1st International Conference on Applied Soft Computing Techniques
22 & 23.04.2017
In association with
International Journal of Scientific Research in Science and Technology

Achieving Security for Data Access Control Using Cryptography Techniques

[1]Dr.V.Vasanthi, [2]S. Akram Saeed Aglan Alhammadi, [3]Ramkumar. S, [3]Sathish Kumar
1Asst.Prof, Dept. of Computer Science,
2Ph.D Research Scholar, Dept. of Computer Science
Rathinam College of Arts and Science, Rathinam Techzone, Coimbatore, Tamil Nadu, India
E-mail :vasanthi.cs@rathinamcollege.com
3Asst.Prof, Dept. of Computer Applications,
3Asst.Prof, Dept. of CS & IT
Kalasalingam University, Madurai, Tamil Nadu, India

ABSTRACT

The amount of data being collected and stored every day by private and public sectors increased dramatically. Almost all industries, organizations and hospitals are maintaining personal information about individuals for decision making or pattern recognition. Security risk is very high while sharing this personal sensitive information among different data collectors. Therefore, privacy-preserving processes have already been developed to sanitize confidential information beginning with the samples while keeping their utility. For that safe and secure distributed computation new privacy preserving data mining algorithm has been developed. The main goal of these algorithms is to prevent that sensible information from hackers, during knowledge extraction from voluminous data. This work presents a protection saving approach that could be connected to decision tree learning, without associative misfortune of precision. This approach changeover the definitive specimen information sets into a gathering of undiscovered information sets, from which definitive information examines can't be remade without the whole assembly of unbelievable information sets. In the mean time, a proficient and precise decision tree might be manufactured straightforwardly from those stunning information sets. This novel methodology might be connected straightforwardly to the information space when the first sample is gathered. The methodology is versatile with other protection preserving approaches, for example, cryptography for extra protection.

Keywords : RSA, Data Mining, DSA, Cryptography, Cloud

I. INTRODUCTION

Data mining is an emerging field which connects different major areas similar databases, artificial intelligence and statistics. Data mining is an efficient tool for investigate and extract previously unknown patterns from voluminous data. The procedure of data mining requires a large amount of information to become collected into a central site. In modern days organizations are extremely depending on data mining in results to provide better resources, to achieving greater profit, and better decision-making. For these purposes organizations collect huge amount of data [1]. Just for example, business organizations collect data concerning the consumers for marketing purposes and improving business strategies, medical organizations collect medical records for better treatment and medical research. When using the rapid ahead of time web, networking, hardware and software technology there is remarkable growth in the quantity of data that could be collected from several sites or organizations. Huge volumes of Data collected in this particular manner also have sensitive data about individuals. It has been obvious that whenever a data mining algorithm is run against the union of various databases, the extracted knowledge simply not only consists of discovered patterns and correlations that might be hidden within the data however it reveals the information which is regarded as to be private. Privacy is a vital issue a lot of data mining applications that cope with health care,

security, financial along with other types of sensitive [2][3].

Privacy preserving knowledge mining is fundamentally new research area, which effectively extracts hidden information without including private knowledge of individuals. Privacy preserving technique becomes the new direction for organizations sharing the knowledge for clustering and also achieving privacy of individuals. Privacy issues are thinking about in situations. They are centralized surroundings and distributed surroundings. In centralized surroundings, database is available in single location. In this surroundings, a privacy preserving knowledge mining techniques are used to hide sensitive knowledge of individuals [4]. In distributed surroundings, knowledge is distributed to multiple sites.

The objective of the computational task is participates the parties to securely process some functions of their distributed and private inputs. An inquiry that emerges here is the thing that it implies that for a computation to be secured. A system of approaching this address is to provide a list of properties that ought to be saved. The headmost such property that regularly comes to mind is that of confidentiality or privacy.

A naive attempt at formalizing privacy would be to require that each party learns nothing about the other parties' inputs, though if it behaves maliciously. However, such a definition is generally unattainable because the defined output of the computation typically reveals some information on other parties' inputs. (For example, a decision tree estimated on two distributed databases reveals some information about both databases.) Therefore, the privacy requirement is generally formalized by saying that the only information learned by the parties in the computation (again, even by those who behave maliciously) is that specified by the function output. Even though privacy is a primary security property, it seldom suffices.

Another important property is that of correctness; this states that the parties' output is really that defined by the function (if correctness is not warranted, then getting a malicious party probably can obtain the specified decision tree while the honest party receives a tree that is modified to offer misleading information). A central query that arises during this method of shaping security properties is: when is our list of properties complete? This query is, of course, application-dependent and this essentially means that for each new drawback, the process of deciding which security properties are required must be re-evaluated. The effort that creating the right selection of properties will often be very challenging and it also may take number of years until we are believed strongly that a definition truly captures the security requirements that might be needed. Additionally, an incomplete of properties could simply contribute to real security failures. This paper concentrates on anticipating such attacks from third parties regarding the whole lifetime of the samples.

Contemporary research in privacy preserving knowledge mining chiefly falls into following categories first perturbation and randomization-based approaches, secondly secure multiparty computation (SMC)-based approaches. SMC approaches use cryptographic tools for collaborative information mining computation by multiple parties. Samples are distributed among different parties and they participate in the information computation in addition to communication technique.

SMC analysis focuses on protocol development for protecting privacy among the involved parties or computation efficiency; however, centralized processing of samples and storage privacy is out of the scope of SMC. A key utility of wide databases today is research, whether it is scientific or economic and merchandise oriented. Thus, for instance, the medical field has much to gain by pooling data for research; as can even competing businesses with mutual interests. Even though there is a potential gain, this is often unattainable because of the confidentiality issues which arise.

II. LITERATURE SURVEY

In recent years, privacy-preserving data mining techniques has been studied extensively, because of the wide proliferation of sensitive information on the internet. Most of the researchers and analyst focused on security and proposed number of algorithmic techniques for privacy-preserving data mining used in many applications such as medical, Bio-terrorism, identity theft, video surveillances, genomic and so on. In noise addition, generally a random number (noise) is drawn from a probability distribution having zero mean and a small standard deviation. This noise is then added to a

numerical attribute in order to mask its original value. Generally noise is added to the confidential attributes, of a micro data file before the data is released, in order to protect the sensitive information of an individual. However, adding noise to both confidential and Non-confidential attributes can improve the level of privacy by making re-identification of the records more challenging. The main objective of noise addition is to protect individual privacy by masking the micro data while introducing the least amount of incorrectness in it. The incorrectness in the statistic of a perturbed data set with respect to the statistic of the unperturbed data set.

Charu C. Aggarwal and Philip S. Yu [5] stated the methods which have been used in the privacy preservation Like: First, Randomization technique has been used to add the noise in the data in order to mask the attribute value of the records.

Hillol Kargupta ,souptik data,Qi wang and Krishnamoorthy sivakumar [6] triggering the randomized data distortion technique to mask the attribute value of the records for preserving the privacy of sensitive data. This work attempts to add the noise in the data values randomly and it's difficult to identify the predictive structure in spectral domain. To defeat this issue, work with the randomized object i.e. random matrix based spectral fitting techniques to retrieve the original values from the dataset and add random values.
Xiaokui Xiao, Yufei Tao and Minghua Chen [7] achieves with two crucial properties. First, collusion is useless, meaning that the colluding recipients cannot learn anything more than what the most trustable recipient (among the colluding recipients) already knows alone. Second, the data each recipient receives and it can be regarded (and hence, analyzed in the same way) as the output of conventional uniform perturbation. The proposed technique is both spaces economical and computationally efficient.

Keke Chen and Ling Liu[8] created data mining service oriented framework to perform the geometric data perturbation approach and developed three protocols for perturbation unification for secure the geometric data with different parties. The unique feature of this proposed work has involves three protocol namely simply protocol, Negotiation protocol and space adaptation protocol. The basic challenge include (1) how to securely generate the same random perturbation in each site, while preventing the service provider knowing the unified perturbation, and (2) how to prevent privacy breach caused by data providers. Negotiation protocol overcomes the challenges of simply protocol by improving the overall privacy guarantee for all data providers. However, if perturbation data is generated randomly in simple protocol in terms of privacy guarantee then the providers may not be satisfied with in it execution. But in the negotiation protocol, each data provider has a chance to review the candidate perturbation and vote for or against the candidate

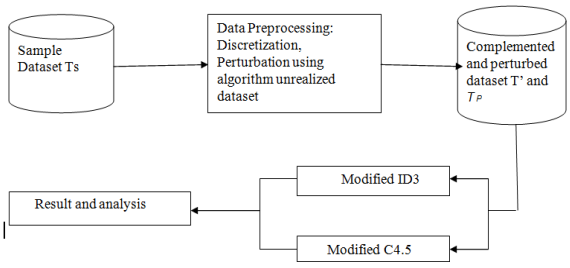
Alka Gangrade, Ravindra Patel[9] evaluated the multi party data to perform the privacy data perturbation approach with the secure manner and introduce a Privacy-preserving decision tree classifier using C4.5 algorithm without involving any third party. In multi party centralized the records are resides only in one party such horizontally or vertically to partitioning the data set. It is based on to calculate the union (frequent items can be evaluated using Secure Multiparty Computation methods for classifying rule mining) of all parties databases without trusted third party.

P.Kamakshi and A.Vinaya Babu[10] performed privacy preservation in the distributed environment using cryptography technique in data mining algorithms. Cryptography, the science of communication and computing in the presence of a malicious adversary extends from the traditional tasks of encryption and authentication to protocols for securely distributing computations among a group of mutually distrusting parties. In an ideal situation, in addition to the original parties there is also a third party called "trusted party" who does not deviate from the activities prescribed for him.

III. OVERVIEW OF THE WORK

The privacy preserving is one technique for changes over the original sample data sets into an aggregation of annoyed data sets. In that, original samples can't be recreated without the entire group of unreal data sets. In the meantime, an exact decision tree could be built directly from those unreal data sets.
This original approach can be applied directly to the data storage as soon as the first sample is collected. The approach is well-suited with other privacy preserving approaches, such as cryptography, for extra security.

The privacy is preserved even if the data is spread across multi parties. Suppose a bank issues the account holders' some of the attributes to more than one insurance agency. Then from the attributes of the table along with the records given to one insurance agency, other agency could not guess or identify the facts regarding the account holders. Likewise, if two agencies give their data set (retrieved from the bank) to other parties, they must not identify the facts by combing both data sets.



The overall frame work of the system involves two main modules first is data pre-processing and second is decision tree generation. In data pre-processing module initially continuous value attribute dataset is converted into discrete value after that the dataset is converted into sanitized version by using algorithm unrealized training set. Then generated complemented dataset and perturbed dataset is given as a input to decision tree generation module in which decision tree is built by using ID3 and C4.5 and result generated by both the algorithm is compared to analyze the algorithm.

IV. OVERVIEW OF THE TECHNICAL DETAILS

1 DEFINING PRIVACY PRESERVATION IN DATA MINING:

Basically there are two major dimensions in privacy preservation is: 1) Users' personal information 2) Information concerning their collective activity. First they have used individual privacy preservation and the latter as collective privacy preservation, which is related to corporate privacy. Individual privacy preservation: The primary goal of data privacy is the protection of personally identifiable information.

The information is considered personally recognizable if it can be connected, directly or indirectly, to an individual person. Thus, when personal information are subjected to mining, the attribute data associated with individuals are private and must be protected from disclosure. Analysts are able to acquire information from global models rather than particular individual characteristics.

2. COLLECTIVE PRIVACY PRESERVATION

It is not enough to protect only personal data. Sometimes, they may need to protect against learning sensitive knowledge representing the activities of a group. The protection of sensitive knowledge is collective privacy preservation. The goal is quite similar to statistical databases; within which security management mechanisms provide aggregate information about groups (population) and, at the same time, should prevent disclosure of confidential information about individuals.

However, unlike statistical databases, another aim of collective privacy preservation is to preserve strategic patterns that are paramount for strategic decisions, rather than minimizing the falsification of all statistics (e.g., bias and precision). In other words, the goal of collective privacy preservation is not only to protect personally identifiable information but also some patterns and trends that are not supposed to be discovered.

In collective privacy preservation, organizations have to deal with some interesting conflicts. For example, when personal information undergoes analysis processes that produce new facts about users' hobbies, shopping patterns, or preferences, these facts may be used in recommender systems to predict or affect 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 identifiable information but also some strategic patterns.

In the business world, this pattern describes about the knowledge that can provide competitive advantages, and therefore must be protected. Protecting the knowledge that is discovered from confidential information is the big challenge (e.g., medical, financial, and crime information). The absence of privacy safeguards can evenly compromise individuals' privacy. While violation of individual privacy is clear, violation of collective privacy will lead to violation of individual's privacy.

Privacy Preserving Data Mining:

Models and Algorithms propose 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 protocols for data sharing, statistical techniques for disclosure and inference control, query auditing methods, randomization and perturbation-based techniques. This edited volume also contains surveys by distinguished researchers in the privacy field. Each survey includes the key research content as well as future research directions of a particular topic in privacy.

3. Privacy Preserving Techniques

There are many approaches which have been adopted for privacy preserving data mining. To classify that based on the following dimensions:

- Data distribution
- Data modification
- Data mining algorithm
- Data or rule hiding
- Privacy preservation

The first dimension refers to the distribution of data. Some of the approaches have been developed for centralized data, while others refer to a distributed data scenario. Distributed data scenarios can also be classified as horizontal data distribution and vertical data distribution. Horizontal distribution refers to these cases where different database records reside in different places, while vertical data distribution, refers to the cases where all the values for different attributes reside in different places.

The second dimension refers to the data modification scheme. In general, data modification is used in order to modify the original values of a database that needs to be released to the public and in this way to ensure high privacy protection. It is important that a data modification technique should be in concert with the privacy policy adopted by an organization. Methods of modification include:

- ✓ Perturbation is accomplished by the alteration of an attribute value by a new value (i.e., changing a 1-value to a 0-value, or adding noise).
- ✓ Blocking is the replacement of an existing attribute value with a "?",
- ✓ Aggregation or merging which is the combination of several values into a coarser category,
- ✓ swapping that refers to interchanging values of individual records, and
- ✓ Sampling is refers to releasing data for only a sample of a population.

The third dimension refers to the data mining algorithm, for which the data modification is taking place. This is actually something that is not known before-hand, but it facilitates the analysis and design of the data hiding algorithm. We have included the problem of hiding data for a combination of data mining algorithms, into our future research agenda.

For the time being, various data mining algorithms have been considered in isolation of each other. Among that, the most important ideas have been developed for classifying data mining algorithms, like decision tree inducers, association rule mining algorithms, clustering algorithms, rough sets and Bayesian networks.

The fourth dimension refers to whether raw data or aggregated data should be hidden. The complexity for hiding aggregated data in the form of rules is of course higher, and for this reason, mostly heuristics have been developed. The lessening of the amount of public information causes the data miner to pro-duce weaker inference rules that will not allow the inference of confidential values. This process is also known as "rule confusion". The last dimension which is the most important refers to the privacy preservation technique used for the selective modification of the data. Selective modification is required in order to achieve higher utility for the modified data given that the privacy is not jeopardized. The techniques that have been applied for this reason are:

Heuristic-based techniques like adaptive modification that modifies only selected values that minimize the utility loss rather than all available values.

- Cryptography-based techniques like secure multi-party computation where a computation is se-cure if at the end of the computation no party knows anything except its own input and the results
- Reconstruction-based techniques where the original distribution of the data is reconstructed from the randomized data.

It is important to realize that data modification results in degradation of the database performance. In order to quantify the degradation of the data, its mainly use two metrics. The first one, measures the confidential data protection, while the second measures the loss of functionality.

V. Proposed Methodology

The existing ID3 (Iterative Dichotomiser 3) decision tree learning algorithm which covers the discrete-valued attributes that are implemented in the proposed system.

To preserve privacy when datasets are given to multiple parties, the proposed system finds the solution for the key problem of applying geometric data perturbation in multiparty collaborative mining which securely unify multiple geometric perturbations that are preferred by different parties, respectively.

In the proposed system, the privacy is preserved even if the data is spread across multi parties. Suppose a bank issues the account holders' some of the attributes to more than one insurance agency. Then from the attributes of the table along with the records given to one insurance agency, other agency could not guess or identify the facts regarding the account holders. Likewise, if two agencies give their data set (retrieved from the bank) to other parties, they must not identify the facts by combing both data sets.

The following are the advantages which are enclosed in the proposed theory

- Present the geometric perturbation approach which helps to multiparty privacy-preserving collaborative mining.
- From the two given data sets, the original facts can not be guessed.
- Privacy is preserved even if the data is spread across multi parties.
- Consider multiple service providers collaboratively providing the privacy preserving mining service to multiple data providers.

ID3 ALGORITHM

In decision tree learning, **ID3** (**Iterative Dichotomiser 3**) is an algorithm invented by Ross Quinlan used to generate a decision tree from a dataset. ID3 is the precursor to the C4.5 algorithm, and is often utilized in the machine learning and natural language processing domains.

Decision tree algorithms are a method for approximating discrete-valued target functions, during which the learned function is represented by a decision tree. These types of algorithms are famous in inductive learning and have been successfully applied to a broad range of tasks. Decision trees classify instances by sorting them down the tree from the root to some leaf node that provides the classification of the instances. Each node in the tree specifies some attribute instance and each branch descending from that node corresponds to one of the possible values for this attribute. The reasons for decision learning tree algorithms to be attractive are:

1. They generalize in a better way for unobserved instances, once examined the attribute value combined with in the training data.
2. They are efficient in computation as it is proportional to the number of training instances observed.
3. The tree elucidation gives an exceptional comprehension of instance and classify instances depend upon attributes

Example

ID3 algorithm is explained here using the classic 'Play Tennis' example.

VI. IMPLEMENTATION OF THE PROPOSED SYSTEM

7.1.1 ADD DATA SET

In this, the data set records are added. The outlook, temperature, humidity and windy column values along with result column is added with both realized and unrealized values. The details are saved into 'DataSet' table.

7.1.2 VIEW ORIGINAL DATA SET

In this, the data set records are viewed using data grid view control. The outlook, temperature, humidity and windy column values along with result column is displayed with realized values. The details are fetched from 'DataSet' table.

7.1.3VIEW UNREALIZED DATA SET

In this, the data set records are viewed using data grid view control. The outlook2, temperature2, humidity2 and windy2 column values are added with unrealized values. The details are fetched from 'DataSet' table.

7.1.4 VIEW ORIGINAL/UNREALIZED DATA SET

In this, the data set records are viewed using data grid view control. The outlook, temperature, humidity and windy column values and outlook2, temperature2, humidity2 and windy2 column values are displayed. The details are fetched from 'Dataset' table.

7.1.5 MODIFY ORIGINAL/UNREALIZED DATA SET

In this, the data set records are viewed using data grid view control. The outlook, temperature, humidity and windy column values and outlook2, temperature2, humidity2 and windy2 column values are displayed. The details are fetched from 'DataSet' table. The data grid records can be modified and saved into database.

7.1.6 DECISION TREE GENERATION

In this, the decision tree is generated based on ID3 algorithm and displayed in the form. The records are taken from 'Dataset' table, best root attribute is selected and tree is constructed.

7.1.7 GEOMETRIC PERTURBATION

This module presents the geometric perturbation approach which helps to multiparty privacy-preserving collaborative mining. From the two given data sets, the original facts cannot be guessed. Privacy is preserved even if the data is spread across multi parties. It considers multiple service providers collaboratively providing the privacy preserving mining service to multiple data providers. The data is encrypted and extra bits are added in the result. In this approach, two results are added with two different bits. The application running in two places discards the unwanted bit and decrypts the data.

VII. Result and Discussion

The output accuracy (the similarity between the decision tree generated by the regular method and by the new approach), the storage complexity (the space required to store the unrealized samples based on the size of the original samples) and the privacy risk (the maximum, minimum, and average privacy loss if one unrealized data set is leaked).

ALGORITHM USED	ACCURACY	COST	EFFICENCY
Data Modification Techniques	LOW		MEDIUM
SMC Approaches	MEDIUM	HIGH	LOW
Perturbation-Based Approaches		HIGH	MEDIUM
Cryptographic Techniques			MEDIUM
Iterative Dichotomizer 3 Decision Tree Learning Algorithm		MEDIUM	HIGH
Multiparty Collaborative Mining	High	Low	High
Geometric Perturbation	High	Low	High

The privacy risk of the dummy attribute values technique, the average privacy loss per leaked unrealized data set is small, except for the even distribution case (in which the unrealized. Samples are the same as the originals). By doubling the sample domain, the average privacy loss for a single leaked data set is zero, as the unrealized samples are not linked to any information provider. The randomly picked tests show that the data set complementation approach eliminates the privacy risk for most cases and always improves privacy security significantly when dummy values are used.

COMPARISON OF FOUR PRIVACY PRESERVING ALGORITHM TECHNIQUE

Algorithm	Probability	Privacy	Accuracy
SMC approaches	**68**	**80**	**80**

Perturbation-based approaches	**73**	**83**	**90**
ID3 algorithm	**99**	**100**	**99**
Geometric perturbation	**100**	**100**	**100**

Table 1 : comparison of four privacy preserving algorithm technique

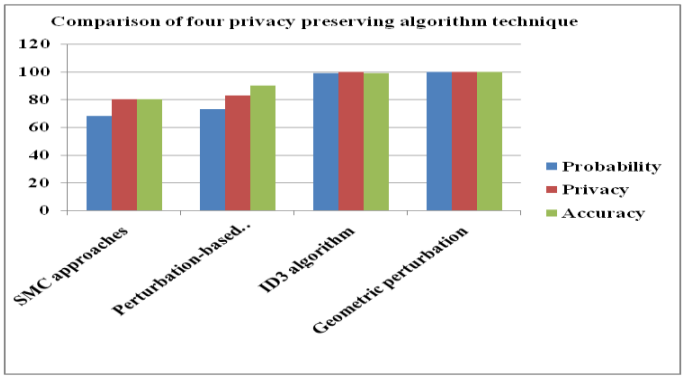


Figure 1 : Chart analysis of four privacy algorithm comparison

Table and fig 1 shows the comparison of four privacy preserving algorithm technique represents SMC approaches, Perturbation-based approaches, ID3 algorithm and Geometric perturbation. The result reveals that the SMC approaches given Probability, privacy and accuracy represents 68, 80 and 80. Perturbation-based approaches produced 73, 83, 90 and ID3 algorithm produced 99, 100, 99 finally Geometric perturbation technique given 100,100,100 we obtained the above results Geometric perturbation algorithm outperform the other remaining approaches.

THEORETICAL ANALYSIS

The theoretical analysis is as follow:

Features	DES	RSA	Triple DES
Key Used	Same key is used for encryption and decryption purpose	Different keys are used for encryption and decryption purpose	Same key is used for encryption and decryption purpose
Scalability	It is scalable algorithm due to varying the key size and value size	No scalability occurs	It is scalable algorithm due to varying the key size and value size
Avalanche effect	No more affected	More affected	No more affected
Power Consumption	Low	High	More than DES
Throughput	Very high	Low	High
Confidentiality	High	Low	Very high

The security of the scalar product protocol is based on the inability of either side to solve k equations in more than k unknowns. Some of the unknowns are randomly chosen, and can safely be assumed as private. However, if enough data values are known to the other party, the equations can be solved to reveal all values. Therefore, the disclosure risk in this method is based on the number of data values that the other party might know from some external source.

The scalar product protocol is used once for every candidate item set. This could introduce extra equations. When the candidate item set contains multiple attributes from each side, there is no question of linear equations so it does not perceptibly weaken the privacy of the data.

Input Size (KB)	**DES**	**3DES**
75	21	57
96	32	55
112	54	81
286	97	173
359	188	198
600	198	202
951	391	327
5345.28	1399	1149
Throughput (MB/sec)	**3.01**	**2.8**

Table 2 : Execution Time (Milliseconds) of Encryption of different data packet size (DES & 3DES)

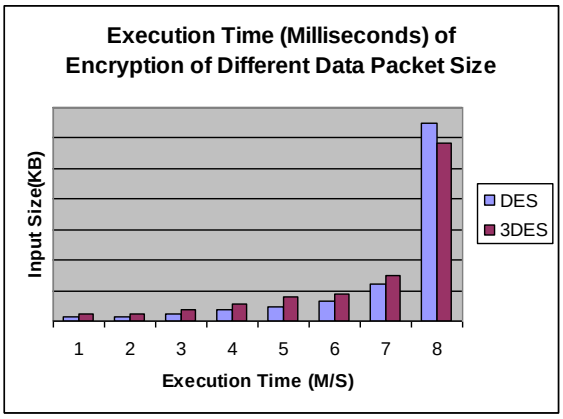


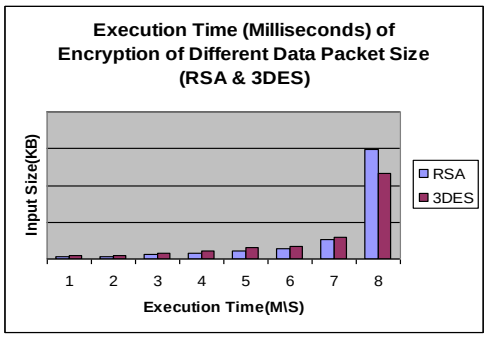
Figure 2: Execution Time (Milliseconds) Of Decryption Of Different Data Packet Size(DES And 3DES)

Input Size(KB)	RSA	3DES
45	42	45
55	37	42
96	58	65
236	88	104
319	152	135
560	148	160
899	252	181
5345.28	893	845
Throughput (MB/sec)	**2.0875**	**1.97125**

Table 2: Execution Time (Milliseconds) of Decryption of Different data packet size (DES & 3DES)

Input Size (KB)	RSA	3DES
75	38	50
96	35	44
112	55	76
286	83	113
359	105	155
600	143	177
951	264	299
5345.28	1296	1166
Throughput (MB/sec)	**2.52**	**2.08**

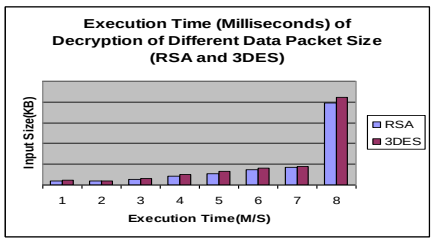
Table3: Execution Time (Milliseconds) of Encryption of Different data packet size
(RSA & 3DES)



Execution Time (Milliseconds) of Decryption of Different Data Packet Size(RSA and 3DES)

Input Size (KB)	DES	3DES
75	25	62
96	34	59
112	56	84
286	102	176
359	192	203
600	205	210
951	402	333
5345.28	1401	1152
Throughput (MB/sec)	**3.03**	**2.84**

Table 4: Execution Time (Milliseconds) of Decryption of Different data packet size (RSA & 3DES)



Execution Time (Milliseconds) of Encryption of Different Data Packet Size
(DES and 3DES)

VIII. CONCLUSION

Privacy preservation via data set complementation fails if all training data sets are leaked because the data set reconstruction algorithm is generic. Therefore, further research is required to overcome this limitation. As it is

very straight forward to apply a cryptographic privacy preserving approach, such as the (anti)monotone frame-work, along with data set complementation, this direction for future research could correct the above limitation.

This paper covers the application of this new privacy preserving approach with the ID3 decision tree learning algorithm and discrete-valued attributes only. The norm in data collection processes, a sufficiently large number of sample data sets have been collected to achieve significant data mining results covering the whole research target. Second, the number of data sets leaked to potential attackers constitutes a small portion of the entire sample database.

This paper covers the applications of this new privacy preserving approach with the ID3 decision tree learning algorithm. In addition geometric perturbation mechanism is used so that the data is secured even distributed to more than one parties. It is suitable for multiparty data distribution.

IX. References

[1] 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.

[2] 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

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

[4] 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.

[5] C. Aggarwal and P. Yu, Privacy-Preserving Data Mining:, Models and Algorithms. Springer, 2008.restoneTireRecall.htm, May 2001.

[6] Hillol Kargupta ,souptik data,Qi wang and Krishnamoorthy sivakumar,"random data Perturbation technique and privacy preserving data mining ",IEEE international conference on data mining, 2003.

[7] Xiaokui Xiao, Yufei Tao and Minghua Chen," Optimal Random Perturbation at Multiple Privacy Levels", ACM. VLDB '09, August 24-28, 2009, Lyon, France

[8] Keke Chen Ling Liu," Privacy-preserving Multiparty Collaborative Mining with Geometric Data Perturbation", IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED COMPUTING, VOL. XX, NO. XX, JANUARY 2009

[9] Alka gangrade, ravindra patel," building privacy-preserving c4.5 Decision tree classifier on multiparties," international journal on computer science and engineering vol.1(3), 2009, 199-205

[10] p.kamakshi , dr.a.vinaya babu, " 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

[11] Mohammad Ali Kadampur, Somayajulu D.V.L.N," A Noise Addition Scheme In Decision Tree For Privacy Preserving Data Mining," Journal Of Computing, Volume 2, Issue 1, January 2010, Issn 2151-9617

[12] Li Liu Murat Kantarcioglu and Bhavani Thuraisingham," Privacy Preserving Decision Tree Mining from Perturbed Data," Proceedings of the 42nd Hawaii International Conference on System Sciences - 2009

[13] M. Naga lakshmi, k sandhya rani," a privacy preserving clustering method based on fuzzy approach and random rotation perturbation." Publications of problems & application in engineering research, vol 04, special issue01; 2013

[14] V.Thavavel and S.Sivakumar," A generalized Framework of Privacy Preservation in Distributed Data mining for Unstructured Data Environment," International Journal of Computer Science Issues, Vol. 9, Issue 1, No 2, January 2012

[15] Pui k. Fong and jens h. Weber-jahnke," privacy preserving decision tree learning using unrealized data sets," ieee transactions on knowledge and data engineering, vol. 24, no. 2, february 2012

[16] S.nithya and p.senthil vadivu," efficient decision tree based privacy preserving Approach for unrealized data sets," international journal of advances in computer science and technology, 2(6), june 2013, 83 - 89

[17] Seema kedar, sneha dhawale, wankhade vaibhav," privacy preserving data mining," international journal of advanced research in computer and communication engineering vol. 2, issue 4, april 2013

[18] Justin Brickell and Vitaly Shmatikov," Privacy-Preserving Classi_er Learning",

[19] Keke chen ling liu," privacy-preserving multiparty collaborative mining with geometric data perturbation", ieee transactions on parallel and distributed computing, vol. Xx, no. Xx, january 2009

[20] S.L. Wang and A. Jafari, "Hiding Sensitive Predictive Association Rules," Proc. IEEE Int'l Conf. Systems, Man and Cybernetics, pp. 164-169, 2005.

[21] Q. Ma and P. Deng, "Secure Multi-Party Protocols for Privacy Preserving Data Mining," Proc. Third Int'l Conf. Wireless Algo-rithms, Systems, and Applications (WASA '08), pp. 526-537, 2008

[22] J. Gitanjali, J. Indumathi, N.C. Iyengar, and N. Sriman, "A Pristine Clean Cabalistic Foruity Strategize Based Approach for Incre-mental Data Stream Privacy Preserving Data Mining," Proc. IEEE Second Int'l Advance Computing Conf. (IACC), pp. 410-415, 2010.

[23] L. Liu, M. Kantarcioglu, and B. Thuraisingham, "Privacy Preser-ving Decision Tree Mining from Perturbed Data," Proc. 42nd Hawaii Int'l Conf. System Sciences (HICSS '09), 2009.

[24] Y. Zhu, L. Huang, W. Yang, D. Li, Y. Luo, and F. Dong, "Three New Approaches to Privacy-Preserving Add to Multiply Protocol and Its Application," Proc. Second Int' l Works hop Knowledge Discovery and Data Mining, (WKDD '09), pp. 554-558, 2009.

[25] M. Shaneck and Y. Kim, "Efficient Cryptographic Primitives for Private Data Mining," Proc. 43rd Hawaii Int'l Conf. System Sciences (HICSS), pp. 1-9, 2010.

[26] J. Vaidya and C. Clifton. Privacy-preserving decision trees over vertically partitioned data. In Proceedings of the 19th Annual IFIP WG 11.3 Working Conference on Data and Applications Security, Storrs,Connecticut, 2005. Springer. L. Liu, M. Kantarcioglu, and B. Thuraisingham, "Privacy Preserving Decision Tree Mining from Perturbed Data," Proc. 42nd Hawaii Int'l Conf. System Sciences (HICSS '09), 2009.

[27] S. Ajmani, R. Morris, and B. Liskov, "A Trusted Third-Party Computation Service," Technical Report MIT-LCS-TR-847, MIT, 2001.

[28] Lian Liu.,Jie Wang.,Jun Zhang., ``Wavelet based data perturbation for simultaneous privacy preserving and statistics preserving.," In Proceedings of IEEE International Conference on Data Mining workshop., 2008.