Available online at http://www.journalcra.com

International Journal of Current Research
Vol. 7, Issue, 03, pp.13788-13790, March, 2015

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

ISSN: 0975-833X

RESEARCH ARTICLE

ENABLING MULTI LEVEL TRUST IN PRIVACY PRESERVING DATA MINING

***Akshay Gadekar, Pranav Nilawar, Amol Divekar and Nalawade, J. E.**

Computer Engineering, Savitribai Phule Pune University, Lonavala, Maharashtra 410401, India

ARTICLE INFO

Article History:
Received 21st December, 2014
Received in revised form
18th January, 2015
Accepted 05th February, 2015
Published online 31st March, 2015

Key words:
Data mining,
Privacy preserving data mining,
multilevel trust,
random perturbation.

ABSTRACT

To propose an additive perturbation based PPDM to address the problem of developing accurate models about all data without knowing exact information of individual values. To preserve privacy, the approach introduces random perturbation to individual values, before the data are published to third parties for mining purposes. In Existing System, the PPDM approach assumes single level trust on data miners. Under the single level trust, a data owner generates only one perturbed copy of its data with affixed amount of uncertainty. In proposed system, the PPDM approach introduces multilevel trust on data miners. Here different perturbed copies of same data are available to data miner at different trust levels & may combine these copies to jointly add additional information about original data & release the data is called diversity attacks. To prevent these attacks, using multilevel PPDM approach where random Gaussian noise is added to the original data with arbitrary distribution. So, the data miners will have no diversity gain in their joint reconstruction of the original data. This allows data owners to generate perturbed copies of its data on demand at arbitrary trust levels. It provides data owner very flexibility.

INTRODUCTION

Data mining is the extraction of hidden predictive information from large databases, is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. *Data* perturbation, a widely used and accepted Privacy Preserving Data Mining (PPDM) approach, it assumes single-level trust on data miners. This approach introduces some uncertainty about individual values before. The owner published or released the data to third parties for data mining purposes. In the single trust level (STL) a data owner generates only one perturbed copy of its data with affixed amount of uncertainty and this copy is released to third parties. There is new technique of Multilevel Trust (MLT) poses new challenges for perturbation-based PPDM. In the single-level trust scenario where only one perturbed copy is released, now multiple different perturbed copies of the same data are available to data miners at different trusted levels. The high level trusted data miner can access less perturbed copy; it may also have access to low trust level perturbed copies. Moreover, if a data miner could access multiple different perturbed copies through various other means, e.g., accidental leakage or colluding with others. By utilizing diversity across differently perturbed copies, the data miner may be able to produce a more accurate reconstruction of the original data than what is allowed by the data owner. We called it as a diversity attack. Preventing diversity attacks is the key challenge in solving the MLT-PPDM problem

Corresponding author: Akshay Gadekar,
Computer Engineering, Savitribai Phule Pune University, Lonavala, Maharashtra 410401, India.

2. Privacy Preserving Data mining

Data mining is the process of finding useful patterns in data. The objective of data mining is to use discovered patterns to help explain current behavior or to predict future outcomes. Several aspects of data mining process can be studied .These includes:-

1. Data gathering and storage
2. Data Selection and preparation
3. Model building and Testing
4. Interpreting and validating results

Data mining is under attack from privacy advocates because of a misunderstanding about what it actually is and a valid concern about how it is generally done**.** Privacy Preserving Data mining, proposes a number of techniques to perform the data mining tasks in a privacy-preserving way. These techniques generally fall into the following categories: data modification techniques, cryptographic methods and query auditing methods, randomization and perturbation-based techniques.

2.1 Individual privacy preservation

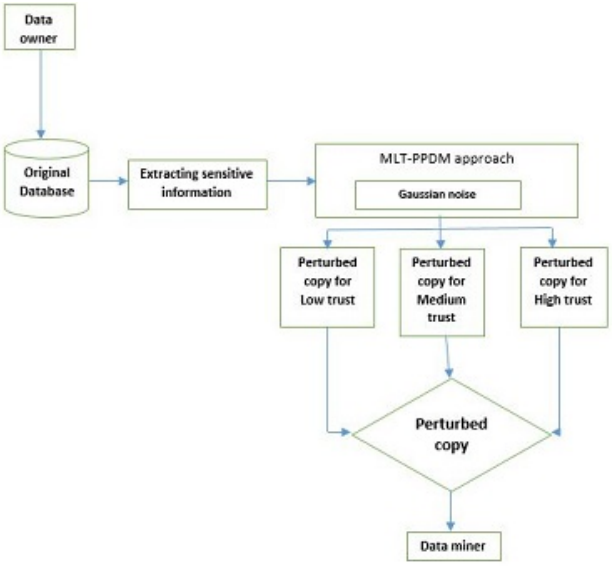
The primary goal of data privacy is the protection of personally identical information. In general, information is

considered personally identical if it can be linked, directly or indirectly, to an individual person. Thus, when personal data are subjected to mining, the attribute values associated with individuals are private and must be protected from disclosure. Miners are then able to learn from global models rather than from the characteristics of a particular individual

2.2. Collective privacy preservation

Protecting personal data may not be enough. Some- times, we may need to protect against learning sensitive knowledge representing the activities of a group. We refer to the protection of sensitive knowledge as collective privacy preservation. The goal here is quite similar to that one for statistical databases, in which security control mechanisms provide aggregate information about groups (population) and, at the same time, should prevent disclosure of confidential information about individuals. However, unlike as is the case for statistical databases, another objective of collective privacy preservation is to preserve (hide) strategic patterns that are paramount for strategic decisions, rather than minimizing the distortion of all statistics (e.g., bias and precision). In other words, the goal here is not only to protect personally identical information but also some patterns and trends that are not supposed to be discovered. In the case of collective privacy preservation, organizations have to cope with some interesting conflicts. For instance, when personal information undergoes analysis, processes that produce new facts about users' shopping patterns, hobbies, or preferences, these facts could be used in recommender systems to predict their future shopping patterns. In general, this scenario is beneficial to both users and organizations. However, when organizations share data in a collaborative project, the goal is not only to protect personally identical information but also to protect some strategic patterns. In the business world, such patterns are described as the knowledge that can provide competitive advantages, and therefore must be protected more challenging is to protect the knowledge discovered from confidential information (e.g., medical, financial, and crime information).

3. System Design



4. Model of Data Miners

In Example, we assume that there are two types of data miners. The first type refers to the conscious data miners. These miners always act legally in that they perform regular data mining tasks and would never intentionally breach the privacy of the data. On the other hand, malicious data miners would purposely breach the privacy in the data being mined. Malicious data miners come in many forms. We focus on a particular sub-class of malicious miners. That is, malicious data miners follow standards but are curious, they follow proper protocols and standard procedures, but they may perform some analysis (i.e., they are curious) to discover private information. This kind of curious (nevertheless malicious) behavior is most common and has been widely adopted as an adversary model in the literature. This is because, in reality, a workable system must be net both the conscious and the malicious data miners. For example, in an online bookstore a data miner may use the association rules of purchase records to make recommendations to its customers (data providers). In addition, this data miner can also perform some analysis to discover private information. This data miner, as a long-term agent, requires large numbers of data providers to collaborate with. In other words, even a malicious data miner desires to build a reputation for trustworthiness. Thus, honest but curious behavior is an appropriate choice for many malicious data miners.

5. Batch Generation

In batch generation, the data owner determines the M trust levels, & generates M perturbed copies of the data in one batch. In this case, all trust levels are predefined & all are given when generating the noise. It generates noise to the perturbed copies of the dataset. The Noise Generation is based on the Gaussian Noise process. Let G1 through GL be L Gaussian random variables. They are said to be jointly Gaussian. It follows linear combination of multiple independent Gaussian random variables. G1 through GL are jointly Gaussian. It is a linear combination of them & also a Gaussian random variable.

5.1. Parallel Generation

In this method the components of noise Z, i.e., Z1 to ZM, are generated simultaneously based on the probability distribution function.

1. Input: X, KX, and σ Z1 to σZM
2. Output: Y
3. Construct KZ with KX and σZ1 to σZM
4. Generate Z with KZ
5. Generate Y = HX + Z
6. Output Y

5.2. Sequential Generation

This method sequentially generates M independent noise Z1, and (Zi −Zi−1) for i from 2 to M. The large memory requirement of Algorithm 1 motivates to seek for a memory efficient solution. Instead of parallel generation, sequentially

generating noise Z1 to ZM, each of which a Gaussian vector of N dimension.

1. Input: X, KX, and σZ1 to σZM
2. Output: Y1 to YM
3. Construct Z1 ∼N(0, σZ1KX)
4. Generate Y1 = X + Z1
5. Output Y1
6. for i from 2 to M do
7. Construct noise Z2 ∼N(0, (σZi − σZi−1)KX)
8. Generate Yi = Yi−1 + Z2
9. Output Yi
10. end for

5.3. On Demand Generation

As opposed to the batch generation, new perturbed copies are introduced on demand in this scenario.

1. Input: X, KX, σZ1 to σZM, and values of Z′:v1
2. Output: New copies Z″
3. Construct KZ with KX and σZ1 to σZM
4. Extract KZ′ ,KZ″Z′ , and KZ″ from KZ
5. Generate Z″ as a Gaussian with mean and variance
6. for i from L + 1 to M do
7. Generate Yi = X + Zi
8. Output Yi
9. end for

Assume L(L < M) existing copies of Y1 to YL, so that the data owner, upon requests, generates additional M−L copies of YL+1 to YM. Among three techniques on-demand generation offers data owner's maximum flexibility where data owners generate perturbed copies of the data at arbitrary trust levels on-demand.

6. Conclusion

In this paper, using additive perturbation based PPDM approach for multilevel trust is used for providing better flexibility & security. MLT-PPDM allows data owners to generate differently perturbed copies of its data for different trust levels. This method address the challenge of preventing the data miners from combining copies at different trust levels to jointly reconstruct the original data more accurate than what is allowed by the data owner. This challenge is addressed by properly correlating noise across copies at different trust levels. So, the data miners will have no diversity gain in their joint reconstruction of the original data. Finally, our solution allows data owners to generate perturbed copies of its data at arbitrary trust levels on-demand. This property offers the data owner maximum flexibility.

REFERENCES

Aggarwal, C. 2008. "Privacy and the Dimensionality Curse," Privacy-Preserving Data Mining, pp. 433-460

Agrawal, R. and R. Srikant, 2000. "Privacy Preserving Data Mining", Proc. ACM SIGMOD *Int'l Conf. Management of Data* (SIGMOD '00).

Kifer, D. and J.E. Gehrke, 2006 "Injecting Utility Into Anonymized Datasets," Proc. ACM SIGMOD *Int'l Conf. Management of Data*.

Xiao, X. and Y. Tao, 2007. "M-Invariance: Towards Privacy Preserving Re-Publication of Dynamic Datasets," Proc. ACMSIGMOD Int'l Conf. Management of Data.

Xiao, X., Y. Tao, and M. Chen, 2009."Optimal Random Perturbation at Multiple Privacy Levels," Proc. *Int'l Conf. Very Large Data Bases*.

****