theory of Cryptography Conf. Theory of Cryptography (TCC), 2009.
24. Li, M., Yu, S., Cao, N., and Lou, W., "Authorized Private Keyword Search over Encrypted Data in Cloud Computing," Proc. 31st Int'l Conf. Distributed Computing Systems (ICDCS '10), pp. 383- 392, June 2011.
25. Wang C., Cao, N., Li J., Ren, K., and Lou, W., "Secure Ranked Keyword Search over Encrypted Cloud Data," Proc. IEEE 30th Int'l Conf. Distributed Computing Systems (ICDCS '10), 2010.
26. Wang, C., Cao, N., Ren, K., and Lou, W., "Enabling Secure and Efficient Ranked Keyword Search over Outsourced Cloud Data," IEEE Trans. Parallel and Distributed Systems, vol. 23, no. 8, pp. 1467- 1479, Aug. 2012.
27. Wong, W. K.,Cheung, D. Kao, W., and Mamoulis,N., "Secure kNN Computation on Encrypted Databases," Proc. 35th B. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD), pp. 139-152, 2009.
28. Yu, S., Wang, C., Ren, K., and Lou, W., "Achieving Secure, Scalable, and Fine-Grained Data Access Control in Cloud Computing," Proc. IEEE INFOCOM, 2010.
29. Wang, C., Wang, Q., Ren, K. and Lou, W, "Privacy-Preserving Public Auditing for Data Storage Security in Cloud Computing," Proc. IEEE INFOCOM, 2010.
30. Zerr, S., Demidova, E., Olmedilla, D., Nejdl, W., Winslett, M., and Mitra, S., "Zerber: r-Confidential Indexing for Distributed Documents," Proc. 11th Int'l Conf. Extending Database Technology (EDBT '08), pp. 287-298, 2008.