International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
www.ijcsjournal.com Volume 4, Issue 1, No 1, 2016. ISSN: 2348-6600
Reference ID: IJCS-110 PAGE NO: 636-641.

Cloud Data Storage Security – A Review Issues & Challenges

Jayant Kumar[#1], Asst. Prof. Nitin Agarwal[*2]

[#]*Department of Computer Science & Engineering,*
NRI Institute of Information Science & Technology
Bhopal, India

[1]jntkumar82@gmail.com
[2]seonitin79@gmail.com

Abstract— **Here in this paper a complete survey of all the techniques used for the security of cloud data storage is analyzed and discussed. Since various techniques are already implemented for the security and privacy of cloud data in terms of attacks, computational cost and privacy preservation. So here a complete set of survey and review is done on these techniques so that a new and efficient technique is implemented in future.**

***Index Terms*—Cloud Computing, Cloud Security, Data Storage, Multi Keywords.**

I. INTRODUCTION

Cloud Computing means a remote server that access through the internet which helps in business applications and functionality along with the convention of system software for respective web application. Cloud computing concept saves capital that cloud users pay out on annual or monthly payment. Due to advantage of cloud services, more and more sensitive information are being centralized into the cloud servers, such as confidential videos and photos, various emails, personal health records information, corporation business data, government documents, etc. So as to privacy problem, data privacy [1] and data loss will be increase in certain circumstances. When users outsource their private onto cloud, the cloud service provider able to monitor the communication between the users and cloud at will trust or untrusted. As cloud computing is promising development in computing concept the confidence increase becomes very important aspect. There are mainly two parameters which can help to get better the confidence on the cloud services. One is to improve efficiency and another for improving security. To improve the efficiency the keyword search method is enhanced as it makes available two way communications between cloud server and the cloud customer. But while deploying security the burden on cloud server gets increased unexpectedly. Consequently it is extremely significant to maintain these two factors so that to improve overall efficiency of the cloud services [2]. Also the world is of mobile devices, so everyone wants to use cloud services on their mobile devices and if the computational cost goes to elevated then it effects into important resource utilization, which is not appropriate for mobile devices. So current scenario is having need of a proficient method is cloud services in the expectations.

Cloud is a service which can be accessed from everywhere if arranged in that way at any path. It causes lots of parties or persons using it for their purpose. In such case the data various parties may contribute to within them on the cloud server can be secret. In addition every cloud user who uses cloud services doesn't like to get followed. In such cases it is very important to maintain their privacy [3]. Thus to maintain their privacy the files and even the search requests are encrypted as soon as the request is sent to the server. This encryption may also affect the efficiency of searching techniques as the search should go on in encrypted manner.
Besides, in cloud computing data owners may allocate their outsourced data with a number of cloud users, who strength want to only get back the data files they are paying attention in cloud server. One of the most fashionable ways to do so is throughout keyword-based retrieval. It is like better to get the retrieval outcome with the most significant files that match users' interest instead of all the files, which indicates that the files should be ranked in the order of relevance by users' interest and only the files with the highest relevance's are sent back to cloud users. To develop security exclusive of give up effectiveness, methods here in [4], [5], give you an idea about that they sustain top-k single keyword retrieval under different circumstances. To protect data privacy, confidential data has to be encrypted before outsourcing, so as to provide end-to-end data confidentiality assurance in the

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
www.ijcsjournal.com Volume 4, Issue 1, No 1, 2016. ISSN: 2348-6600
Reference ID: IJCS-110 PAGE NO: 636-641.

cloud. Clouds enable customers to remotely store and access their data by lowering the cost of hardware ownership while providing robust and fast services [6]. The importance and necessity of privacy preserving search techniques are even more pronounced in the cloud applications. Due to the fact that large companies that operate the public clouds like Google or Amazon may access the sensitive data and search patterns, hiding the query and the retrieved data has great importance in ensuring the privacy and security of those using cloud services. We aim to achieve an efficient system where any authorized user can perform a search on a remote database with multiple keywords, not including exposing neither the keywords he/she searches for, nor the pleased of the documents he/she get backs. The main confront of cloud storage is guaranteeing have power over, and the essential integrity and confidentiality of all stored cloud data.

II. THEORETICAL BACKGROUND

Basically, a cloud storage system can be considered to be a network of distributed data centers which typically uses cloud computing technologies like virtualization, and offers some kind of interface for storing cloud data. To enhance the ease of use of the cloud data, it possibly unnecessarily stored at dissimilar positions. On the whole, all of this is not able to be seen to the cloud user. This cloud storage system requires secure methods of preserving important data in order to prevent unrecoverable data loss, whilst constantly keeping up with increasing demands for storage space. It is necessary to regularly make extra copies of the information, so as to be able to restore it to an earlier version if need be. These copies of cloud additional raise the demand for cloud storage space on cloud server. Additional requirements arise from the variety of devices used to access the data simultaneously. Private and business users demand an easy way to synchronize and access their data independent of both device and location. The software providing these features must also be tailored to the needs of the individual with no technical background.

With the intention of get together these requires; companies make huge assets into their IT infrastructure on cloud storage. Additional hardware and software is required, as well as state for its operation and maintenance. Larger company's capacity has to think about building a contributed data center. These expenses conflict with the continuing need to reduce costs in order to stay competitive.

Cloud storage services offer user-friendly, easily accessible and money-saving ways of storing and automatically backing up arbitrary data. These services are available on-demand on the Internet. A customer simply accesses the website of a cloud storage provider and rents storage space as necessary by selecting one of the provider's packages. A precondition for using this service is Internet access from the customer's computers or mobile devices. Depending on the amount of data to be transmitted to the cloud, sufficient bandwidth must be available; otherwise the transfer could be very time-consuming.

Usage of a cloud storage provider basically means entrusting data to a third party where no prior relationship based on trust has been established. Individuals who upload personal information to the cloud want to be sure that only certain people are able to access it. This should also exclude the provider, since there is no justifiable reason for it to access the data.

Companies may entrust files containing sensitive business data and valuable intellectual property which may be of great interest for industrial espionage. The unauthorized disclosure of customer information, business secrets or research data poses a serious threat to a company's business. In addition, compliance requirements with both internal security guidelines and legal regulations have to be met. The cloud storage provider may be subject to different legal regulations than the user.

In Cloud Computing, data owners may share their outsourced data with a large number of users, who might want to only retrieve certain specific data files they are interested in during a given session. One of the most popular ways to do so is through keyword-based search. Such keyword search technique allows users to selectively retrieve files of interest and has been widely applied in plaintext search scenarios. Unfortunately, data encryption, which restricts user's ability to perform keyword search and further demands the protection of keyword privacy, makes the traditional plaintext search methods fail for encrypted cloud data. Although traditional searchable encryption schemes allow a user to securely search over encrypted data through keywords without first decrypting it, these techniques support only conventional Boolean keyword search, without capturing any relevance of the files in the search result. Ranked search improves system usability by normal matching files in a ranked order regarding to certain relevance criteria (e.g., keyword frequency), as directly outsourcing relevance scores will drips a lot of sensitive information against the keyword privacy.

An easy way to comply with the conference paper formatting requirements is to use this document as a template and simply type your text into it.

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
www.ijcsjournal.com Volume 4, Issue 1, No 1, 2016. ISSN: 2348-6600
Reference ID: IJCS-110 PAGE NO: 636-641.

III. LITERATURE SURVEY

In this paper [8], we introduce the concepts of similarity relevance and scheme robustness to formulate the privacy issue in searchable encryption schemes, and then solve the insecurity problem by proposing a two-round searchable encryption (TRSE) scheme. Novel technologies in the cryptography community and information retrieval (IR) community are employed, including homomorphic encryption and vector space model. In the proposed scheme, the majority of computing work is done on the cloud while the user takes part in ranking, which guarantees top-k multi-keyword retrieval over encrypted cloud data with high security and practical efficiency. Our contributions can be summarized as follows:

We propose the concepts of similarity relevance and scheme robustness. We, thus, perform the first attempt to formulate the privacy issue in searchable encryption, and we show server-side ranking based on order-preserving encryption (OPE) inevitably violates data privacy.

We propose a TRSE scheme, which fulfills the secure multi-keyword top-k retrieval over encrypted cloud data. Specifically, for the first time, we employ relevance score to support multi-keyword top-k retrieval.

Thorough analysis on security demonstrates the proposed scheme guarantees high data privacy. Furthermore, performance analysis and experimental results show that our scheme is efficient for practical utilization. By security analysis, we show that the proposed scheme guarantees data privacy. According to the efficiency evaluation of the proposed scheme over a real dataset, extensive experimental results demonstrate that our scheme ensures practical efficiency.

Ranked search greatly enhances system usability by returning the matching files in a ranked order regarding to certain relevance criteria (e.g., keyword frequency), thus making one step closer towards practical deployment of privacy-preserving data hosting services in the context of Cloud Computing. To achieve our design goals on both system security and usability, we propose to bring together the advance of both crypto and IR community to design the ranked searchable symmetric encryption scheme, in the spirit of "as-strong-as-possible" security guarantee. Specifically, we explore the statistical measure approach from IR and text-mining to embed weight information (i.e. relevance score) of each file during the establishment of searchable index before outsourcing the encrypted file collection.

In this paper [9], for the first time we define and solve the problem of effective yet secure ranked keyword search over encrypted cloud data. Ranked search greatly enhances system usability by returning the matching files in a ranked order regarding to certain relevance criteria (e.g., keyword frequency), thus making one step closer towards practical deployment of privacy-preserving data hosting services in Cloud Computing. We first give a straightforward yet ideal construction of ranked keyword search under the state-of-the-art searchable symmetric encryption (SSE) security definition, and demonstrate its inefficiency. To achieve more practical performance, we then propose a definition for ranked searchable symmetric encryption, and give an efficient design by properly utilizing the existing cryptographic primitive, order-preserving symmetric encryption (OPSE). Thorough analysis shows that our proposed solution enjoys "as-strong-as-possible" security guarantee compared to previous SSE schemes, while correctly realizing the goal of ranked keyword search. Extensive experimental results demonstrate the efficiency of the proposed solution.

In this paper [10], we propose a secure cloud storage system supporting privacy-preserving public auditing. Our work is among the first few ones to support privacy-preserving public auditing in cloud computing, with a focus on data storage. Besides, with the prevalence of cloud computing, a foreseeable increase of auditing tasks from different users may be delegated to TPA. We utilize the homomorphic linear authenticator and random masking to guarantee that the TPA would not learn any knowledge about the data content stored on the cloud server during the efficient auditing process, which not only eliminates the burden of cloud user from the tedious and possibly expensive auditing task, but also alleviates the users' fear of their outsourced data leakage. Considering TPA may concurrently handle multiple audit sessions from different users for their outsourced data files, we further extend our privacy-preserving public auditing protocol into a multiuser setting, where the TPA can perform multiple auditing tasks in a batch manner for better efficiency. Extensive analysis shows that our schemes are provably secure and highly efficient. Our preliminary experiment conducted on Amazon EC2 instance further demonstrates the fast performance of our design on both the cloud and the auditor side. We leave the full-fledged implementation of the mechanism on commercial public cloud as an important future extension, which is expected to robustly cope with very large scale data and thus encourage users to adopt cloud storage services more confidently.

In this paper [11], we present such a system—Cloud Capacity Manager (CCM)—an on-demand compute capacity management system and its methods for dynamically

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
www.ijcsjournal.com Volume 4, Issue 1, No 1, 2016. ISSN: 2348-6600
Reference ID: IJCS-110 PAGE NO: 636-641.

multiplexing the compute capacity of virtualized data centers at scales of that combines various low-overhead techniques, motivated by practical on-field observations, to achieve scalable capacity allocation for thousands of machines. CCM achieves this scale by employing three-level hierarchical management architecture. CCM also sheds light on the tradeoffs due to two unavoidable issues in large-scale commodity data centers: 1) maintaining low operational overhead, given variable cost of performing management operations necessary to allocate resources, and 2) coping with the increased incidences of these operations' failures. The capacity managers at each level continuously monitor and aggregate black-box VM CPU and memory usage information, and then use this aggregated data to make independent and localized capacity allocation decisions. An experimental evaluation on a fairly large infrastructure, that to achieve better capacity multiplexing, the focus needs to not only be on the accurate prediction of workload demand and aggressive optimization of the allocation algorithms, but also on dealing with the practical limitations of real-life infrastructures.

The advent of cloud computing, data owners are motivated to outsource their complex data management systems from local sites to commercial public cloud for great flexibility and economic savings. But for protecting data privacy, sensitive data has to be encrypted before outsourcing, which obsoletes traditional data utilization based on plaintext keyword search. Thus, enabling an encrypted cloud data search service is of paramount importance. Considering the large number of data users and documents in cloud, it is crucial for the search service to allow multi-keyword query and provide result similarity ranking to meet the effective data retrieval need. Related works on searchable encryption focus on single keyword search or Boolean keyword search, and rarely differentiate the search results. In this paper [12], for the first time, we define and solve the challenging problem of privacy-preserving multi-keyword ranked search over encrypted cloud data (MRSE), and establish a set of strict privacy requirements for such a secure cloud data utilization system to become a reality. Among various multi-keyword semantics, we choose the efficient principle of "coordinate matching", i.e., as many matches as possible, to capture the similarity between search query and data documents, and further use "inner product similarity" to quantitatively formalize such principle for similarity measurement. We first propose a basic MRSE scheme using secure inner product computation, and then significantly improve it to meet different privacy requirements in two levels of threat models. Thorough analysis investigating privacy and efficiency guarantees of proposed schemes is given, and experiments on the real-world dataset further show proposed schemes indeed introduce low overhead on computation and communication.

This paper [13] aims to provide searching a file over cloud environment using multiple keywords representing the file with various probable situations. The aim is to provide the security to its maximum extent by including encryption and decryption methods. Authorization of the users directly by the administrators allows the files involved to transfer more securely. Encryption and decryption of both file name and file which uses asymmetric and symmetric key algorithms respectively. The secret key is generated for each user to prevent any other user to misuse the file. The data that are stored in the cloud has to be protected completely from any attack that is caused both by external and internal attackers. Most of the internal attacks are used by the cloud providers by using similarity relevance and analysing the statistical leakage. Based on the usage of the file over ranked manner, it is easy to get all the details of the most used files through probability prediction. This kind of data leakage should be completely avoided and maximum protection to the data is given. The solution suggests the same by applying some new concepts to increase the data security.

IV. Benefits of a Cryptographic Storage Service

The core properties of a cryptographic storage service are that (1) control of the data is maintained by the customer and (2) the security properties are derived from cryptography, as opposed to legal mechanisms, physical security or access control. Therefore, such a service provides several compelling advantages over other storage services based on public cloud infrastructures. In this section, we recall some of the main concerns with cloud computing as outlined in the Cloud Security Alliance's recent report [14] and highlight how these concerns can be mitigated by such an architecture.

A. *Regulatory compliance:* Most countries have laws in place that make organizations responsible for the protection of the data that is entrusted to them. This is particularly so for the case of personally identifiable information, medical records and financial records. And since organizations are often held responsible for the actions of their contractors the use of a public cloud storage service can involve significant legal risks. In a cryptographic storage service, the data is encrypted on-premise by the data processor(s). This way, customers can be assured that the confidentiality of their data is preserved irrespective of the actions of the

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
www.ijcsjournal.com Volume 4, Issue 1, No 1, 2016. ISSN: 2348-6600
Reference ID: IJCS-110 PAGE NO: 636-641.

cloud storage provider. This greatly reduces any legal exposure for both the customer and the provider.

B. *Geographic restrictions:* Data that is stored in certain legal jurisdictions may be subject to regulations even if it was not collected there. Because it can be difficult to ascertain exactly where one's data is being stored once it is sent to the cloud (i.e., many service providers have data centers deployed throughout the world) some customers may be reluctant to use a public cloud for fear of increasing their legal exposure. In a cryptographic storage service data is only stored in encrypted form so any law that pertains stored data has little to no effect on the customer. This reduces legal exposure for the customer and allows the cloud storage provider to make optimal use of its storage infrastructure, thereby reducing costs.

C. *Subpoenas:* If an organization becomes the subject of an investigation, law enforcement agencies may request access to its data. If the data is stored in a public cloud, the request may be made to the cloud provider and the latter could even be prevented from notifying the customer. This can have severe consequences for customers. First, it preempts the customer from challenging the request. Second, it can lead to law enforcement having access to data from clients that are not under investigation (Wired 2009). Such a scenario can occur due to the fact that service providers often store multiple customer's data on the same disks. In a cryptographic storage service, since data is stored in encrypted form and since the customer retains possession of all the keys, any request for the data must be made directly to the customer.

D. *Security breaches:* Even if a cloud storage provider implements strong security practices there is always the possibility of a security breach. If this occurs the customer may be legally responsible. In a cryptographic storage service data in encrypted and data integrity can be verified at any time. Therefore, a security breach poses little to no risk for the customer.

E. *Electronic discovery:* Digital information plays an important role in legal proceedings and often organizations are required to preserve and produce records for litigation. Organizations with high levels of litigation may need to keep a copy of large amounts of data on-premise in order to assure its integrity. This can obviously negate the benefits of using a cloud storage service. Since, with a cryptographic storage service, a customer can verify the integrity of its data at any point in time (e.g., every hour) a provider has every incentive to preserve its integrity.

F. *Data retention and destruction:* In many cases a customer may be responsible for the retention and destruction of data it has collected. If this data is stored in the cloud, however, it can be difficult for a customer to ascertain the integrity of the data or to verify whether it was properly discarded. A cryptographic storage service alleviates these concerns since data integrity can be verified and since the information necessary to decrypt data (i.e., the master key) is kept on-premise. Secure data erasure can be effectively achieved by just erasing the master key.

REFERENCES

[1] Cloud Security Alliance, "Top Threats to Cloud Computing," http://www.cloudsecurityalliance.org, 2010.

[2] Kui Ren, Cong Wang and Qian Wang, "Toward Secure and Effective Data Utilization in Public Cloud", IEEE Network, November/December 2012.

[3] Cong Wang, Sherman S.M. Chow, Qian Wang, Kui Ren, and Wenjing Lou, " Privacy-Preserving Public Auditing for Secure Cloud Storage", IEEE Transactions on Computers, Vol. 62, No. 2, February 2013.

[4] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-Preserving Multikeyword Ranked Search over Encrypted Cloud Data," Proc.IEEE INFOCOM, 2011.

[5] H. Hu, J. Xu, C. Ren, and B. Choi, "Processing Private Queries over Untrusted Data Cloud through Privacy Homomorphism," Proc. IEEE 27th Int'l Conf. Data Eng. (ICDE), 2011.

[6] L. M. Vaquero, L. Rodero-Merino, J. Caceres, and M. Lindner. A break in the clouds: towards a cloud definition. SIGCOMM Comput. Commun. Rev., 39:50{55, December 2008.

[7] Everaldo Aguiar, Yihua Zhang, and Marina Blanton, "An Overview of Issues and Recent Developments in Cloud Computing and Storage Security" 2012.

[8] Jiadi Yu, Peng Lu, Yanmin Zhu, Guangtao Xue and Minglu Li, "Toward Secure Multi-keyword Top-k Retrieval over Encrypted Cloud Data" IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, VOL. 10, NO. 4, JULY/AUGUST 2013.

[9] Cong Wang, NingCao, JinLi, KuiRen, and Wenjing Lou, "Secure Ranked Keyword Search over Encrypted Cloud Data" 2011.

[10] Cong Wang, Sherman S.M. Chow, Qian Wang, Kui Ren, and Wenjing Lou, "Privacy-Preserving Public Auditing for Secure Cloud Storage" IEEE TRANSACTIONS ON COMPUTERS, VOL. 62, NO. 2, FEBRUARY 2013.

[11] Mukil Kesavan, Irfan Ahmad, Orran Krieger, Ravi Soundararajan, "Practical Compute Capacity Management for Virtualized Data Centers" IEEE TRANSACTIONS ON

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
www.ijcsjournal.com Volume 4, Issue 1, No 1, 2016. ISSN: 2348-6600
Reference ID: IJCS-110 PAGE NO: 636-641.

CLOUD COMPUTING, VOL. 1, NO. 1, JANUARY-JUNE 2013.

[12] Ning Caoy, Cong Wangz, Ming Li, Kui Renz and Wenjing Lou, "Privacy-Preserving Multi-keyword Ranked Search over Encrypted Cloud Data" 2010.

[13] Hussain Abo Surrah, "Multi Keyword Retrieval On Secured Cloud" Asian Journal of Technology & Management Research ISSN: 2249 –0892 Vol. 04 – Issue: 01 Jan - Jun 2014.

[14] Cloud Security Alliance. «Security Guidance for Critical Areas of Focus in Cloud Computing.» April 2009. http://www.cloudsecurityalliance.org/guidance/csaguide.pdf.