International Engineering Journal For Research & Development

Vol.4
Issue 5

E-ISSN NO:-2349-0721

Impact factor : 6.03

A PRACTICAL APPROACH SCHEMA OF PRIVACY PRESERVING IN DATA STREAMS

Miss A.G.Raut
PG Department of Computer Science
SGBAU,Amravati, India
lanushruti.raut@gmail.com

Dr . S. S. Sherekar
PG Department of Computer Science
SGBAU,Amravati , India
swatisherekar@gmail.com

Dr. V. M. Thakare
PG Department of Computer Science
SGBAU,Amravati, India
vilthakare@gmail.com

Abstract-

Recently, data mining over transactional data streams has become an attractive research area. However, releasing raw transactional data streams, in which only explicit identifying information must be removed and may compromise individual privacy. Many privacy-preserving approaches have been proposed for publishing static transactional data. Due to the characteristics of data streams, which must be processed quickly, static data anonymization methods cannot be directly applied to data streams.

Shadow Coding is use to preserve the privacy in data transmission and ensure the recovery in data collection, it achieves privacy preserving computation in a data-recoverable, efficient, and scalable way. This paper provide practical techniques to make Shadow Coding efficient and safe in data streams.

Keywords— Transactional data streams , Shadow coding.

INTRODUCTION (*HEADING 1*)

Collecting data from distributed data providers is a challenging task to big data related research communities and industries. The protection of privacy is social policy issue due to the contemporary developments in information technology which has made probable the compilation and investigation of millions of transactions containing personal data. Siyuan Liu, Proposed a practical method, Shadow Coding, to preserve the privacy in data transmission and ensure the recovery in data collection, which achieves privacy preserving computation in a data-recoverable, efficient, and scalable way [1]. Method of Data stream mining techniques are classified into three categories according to the stream processing models: the landmark window model, the damped window model and the sliding window model [2]. Also novel privacy-preserving k-means algorithm based on a simple yet secure and efficient multiparty additive scheme that is cryptography-free. [3]. The method is proposed for solution to tarmac way for investigation track and toil well according to the evaluation metrics including hiding effects, data utility, and time performance. [4]. In propose method a light-weight anonymous data perturbation method for efficient privacy preserving in distributed data mining is proposed first define where the privacy constraints for data perturbation based PPDM in a semi-honest distributed environment. [5]. This paper implements the proposed protocols and analyze the computation and communication cost, and security.

EASE OF USE

II) BACKGROUND

As per the studies on Privacy Preserving many models and approaches have been develop for the Providing Privacy in data mining in recent past years. Such approaches are:

Two Privacy-Preserving Approaches for publishing Transactional Data Streams analyze the privacy problem in publishing transactional data streams based on sliding window. [1]. Efficient and Privacy-Preserving k-means clustering For Big Data Mining in technique proposed a novel privacy-preserving k-means algorithm based on a simple yet secure and efficient multiparty additive scheme that is cryptography-free. [2]. SMC: A Practical Schema for Privacy-Preserved Data Sharing approach is based on shadow coding to preserve the privacy in data transmission and ensure the recovery in data collection. [3]. A Pristine Clean Cabalistic Foruity Strategize Based Approach for Incremental Data Stream proposed a new technique called Cabalistic fortuity strategize based approach for Incremental data stream based PPDM. Our technique optimizes the privacy level by toughening the re-identification of original data without compromising the processing speed and data utility [4].Distributed anonymous data perturbation method for privacy- preserving data mining method propose a light-weight anonymous data perturbation method for efficient privacy preserving in distributed data mining [5]. This paper analysis five technique for privacy preserving Two Privacy-Preserving Approaches for publishing Transactional Data Streams, Efficient and Privacy-Preserving k-means clustering For Big Data Mining, SMC: A Practical Schema for Privacy-Preserved Data Sharing, A Pristine Clean Cabalistic Foruity Strategize Based Approach for Incremental Data Stream, Distributed anonymous data perturbation method for privacy-preserving data mining, This paper is organizes as follows.

Section I Introduction. **Section II** discusses Background. **Section III** discusses previous work. **Section IV** discusses existing methodologies. **Section V** discusses attributes and parameters and how these are affected on privacy preserving technique. **Section VI** proposed method and outcome result possible. Finally **section VII** Conclude this review paper.

.

PREVIOUS WORK DONE

In research literature, many privacy Preserving techniques are used for provide effectiveness, scalability.

Fosca Giannotti et al. (2018) [1] have proposed an Two Privacy-Preserving Approaches for publishing Transactional Data Streams in that developed an the proposed scheme presents a the sliding window is an effcient and commonly used approach to handle or mine continuously generated data streams. The above anonymization methods for data streams do not guarantee that every sliding window satisfies a given privacy requirement, and still lead to privacy leakage.

Zakaria Gheid et al. (2016) [2] has proposed an Efficient and Privacy-Preserving k-means clustering For Big Data Mining in this method, propose an efficient k-means protocol that aims to ensure total privacy protection under a given security model without using any cryptographic scheme.

Siyauan Liu et al.(2015) [3] has proposed an SMC: A Practical Schema for Privacy-Preserved Data Sharing over distributed data streams in the anonymity-based method that utilize a substitution schema to blur data providers information and thus preserves the correlation .

J.Gitanjali1,et al.(2014) [4] have proposed A Pristine Clean Cabalistic Foruity Strategize Based Approach for Incremental Data Streams in that method the proposed scheme focus on the value distortion techniques which uses multiplicative noise since additive noise can easily filtered out in many cases which leads to compromise in privacy.

Feng Li et al.(2008) [5] has worked on Distributed anonymous data perturbation method for privacy-preserving data mining in this method, propose a distributed anonymous perturbation method to realize efficient PPDM in a

distributed environment. Key privacy constraints of consistency, privacy and integrity, and robustness in the distributed process are defined.

IV. EXISTING METHODOLOGIES

Many Privacy preserving approach have been implemented over the last several decades. There are different technique that are implemented for providing privacy in data mining i. e SMC: A Practical Schema for Privacy-Preserved Data Sharing, Two Privacy-Preserving Approaches for publishing Transactional Data Streams, Efficient and Privacy-Preserving k-means clustering For Big Data Mining, Distributed anonymous data perturbation method for privacy-preserving data mining, A Pristine Clean Cabalistic Foruity Strategize Based Approach for Incremental Data Streams,

1) Two Privacy-Preserving Approaches for publishing Transactional Data Streams: The proposed scheme presents a the sliding window is an effcient and commonly used approach to handle or mine continuously generated data streams. The above anonymization methods for data streams do not guarantee that every sliding window satisfies a given privacy requirement, and still lead to privacy leakage [1].

$$InfoLoss(TSW_i) = \frac{\sum_{a \in I} sup(a) \times InfoLoss(a)}{\sum_{a \in I} sup(a)}, \qquad (1)$$

Where, sup(a) is the number of transactions in TSW i that contain a.

2)Efficient and Privacy-Preserving k-means clustering For Big Data Mining: In this method, propose, propose an efficient k-means protocol that aims to ensure total privacy protection under a given security model without using any cryptographic scheme, works attempted to reach a partial privacy protection by adding some noise before sharing data. However, the minimum error rate raised by the added noise is intolerable for applications needing a high accuracy level, such as healthcare. [2].

Algorithm 1: k-means clustering

1: Randomly select k cluster centers $\{c_1, ...c_k\}$.
2: **repeat**
3: Assign each data entity to the closest cluster center c_i.
4: Replace each cluster center c_i by the mean of th cluster i.
5: **until** cluster centers do not change.

3) SMC: A Practical Schema for Privacy-Preserved Data Sharing: The proposed scheme implement propose Shadow Coding which is privacy-preserving and data recoverable for data collection over the distributed data streams. First, map the locations of base stations into a matrix for each data provider. Second, align the matrices of the n data providers. Third, when a data provider sends data to the demander, the provider randomly selects a matrix in a Shadow Matrix set and adds it to the original data. During this data transmission, the ID information is protected by the ID-based encryption by a trusted third party.[3]

4) A Pristine Clean Cabalistic Foruity Strategize Based Approach for Incremental Data Streams:

In the proposed scheme focus on the value distortion techniques which uses multiplicative noise since additive noise can easily filtered out in many cases which leads to compromise in privacy. Proposed value distortion technique to protect the privacy by adding random noise from a gaussian distribution to the actual data. [4].

$$r_{ij} = \sqrt{s} \times \begin{cases} 1 & \text{with prob. } \frac{1}{2s} \\ 0 & \text{with prob. } 1 - \frac{1}{s} \\ -1 & \text{with prob. } \frac{1}{2s} \end{cases}$$

Where, the equation states that the random matrix *r* with *I* rows and *j* columns and the distribution occurs from lower limit to upper.

5) Distributed anonymous data perturbation method for privacy-preserving data mining: The proposed scheme the adaptive privacy-preserving summary protocol in this subsection is proposed to address constraint consistency. Two entities can generate random data: DC and DPs. Generating a random dataset by DPs would lead to unpredictable distribution. One solution is to generate the random dataset by DC, and deliver it to DPs separately. It is beyond doubt that DC can generate a random dataset with uniform distribution.

$$F_{X_1}{}'(a) = \frac{\int_{-\infty}^{a} f_Y(\omega_1 - z) f_X(z) \mathrm{d}z}{\int_{-\infty}^{+\infty} f_Y(\omega_1 - z) f_X(z) \mathrm{d}z},$$

where , *fY*() is the probability density function of *Yi* and *fX*() is the probability density function of *Xi*.

V.ANALYISIS AND DISCUSSION

Random Projection-Based Multiplicative Data Perturbation for Privacy Preserving Distributed Data Mining shows how much privacy perturbation technique can preserve when the adversary has different kinds of prior knowledge of the data and when the basic assumptions of this technique are not satisfied.[1].Privacy-preserving mining of association rules from outsourced transaction databases method would be interesting to enhance the framework and the analysis by appealing to cryptographic notions such as perfect secrecy [2]. Incentive compatible privacy preserving distributed classification in this method discuss how to incentivize data sharing in privacy-preserving distributed data mining applications. [3]. Privacy Preserving outsourced association rule mining on vertically partitioned databases uses comparison scheme is based on the symmetric homomorphic encryption scheme [4].

A random decision tree framework for privacy preserving data mining in that derive & the computation and communication cost, assuming k sites, n attributes, j instances, p class values, and m random trees, and then go through the security analysis. [5].

Proposed Methods and Techniques	Characteristics	
	Advantages	**Disadvantages**
Two Privacy Preserving Approaches For Publishing Transactional Data Strems	A major advantage of using the proposed approach is more efficient than sliding window.	The drawback of this method is Not working for Cryptography and data security.
Efficient and Privacy-Preserving k-means clustering For Big Data Mining.	A major advantage of using the proposed approach is It provides simple efficient and privacy preserving k-means protocol based on multiparty additive scheme, It provides the high level of security.	The drawback of this method This technique is encryption-based techniques are not suitable for big data sets.
SMC:A Practical Schema for Privacy-Preserved Data Sharing over Distributed Data Streams.	This Proposed method provide Data recovery. Data demander anonymity, Transmission synchronicity.	The proposed scheme is unable to the distributed data sharing problem in an asynchronous distributed environment.
A Pristine Clean Cabalistic Foruity Strategize Based Approach for Incremental Data Stream.	It provides simple efficient and privacy preserving k-means protocol based on multiparty additive scheme.	This method is not used for most robust encryption algorithm.
Distributed anonymous data perturbation method for privacy-preserving data mining	It provides Consistency, Privacy and Integrity, Robustness	This method use centralized PPDM methods based on perturbation.

TABLE 1: Comparison between different Practical approach schema for Privacy Preserving Techniques.

I. Proposed Methodology

Providing privacy in data mining is important and difficult task to analyse and discuss about various methods based on different parameters i.e. accuracy, quality, cost, time, flexibility, effectiveness, etc for different privacy preserving models. There are still problems which trouble in this field. New privacy preserving method called "Privacy Preserving for data mining " this method for more effective and more accurate privacy preserving is

propose here to overcome the problems of previous models.. The proposed method uses encryption decryption algorithm that is easier as compare to other methods.

Diagrammatic representation of proposed method is shown as follows:

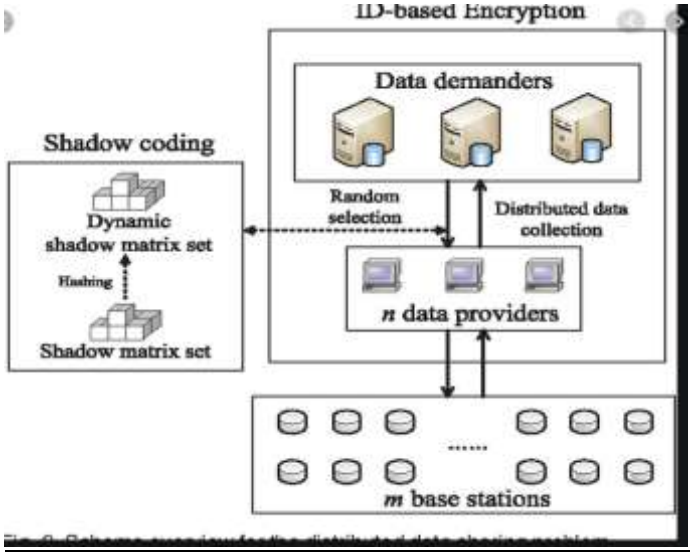

Fig1: Block Digram of SMC Technique

ACKNOWLEDGMENT AND POSSIBILITY RESULT

In this way the proposed method is provides experimental results for a the valid period of a shadow matrix set, that is, going to study how long to generate a new set also it provides experimental results for Collusion participants (CPs) exist in DPs in a semi-honest environment.

CONCLUSION

This paper focused on the proposed scheme implements a privacy-preserved data sharing problem is studied in the context of real-life distributed mobile phone networks. Formulate the problem as a distributed data sharing problem, and propose a shadow coding method with shadow matrix computation, which is privacy-preserving, efficient, and data-recoverable

FUTUREWORK

From observations of the proposed method the future concern is at will study attack models and address the distributed data sharing problem in an asynchronous distributed environment.

REFERENCES

Jinyan Wang, Chaoji Deng, Xianxian Li, Two Privacy-Preserving Approaches for Publishing Transactional Data Stream, vol. .6 Oxford: Clarendon,May 2018, pp.68–73.

Zakaria Gheid , Yacine Challal, “Efficient and Privacy-Preserving k-means clustering for big data mining”,, ieee trustcom -big datase ispa, vol. 6, 2016, pp. 271–350.

Siyuan Liu, Qiang Qu, Lei Chen, and Lionel M. Ni, “SMC: A Practical Schema for Privacy-Preserved Data Sharing over Distributed Data Streams,” ieee transactions on big data, vol. 02, pp. , June 2015.

J.Gitanjali1, Dr.J.Indumathi, Dr.N.Ch.Sriman Narayana Iyengar, “IEEE, Danish Mehmood, and David Lorenzi,” IEEE”, vol. 6, pp. 740–741,Vol.6 August 2010.

[5] Feng LI, Jin MA, Jian-hua LI, “Distributed anonymous data perturbation method for privacy-preserving data mining,” journal of zhejiang university science, April 2008.