Integrated Intelligent Research (IIR)
International Journal of Data Mining Techniques and Applications
Volume: 06 Issue: 01 June 2017, Page No.33-35
ISSN: 2278-2419

A Novel Cryptographic Approach for Privacy Preserving in Big Data Analysis for Sensitive Data

Sujatha K[1], Rajesh.N[2], Udayarani V[3]
[1]Research Scholar, School of computing& IT, REVA University, Bangalore, India.
[2]Lecturer, Information Technology, Shinas college of Technology, Sultanate of Oman
[3]Senior Associate Professor, School of computing& IT, REVA University, Bangalore, India
sujathak@reva.edu.in, rajesh.natarajan@shct.edu.om, udayaraniv@reva.edu.in

Abstract : Privacy preservation of data mining has developed as a widespread study area in order to secure the private information over the network. Privacy preserving methods like L-diversity, Randomization, Data distribution, and Kanonymity have been recommended in directive to perform privacy preservation of data mining. The Privacy preservation of data mining (PPDM) methods protects data by masking or by erasing the original sensitive one to be masked. The private data can be preserved efficiently using cryptographic techniques .In our work we propose a modified two fish algorithm with 256 keys as a cryptographic approach to provide security for PPDM.

Keywords: Privacy Preserving Data Mining (PPDM), Big data, cryptographic, Data privacy

I. INTRODUCION

Big data plays a crucial role in all the fields of engineering .In the area of developing an internet based applications in which big data faces information security risks and privacy protection while accumulating, storing, analyzing and utilizing. Big data applications are very crucial assets to organizations, business, and companies. The accumulating of big data certainly increases the risk of leakage of user privacy. Big data [1] contains a huge amount of user's information in which the development and usages of big data is certainly crucial and easy to break the privacy by technical loopholes. Generally, when big data is transmitted over various areas and department, it causes the leakage of sensitive data which is a threat to privacy. In most of the organizations such as medical databases it is very crucial to keep some of the confidential data as private, and this type of data needs to be preserved. The private data can be preserved efficiently using cryptographic techniques [2]. In most of the organizations such as medical databases it is very much crucial to keep some of the confidential data as private, and this type of data needs to be preserved. The private data can be preserved efficiently using cryptographic techniques [2]. The Privacy preservation of data mining (PPDM) methods[3] protects the data by masking or by erasing the original sensitive one to be masked In our work we propose a modified two fish algorithm with 256 keys as a cryptographic approach to provide security for PPDM.

II. BACKGROUND OF RESEARCH

Privacy preservation of data mining has developed as a widespread study area in directive to secure the sensitivity information over the network. Data mining is a process of mining useful patterns of data. Usually data mining is very much essential because of pattern recognition on huge data storage in either organizations or in the storage networks. The data storage contains several terabytes of data or petabytes of data especially in hospital databases or in social network databases, sales data etc. It is very important to classify the data based on some criteria i.e. pattern so that we can easily arrange the data access and retrieval. Privacy-preservation is considered as a problem of securing private data of data mining systems. The core considerations of privacy preservation are two ways: In the early stage [4] sensitive raw data like personal information including name, address, age so on, should be transformed or clipped out from the database, in order ,not to compromise another entity's privacy. Next, sensitive information is not supposed to be mined using familiar data mining techniques from a database, failing which leads to a compromise in data privacy. At present, many privacy preservation methods [5] for data mining are presented. These include K-anonymity [6], classification, cluster analysis, association rules, distributing privacy preservation, Ldiversity, randomization, taxonomy tree, condensation techniques [4, 7].

The Privacy Preservation of Data Mining methods secures the data by masking or erasing the sensitive data from the original. Privacy preservation is based on the concepts of privacy failure, the ability to determine the unique user data from the changed one, data loss and estimate the data accuracy loss. The primary purpose of these methods is to reduce a trade-off among accuracy and privacy [8]. New approaches that employ cryptographic techniques to avoid information leakage are computationally very classy. Similarly, PPDMs use data distribution methods and it also includes partitioned data in longitude or in latitude fashion through multiple individuals. Beside with these techniques Two fish Algorithm [9] is very useful in securing private data .It consists of 128-bit block cipher [4, 2] that accepts a variable length key up to 256 bits. The cipher is a 16-round Feistel system through a bijective function made up of four key-dependent 8-by-8-bit S-boxes,a static 4-by-4 maximum distance distinguishable matrix over GF(28), a pseudo-Hadamard transform, bitwise rotations, and a reasonably designed key schedule. The strategy of both the round function and the key schedule permits a wide variety of balances among speed, [3] software size, key setup time, gate count, and memory.

III. CRITICAL REVIEW OF LITERATURES AND IDENTIFICATION OF RESEARCH GAPS

Various societies like banking, medical treatment organizations, and search engines accumulate huge data. This data can be shared for the purpose of analysis which in turn helps the organization to gain valuable knowledge which might contain sensitive information about any individual. For

International Journal of Data Mining Techniques and Applications
Volume: 06 Issue: 01 June 2017, Page No.33-35
ISSN: 2278-2419

example, organizations such as hospitals contains health records of the recipients admitted as patients. They share these health records with the scholars for purpose analysis [9]. There is a chance that the person/scholar availing the data which may be sensitive may obtain individuals information. When we speak of sensitive data privacy is of a major concern. As a result, a new path in the field of data mining is created ,known as privacy preservation in data mining (PPDM) [5,9].

In the literature review and in the summary we have stated the importance of privacy preservation. Privacy preservation of data mining [10] has been developed as a wide spread research study in order to secure sensitive information .Such as anonymization should not only satisfy original privacy requirements but also safeguard the utility of the data. Certainly, K-anonymity is an efficient means of privacy preservation in data mining. However, several demonstrated that the data processed by this system often failed to overcome some attacks and are susceptible to internet phishing. As a result, further privacy preservation in data mining using K-anonymity needs an advance data infrastructure to support the combination of present data functionality. This would certainly fulfill the requirements of different kinds of clients and communities. Even though the present search algorithms are capable to speed up the retrieval process, they do not scale up to large volume of data since the linear increase in response time with the amount of the searched datasets. The proposed technique for the searching of distributed large data amongst many cloud providers must possess the ability to preserve privacy, and must be scalable, efficient, well-suited, and good for utility as well as integrity.

The objective of this technique is to secure the sensitive information while retrieving the data. Privacy preservation of data mining methods are separated and split into widespread zones, data securing and information securing. Data securing is elimination or changing sensitive data from the database before revealing it to an external entity. Knowledge security focuses securing the important data which can be extracted from the databank by several data mining systems [8]. Current methods for preserving security in data mining [7] are classified as follows: Data Discomposure, Cryptography techniques and Query auditing [11]. The main constraint lies in the altered data concentrating on the noise that is being added. Usually in the data discomposure method, the transformation of the sensitive data is carried out once the noise is added to the original data. Encryption and decryption contains the use of key for original data. In order to overcome the drawbacks of the existing techniques we are introducing a modified Two fish Algorithm.

Table-1: PPDM METHODS

PPDM:	**Methods Description**
Data distribution:	This might include partitioned data in longitude or in latitude fashion.
Data falsification:	Includes discomposure, delaying, combining or gathering, interchanging and selection.
Knowledge mining algorithms	Includes classifying, Associating, cluster analysis.
Data or rules hidden:	Points to secure main data or rule of creative or new facts.
Kanonymity:	Obtain anonymization.
Ldiversity:	Preserve the diversity of the private data.
Classification Tree:	Attribute are generalized to bind the info leak.
Randomization:	An un-sophisticated and important technique to hide the individual data in PPDM.
Protecting Private data:	Protects the confidentiality, it should adjust data cautiously to achieve ideal data utility.

IV. STATEMENT OF RESEARCH PROBLEM

Identify the data which needs to be preserved from Big Data which is collected by data collector. We propose a modified Two-Fish algorithm in order to preserve the privacy of individual sensitive information.

V. METHODOLOGY OR APPROACH INTENDED TO BE ADOPTED IN THE EXECUTION OF THE RESEARCH

Privacy preservation of data mining [10] is a wide spread research study in order to secure the sensitive information .The Methodology is shown in figure 1.

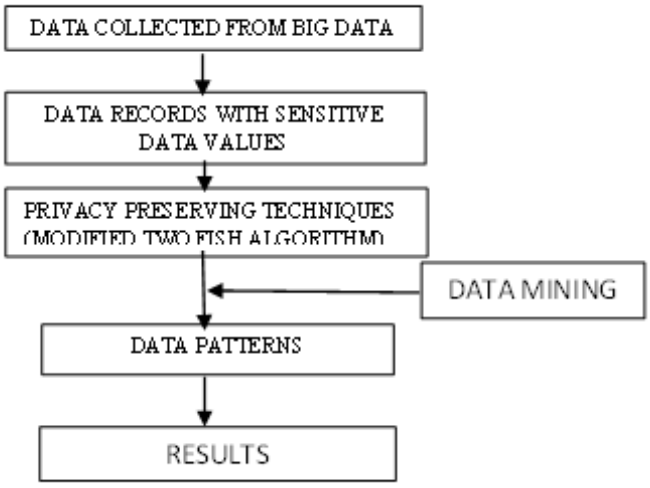

Figure 1. Methodology

VI. OBJECTIVES OF RESEARCH

The main considerations of privacy preservation are in two fields. In the early stage [4] sensitive raw data like personal information including name, address, age so on, should be transformed or clipped out from the original database, in order for the beneficiary of the data to not compromise another entity's privacy.

- Next, sensitive knowledge is not supposed to be mined using familiar data mining techniques from a database, failing which leads to a compromise in data privacy.
- Identify the data which needs to be preserved from big data which is collected by data collector.

International Journal of Data Mining Techniques and Applications
Volume: 06 Issue: 01 June 2017, Page No.33-35
ISSN: 2278-2419

- We propose a modified Two-Fish algorithm in order to preserve the sensitive information.
- We try to increase the key length from 256 to 512.
- Verify the results.

VII. EXPECTED OUTCOMES

The main goal is to protect data privacy and confidentiality. Develop a modified two fish algorithm to secure the sensitive data.
Big data analysis for privacy preservation of sensitive information.

References

[1] Big Data Privacy Preservation, Ericsson Labs,http://labs.ericsson.com/blog/privacy-preservation-in-big-data-analytics.
[2] Benny Pinkas HP Labs benny.pinkas@hp.com , “Cryptographic techniques for privacy-preserving data mining”, SIGKDD Explorations, Volume 4, Issue 2 - page 12
[3] Sharma, Manish, Atul Chaudhary, Manish Mathuria, Shalini Chaudhary, and Santosh Kumar. "An efficient approach for privacy preserving in data mining", 2014 International Conference on Signal Propagation and Computer Technology (ICSPCT 2014), 2014.
[4] S.Bhanumathi and Sakthivel “A New Model for Privacy Preserving Multiparty Collaborative Data Mining” ICCPCT-2013, (845).
[5] Agarwal, R. and Shrikant, R. "Privacy Preserving Data Mining", Proceeding of special interest group on management of data pp.439-450, 2000.
[6] L. Sweeney, "k-anonymity: A model for protecting privacy," International Journal on Uncertainty, Fuzziness and Knowledge-Based Systems, pp. 557–570, 2002.
[7] Elisa Bertino, Dan Lin, and Wei Jiang, A Survey of Quantification of Privacy Preserving Data Mining Algorithms, Springer ,volume 34 pp183-205
[8] V.S Verykios, A.K Elmagarmid, E. Bertino, Y. Saygin and E. Dasseni, "Association Rule Hiding", IEEE Transaction Knowledge and Data Engineering, 16(4): 434 - 447, 2004.
[9] Jian Wang, Yong Cheng Lou, Yen Zh Jiajin Le, "A Survey on Privacy Preserving Data Mining",International Workshop on Database Technology and Application pp. III - 114, 2009.
[10] Anjana Go Sain, Nikita chugh ,”Privacy Preservation in Big Data”,IJCA,Volume 100-No 17,August 2014.
[11] R.L. Rivest, A. Shamir, and L. Adleman , A Method for Obtaining Digital Signatures and Public-Key Cryptosystems