Analysis on Privacy Preserving and Data Security for Cloud Data Storage

Dr. Amit Kr. Chaturvedi[#1], Meetendra Singh Chahar[*2], Dr. Kalpana Sharma[#3]
[#]*Assistant Prof.,MCA Deptt., Govt. Engineering College, Ajmer, Rajasthan, India*
**Ph.D. Scholar, CSE Deptt., Bhagwant Univ., Ajmer*

Abstract

Cloud computing provides massive computation power and storage capacity, which enable users to deploy applications without infrastructure investment. Another advantage of using the cloud services are the scalability, accessibility, and maintainability from anywhere, anytime. Many services existing on the cloud servers because of economic benefits and operational convenience. So, Privacy-preserving security solutions for cloud services are always a point of discussion and need to secure the personal sensitive data of the individuals or institutions. In this paper, we analyze current privacy preserving solutions for cloud services and summarize in the section 4 - Important Findings. The various solutions proposed includes application of encryption on the data, anonymous authentication for registered users, quasi-identifier index based approach for privacy preservation over incremental data sets, non-bilinear group signature scheme, etc.

Keywords – *Privacy , data security, cloud data storage, anonymizing user identification.*

I. INTRODUCTION

As computing in public based data management system changed and shifted towards cloud computing. Most of the organizations working on the public data use cloud data storage and management system. When we shift the personal data to the cloud storage, the question of privacy preserving is raised first. If this private data wrongly interpreted or used, it will be serious issue and in the present many cases happened on such personal data saved on a public data storage system. Many researchers even discussed , presented surveys and proposed systems for handling and privacy preserving of such important public data. Here in this paper we are going to present a survey on this issue and discuss the already proposed systems.

As the awareness in the society increases and people are more accessing the internet based services, individuals are paying now more attention to the issue of privacy in cloud computing. Top database vendors are adding cloud support for their database and so more data is moving into the cloud. Privacy concerns will continue to grow, because these databases often contain sensitive and personal information related to companies and/or individuals. There are various issues of privacy and security of your information and documents, reliability of the remote servers, vendor choice and reliability, and portability of data in the cloud computing. There are some safeguards also in the cloud computing and should be used to prevent unauthorized access, disclosure of information, copying, use of modification of personal information.

II. PRIVACY AND DATA SECURITY ISSUES

There are some privacy and data security issues that are discussed below:

(i) Disclosure of sensitive private information
(ii) Lack of access control enforcement
(iii) Dynamic exchange of data
(iv) Cloud data storage security

Individuals, Institutions, or the companies have private and sensitive information and if such information are stored and operated with the cloud based service providers (CSPs). There are the chances of leakage of the sensitive and private information that may include personally identifiable information, Usage data, Unique device identities and so on.

In cloud computing, people getting inappropriate or unauthorized access to personal data in the cloud by taking advantage of certain vulnerabilities. This is the Lack of access control enforcement and it creates some security holes. Adversaries take the advantage of such security hole and fetch personally identifiable information of individuals or institutions.

As cloud computing is the dynamic environment and services access by the individual or institution on pay per use basis. So, the computing is done in a virtual dynamic environment and service interactions can be created in a more dynamic way than traditional e-commerce scenarios. Services can potentially be aggregated and changed dynamically by service providers and can also change the provisioning of services. The exchange of data should takes place in such a dynamic arrangements. So, chances of security holes creation also increase in such dynamic exchange of data and hence the security provisions are also required.

A cloud data storage is a collection of softwares, videos, pictures, documents, contact and files. This storage may be encrypted and indexed by the Data Owners when placed on the cloud servers. Data Users may search results and ranked results by accessing these cloud storage items. There are the provisions of having search controls like trapdoor and Access Control like Decryption Keys. But the cloud computing does not provide control over the stored data in the cloud data centres. The cloud service providers have full of control over the data, they can perform any malicious tasks such as copy, destroying, modifying, etc. The only encryption doesn't give full control over the stored data but it gives somewhat better than plain data.

III. REVIEW OF THE LITERATURE

Xuyun Zhang, Chang Liu, Surya Nepal, Jinjun Chen proposed an efficient quasi-identifier index based approach for privacy preservation over incremental data sets on cloud. They state that Cloud computing provides massive computation power and storage capacity which enable users to deploy applications without infrastructure investment. Many privacy-sensitive applications like health services are built on cloud for economic benefits and operational convenience. Usually, data sets in these applications are anonymized to ensure data owners' privacy, but the privacy requirements can be potentially violated when new data join over time. Most existing approaches address this problem via re-anonymizing all data sets from scratch after update or via anonymizing the new data incrementally according to the already anonymized data sets. However, privacy preservation over incremental data sets is still challenging in the context of cloud because most data sets are of huge volume and distributed across multiple storage nodes. Existing approaches suffer from poor scalability and inefficiency because they are centralized and access all data frequently when update occurs. In this paper, we propose an efficient quasi-identifier index based approach to ensure privacy preservation and achieve high data utility over incremental and distributed data sets on cloud. Quasi-identifiers, which represent the groups of anonymized data, are indexed for efficiency. An algorithm is designed to fulfil our approach accordingly. Evaluation results demonstrate that with our approach, the efficiency of privacy preservation on large-volume incremental data sets can be improved significantly over existing approaches [1].

Vanga Odelu, Ashok Kumar Das, Adrijit Goswami presented a secure effective dynamic group password-based authenticated key agreement scheme for the integrated EPR information system. With the rapid growth of the Internet, a lot of electronic patient records (EPRs) have been developed for e-medicine systems. The security and privacy issues of EPRs are important for the patients in order to understand how the hospitals control the use of their personal information, such as name, address, e-mail, medical records, etc. of a particular patient. Recently, Lee et al. proposed a simple group password-based authenticated key agreement protocol for the integrated EPR information system (SGPAKE). However, in this paper, we show that Lee et al.'s protocol is vulnerable to the off-line weak password guessing attack and as a result, their scheme does not provide users' privacy. To withstand this security weakness found in Lee et al.'s scheme, we aim to propose an effective dynamic group password-based authenticated key exchange scheme for the integrated EPR information system, which retains the original merits of Lee et al.'s scheme. Through the informal and formal security analysis, we show that our scheme provides users' privacy, perfect forward security and known-key security, and also protects online and offline password guessing attacks. Furthermore, our scheme efficiently supports the dynamic group password-based authenticated key agreement for the integrated EPR information system. In addition, we simulate our scheme for the formal security verification using the widely-accepted AVISPA (Automated Validation of Internet Security Protocols and Applications) tool and show that our scheme is secure against passive and active attacks [2].

L. Malina, J. Hajny, P. Dzurenda and V. Zeman discussed Privacy-preserving security solution for cloud services. We propose a novel privacy-preserving security solution for cloud services. Our solution is based on an efficient non-bilinear group signature scheme providing the anonymous access to cloud services and shared storage servers. The novel solution offers anonymous authentication for registered users. Thus, users' personal attributes (age, valid registration, successful payment) can be proven without revealing users' identity, and users can use cloud services without any threat of profiling their behavior. However, if a user breaks provider's rules, his access right is revoked. Our solution provides anonymous access, unlinkability and the confidentiality of transmitted data. We implement our solution as a proof of concept application and present the experimental results. Further, we analyze current privacy preserving solutions for cloud services and group signature schemes as basic parts of privacy enhancing solutions in cloud services. We compare the performance of our solution with the related solutions and schemes [3].

D. Chandramohan, T. Vengattaraman, D. Rajaguru, P. Dhavachelvan presented a new privacy preserving technique for cloud service. In data analysis the present focus on storage services are leveraged to attain its crucial part while user data get compromised. In the recent years service user's valuable information has been utilized by

unauthorized users and service providers. This paper examines the privacy awareness and importance of user's secrecy preserving in the current cloud computing era. Gradually the information kept under the cloud environment gets increased due to its elasticity and availability. However, highly sensitive information is in a serious attack from various sources. Once private information gets misused, the probability of privacy breaching increases which thereby reduces user's trust on cloud providers. In the modern internet world, information management and maintenance is one among the most decisive tasks. Information stored in the cloud by the finance, healthcare, government sectors, etc. makes it all the more challenging since such tasks are to be handled globally. The present scenario therefore demands a new Petri-net Privacy Preserving Framework (PPPF) for safeguarding user's privacy and, providing consistent and breach-less services from the cloud. This paper illustrates the design of PPPF and mitigates the cloud provider's trust among users. The proposed technique conveys and collaborates with Privacy Preserving Cohesion Technique (PPCT), to develop validate, promote, adapt and also increase the need for data privacy. Moreover, this paper focuses on clinching and verification of unknown user intervention into the confidential data present in storage area and ensuring the performance of the cloud services. It also acts as an information preserving guard for high secrecy data storage areas [4].

Y. A. A. S. Aldeen, M. Salleh, Y. Aljeroudi proposed an innovative privacy preserving technique for incremental datasets on cloud computing. Cloud computing (CC) is a magnificent service-based delivery with gigantic computer processing power and data storage across connected communications channels. It imparted overwhelming technological impetus in the internet (web) mediated IT industry, where users can easily share private data for further analysis and mining. Furthermore, user affable CC services enable to deploy sundry applications economically. Meanwhile, simple data sharing impelled various phishing attacks and malware assisted security threats. Some privacy sensitive applications like health services on cloud that are built with several economic and operational benefits necessitate enhanced security. Thus, absolute cyberspace security and mitigation against phishing blitz became mandatory to protect overall data privacy. Typically, diverse applications datasets are anonymized with better privacy to owners without providing all secrecy requirements to the newly added records. Some proposed techniques emphasized this issue by reanonymizing the datasets from the scratch. The utmost privacy protection over incremental datasets on CC is far from being achieved. Certainly, the distribution of huge datasets volume across multiple storage nodes limits the privacy preservation. In this view, we propose a new anonymization technique to attain better privacy protection with high data utility over distributed and incremental datasets on CC. The proficiency of data privacy preservation and improved confidentiality requirements is demonstrated through performance evaluation [5].

Jian Wang Yan Zhao Shuo Jiang Jiajin Le, expressed that people can only enjoy the full benefits of Cloud computing if we can address the very real privacy and security concerns that come along with storing sensitive personal information in databases and software scattered around the Internet. There are many service provider in the internet, we can call each service as a cloud, each cloud service will exchange data with other cloud, so when the data is exchanged between the clouds, there exist the problem of disclosure of privacy. So the privacy disclosure problem about individual or company is inevitably exposed when releasing or sharing data in the cloud service. Privacy is an important issue for cloud computing, both in terms of legal compliance and user trust, and needs to be considered at every phase of design. Our paper provides some privacy preserving technologies used in cloud computing services [6].

N.M. Joseph, E. Daniel, N.A. Vasaanthi, presented a survey on Privacy-Preserving Methods for Storage in Cloud Computing. At present the mankind are progressively relying more on a number of online storage stores to back up our data or for using it in real time which gives an anywhere, anytime access. All these services bring with it, concerns of security and privacy weaknesses for all the services provided by them since the user's data are stored and maintained out of user's premises. This paper portrays the various issues associated to privacy while storing the user's data on third party service providers, which is more commonly termed as cloud service. Cloud computing refers to the fundamental infrastructure for an up-coming model of service provision that has the benefit of dropping cost by sharing computing and storage resources, united with an on-demand provisioning mechanism depending on a pay-per-use business model. Without appropriate security and privacy solutions designed for clouds this computing paradigm could become a huge failure. There is a lot of research being made to spot out the issues with these cloud service providers and cloud security in general. This paper is on regard of one of the key issue -privacy that occur in the context of cloud computing and analyze the various works being done to solve the issues in privacy and thus to ensure privacy to outsourced data on cloud storage [7].

Dr. K. Kartheeban, A. Durai Murugan also discussed Privacy Preserving Data Storage Technique in Cloud Computing. During the end of this decade, cloud computing is becoming an important IT buzz

word and is enhanced by way of a paradigm shift of the commercial capabilities technological know-how towards a subscription centered or pay-on-demand provider business mannequin to deal with; in many IT booms around the world. Cloud computing offering more services and data storage is one such important services offered by cloud computing. Since data is now not saved in the costumers own servers, cloud information storage requires the security issues distinctive on purchasers outsourced their data in the cloud service provider (CSP). Meanwhile depending on a single CSP for their external storage information shouldn't be very promising one from a client point of view. In addition, ensuring data availability as well as providing better privacy may also be done through dividing the consumer's input information into data modules and distributing them among the several available CSPs. In this algorithm a minimum number of CSPs should work together in order to receive entire information block. Therefore a Privacy Preserving Multi-Cloud Storage (PPMCS) algorithm is proposed which preserves privacy and availability of information in cloud computing by distribution of information among the multiple available CSPs [8].

Hui Wang presented the Privacy-Preserving Data Sharing in Cloud Computing. Storing and sharing databases in the cloud of computers raise serious concern of individual privacy. We consider two kinds of privacy risk: presence leakage, by which the attackers can explicitly identify individuals in (or not in) the database, and association leakage, by which the attackers can unambiguously associate individuals with sensitive information. However, the existing privacy-preserving data sharing techniques either fail to protect the presence privacy or incur considerable amounts of information loss. In this paper, we propose a novel technique, Ambiguity, to protect both presence privacy and association privacy with low information loss. We formally define the privacy model and quantify the privacy guarantee of Ambiguity against both presence leakage and association leakage. We prove both theoretically and empirically that the information loss of Ambiguity is always less than the classic generalization-based anonymization technique. We further propose an improved scheme, PriView, that can achieve better information loss than Ambiguity. We propose efficient algorithms to construct both Ambiguity and PriView schemes. Extensive experiments demonstrate the effectiveness and e±ciency of both Ambiguity and PriView schemes [9].

Ms. P. R. Bhuyar, Dr. A. D. Gawande, Prof. A. B. Deshmukh, presented Horizontal Fragmentation Techniques in Distributed Database. Distributed database technology is expected to have a significant impact on data processing in the upcoming years. Today's business environment has an increasing need for distributed database and Client/server applications as the desire for consistent, scalable, reliable and accessible information is steadily growing. Distributed processing is an effective way to improve reliability and performance of a database system. Distribution of data is a collection of fragmentation, allocation and replication processes. Previous research works provided fragmentation solution based on empirical data about the type and frequency of the queries submitted to a centralized system. These solutions are not suitable at the initial stage of a database design for a distributed system. The purpose of this work is to present an introduction to Distributed Databases which are becoming very popular now days with the description of distributed database environment, fragmentation and horizontal fragmentation technique. Horizontal fragmentation has an important impact in improving the applications performance that is strongly affected by distributed databases design phase. In this report, we have presented a fragmentation technique that can be applied at the initial stage as well as in later stages of a distributed database system for partitioning the relations. Allocation of fragments is done simultaneously in the algorithm. Result shows that proposed technique can solve initial fragmentation problem of relational databases for distributed systems properly [10].

N. Vurukonda, B.T. Roa discussed that Cloud computing is a revolutionary mechanism that changing way to enterprise hardware and software design and procurements. Because of cloud simplicity everyone is moving data and application software to cloud data centers. The Cloud service provider (CSP) should ensure integrity, availability, privacy and confidentiality but CSP is not providing reliable data services to customer and to stored customer data. This study identifies the issues related to the cloud data storage such as data breaches, data theft, and unavailability of cloud data. Finally, we are providing possible solutions to respective issues in cloud [11].

S. Sehgal, S. Chaudhry, P. Biswas, S. Jain, presented a new genre of recommender system based on modern paradigms of data filtering. In this era of web, we have a huge amount of information overload over internet. To extract useful information, filtering is required. Search engines help to solve this problem to some extent but they do not provide personalization of data. Hence, there is a need of recommendation engine. With the help of recommender software the preferences of user for a particular product can be foreseen. Recommender systems help in pinpointing the required information thereby deescalating unwanted information. Collaborative filtering is the most efficient approach to create recommendations so that the identified choices of a user's group can be used to envisage the preferences for other users which are not yet known to them. Through this paper we endeavor to present a

thorough survey of collaborative filtering methods that can help in future for further research in this field and thereby propose a solution to enhance the precision and recall measures of recommendations [12].

Cloud Computing is defined as an environment in which users can share their resources with others in pay per use model. The resources are stored centrally and can access from anywhere. Despite these advantages, there still exist significant issues that need to be considered before shifting into cloud. Security stands as major obstacle in cloud computing. This paper gives an overview of the security issues on data storage along with its possible solutions. It also gives a brief description about the encryption techniques and auditing mechanisms [13].

Akhil K.M., Praveen Kumar M, Pushpa B.R discussed that Cloud computing is the revolution through which individuals can share resources, services and data among the users through the network. Since millions of users uses the same network for data transfer, the data becomes more vulnerable to different security attacks from intruders. Providing security to these data has become the critical area of concern. The current system for data security concentrates on providing security to the stored data in cloud storage but concerns less on securing the data while it is being transferred. The data becomes prone to intruder attacks while being transferred. Also, in the current existing trend, the third party auditor is given access to data during data transfer. This also increases the access vulnerability of data as the intruder could act as third party and gain access to the data. Considering security as a crucial issue, the system proposed concentrates on providing security to transferring data using encryption technique. The system also takes into consideration the issue concerned with the third party auditor, that in the proposed approach, the auditor is denied access to the user data. Experiments are conducted and has shown that the proposed approach increases the overall security of system by making it difficult for intruders to crack the data being transferred [14].

M. Derfouf, A. Mimouni, M. Eleuldj presented that nowadays, Cloud computing is a major trend. It is a new data hosting technology that becomes very popular lately thanks to the amortization of costs induced to companies. Since this concept is still in its first stages, new security risks start to appear. This paper will set forth the major security issues in cloud computing and propound a new solution to secure data storage in the cloud environment [15].

IV. IMPORTANT FINDING

Security is major concern to the cloud computing. There is strong thrust to provide security at infrastructure -network level, Host level, application level and data. The data is associated with each level like network, host and Application level. In this paper, security of cloud data is focused. Cloud computing uses several technologies. The security issues related to different type attacks related to several technologies needs to be addressed. Some data security issues in cloud computing includes availability, data persistence, third-party control, legal issues and privacy.

Researchers proposed variety of solution for improving privacy and security of data on the cloud servers. One solution is based on an efficient non-bilinear group signature scheme and it provides the anonymous access to the cloud services and shared storage servers. The anonymous authentication for registered users will be a good approach and it proves that without revealing the user's identity, they can access the cloud services without any threat of profiling their behaviour.

Another approach is Quasi-identifier index based approach for privacy preservation over incremental data sets on cloud. This approach also works on the anonymized data sets. Most existing approaches address this problem via re-anonymizing all data sets from scratch after update or via anonymizing the new data incrementally according to the already anonymized data sets. Existing approaches suffer from poor scalability and inefficiency because they are centralized and access all data frequently when update occurs. Another approach proposed by L. Malina, J. Hajny, P. Dzurenda and V. Zeman uses non-bilinear group signature scheme to provide the anonymous access to the cloud services and shared storage servers.

V. CONCLUSION

Personal and institution's important data, which resides on the cloud servers, is valuable for them. If it is shared with others without the knowledge of the owner and/or such private data accessed by other unauthenticated users anyhow then it is a big stumbling block of accepting the cloud computing benefits. As cloud computing have lot of benefits for the society and the computing world, research on privacy preserving and data storage security is very much required. Even researches have presented various approaches for anonymizing the private data for the registered users and hides the user's identity form the data sets. Some of the important contributions in this direction are non-bilinear group signature scheme, Quasi-identifier index based approach, etc.

Further research is required in this field of securing the private data, as computer world now shifting towards placing the important data on cloud servers and using the important services of cloud computing.

ACKNOWLEDGMENT

The authors are thankful to all the people, who help or encourage us directly or indirectly, for the preparation of this research paper.

REFERENCES

[1] Xuyun Zhang, Chang Liu, Surya Nepal, Jinjun Chen, "An efficient quasi-identifier index based approach for privacy preservation over incremental data sets on cloud", Journal of Computer and System Sciences 79 (2013), Elsevier, pp. 542–555

[2] Vanga Odelu, Ashok Kumar Das, Adrijit Goswami, "A secure effective dynamic group password-based authenticated key agreement scheme for the integrated EPR information system", Journal of King Saud University – Computer and Information Sciences (2016) 28, pp. 68–81

[3] L. Malina*, J. Hajny, P. Dzurenda and V. Zeman, "Privacy-preserving security solution for cloud services", Journal of Applied Research and Technology, Science direct, Volume 13, Issue 1, February 2015, Pages 20-31

[4] D. Chandramohan, T. Vengattaraman, D. Rajaguru, P. Dhavachelvan, "A new privacy preserving technique for cloud service user endorsement using multi-agents", Journal of King Saud University – Computer and Information Sciences (2016) vol. 28, pp. 37–54

[5] Y. A. A. S. Aldeen, M. Salleh, Y. Aljeroudi, "An innovative privacy preserving technique for incremental datasets on cloud computing", Elsevier, Journal of Biomedical Informatics, vol. 62 (2016), pp. 107–116

[6] Jian Wang Yan Zhao Shuo Jiang Jiajin Le, "Providing Privacy Preserving in cloud computing", 2009 IEEE International Conference on Test and Measurement, ISSN 978-1-4244-4700-8, pp. 213-216

[7] N.M. Joseph, E. Daniel, N.A. Vasaanthi, "Survey on Privacy-Preserving Methods for Storage in Cloud Computing", Amrita International Conference of Women in Computing (AICWIC'13), Proceedings published by International Journal of Computer Applications® (IJCA), pp. 1-4

[8] Dr. K. Kartheeban, A. Durai Murugan, "Privacy Preserving Data Storage Technique in Cloud Computing", 2017 IEEE INTERNATIONAL CONFERENCE ON INTELLIGENT TECHNIQUES IN CONTROL, OPTIMIZATION AND SIGNAL PROCESSING.

[9] Hui Wang, "Privacy-Preserving Data Sharing in Cloud Computing", JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY, vol. 25, issue 3, pp. 401-414 May 2010

[10] Ms. P. R. Bhuyar, Dr. A. D. Gawande, Prof. A. B. Deshmukh, "Horizontal Fragmentation Techniques in Distributed Database", International Journal of Scientific and Research Publications, Volume 2, Issue 5, May 2012

[11] N. Vurukonda, B.T. Roa, "A study on Data Storage Security Issues in Cloud Computing", presented in the 2nd International Conference on Intelligent Computing, Communication & Convergence (ICCC-2016), published in Elsevier, Procedia Computer Science, vol. 92 (2016), pp. 128 – 135

[12] S. Sehgal, S. Chaudhry, P. Biswas, S. Jain, "A New Genre Of Recommender Systems Based On Modern Paradigms Of Data Filtering", 2nd International Conference on Intelligent Computing, Communication & Convergence (ICCC-2016), Procedia Computer Science, vol. 92 (2016), pp. 562 – 567

[13] M. B. Jayalekshmi and S. H. Krishnaveni, "A Study of Data Storage Security Issues in Cloud Computing", Indian Journal of Science and Technology, Vol 8 (24), 2015, pp. 1-5.

[14] Akhil K.M., Praveen Kumar M, Pushpa B.R, "Enhanced Cloud Data Security Using AES Algorithm", 2017, International Conference on Intelligent Computing and Control (I2c2)

[15] M. Derfouf, A. Mimouni, M. Eleuldj, "Vulnerabilities and storage security in Cloud Computing", 2015, IEEE conference, ISSN : 978-1-4673-8149-9