INTERNATIONAL JOURNAL OF PURE AND APPLIED RESEARCH IN ENGINEERING AND TECHNOLOGY

A PATH FOR HORIZING YOUR INNOVATIVE WORK

INFORMATION SECURITY IN BIG DATA USING AES TECHNIQUE

SNEHAL R. AWACHAT, PROF S. W. MOHOD

Department of computer engineering, BDCOE, Wardha.

Accepted Date: 15/03/2016; Published Date: 01/05/2016

Abstract: Data mining technology means identifying patterns and trends from large collection of data. Privacy protection in data mining is crucial issue that has captured the attention of many researchers. Current studies of privacy preserving data mining mainly focus on how to reduce privacy risk brought by data mining operation. A number of methods and techniques has been proposed for privacy preserving in data mining .In this paper represent generalization base technique to privacy preserving and proposing system "Information security in big data using AES technique" which is more strong to preserve privacy.

Keywords: Privacy preserving data mining, Diffie Hellman algorithm, AES Technique.

Corresponding Author: MS. SNEHAL R. AWACHAT

Access Online On:

www.ijpret.com

How to Cite This Article:

Snehal R. Awachat, IJPRET, 2016; Volume 4 (9): 23-30

PAPER-QR CODE

INTRODUCTION

Today's everyone have data collector and collect the data in digital form. A number of recorded databases are available with detail information about relative field. Annonymization mean identifying information is removed from the original data to protect personal information. To understand the concept of data annonymization suppose we take simple example of medical patient. The information of single patient is stored in a single line. If the database is anonymized it is not possible to identify the patients records In suppression based approach we are using use difffie hellman key exchange algorithm to generate private secure key then applying AES algorithm to encrypt and decrypt data using key generated by diffie hellman key exchange algorithm.

Literature survey

Abouelela Abdou Hussien, Nagy Ramadan Darwish, Hesham A. Hefny [1] "Utility Based Anonymization Using Generalization Boundaries to Protect Sensitive Attributes" in June 2015 describes Utility-Based Anonymization using Generalization Boundaries to protect Sensitive Attributes Depending on Attributes Sensitivity Weights. In this technique researchers start with considering the sensitivity of values in queries and then only quires having sensitive values (taking in account Utility based Anonymization) are generalized using Generalization Boundaries and the other quires that doesn't have sensitive values can be directly published.

Dr. K. Sagar [2] "A Novel Anonymization Technique for Privacy Preserving Data Publishing" in August 2015 describe Slicing partitions the data set both vertically and horizontally. Vertical partitioning is done by grouping attributes into columns based on the correlations among the attributes. Each column contains a subset of attributes that are highly correlated. Horizontal partitioning is done by grouping tuples into buckets. Finally, within each bucket, values in each column are randomly permutated (or sorted) to break the linking between different columns. In data slicing the attributes of tables are used for slicing and bucketized, in a generalized table each attribute value is replaced with the multi set of values in the bucket. Slicing first partitions attributes into columns. Each column contains a subset of attributes.

Guang Li, Yadong Wang, Xiaohong Su [3] "A New Algorithm-independent Method for Privacy preserving Data Mining" in Feb 2014 describe PPDM method is based on data perturbation. Only part of the data-perturbation-based methods is algorithm-irrelevant, which are favorable because common data mining algorithms can be used directly. A new algorithm-irrelevant PPDM method based on data perturbation. To maintain data utility, the generated data should have the same distribution as the original data. Method uses a two-step strategy to solve this

problem. First, it generates a set of independent and identically distributed data for each attribute without considering the correlation of attributes. Then it restores the attributes correlation with the ranking order.

Ms. R. Kavitha, Prof. D. Vanathi [4] "A Study of Privacy Preserving Data Mining Techniques" in August 2014 discuss method for Perturbation, K-Anonymization, condensation, and Distributed Privacy Preserving Data mining. A review of the state-of-the-art methods for privacy and analyze the representative technique for privacy preserving data mining and point out their merits and demerits.

Tamas Zoltan Gal, Gabor Kovacs, Zsolt T. Kardkov Acs [5] "Survey on privacy preserving data mining techniques in health care databases" in 2014 discuss an outlook on data anonymization problems by case studies, give a summary on the state-of-the-art health care data anonymization issues including legal environment and expectations, the most common attacking strategies on privacy, and the proposed metrics for evaluating usefulness and privacy preservation for anonymization.

Wei Peng, Feng Li, Xukai Zou, Jie Wu, Fellow [6] "A Two-stage Deanonymization Attack Against Anonym zed Social Networks" in 2014 propose an algorithm, Seed and Grow, to identify users from an anonymized social graph, based solely on graph structure. The algorithm first identifies a seed sub-graph, either planted by an attacker or divulged by a collusion of a small group of users, and then grows the seed larger based on the attacker's existing knowledge of the users' social relations. It identifies and relaxes implicit assumptions taken by previous works, eliminates arbitrary parameters, and improves identification effectiveness and accuracy. Simulations on real-world collected datasets verify our claim.

Prachi Kohale, Sheetal Girase [7] "Privacy Preservation of Data in Data Mining" in June 2014 describe Angelization is new anonymization technic for privacy preserving publication which is applicable to any monotonic anonymization principle. It produces anonymized relation that achieve privacy guarantee as conventional generalization but permits much more accurate reconstruction of original data distribution. new anonymization technique that is effective as generalization in privacy protection but able to retain significantly more information in micro data. It is applicable to any principle l-diversity, t-closeness and to overcome the drawback of several methods. Compared to traditional generalization it ensures same privacy guarantee preserve significantly more information in micro data.

Abouelela Abdou Hussien, Nermin Hamza, Hesham A. Hefny [8] "Attacks on Anonymization-Based Privacy-Preserving: A Survey for Data Mining and Data Publishing" in Feb 2013 evaluate a

survey for most of the common attacks techniques for anonymization based PPDM & PPDP and explain their effects on Data Privacy. k-anonymity is used for security of respondents identity and decreases linking attack in the case of homogeneity attack a simple k-anonymity model fails and we need a concept which prevent from this attack solution is l-diversity.

Sridhar Mandapati, Dr. Raveendra Babu Bhogapathi, Ratna Babu Chekka [9] "A Hybrid Algorithm for Privacy Preserving in Data Mining" in July 2013 describe Evolutionary Algorithms (EAs) provides effective solutions for various real-world optimization problems. Evolutionary Algorithms are efficiently employed in business practice. In privacy preserving domain, the existing EA solutions are restricted to specific problems such as cost function evaluation. In this work, it is proposed to implement a Hybrid Evolutionary Algorithm using Genetic Algorithm (GA) and Particle Swarm Optimization (PSO). Both GA and PSO in the proposed system work with the same population, k-anonymity is accomplished by generalization of the original dataset. The hybrid optimization is used to search for optimal generalized feature set.

Madhusmita Sahu, Debasis Gountia, Neelamani Samal [10] "Privacy Preservation Decision Tree Based on Data Set Complementation" in April 2013 describe algorithm to protect the sensitive information in data from the large amount of data set. The privacy preservation of data set can be expressed in the form of decision tree, cluster or association rule. Evaluate a privacy preservation based on data set complement algorithms which store the information of the real dataset. So that the private data can be safe from the unauthorized party, if some portion of the data can be lost, then we can reconstructed the original data set from the unrealized dataset and the perturbing data set.

Sailaja. R. J. L, P. Dayaker [11] "Preventing Diversity Attacks in Privacy Preserving Data Mining" in September 2013 implement multilevel trust based PPDM which enables data owners to have freedom to choose the level of privacy needed. Based on this trust level perturbations are made. We built a prototype application that demonstrates the proof of concept. The existing perturbation based PPDM models assume single level trust on data miners. It focus on multilevel trust based PPDM which provides more flexibility to data owner in choosing the level of privacy to data. A challenge in doing so is that malicious data miners can use multiple copies of perturbed data in order to establish the original data. This kind of attack is prevented by using noise correlation matrix across the copies to deny the attackers not to have diversity option.

Seema Kedar, Sneha Dhawale, Wankhade Vaibhav, Pavan Kadam, Siddharth Wani, Pavan Ingale [12] "Privacy Preserving Data Mining" in April 2013 describe survey paper to understand the

existing privacy preserving data mining techniques and to achieve efficiency. Hide sensitive item sets so that the adviser cannot extract the modified database. To solve such problems there are some algorithms presented by various authors worldwide. This survey paper on PPDM can be helpful for finding the loopholes and drawbacks of existing data mining techniques. This survey ensures efficient privacy preserving of data. The use of existing algorithms works towards the direction to reduce the impact of PPDM on the source database. A comparative study all these systems would definitely help in developing a new system that combines all the advantages and overcomes the drawbacks of systems.

P. usha, R. Shriram, W. Aisha Banu [13] "Modified Anonymity Model for Privacy Preserving Data Mining" in February 2013 describe k-anonymization method, every tuple in the dataset released be indistinguishably related to no fewer than k respondents. But the distribution preservation is not guaranteed a modified k-anonymity model is introduced where the privacy in a dataset is preserved while preserving the distribution also By this model a way is found to preserve the privacy of any dataset and also maintain the distribution as well as the utility.

Pawan R Bhaladhare, Devesh C Jinwala [14] "A Sensitive Attribute based Clustering Method for k anonymization" in March 2012 describe method grouping similar data together based on sensitive attribute and then anonymizes them. The main intention is to minimize information loss and data utility, while also protecting the sensitive attributes and private information of an individual, a sensitive attribute based clustering approach for k-anonymization was proposed, which shows comparable result with respect to information loss and execution time. Based on the investigations of the information loss and data utility with respect to the current privacy preserving approaches. To protect the data many methods are modifying the quasi-identifier data and some are using rules to hide the data from disclosure. Currently, quasi-identifier has to be modified in order for no information to be lost and data utility to be maintained; as a result new methods need to be developed without any modification. To investigate a hybrid approach: Many algorithms are based on classification approaches and some are based on clustering approaches. Such hybrid approach can potentially provide new and better ways to protect the privacy.

A.K. Ilavarasi, D. Jeniffa, Dr. B. Sathiyabhama [15] "Post Anonymization Techniques in Privacy Preserved Data Mining" in May 2012.Propose a model in which a multi-decision tree classifier is built on the better than and training duration shorter than the normal ID3 based ADABOOST classifier.

Proposed work

K-Anonymity is a method for providing privacy preservation by ensuring that data cannot be displayed to an individual. The main purpose is to protect individual privacy. In a k-anonymous dataset, if any identifying information is found in the original dataset with k tuples then first we identifies quasi-identifiers i.e. the tuple that clearly distinguish the given tuple in database. Then we are applying MW algorithm for suppression based Approach.In this algorithm we are identifying quasi-identifiers and we are computing A k-partition which is a collection of disjoint subsets of rows in which each subset contains at least k rows and the union of these subsets is the entire table. And next we are replacing each record having with — * ‖. In suppression based approach we are using diffie Hellman key exchange algorithm to generate private secure key. Then we are applying AES (Advanced Encryption Standard) algorithm to encrypt and decrypt data by using the key generated by the diffie Hellman key exchange algorithm. In this approach we are dealing with encrypted data not directly with the original data. When user enters his information then we are encrypting his information by using AES and we are also encrypting all data in table using same algorithm. If information from user matches with table information this tuple will decrypted and inserted into table. In Generalization based Approach we are replacing the value in table with the more general values. If the data entered by the user matches with the value being replaced by the general value then this record will replaced by the general value and these general values being inserted into table. In this proposed system we use AES algorithm and Diffie–Hellman key exchange algorithm. We use AES algorithm for improving the quality of overall system. The major reason for using AES is that AES working under three approved key length: 128 bits, 192 bits, and 256 bits. An algorithm starts with a random number, in which the key and data encrypted with it are scrambled though four rounds of mathematical processes and make system stronger. The another Diffie–Hellman key exchange algorithm is used for exchanging cryptographic keys. This algorithm allows two parties that has no prior knowledge of each other can share a shared key for communications by exchanging data over a public network.

CONCLUSION

In the literature survey, I have studied that lot many work is done only for privacy preserving data mining. For that purpose authors have used different methods. So new method can be implemented for improving security in annonymization method.

ACKNOWLEDGEMENT

I represent my sincere gratitude to Prof. S. W. Mohod, Assistant Professor, Sr. Gr., BDCE, Sewagram, for her constantly guidance throughout the work and support.

REFERENCE:

1. Abouelelaabdou Hussain, Nagy Ramdan, Hesham "Utility based annonymization using generalization boundries to protect sensitive attributes" in journal of information security.
2. Dr. K Sagar "A novel annonymization technique for privacy preserving data publishing" in international journal of emerging technology and advance engineering.
3. Guang Li, Yadong Wang, Xiahong Su "A new algorithm independent method for privacy preserving data mining" in journal of computational information system.
4. Ms. R. Kavitha, Prof .D. Vanathi "A study of privacy preserving data mining techniques" in international journal of science and applied information technology.
5. Tamaszoltan G al, Gabor Kovacs"Survey on privacy preserving data mining techniques in health care database" in acta univ.
6. Wei Peing, Feng Li, Xukaizou, Jie, Fellow "A two stage deannonymization attack against annonymized social network" in IEEE Transaction.
7. Prachi kohale, Sheetal girase "Privacy preservation of data mining "In journal of engineering reaserch and application.
8. Abouelelaabdouhussien, Hesham "Attacks on annonymization based privacy preserving "In journal of information security.
9. Sridhar mandapati,Dr. Raveendrababu, Ratnababuchekka" A hybrid algorithm for privacy preserving in data mining " in I. J. Intelligent system and application.
10. Madhu susmita sahu, Debasis gountia, Neelam anisamal "privacy preservation decision tree based on data set complementation" in international journal of innovative reaserch in computer.
11. Sailaja R. J. L , P.Dayakar "Preventing diversity attack in privacy preserving data mining" in international journal of trends and technology.
12. Seemakedar, Snehal dhawale, Pavan kadam "Privacy preserving data mining" in international journal of advanced reaserch in computer.
13. P. Usha, R. Shriram, W. Aishabanu"Modified anonymity model for privacy preserving data mining " in international journal of computer application
14. Pawan Bhaldhare, Devesh Jinwala" A Sensitive attribute based on clustering method for k annonymization" article.
15. A. K. Ilavarsi, D. jennifer" Post annonymization techniques in privacy preserved data mining "in international journal of computer application.