ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
Volume 1, Issue 8, October 2012

SEARCHING TECHNIQUES IN ENCRYPTED CLOUD DATA

Deepa P L, S Vinoth Kumar, Dr S Karthik

Abstract— **Cloud computing can be defined as a new style of computing in which the resources are provided online through the internet. It provides storage as well as service. It uses the technique of virtualization. Virtualization provides the abstraction of data. Large amount of data can store in the cloud. Cloud provider encrypts the sensitive data and stores it in the cloud so that only the authenticated users can access the data. Thus the keyword privacy is maintained. Searching is very difficult in encrypted data. In this paper we focus on different searching techniques and at the end a better solution is identified.**

Index Terms— **Cloud computing, Confidential Data, Ranked Keyword Search, Searchable Encryption**

I.INTRODUCTION

Cloud computing is the process of accessing the services without knowing the exact location of the data. Cloud computing is 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 a 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.

The first computing paradigm developed was mainframe computing. Then comes the pc computing, network computing, internet computing and grid computing. Grid computing is similar to the cloud computing. The primary difference is that cloud is utility computing and grid is not. Cloud uses the concept of virtualization which is not used by grid. Cloud is a centralized system approach while grid is distributed system. Resource discovery in grid system is accomplished by offline. The resource allocation is not time based in cloud. Grid computing is based on time based resource allocation. Level of customer satisfaction is important in cloud computing.

II. CLOUD COMPUTING ADVANTAGES

A.Reduced software costs

With the technology of cloud computing you can avoid purchasing of software applications and thereby reduce the cost. Some sites provide free services. Software licensing is not necessary.

B.Lower computer costs

There is no need for a computer with high processing capability. The cloud applications can run on a simple computer having less memory, smaller hard disk etc.

C.Unlimited storage capacity

Cloud computing doesn’t have any storage constraints. Large amount of data can stored on the cloud.

D. *Universal document access*

In cloud, the user can access the data from anywhere at any time. Integrity of the data is also preserved

E.Increased data reliability

Cloud computing provides a safe computing technique since it prevents the destruction of data even if the computer crashes.

III. CLOUD COMPUTING DISADVANTAGES

IV.Security issues

Potential risk on cloud is security of the data stored. Cloud has less control over access the information [11]. The user has no knowledge about the location of data. Unauthorized accessing of data can be takes place in private cloud. The better solution is to store the sensitive data in an encrypted format so that only the authorized user can access the data [8].

V.Constant Internet connection is needed

The cloud services can use only with a constant internet connection.

IV. UNITS

Cloud computing offers three services.

A. *Software as a service*

Using this service, the cloud users can access resource as well as applications from the cloud. Thus user can avoid the cost of purchasing the software. So there is no need to install and run the software on the user’s system. The user is

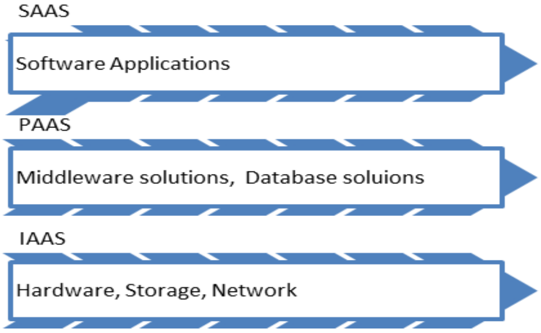
SAAS
Software Applications
PAAS
Middleware solutions, Database soluions
IAAS
Hardware, Storage, Network

Fig.1.Cloud computing service models

Manuscript received Sep 15 , 2012.

Deepa P L, *Software Engineering, Anna University/SNS College of Technology, Coimbatore-641035, Tamil Nadu, (e-mail: deepadnair1988@gmail.com, India, 8220840338*

S VinothKumar, *Computer Science & Engineering, Anna University/SNS College of Technology, Coimbatore-641035, Tamil Nadu, India (e-mail: vinothmepco@gmail.com).*

S Karthik, *Computer Science & Engineering, Anna University/SNS College of Technology, Coimbatore-641035, Tamil Nadu, India (e-mail:profskarthik@gmail.com).*

able to use the application provided by the cloud, but not able to control the operating system which it is provided.

B. Platform as a service

The cloud provider provides a computing platform for the users. This service provides access to operating system and associated service. The developer will develop the application and run on the cloud platform. Thus the consumer can use the environment for their applications.

C. Infrastructure as a service

This is the most basic service provided by the cloud. The storage, network components can access by the user using infrastructure as a service. The service providers supply the resources depending upon the demand.

V. CLOUD STORAGE INFRASTRUCTURE

The cloud storage is an abstraction behind an interface. The storage can access on demand. Cloud provides geographically distributed storage capacity. Cloud storage architecture provides the delivery of service on demand in a scalable and multitenant way [3, 4].

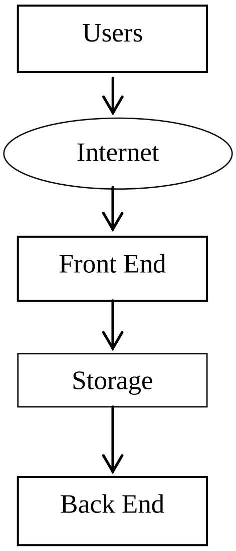

Fig.2.Cloud Storage Architecture

VI.CHARACTERISTICS

A.Manageability

It is the ability to manage a system with minimal resources. The user can minimize the cost by storing data in the cloud. The cost includes the cost of the physical storage as well as managing it.

B.Access Method

The difference between cloud storage and the normal storage lies in the way by which it is accessed. Different providers implement different access methods.

C.Multitenancy

This is one of the major characteristics of cloud storage architecture. Different users can use the single storage. Multitenancy is the essential attribute of cloud computing which helps to share a server by different customers.

D.Scalability

This is the ability to scale to meet higher demands or load in a graceful manner.

E.Availability

Whenever the data is requested by the user, the cloud storage provider has to provide the data.

F.Efficiency

The efficient use of available resources is important. The sensitive data can be stored in an encrypted format thereby it become available only to authenticated users.

VII.DEPLOYMENT MODELS OF CLOUD

A.Public Cloud

Public cloud computing services can be provided by the third party providers. Through the public cloud, the general public can access the storage, applications, services. Google is a public cloud [5].

B.Private Cloud

Private Cloud is cloud infrastructure developed for a single organization. Private cloud require high development costs.

C.Hybrid Cloud

This is a combination of two or more clouds. Load sharing is done in hybrid cloud. This is a mixture of public, private and community cloud.

D.Community Cloud

This cloud is shared among two or more organizations that have similar cloud requirements. Facebook is a community cloud.

VIII.SEARCHING IN CLOUD DATA

Data can be stored both as public as well as private. Different searching strategies are available for both types of data. The confidential data are stored in the cloud using encryption technique [1, 9]. So only the authenticated members who know the key can access the data. Accessing data from encrypted storage is very difficult. A different type of searching technique is used to search for encrypted data [10].

IX.EXISTING SOLUTIONS FOR KEYWORD SEARCH

A. FUZZY KEYWORD SEARCH

In this paper, the problem of effective fuzzy keyword search over the encrypted data is solved. Fuzzy keyword search enhances the system usability by returning exact matching results. If the exact match fails, it returns the closest match as the result. Edit distance is used to quantify the keyword similarity [14].

Main modules in Fuzzy keyword search are

1. Wildcard-based technique
   A wildcard is used to edit the operations at the same position. The edit distance can be calculated using substitution, deletion and insertion [16].
2. Gram-based technique
   Here the fuzzy set is constructed based on grams. The gram of a string is a substring and can be used for effective approximate search. The order of the

characters after the primitive operation is always kept the same before the operations.

3. Symbol-based trie-traversed scheme

   In this technique, a multi-way tree is constructed for storing the fuzzy keyword set over a finite symbol set

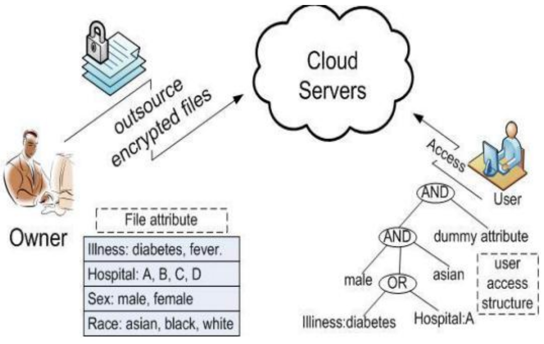

Fig.3.Symbol-based trie-traversed scheme

All the trapdoors sharing a common prefix have common nodes. The fuzzy keyword in the trie can be found by depth first search approach [14].

Advantages:

- Maintaining keyword privacy
- Effective utilization of remotely stored encrypted data

B. IMPROVING EFFICIENCY OF SEARCH USING CONJUNCTION OF KEYWORD

In this technique, conjunction of keywords is implemented for searching. The conjunctive keyword search mechanism will retrieve most efficient and relevant data files. The conjunctive keyword search automatically generates ranked results so that the searching flexibility and efficiency will be improved [15].

This technique uses the wildcard based method and gram based method for constructing fuzzy keyword sets and symbol based trie- traverse scheme for generating a multi way tree to store the fuzzy keyword sets generated. This reduces the storage overhead. It also uses the Edit distance concept to quantify the keyword similarity.

Advantages:

- Conjunction/sequence of keywords automatically generates a ranking mechanism.
- This retrieves a highly relevant search result.

Disadvantages:

- The integrity of rank order in search result is not checked.

C. VERIFIABLE FUZZY KEYWORD SEARCH

The user generates a symbol based index tree with encrypted documents and outsources it in the cloud server. When the search request is received by the server, the server maps the searching request to a set of documents. Each document is assigned an identifier and a set of keywords [6].

After searching, the server retrieves the search request and the proof for the result to the user. Using the proof, the user can verify the correctness and completeness of the result.

The searching is done based upon some rules.

(a) The searching input exactly matches the preset keyword.

(b) If there exist format inconsistencies in the searching input, it will return the closest possible matches available.

Search privacy as well as the document privacy is ensured. The document privacy is ensured by the encryption algorithm [12].

Advantages:

- More efficient for real application.
- Provide security and verifiability.

Disadvantages:

- Multi keyword query is not supported.

D. EMBEDDING EDIT DISTANCE TO ENABLE PRIVATE KEYWORD SEARCH

Here the private identification scheme is combined with classical embedding of edit distance into the hamming distance. This is to obtain a fuzzy keyword search for the edit distance. This method does not need to a priori define the set of words which are considered as acceptable for the search. It increases security in this model [7].

Managing the nearest neighbour search in encrypted domain is the principle of a private identification scheme. This technique associates a message into a set of keywords and to consider each keyword as a virtual address. Receiver can recover link toward the associated messages. Information retrieval enables to retrieve a block from the database without knowledge about the query and answer.

Advantages:

- When looking for close keywords, this technique enables flexibility on the tolerated edit distance.
- This technique preserves the confidentiality of the keyword given by the user to search.

E. INDEX MANAGEMENT SCHEME

A searchable re-encryption scheme is introduced. Using this technique, user can share the data with others safely by generating searchable encryption index and re-encrypting it.

The security requirements are set up and it uses two techniques-Proxy re-encryption function and searchable encryption function. These methods provide efficiency. The search method uses multiple keywords and thus the flexibility is provided [13].

Advantages:

- Efficiency is provided in terms of calculation volume.
- Traffic efficiency is provided by using only one round of communication process for keyword search and encryption.
- This method provides quick search speed.
- Regardless of time usage, the encrypted data saved in unreliable distant data server can be shared safely and efficiently.

Disadvantages:

- Index composed of multiple keywords with variable length is not possible.

F. AUTHENTICATING RESULTS OF RANKED KEYWORD SEARCH OVER CLOUD DATA

Developing a private cloud is very expensive. This paper provides a provision to store the sensitive data in the public cloud. The sensitive data is stored in an encrypted format to avoid unauthorized access. Enabling search in the encrypted data is very difficult. A ranked search is implemented inorder to implement the top k retrieval.

Term frequency (TF) is the number of times a particular keyword appears within the file. Inverse document frequency (IDF) is calculated by dividing total number of files by the number of files that contain the particular keyword. Ranking function is calculated using TF*IDF rule [2].

One to many order preserving mapping is used to avoid information leakage. In this technique, the random plaintext-to-bucket mapping of Order Preserving Symmetric Encryption (OPSE) is used but also incorporates the unique file id together with the plaintext. The results generated are authenticated using a one way hash chain technique. One way hash function is an algorithm that turns messages or text into a fixed string of digits, mainly for the security purpose[2].

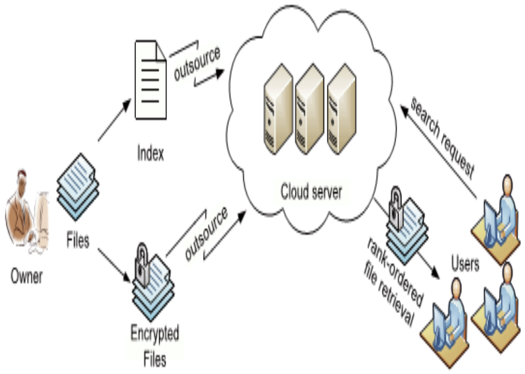

Fig.4.Architecture for search over encrypted cloud data

One to many order preserving mapping is used to avoid information leakage. In this technique, the random plaintext-to-bucket mapping of Order Preserving Symmetric Encryption (OPSE) is used but also incorporates the unique file id together with the plaintext. The results generated are authenticated using a one way hash chain technique. One way hash function is an algorithm that turns messages or text into a fixed string of digits, mainly for the security purpose.

The papers discussed above provide storage for the sensitive data in the cloud. Every technique provides the data to be store in the encrypted format. In this paper the searching in the encrypted data is done by using ranking function and the retrieved results are authenticated.

Advantages:

- Displays the most relevant results
- The order of results retrieved is maintained
- Hash chain is a light weight technique

X. CONCLUSIONS

Above six papers are referred for this survey. In the beginning, cloud computing basis and its storage infrastructure is explained. After that different searching techniques in the cloud data are discussed. Each has its own advantages and disadvantages. Inorder to overcome the disadvantages, ranked keyword searching scheme with authentication of results is implemented.

ACKNOWLEDGMENT

The authors would like to thank the Editor-in-Chief, the Associate Editor and anonymous Referees for their comments.

REFERENCES

1) B. Zhang and F. Zhang, “An efficient public key encryption with conjunctive- subset keywords search”, Journal of Network and Computer Application s, vol. 34, no. 1, (2011)

2) Cong Wang,Ning Cao,Kui Ren and Wenjing Lou,”Enabling Secure and Efficient Ranked Keyword Search over Outsourced Cloud Data”,*Proc.IEEE Transactions on parallel and distributed system,Aug2012*

3) C. Wang, K. Ren, S. Yu, K. Mahendra, and R. Urs, “Achieving Usable and Privacy-Assured Similarity Search over Outsourced Cloud Data,”Proc. *IEEE INFOCOM*,2012..

4) D. Boneh, G.D.Crescenzo, R. Ostrovsky, and G.Persiano,“Public key encryption with keyword search,” in Proc. of EUROCRYP’04, 2004. (2002) The IEEE website. [Online]. Available: http://www.ieee.org/

5) *E. Shi, J. Bethencourt, H. Chan, D. Song, and A. Perrig, “Multi-Dimensional Range Query over Encrypted Data,”Proc. IEEE Symp. Security and Privacy,2007.*

6) Jianfeng Wang,Xiaofeng Chen,Hua Ma,Qiang Tang and Jin Li, “A Verifiable Fuzzy Keyword Search Scheme Over Encrypted Data“,*Journal of Internet Services and Information Security (JISIS), volume: 2, number: 1/2, pp. 49-58*

7) Julien Bringer and Hervé Chabanne,” Embedding edit distance to enable private keyword search”*Springeropen journal 2012*

8) K. Ren, C. Wang, and Q. Wang, “Security Challenges for the Public Cloud,”IEEE Internet Computing, vol. 16, no. 1, pp. 69-73, 2012.

9) M.Belare, A.Boldyreva, and A.O’Neil, “Deterministic and efficiently searchable encryption,” in Proceedings of rypto 2007, volume 4622 of LNCS. Springer- Verlag, 2007.

10) N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, “Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data,”Proc. IEEE INFOCOM ’11,2011

11) Q. Wang, C. Wang, K. Ren, W. Lou, and J. Li, “Enabling PublicVerifiability and Data Dynamics for Storage Security in

12) CloudComputing,"IEEE Trans. Parallel and Distributed Systems, vol. 22,no. 5, pp. 847-859, May 2011

13) S. Ji, G. Li, C. Li, and J. Feng. Efficient interactive fuzzy keyword search. In Proc. of 18th International World Wide Web Conference(WWW'09), Madrid, Spain. ACM, April 2009

14) Sun-Ho Lee and Im-Yeong Lee,"Secure Index Management Scheme on Cloud Storage Environment" ,*International Journal of Security and Its Applications Vol. 6, No. 3, July, 2012*

15) T.Balamuralikrishna,C.Anuradhaand N.Raghavendrasai, "Fuzzy keyword search over encrypted data over cloud computing",Asian Journal of Computer Science and Information Technology 2011

16) T. M Nisha and V. P Lijo ,"Improving the Efficiency of Data Retrieval in Secure Cloud by Introducing Conjunction of Keywords", *Proceedings published in International Journal of Computer Applications (IJCA)25*.

Deepa P L received B-tech degree in Computer Science and Engineering from Toc H institute of Technology, Cochin University, Kerala, India in 2010. She is presently doing ME Software Engineering in SNS College of Technology, Anna University, Chennai. Her area of interest includes Cloud Computing and software Engineering.

S. Vinoth Kumar received B.E degree in Computer Science and Engineering from Anna University, Chennai, India in 2006,M.E degree in Computer Science and Engineering from Anna University, Tirunelveli, India in 2009. He is at present working as Assistant Professor in the department of Computer Science Engineering at SNS College of Technology, Coimbatore, India. His area of interest include Image Processing, Network Security and Mobile Computing.

Professor Dr.S.Karthik is presently Professor & Dean in the Department of Computer Science & Engineering, SNS College of Technology, affiliated to Anna University- Coimbatore, Tamilnadu, India. He received the M.E degree from the Anna University Chennai and Ph.D degree from Ann University of Technology, Coimbatore. His research interests include network security, web services and wireless systems. In particular, he is currently working in a research group developing new Internet security architectures and active defense systems against DDoS attacks. Dr.S.Karthik published more than 35 papers in refereed international journals and 25 papers in conferences and has been involved many international conferences as Technical Chair and tutorial presenter. He is an active member of IEEE, ISTE, IAENG, IACSIT and Indian Computer Society.