Privacy Preservation of Data using Hybrid Approach

Gunjan S. Bonde

Department of Computer Science and engineering

SSBT's College of engineering and technology ,

Kaviyatri Bahinabai Chaudhari N.M.U, Jalgaon ,[M.S], India
Email:gunjansbonde11@gmail.com

Akash D. Waghmare

Department of Computer Science and engineering

SSBT's College of engineering and technology,

Kaviyatri Bahinabai Chaudhari N.M.U, Jalgaon ,[M.S],India
Email: aakashjan8@gmail.com

Abstract–**During last few years, organizations which collected data has the requirement that they need to preserve the privacy of the users sensitive data. Users sensitive data is transmitted online over the internet because of the enhance use of E-commerce, and it needs to maintain privacy of the individuals data in document. Also needs to preserve authenticity. There is an approach in which data is transmitted in perturbed form so that users identity or sensitive information cannot be exposed. In proposed system, perturbation technique is applied to users sensitive data and authentication protocol based on MAC of receiver is used so only authenticated user can access the data. It is the process in which data in the document is perturbed and secured by secrete key. In such technique, after authentication original data will be retrieved only when key, file and authenticated MAC address matches. The evaluation result shows reduce the information loss while reconstructing original data.**

Keywords **– Anonymization, Perturbation, Authentication, Privacy, MAC**.

I .INTRODUCTION

Privacy preserving plays an important role to provide security tousersimportatnt data. Data exposure means unauthorized transmission of sensitive data to an unknown destination where the condentiality of information is compromise. In the growing technology, todays generation is more conscious about their privacy being preserved while use of their data in any way. The objective of this technique is sensitive data of user not being reveal and misused. In proposed system, perturbation technique is used to provide security to users sensitive information in online document transmission. In this technique original sensitive data in file such as PAN number, aadhar number, email address, mobile number can be perturbed with some other random data. Files are exchanged via email, so for safe and reliable information exchange between parties, it needs a technique to verify that the other party requesting access is actual authenticate user. Such technique uses MAC address of receiver for authentication. Authenticated user can access original document when key, file and registered MAC address matches. Also validates that file is shared with particular user who tries to access it.

Online transactions and purchasing is one of the important reason for exposure of users sensitive information because original information may be used by any unauthorized user so it needs to maintain privacy of such information. Common techniques like Condensation , Anonymization are used to preserve the privacy, but unfortunately they leads to information loss while reconstructing original data. So needs a technique which hides original sensitive information of users before transmit. Also needs a technique to verify that the receiver requesting access is actual authentic user. To provide security against data violation perturbation technique based on MAC address may be the solution.

The contribution in the propose system is to preserve the privacy of users original sensitive information in document by perturbation technique. In which perturbation method applies to original document before exchanging to make the sensitive data perturbed in such document so the unauthorized user cannot access it and also original data is reconstructed without any information loss.It needs to validate that file is shared with particular user who tries to access it. Authentication protocol based on MAC address is proposed so the only authenticated user allows to access original document.

The rest of the paper is organized as follows: Section II described the Related Work. Section III described the Methodology. Section IV described the Result and Discussion, while Section V Conclusion and Future Work of the paper.

II. LITERATURE SURVEY

The structure of literature survey is shown in Figure 1. Security plays an important role during the data transmission.

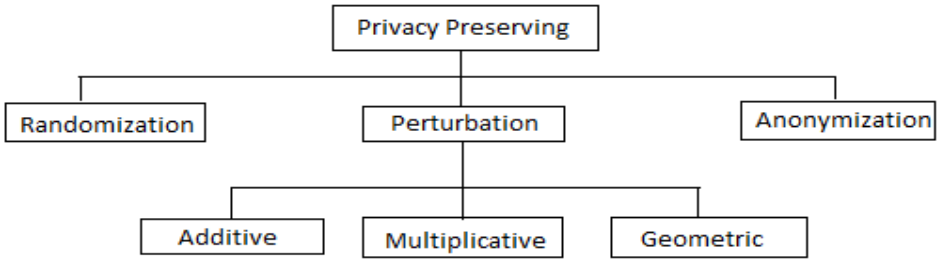

Figure 1. Structure of Literature Survey

R. Mahesh and T. Meyyappan, in [3], proposed a method such as generalization of quasi-identifier to maintain the privacy by setting quasi- identifier in certain range and duplicate records are deleted. Due to deletion of duplicate records they improve in reducing information loss and provides more privacy and better performance when compared with l-diversity and k-anonymity. Such method provides security from record linkage and attribute linkage attack. This method only works well with numeric data.

Neha Gupta, et. al., in [4], proposed a method for maintain the privacy in translation and rotation based perturbation. In translation based perturbation technique particular noise is added to data for alteration and in rotation based the data is replaced by the angles. Value of k in the dataset is take as median for number of records in rotation based techniques. k-records are takes at a time and for rotate the data threshold value is used. Value of noise is selected for translation based techniques and such noise is added to record if noise value is smaller than attribute value. Otherwise it is subtracted to replaced the data.

MebaeUshida, Kouichi Itohet. al., in [5] , proposed a method of data aggregation for satisfying the needs of cloud which gives assurance against the exposure of information which is stored and gives results as per ability. Before any data stored to cloud it needs such data is altered by user and then such altered results are retrieved. Each user has secrete private key and user used such key for obtaining the original data from altered data.

L. Sweeney, in [6], proposed a method to maintain the privacy during publishing data based on anonymization. Data holders can encrypt or alter users sensitive information such as name and unique security numbers to protect users identity when releasing sensitive information. plain text data provides no guarantee for anonymity. K-anonymity model is used to achieves k-anonymity using suppression and generalization to preserve privacy. It is di□cult for deceiver to decide the identity of users personal information in the collection of data set. Each combination of values of quasi-identifiers which is indistinctly matches to at least k-1 respondents. In generalization, values are replaced with less specific values but semantically reliable value. For example, age of person is replaced by the range such as youth, middle age or adult so as to reduce risk of identification.

Jianming Zhu et. al., in [7], proposed a ElGamal encryption, K-nearest neighbour and homomorphic encryption based on cryptography for maintain the privacy of data. Homomorphic encryption method is finest method for maintaining the privacy as a result because it supports the operations of application on encrypted data. Results shows that even if such technique is doing well for maintain the privacy but it has some limitation such as implementation and time complexity.

M. Suriyapriyaet. al., in [8], proposed the method using symmetric key encryption for maintain privacy. Symmetric key is used for data encryption and key distribution centre generate similay key for encryption and decryption. After the data is encrtpted no one can predict such data. The data is secured only until the key is secured. Original values are retrieved if key is found by unauthorized user they can decrypt the encrypted data. Experimental results shows such technique reduces information loss for preserving the privacy but such method is not successful because of its complicated working and inefficiency in performance when size of data is increases.

III. METHODOLOGY

The propose approach focuses on preserving the privacy of users sensitive data while transmission of documents.Today's generation is more conscious about their privacy being preserved while use of their data in any way. Disclosure of plaintext sensitive data of users in the file is a serious security issue. To address such problem, privacy of users sensitive data is preserved by perturbing the original data so such data does not reveal sensitive information while transmitting over internet. It needs to verifies that the file is shared with particular user who tries to open the file. Also for safe and reliable information exchange between parties it needs a procedure to verify that receiver that requesting access is actual authentic user. To insure security, authentication protocol based on MAC address of receiver is used so only authenticated user has access to original file..

Architecture

The purpose of hybrid approach is to maintain the privacy of users sensitive data in document. So documents are transmitted in perturbed form and access by only authenticated user. The architecture of proposed hybrid approach is shown in Figure 2.

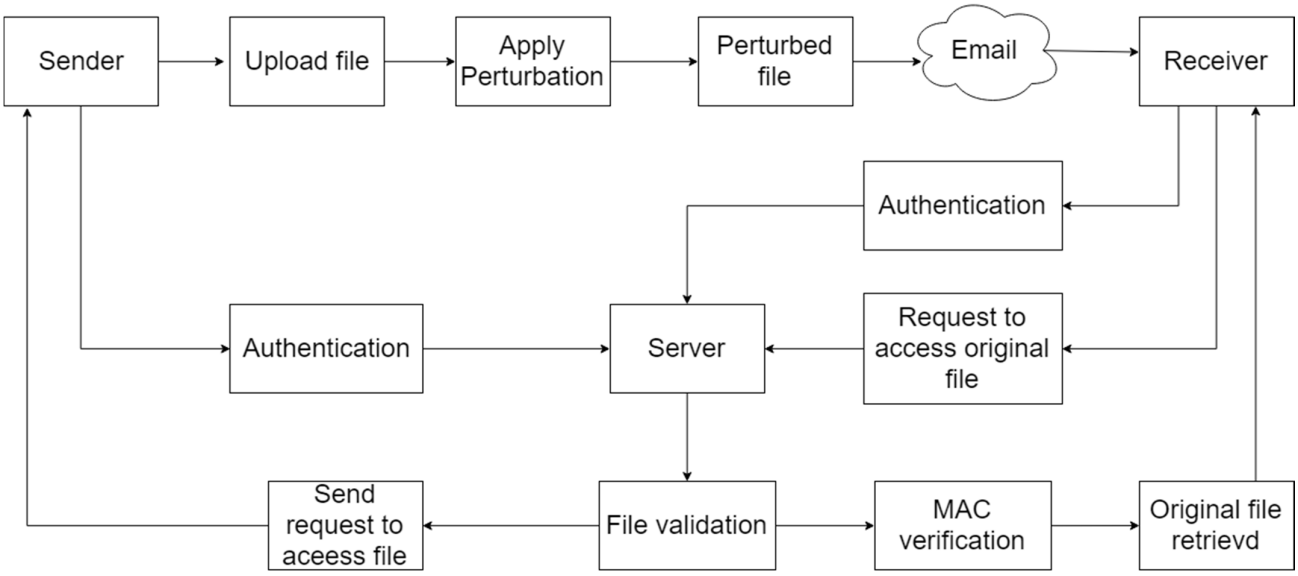

Figure 2. Architecture of the Hybrid Approach

The sender first upload data file and apply perturbation to make it perturbed and file and key send to receiver via email. The sensitive data in file such as PAN number, aadhar number, mobile number and email address is replaced so as original data is altered by some another data. When receiver tries to open file, server checks whether the particular file is shared with receiver who tries to access it and if file is shared with such receiver then checks for MAC address with registered MAC address. If file is not shared with user then request sends to sender for access that file. If file, key and MAC address matches then original file is retrieved otherwise perturbed file is generated.

Perturbation process for alteration of original data is shown in algorithm 1. The purpose of algorithm is to transmit sensitive data in perturbed form so unauthorized user cannot access it.

Algorithm: - Perturb Process

Procedure Perturb file

Require: original pdf document containing sensitive information

1: Select file for perturbation

2: Extract the contents of pdf using streams

3. **Repeat**

Search for sensitive information

Replace the sensitive information with additive data

Encrypt the data and store in pdf header

4: **until** each sensitive information is perturbed

5. perturb file generated

6: **end procedure**

Sender process is given in algorithm 2. In which sender send perturbed file and key to receiver via mail.

Algorithm :-Sender Process

Procedure Sending Perturb File to User

Require: Perturb file from procedure 1

1: Read perturb file

2. Generate key file.

3: Enter email address of receiver& Send file with key to user via http protocol

end procedure

Reverse perturbation is given in algorithm 3. In which original data is generated if key, file and MAC address matches.

Algorithm :-Re-Perturb Process

Procedure Re-Perturb File Received from User

Require: Perturb file received from procedure 2

1: Read perturb file & key received from sender

2. submit file and key to server

3. **Repeat**

Search for pdf header

4.**if**

MAC address verifies

Decrypt the data stored in pdf header

Replace the additive data with sensitive information

5. **End if**

6: **until** every sensitive information is retrieved

7. Store the data into new file

Replace the additive data with sensitive information

8: **until** every sensitive information is retrieved

9. store the data into new file

end procedure

IV.RESULT AND DISCUSSION

Performance Metrics is used in reconstruction of original data from perturbed data for finding the processing time of system. Processing time is measured based on how much sensitive data is re-perturbed. Processing time is calculated based on formula-

Processing time= t_responce – t_request

Where, t_responce= Time of response received

t_request= Time of request submitted

Experimental result consist of the processing time of execution based on how much sensitive data in file and how much time require with respect to file size.

Table 1 shows processing time of proposed approach for 200 KB file . The graph shows in Figure 3 is plotted by consider the processing time. Processing time is calculation of both request time and response time.

Number of Sensitive Data	10	20	30	40	50
Proposed Approach	4500	4750	4800	5000	5500

Table 1. Processing time for different number of sensitive data

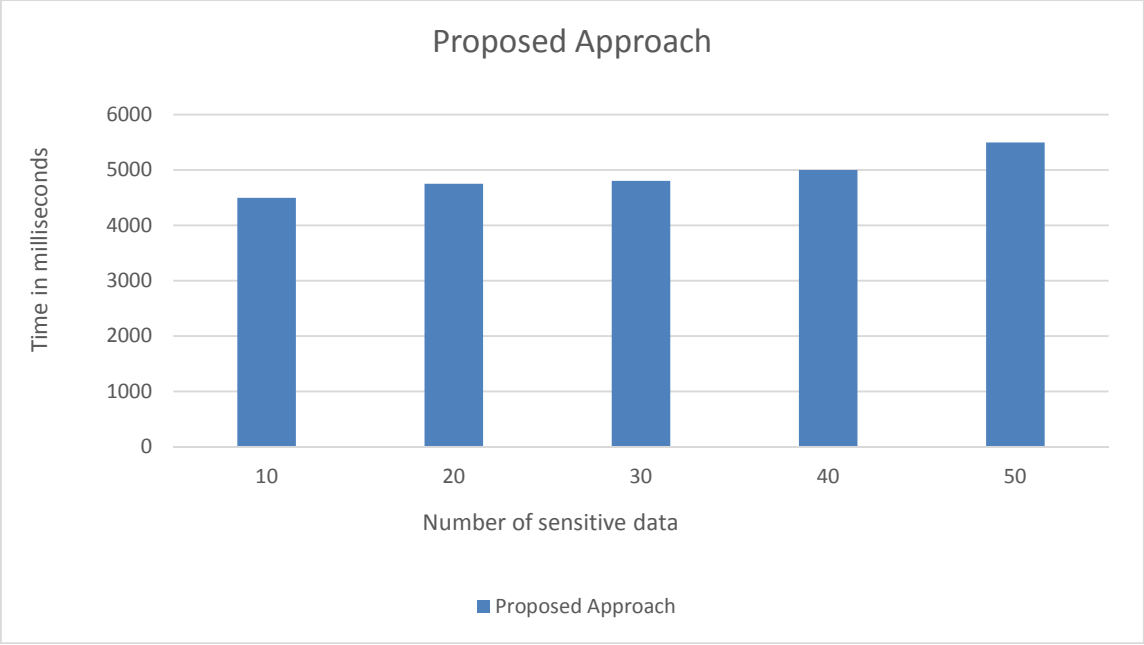

Fig.3 Processing time for different number of sensitive data

Table 2 shows processing time of proposed approach for different file size. The graph shows in Figure 4 is plotted by consider the processing time. Processing time is calculated based on of both request time and response time.

Size of File	100	200	300	400	500
Proposed Approach	4000	4500	4900	5280	5700

Table 2. Processing time for different file size

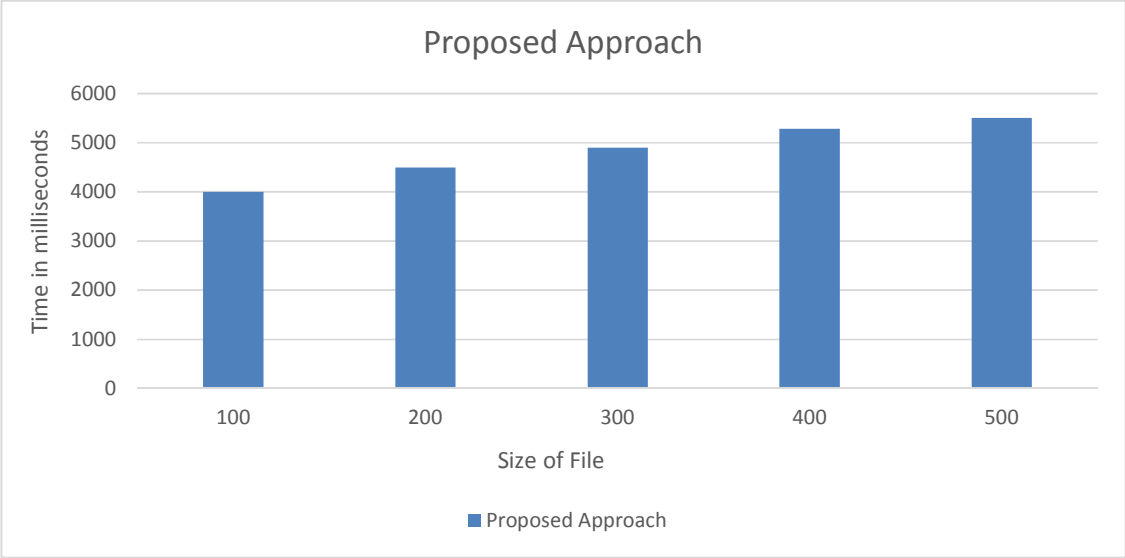

Fig.4 processing time for different file Size

The implementation result of the proposed system shows processing time for different number of sensitive data in same sized file. It shows the performance of system which states processing time to altersoriginal sensitive data to perturbed data. Also calculates processing time for different file size for same number of sensitive information. Time may differs on different processor.

V.CONCLUSION

The privacy preservation of data transmission require more authentication and security. The most important issue of sensitive data transmission is such data is access by unauthorized user. To avoid this the propose system use perturbation method to altered sensitive data with some random data during exchange ofdocument. Also needs that only authentic user can access such data so MAC based authentication is used. Propose system successfully perturbed data and reconstruct original data. The experimental result of the system shows processing time requires to reconstruct different numbers of sensitive data.Future work will be aimed at improving the security and performance so apply to large sensitive dataset.

REFERENCE

[1] Arshveer Kaur, "A Hybrid Approach of Privacy Preserving Data Mining using Suppression and Perturbation Techniques", International Conference on Innovative Mechanisms for Industry Applications IEEE, 2017.
[2] Alpa Shah and Ravi Gulati, "Evaluating Applicability Of Perturbation Techniques For Privacy Preserving Data Mining By Descriptive Statistics", International Conference on Advances in Computing, Communications and Informatics, Sept 2016.
[3] R.Mahesh and T. Meyyappan , "Anonymization technique through record elimination to preserve privacy of published data", International Conference on (pp. 328-332). IEEE., 21-22 Feb, 2013.
[4]N. Gupta, L. Rajput.," Preserving Privacy Using Data Perturbation in Data Stream", International Journal of Advanced Research in Computer Engineering and Technology (IJARCET), 2013.
[5]M. Ushdia, Itoh, K., Y. Katayama, F. Kozakura, &H. Tsuda., "A Proposal of Privacy-Preserving Data Aggregation on the Cloud Computing. In Network-Based Information Systems", 16th International Conference on (pp. 141-148), 2013.
[6]L. Sweeney, "k-Anonymity: A Model for Protecting Privacy", International Journal on Uncertainty, Fuzziness and Knowledge-based Systems", 10(5), 2002.
[7]Zhu, J., "A new scheme to privacy-preserving collaborative data mining", Information Assurance and Security, IAS'09, Fifth International Conference on (Vol. 1, pp. 468-471). IEEE, 2009, August.
[8] M. Suriyapriya, A. Joicy, "Attribute Based Encryption with Privacy Preserving In Clouds", International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169 Volume: 2 Issue: 2, February 2014.
[9] Zhang, X., Yang, L. T., Liu, C., Chen, J., "A scalable two phase top-down specialization approach for data anonymization using map reduce on cloud`", Parallel and Distributed Systems, IEEE Transactions on, 25(2), 363-373, 2014.