A novel approach for cryptography technique on Perturbed data for Distributed Environment

Nishant Doshi

National Institute of Technology, Surat, India
doshinikki2004@gmail.com

Abstract

Using the methods of data mining, one can search for large patterns in the huge database system. Now with the amassedevolution of technology, the data requirements and the amount of the data will significantly increase. Therefore, the data mining users,uses new methods for pattern matching which can be used for decision making. The data stored by the organizations are in a bulk. Therefore, when user gives a particular query, the amount of important or secure data can also be revealed as an answer of a query. This can harm to reputation of an organization. Therefore, privacy can concern to the above issue that not reveals any such kind of information about data provider and vice versa. Therefore, data needs to be modified without losing the data integrity. This paper outlines a method that achieve confidentiality from client and owner side which relatively less size of cipher text through mediator. We have extended the paper by Nishant et al. in CNSA-2012 to prevent the long term secret attack.

Keywords:

Data mining, Cryptography, data perturbation, privacy, sensitive data.

1 Introduction

Data mining is one of the useful fields that connects different major areas like Artificial intelligence, databases etc. To investigate the unidentified data pattern from huge database, data mining can be act as dominant tool as contended by authors in [14-15, 21, 24]. In [31-32] authors said that organization depend on data mining, gives better throughput to their customers.[22] shows an example which uses hospital record for collecting large data for patients. With the increasing use of technologies like internet, networking, hardware and software the amount of data with different organizations is collected in huge, which also include the sensitive data also. It may be possible that the answer of query issued by customer, can revealed the important data regarding health care, finance , security etc. The common tendency of people is to hide the sensitive information. In real scenario like hospital records, the analyst requires the records for more than one hospital as it will provide mutual benefits to the hospitals. In this example each hospital want to share the data but neither of them want to share the data of their patients or nor disclose it. In this situation, use of the privacy preserving data mining will enhance the integrity of data [20].

The rest of the paper is organized as follows. The section 2 and 3 we have given background study or literature study. In section 4, we have given the long term secret attack on our existing approach .In section 5, we discussed the proposed system. Related work and future expansion are given in section 6 and References are at the end.

DOI:10.5121/ijcis.2012.2309

2 Background Study

In general, there are two main approaches for the given problem one is use data transformation based approach and another is the cryptographic based approach. In firstapproach, the sensitive data is modified in such a way that it maintains its sensitive information. There are various data modification techniques as given in [1-5, 8, 11]. In cryptography technique data is modified using encryption techniques. In this one the communicating parties uses secured multiparty protocols as given in [6-7,17-19] which is not released any information to the third party. The basic techniques used were secure sum, secure size, set union etc.[27] suggest that in presence of adversary, which get some leakage data or gain some access to sensitive data. So in order to prevent this we require the third party that called the "trusted party". All the parties send their query or data to the trusted party and vice versa. [28] Suggests that any cryptography technique do not reveal any data with presence of the trusted party. [25-26] suggests the privacy preserving rules.

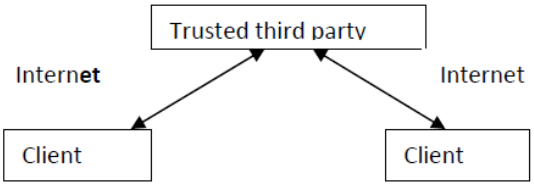

Fig 1. System using trusted party [29]

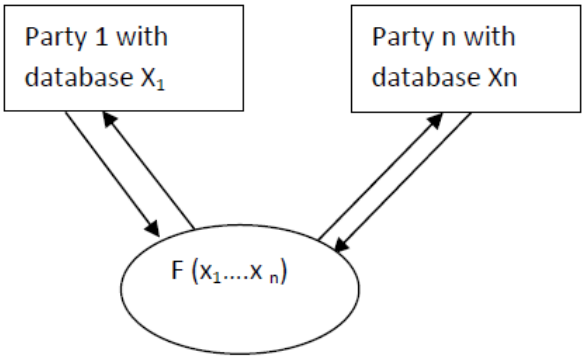

Fig 2. Secure multiparty communication [29]

As shown in figure 1 there are clients who are on the internet and they are communicating with the trusted server through the communication link. So in real scenario all the data which communicate between client and server must be encrypted otherwise an adversary can gain the view or access or modify the data. Now consider the figure 2 in which there are n party each of which had their database and they will merge their data using function F and get the perturbed data. The data sent by each data provider do not contain any sensitive information. We can enhance our scheme by incorporating features as given in [9-10].

3 Overview of randomization perturbation technique

In this approach the privacy of data is maintained by perturbed data[12,13,15,23] with randomization algorithm approach. [16] Suggest adding noise in this method. The Gaussian distribution technique is used for this one. As shown in figure 3 first the Gaussian algorithm and

using the random variables applied to data then different conditions for data integrity will be checked than the noise will be given to data and lastly the perturbed data is ready for communications. Figure 4 represent the framework which use to share data using the perturbation and encryption techniques.

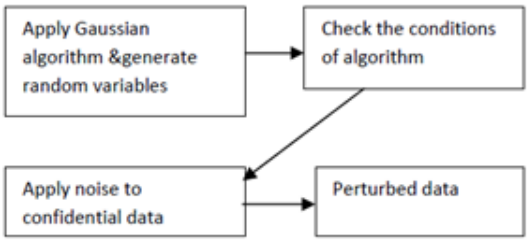

Fig. 3. Perturbation technique [29]

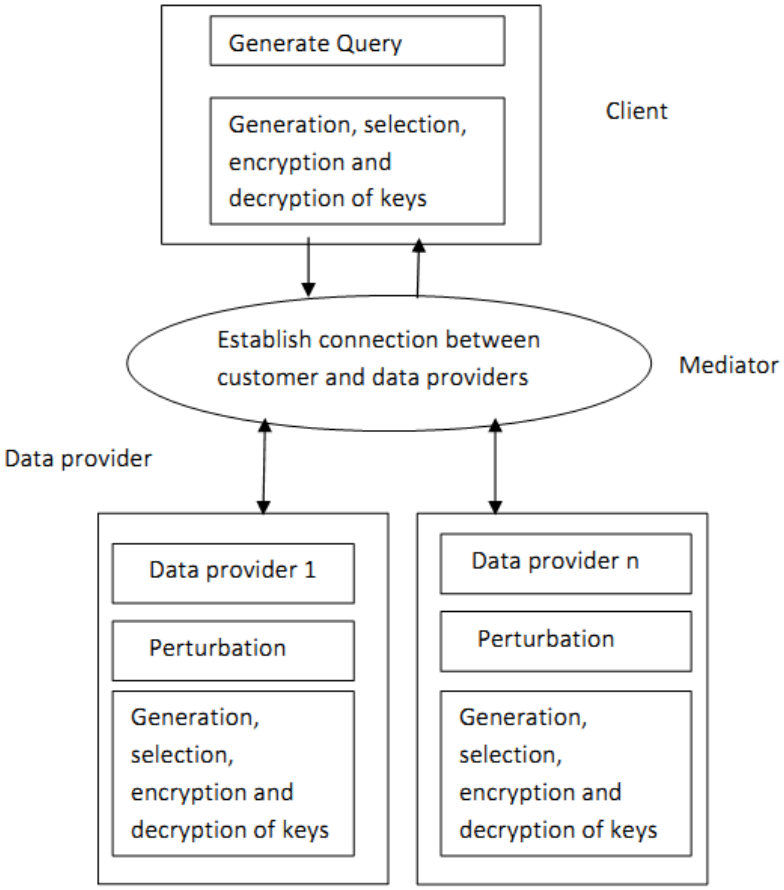

Fig. 4. Framework that represent sharing of data using encryption and perturbation method[29].

4 Attack on Our existing approach

As we have observed the one serious attack on the existing system of [33] as follows. In their algorithms of the scheme of [33], if the long term secret of the mediator is compromise than the all previous transactions will be compromised. Here we assume that attacker have already recorded the transactions when they happens. One may think that by compromising previous transactions can give the attacker the history for a particular users. The one key point of cryptographic security is that, even if long term secret compromise than that must not compromise the past conversations. With this as aim, we have proposed the more details scheme that prevents this attack.

5 Proposed System

Here we assume that mediator will change its key so that public key every day as follows. There is a hash function (one-way collusion resistance)which publically available. Every day, from beginning mediator will hash the key of previous day and discard the previous key. According to secret key,the mediator can generate the public key.

The steps of the proposed algorithm are as follow. Here each client contain its *PKc* (Public key) and *SKc*(secret key) same as each mediator contains *PKm* and *SKm*. Data providers do not need any keys but they know the *PKm*.

1. Mediator computer *secret key=Hash(previous day's key)* and generate corresponding public key.
2. Send the public key to related data providers.
3. Client *c* sends the query to mediator to get the data.
4. Mediator maintains the table in which a random generated number (or sequential number) *Rc* is associated with each incoming client request. Now mediator send query of client with random number to all data providers.
5. The data provider which satisfies the client requirements sends the perturbed data M with *Rc* and whole encrypted under pubic key *PKm* to mediator.
6. The mediator decrypts the data using own secret key *SKm* and checks the corresponding client for *Rc* and sends data back to client which encrypted under public key of client *c* e.g. *PKc*.
7. Client decrypts the data using its private key *SKc* and gets the required perturbed data.

As one can see that even if attacker can have key for particular day, he cannot get the key of previous day, as the hash function is one way only. As we have assumed that, the mediator will securely discard the key for previous day there is no way that attacker can retrieve the old keys.

6 Conclusion

This paper gives the proposed algorithm to reduce the network overheads between clients and data providers and still maintaining the privacy of data providers and clients with each other. Therefore, the main aim of this paper is to compare with [33] and suggest a new method that reduces the network overheads and long term secret compromise attack. The current open problem is if mediator is compromised than all the forward communication will be compromised. We will also look at concept of [32] to enhance our paper.

References

[1] Agrawal R.,Srikant R., ``Privacy Preserving Data Mining.," In the Proceedings of the ACM SIGMOD Conference. 2000.

[2] K.Muralidhar.,R.Sarathi, "A General additive data perturbation method for data base security" journal of Management Science. ,45(10):1399-1415,2002

[3] Agrawal D. Aggarwal C.C. " On the Design and Quantification of Privacy Preserving Data mining algorithms." ACM PODS Conference, 2002

[4] Muralidhar K. and Sarathy R. " Data Shuffling- a new masking approach for numerical data." Management Science, forthcoming, 2006.

[5] V.S. Iyengar."Transforming data to satisfy privacy constraints" In Proc. of SIGKDD'02, Edmonton, Alberta,Canada,

[6] Lindell Y., PinkasB."Privacy preserving Data Mining" CRYPTO 2000.

[7] Yu.H.,VaidyaJ.,JiangX."Privacy preserving SVM Classification on vertically partitioned data" PAKDD conference, 2006.

[8] D. Agarwal and C.C.Aggarwal, " On the design and quantification of privacy preserving data mining algorithms", In Proceedings of the 20th Symposium on Principles of Database systems,Santabarbara,California,USA, May2001

[9] Mrs. PrachiKarandikar, Prof. Sachin Deshpande "Preserving Privacy in Data Mining using Data Distortion Approach", ISSN : 2250-3439, International Journal of Computer Engineering Science, Volume 1 Issue 2, 2011.

[10] G.Ravi Kumar, Dr.G.A. Ramachandra and G.Sunitha An Evolutionary Algorithm for Mining Association Rules Using Boolean Approach, ISSN : 2250-3439, International Journal of Computer Engineering Science, Volume 1 Issue 3, 2011.

[11] M. Kantarcioglu and C. Clifton, "Privacy-preserving distributed mining of association rules on horizontally partitioned data", In the ACM SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery (DMKD'02), pp. 24–31,Madison, June 2002,

[12] K. Muralidhar, R. Sarathy, and R. A. Parsa, "A general additive perturbation method for database security," Management Science, vol. 45, no. 10, pp. 1399–1415, 1999.

[13] R.Agrawal, A.Evfimievski, R.Srikant, "Information sharing across private databases", In Proc.of ACM SIGMOD, 2003.

[14] J. Han and M. Kamber, "Data Mining: Concepts and Techniques". Morgan Kaufmann Publishers, 2000.

[15] H. Kargupta, S. Datta, Q. Wang, and K. Sivakumar, "On the privacy preserving properties of random data perturbation techniques" In Proc. of 3rd IEEE Int. Conf. on Data Mining, Washington, DC, USA,, pages99–106, 2003.

[16] K. Muralidhar, R. Parsa, and R. Sarathy, " A general additive data perturbation method for database security",Management Science, 19:1399–1415, 1999.

[17] B. Pinkas, " Cryptographic techniques for privacy preserving data mining " SIGKDD Explorations, 12–19, 2002.

[18] A. Evfimievski, " Randomization in privacy preserving data mining", In ACM SIGKDD Explorations Newsletter, volume 4, pages 43–48, 2002.

[19] Vaidya, J, Clifton, C., " Privacy-Preserving Data Mining: Why, How, and When", IEEE Security and Privacy, 2, 19- 27, 2004.

[20] Clifton, C, Kantarcioglu, M, Vaidya,J Lin, X, Zhu, M Y. , "Tools for privacy preserving distributed data mining" , SIGKDD Explor. News., 28-34, 2002

[21] Weiss G M, "Data Mining in Telecommunications", In Data Mining and Knowledge Discovery Handbook, A Complete Guide for Practitioners and Researchers.Kluwer Academic Publishers, 1189-1201, 2005.

[22] H. Kargupta, S. Datta, Q. Wang, and K. Sivakumar, "On the privacy preserving properties of random data perturbation techniques", In Proceedings of the 3rd IEEE International Conference on Data Mining, pages 99–106, Melbourne, Florida, November 19-22, 2003.

[23]Muralidhar, K., Parsa, R. and Sarathy, R " A General Additive Data Perturbation Method for database Security, Management" , 1399-1415,1999.

[24]Li Liu , Murat Kantarcioglu, BhavaniThuraisingham "The applicability of the perturbation based privacypreserving data mining for real-world data", Data & Knowledge Engineering 65 (2008) 5–21.

[25] A. Evfimievski, R. Srikant, R. Agrawal, and J. Gehrke. "Privacy preserving mining of association rules" In Proceedings of 8th ACM SIGKDD International Conference on Knowledge Discovery Data Mining, July 2002.

[26] Y. Lindell and B. Pinkas" Privacy preserving data mining". In Advances in Cryptology - crypto2000, Lecture Notes in Computer Science, volume 1880, 2000.

[27] J. Vaidya and C. Clifton, " Privacy preserving association rule mining in vertically partitioned data". In Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, July 23-26 2002.

[28] Kargupta. H., Datta,.S.,Wang.Q, And Sivakumar. K. " On the privacy preserving properties of random data perturbation techniques", Proc. of Intl. Conf. on Data Mining (ICDM) (2003).

[29] P.Kamakshi , Dr.A.VinayaBabu, "Preserving Privacy and Sharing the Data in Distributed Environment using Cryptographic Technique on Perturbed data" Journal of Computing, Volume 2, Issue 4, April 2010,ISSN 2151-9617.

[30] R.Agarwal and R.Srikant, "Privacy preserving data mining", In Procseedings of the 19th ACM SIGMOD conference on Management of Data ,Dallas,Texas,USA, May2000

[31] J. Canny, "Collaborative filtering with privacy". In IEEE Symposium on security and privacy , pages 45-57 Oakland,May 2002

[32] K Jothimani, S. Antony SelvadossThanmani MS: Multiple Segments with Combinatorial Approach for Mining Frequent Item sets Over Data Streams, International Journal of Computer Engineering Science, ISSN : 2250-3439,Volume 2 Issue 2, 2012.

[33] N. Goswami, T. Chauhan and N. Doshi. "Efficient Cryptographic Technique on Perturbed Data in Distributed Environment". In the proceedings of The Fifth International Conference on Network Security & Applications (CNSA-2012). Ser to Advanced in Intelligent Systems and Computing, Vol. 176. Eds. N. Meghanathan, D. Nagamalai, N. Chaki. pp:239-244. Springer Berlin / Heidelberg,2012.