Privacy Preserving and Deterministic Data Retrieval In Outsourced Cloud Data

ASAITHAMBI.N
School Of Computer Science and Engineering, Bharathidasan University, Trichy

Abstract
Searching a keyword using an encryption method allows one to upload strongly encrypted documents on the cloud environment and it facilitate the query to locate the data on the server without need of the documents to be decrypted before doing the searching. In this proposal we have proposed novel secure and efficient keyword searching that produces the result based on the matching data items in the ranked order. Unlike other previous methods our searching complexity is sub linear to the total documents that are relevant to the keywords. This research shows that proposed approach is proved to be secured against the adaptive chosen-keyword attacks and it proves that our proposed approach is highly efficient and ready to implement in the real world cloud systems.

Keywords
Cloud Computing; Multi-keyword ranked Search; Privacy Preserving; Confidential Data

I. Introduction

Cloud computing involved highly available massive compute and storage platforms offering a wide range of services. It enables the convenient, on-demand network access to centralized resources that can be rapidly brings into effective action with the great efficiency and minimum management overhead. The advantages of Cloud Computing include: On-demand self-service, easy network access, frequent resource elasticity, pooling, usage-based pricing. As Cloud Computing becomes predominant, more easily offended information to be centralized into the cloud. So, the fact that data owners and cloud server are not in equal trusted domain may put the sourced data at risk. Thus, data encryption makes effective utilization of a very challenging task given that there could be a large amount of outsourced data files.

Cloud computing also called as a utility computing since it uses pay per use paradigm. Users have to pay for the usages. With the technology of cloud computing, users can access variety of resources like programs, storage and application development platforms. Cloud is the extension of object oriented programming and it uses the concept of abstraction.

Cloud computing is an emerging technology which helps as an utility, through which clients are going to store their data in the cloud server and using applications from a set of computing resources[1]. Here sensitive data is going to be centralized in the server. In some times the cloud server may leaks the data to hackers [2]. The data is going to encrypted before outsourced to achieve privacy. The encryption techniques increase the data utilization from a large amount of data. To retrieve data files we introduced keyword search mechanism. By this mechanism the users are going to retrieve the data files of their interest. In traditional search, encryption techniques the users are going to search data by using keywords without decrypting it, they support only Boolean keyword search only [2][10]. In cloud computing graded keyword search enhances the system usability by displaying the matching files by the help of relevance score. To achieve security and usability we introduce advanced cryptographic and information retrieval techniques, and using one-to-many order preserving symmetric encryption [3].

A. Types Cloud Computing:

Basically there are three types of public cloud Services:

1. Infrastructure as a service (IaaS):

In this most basic cloud service model, IaaS providers offer computers, as physical or more often as virtual machines, and other resources.

2. Platform as a service (PaaS):

In the PaaS model, cloud providers deliver a computing platform typically including operating system, programming language execution environment, database, and web server. Application developers can develop and run their software solutions on a cloud platform without the cost and complexity of buying and managing the underlying hardware and software layers.

3. Software as a service (SaaS):

In the SaaS model, cloud providers install and operate application software in the cloud and cloud users access the software from cloud clients. The cloud users do not manage the cloud infrastructure and platform on which the application is running. This eliminates the need to install and run the application on the cloud user's own computers simplifying maintenance and support. What makes a cloud application different from other applications is its scalability.

Manuscript received September 5, 2016
Manuscript revised September 20, 2016

II. Problem Formulation

To activate ranked search for effective utilization of outsourced cloud data, our system design should simultaneously achieve security and performance guarantees as follows.

Privacy and Confidentiality:
When the user host data to the cloud there should be some guarantee that the data will have limited authorized access. Unauthorized access to customer sensitive data by is another risk that can create potential threat to cloud data. Assurance should be provided with various security and privacy policies to secure cloud data.

Data Integrity:
While preserving the privacy of the data, the cloud provider should make sure that the cloud data will persist the data integrity. Since while hosting data on cloud, it undergoes many changes and mechanisms, so cloud provider should ensure the data integrity. For compliance purposes, it may be necessary to have exact records as to what data was placed in a public cloud, when it occurred, what virtual memories (VMs) and storage it resided on, and where it was processed.

Secured Multi-keyword Ranked Search:
To design search schemes which allow multi-keyword query and provide result similarity ranking for valuable data retrieval, instead of returning undifferentiated results.

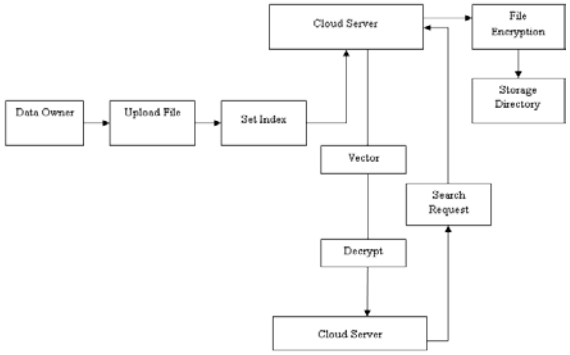

Fig: 1. System Architecture

Effectiveness with high performance:
Above goals on functionality and privacy should be achieved with low communication and computation overhead.

Algorithm Specification:
Our access control scheme is constructed on prime order groups, because the group operations on prime order groups are much faster than the ones on composite order groups. In this section, we will prove that our dynamic policy access control scheme is secure in the generic bilinear group model and random model.

Input: {Vi} → Set of piece vectors for each attribute Attri
Input: s → the secret to be shared
Output: M → Monotone Span Program
1: let ~Z be a vector and set ~Z (0) = s;
2: let M be a matrix;
3: let r be a labelling function;
4: for all Attri do
5: for each piece vector ~Vi for Attri; do
6: apppend each random value in ~Vi to ~Z;
7: construct the position vector ~vi for Attri;
8: append ~vi to M;
9: let r(Mvi) to Attri;
10: end for
11: end for
12: pad M with the same row size;
13: return (~Z;M;r);

III. Literature Review

Mikhail Strizhov and Indrajit Ray "Multi-keyword Similarity Search Over Encrypted Cloud Data", International Journal of Emerging Technology and Advanced Engineering, 2010.
Searchable encryption allows one to upload encrypted documents on a remote honest-but-curious server and query that data at the server itself without requiring the documents to be decrypted prior to searching. In this work, we propose a novel secure and efficient multi-keyword similarity searchable encryption (MKSim) that returns the matching data items in a ranked ordered manner. Unlike all previous schemes, our search complexity is sublinear to the total number of documents that contain the queried set of keywords. Our analysis demonstrates that proposed scheme is proved to be secure against adaptive chosen-keyword attacks. We show that our approach is highly efficient and ready to be deployed in the real-world cloud storage systems.

Prof. C. R. Barde, Pooja Katkade, Deepali Shewale, Rohit Khatale "Secured Multiple-keyword Search over Encrypted Cloud Data" International Journal of Emerging Technology and Advanced Engineering, 2014.
As cloud computing become more flexible & effective in terms of economy, data owners are motivated to outsource their complex data systems from local sites to commercial public cloud. But for security of data, sensitive data has to be encrypted before outsourcing, which overcomes method of traditional data utilization based on plaintext keyword search. Considering the large number of data users and documents in cloud, it is necessary for the search service to allow multi-keyword query and provide result similarity ranking to meet the effective data retrieval need. Retrieving of all the files having queried keyword will not be affordable in pay as per use cloud paradigm.

Zhihua Xia, Li Chen, Xingming Sun, and Jin Wang "An Efficient and Privacy-Preserving Semantic Multi-Keyword Ranked Search over Encrypted Cloud Data", ISSN: 2287-1233 ASTL-2013.
As so much advantage of cloud computing, more and more data owners centralize their sensitive data into the cloud. In this paper, we propose a semantic multi-keyword ranked search scheme over the encrypted cloud data, which simultaneously meets a set of strict privacy requirements. Firstly, we utilize the "Latent Semantic Analysis" to reveal relationship between terms and documents. The relationship between terms is automatically captured. Secondly, our scheme employ secure "k-nearest neighbor (k-NN)" to achieve secure search functionality. The proposed scheme could return not only the exact matching files, but also the files including the terms latent semantically associated to the query keyword. Finally, the experimental result demonstrates that our method is better than the original MRSE scheme.

Shih-Ting Hsu, Chou-Chen Yang, and Min-Shiang Hwang "A Study of Public Key Encryption with Keyword Search", International Journal of Network Security, Vol.15, No.2, PP.71-79, Mar. 2013.
Public Key Encryption with Keyword Search (PEKS) scheme enables one to search the encrypted data with a keyword without revealing any information. The concept of a PEKS scheme was proposed by Boneh et al. in 2004 and Baek et al. who extended PEKS scheme into a secure channel free PEKS scheme (SCF-PEKS) which removes the assumption, a secure channel between users and a server. In this paper, we show an overview of six existing security models of PEKS/SCF-PEKS scheme and conclude five security requirements that must satisfy to construct a secure PEKS/SCF-PEKS scheme. Then we compare the security and efficiency of the security models and discuss the future researches of PEKS/SCF-PEKS.

IV. Proposed Method

A. Keyword Search Technique

Keyword-based search is a well-studied problem in the world of text documents and Internet search engines. Inverted lists are common data structures used for solving keyword queries. An interesting post search activity is the ranking of results the problem of keyword search over all documents.

B. Attribute Based Encryption

In an ABE system, a user's keys and ciphertexts are labeled with sets of descriptive attributes and a particular key can decrypt a particular ciphertext only if there is a match between the attributes of the ciphertext and the user's key. Using ABE, access policies are expressed based on the attributes of users or data, which enables a user to selectively share the data among a set of users by encrypting the file under a set of attributes, without the need to know a complete list of users. The complexities per encryption, key generation and decryption are only linear with the number of attributes involved. However, to integrate ABE into a large-scale system, important issues such as key management scalability, dynamic policy updates, and efficient on-demand revocation.

C. Multi-Keyword Ranked Search

For their first time, we define and solve 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 choose the efficient similarity measure of "coordinate matching", i.e., as many matches as possible, to capture the relevance of data documents to the search query.

V. Results and Discussion

A. Efficiency Analysis

Index Construction: To build a searchable sub index Ii for each document Fi in the dataset F, the first step is to map the keyword set extracted from the document Fi to a data vector Di, followed by encrypting every data vector. The time cost of mapping or encrypting depends directly on the dimensionality of data vector which is determined by the size of the dictionary, i.e., the number of indexed keywords. And the time cost of building the whole index is also related to the number of sub index which is equal to the number of documents in the dataset. Fig: 2. shows that, given the same dictionary where |W| = 4000, the time cost of building the whole index is nearly linear with the size of dataset since the time cost of building each sub index is fixed. Number of keywords indexed in the dictionary determines the time cost of building a sub index. The major computation to generate a sub-index in MRSE I includes the splitting process and two multiplications of a (n + 2) × (n + 2) matrix and a (n + 2) 2).

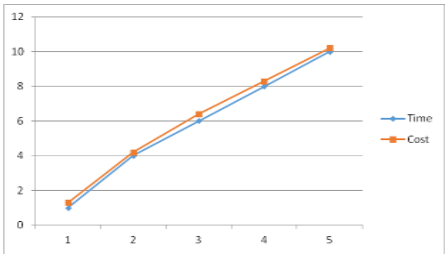

Fig: 2. Time cost of building Index

Trapdoor Generation: Fig: 2. shows that the time to generate a trapdoor is greatly affected by the number of keywords in the dictionary. Like index construction, every trapdoor generation incurs two multiplications of a matrix and a split query vector, where the dimensionality of matrix or query vector is different in two proposed schemes and becomes larger with the increasing size of dictionary. Fig: 3. demonstrates the trapdoor generation cost in the scheme is about 20 percentages larger than that in the scheme. Like the sub index generation, the difference of costs to generate trapdoors is majorly caused by the different dimensionality of vector and matrices in the two schemes. More importantly, it shows that the number of query keywords has little influence on the overhead of trapdoor generation, which is a significant advantage over related works on multi-keyword searchable encryption.

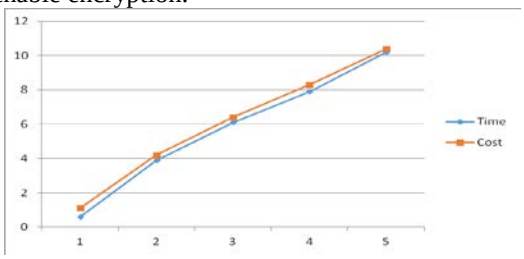

Fig: 3. Time cost of generating trap door

Query: Query execution in the cloud server consists of computing and ranking similarity scores for all documents in the dataset. Fig: 4. shows the query time is dominated by the number of documents in the dataset while the number of keywords in the query has very slight impact on it like the cost of trapdoor generation above.

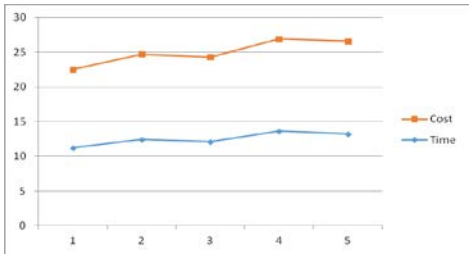

Fig: 4. For different number of query keywords

VI. Conclusion

In this paper, for the first time define and solve the problem of multi-keyword ranked search over encrypted cloud data, and establish a variety of privacy requirements. Among various multi-keyword semantics, 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 propose a basic idea of MRSE using secure inner product computation. Then we give two improved MRSE schemes to achieve various stringent privacy requirements in two different threat models. Thorough analysis investigating privacy and efficiency guarantees of proposed schemes is given, and experiments on the real world dataset show our proposed schemes introduce low overhead on both computation and communication.

References

[1] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data," Proc. IEEE INFOCOM, pp. 829- 837, Apr, 2011.

[2] L.M. Vaquero, L. Rodero-Merino, J. Caceres, and M. Lindner, "A Break in the Clouds: Towards a Cloud Definition," ACM SIGCOMM Comput. Commun. Rev., vol. 39, no. 1, pp. 50-55, 2009.

[3] N. Cao, S. Yu, Z. Yang, W. Lou, and Y. Hou, "LT Codes-Based Secure and Reliable Cloud Storage Service," Proc. IEEE INFOCOM, pp. 693-701, 2012.

[4] S. Kamara and K. Lauter, "Cryptographic Cloud Storage," Proc. 14th Int'l Conf. Financial Cryptograpy and Data Security, Jan. 2010.

[5] A. Singhal, "Modern Information Retrieval: A Brief Overview," IEEE Data Eng. Bull., vol. 24, no. 4, pp. 35- 43, Mar. 2001.

[6] I.H. Witten, A. Moffat, and T.C. Bell, Managing Gigabytes: Compressing and Indexing Documents and Images. Morgan Kaufmann Publishing, May 1999.

[7] D. Song, D. Wagner, and A. Perrig, "Practical Techniques for Searches on Encrypted Data," Proc. IEEE Symp. Security and Privacy, 2000.

[8] E.-J. Goh, "Secure Indexes," Cryptology ePrint Archive, http:// eprint.iacr.org/2003/216. 2003.

[9] Y.-C. Chang and M. Mitzenmacher, "Privacy Preserving Keyword Searches on Remote Encrypted Data," Proc. Third Int'l Conf. Applied Cryptography and Network Security, 2005.

[10] R. Curtmola, J.A. Garay, S. Kamara, and R. Ostrovsky, "Searchable Symmetric Encryption: Improved Definitions and Efficient Constructions," Proc. 13th ACM Conf. Computer and Comm. Security (CCS '06), 2006.

[11] D. Boneh, G.D. Crescenzo, R. Ostrovsky, and G. Persiano, "Public Key Encryption with Keyword Search," Proc. Int'l Conf. Theory and Applications of Cryptographic Techniques (EUROCRYPT), 2004.

[12] M. Bellare, A. Boldyreva, and A. ONeill, "Deterministic and Efficiently Searchable Encryption," Proc. 27th Ann. Int'l Cryptology Conf. Advances in Cryptology (CRYPTO '07), 2007.

[13] Unique Sense: Smart Computing Prototype S. Vijaykumara, , S.G. Saravanakumarb, , M. Balamurugan, Dr 2nd International Symposium on Big Data and Cloud Computing (ISBCC'15)