International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2017 IJSRCSEIT | Volume 2 | Issue 5 | ISSN : 2456-3307

Review of Various Data Storage and Retrieval Method for Cloud Computing

Ajeet Mishra[1], Prof. Umesh Kumar Lilhore[2], Prof. Nitesh Gupta[3]
M. Tech. Research Scholar[1], Head PG[2], Assistant Professor[3]
NRI Institute of Information Science & Technology Bhopal, Madhya Pradesh, India, India

ABSTRACT

Cloud computing is a widely used computing technology by IT world. It is a fast-growing technique, which serves computing resources such as IaaS, PaaS and SaaS to cloud user on "Pay per use" basis. A Cloud user can store their private data over cloud server and can access securely at any time. The User doesn't have to worry about storage and maintenance of cloud data. This unique data accessibility feature of cloud attract user to utilize cloud services. Due to the high availability of various IT computing resource over cloud, attracts cloud user to utilize its services. Day by day size of data and services are getting increases over the cloud. It is quite challenging job for cloud service provider to maintain the data integrity and privacy of the stored user data. Another challenge is encounter during retrieval of encrypted stored data. Various cryptography methods are used to maintain the data privacy and integrity. Overcloud server data are stored in encrypted form. After placing the data on the cloud, retrieving the same is also a quite tedious job. In order to retrieve the data, several methods are available suggested by various researchers. Most of the existing techniques are limited to handle a single keyword search with its own limitation. To enhance searching in terms of efficiency and fastness, a multi-keyword search technique can be adapted to retrieve a corresponding document from the cloud. This paper proposes a survey on a secure search scheme supporting single-keyword or multi-keyword ranked search over encrypted cloud data.

Keywords : Cloud computing, data retrieval, Single keyword, Multi keyword and Racked Search.

I. INTRODUCTION

Cloud Computing is an on-demand service over network servers which are hosted on the Internet to process, store and organize the data, rather than a local server or personal computer. The Cloud services and applications run on the distributed network which provides a virtual resource for the end user. These resources could be accessed by standard Internet and networking protocols. Data integrity verification is one of the massive responsibilities with cloud data because the probability of involvement in the malicious activity of a cloud user and cloud provider is very high [11, 15]. There are many ways to address this problem. The user can use encryption and decryption process. However, it requires huge computing time and functional overheads. Applying data auditing may be the other way to address this problem. Even if cloud provides such amazing services to its clients, there are some problems related to cloud such as security of data stored in cloud and integrity of data. The data security can be guaranteed using encryption technique before sending data to a cloud server and integrity of data can be guaranteed by signing data blocks using users signature such that, except the user, no one can be able to generate similar signature [8].

Ranked search can also elegantly eliminate unnecessary network traffic by sending back only the most relevant data, which is highly desirable in the "pay-as-you-use" cloud paradigm. For privacy protection, such ranking operation, however, should not leak any keyword related information. On the other hand, to improve the search result accuracy as well as to enhance the user searching experience, it is also necessary for such ranking system to support multiple keywords search; a single keyword search often yields far too coarse results. Even with this provision, there is the possibility of leakage of data, as the integrity of data is verified by third-party auditor thus the data needs to be copied

from cloud server to third-party auditor and problem starts. This increases the complexity of auditing process [2, 4]. In this paper, we are presenting a review of various data storage and retrieval method for cloud computing. This paper is organized in various chapters which include the introduction of cloud computing, existing work, data retrieval methods and conclusions and future scope.

II. CLOUD COMPUTING

The term cloud computing is referred as "The Cloud" which used as a "Metaphor" for the "internet" so cloud computing is a "type of Internet-based computing". It is a new and an innovative idea of 21st Century for IT industries. The purpose of the cloud computing dynamically delivers the computing resources and capabilities as services over the web. It is a new technology of computing in which dynamically scalable and often virtualized resources are provided as a service over the internet [2, 4]. Cloud computing involved different hosted services over the internet which is IaaS (Infrastructure-as-a-service), PaaS (Platform-as-a-service), and SaaS (Software-as-a-service).IaaS service provides the infrastructure like memory, space, storage etc to users. Saas service provides different application rather installing to its own system. PaaS service provides cloud application to the developer and responsible for virtualization of resources and makes it as a single layer.

2.1 Types of Cloud-Cloud deployment models are-

a) **Public cloud-** A public cloud is one standard of cloud computing, in which a cloud service provider provides a virtual environment, make a pool shared resources, such as applications and storage, offered to the general public over the web. Public cloud services are offered to users as pay-per-usage.
b) **Private cloud-** Organizations choose to build their private cloud as to keep the strategic, operation and other reasons to themselves and they feel more secure to do it.
c) **Hybrid model-** It consists of multiple service providers. It provides the services of both public and private cloud. It is used by organizations when they need both private and public clouds both.

2.2 Cloud Computing Services- Users can access these services in a pay per use on-demand model. Users can access these services over the internet. Cloud computing provides following services.

a) **IaaS-** In Infrastructure as a service, the service provider shares infrastructure resources to support operations done by the end user. The examples for infrastructure are CPU, memory, network, server etc. Software as a Service is the capability provided to the consumer is to use the provider's applications running on a cloud infrastructure.
b) **PaaS-**This service is used to complete the life cycle of building and delivering web applications, which are available over the internet. The developers of this service are concerned only with web-based development and do not care about the operating system. Anyone who has the internet can do this and deploy over the internet. Some examples of PaaS model are eBay, Google, iTunes, YouTube.
c) **SaaS-** The capability provided to the consumer is to use the provider's applications running on a cloud infrastructure. The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, storage, or even individual application capabilities, with the possible exception of limited user-specific application configuration settings.

2.3 Challenges in Cloud computing-Cloud computing have following issues-

a) **Data protection-** To be considered protected, data from one customer must be properly segregated from that of another; it must be stored securely when "at rest" and it must be able to move securely from one location to another. Cloud providers have systems in place to prevent data leaks or access by third parties. Proper separation of duties should ensure that auditing or monitoring cannot be defeated, even by privileged users at the cloud provider.

b) **Authentication-** The authentication of the respondent device or devices like IP spoofing, RIP attacks, ARP poisoning (spoofing), and DNS poisoning are all too common on the Internet. TCP/IP has some "unfixable flaws" such as "trusted machine" status of machines that have been in contact with each other, and tacit assumption that routing tables on routers will not be maliciously altered. One way to avoid IP spoofing by using encrypted protocols wherever

possible. They also suggest avoiding ARP poisoning by requiring root access to change ARP tables; using static, rather than dynamic ARP tables; or at least make sure changes to the ARP tables are logged.

c) **Data Verification**- Things like tampering, loss, and theft, while on a local machine, while in transit, while at rest at the unknown third-party device, or devices, and during remote backups. Resource isolation ensures the security of data during processing, by isolating the processor caches in virtual machines, and isolating those virtual caches from the Hypervisor cache.

d) **Infected Application**- Vendor should have the complete access to the server for monitoring and maintenance, thus preventing any malicious user from uploading any infected application onto the cloud which will severely affect the customer. Cloud providers ensure that applications available as a service via the cloud are secure by implementing testing and acceptance procedures for outsourced or packaged application code. It also requires application security measures (application-level firewalls) be in place in the production environment

e) **Availability**- Cloud providers assure customers that they will have regular and predictable access to their data and applications.

f) Data Retrieval- Efficiently accessing of cloud data is always desirable and challenging.

III. Existing Work in Cloud Data Retrieval

Existing systems are based on following data retrieval schemes -

3.1 Boolean Keyword Search-Boolean systems allowed customers to specify their information need using a combination of Boolean operators AND, OR and NOT. Boolean systems have several disadvantages, for example, there are no any features of document ranking, and it is very difficult for a customer to make a good search request. Thus, the drawback of existing system specifies the important need for new techniques that support searching flexibility.

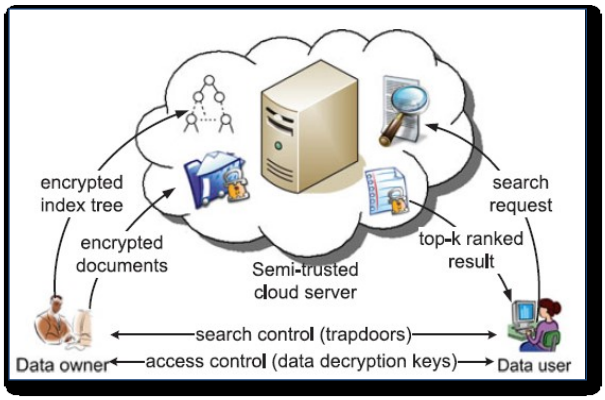

Figure 3. Cloud Data Retrieval and Indexing

3.2 Single Keyword Based Searchable Encryption-In single keyword searchable encryption schemes encrypted searchable indexes are uses and its contents hidden to the server unless it's given appropriate trapdoors generated via secret key(s)[3]. In the public key setting, anyone with a public key can write to the data stored on server but only authorized users with the private key can search. Public key solutions are usually very computationally expensive, however. Furthermore, the keyword privacy could not be protected in the public key setting since server could encrypt any keyword with public key and then use the received trapdoor to evaluate this cipher text[6].

3.3 Searchable Encryption- It allows users to securely search complete encrypted data through keywords. This method supports only Boolean search, without capturing any relevant data [2]. This approach suffers from two main drawbacks when directly applied in the context of Cloud Computing. First one, users who do not necessarily have pre-knowledge of the encrypted cloud data, have to post-process every file got, in order, to find ones most matching their interest; another drawback, regularly getting all files containing the queried keyword further incurs unnecessary network traffic, when retrieving more than one files.

3.4 MKSE or Multi Keyword Searchable Encryption- In cloud computing to search functionalities, conjunctive keyword search over encrypted data had been proposed. These schemes incur large overhead caused by their fundamental primitives, such as computation cost by bilinear map [5].

3.5 TOP-K data retrieval method- In order to efficiently solve data search problem, existing self-indexing algorithms are not sufficient [14]. Two

approaches to enhance the retrieval of self-index have been proposed. The first uses a document-based array, which uses to map in between every suffix in set T to its corresponding document identifier. The second is to store, in addition to the global self-index, one self-index of each individual document in the collection.

3.6 Ranked Keyword Search- Ranked search greatly enhances system usability by returning the matching files in a ranked order regarding certain relevance criteria (eg. keyword frequency) thus; making one step closer toward practical deployment of privacy-preserving data hosting services in the context of cloud computing [6]. To the best of knowledge, it gives a legal status for the first time the problem of effective ranked keyword search over encrypted cloud data. Ranked keyword search strongly provides system usability by returning the matching files in ranked order concerning to certain relevance criteria, thus moving closer towards the practical action of privacy-preserving data presenting services in the cloud.

3.7 Fuzzy Keyword Searchable Encryption: Fuzzy keyword search greatly enhances system usability by returning the matching files when users searching inputs exactly match the predefined keywords or the closest possible matching files based on keyword similarity semantics when an exact match fails [5]. More specifically, it uses edit distance to quantify keywords similarity and develop a novel technique that is a wildcard-based technique, for the construction of fuzzy keyword sets. This technique eliminates the need for enumerating all the fuzzy keywords and the resulted size of the fuzzy keyword sets is significantly reduced [9].

3.8 Plaintext Fuzzy Keyword Search- The importance of fuzzy search has received attention in the context of plaintext searching in information retrieval community [12]. The problem is addressed in the traditional information access paradigm by allowing users to search without using try-and-see approach for finding relevant information based on approximate string matching. At the first glance, it seems possible for one to directly apply these string matching algorithms to the context of searchable encryption by computing the trapdoors on a character base within an alphabet. This trivial construction suffers from the dictionary and statistics attacks and fails to achieve the search privacy.

IV. Challenges in Existing System

In Cloud computing, efficient data retrieval for encrypted data is always challenging. Based on literature survey following challenges were found in existing work

1) High data retrieval time
2) Privacy of data
3) Data indexing and ranking
4) Integrity of Data

V. Conclusions and Future Work

Cloud computing is a widely used technology. This paper proposed a brief literature survey on efficient data retrieval techniques for encrypted cloud data. From the survey, the multi-keyword search technique sounds to be more efficient than other available searching technique. Many search schemes over encrypted data, supports multi-keyword query and similarity ranking simultaneously for data retrieval in cloud computing. In the future work, we will develop a more efficient data retrieval method and will compare with various existing methods.

VI. References

[1]. Chi Chen, Xiaojie Zhu, Peisong Shen, J.Hu, S.Guo, Z.Tari, and Albert Y. Zomaya, Fellow, "An Efficient Privacy-Preserving Ranked Keyword Search Method", IEEE Transactions on Parallel and Distributed Systems, year 2015 PP 1-15.

[2]. V.Saiharitha, S.J.Saritha "A Privacy and dynamic Multi-keyword Ranked Search Scheme over Cloud Data Encrypted", IEEE Year 2016, PP 131-135

[3]. Zhihua Xia, Member, Xingming Sun, and Qian Wang," A Secure and Dynamic Multi-Keyword Ranked Search Scheme over Encrypted Cloud Data", IEEE transactions on parallel and distributed systems, vol. 27, no. 2, February 2016, PP 340-353.

[4]. Aashi Qul Huq A, Bhaggiaraj S,"Improving Privacy Multi-Keyword Top-K Retrieval Search Over Encrypted Cloud Data", In International Journal Of Engineering And Computer Science ISSN:2319-7242 Volume 4 Issue 4 April 2015, Page No. 11385-11390.

[5]. Cong Wang, Qian Wang, Kui Ren, Member, Ning Cao, and Wenjing Lou," Towards Secure and Dependable Storage Services in Cloud Computing", In proc. the 17th IEEE International Workshop on Quality of Service (IWQoS'09) IEEE 2009, Page No. 999-1013.

[6]. Cheng-Kang Chu, Sherman S. M. Chow, Wen-Guey Tzeng, Jianying Zhou, and Robert H. Deng," Key-Aggregate Cryptosystem for Scalable Data Sharing in Cloud Storage", In IEEE Transactions on Parallel and Distributed Systems. Volume: 25, Issue: 2. Year: 2014, Page No. 521-533.

[7]. D. Pratiba, Dr. G.Shobha and Vijaya Lakshmi.P.S," Efficient data retrieval from cloud storage using data mining technique", International Journal on Cybernetics & Informatics (IJCI) Vol. 4, No. 2, April 2015, Page No.271-280.

[8]. Ahmed Shawish and Maria Salama," Cloud Computing: Paradigms and Technologies", In Proc. Inter cooperative Collective Intelligence: Techniques and Applications, Studies in Computational Intelligence 495, Springer-Verlag Berlin Heidelberg 2014, Page No. 642-671.

[9]. Danan Thilakanathan, Shiping Chen, Surya Nepal and Rafael A. Calvo," Secure Data Sharing in the Cloud", In Proc. Security, Privacy, and Trust in Cloud Systems, DOI: 10.1007/978-3-642-38586-5_2, Springer-Verlag Berlin Heidelberg 2014, Page No. 888-921.

[10]. Keiko Hashizume, David G Rosado and Eduardo Fernández," An analysis of security issues for cloud computing", In proc. Journal of Internet Services and Applications, Springer 2013, Page No.81-84.

[11]. Nelson Gonzalez, Charles Miers, Fernando Red´ıgolo, Marcos Simpl´ıcio, Tereza Carvalho, Mats N¨ Naslund and Makan Pourzandi," A quantitative analysis of current security concerns and solutions for cloud computing", In proc. Journal of Cloud Computing: Advances, Systems, and Applications , Springer 2012, Page No. 178-196.

[12]. Cong Wang, Ning Cao, Kui Ren, Wenjing Lou," Enabling Secure and Efficient Ranked Keyword Search over Outsourced Cloud Data", In proc. 30th International Conference on Distributed Computing Systems (ICDCS'10) IEEE, Page No. 98-112.

[13]. J. Shane Cul pepper, Matthias Petri and Falk Scholer," Efficient In-Memory Top-k Document Retrieval", In ACM 2012, 978-1-4503, Page No.1472-1482.

[14]. Revathy B.D, Chetan S.P," Enabling Secure and Efficient Multi Keyword Ranked Search over Encrypted Cloud Data", International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE) ISSN: 0976-1353, Volume 13 Issue 1 –March 2015, Page No.88-92.

[15]. Zhihua Xia, Li Chen, Xingming Sun, and Jin Wang," An Efficient and Privacy-Preserving Semantic Multi-Keyword Ranked Search over Encrypted Cloud Data", In Proc. Advanced Science and Technology Letters , Vol.31 (MulGraB 2013), Page No.284-289.

[16]. Li Chen, Xingming Sun, Zhihua Xia and Qi Liu," An Efficient and Privacy-Preserving Semantic Multi-Keyword Ranked Search over Encrypted Cloud Data", In Proc. International Journal of Security and Its Applications, Vol.8, No.2 (2014), Page No. 323-332.

[17]. R. Sharmila," Secure retrieval of files using homomorphic encryption for cloud computing", In proc. IJRET: International Journal of Research in Engineering and Technology 2014, Volume: 03 Special Issue: 07, Page No. 845-849.

[18]. Revathy B.D, Anbumani .A, Rohith .V," Enabling Secure and Efficient Multi Keyword Ranked Search over Encrypted Cloud Data", In proc. International Journal of Science, Engineering and Technology Research (IJSETR), Volume 4, Issue 2, February 2015, ISSN: 2278 – 7798, Page no. 389-394.