ijecs
open access www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume – 4 Issue - 12 December, 2015 Page No. 15392-15396

Providing Multi Security In Privacy Preserving Data Mining

[1]s.Nathiya, [2]c. Kuyin, [3]j.D.Sundari

[1] Assistant Professor, PG Department of Computer Sciecne,
Jayaraj Annapackiam College for women (Autonomous), Periyakulam. Theni Dt.
s.nathiyamsc@gmail.com
.&
[2]Assistant Professor, Department of Commerce with CA,
Jayaraj Annapackiam College for women (Autonomous), Periyakulam. Theni Dt.
mesiachinna@gmail.com
&
[3] Assistant Professor, Department of Mathematics,
Jayaraj Annapackiam College for women (Autonomous), Periyakulam. Theni Dt.
jd.sundari@yahoo.com

Abstract: Privacy Preserving Data Mining (PPDM) addresses the problem of developing accurate models about aggregated data without access to precise information in individual data record. In our setting, the more trusted a data miner is the less perturbed copy of the data (original) it can access.

Under this setting, a malicious data miner may have access to differently perturbed copies of the fake data through various means, and may combine these diverse copies to jointly infer additional information about the fake data and the data owner does not intend to release.

To Preventing diversity attacks is the key challenge of providing multi security in Privacy Preserving Data Mining services. We address this challenge by properly correlating perturbation across copies at different trust levels. We prove that our solution is robust against diversity attacks with respect to our privacy goal. Our solution allows generating perturbed copies of fake data for arbitrary trust levels on demand.

Keywords: Diversity attack, perturbed approach, Data miner, PPDM, Multi Security, Attribute based key, Hash Key.

2. INTRODUCTION

Privacy Preserving Data Mining is a method which ensures privacy of individual information during mining. Privacy Preserving Data Mining (PPDM) is used to extract relevant knowledge from large amount of data and at the same time protect the sensitive information from the data miners.

The problem in privacy-sensitive domain is solved by the development of the Multi security in Privacy Preserving Data Mining (MS-PPDM) where multiple differently perturbed copies of the fake data are available to data miners at different trusted levels.

Data perturbation, a widely employed and accepted Privacy Preserving Data Mining (PPDM) approach, tacitly assumes single-level trust on data miners. This approach introduces uncertainty about individual values before data are published or released to third parties for data mining purposes. Under the single trust level assumption, a data owner generates only one perturbed copy of its data with a fixed amount of uncertainty. This assumption is limited in various applications where a data owner trusts the data miners at different levels.

The government or a business might do internal (most trusted) data mining, but they may also want to release the data to the public, and might perturb it more. The mining department which receives the less perturbed internal copy also has access to the more perturbed public copy. It would be desirable that this department does not have more power in reconstructing the data by utilizing both copies than when it has only the internal copy.

Conversely, if the internal copy is leaked to the public, then obviously the public has all the power of the mining department. However, it would be desirable if the public cannot reconstruct the original data more accurately when it uses both copies than when it uses only the leaked internal copy.

3. MULTI SECURITY IN PPDM

This new dimension of Multi Security poses new challenges for perturbation based PPDM. In contrast to the single-level trust scenario where only one perturbed copy is released, now multiple differently perturbed copies of the fake data are available to data miners at different trusted levels.

The more trusted a data miner is the less perturbed copy it can access which means original data. Moreover, a data miner could access multiple perturbed copies through various other means, e.g., accidental leakage or colluding with others.

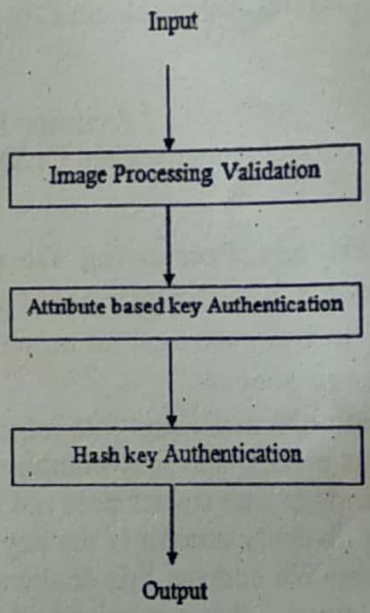
By utilizing diversity across differently perturbed copies, the data miner may be able to produce a more accurate reconstruction of the original data than what is allowed by the data owner. To refer this attack as a diversity attack.

It includes the colluding attack scenario where adversaries combine their copies to mount an attack; it also includes the scenario where an adversary utilizes public information to perform the attack on its own. Preventing diversity attacks is the key challenge in solving the Multi Security in PPDM problem.

Privacy preserving data mining help to achieve data mining goals without scarifying the privacy of the individuals and without learning underlying data values. Privacy-preserving data mining (PPDM) refers to the area of data mining that seeks to safeguard sensitive information from unsolicited or unsanctioned disclosure. Privacy is becoming an increasingly important issue in many data mining applications.

In this research, we address this challenge in enabling Multi security in PPDM services. In particular, we focus on the additive perturbation approach where random Gaussian noise is added to the fake data with arbitrary distribution, and provide a systematic solution.

Through a one-to-one mapping, our solution allows to generate distinctly perturbed copies of its data according to different trust levels. Defining trust levels and determining such mappings are beyond the scope of this research.


Input
Image Processing Validation
Attribute based key Authentication
Hash key Authentication
Output

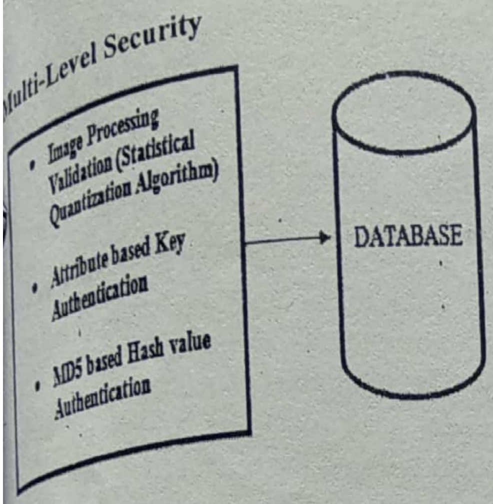
4. METHODOLOGY

The new dimension of Multi Security poses new challenges for perturbation-based PPDM. In contrast to the single-level trust scenario where only one perturbed copy of the fake data is released, now multiple differently perturbed copies of the fake data are available to data miners at different trusted levels. The more trusted a data miner is the less perturbed copy it can access which means the original data.

Preventing diversity attacks is the key challenge in solving the Multi Security in PPDM problem. In this research we address these challenges in enabling Multi Security in PPDM services.

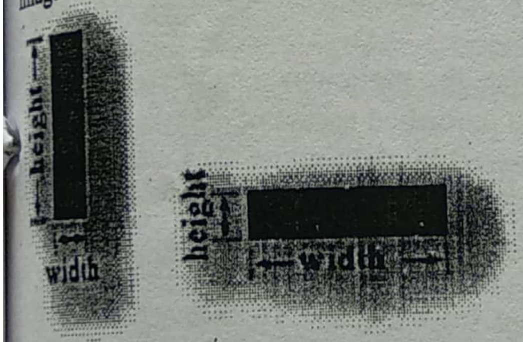
In this paper there are three different levels are used to indicate who is the trusted data miner and the malicious data miner. The trusted data miner can access the original data which is allowed by the data owner who is called the valid user. Then the malicious data miner can access differently perturbed copies of the fake data which is not allowed by data miner. The malicious data miner is also called the invalid user.

DOI: 10.18535/Ijecs/v4i12.50

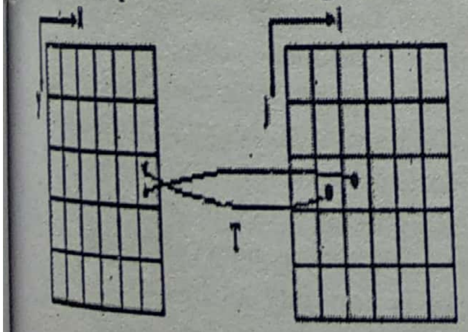

Multi-Level Security
Image Processing Validation (Statistical Quantization Algorithm)
Attribute based Key Authentication
MD5 based Hash value Authentication
DATABASE

4.1 Image processing

The first level is Image Processing. Some conditions are used to verify the valid user. First, to get the input as an image from the user. And to extract User data who is Valid from data base. To validate the user's input image and the original Image which has the same Height andWidth.



If it is valid means continue to validate both image pixel values. Again if it is valid means the User Input is Valid and goes to the next level.



4.2 Attribute based Key Authentication

Authentication Methods

There are several authentication methods including: password, public-key, anonymous, remote and certificate-based authentication. Some of the Followings,

Password authentication - the oldest and the easiest to implement. It includes reusable passwords, one-time passwords, challenge response passwords, and combined approach passwords.

Public Key Authentication – This requires each user of the scheme to first generate a pair of keys and store each in a file. Each key is usually between 1024 and 2048 bits in length. Public-private keys pairs are typically created using a key generation utility. The server knows the user's public key because it is published widely. However, only the user has the private key.

The second level is attribute based key authentication. The two methods are used in this level.

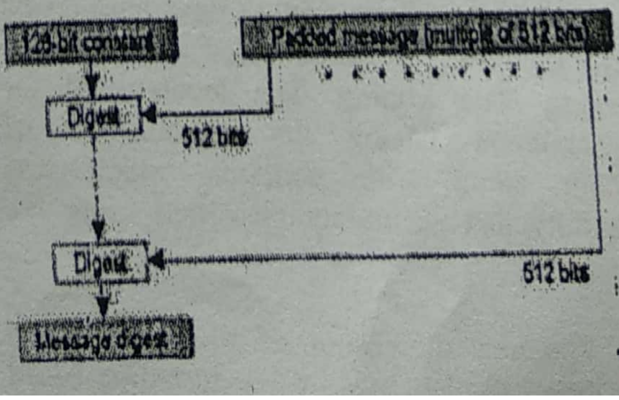
i) Key Allocation: In this method to get the input from valid user and to allocate the particular key to the particular user who is valid.

ii) Key Authentication: First to receive the user key. And to validate with the original User key and to submitted the particular key. If key is matched it is used to extract the original data otherwise not. Suppose if it is identifying the Malicious Key, to call the readymade fake data.

4.3 Hash Key Authentication

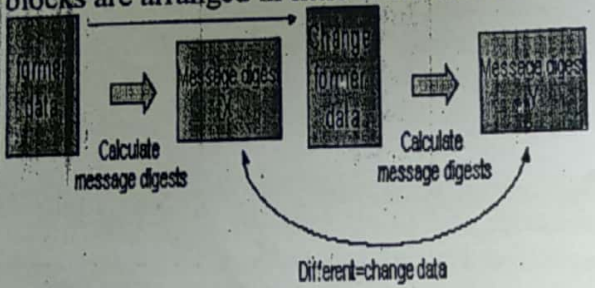
The third level is the hash key authentication. In this level is used to check the above two levels are completed successfully then proceeding with this step. To receive all the user inputs in the first and second levels. Then to perform hash function it to make encrypt (or) decrypt the user value using MD5.

MD stands for message digest. MD5 processes a variable-length message into a fixed-length output of 128 bits. The input message is broken up into chunks of 512-bit blocks (sixteen 32-bit little endian integers); the message is padded so that its length is divisible by 512.



DOI: 10.18535/Ijecs/v4i12.50

The padding works as follows: first a single bit, 1, is appended to the end of the message. This is followed by as many zeros as are required to bring the length of the message up to 64 bits fewer than a multiple of 512. The remaining bits are filled up with a 64-bit little endian integer representing the length of the original message, in bits, modulo 2^{64}. The bytes in each 32-bit block are big endian, but the 32-bit blocks are arranged in little endian format.



The main MD5 algorithm operates on a 128-bit state, divided into four 32-bit words, denoted *A*, *B*, *C* and *D*. These are initialized to certain fixed constants. The main algorithm then operates on each 512-bit message block in turn, each block modifying the state. The processing of a message block consists of four similar stages, termed *rounds*; each round is composed of 16 similar operations based on a non-linear function *F*, modular addition, and left rotation. There are four possible functions *F*; a different one is used in each round:

$$
\begin{aligned}
F(X,Y,Z) &= (X \wedge Y) \vee (\neg X \wedge Z)\\
G(X,Y,Z) &= (X \wedge Z) \vee (Y \wedge \neg Z)\\
H(X,Y,Z) &= X \oplus Y \oplus Z\\
I(X,Y,Z) &= Y \oplus (X \vee \neg Z)
\end{aligned}
$$

5. RESULT AND DISCUSSION

The proposed approaches have been tested successfully on the levels extended in the Privacy Preserving Data Mining. The levels used for experimentation having the combination of algorithm along with methods which was successfully and accurately classified using the approach.

The new dimension of Multilevel Trust poses new challenges for perturbation-based PPDM. In contrast to the single-level trust scenario where only one perturbed copy is released, now multiple differently perturbed copies of the data are available to data miners at different trusted levels.

In this system three levels were implemented for multi security. Multi Security leads to the multilevel trust for the users. The more trusted data miner can access the original data when completed all the three levels successfully. The malicious data miner can access the readymade fake data set which has the unwanted noise without the original data.

6. CONCLUSION

Preventing diversity attacks is the key challenge in solving the Multi security in PPDM problem. In this research we address these challenges in enabling multi security in Privacy Preserving Data Mining services. The trusted data miner can access the original data which is allowed by the data owner who is called the valid user.

Then the malicious data miner can access differently perturbed copies of the fake data which is not allowed by data miner. The malicious data miner is also called the invalid user.

So, the solution allows generating distinctly perturbed copies of fake data according to different trust levels. To defining the trusted levels are beyond the scope of this research.

7. FUTURE WORK

To believe that multi security in privacy preserving data mining can find many applications. In this work took the initial step to enable MS-PPDM services. Many interesting and important directions were worth exploring. Studying the MS-PPDM problem under a relaxed setting where adversaries may also carry out diversity attacks is certainly an interesting future direction.

So, the fruitful direction for future data mining research will be the development of techniques that incorporate privacy concerns, and then to be implemented using biometrics, adding extra levels for the strong security to preserve the original data from the malicious data miner.

8. REFERENCES

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

2.K. Liu, H. Kargupta, and J. Ryan, "Random Projection-Based Multiplicative Data Perturbation for Privacy Preserving Distributed Data Mining," IEEE Trans. Knowledge and Data Eng., vol. 18,no. 1, pp. 92-106, Jan. 2006.

3. A. Evfimievski, J. Gehrke, and R. Srikant, "Limiting Privacy Breaches in Privacy Preserving Data Mining," Proc. ACM Symp. Principles of Database Systems, 2003.

4. A. Iliev and S. Smith, "More Efficient Secure Function Evaluation Using Tiny Trusted Third Parties," Technical Report TR2005-551, Dept. of Computer Science, Dartmouth Univ., 2005.

5. K. Chen and L. Liu, "Privacy Preserving Data Classification with Rotation PertuK. Chen and L. Liu, "Privacy Preserving Data Classification with Rotation Perturbation," Proc. IEEE Fifth Int'l Conf. Data Mining, 2005.

6. R.Agrawal, A. Evfimievski, and R. Srikant, "Information Sharing across Private Databases," Proc. ACM SIGMOD Int'l Conf.Management of Data, 2003.
7.G. Wang, Z. Zhu, W. Du, and Z. Teng, "Inference Analysis in Privacy-Preserving Data Re-Publishing," Proc. Int'l Conf. Data Mining, 2008.

8. Umesh Kumar Singh, Bhupendra Kumar Pandya , Keerti Dixit, "An Overview on Privacy Preserving Data Mining Methodologies", International Journal of Engineering Trends and Technology, October 2011.

9. Xiaolin Zhang and Hongjing Bi, "Research on privacy preserving classification data mining based on random perturbation", International Conference on Information Networking and Automation (ICINA) , pp v1 -173-v1-178, oct 2010.

AUTHOR PROFILE

[1] **S. Nathiya** , Assistant Professor in the PG department of Computer Science, Jayaraj Annapackiam College for women (Autonomous), Periyakulam, Theni Dt, Tamilnadu, India. She received her UG degree from Sri Kaliswari College of Arts and Science, Sivakasi, Tamilnadu, India. PG degree from Jayaraj Annapackiam College for women, Theni Dt, Tamilnadu, India. And M.Phil degree from Nadar Saraswathi College of Arts and Science, Theni Dt, Tamilnadu, India. She has published two papers in reputed journal.

[2] **C. Kuyin** , Assistant Professor in the department of commerce with Computer Application, Jayaraj Annapackiam College for women (Autonomous), Periyakulam, Theni Dt, Tamilnadu, India. She received her UG degree from Manonmaniam Sundaranar University ,Tirunelveli, Tamilnadu, India.PG degree from Sri Ramasamy Naidu Memorial College,Sattur. And M.Phil degree from Manonmaniam Sundaranar University ,Tirunelveli, Tamilnadu, India.

[3] **J. D. Sundari** , Assistant Professor in the department of mathematics, Jayaraj Annapackiam College for women (Autonomous), Periyakulam, Theni Dt, Tamilnadu, India. She received her UG and PG degree from Jayaraj Annapackiam college for women (Autonomous), Periyakulam, Theni Dt, Tamilnadu, India. B.Ed degree from women's university college of education, Kodaikannal, Dindigul Dt, Tamilnadu, India. M.Ed degree from Sakthi College of teacher training and research Institution, Oddanchatram, Dindigul Dt, Tamilnadu, India.