ISSN:0975-9646

A Survey on Searching Techniques over Encrypted Data

Ms. Archana D. Narudkar[#1], Mrs. Aparna A. Junnarkar[*2]

[#1,2] *Department of Computer Engineering*
P.E.S Modern College of Engineering,Pune,India
*

***Abstract*— Exponential growth of internet users throughout the world raises the issue of data storage in the industry. As the answer to this cloud system plays a vital role. Due to lack of internal security at the cloud service provider end, cloud data storage is largely depending on the cryptographic techniques. Although traditional searchable encryption schemes allow users to securely search over encrypted data through keywords using Boolean values. So there is a much urge is arises to study these systems which actually performs searching techniques and to find their weakness. Many searching techniques are existed based on similarity, keyword matching, fuzzy logic and many more ideas as discussed in this paper. But most of them are concentrating on fetching as many as more documents for the given query. These techniques are actually increases the time complexity of the searching techniques even though they provide accurate results.**

***Keywords*— Trapdoor, indexing, cloud service provider , encryption.**

I. Introduction

In the late 1960’s the idea of “Utility computing” that was coined by MIT computer scientist and Turing award winner John McCarthy was preferably known as the concept of cloud computing over a network. Industries were looking for some sort of major solution, since utility computing ended up becoming something of a big business for companies such as IBM. Indeed, Martin Greenberger pointed out the concept that “advanced arithmetical machines of the future” were now being used not only institutionally for scientific calculation and research but also for business functions such as accounting and inventory. Further, he anticipated his piece of work in which computers would be universal almost like the major power companies running wires everywhere in due time.

As the technology enhances, the question was immediately raised whether “Information utility” would become a regulate like the power industry or be a private entity in and of itself. Later on IBM saw the potential for enormous profit to be made in this type of business and took into consideration by providing computing services to companies for top dollar. The technical limitations on bandwidth as well as disk space were a huge constraint on what could have been developed. The paradigm for this type of knowledge was simply not in place to evaluate yet for cloud computation to take into consideration, though the use of mainframe processing still proved to be profitable for quite some time. The companies such as Sun Microsystems began outing the concept in the market that “the network is the computer” successfully. Further, Larry Ellison implemented an idea (who invested in Salesforce.com) that had for terminal machines that would cost less than $300. These ideas were really appreciable accordingly, but they never implemented as consumers were looking for more complete personal computer solutions that can be affordable, like some storage device that are available. Within a changeover in past decade in indexing the internet that had given rise to first Yahoo, and then Google, has shown us how working in to a vast network area of knowledge was likely the ancestor of the interactivity that we can enjoy today with cloud computing. Indexing an internet may be thought as of fun, but these search engines were composing of vast amount of information over network of server present around the world.

Due to a revolutionary change in the field of industries over past decade, there has been increase in demand of outsourcing of data over a wide range of network. In order to manipulate this huge amount of data in cost effective manner enterprise has adapted a prevalent technology called cloud computing that remove the burden of data management. In this data driven environment enterprise tend to store their data onto cloud that compromise of valuable asset of customer data like emails, personal health data etc. Cloud computing is turning out to be most essential paradigm in the development of information technology which offer flexible access , ubiquitous, on demand access and capital expenditure saving .

Despite its technical advantage in business, enterprise should always keep concern of its privacy from the prying eyes over a network. Privacy preserving is one of the major hurdles in cloud for user, especially when the user data that reside in local storage is outsourced and computed onto cloud. The sensitive data that a cloud service provider is holding could be secure by firewalls ,intrusion detection system also CSP has full control over the infrastructure of cloud including lower level of system stack and system hardware. Although mitigate concern are taken still privacy breaches is likely to occur in the paradigm. In few cases the service provider is not fully trusted, but still we need the service. Therefore, some method should be empowered to protect the user data and user queries from unauthorized person in the cloud environment. Thus, before sending data onto the cloud, data must be encrypted to protect from data privacy and unsolicited access.

The working pattern of the cloud for data storage and accessibility can be shown in figure 1.

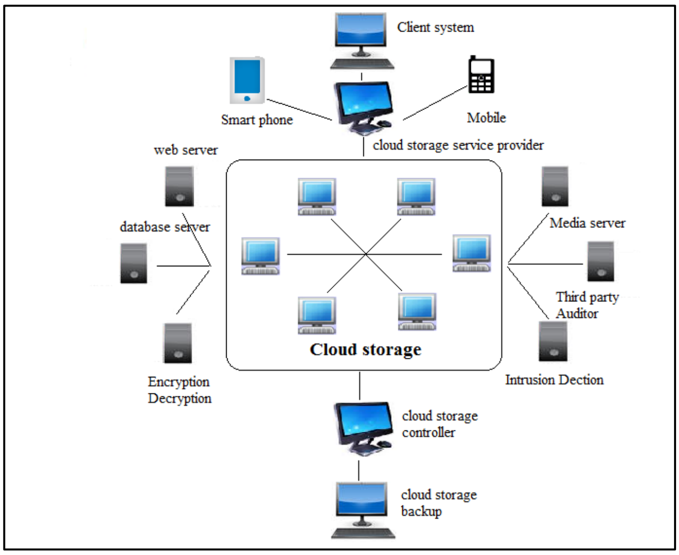

Fig 1: Cloud storage working pattern

However many user believe that encryption of data before outsourcing provide a strong guarantee that the data privacy would be maintain against the cloud service providers. For example, the user may encrypt a email body by using a public key before sending it to the service provider and then send the data to the service provider. Since, public key is known only to the user the service provider could not breach the privacy of email. Though encryption provides privacy, it makes data utilization a challenging task such that it complicates the computation on the data such as the fundamental search operation being carried out on cloud. Without keyword search function the cloud will become a remote storage which provides limited value to all the enterprises that store its data on the cloud. Still, cloud would not provide a efficient search on the encrypted data to approve the benefits of a full-fledged cloud computing environment.

The rest of the paper is organized as follows. Section 2 discusses some related work and section 3 concludes the paper.

II. Related Work

Encryption is an easiest route to keep the privacy of the data. On another side search operation on such data is very challenging task. A number of search techniques had been implemented to perform this one. In spite of this searching data on this cloud is facing some severe problems. Now a days cloud computing is on glance. Normally the size of the standalone system is not enough to store large data, this problem of storage can be easily overcome by the cloud computing. A survey was conducted by David Simms and he found that almost 95% of peoples are rely on cloud for there storage. The biggest advantage offers by the clouds are data outsourcing. Generally third party vendors are there to provide the cloud services which reduce the burden of maintaining cloud as it maintains by that third party only. With this great advantage, if utilization of such data is not proper then there will be no use of these advantages.

As security is one of the main concerns of the cloud data, NIST a cloud community offers some features to protect data from trapdoors. Therefore the main challenge in cloud computing is effective searching. Previously searching is suitable for only unencrypted data which does not offers security over cloud hence encryption of data came into existence. And to search over such encrypted data is challenging task. This paper focuses on the same problem , it gives different keyword search methods for encrypted data over cloud.

Traditionally it has seen that plaintext method cannot be applied directly to search the desire data. Multi - key word ranked search has traditionally been provided by Information Retrieval system (IRS) for data user. To overcome the issue of searching on encrypted data a algorithm has been proposed by [1], which depicted a problem of multi-keyword search over encrypted data using Latent- Semantic -Analysis, which not only return the file including the terms latent semantically associated with keyword query but also return the exact matching file. It uses the vector consisting of TF values that analyse the latent semantic association between terms and documents by LSA .Security and privacy is enforce by using a splitting *k-NN* technique to encrypt the index and the queried vector, so that we can obtain the accurate result. [2] Address, content-based multimedia retrieval over encrypted databases that enable client retrieval directly in the encrypted domain. In order to secure index scheme such as mini-Hash sketches and secure inverted index it uses jointly exploiting technique like cryptography, image processing, and information retrieval. The first schema exploits randomized hash functions and the second schema makes use of inverted indexes of visual words. This model is further enhanced to overcome mini- Hash scheme that require longer sketches to achieve better performance in order to achieve performance similar to that of the inverted index scheme.

To retrieve a document containing only a word [3] describe, a cryptographic model that focuses on problem of searching an encrypted data and provide a secure crypto system. This technique focus on hidden query so that the untrusted server cannot search for a word without the user's authorization and it also support query isolation means, the server learns nothing more than the result. It Gives an approach to search over remotely located data .In this method document retrieval is done in two phases. Here independent public key encryption method can be choose, also it is suitable for different file formats which add more weightage to this method. But it requires additional storage overhead and will not guarantees the security of the data.

Quin Liuy, Guojun Wangyz, and Jie Wuz [4] say if CSP is used on searching then system will face the problem of security. One solution on this is to use cryptographic approach that will manage the authorized users which have access to such data. The authors try to preserve the privacy of the data, keyword and the semantics of the data. All technique is based on the public key encryption. It makes use of CSP for the decryption purpose.Unlike another methods it discussed an SPKS approach where cloud service provider is actually gets involved. The role of the cloud service provider is to do decipher text of cipher text. By doing this computational cost of user for the decryption of encrypted data can be reduce significantly. Here decryption is done by the cloud service provider but he really not aware about the keyword and data in which keyword is to be searched. To do experimental evaluation system performance is compared with the Boneh eta al system. It makes use of CSP for the decryption purpose. Ming li [5] states the authorized private keyword search techniques that offers the privacy of the data, query also a multidimensional keyword searching operations. Advanced to this he proposed APKS+. The first method enhances the efficiency of searching algorithms and the second method preserves the privacy of the query. Here the problem of authorised searching of keyword is discussed. In this paper two approaches cryptographic primitive and Hierarchical Predicate Encryption are proposed. This system makes use of personal health records for the purpose of testing the system. The above techniques will get fails if entity synonyms or morphological variants are used.

On contrast to this Cong [6] states search operations on encrypted data will increase the cost of processing and traffic of network. Cong Wang proposed a new searching theory that reduces the processing overhead that generally obstacles the search system. Author uses build index along and the keyword frequency based relevance score. It implements a secure ranked based keyword search method. In this method order preserving mapping scheme is used where small encrypted files are processed first then large encrypted files are processed. Authors present a theory for retrieving documents that makes use of Ranked Searchable Symmetric Encryption, Order Preserving Symmetric Encryption and One Many Order Preserving Mapping. This method is used to achieve a great accuracy and security. Also it avoids unwanted retrieval and traffic problem. But system fails if multiple keywords are fed as input, with such input searching speed also increases. The above stated methods are based on exact query matching but it did not implement the similarity matching .

Boneh et al. [7] proposed a Very first PKC based search scheme when he inspired from the identity based encryption. This scheme is initially well suited for single query only. By using this scheme anyone with public can write to cloud but the user having a private key can only allow performing searching operations on cloud. By taking base of this technique, a number of methodologies have been implemented to filter the searching techniques.

One of the great author Li [8] proposed another predictive encryption technique which is based on the hierarchical encryption. This technique build authorized keyword search technique over cloud. Like another techniques this technique also gives search access to the authorized users and non-authorized users will not get access to search. In spite of effectiveness of these schemes it has biggest drawback that it is computationally expensive.

In [9] authors try to solve the problem of searching over encrypted data in cloud. It makes use of confidentiality preserving rank order technique. This method forms the framework by using secure index, encrypted domain search and ranked retrieval for the extraction of data from the cloud. Depending on encrypted queries it ranks the documents and document having most rank will be pushed up using ranked method. The given method is well suited for large documents and also It provides higher accuracy and security. But for this method computational cost is high and protecting communication link is bit difficult task.

Dan Boneh et.al. [10] Presents an attribute based encryption approach with prediction encryption scheme. The drawbacks of one technique are easily overcome by another scheme. Since it makes use of two different schemes it is highly secure and faster. Since on cloud the data is located form remote locations, hence it is challenging task to access and retrieve data form such remotely located information. Here Smith generate a one key for the email gateway and by using this key email gateway get access to check “urgent” keyword in email without reading the complete email. By doing so the desirable work of both parties can be done and the privacy of the system also will not get compromised. Here identity based encryption is used for the purpose of the encryption. The disadvantages of the above systems are: 1.refreshing keywords, 2.secure channel removal, 3.multiple keyword processing. [11] Makes use of PIR queries for searching over cloud. This method uses bloom filter gives storage space which can be useful to store some extra information. It hides the identity of the communication also keeps the semantic of the encrypted data. But it will not preserve the privacy and correctness of the data.

Mehmet Kuzu [12] introduces a method of locality sensitive hashing which is a high dimensional space searching technique .which uses a hashing technique to create trap door for searching encrypted documents in the cloud. As the hashing technique is one way it can not reverse engineer to recheck the outcomes and also this method takes little while to search the document due to granular hashing process.

Ming Li et al. [13] Gives a novel approach for privacy preserving searching paradigm. The searching approaches are used to search keyword over cloud data that are outsourced from the third parties. Here two types of Searching approaches are used: 1.Ranked over keyword search and 2. Search over structured data. The reason behind using of these techniques is their popularity in the field of information retrieval in plain domain. In above technique confidentiality is obtained buy it failed to give variability once searching on encrypted data is done.

Jianfeng Wang et al. [14] Presents an important approach that not only guaranty the confidentiality and security but also the verifiability of the searching method. Verifiability refers to cross check condition which usually done to be safe on our side. A symbol tree based searching is done encrypted data to achieve the goal. The reason to use fuzzy logic for the proposed approach is same as said above. Here for the very first time fuzzy logic based keyword search technique for encrypted data is proposed. A great security of data can be achieved while preserving the privacy of the data. An experiment shows that for each query system will give the constant cost of complexity. Also the computational cost is reduced from the O(L) to O(1) where L is a length of query. Author makes use of Linux machine with Pentium dual core processor for the purpose of experiments.

Wenhai Sun et al. [15] elaborates another similarity based ranking approach for the said purpose. To get done with the task frequency of each word is find out. And in advance a vector space model with cosine similarity is used to increase the efficiency of the searching algorithm. Vector space model is used to support conjunctive and disjunctive searching also. A tree based index searching with multidimensional algorithm is used. The reason behind using this is to speed up the process as this speed is one of the main drawbacks in linear search. Further to increase the privacy of the searching algorithms two more indexes known as cipher text model and background model is used. So finally author concludes that efficiency of searching and preciseness of searching can be well balanced by the approach.

Suppose Sam wants to send email to the Smith which is encrypted by the public key of Smith. This email is sent through the email gateway. Email gateway wants to check whether email contains “urgent” keyword or not. If it contains “urgent” keyword then first priority is given to that message but at the same time Smith don’t want to decrypt complete message by the email gateway.

While searching query on given domain the nature of query plays an important role. The more accurate the query the more precise will be the output.

III Conclusion

The above studied systems are been proposed to conduct the search over the encrypted data that are been stored on the cloud side. And in majority of the systems the accumulated problem with them is, they are taking more time to search the data. Authors like Jianfeng Wang et al. depicts a method of searching data using tree based method with fuzzy logic where system is emphasis to reduce the cost. Boneh et al. shows method of key based searching technique. Whereas Y.-C. Chang, uses PIR query with bloom filters to search over the encrypted data .So all these methods are never talking about reducing the time complexity of the system. A very few finger counting systems are been proposed to decrease the time of searching but they are also not successfully implemented this on huge amount of data.

So a considerable research is required to minimize the searching time over the encrypted data in cloud.

References

[1] Li Chen,Xingming Sun,Zhihua Xia ,Qi Liu,” An Efficient and Privacy-Preserving Semantic Multi-Keyword Ranked Search over Encrypted Cloud Data”in International Journal of Security and Its Application in 2014 .

[2] Wenjun Lu, Ashwin Swaminathan, Avinash L. Varna, and Min Wu, “Enabling search over encrypted multimedia databases,” in proc.of SPIE Media Forensicsand Security,09,2009.

[3] D. X. Song, D. Wagner and A. Perrig,” Practical Techniques for Searches on Encrypted Data”., Proceedings of the 2000 IEEE Symposium on Security and Privacy, pp. 44-55.

[4] Qin Liuy, Guojun Wangyz, and Jie Wuz,”Secure and privacy preserving keyword searching for cloud storage services”, ELSEVIER Journal of Network and computer Applications, March 2011

[5] Ming Li et al.,” Authorized Private Keyword Search over Encrypted Data in Cloud Computing, IEEE proc. international conference on distributed computing systems, June 2011, pages 383-392

[6] Cong Wang et al.,”Enabling Secure and Efficient Ranked Keyword Search over Outsourced Cloud Data”, IEEE Transactions on parallel and distributed systems, vol. 23, no. 8, August 2012

[7] Boneh, D., Di Crescenzo, G., Ostrovsky, R., Persiano, G.: Public key encryption with keyword search. In: Advances in Cryptology-Eurocrypt 2004, pp. 506–522. Springer (2004)

[8] Li, M., Yu, S., Cao, N., Lou, W.: Authorized private keyword search over encrypted data in cloud computing. In: Distributed Computing Systems (ICDCS), 2011 31st International Conference on, pp. 383–392. IEEE (2011)

[9] Swaminathan A, Mao Y, Su G-M, Gou H, Varna AL, He S, M. Wu, Oard D “Confidentiality Preserving Rank-Order search” Storage security and survivability 2007-Conference Papers (Proceedings of the 2007 ACM workshop on Storage security and survivability) pp.7-12.

[10] Dan Boneh, Giovanni Di Crescenzo, Rafail Ostrovsky, Giuseppe Persiano “Public Key Encryption with Keyword Search” Springer Berlin Heidelberg; EUROCRYP’04, Volume: 3027 of LNCS- 2004 pp.506-522

[11] Y.-C. Chang, M. Mitzenmacher, “Privacy preserving keyword searches on remote encrypted data” conference 2005.

[12] Mehmet Kuzu, Mohammad Saiful Islam, Murat Kantarcioglu “Efficient Similarity Search over Encrypted Data “, IEEE trasaction ,2012.

[13] Ming Li et al., ”Toward Privacy-Assured and Searchable Cloud Data Storage Services”, IEEE Transactions on Network, volume 27, Issue 4, July/August 2013

[14] Jianfeng Wang et al., “Efficient Verifiable Fuzzy Keyword Search over Encrypted Data in Cloud Computing” , Journal of Computer Science and Information system, volume 10, Issue 2, April 2013

[15] Wenhai Sun et al., "Privacy-Preserving Multi- keyword Text Search in the Cloud Supporting Similarity-based Ranking", the 8th ACM Symposium on Information,Computer and Communications Security , Hangzhou, China, May 2013.