International Journal of Recent Trends in Engineering & Research
ISSN (ONLINE) : 2455 – 1457
IMPACT FACTOR : 4.101

ICNSCET20- International Conference on New Scientific Creations in Engineering and Technology

DATA PRIVACY PROTECTION USING VERTICAL FRAGEMENTATION AND DATA SUPPRESSION TECHNIQUES

M.r.K.JEEVA[1] M.E (Assistant professor) , P.RADHIKA[2] M.E (final year) *[1]Computer science, kurinji college of Engineering and Technology , [2]Computer science, kurinji college of Engineering and Technology*

ABSTRACT

To a rapid advancement in the electronic commerce technology, the use of online banking has dramatically increased. Online banking is used for purchasing goods and services with the help of virtual card and physical card whereas virtual card for online transaction and physical card for offline transaction. As net banking becomes the most popular mode of payment for both online as well as regular purchase, it provides cashless shopping. It will be the most convenient way to do online shopping, paying bills etc. In online payment mode, attackers need only little information for doing fraudulent transaction (secure code, card number, expiration date etc.). In this purchase method, mainly transactions will be done through Internet or telephone. To commit fraud in these types of purchases, a fraudster simply needs to know the card details. Most of the time, the genuine cardholder is not aware that someone else has seen or stolen his card information. Hence, risks of fraud transaction using banking information have also been increasing. In the existing cyber security system, fraudulent transaction will be detected after transaction is done. It is difficult to find out fraudulent and regarding loses will be barred by issuing authorities. So in this project we can implement vertical level server system to partition the intermediate gateway with improved security. Transaction details are splited and stored as sensitive attributes in primary and secondary servers. And also implement data suppression scheme to replace the string and numerical characters into special symbols to overcome the traditional cryptography schemes.

Keywords—component; formatting; style; styling; insert (keywords) (minimum 5 keyword require)

I. INTRODUCTION

1.1 OVERVIEW

Data mining is the computing process of discovering patterns in large data sets involving methods at the intersection of machine learning, statistics, and database systems.

It is an interdisciplinary subfield of computer science. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. Data mining is the analysis step of the "knowledge discovery in databases" process, or KDD. Data mining (the analysis step of the "Knowledge Discovery in Databases" process, or KDD),

a field at the intersection of computer science and statistics, is the process that attempts to discover patterns in large data sets.

It utilizes methods at the intersection of artificial intelligence, machine learning, statistics, and systems. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use Aside from the raw analysis step, it involves database and data management aspects, data pre-processing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures, visualization, and online updating.

The actual data mining task is the semi-automatic or automatic analysis of large quantities of data to extract previously unknown, interesting patterns such as groups of data records (cluster analysis), unusual records (anomaly detection), and dependencies (association rule mining, sequential pattern mining).

This usually involves using database techniques such as spatial indices. These patterns can then be seen as a kind of summary of the input data, and may be used in further analysis or, for example, in machine learning and predictive analytics.

For example, the data mining step might identify multiple groups in the data, which can then be used to obtain more accurate prediction results by a decision support system. Neither the data collection, data preparation, nor result interpretation and reporting is part of the data mining step, but do belong to the overall KDD process as additional steps

2. LITERATURE REVIEW

2.1 TITLE: SECURITY AND PRIVACY CONCERNED ASSOCIATION RULE MINING TECHNIQUE FOR THE ACCURATE FREQUENT PATTERN IDENTIFICATION

AUTHOR: T. NUSRAT JABEEN YEAR: 2018

Data mining is typically an inter-disciplinary field with its basis formed in enterprise decision support. Data mining is not dedicated for the analysis of small datasets. It is regarded to be the job of finding fascinating and concealed patterns/data from huge chunks of information in cases where the data is found in databases, data repositories, OLAP or other information present in the repository.Data mining is also involved with a combination of methodologies from several disciplines like database technology, statistics, machine learning, neural networks, fuzzy and rough set theory, knowledge representation, inductive logic programming, information retrieval and etc. In data mining, the elementary issue is to discover the frequent item set found in the massive database.

The mining of frequent item set finds significance in a broad array of application fields like bioinformatics, web usage mining etc. Multiple numbers of diverse algorithms has been introduced for discovering the frequent item set. The Apriori and FP- Growth algorithm are the most widely accepted algorithms used in association rule mining.

The Apriori algorithm is basically a bottom-up approach algorithm or level-wise search algorithm. A subset consisting of frequent item set should also be a frequent item set i.e in case {AB} is a frequent item set, then both A and B must be frequent item sets, known as Apriori property.Extracting Association Rules for Distributed Association Rules (EAR4DAR) Algorithm was suggested by salih that does the extraction of the association rule from local association rules into global association rules over distributed systems.

Distributed Frequent Item Mining algorithm was formulated by lamine and ke-chadi that creates diverse clusters and grid environments.

Demerits: Databases might collude with one another

2.2TITLE : SET-VALUED DATA ANONYMIZATION MAINTAINING DATA UTILITY AND PROPERTY

AUTHOR: DEDIGUNAWAN YEAR: 2018

Publishing a set-valued database such as those containing web click log, customer transaction and trajectory data allows one to analyze its data and is beneficial for the public. Anun solved problem of publishing database to the public is that sensitive information such as personal private preference can be linked to a specific individual.

Therefore, an action such as data modification to achieve data anonymity should be taken before publishing the database. Pseudonym is one simple way to protect the data. It alters certain attributes such as name or personal identification number into pseudonym to create anonymous data.

These techniques have been proved insecure since an adversary can commit identity linkage attack by comparing some attributes in the released database with those in other publicly available data bases such as voter-lists and associating are cord with a specific individual. Performing data modification to achieve maximum anonymity with keeping data utility known to be NP-hard problem.

Set-valued database publication has been attracting much attention due to its benefit for various applications like recommendation systems and marketing analysis.

However, publishing original database directly is risky since an unauthorized party may violate individual privacy by associating and analyzing relations between individuals and set of items in the published database, which is known as identity linkage attack.

Generally, an attack is performed based on attacker's background knowledge obtained by a prior investigation and such adversary knowledge should be taken into account in the data anonymization. Various data anonymization schemes have been proposed to prevent the identity linkage attack.

However, in existing data anonymization schemes, either data utility or data property is reduced a lot after excessive database modification and consequently data recipients become to distrust there leased database.

Demerits:

Adversary get knowledge from database

2.3 TITLE: FREQUENT ITEMSETS MINING WITH DIFFERENTIAL PRIVACY OVER LARGE-SCALE DATAAUTHOR AND YEAR: XINYU XIONG, 2018

In recent years, with the explosive growth of data and the rapid development of information technology, various industries have accumulated large amounts of data through various channels. To discover useful knowledge from large amounts of data for upper-layer applications (e.g. business decisions, potential customer analysis, etc.), data mining has been developed rapidly. It has produced a positive impact in many areas such as business and medical care. Along with the great benefits of these advances, the large amount of data also contains privacy sensitive information, which may be leaked if not well managed. For instance, smart phone applications are recording the whereabouts of users through GPS sensors and are transferring the data to their servers.

Demerits: High level relative error

2.4 TITLE: TWO PRIVACY-PRESERVING APPROACHES FOR PUBLISHING TRANSACTION DATA STREAMSAUTHOR AND YEAR: JINYAN WANG, 2018

Privacy-preserving data publishing provides methods and tools for publishing useful information while preserving individual privacy. Recently, it has received wide attention from academia and industry, and many approaches have been proposed for different data publishing scenarios.

Demerits: Difficult to support batch processing

2.5TITLE: IDENTITY-BASED PROXY-ORIENTED DATA UPLOADING AND REMOTE DATA INTEGRITY CHECKING PUBLIC CLOUD YEAR: H WANG 2016

In public cloud computing, the clients store their massive data in the remote public cloud servers. Since the stored data is outside of the control of the clients, it entails the security risks in terms of confidentiality, integrity and availability of data and service.

Remote data integrity checking is a primitive which can be used to convince the cloud clients that their data are kept intact. In some special cases, the data owner may be restricted to access the public cloud server, the data owner will delegate the task of data processing and uploading to the third party, for example the proxy.

Demerits: Provide formal security Pro

2.6 TITLE: PRIVACY-PRESERVING AND REGULAR LANGUAGE SEARCH OVER ENCRYPTED CLOUD DATA AUTHOR: K. LIANG

Much like the popularity of portable personal electronic devices, cloud storage service has been booming over the last decade. Its outstanding advantages, such as considerable storage space, flexible accessibility and convenient data retrieval, strongly catch the attention of Internet users. Accordingly, to date not only individuals but also industries, research institutes prefer to remotely store their data to cloud servers, such that they can get rid of the burden of local data management and maintenance. This makes cloud storage service share a great piece of market cut in the field of data management even in the ear of big data.

Demerits: Size of search token is high

2.7 TITLE: TWO BIRDS WITH ONE STONE: TWO-FACTOR AUTHENTICATION WITH SECURITY BEYOND CONVENTIONAL BOUND AUTHOR: D. WANG

To address the issue of password leakage from a compromised server, threshold password-only authentication schemes have recently been proposed. In such schemes, the password

Demerits: Difficult to detect user card corruption

2.8 TITLE: FULLY HOMOMORPHIC ENCRYPTION USING IDEAL LATTICES AUTHOR: C. GENTRY

We propose a fully homomorphic encryption scheme – i.e., a scheme that allows one to evaluate circuits over encrypted data without being able to decrypt.

Our solution comes in three steps. First, we provide a general result – that, to construct an encryption scheme that permits evaluation of arbitrary circuits, it suffices to construct an encryption scheme that can evaluate (slightly augmented versions of) its own decryption circuit; we call a scheme that can evaluate its (augmented) decryption circuit bootstrappable.

Demerits:

Provide -polynomials larger identifier

2.9 TITLE: HOMOMORPHIC ENCRYPTION FROM LEARNING WITH ERROR CONCEPTUALLYSIMPLER,ASYMPTOTICALLYFASTER,ATTRIBUTE-BASED AUTHOR: C. GENTRY

We describe a comparatively simple fully homomorphic encryption (FHE) scheme based on the learning with errors (LWE) problem. In previous LWE-based FHE schemes, multiplication is a complicated and expensive step involving "linearization". In this work, we propose a new technique for building FHE schemes that we call the approximateeigenvector method. In our scheme, for the most part, homomorphic addition and multiplication are just matrix addition and multiplication.

Demerits: Large size of cipher text

2.10 TITLE: EFFICIENT FULLY HOMOMORPHIC ENCRYPTION FROM (STANDARD) LWE AUTHOR: ZVIKA BRAKERSKI

The main building block in Gentry's construction (a so-called somewhat homomorphic encryption scheme) was based on the (worst-case, quantum) hardness of problems on ideal lattices. 1 Although lattices have become standard fare in cryptography and lattice problems have been relatively well studied, ideal lattices is a special breed that we know relatively little about. Ideals are a natural mathematical object to use to build fully homomorphic encryption in that they natively support both addition and multiplication (whereas lattices are closed under addition only). Indeed, all subsequent constructions of fully homomorphic encryption relied on ideals in various rings in an explicit way. Our first contribution is the construction of a "somewhat" homomorphic encryption scheme whose security relies solely on the (worst-case, classical) hardness of standard problems on arbitrary (not necessarily ideal) lattices.

Demerits

Computationalcomplexityishigh

3.RESEARCH METHODLOGY

3.1 OVERVIEW OF THE PROJECT

A credit network m works form the basis of several Sybil-tolerant social networks, spam-resistant communication protocols, and payment systems. Existing systems, however, expose agents' trust links as well as the existence and volumes of payment transactions, which are considered sensitive information in social environments or in the financial world.

One of the most important challenge is, using supervised data mining technique like learning machine or classification relies on accurate identification of fraudulent and non-fraudulent transactions, however these information usually do not exists or limited or confidential . Financial institutions prefer to not disclose this kind of information and categorize them in high-risk data, so accessing to this kind of data is very restricted.

Therefore the process has difficulty in step "Applying Models" and "Pattern evaluation", so the extracted knowledge might not be cover all fraud scenarios and it increase the error and decrease the accuracy and finally the Decision Support System (DSS) accuracy is decreased as well.

So many researches have done to fill this gaps and present models or techniques to overcome these issues and enhance the DSS models trust between agents in a distributed environment and enables payments between arbitrary pairs of agents. With their flexible design and robustness against

intrusion, credit net. So we can implement vertical clustering algorithm and construct the rules as K-Anonymity approach

3.1.1 EXISTING SYSTEM

A credit network models trust between agents in a distributed environment and enables payments between arbitrary pairs of agents. With their flexible design and robustness against intrusion, credit networks form the basis of several Sybil-tolerant social networks, spam-resistant communication protocols, and payment systems.

Existing systems, however, expose agents' trust links as well as the existence and volumes of payment transactions, which is considered sensitive information in social environments or in the financial world.

This raises a challenging privacy concern, which has largely been ignored by the research on credit networks so far.

Privacy preserving standards have been created recentlybecause sensitive information is now frequently stored on computers that are attached to the Internet.

Also many tasks that were once done by hand are carried out by computer; therefore there is a need for Information Assurance (IA) and security.

Privacy preserving is an important in order to guard against identity theft. Businesses also need security because they need to protect their trade secrets and proprietary information.

Cyber-terrorism is one of the major terrorist threats posed to our nation today. As we have mentioned earlier, this threat is exacerbated by the vast quantities of information now available electronically and on the web.

Homomorphic encryption is a form of encryption which allows specific types of computations to be carried out on cipher text and obtain an encrypted result which decrypted matches the result of operations performed on the plaintext.

For instance, one person could add two encrypted numbers and then another person could decrypt the result, without either of them being able to find the value of the individual numbers

3.1.1.1 DISADVANTAGES

- Unauthorized person can view the details easily. So security was less
- Maintain the details in single server
- Need large amount of storage space for store the encrypted data.
- Easily hack the details.

3.1.2 PROPOSED SYSTEM

With the advent of communications techniques, e-commerce as well as online payment transactions are increasing day by day.

Along with this financial frauds associated with these transactions are also intensifying which result in loss of billions of dollars every year globally.

Also the various types of benefits like cash back, reward points, interest-free credit, discount offers on purchases made at selected stores, and so forth tempt the customers to use credit card instead of cash for their purchases.

The major problem for e-commerce business today is that fraudulent transactions appear more and more like legitimate ones and simple pattern matching techniques are not efficient to detect fraud.

We can implement vertical clustering algorithm to cluster the datasets into more than one level. Subsets of attributes (that is, columns) form the fragments. Rows of the fragments that correspond to each other have to be linked by a tuple identifier. A vertical fragmentation corresponds to projection operations on the table.

Data from the fragments can be recombined to result in the original data set. For vertical fragmentation, the join operator is used on the tuple identifier to link the columns from the fragments; in horizontal fragmentation, the union operator is used on the rows coming from the fragments.

And also implement K-Anonymity algorithm which is a property possessed by certain anonymized data. Given person-specific field-structured data, produce a release of the data with scientific guarantees that the individuals who are the subjects of the data cannot be re-identified while the data remain practically useful. A release of data is said to have the k-anonymity property if the information for each person contained in the release cannot be distinguished from at least k-1 individuals whose information also appear in the release.

The various procedures and programs for generating anonymised data providing k-anonymity protection have been patented

3.1.2.1 ADVANTAGES

- Hackers can be difficult to hack the data from server
- Reduce the time complexity and computational complexity
- Reduce the manual work
- Overcome the guessing attacks, man-in-middle attacks and reply attacks

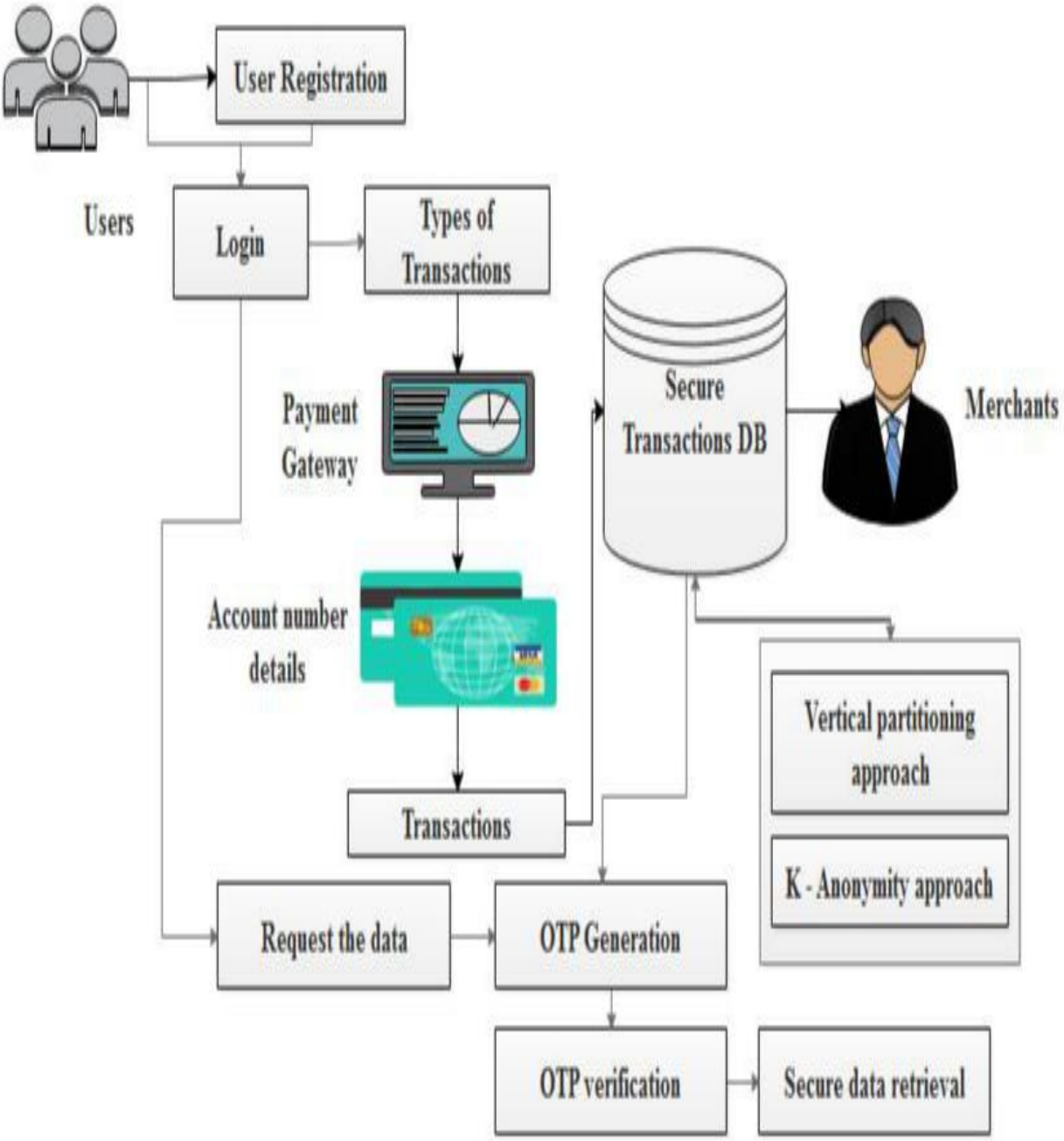

Fig 3.1 SYSTEM ARCHITECTURE

3.2 MODULES DESCRIPTION

1. BANK INTERFACE CREATION
2. TRANSACTION DETAILS
3. VERTICAL PARTITIONING APPROACH
4. DATA SUPPRESSION
5. AUTHORIZED ACCESS

4.CONCLUSION AND FUTURE ENHANCEMENT

CONCLUSION

The primary goal of data privacy is the protection of personally identifiable information. In general, information is considered personally identifiable if it can be linked, directly or indirectly, to an individual person. Thus, when personal data are subjected to mining, the attribute values associated with individuals are private and must be protected from disclosure. Miners are then able to learn from global models rather than from the characteristics of a particular individual.

FUTURE ENHANCEMENT

In future work, we can extend framework to implement various algorithms to enhance the security with dynamic splitting. Multiple K-anonymity is implemented for anonymized data with improved security.

REFERENCES

1. Jabeen, T. Nusrat, M. Chidambaram, and G. Suseendran. "Security and privacy concerned association rule mining technique for the accurate frequent pattern identification." International Journal of Engineering & Technology 7.1.1 (2018): 19-24.

2. Gunawan, Dedi, and Masahiro Mambo. "Set-valued Data Anonymization Maintaining Data Utility and Data Property." Proceedings of the 12th International Conference on Ubiquitous Information Management and Communication.ACM, 2018.

3. Xiong, Xinyu, et al. "Frequent Itemsets Mining with Differential Privacy over Large-scale Data." IEEE Access (2018).

4. Wang, Jinyan, Chaoji Deng, and Xianxian Li. "Two Privacy-Preserving Approaches for Publishing Transactional Data Streams." IEEE Access 6 (2018): 23648-23658.

5. H. Wang, D. He, and S. Tang, “Identity-based proxy-oriented data uploading and remote data integrity checking in public cloud,” IEEE Trans. Inf. Foren.Secur., vol. 11, no. 6, pp. 1165–1176, 2016.

6. K. Liang, X. Huang, F. Guo, and J. K. Liu, “Privacy-preserving and regular language search over encrypted cloud data,” IEEE Trans. Inf. Foren.Secur., vol. 11, no. 10, pp. 2365–2376, 2016.

7. D. Wang and P. Wang, “Two birds with one stone: Two-factor authentication with security beyond conventional bound,” IEEE Trans. Depend. Secur.Comput., 2016.

8. C. GENTRY, “Fully homomorphic encryption using ideal lattice,” Proc. ACM STOC 2009, pp. 169–178.

9. C. Gentry, A. Sahai, and B. Waters, "Homomorphic encryption from learning with errors: Conceptually-simpler, asymptotically faster, attribute-based," in Proc. CRYPTO 2013, pp. 75–92.

10. Z. Brakerski and V. Vaikuntanathan, "Efficient fully homomorphic encryption from (standard) LWE," in Proc. IEEE FOCS 2011, pp. 97–106.

11. Radha Krishnan, B., Vijayan, V., Parameshwaran Pillai, T. and Sathish, T., 2019. Influence of surface roughness in turning process—an analysis using artificial neural network. Transactions of the Canadian Society for Mechanical Engineering, 43(4), pp.509-514.

12. Krishnan, B.R., Ramesh, M., Giridharan, R., Sanjeevi, R. and Srinivasan, D., Design and Analysis of Modified Idler in Drag Chain Conveyor. International Journal of Mechanical Engineering and Technology, 9(1), pp.378-387.

13. Krishnan, B.R., Vijayan, V. and Senthilkumar, G., 2018. Performance analysis of surface roughness modelling using soft computing approaches. Applied Mathematics & Information Sciences, 12(6), pp.1209-1217.

14. KRISHNAN, B.R. and PRASATH, K.A., 2013. Six Sigma concept and DMAIC implementation. International Journal of Business, Management & Research (IJBMR), 3(2), pp.111-114.

