Enhanced Batch Generation based Multilevel Trust Privacy Preserving in Data Mining

B.Anitha
M.Tech.(SWE)
Dept.of CSE
Kakatiya Institute of Technology & Science
Warangal-15,A.P.,India.

B.Hanmanthu
Assistant Professor,
Dept.of CSE
Kakatiya Institute of Technology & Science
Warangal-15,A.P.,India.

B.Raghu Ram
Assistant Professor,
Dept.of CSE
Kakatiya Institute of Technology & Science
Warangal-15,A.P.,India.

ABSTRACT

The motivation of Privacy Preserving Data Mining (PPDM) is to obtain valid data mining results without access to the original sensitive information. The different privacy preserving technique on Perturbation based PPDM approach introduces random perturbation to individual values to preserve privacy before data are published. This proposed work is based on perturbation based privacy preserving data mining. Here random perturbation approach is applied to provide privacy on the data set. Previously privacy is limited to single level trust in providing privacy to the data but now it is enhanced to multi level trust. The problem with existing multi level trust PPDM algorithms is that they fail to protect form non linear attacks. Considering that this proposed work make uses enhanced batch generation to provide privacy in the multi level trust in which data will perturb multiple times so that it can avoid non linear attacks.

Keywords: Privacy Preserving Data Mining, Multi Level Trust, Batch generation based perturbation.

1. INTRODUCTION

Data perturbation, a widely utilized Privacy Keeping Data Seek approach, tacitly assumes single-level rely on information or data miners. This strategy introduces doubt about specific values before data tend to be published in addition to be revealed for the third get-togethers for information mining specifications. Under the actual single rely on level presumption information owner generates only one perturbed information of their data using a fixed quantity of uncertainty. This premise is fixed in various applications in which a data administrator trusts the details miners in various quantities. We present below a two rely on level scenario such as a motivating case.

The government or perhaps a business may well do internal (most trusted) information mining, however they will often release the details to all people, and may well perturb that more. The mining department which regularly receives the actual less perturbed internal copy now offers access for the more perturbed community copy.

It is going to be desirable that department don't get more power in reconstructing your initial data by means of both replicates in comparison with when it's got only the inside copy.

Nonetheless, if the inside copy is going to be leaked for the public, then naturally everyone has the power within the mining scale. However, it could be desirable if the public are not able to reconstruct the initial data accurately any time by using both replicates than by using only the actual leaked internal copy.

This particular new sizing of Multilevel Believe offers new troubles for perturbation-based PPDM. In contrast to the single-level rely on scenario where only one perturbed information is introduced, now a number of differently perturbed copies on the same data can be found to information miners in various trusted levels. Greater dependable a files miner is going to be, the fewer perturbed copy it could actually access; additionally, it can have use of the perturbed reports at reduced trust levels. Additionally, an information miner can easily access a number of perturbed replicates through other sorts of means, unintentional leakage in addition to colluding applying others.

By using diversity around differently perturbed replicates, the information miner could possibly produce an accurate reconstruction within the original information than what on the planet is allowed from the data administrator. We speak about this attack such as a diversity invasion. It is made of the colluding invasion scenario during which adversaries incorporate their replicates to guide an invasion; it also includes the actual scenario during which an adversary utilizes community information to try and do the attack with no treatment. Preventing multiplicity attacks will be the key issue in getting rid of the challenge.

In this particular paper, all of us address this particular challenge with enabling solutions. In certain, one should concentrate on the additive perturbation strategy where hit-or-miss Gaussian noises is put into the original data using arbitrary supply, and provide a systematic remedy. Through a one-to-one mapping, this remedy allows a data owner to create distinctly perturbed reports of their data based on different rely on levels. Defining rely on levels in addition to determining this kind of mappings usually are beyond your scope of this paper.

2. PREVIOUS WORK

The information technology evolution has enabled an organization (e.g., hospitals, institutes) to collect large amount of sensitive data (e.g., medical records, transaction history), which is referred as *microdata*. To facilitate research, these organizations should provide public access to their microdata, which, may pose a risk to individual privacy. For example, assume Census Bureau which maintains an online database for answering queries of count on the microdata T, which has three columns, Age, Zipcode, and Income (Name is included to provide row referencing). Consider a case who knows the age 20 and zipcode 15000 of Alice, and the fact of Alice is involved with T. The permission to create digital or hard copies of this work for personal use is granted without fee such that the copies are not created or distributed for the sake of direct commercial advantage, the VLDB notice of copyright and the publication title and the date appear.

Various output perturbation techniques are available in the literature of statistics. Those are not based on a rigorous privacy definition. To overcome this, authors in [1] developed *differential privacy* principle, and provides it to avoid queries that reveals sensitive information. Assume Q as the set of previously answered queries. Provided a new query q, the database finds whether {q} [Q violating differential privacy. If yes, then q is rejected; otherwise, the database results a noisy result. As shown, this approach confirms that an adversary may recover any sensitive information with low probability, even if one has audited all the queries results in history

Despite the art state, differential privacy has two drawbacks that reduces the practical applicability. First, there is no existing solution for testing differential privacy. The difficulty stems from the so-called computation L1 *sensitivity*, which is an important component in verifying differential privacy. The best efforts are because of [1], who shows out several special cases where L1 sensitivity can be determined. Same attempts have already been made. Unfortunately, the calculation problem is still open. In other terms, differential privacy is inapplicable when arbitrary queries are allowed. The second defect in differential privacy exists in all the previous solutions of output perturbation. When the database rejects a query, it returns nothing. This generates negative user experience, because a legitimate user would spend a lot of time with different queries for getting an answer. Even worse, differential privacy provides only finite queries. In other terms, the statistical database should go offline, and all the future queries are directly refused. [1]

Privacy is becoming an important issue in various data-mining applications such as health care, security, financial, and other types of sensitive data. It has become important in counter terrorism and homeland defence-related applications. These applications require creating of profiles, constructing of social network models, and detecting terrorist communications from privacy sensitive data.

However, combining such diverse data sets may violate the privacy laws. Even though health organizations are allowed to give data as long as the identifiers (e.g., name, SSN, address, etc.,) are eliminated, it is considered unsafe since re-identification attacks can be constructed for linking different sets of public data to identify the initial subjects. This requires a well designed technique that provides careful attention for hiding privacy-sensitive information, while securing the inherent statistical dependencies which are important for the applications of data mining.

The problem we are interested in can be defined as follows: Assume there are N organizations 01; O2; . . .; ON; each organization Oi has a database DBi of private transaction. A third party data miner want to study union statistical properties of these databases SN. Organizations are satisfied with this, but they are strict to disclose raw data. How could a data miner perform analysis of data without compromising on the privacy data? This is referred as the census problem. In this scenario, the data is distorted and its new representation is shown; everybody has arbitrary access to the published data. Fig. illustrates a distributed two party input case and a single-party-input case.

A randomized multiplicative data perturbation technique is considered for this problem in this paper. It is motivated by the work provided somewhere else that pointed out some of the additive random perturbation problems. This paper explores using multiplicative random projection matrices for constructing a new data representation. The transformed data is given to the data miner. It can be shown that the inner product and distance of Euclidean are preserved in the new data. The approach is dependent on the Johnson-Linden Strauss lemma which states that any point sets in m-dimensional Euclidean space can be provided into k-dimensional subspace, such that any two-point pair-wise distance is maintained within an arbitrary small factor. Therefore, by data projecting onto a random subspace, we can change its original form while keeping much of underlying distance-related characteristics of statistics.

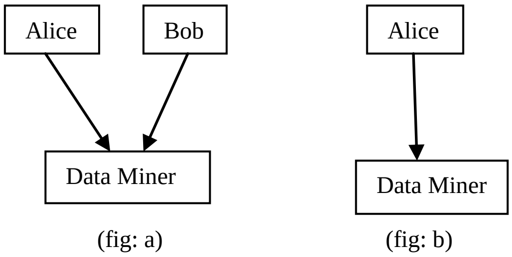

Fig: a: Distributed Two-Party-input Computation Model

b: Single-Party-input Computation Model

Data Perturbation

The approaches of Data perturbation can be grouped into two categories: the probability distribution approach and the distortion of value approach. The probability distribution approach replaces the data with a sample from the same distribution or by the distribution itself, and the distortion of value approach perturbs elements of data or attributes by either additive noise, multiplicative noise, or some other procedures of randomization.

In this paper, we focus on the distortion of value approach. An additive data perturbation technique for building decision tree classifiers is proposed in work. Some random noise is added to each data element for randomization independently from a known distribution such as Gaussian distribution. The data miner reconstructs the original data distribution from its perturbed version and builds the classification models.

The additive noise drawback makes one think about using multiplicative noise for protecting the data privacy. Two basic multiplicative noise forms have been well studied. One is to multiply each data element with a random number that is truncated Gaussian distribution with a mean one and small variance. Another one is a logarithmic transformation of the data first, add multivariate Gaussian noise which is predefined, and then take the antilog of the data noise-added. In general, the first method is good if the data miner only wants to make minor changes to the original data; the second method provides higher security than the first one while maintaining the data utility in the scale of log.

Data Swapping

The data swapping basic idea, which was first proposed in [2], is to transform the database by practically switching attributes subset between selected record pairs so that the lower order frequency counts are preserved and the data confidentiality is not compromised. This technique could be categorized under the data perturbation category. A large variety of refinements and data swapping applications have been addressed since its initial appearance.

K-Anonymity

The k-Anonymity model calls the problem that a data owner wishes to share a person-specific data collection without releasing the identity of an individual person. To reach this goal, generalization of data and techniques of suppression are used to protect the private sensitive information. All the attributes (referred as quasi-identifier) in the private database that can be used for linking the external information would be found, and the data is revealed only if the information for each individual contained in the revelation cannot be distinguished from at least k other people.

Secure Multiparty Computation

The Secure Multiparty Computation (SMC) technique considers the evaluating problem of a function of the secret inputs from one or two or more parties, so that no party learns anything except the function designated output. A large body of protocols of cryptographic, including evaluation of circuit protocol, oblivious transfers, encryption of homomorphism, and a commutative encryption, serves as the building blocks of the Secure Multiparty Computation (SMC). The work has provided a broad view of SMC framework and its applications in the field of data mining. The work provides a detailed and rigorous introduction to SMC. It was earlier shown that any functions that are to be expressed by the circuit of arithmetic are privately computable using a protocol of generic circuit evaluation. However, the communication and complexity of computational doing, makes this general approach infeasible for large volumes of data sets

Distributed Data Mining

The distributed data mining (DDM) approach considers data mining models computations and "patterns" extraction at a given (chosen) node by providing only the minimal required information among the other participating nodes. The work proposes a paradigm for clustering of distributed privacy data which is sensitive in an unsupervised or semi-supervised cases. In this algorithm, each site of local data builds a general model and transmits only the parameters of the model to the site which is central and where a global clustering model is constructed. An algorithm of distributed privacy-preserving for Bayesian network parameter learning is reported elsewhere.

Rule Hiding

The main objective behind the rule hiding is to transform the database so that all the underlying patterns can be discovered and the sensitive rules are masked. The work has already given a formal proof that the sanitization of optimality is an NP-hard problem for the hiding of the large volumes of item sets which are sensitive in the context of association rule mining. For the same reason, some of the heuristic approaches have been provided to address the issues of complexity.

For example, the association rule hiding technique which is perturbation-based is implemented by changing a selected set of 1-values to 0-values or vice versa such that the frequent sets of item that generates the rules are hidden or with the support of sensitive rules is lowered to a user-specified threshold. The association rule hiding approach which is blocking based replaces certain number of attributes of the data with a question mark. In this case, the minimum support and confidence will be modified into a minimum interval. As long as possible if the support and/or the confidence of a sensitive rule lie below the middle in the two ranges, the data confidentiality is expected to be protected. [2]

An issue that has gained crucial importance most recently is the ability to cope with the requirement to share data across multiple enterprises on one hand and the contradictions between security and privacy on the other hand. This type of collaboration helps in the relevant trends detection and the data anomalies. The collaboration requirement is most often mandated by legislations. For example in the USA, the Act of patriotism requires banks for the analysis of the transaction data of the customer for anti-money laundering and other purposes. On the other hand, the Graham-Leach-Bailey Act prohibits, the data sharing for any other different purposes. This requirement could even be extended to the interactions between two or more lines of business (LOBs) in a company in some of the cases. In the same way, in the field of health care scenario, doctors and companies of insurance need to protect the privacy of individual patient data, while the community of the health care can gain more if that individual patient data can be pooled and analysed for common good.

In this paper, a solution is presented for privacy preserving data sharing and mining for such various application areas. This solution provides the data to reside with the enterprises and to be processed in a federated setup in secure privacy but with the resource constrained coprocessors like the IBM PCIXCC. The data transmission to the coprocessors is encrypted and is then protected from the eavesdroppers (malicious data miners). Again, since the coprocessor is secure privacy, the data which is decrypted in the process before and which is processed later inside is protected from those sharing of the data and with any third party administering the solution.

Further many operations can be performed on the data such as the data can be joined, mined and analysed in its initial plain text. With the resource constrained coprocessors if one needs to facilitate mining, this has developed and embedded data mining methodologies. These provide the required ability to detect trends and anomalies in the shared sets of data. These are run on the data managed with the a light weight database engine with secure federation extensions.

Anti Money Laundering (AML) is a possible application area in an inter bank service centre for due diligence. Money Laundering practically involves money transfer from one account to another account, often spanning finance establishments. To detect this, these financial establishments will have to collaborate and try to determine the suspicious patterns in the database transactions. However lately, they are also concerned about the privacy of their individual account data. While maintaining privacy, a solution for sharing and mining of data would help in these cases.

In many of the application areas, another application area is credit rating in an intra bank service centre which has got multiple LOBs. Here there are two contradicting imperatives, namely, data analysing from multiple LOBs to known customers better and to keep the matter in mind that there are legislations which limit LOBs from sharing of the data depending on the business line. [3]

While protecting the identity of individuals, *K-anonymization* is a promising approach to publishing of the data. In the database the data holder has a full table of the form *D* (Explicit identifier, quasi identifier, Sensitive attribute). Explicit identifier has got identifying information (such as SSN and names). Quasi identifier (such as the date of birth, gender, and zip code) does not reveal individual identity, but can also be used to link to another person or an explicit individual identity in some other external sources. The Sensitive attributes consists of specific information of another person (such as medical and DNA entries). Instead of the original table *D* publishing, in the database, the data holder publishes in an anonymized release *R*(Quasi identifier *QID*, Sensitive attribute), where *QID* is a Quasi identifier which is *k-anonymized* version .Each record should belong to an equivalence class of size *k* at the least and all the records are made indistinguishable in an equivalence class with respect to Quasi identifier *QID* by hiding some of the details. In a *k*-anonymized release, if an individual account is linked to a record through Quasi identifier *QID*, it is then also linked to at least k¡1 of other records.

Modifications are not made to the sensitive attributes because the data usefulness depends on the exact information of this sensitive attribute. *K-anonymization* has been basically studied for a single static data release. In general, however, new data arrival comes continuously and up-to-date data has required to be published to the researchers in an obedient timely manner.

One approach to this process is to anonymize and publish the new arrived records individually each time when they arrive. This approach has a disadvantage of severe data distortion because of the small increments are anonymized in-dependently. Moreover practically, it is more difficult to analyse a whole collection of independently anonymized sets of data.

For example, if the country name (e.g., India) is used in the release for the records of the first month and if the city name (e.g., Hyderabad) is used in the release for the records of the second month, counting the whole total number of persons born in Hyderabad is impossible. Rather than this, another approach is to enforce the later release than the previous release to be no more specialized. Each subsequent release gets increasingly distorted, even if more new data are available [4]. This is the major drawback.

3. PROPOSED WORK

It is very necessary to publish an individuals personal information for research purposes. For example, a hospital may reveal patients' records of diagnosis such that the researchers can study various diseases characteristics. The raw data or the original data, also called as micro data, contains the identities (e.g. names) of individual's record, which are not released to protect their privacy information. However, there may exist another attributes that are to be used, in combination with some other external database, to recover the original personal identities.

For example, consider that the hospital publishes the data which does not originally indicate the patients' name. However, if an adversary has the ability to access voter registration list, she can easily find the identities of all patients by joining any of the two tables. These three attributes are the quasi-identifier (QI) attributes.

To avoid the above problem, Generalization is a common approach which is performed by transforming the quasi-identifier QI values into a very less specific forms such that they do not uniquely represent individuals. However, in particular, if the QI values of each tuple are identical to those of at least k −1 other tuples then the table is k-anonymous

In this paper, a novel privacy preserving technique that overcomes the above problems is developed. The concept of personalized anonymity is the core of our solutions, i.e., the degree of privacy protection can be specified by a person for his/her sensitive values. In general, the taxonomy is accessible by the public, and organizes all the diseases as leaves of an individual tree. An intermediate node in the tree contains a name providing the diseases in its subtree. Here some part of the tree is omitted since it is not relevant to our discussion.

An individual preference can easily be solicited from an individual when he/she is providing his/her data. In this approach, an individual preference is formulated through a node in the taxonomy. Therefore, nobody may be able to infer, with a significant confidence that he/she suffered from any of the disease in the subtree of the node in the tree. In other terms, in Andy's opinion, allowing the public to associate with him accessing data.

The empty-set preference points out that he/she is willing to reveal her result of actual diagnosis; therefore, tuple should be published directly. However, in general, attribute may not be "sensitive" for many of the people, so that it is often not required to apply any privacy protection to this value.

In fact, personalization is an inherent notion of privacy preservation. The objective of personalization is to protect the interests of individual person at the first place. Somewhat surprisingly, it is so far the case where the literature has focused on a universal approach that provides the same amount of privacy preserving for all the individual persons, without catering for their concrete needs.

The consequence is that we may not be providing sufficient protection to a sub set of people (such as Andy in the above example), while applying excessive control over privacy to another subset of people. This method is rather more flexible, as it decides the minimum amount of generalization which is necessary for satisfying everybody's requirements, and therefore, gains the maximum amount of information from the microdata.

3.1 Batch Generation

In this first scenario, the owner of the data determines the M trust levels a priori, and generates the M perturbed copies of the data in one batch at the same time. In this scenario, all trust levels are predefined prior to the perturbed copies of the data and then are given for Z1 to Zm when generating the noise. This refers to the scenario as the batch generation. Here two batch algorithms are proposed. One of the algorithms is that which generates noise Z1 to ZM in parallel while the other algorithm generates sequentially.

3.2 Parallel Generation

Without the loss of generality, we assume that Zi < Zi+1 where 1 < I < M. In practice, this algorithm generates the components of noise ZZ, i.e., Z1 to ZM, simultaneously it is based on the probability distribution function which is shown below, for any real N * M dimension vector v.

$$\mathrm{Fz}\ (\mathrm{v}) = \frac{1}{\sqrt{}(2\pi)^{M}\det{(kz)}}\, e^{\frac{-1}{2}v^{T}Kz^{-1}v}$$

```
Parallel_Generation
{
        Input:   X, K
        Output: Y
        Construct Kz ← Kx
        Generate Z ← Kz
        Generate Y ← H X + Z
        Output ← Y
}
```

This algorithm then constructs YY as HX + ZZ and outputs the result. Here this algorithm is referred to as parallel generation. This parallel generation algorithm serves as a baseline algorithm for the next two algorithms i.e., sequential generation and on-demand generation

3.3 Sequential Generation

The large memory requirement for the parallel generation algorithm motivates one to seek for a memory efficient solution rather than this algorithm. Instead of parallel generation algorithm, one can sequentially start generating noise Z1 to ZM, each of which is a Gaussian vector of N dimension. The validity of the alternative procedure is based on the insight by using the sequential process.

```
Sequential_Generation
{
        Input:   X, K
        Output: Ym
        Construct Z ← N (0...Kx)
        Generate Yi ← X + Z1

        For i = 2 to M do
                Create Noise
                Generate Y ← Yi-1 + E
                Output ← Yi
}
```

3.4 On Demand Generation

Rather than the previous algorithms, the new perturbed copies are introduced on demand in this second scenario as opposed to the batch generation. As the requests may be arbitrary, the trust levels corresponding to the new copies would be arbitrary as well.

The new copies can be either lower or higher levels rather than the existing trust levels. Here this scenario is referred to as on-demand generation. Achieving the goal of privacy in this scenario will give data owners the maximum flexibility in providing MLT-PPDM services.

3.5 Enhanced Batch Generation

In this approach we apply the random perturbation approach for enhancing the privacy of the data. In this scenario the data owner determines the M trust levels a priori to the perturbed copies of data, and then generates random M perturbed copies of the data in multiple batches. In this scenario, all trust levels are predefined and for Z1 to Zm are given when generating the random noise. Here this scenario is referred to as the enhanced batch generation. Through this approach non linear attacks can also be avoided by protecting the data in generating random noise and applying retention replacement.

The time complexity of the above proposed algorithms O (n) for all the above algorithms and the space complexity is determined in terms of the perturbated data or the original data.

```
EnhancedBatch_Generation
{
        Input:   X, K
        Output: Ym
        Generate Random M
        Apply Priori
        Construct Z ← N (0...Kx)
        Generate Yi ← X + Z1
        Create Random Noise
        Generate Y ← Yi-1 + E
        Output ← Yi
}
```

In the enhancement of the proposed model, the advantage is that the data is perturbed evenly such that it is difficult to break or trace back the privacy of the original content. This is evident from the fig 4 which shows enhanced scheme in blue representing the data proportions evenly distributed as compared to drastic peaks in the existing batch generation scheme.

4. RESULTS

The concept of this paper is implemented and different results are shown below, The proposed paper is implemented in Java technology on a Pentium-IV PC with minimum 20 GB hard-disk and 1GB RAM. The proposed paper's concepts show efficient results and has been efficiently tested on different Datasets.

These experiments are run on a real data set CENSUS, which is commonly used in the literature of privacy preservation, for carrying out the experiments and evaluating their performance. This data set contains thousands of tuples with different attributes. The distributions of the data are shown figures below respectively.

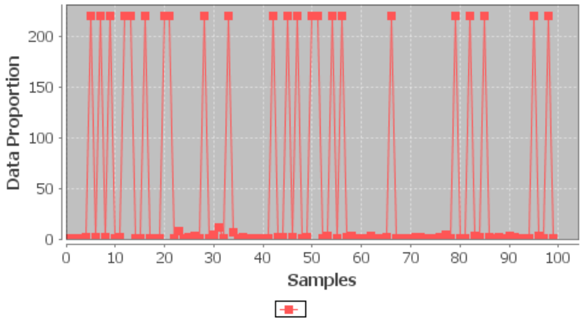

Fig. 1 Proposed system performing Batch generation.

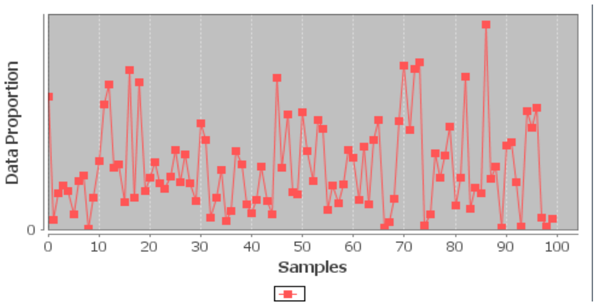

Fig. 2 Proposed system performing Enhanced batch generation

The above graph shows the enhanced batch generation based on the 100 samples. We observe that the data is very much randomly perturbated and noise added randomly.

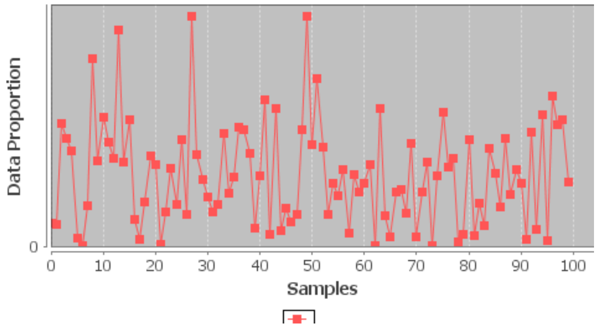

Fig. 3 Proposed System Batch Processing Graphs

The above graph shows the enhanced batch generation based on the 100 different samples. We observe that the data is very much randomly perturbated and noise added randomly

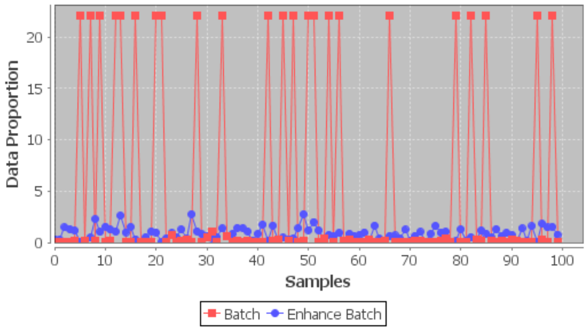

Fig. 4 Comparative Graphs batch and enhanced batch generation

The above graph shows the comparison between the normal batch and enhanced batch generation based on the 100 samples. Here observe that the data is very much closely perturbated in the enhanced generation and noise added which will be difficult to backtrack the original data.

5. CONCLUSIONS

In this work, the scope of additive perturbation based PPDM to multilevel trust (MLT) is expanded, by relaxing an implicit assumption of single-level trust in the exiting work. MLT-PPDM provides data owners to generate differently perturbed copies of its data for different levels of trust. The most important challenge lies in preventing the data miners from combining the copies at different trust levels to jointly reconstruct the initial data more accurately than what is allowed by the trusted data owner. The problem is addressed by properly correlating noise across the copies at different trust levels in the database. It is to prove that if one can design the noise covariance matrix to have corner-wave property, then data miners will have no diversity gain in their joint reconstruction of the original data. This claim is verified and demonstrates the effectiveness of the solution through numerical evaluation. Last but not the least, this solution allows data owners to generate perturbed copies of its data at arbitrary trust levels on-demand. This offers the data owner maximum flexibility as a property. One should believe that multilevel trust privacy preserving data mining can find applications in many of the fields. This work takes the initial step to enable the MLT-PPDM services. There are many more interesting and important directions worth exploring. For example, it is not least clear how to expand the scope of other approaches in the areas of partial information hiding, such as random rotation-based data perturbation, retention replacement and k-anonymity, to multilevel trust. It is also of great interest to extend this approach to handle evolving data streams.

6. ACKNOWLEDGEMENT

Our thanks to the management members and principal of Kakatiya Institute of Technology and Science-Warangal who have facilitated resources to read and compute in order to develop this article and our sincere thanks to Head of the Department Prof.P.Niranjan and Coordinator Sri V.Shankar who encouraged us research and publish this paper.

7. REFERENCES

[1] X.Xiao and Y. Tao, “Output Perturbation with Query Relaxation,” Proc. Int’l Conf. Very Large Data Bases, 2008.

[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] B. Bhattacharjee, N. Abe, K. Goldman, B. Zadrozny, V.R. Chillakuru, M.del Carpio, and C. Apte, “Using Secure Coprocessors for Privacy Preserving Collaborative Data Mining and Analysis,” Proc. Second Int’l Workshop Data Management on New Hardware (DaMoN ’06), 2006.

[4] B. Fung, K. Wang, A. Fu, and J. Pei, “Anonymity for Continuous Data Publishing,” Proc. Int’l Conf. Extending Database Technology (EDBT), 2008.

[5] F. Li, J. Sun, S. Papadimitriou, G. Mihaila, and I. Stanoi, “Hiding in the Crowd: Privacy Preservation on Evolving Streams Through Correlation Tracking,” Proc. IEEE 23rd Int’l Conf. Data Eng. (ICDE), 2007.

[6] G. Wang, Z. Zhu, W. Du, and Z. Teng, “Inference Analysis in Privacy-Preserving Data Re-Publishing,” Proc. Int’l Conf. Data Mining, 2008.

[7] D. Kifer and J.E. Gehrke, “Injecting Utility Into Anonymized Datasets,” Proc. ACM SIGMOD Int’l Conf. Management of Data, 2006.

[8] A.Machanavajjhala, J. Gehrke, D. Kifer, and M. Venkitasubramaniam, “L-Diversity: Privacy Beyond K-Anonymity,” Proc. Int’l Conf. Data Eng., 2006.

[9] X. Xiao and Y. Tao, “Personalized Privacy Preservation,” Proc. ACM SIGMOD Int’l Conf. Management of Data, 2006.

[10] X. Xiao and Y. Tao, “M-Invariance: Towards Privacy Preserving Re-Publication of Dynamic Datasets,” Proc. ACM SIGMOD Int’l Conf. Management of Data, 2007.

[11] S. Papadimitriou, F. Li, G. Kollios, and P.S. Yu, “Time Series Compressibility and Privacy,” Proc. 33rd Int’l Conf. Very Large Data Bases (VLDB ’07), 2007.

[12] X. Xiao, Y. Tao, and M. Chen, “Optimal Random Perturbation at Multiple Privacy Levels,” Proc. Int’l Conf. Very Large Data Bases, 2009.

8. AUTHOR'S PROFILE

B.Anitha Currently persuing Master of Technology in Computer Science and engineering with specilization in Softwatre Engineering.Computer Science and Engineering Department, Kakatiya Institute of Technology & Science (KITS), Kakatiya University-Warangal.A.P.,India.

B.Hanmanthu obtained his Bachelor's degree in Computer Science and Engineering from JNT University of India. Then he obtained his Master's degree in Computer Science and Engineering with specialization Software Engineering from JNT University Hyderabad, and he is also life member of ISTE. He is currently Assistant Professor of Computer Science and Engineering, Kakatiya Institute of Technology & Science (KITS), Kakatiya University-Warangal. His specializations include Data mining and Data warehousing, Databases and networking.

B.Raghuram obtained his Bachelor's degree in Computer Science and Engineering from JNT University of India. Then he obtained his Master's degree in Computer Science and Engineering from Pondicherry Central University Pondecherry, and he is also life member of ISTE. He is currently Assistant Professor of Computer Science and Engineering, Kakatiya Institute of Technology & Science (KITS), Kakatiya University-Warangal. His specializations include Data mining and Data warehousing, Databases and networking.