Proposing PDM Model for Securing Data Storage on Cloud Servers

Amit Kumar Chaturvedi, Meetendra Singh Chahar, Kalpana Sharma

Abstract*: Clouds are the group of resources like data storage, processors, security tools, etc. that are shared by the big resource providers like Amazon, Google, Yahoo etc. and the users of these resources. There is the requirement of privacy policy when we work in such a shared, unknown, untrusted, and pay per use environment. Computing industry is now shifted their orientation from arranging first the resources to developing new applications or application of new ideas. Because in the cloud computing world , every resource required for developing and executing an idea or application is available or pay per use basis. But even though everything is available with scalability or these resources, the data of the business transaction with authenticity is prime for all either business organization or customer. As we know that user's and transaction data is very important and unauthorized access is illegal and harmful for everyone. Because worldwide the use of online services is increasing exponentially and the use of cloud computing for these solutions is also increasing. We have considered this problem for our research work and in this paper, we are proposing a PDM privacy preserving model for more securing the cloud data.

***Keywords* : Perturbation, Data Security, encryption, decryption, cryptography.**

I. INTRODUCTION

Cloud computing is such a kind of system that uses sharing of computing resource on pay per use basis and there are unlimited resources and services in all dimensions of computing as per the requirement and paying capability of the user. The payer have the multiple choices also for the same resources. The important thing is that the business organization may provide the services worldwide through the internet network. These all resources and services like data storage, security, processors, virtual machines, load balancers, APIs, etc. are available online for 24x7 from anywhere, anytime. Internet access is the backbone for getting and providing these services. Cloud computing uses a layered architecture and the three layers of it are IaaS, PaaS, and SaaS.

IaaS layer deals with providing the infrastructure services like virtual machines, servers, load balancers, network related services, etc. PaaS layers deals with providing platform related services like execution runtime environments, databases or storage space, web servers, and various development tools to the application developers. SaaS is very important layer of the cloud computing for accessing and service administration of the various applications. As cloud computing is mostly used for CRM (Customer-relationship Management) services, various email applications are developed using cloud computing and the popular example of it is the Gmail, Virtual Desktops are also available for the users with bunch of application services, various communication or chat applications are also developed using the cloud infrastructure example whatsapp, facebook etc., various online game applications are also developed using the cloud computing. So, it is very useful computing environment with lot of application and with the capability to develop new and innovative application with very less efforts in pay per use basis. It has lot of scopes for scalability of these application if adopted by the user. Its services may be started 24 x 7 by just paying for it from anywhere, anytime and closing the service is so easy as we switch off the lights.

In cloud computing, the source of the services are unknown, we will pay and get the services through the online solutions. This mode of service in one point of view increases the security level, but if fraud happened then it is very difficult to identify the culprit. Now, it is regularly a matter of discussion between the researcher, users, and service providers that how to work in such an efficient, easy to start, economic, scalable, untrusted cloud environment with the security of user's or organizational personalize and sensitive data. The importance and demand of such security solutions also increases when we judge the trust of the users with these solutions, because data is prime for all either a customer or organization.

Main feature of the cloud services is that user's data are usually processed online on machines that users does not know. It can become a main problem for the growth of cloud services, where data is sensitive. In this research work, we are proposing a highly decentralized framework to keep track of the actual usage of the data owner's data in the cloud. The Cloud Information Accountability framework given in this paper conducts automated logging and distributed auditing of access of data by any user, carried out at any time. It has two major components: logger and log harmonizer. Data owner will get confirmation that his data is handled according to his desire.

II. NEEDS OF SECURING CLOUD DATA STORAGE

Data security is classified into three different categories:

1. Privacy Preservation defines that Privacy of personal and important information in cloud is crucial as the cloud servers are not trusted. Confidentiality and authorization are main requirements of privacy preservation in cloud.

Revised Manuscript Received on February 05, 2020.

***** Correspondence Author

Dr. Amit Kumar Chaturvedi*, CS Dept, Engineering College, Ajmer, Rajasthan, India. Email: amit0581@gmail.com

Meetendra Singh Chhahar, Ph.D. Scholar, CS Dept., Bhagwant Univ., Ajmer , Raj. India. Email: meetendra26@gmail.com

Dr. Kalpana Sharma, CS Dept., Bhagwant Univ. Ajmer, Raj., India. Email: kalpanasharma26@gmail.com

2. Storage Security defines that Cloud storage Security is the task of providing integrity to the shared data stored at dishonest cloud servers.

3. Data Security defines that the data or information security is the process of protecting the data from unauthorized users, preventing alterations and restricting the access of sensitive information.

There is consistent growth in the field of cloud computing. These services can be utilized on pay-per-use basis. The problem of disclosure of privacy occurs when data is exchanged in the cloud. The idea is to build privacy preserving model of storage, where data sharing services can update and control the access and usage of their shared data. In cloud computing, privacy preserving is an important issue and it needs to be considered at every phase of design. The paper proposes a metadata based storage methodology along with an encryption technique to provide additional security.

Data perturbation is classified into two kinds: **random perturbation (RP)** and **randomized response (RR)**. The order relationship of exchanging number in RP achieves privacy-preserving goal by hiding the corresponding relation between numbers and their objects; RR is to add appropriate amount of random noise into data under the condition of no change in raw data distribution. Data can be centralized data and distributed data or horizontal partition and vertical partition. At present centralized privacy-preserving method mainly adopts RR and RP to realize. Distributed privacy preserving data mining algorithm is realized through the employment of SMC.

Preserving privacy using data perturbation in data stream: The proposed hybrid algorithms for data perturbation that is the data perturbation for privacy preserving in data stream clustering. Perturbation techniques are often evaluated with two basic metrics: ***level of privacy guarantee*** and ***level of model-specific data utility preserved***. The main idea of Perturbation-Based technique involves increasing a noise in the raw data in order to perturb the original data distribution and to preserve the content of hidden raw data. Data perturbation algorithms have been proposed for data set perturbation also included permutation techniques like Translation Based Perturbation and Rotation Based Perturbation.

III. PERTURBATION BASED DATA MODEL [PDM] AND METHODOLOGY

There are variety of proposals submitted by the researchers to convert the data from its original form to some another form before saving or transferring to the third party servers like cloud server storage. Perturbation is also one of the technique that we are using here in this paper. Besides perturbation we may also increase the security by applying a suitable cryptographic technique on the data, which definitely increases the complexity and obviously time consuming, but because data is prime, so it needs to be applied. So, in our proposal, we have used both the perturbation and cryptography to secure the data. Now, let us understand the PDM model.

In perturbation the original values are changed with some synthetic data values, so that the statistical information computed from the perturbed data does not differ from the statistical information computed from the original data to a larger extent. The perturbed data records do not agree to real-world record holders, so the attacker cannot perform the thoughtful linkages or recover sensitive knowledge from the available data. Perturbation can be done by using additive noise or data swapping or synthetic data generation.

The basic idea behind proposing this PDM model is to add some noise in the sensitive and personal data called datasets in the model. After adding the noise to the original data, it is not meaningful for the others. This noise can be categorized into two: (1) additive and (2) multiplicative noise. Because according to our proposed model the original data first converted into the ASCII values and then we add noise using our perturbation addition algorithm add_perturb() to this data, which is in the ASCII form. There are two modes of adding this noise either in the additive or in the multiplicative form. After adding noise, the data is encrypted using DES algorithm in this example, but we may use other encryption algorithms also like AES, TrippleDES, Twofish, etc. and the whole data will be encrypted. Now data perturbation is applied successfully. This encrypted data is now safe to send it on the cloud data stores.

When we have to use the data, we fetch this data from the cloud data stores and first do the decryption using DES, the outcome of the decryption is the ASCII values but with additive noise. So, before proceeding the noise has to be removed first from this ACSII data. After removing the noise, convert the ASCII value into respective data or value. This is the final data that we want and is the original data.

Hence, this PDM model provides a very secured form of data to be saved on the third party cloud servers or storage. Nobody can theft the original data, if we use this model. This model uses an innovative technique to secure the data.

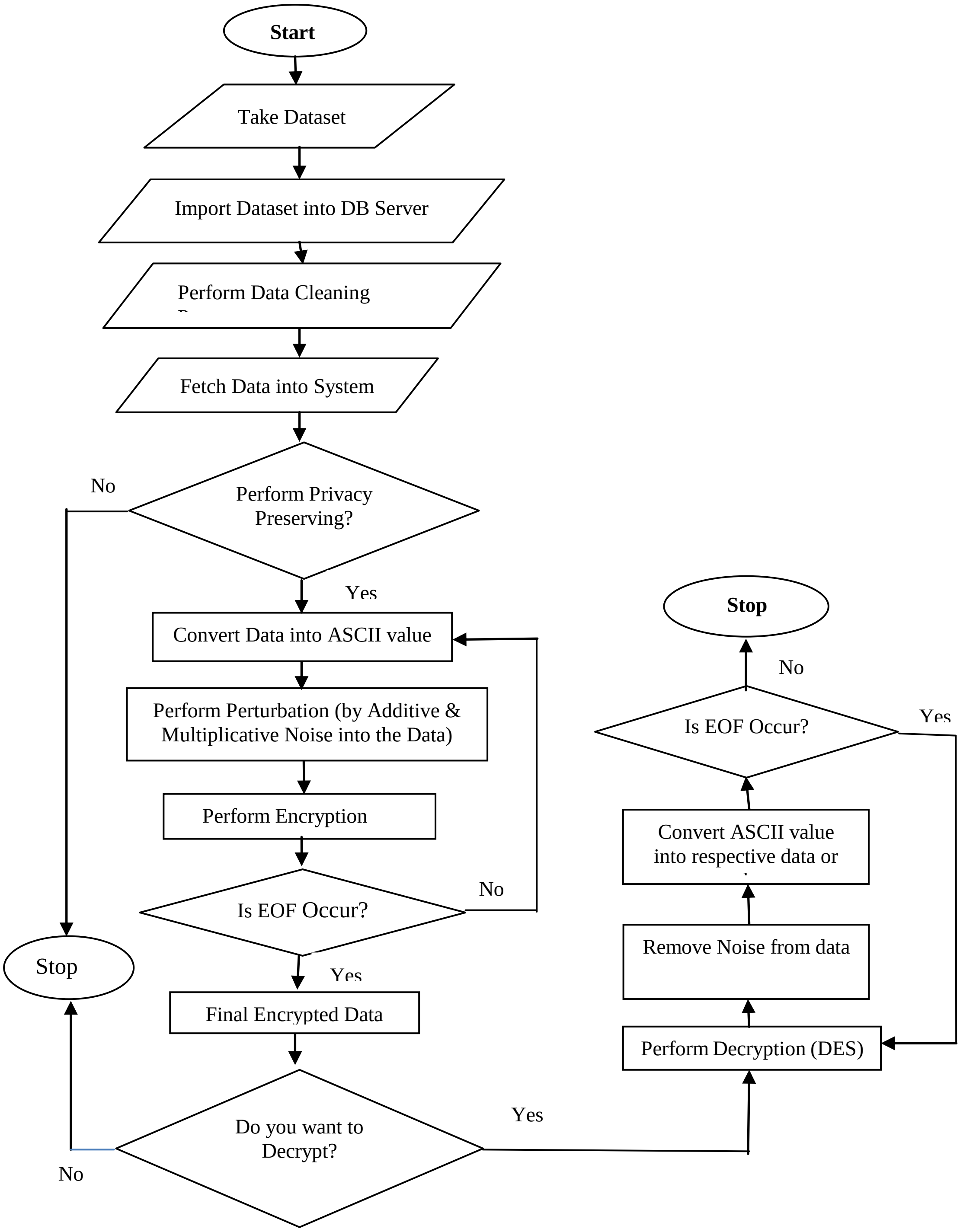

Figure 1 : flowchart of Perturbation based data model [PDM]

IV. RESULT ANALYSIS

As securing data is a strategic challenge, most of the research proposes solutions with direct applying the cryptography or by using the perturbation technique.

We have segregated the data in two forms: (1) personal information, and (2) sensitive information. But in this proposal, we have first take the dataset or plain data and before adding the noise or perturbing the data. We first converted the data in the ASCII form. Because we have added the noise in the ASCII form of data in additive form and in this form the size of the data file is little bit increased. We may also add the noise in the multiplicative form and that form will increase the size of the data file in multiples i.e. 4 times, 8 times or more, as per the key size. The increase in the size of the file matters and it directly related to its execution time, consumed memory, waiting time, etc. and is directly related to the incurred cost.

So, if we compare the additive and multiplicative form of adding noise using the perturbation techniques, the additive form of adding noise is cost effective and hence for this reason we have used the additive form in our proposal.

After adding the noise using the perturbation technique or using add_perturb() algorithm. The outcome file is then encrypted using an encryption algorithm like DES, AES, etc. and in our proposal we have used the DES algorithm.

The outcome of this encryption is the final that may be saved to the third party server or cloud data storage.

Hence, if we compare our PDM model with the existing models. The results are definitely good and there will be 0% chance of unauthorized access of the data secured using this model.

V. CONCLUSION

Here, in the PDM model, we have applied privacy preservation on the datasets. First of all, we will convert all data in to their respective ASCII values. Now after that we will apply perturbation techniques means we are going to add noise on the data. Next we will apply cryptography technique on the data. In the cryptography technique we have applied DES algorithm for the next procedure. Now data perturbation is applied successfully. Now to achieve original data back we have to perform the reverse process again as explained in the PDM model.

Privacy preserving can be achieved by using two techniques, by adding the noise and using the cryptography we can protect the data. In this PDM model, we have used both. So, here data loss will be 0%. But it will some time while performing encryption as well as decryption. There are no chances of data loss. While if we apply only Perturbation technique then there will be chances of data loss. If we apply only cryptography technique, then quality of data will not that much good, we had improved quality of the data here also.

The main objective of privacy preserving data mining is developing algorithm to hide or provide privacy to certain sensitive information so that they cannot be disclosed to unauthorized parties or intruder. Although a Privacy and accuracy in case of data mining is a pair of ambiguity. Succeeding one can lead to adverse effect on other. In this, we made an effort to review a good number of existing PPDM techniques. Finally, we conclude there does not exists a single privacy preserving data mining algorithm that outperforms all other algorithms on all possible criteria like performance, utility, cost, complexity, tolerance against data mining algorithms etc. Different algorithm may perform better than another on one particular criterion.

ACKNOWLEDGMENT

Many people have supported us for the preparation of this research papers, we are highly thankful for them. We are also thankful for the faculty and staff of the Bhagwant Univ., Ajmer for supporting us at every step of our work.

REFERENCES

1. Kamakshi, P. , "A survey on privacy issues and privacy preservation in spatial data mining",. 2014, International Conference on Circuits, Power and Computing Technologies [ICCPCT-2014]. doi:10.1109/iccpct.2014.7054961.
2. Akash Siddhpura, Prof. Daxa V. Vekariya, "An approach of Privacy Preserving Data mining using Perturbation & Cryptography Technique", International Journal on Future Revolution in Computer Science & Communication Engineering, ISSN: 2454-4248, 2018, Volume: 4 Issue: 4 255 - 259.
3. Gurevich, A., & Gudes, E., "Privacy preserving Data Mining Algorithms without the use of Secure Computation or Perturbation", 2006, 10th International Database Engineering and Applications Symposium (IDEAS'06). doi:10.1109/ideas.2006.37
4. Li. Liu, M. Kantarcioglu and B. Thuraisingham, "Privacy Preserving Decision Tree Mining from Perturbed Data", 2009, 42nd Hawaii International Conference on System Sciences. doi:10.1109/hicss.2009.353
5. Vulapula Sridhar Reddy, Barige Thirumala Rao, " A Combined Clustering and Geometric Data Perturbation Approach for Enriching Privacy Preservation of Healthcare Data in Hybrid Clouds", 2017, Department of Computer Science and Engineering, Koneru Lakshmaiah University, Vijayawada, India
6. Kaur, A., "A hybrid approach of privacy preserving data mining using suppression and perturbation techniques", 2017, International Conference on Innovative Mechanisms for Industry Applications (ICIMIA). doi:10.1109/icimia.2017.7975625
7. A. Shinde, K. Saxena, A. Mishra, S.K.Sahu, "Privacy prevention of sensitive rules and values using perturbation technique",2016, ISSN: 978-9-3805-4421-2/16/$31.00_c 2016, IEEE, pp. 577-581
8. Dhiman, E. V., Himakshi, E., Kaur, E. A., & Kumar, M. "Pragmatic approach to conquer security perturbation in cloud computing using level classification", 2017, 2nd International Conference for Convergence in Technology (I2CT). doi:10.1109/i2ct.2017.8226119
9. Kaur, A., & Sofat, S., "A proposed hybrid approach for privacy preserving data mining", 2016, International Conference on Inventive Computation Technologies (ICICT), doi:10.1109/inventive.2016.7823283
10. Yonezawa, C., & Takeuchi, S., "Perturbation caused by cloud in ERS SAR interferogram", IGARSS-2003, 2003, IEEE International Geoscience and Remote Sensing Symposium. Proceedings (IEEE Cat. No.03CH37477). doi:10.1109/igarss.2003.1295517, pp. 4365-4367
11. Yan, K., Du, Y., & Ren, Z. , "MPPT Perturbation Optimization of Photovoltaic Power Systems Based on Solar Irradiance Data Classification", IEEE Transactions on Sustainable Energy, 2018, pp. 1-8, doi:10.1109/tste.2018.2834415
12. Pariselvam, S., & Swarnamukhi, M., "Encrypted Cloud Based Personal Health Record Management Using DES Scheme", 2019, IEEE International Conference on System, Computation, Automation and Networking (ICSCAN), doi:10.1109/icscan.2019.8878773
13. Kale, P. V., & Welekar, R., "A survey on different techniques for encrypted cloud data", 2017, International Conference on Intelligent Computing and Control Systems (ICICCS), doi:10.1109/iccons.2017.8250718, pp. 245-247
14. Kaur, M., Jain, A., & Verma, A., "Optimized cloud storage capacity using data hashes with genetically modified SHA3 algorithm", 2017, International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), pp. 2980-2984, doi:10.1109/icecds.2017.8390002
15. Shen, J., Liu, D., Shen, J., Tan, H., & He, D., "Privacy Preserving Search Schemes over Encrypted Cloud Data: A Comparative Survey", 2015, First International Conference on Computational Intelligence Theory, Systems and Applications (CCITSA), pp. 197-202, doi:10.1109/ccitsa.2015.46
16. Tangwongsan, S., & Itthisombat, V., "A highly effective security model for privacy preserving on cloud storage", 2014, "IEEE 3rd International Conference on Cloud Computing and Intelligence Systems", pp. 505-509, doi:10.1109/ccis.2014.7175788
17. Saxena, V. K., & Pushkar, S., "Privacy preserving model in cloud environment", 2014, "Conference on IT in Business, Industry and Government (CSIBIG)", doi:10.1109/csibig.2014.7056953
18. Wang, H., Zheng, Z., & Wang, Y., "A New Privacy-Preserving Broadcast Encryption Scheme from DPVS", 2013, 5th International Conference on Intelligent Networking and Collaborative Systems, pp. 329-334, doi:10.1109/incos.2013.61
19. Gui, Q., & Cheng, X., "A Privacy-Preserving Distributed Method for Mining Association Rules", 2009, International Conference on Artificial Intelligence and Computational Intelligence, pp.294-297, doi:10.1109/aici.2009.486
20. Kun Peng, Feng Bao "Trust Management In Privacy - Preserving Information System", IEEE conference - 2010, pp. 1-4, 978-1-4244-5540-9/10/$26.00 ©2010 IEEE
21. Biswal, B., "Privacy Preserving Data Communication Model", 2009, Second International Conference on Emerging Trends in Engineering & Technology, pp. 333-336, doi:10.1109/icetet.2009.185
22. Mochizuki, Y., & Manabe, Y., "A privacy-preserving collaborative filtering protocol considering updates", 2015, 10th Asia-Pacific Symposium on Information and Telecommunication Technologies (APSITT), pp. 142-144, doi:10.1109/apsitt.2015.7217100

23. Stant, O., Sirdey, R., Gouy-Pailler, C., Blanchart, P., BenHamida, A., & Zayani, M.-H. , "Privacy-Preserving Tax Calculations in Smart Cities by Means of Inner-Product Functional Encryption", 2018, 2nd Cyber Security in Networking Conference (CSNet), doi:10.1109/csnet.2018.8602714
24. Drosatos, G., Efraimidis, P. S., Athanasiadis, I. N., D'Hondt, E., & Stevens, M., "A Privacy-Preserving Cloud Computing System for Creating Participatory Noise Maps", 2012, IEEE 36th Annual Computer Software and Applications Conference, pp. 581-586, doi:10.1109/compsac.2012.78
25. Pooja HP , Nagarathna N "Privacy Preserving Issues and their Solutions in Cloud Computing: A Survey", International Journal of Computer Science and Information Technologies (IJCSIT), Vol. 6 (2) , 2015, pp. 1588-1592.

AUTHORS PROFILE

Dr. Amit Kumar Chaturvedi is presently working as Assistant Prof in CS Dept of Engineering college, Ajmer. He has around 19 years of long PG teaching and 10 years of research experience. He has published 80 research papers in national and international journals, conferences, seminars. His main research areas are adhoc networks, cloud computing, cryptography, steganography, and image processing.

Meetendra Singh Chahar received his M.C.A. degree in 2005. He is a research scholar of Bhagwant University, Ajmer and pursuing Ph.D. degree in Computer Science. He has published over 3 research papers in refereed Journal. He is a member of Computer Society of India. His research interests include Cloud Computing and Cryptography.

Dr. Kalpana Sharma [Ph.D (Computer Science), MCA, M.Sc.(CS)] is presently working as an assistant professor in department of Computer Science & Application, Bhagwant University, Ajmer. She has authored more than 25 research papers in international and national journal. She has guided 20 PG in their project work. She is also member of Board of Studies, member of DRC, at Bhagwant University, Ajmer. Apart from this she is a convener of Rural Computer Literacy Program (inspired by Digital India Campaign) held twice a year in Bhagwant University Ajmer for the students of nearby rural area.