Survey on Secured Searching Techniques in Encrypted Cloud Data

[1]Bismi M, [2]Sulphikar A

[1]PG Scholar, [2]Associate Professor
[1]Department of Computer Science and Engineering,
[1]LBSITW, Poojappura, Trivandrum, Kerala, India

Abstract— **Cloud Computing can be defined as a new paradigm in which the resources are provided online through the internet. This paradigm shifts the location of infrastructure to the internet and thus reduces the costs associated with hardware and software resources. The cloud provider encrypts the confidential data stored in cloud and keyword search is used for effective file retrieval. Cryptographic techniques are used to solve the security and privacy problems. Multiple keywords are used in the search request for efficiency and co-ordinate matching algorithm is used to find the similarity between data in the cloud. The resulting files can be retrieved according to the total relevance score.**

Index Terms— **Cloud computing, data retrieval, keyword search, searchable encryption**

I. Introduction

Cloud computing is an emerging paradigm which can be defined as a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort(NIST definition)[1]. Cloud computing can be another term for distributed computing over a network which means the ability to run a program on many connected computers at the same time. Cloud follows a 'pay-as-you-use' model. The great flexibility and economy savings are motivating not only individuals but also the enterprises to store their local data in the cloud. Thus the main purpose of cloud system is to maximize the use of shared resources. Figure 1 gives the cloud storage architecture.

According to NIST definition, cloud has got five essential characteristics[2] such as broad network access, rapid elasticity, measured service, on-demand self-service and resource pooling, three service models such as software as a service(SaaS), platform as a service(PaaS), infrastructure as a service(IaaS) and four deployment models such as public, private, hybrid and community cloud.

For providing data privacy, encryption technique is used to encrypt the owner data before uploading to the public cloud. Searchable encryption is the technique used to encrypt the original data in plain text into cipher text using a key. Previous traditional searchable encryption techniques are using single/ Boolean keyword search to find out the results, but with low efficiency. Further, secure ranked keyword search methods are proposed which increases system usability and file retrieval accuracy. But this method supports only single keyword. Considering the large number of data users and documents in cloud, it is necessary to allow multiple keywords in the search request. For more efficiency in data retrieval, we can add an ontology based search along with the multi keywords. Ontology based search will give meaningful words related to the query keyword. Thus, the system could be able to return exactly matching files along with the words meaningfully related to the query keyword.

The rest of the paper is organized as follows. Section 2 explains the basic terms and techniques involved. Section 3 gives the literature survey of existing works along with their pros and cons. Finally, in section 4, we conclude our survey paper with further research directions.

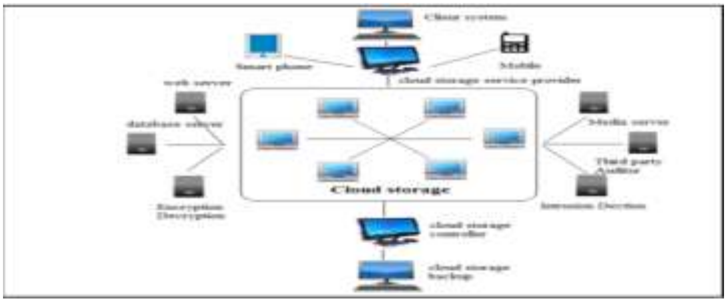

Figure 1: Cloud Storage Architecture

II. TERMS AND TECHNIQUES

This section explains some of the terms, models and techniques used in the existing systems. They are:

A. *Vector Space Model:* In this model [3], text is represented by a vector of terms. If words are chosen as terms, then every word in the vocabulary becomes an independent dimension in a very high dimensional vector space. No term is assigned a negative value. To assign numeric score to a document for a query, the model finds out the similarity between query vector and the document vector. Inner-product (dot product) between two vectors is often used as a similarity measure.

B. *Probabilistic Model:* This model is based on the general principle that documents in a collection should be ranked by decreasing probability of their relevance to a query. This is called probabilistic ranking principle (PRP). The ranking criterion is based on simple Bayes transform.

C. *Inference Network Model:* In this model, a document instantiates a term with certain strength and the credit from multiple terms is accumulated given a query to compute the equivalent of a numeric score for the document.

D. *Term Weighting:* The technique that depends upon the better estimation of various probabilities. Three factors that come into play in final term term weight formulation are:

i. Term Frequency (or tf): Worde that repeat multiple times in a document.

ii. Document Frequency (or df): Words that appear in many documents are considered common. A weighting method based on this is called inverse document frequency (or idf).

iii. Document Length: When collections have documents of varying lengths, longer documents tend to score higher since they contain more words and word repetitions.

E. *Searchable Encryption Algorithm:* An algorithm that consists of the polynomial time randomized algorithms. They are

i. KeyGen(s): s is a security parameter taken and used to generate a key pair either public or private.

ii. PEKS (Apub, w): Apub is a public key and w is a word which is used to produce a searchable encryption.

iii. Trapdoor (Apriv, w): Apriv is a private key and w is a word which is used to produce a trapdoor Tw.

iv. Cipher text Security*:* It is a technique that is used to provide security for the encrypted data. A cipher text attacker could easily break semantic security by reordering the keywords and submitting the resulting cipher text for decryption. A standard technique is used to break this and this technique is called the cipher text security.

v. Private Key Searchable Encryption*:* A model called private key searchable encryption is used to search on a private key encrypted data. The user himself encrypts data, so as to organize in an arbitrary way.

vi. Public Key Searchable Encryption: Public key searchable encryption is a model that allows user to encrypt data and send it to the server. The owner provides decryption key may be different.

III. LITERATURE SURVEY

Traditional data utilization methods are using plain text keyword search. But the large amount of data in cloud is a problem and the main threat in these systems is data security. To ensure user data privacy, sensitive data has to be encrypted before uploading to the public cloud. Search over encrypted data is a basic and common form of data utilization method, which enables users to quickly sort out their required data files from the huge amount of data. For efficient results, we are using multiple keywords in the search request along with ontology based search. Here co-ordinate matching is used for finding the similarity measure between query keyword and documents in the cloud. More the number of query keywords appear in a document, more relevant is the document to query. During index construction, each document and query keywords are associated with binary vectors as sub index, where each bit represents whether corresponding keyword is contained in the document. Thus the similarity can be measured by taking the inner (dot) product of query and document vectors.Some of the existing systems that explains about various searching techniques in encrypted cloud data are discussed below.

S. Kamara and K. Lauter[5] studied the features of new economic and computing model , ie, cloud, which is becoming popular day by day. Public clouds such as Microsoft Azure and Amazon S3 provide scalable and dynamic storage to customers. As the use of cloud services increases, it is necessary to think about the security and privacy. This paper introduces a virtual private storage service based on some cryptographic techniques. The architecture for the model is explained both in consumer and enterprise level. Searchable encryption is used for security and files can be retrieved by using keyword search.Various types of searchable encryption and the provisions for security are also discussed. The results prove that the searching is efficient and secure.

D. Boneh, Di Crescenzo, R. Ostrovsky, G. Persiano[6] introduces the public key encryption technique(PEKS) for searching in cloud data. The cloud server (CS) contains encrypted files and keyword. User creates keyword trapdoor Tw using its private key to search W and the CS checks Tw with existing encrypted keywords and finds out the matching encrypted files. Here the sender makes encryption of files and server makes authentication of user. So a secure channel can be made between the sender, server and user. Initial scheme was only for single keywords. Later a PEKS scheme is proposed for encrypting multiple keywords efficiently.

In the paper "Searchable Symmetric Encryption: Improved Definitions and Efficient Construction", R.Curtmola, Juan Garay, Seny Kamara, R.Ostrovsky[7] explains about searching and retrieving of outsourced data by multi user SSE. This method gives security definitions by non-adaptive setting and adaptive adversary. In former method, the set of corrupted parties is chosen in advance before the interaction begins. In latter method, the adversary chooses whom to corrupt during the course of computation. This method supports multi-user settings and requires no authentication. But this is not suitable for very large cloud data.

The paper "Privacy Preserving Keyword Search on Remote Encrypted Data" by Y.C. Chang and M. Mitzenmacher [8] gives a solution for the issues regarding privacy concerns about the sensitive data outsourced in cloud. Here a keyword index is used for securely searching in remote encrypted data. This scheme prevents the cloud server from learning any possible sensitive plaintext in uploaded databases. The scheme also supports private querying in which neither the database nor the cloud server learns query details. Here no public key crypto system is required and the scheme is independent of encryption method chosen for remote files.

Next Jin Li, Qian Wang, Cong Wang, Ning Cao, Wenjing Lou [9] introduces the fuzzy keyword searching strategy for searching in encrypted cloud data. It enhances system usability by returning exact matching results. If exact match fails, it returns the closest match as result. Here, edit distance is used to quantify keyword similarity. Edit distance is nothing but a way of quantifying how dissimilar two strings (words) are to one another by counting the maximum number of operations required to transform one string to another. Techniques used to obtain fuzzy keyword search are wildcard based and gram based technique. Wild card technique is used to edit the operations at same position. Edit distance is calculated by using substitution, deletion and insertion. In gram based technique, fuzzy set is constructed based on grams. Gram of a string is a substring and can be used for effective approximate search. This scheme is highly efficient and increases searching effectiveness. But it supports only Boolean keywords.

Later C. Wang, N. Cao, J. Li, K.Ren and W.Lou[10][11] introduces the ranked keyword search. The traditional methods for keyword search use single or Boolean queries. Boolean search allows us to combine words and phrases using AND, OR, NOT. Boolean search is used only for searching Boolean queries. This paper introduces a ranking technique for searching in encrypted data. Ranked SSE and Order preserving Symmetric Encryption are used. This ranking avoids network traffic and also the communication and computation overhead will be low. The user retrieval files are ranked according to their relevance score.

The paper "Privacy Preserving Multi Keyword Ranked Search" by Ning Cao, Cong Wang, Kui Ren, Wenjing Lou [12] introduces multiple keywords in the search request and gives a ranked search. The scheme provides efficiency in known cipher text model and background model with low overhead in computation. The technique used for similarity checking is coordinate matching. Inner product similarity is also used to quantitatively evaluate the similarity for ranking files. This scheme overcomes the disadvantages of single or Boolean keyword searches. Both data and keyword privacy are maintained in this scheme.

.

Table 1: Pros & Cons of searching techniques

Year	Proposed work	Advantage	Disadvantage
2004 [6]	Public Key Encryption with Keyword Search	• Use of searchable encryption • Secure channel between sender and user	• Suited for single query only • Needs support for gateway to identify keyword alone
2005 [8]	Privacy Preserving Keyword Search on Remote Encrypted Data	• Use of bloom filter gives extra storage space	• Did not preserve the privacy and correctness of data
2006 [7]	Searchable symmetric encryption: improved definitions and efficient constructions	• Efficient Storage • Access of sparse tables	• Does not provide accurate documents • Multi user SSE
2010 [10][11]	Enabling Secure and Efficient Ranked Search over Outsourced Cloud Data	• Avoids unwanted retrieval and network traffic • Little overhead in index building	• Did not support multiple keywords • Increased search time and cost
2012 [9]	Fuzzy Keyword Search over Encrypted Data in Cloud Computing	• Enhanced system usability • Improved searching effectiveness	• Creation of fuzzy keyword index and exact keyword index is too difficult for large databases

IV. Conclusion

With the advent of cloud computing, more and more sensitive data are stored in the cloud server to reduce management costs. Data is encrypted before uploading to the cloud for security. Encrypted data makes data utilization a very challenging task. In this study, we first provide a brief introduction and some terms and techniques that are being used in our study area. Next we have summarized various searching techniques in encrypted cloud data and also the basics of cryptographic cloud storage. We have understood that the main problems that are to be faced for secured data utilization are data privacy, keyword privacy, index privacy and query privacy. The disadvantages of previous schemes were overcome in the new MRSE scheme. Ranking and use of multiple keywords are also incorporated to improve the efficiency. Coordinate matching and inner product similarity are the techniques used in MRSE scheme. As a future work, an ontology based search can be suggested. It will result more accurate and relevant files according to the query request.

V. Acknowledgment

I am thankful to my guide Mr. Sulphikar A, Associate Professor of Computer Science and Engineering, for his guidance and encouragement for this paper work and also to all the authors of the papers which i refer, to get enough information to write this paper.

References

[1] Mell P, Grance T.The NIST definition of cloud computing. NIST Special Publication. 2011: 800-145.

[2] PENG Yong et al. / Secure cloud storage based on cryptographic techniques/sciencedirect/2012

[3] A.Singhal, “Modern information retrieval: A brief overview,” IEEE Data Engineering Bulletin

[4] 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

[5]] S. Kamara and K. Lauter, “Cryptographic cloud storage,” in RLCPS, January 2010, LNCS.

[6] D. Boneh, G. D. Crescenzo, R. Ostrovsky, and G. Persiano, “Public key encryption with keyword search,” in Proc. of EUROCRYPT, 2004.

[7] R. Curtmola, J. A. Garay, S. Kamara, and R. Ostrovsky, “Searchable symmetric encryption: improved definitions and Efficient constructions,” in Proc. of ACM CCS, 2006.

[8] Y.-C. Chang and M. Mitzenmacher, “Privacy preserving keyword searches on remote encrypted data,” in Proc. of ACNS, 2005.

[9] J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou, “Fuzzy keyword search over encrypted data in cloud computing,” in Proc. of IEEE INFOCOM’10 Mini - Conference, San Diego, CA, USA, March 2012.

[10] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, “Secure ranked keyword search over encrypted cloud data,” in Proc. of ICDCS’10, 2010.

[11] N. Cao, C. Wang, M. Li, W. Lou and K. Ren, “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.

[12] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, “Privacy-preserving multi-keyword ranked search over encrypted cloud Data,” Proceedings of IEEE INFOCOM 2014, pp. 829-837, 2014.