International Journal of Advance Research in Science and Engineering
Vol. No.6, Issue No. 05, May 2017
www.ijarse.com
IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

SECURITY ON DATA MINING TO APPLY SECURITY MECHANISM ON DATA CHUNKS

Namrata Chandrakar [1], Shrikant Tiwari[2], Swarnim Singh Chandel[3]

[1,2,3] *Department of Computer Science & Engineering, Shri Shankaracharya Technical Campus*

SSGI (Faculty of Engineering and Technology) – Bhilai (Chhattisgarh)

ABSTRACT

Security is most important part of every organization. It may be in terms of electronically stored data or may be in terms of hard copy of data. Both or very important for security of organization data. Today most of the data stored electronically and it can be accessible by any one if there is no security on stored data. In other way all data are not important until they mined are analyzed. Instead of applying security on single dataset of mined data, there can be an another way to apply the security on mined. Create an N number of chunks and then apply security on all data chunks with different keys. By apply these process an organization can enhance the security on mined data. Because of all data chunks are secured with different keys.

I. INTODUCTION

Data mining is a technique where, from the raw dataset meaningful information can be extracted. Further by applying the some tools it be analyzed based on the requirement. Pattern or graph can evaluated from these dataset for better understanding. As all know that the data is very important for the organization as well as for personal. It contains private and confidential information related to organization. To retain the privacy of data, organization applied different types of security mechanism.

Every day an organization generates a lot of data. These data is used for future purpose to analyze the business strategy. With the invent of e-commerce site, every site maintain their customer data. These customer data is very essential with respect to privacy and security [1]. To apply the privacy and security policy they applied some security constraints and policy.

Security concerns are directly related with the database and applied tools, security, to protect the data deals many things for the data mining applications and process [2]. Today a lot of data generated every second and the generated data are stored as historical data for every organization, the organization may be private or government. All the generated data is not important for security purpose, but some of them are very important for their organization. These data may contains there business strategy, or some information which is related to privacy. To protect them from intruders there is need to apply some security setting and every organization use their own security policy and some rules.

There are many ways to apply the security and privacy in mined data. Here in the proposed system, the applied algorithm first divide the mined data into chunks after dividing the data into the security mechanism will be applied. For every divided chunk, different keys can be used. Are the same keys also can be used. It depends on

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

the algorithm.

Some Areas of Data Mining and Need to Apply Security:

There are many fields where the data mining techniques can be applied to gain or discover some knowledge based on the organization. The mined information is very essential, so they must contains some security constraints and credentials in terms of security and privacy to access the data. The following are the major subfields where data mining is successfully applied with the security and privacy constraints:

Administration of Health Services: Administrators of health care organizations make hundreds of critical decision on daily basis. The quality of these decisions directly depends on the quality of the information. This information is very essential and concerned with the privacy issue so there in need to apply the security constraints on these type of data to preserve the privacy.

Medical Research and Clinical Care: Most current successful applications of data mining in health informatics are in the subfield of medical research. Physicians and nursing practitioners make diagnostic decision and treatment recommendations based on history, medical imaging, lab results and other text or multimedia records of patients. The applications of health informatics in clinical care decision-making are known as Clinical Decision Support System. Clinical Decision Support System uses prediction models for diagnosis (aiding in the determination of the existence or nature of a disease) and prognosis (the forecast of the probable outcomes of an illness). The all the research for the medical science, are the data related to field where the misuse of these type of data may generate a big problem, so there must be a system where need to apply the security and privacy.

Location Based Data Analysis: Location based data contains the information related to the location of user's. It contains the routing information, saved location and others searching place data related to the user. So while the mining of the location data there is need to apply the privacy and security of the data. Because these type of data can be misused to track the user, if someone hacked.

Marketing Data Analysis and Sales Analysis of Organization: The organization mine their data to extract the some useful information based on the previous analyzed data or from the raw data. They perform various types of comparisons from the different dataset and with comparing previously analyzed data to plan for the market strategy. As all know that the ultimate of data mining is to predict and plan some business strategy for marketing and selling the products. Generally the ecommerce site perform the analysis based on the different criteria like frequent purchased item based on product rating and many more. These type of analysis contains the essential and confidential information related to organization. So there must me need to apply some security on these mined data.

II. SECURITY ANALYSIS

It is very important to store the data securely and manage them [6]. All mined and raw data set are very important and essential component for all the organization because, it contains the data related to that organization which confidential. The data mining security concerns the like Physical Database Integrity, Logical Database Integrity, Element Integrity, Audit ability, Access [2] and may more. The security aspects for the data

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

mining deals with the many things. The most important are the privacy of data and security on data to protect them from intruders.

For the cyber security mining, the anomaly detection techniques are used to detect unusual patterns and behaviors of the data, Link analysis may also be used to trace self-propagating malicious code to its authors and the classification process may be used to find and group various cyber attacks and then use the profiles to detect an attack when it occurs (Bhavani, Latifur, Masud, & Hamlen, 2008). The privacy and security implementation are very important for the data mining application. With the help of the limiting the access of data by the user or by providing some augmentation on data by adding some extra text or by auditing the data also the security can be applied on the data (Chris & Don, 1996).

Security issues and its measures for data mining is major problem now a day. Data mining provides facts and this is not oblivious to the human beings to analyze the data. It also enables the inspection and analysis of huge amount of data. Due to this activity the analyst can leak the information and data of enterprise. Databases are important and essential components of different government and private organizations. To protect the data of the databases used in data warehouse and then data mining is central theme of security system.

III. LITERATURE REVIEW

As stated that security and privacy is most important part of organization to maintain the originality of data and to secure the business related essential information and reports. Here the research published by some author is analyzed based on data mining and security.

Author explains the security issues and concerns the role of self-regulation and the user on privacy and security protections, data protection laws, regulatory trends, and the outlook for privacy and security legislation. Author focuses on online privacy and the security related issues. Although data mining is a comparatively new term but the all technology and tools are not able to mine the all type of data [3]. Author explains the functionality of data mining which are used for specific kind of patterns. Also outline issues related to data mining such as Interactive mining of knowledge at multiple levels of abstraction, Incorporation of background knowledge, Data Mining query language and ad-hoc data mining, Expression and visualization of data mining results, Handling noise and incomplete data.

Author compares the results with the different data mining techniques applied on various security threats and proposed the solution strategy for further requirement. Security applications can be for national security to fight against terrorism attacks or for cyber security to protect computers and networks against corruption (worms and viruses), intrusion, attack, malware and denial of services. Author outlines various advantages and disadvantages of data mining algorithms. Through a comparison based on key criterions the learning algorithms are adaptable and good for unknown pattern recognition but with long process time while linear algorithms are faster but not proper for unknown attacks[4].

Here author [2] deliberates the fact that data mining is a technique to dig the data from the large databases for analysis and executive decision making. Security aspect is one of the measure requirements for data mining applications. Author has given the different security measures likes privacy, correctness, integrity, sensitivity

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

and mistaken data. Author concluded that the data mining security measures are very important for the data mining applications. A security measures should be implemented on behalf of the company policies.

Data mining is the process of discovering insightful, interesting, and novel patterns, as well as descriptive, understandable and predictive models from large-scale data. In this paper, different tasks of data mining are introduced. The goal of the analysis is to specify a relationship between the dependent variable and explanatory variables the as it is done in regression analysis. To proceed with directed data mining techniques the values of the dependent variable must be known for a sufficiently large part of the data set[5].

IV. PROPOSED TECHNIQUES AND PROCESS

One of the key issues raised by data mining technology is not a business or technological one, but a social one. It is the issue of individual privacy. Data mining makes it possible to analyze routine business transactions and gleans a significant amount of information about individuals buying habits and preferences. Another issue is that of data integrity. Clearly, data analysis can only be as good as the data that is being analyzed. A key implementation challenge is integrating conflicting or redundant data from different sources.

Applying the Security on Mined Data: The most important task after successfully mining the data is to how to protect them from intruders. The organization data contains the essential information so there is need to apply the security on data by applying some security mechanism. Here the proposed system provides a way to protect the data from the intruders and unauthorized user's. The proposed is based on the dividing the data into number of parts called chunks and then encrypt them with some keys. The block diagram is represented to understand the flow of applying the security of mined data.

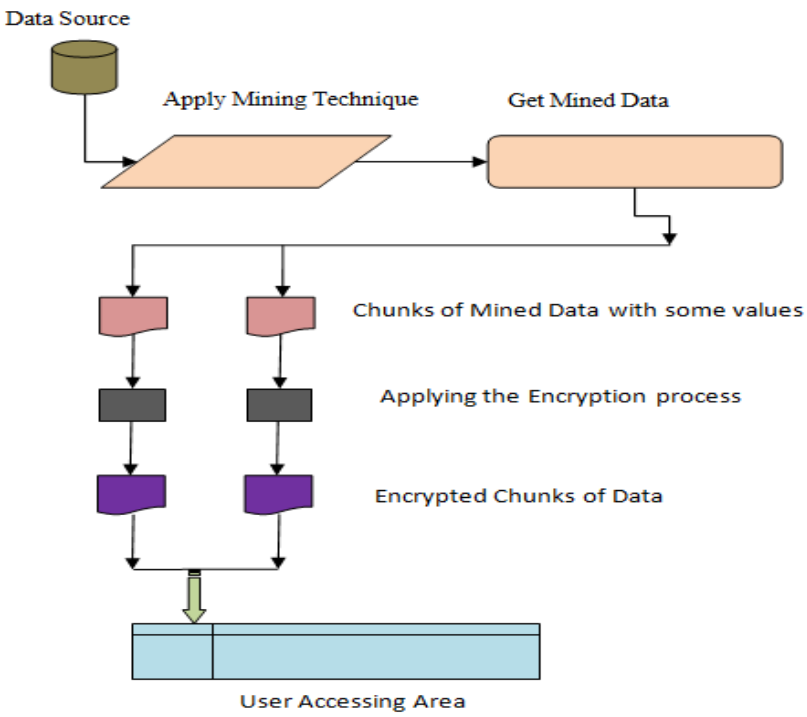

Figure 1: Block Diagram of Creating Data Chunks and Encryption of Chunks

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

The above figure show that how the data will be divide into chunks and then encrypted. The decryption process is just an opposite the above process.

Algorithm: The chunks are the divided parts of mined data. It may any number, depends on the input provided by the user. To apply the security in mined data the following algorithms and process are used.

Pseudo Code for Chunks Creation Encryption.

1. Input the number of chunks, n.
2. Apply the methods to creating the chunks based on input.
3. Apply the encryption techniques to encrypt the chunks.
4. End.

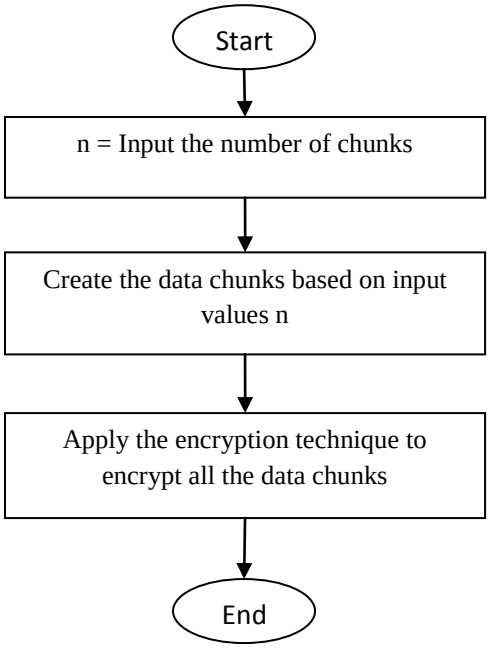

Figure 2: Flow chart pseudo code to creating chunks and encryption the chunks.

Pseudo Code for Decryption and Merging the Chunks .

1. Get the chunks directory.
2. Apply the methods to decrypt the chunks.
3. Merge the chunks.
4. Get the data from the chunks to display the user.
5. End.

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

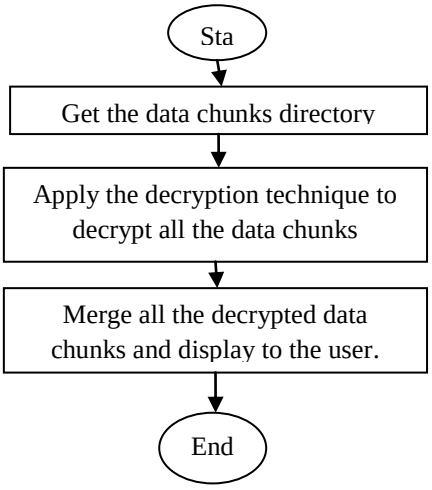

Figure 3: Flow chart pseudo code to decrypting the chunks and merging the chunks.

The above flow diagram shows the overall process of creation of chinks and encryption of chunks also the decryption of chunks and merging the chunks. In this process first the file is divided into different chunks and then encrypt.

V.CONCLUSION

Security and data privacy is one the most challenging part of the data mining process. As it is known that the data contains the organization business data, as well it contains some personal data related to employee. To secure these data is very important by applying some security mechanism. The proposed methodology helps to gain this problem by applying the security on data. First the data is divided into chunks and then after by applying the encryption method with some key, the data chunks will be encrypt. In decryption process, the encrypted data is decrypt and then chunks will be merged.

REFERENCES

[1] L. XU, C. JIANG, J. WANG, J. YUAN and Y. REN, "Information Security in Big Data: Privacy and Data Mining", *IEEE ACCESS*, vol. 2, pp. 1149-1176, 2014.

[2] A. Gupta, V. Bibhu and R. Hussain, "Security Measures in Data Mining", *International Journal of Information Engineering and Electronic Business*, vol. 4, no. 3, pp. 34-39, 2012.

[3] D. Kumar Singh and V. Swaroop, "Data Security and Privacy in Data Mining: Research Issues & Preparation", *International Journal of Computer Trends and Technology*, vol. 4, no. 2, pp. 194-200, 2013.

[4] M. Monshizadeh and Z. Yan, "Security Related Data Mining", *IEEE International Conference on Computer and Information Technology*, no. 775-782, 2014.

[5] A. V. Saurkar, V. Bhujade, P. Bhagat and A. Khaparde, "A Review Paper on Various Data Mining Techniques", *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 4, no. 4, pp. 98-101, 2014.

[6] V. Inukollu, S. Arsi and S. Rao Ravuri, "Security Issues Associated with Big Data in Cloud Computing", International Journal of Network Security & Its Applications, vol. 6, no. 3, pp. 45-56, 2014.