ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 2, February 2017

Confidentiality Preserving Efficient and Dynamic Multi-Keyword Ranked Search over Encoded Multi Cloud Data

J. Ajitha[1], S. Chitra Devi[2]

M.E. Computer Science and Engineering, University College of Engineering, Tiruchirapalli,Bit Campus, Mandaiyur, Trichy-620024, India

Assistant Professor, Department of Computer Science and Engineering, University College of Engineering, Tiruchirapalli, Bit Campus, Mandaiyur, Trichy, India

ABSTRACT: In this system a special tree-based index structure is constructed and proposes a "Greedy Depth-first Search" algorithm to provide efficient multi-keyword ranked search. The proposed scheme can achieve sub-linear search time and deal with the deletion and insertion of documents flexibly. Extensive experiments are conducted to demonstrate the efficiency of the proposed scheme. Due to the increasing popularity of cloud computing, more and more data owners are motivated to outsource their data to cloud servers for great convenience and reduced cost in data management. However, sensitive data should be encrypted before outsourcing for privacy requirements, which obsoletes data utilization like keyword-based document retrieval. In this system, we present a secure multi-keyword ranked search scheme over encrypted cloud data, which simultaneously supports dynamic update operations like deletion and insertion of documents. Specifically, the vector space model and the widely-used TFIDF model are combined in the index construction and query generation. We construct a special tree-based index structure and propose a "Greedy Depth-first Search" algorithm to provide efficient multi-keyword ranked search. The secure kNN algorithm is utilized to encrypt the index and query vectors, and meanwhile ensure accurate relevance score calculation between encrypted index and query vectors. In order to resist statistical attacks, phantom terms are added to the index vector for blinding search results. Due to the use of our special tree-based index structure, the proposed scheme can achieve sub-linear search time and deal with the deletion and insertion of documents flexibly. Extensive experiments are conducted to demonstrate the efficiency of the proposed scheme.

KEYWORDS: Attribute Based Encryption, Secure Data Maintenance, Dynamic Searching Scheme, Cloud Computing.

I. INTRODUCTION

Cloud computing has been considered as a new model of enterprise IT infrastructure, which can organize huge resource of computing, storage and applications, and enable users to enjoy ubiquitous, convenient and on-demand network access to a shared pool of configurable computing resources with great efficiency and minimal economic overhead . Attracted by these appealing features, both individuals and enterprises are motivated to outsource their data to the cloud, instead of purchasing software and hardware to manage the data themselves.

Despite of the various advantages of cloud services, outsourcing sensitive information (such as e-mails, personal health records, company finance data, government documents, etc.) to remote servers brings privacy concerns. The cloud service providers (CSPs) that keep the data for users may access users' sensitive information without authorization. A general approach to protect the data confidentiality is to encrypt the data before outsourcing. However, this will cause a huge cost in terms of data usability. For example, the existing techniques on keyword-based

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 2, February 2017

information retrieval, which are widely used on the plaintext data, cannot be directly applied on the encrypted data. Downloading all the data from the cloud and decrypt locally is obviously impractical.

II. EXISTING SYSTEM

A general way to deal with secure the information privacy is to scramble the information before outsourcing. Searchable encryption plans empower the customer to store the encoded information to the cloud and execute watchword seek over ciphertext space. As such, bottomless works have been proposed under various danger models to accomplish different inquiry usefulness, for example, single catchphrase pursuit, closeness seek, multi-watchword boolean hunt, positioned look, multi-watchword positioned seek, and so forth. Among them, multi-watchword positioned look accomplishes increasingly consideration for its down to earth pertinence. As of late, some dynamic plans have been proposed to bolster embeddings and erasing operations on archive accumulation. These are critical acts as it is exceptionally conceivable that the information proprietors need to redesign their information on the cloud server.

Disadvantages

- Huge cost in terms of data usability. For example, the existing techniques on keyword-based information retrieval, which are widely used on the plaintext data, cannot be directly applied on the encrypted data. Downloading all the data from the cloud and decrypt locally is obviously impractical.
- Existing System methods not practical due to their high computational overhead for both the cloud sever and user.

III. PROPOSED METHODOLOGY

This system proposes a secure tree-based search scheme over the encrypted cloud data, which supports multi-keyword ranked search and dynamic operation on the document collection. Specifically, the vector space model and the widely-used “term frequency (TF) × inverse document frequency (IDF)” model are combined in the index construction and query generation to provide multi-keyword ranked search. In order to obtain high search efficiency, we construct a tree-based index structure and propose a “Greedy Depth-first Search” algorithm based on this index tree.The secure kNN algorithm is utilized to encrypt the index and query vectors, and meanwhile ensure accurate relevance score calculation between encrypted index and query vectors. To resist different attacks in different threat models, we construct two secure search schemes: the basic dynamic multi-keyword ranked search (BDMRS) scheme in the known ciphertext model, and the enhanced dynamic multi-keyword ranked search (EDMRS) scheme in the known background model.

Advantages

- Due to the special structure of our tree-based index, the proposed search scheme can flexibly achieve sub-linear search time and deal with the deletion and insertion of documents.
- We design a searchable encryption scheme that supports both the accurate multi-keyword ranked search and flexible dynamic operation on document collection.
- Due to the special structure of our tree-based index, the search complexity of the proposed scheme is fundamentally kept to logarithmic. And in practice, the proposed scheme can achieve higher search efficiency by executing our “Greedy Depth-first Search” algorithm. Moreover, parallel search can be flexibly performed to further reduce the time cost of search process.

IV. LITERATURE SURVEY

In the year of 2012, the authors K. Ren, C.Wang, Q.Wang et al. revealed a paper titled "Security challenges for the public cloud" and describe into the paper such as Distributed computing speaks to today's most energizing processing outlook change in data innovation. In any case, security and protection are seen as essential deterrents to its

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 2, February 2017

wide reception. Here, the creators plot a few basic security challenges and rouse promote examination of security answers for a dependable open cloud environment.

In the year of 2010, the authors S. Kamara and K. Lauter Cryptographic cloud storage in Financial Cryptography and Data Security" and describe into the paper such as We consider the issue of building a protected distributed storage benefit on top of an open cloud framework where the specialist organization is not totally trusted by the client. We portray, at an abnormal state, a few designs that join later and non-standard cryptographic primitives with a specific end goal to accomplish our objective. We review the advantages such engineering would give to both clients and specialist organizations and give an outline of late advances in cryptography inspired particularly by distributed storage.

In the year of 2009, the author C. Gentry revealed a paper titled "A fully homomorphic encryption scheme" and describe into the paper such as We propose the primary completely homomorphic encryption plot, taking care of an old open issue. Such a plan permits one to register subjective capacities over encoded information without the decoding key—i.e., given encryptions E(m1), ..., E(mt) of m1, ..., m t, one can effectively process a conservative ciphertext that scrambles f(m1, ..., m t) for any proficiently calculable capacity f. Completely homomorphic encryption has various applications.

For instance, it empowers encoded internet searcher questions—i.e., a web crawler can give you a compact scrambled solution for your (boolean) inquiry without recognizing what your inquiry was. It additionally empowers seeking on encoded information; you can store your scrambled information on a remote server, and later have the server recover just records that (when unscrambled) fulfill some boolean requirement, despite the fact that the server can't decode the documents all alone.

All the more comprehensively, it enhances the proficiency of secure multiparty calculation. In our answer, we start by planning a to some degree homomorphic "boostrappable" encryption plot that works when the capacity f is the plan's own unscrambling capacity. We then show how, through recursive self-implanting, bootstrappable encryption gives completely homomorphic encryption.

V. SYSTEM ARCHITECTURE DESIGN

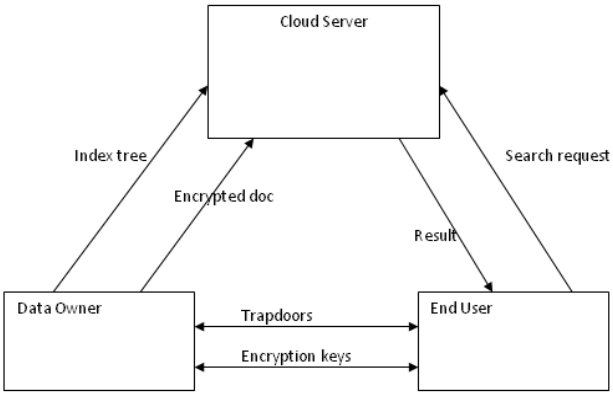

Fig.1 System Architecture Design

Cloud Server

Cloud server stores the encrypted document collection C and the encrypted searchable tree index I for data owner. Upon receiving the trapdoor TD from the data user, the cloud server executes search over the index tree I, and finally returns the corresponding collection of top-k ranked encrypted documents. Besides, upon receiving the update information from the data owner, the server needs to update the index I and document collection C according to the received information.

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 2, February 2017

Data Owner

Data owner has a collection of documents that he wants to outsource to the cloud server in encrypted form while still keeping the capability to search on them for effective utilization. In our scheme, the data owner firstly builds a secure searchable tree index I from document collection F, and then generates an encrypted document collection C for F. Afterwards, the data owner outsources the encrypted collection C and the secure index I to the cloud server, and securely distributes the key information of trapdoor generation (including keyword IDF values) and document decryption to the authorized data users.

Data User

Data users are authorized ones to access the documents of data owner. With t query keywords, the authorized user can generate a trapdoor TD according to search control mechanisms to fetch k encrypted documents from cloud server. Then, the data user can decrypt the documents with the shared secret key.

KNN Algorithm

Knn.java: This is the main driver of the code. To do the classification, we are essentially interested in finding the distance between the particular instance we are trying to classify to other instances. We then determine the classification of the instance we want from a "majority vote" of the other k closest instances. Each feature of an instance is a separate class that essentially just stores a continuous or discrete value depending on if you are using regression or not to classify your neighbors. The additional feature classes and file reader are left to the reader as an exercise. Note that it would be fairly easy to weight features using this model depending on if you want to give one feature more clout than another in determining the neighbors.

The GreedyDP Algorithm

Given the complexity of our problem, a more practical solution would be a near-optimal greedy algorithm. As preliminary, we introduce an operator _t _! Called prune-leaf, which indicates the removal of a leaf topic t from a profile. Formally, we denote by Gi_t _! Giþ1 the process of pruning leaf t from Gi to obtain Giþ1. Obviously, the optimal profile G _ can be generated with a finite-length transitive closure of prune-leaf.

The first greedy algorithm GreedyDP works in a bottomup manner. Starting from G0, in every ith iteration, GreedyDP chooses a leaf topic t 2 TGi ðqÞ for pruning, trying to maximize the utility of the output of the current iteration, namely Giþ1. During the iterations, we also maintain a best profile- so-far, which indicates the Giþ1 having the highest discriminating power while satisfying the _-risk constraint. The iterative process terminates when the profile is generalized to a root-topic. The best-profile-so-far will be the final result (G_) of the algorithm. The main problem of GreedyDP is that it requires recompilation of all candidate profiles (together with their discriminating power and privacy risk) generated from attempts of prune-leaf on all t 2 TGi ðqÞ. This causes significant memory requirements and computational cost.

VI. EXPERIMENTAL RESULTS

Fig.2 Uploading File into Server

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 2, February 2017

Fig.3 View File Status

Fig.4 Search for Data

VII. CONCLUSION AND FUTURE SCOPE

In this system, a secure, efficient and dynamic search scheme is proposed, which supports not only the accurate multi-keyword ranked search but also the dynamic deletion and insertion of documents. We construct a special keyword balanced binary tree as the index, and propose a "Greedy Depth-first Search" algorithm to obtain better efficiency than linear search. In addition, the parallel search process can be carried out to further reduce the time cost. The security of the scheme is protected against two threat models by using the secure kNN algorithm. Experimental results demonstrate the efficiency of our proposed scheme. There are still many challenge problems in symmetric SE schemes. In the proposed scheme, the data owner is responsible for generating updating information and sending them to the cloud server. Thus, the data owner needs to store the unencrypted index tree and the information that are necessary to recalculate the IDF values. Such an active data owner may not be very suitable for the cloud computing model.

It could be a meaningful but difficult future work to design a dynamic searchable encryption scheme whose updating operation can be completed by cloud server only, meanwhile reserving the ability to support multi-keyword ranked

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 2, February 2017

search. In addition, as the most of works about searchable encryption, our scheme mainly considers the challenge from the cloud server. Actually, there are many secure challenges in a multi-user scheme. Firstly, all the users usually keep the same secure key for trapdoor generation in a symmetric SE scheme. In this case, the revocation of the user is big challenge. If it is needed to revoke a user in this scheme, we need to rebuild the index and distribute the new secure keys to all the authorized users. Secondly, symmetric SE schemes usually assume that all the data users are trustworthy. It is not practical and a dishonest data user will lead to many secure problems. For example, a dishonest data user may search the documents and distribute the decrypted documents to the unauthorized ones. Even more, a dishonest data user may distribute his/her secure keys to the unauthorized ones. In the future works, we will try to improve the SE scheme to handle these challenge problems.

REFERENCES

[1] K. Ren, C.Wang, Q.Wang et al., "Security challenges for the public cloud," IEEE Internet Computing, vol. 16, no. 1, pp. 69–73, 2012.
[2] S. Kamara and K. Lauter, "Cryptographic cloud storage," in Financial Cryptography and Data Security. Springer, 2010, pp. 136–149.
[3] C. Gentry, "A fully homomorphic encryption scheme," Ph.D. dissertation, Stanford University, 2009.
[4] O. Goldreich and R. Ostrovsky, "Software protection and simulation on oblivious rams," Journal of the ACM (JACM), vol. 43, no. 3, pp. 431–473, 1996.
[5] D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search," in Advances in Cryptology-Eurocrypt 2004. Springer, 2004, pp. 506–522.
[6] D. Boneh, E. Kushilevitz, R. Ostrovsky, and W. E. Skeith III, "Public key encryption that allows pir queries," in Advances in Cryptology-CRYPTO 2007. Springer, 2007, pp. 50–67.
[7] D. X. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in Security and Privacy, 2000. S&P 2000. Proceedings. 2000 IEEE Symposium on. IEEE, 2000, pp. 44–55.
[8] E.-J. Goh et al., "Secure indexes." IACR Cryptology ePrint Archive, vol. 2003, p. 216, 2003.
[9] Y.-C. Chang and M. Mitzenmacher, "Privacy preserving keyword searches on remote encrypted data," in Proceedings of the Third international conference on Applied Cryptography and Network Security. Springer-Verlag, 2005, pp. 442–455.
[10] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: improved definitions and efficient constructions," in Proceedings of the 13th ACM conference on Computer and communications security. ACM, 2006, pp. 79–88.
[11] J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou, "Fuzzy keyword search over encrypted data in cloud computing," in INFOCOM, 2010 Proceedings IEEE. IEEE, 2010, pp. 1–5.
[12] M. Kuzu, M. S. Islam, and M. Kantarcioglu, "Efficient similarity search over encrypted data," in Data Engineering (ICDE), 2012 IEEE 28th International Conference on. IEEE, 2012, pp. 1156–1167.
[13] C. Wang, K. Ren, S. Yu, and K. M. R. Urs, "Achieving usable and privacy-assured similarity search over outsourced cloud data," in INFOCOM, 2012 Proceedings IEEE. IEEE, 2012, pp. 451–459.
[14] B. Wang, S. Yu, W. Lou, and Y. T. Hou, "Privacy-preserving multikeyword fuzzy search over encrypted data in the cloud," in IEEE INFOCOM, 2014.
[15] P. Golle, J. Staddon, and B. Waters, "Secure conjunctive keyword search over encrypted data," in Applied Cryptography and Network Security. Springer, 2004, pp. 31–45.
[16] Y. H. Hwang and P. J. Lee, "Public key encryption with conjunctive keyword search and its extension to a multi-user system," in Proceedings of the First international conference on Pairing-Based Cryptography. Springer-Verlag, 2007, pp. 2–22.
[17] L. Ballard, S. Kamara, and F. Monrose, "Achieving efficient conjunctive keyword searches over encrypted data," in Proceedings of the 7th international conference on Information and Communications Security. Springer-Verlag, 2005, pp. 414–426.
[18] D. Boneh and B. Waters, "Conjunctive, subset, and range queries on encrypted data," in Proceedings of the 4th conference on Theory of cryptography. Springer-Verlag, 2007, pp. 535–554.
[19] B. Zhang and F. Zhang, "An efficient public key encryption with conjunctive-subset keywords search," Journal of Network and Computer Applications, vol. 34, no. 1, pp. 262–267, 2011.
[20] J. Katz, A. Sahai, and B. Waters, "Predicate encryption supporting disjunctions, polynomial equations, and inner products," in Advances in Cryptology–EUROCRYPT 2008. Springer, 2008, pp. 146–162.