

To Enhance The Security In Data Mining Using Integration Of Cryptographic And Data Mining Algorithms

Gurpreet Kaundal

Student, Lovely Professional University, Phagwara, Pin no. 144402

Abstract: - Data mining is used for to process the information and obtained the knowledge or interesting patterns from the large amount of data. Data mining is a ground breaking way to leverage the information. The information that data mining provides can lead to an immense improvement in the quality and dependability of business decision making. The objective of this work is that to improve the security of the data. Everyday customers or client's information is updated, this information is very important data for the clients or the customers, they never want to disclose with anyone. So, security of information/data is required, needs to maintain it and also needs to reduce the complexity of data, and to make the sensitive data more reliable

Keywords: - Data Mining, Security, Hill cipher, Attacks.

I. INTRODUCTION

Data mining is the process of discovering interesting patterns from massive amounts of data or database. A Pattern is interesting or if it valid then easily understood by humans. In data mining knowledge obtained from patterns which are interested, data mining is a form of knowledge mining [1]. The process of Knowledge discovery contains the cleaning of data, integration of data, selection of data, and transformation of data, evaluate the patterns and present the knowledge. Data mining is the innovative way to gaining new and valuable information. Data mining is a ground breaking way to leverage the information. The information that data mining provides can lead to an immense improvement in the quality and dependability of business decision making. Data mining can identify patterns in company data.

II. TOOLS USED IN DATA MINING:

There are different types of data mining tools that are available in marketplace. Most data mining tools can be classified into three categories:

- Traditional Data Mining Tools
- Dashboards Data Mining Tools
- Text Mining Tools

Traditional Data Mining Tools: The data mining programs helps the many companies to establish their data patterns. It can be done with the help of complex algorithms and techniques. These tools are installed on the desktop to monitor the data. It also helps to capture the information, which is residing outside a database.

Dashboards: These tools are installed in computers. It helps to monitor information in database. It also provides the information regarding data change, and data updates. This functionality makes dashboards easy to use and particularly appealing to managers who wish to have an overview of the company's performance.

Text Mining Tools: The third type of data mining tool is called text mining tool. It has the ability to mine data from different kind of text. These tools scan content and convert the selected data into a format that is compatible with the tool database. Scanned content can be unstructured or structured. Capturing these inputs can provide organizations with a wealth of information that can be mined to discover trends, concepts, and attitudes.

III. SECURITY IN DATA MINING:

The Security of data is that to protect the data from the any third party which are not authorized. In simple words the data security is that to protect the data from the unauthorized access of the users. In database very important data are stored; this data is very useful for making the decisions. Security is a way which is used to provide protection again loss of data, damage, crime and danger. In the data mining, security is a defined in the form of protection that provides the conditions which are related with security or improves the security conditions. The main objective of the security is that to protect the data from the any disaster, corruption and from the theft, and allows the authorized users to access the information [2]. The system security is that to

protect the system or system data from any loss or danger, and allows to users which are authorized access the data in easy way.

IV. RELATED WORK

Dansana J. *et al*/Tools Used in Data Mining:

There are different types of data mining tools that are available in marketplace. Most data mining tools can be classified into three categories:

- Traditional Data Mining Tools
- Dashboards Data Mining Tools
- Text Mining Tools

Traditional Data Mining Tools: The data mining programs helps the many companies to establish their data patterns. It can be done with the help of complex algorithms and techniques. These tools are installed on the desktop to monitor the data. It also helps to capture the information, which is residing outside a database.

Dashboards: These tools are installed in computers. It helps to monitor information in database. It also provides the information regarding data change, and data updates. This functionality makes dashboards easy to use and particularly appealing to managers who wish to have an overview of the company's performance.

Text Mining Tools: The third type of data mining tool is called text mining tool. It has the ability to mine data from different kind of text. These tools scan content and convert the selected data into a format that is compatible with the tool database. Scanned content can be unstructured or structured. Capturing these inputs can provide organizations with a wealth of information that can be mined to discover trends, concepts, and attitudes. (2013) described the privacy preserving of data and also described the data mining decision tree algorithm and CART algorithm. The CART algorithm is uses as a centralized algorithm; in the distributed environment it is not used or difficult to use it. In which uses the set of protocols for secure size and sum which are related with security. In which the author proposed the CART algorithm for security of data in between the multiple parties in vertically environment [3]. The CART algorithm implements in vertically partitioned environment and also use the privacy preserving techniques for data leakage. The privacy preserving techniques are used for to minimize the leakage of data. In data mining process the all attributes of the gini index are securely compute by using this algorithm. In future many applications of data mining such as medical, academics and library etc. uses the CART algorithms.

Karthikeswaran D. *et al* (2012) discussed about the data mining privacy preserving, the data is very sensitive information or data, every data many transactions occur and every time the transactional data updated, so need to protect this data from the third party. In which uses the sensitive rule hiding approach for modified the transactions, most of the cases use the hidden rule (sensitive rule) without generated the false rule [4]. The author proposed the sensitive rule for modification of transactional data, this sensitive rule hiding processes more efficient for large databases; rule hiding provides the great impact of the performance. The other issue is that it not work according to user queries and the user specified constraint. This issue is removing by using the sensitive rules for hiding, the main goal to build a system which is used to find the sensitive rule for hiding by the database administrator.

Kaur G. *et al* (2013) discussed the cryptography scheme to increase the security of data and also discussed about the information transfer with confidentiality. The cryptography schemes are used to enhance the security and to protect the data from the security risks. The author proposed the technique for calculating the PSNR and MSE values with the help of hill cipher with DES for enhance the security. For hiding the text data behind the image file by using this technique and increase the security [5]. This scheme uses the any image as a input that are used for calculate the time for text hide or data hide, in which uses the secret text hiding approach, this is combination of advance hill cipher and the DES techniques for security the data in form of confidential, secure the data from the unauthorized access. In future, designed the new algorithm using the artificial intelligence technique for embedding process, this is used for defense research department for ERP systems and also used for data mining.

V. RESEARCH DESIGN AND RESULTS

RESEARCH DESIGN

To protect the important information first of all need to hide the information so that an attacker can't watch it. So store the information by hiding it into images. The information stores inside an image using hill cipher algorithm and store the image into database. Then data mining iterative algorithm uses for to fetch the data/information from ciphered image and data. Data mining iterative algorithm is used to discover information on the basis of iterations. [7] When access the information and retrieve it in form of image, now in this case if an attacker perform any attack successfully then attacker will take only image because information is hidden inside the image, only user can discover that information from image which are ciphered. It will make

information secure. Let's take example of man-in-middle attack, suppose employees are working in a XYZ Pvt. Ltd. And they are accessing their accounts information by using data mining from their database, suddenly an attacker performs man-in-middle attack and attacker discloses all important information.

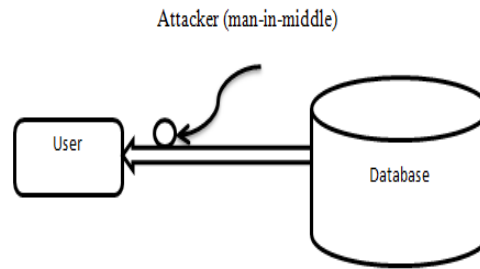


Fig 1 Attack (man-in-middle)

Here in this Fig 1 shown that user is discovering some information and an attacker is also accessing same information from path by using man-in-middle attack. In this research work proposes an idea for to prevent these types of attacks. Stores the information inside an image using hill cipher algorithm and then store it into database. After that uses the iterative algorithm for data mining to fetch the data/information from ciphered image and data. An organization wants to mine the data and generates the knowledge from it. This generated knowledge is important for decision making for an organization. The organizational data may be present in distributed environment and an organization wants to mine the data at one location, so transfer the data through the network and collect the data (or information) at one location. In the network the data are not secure because of security attacks. So, needs to encrypt the data and then transfer the data one location to another location via network. The organization mines the data and generated the knowledge or extracts the knowledge from it. This knowledge is very important for an organization in decision making process. After that uses the data mining iterative algorithm. The Iterative data mining, it will discover information on the basis of iterations such as Preserving Row and Column Margins, Clustering Preserving Structure and Item set Frequencies. The information hidden inside the image, the attacker takes the only image and the user can discover that information from the image, so it will make the information secure as shown in below figure:

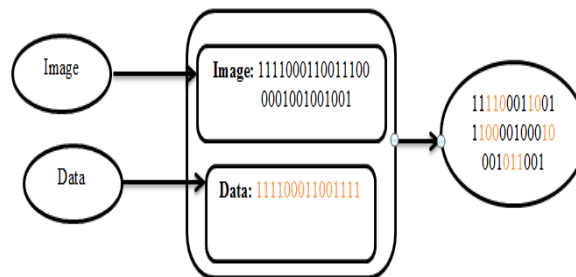


Fig 2Data Hiding Inside the Image

The data mining algorithms and techniques are uses for mining the information. Firstly, take the dataset as input then encrypts the data (or information) by using the hill cipher technique. Then mine the encrypted data with iterative algorithm and generates the knowledge. This knowledge is most important for an organization in decision making.

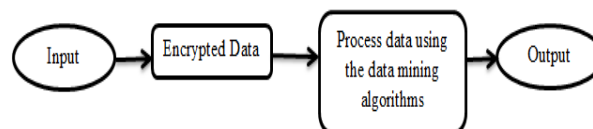


Fig 3 The Research Design

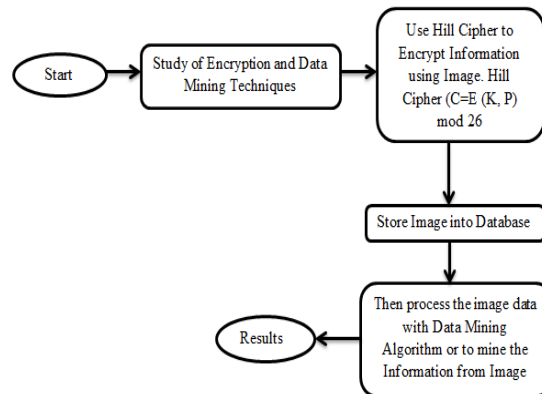


Fig 4 The Research Design Process

This research work will use the hill cipher algorithm for encrypt the information into the image and store the image into the database. The hill cipher algorithm contains the plaintext and the key for encrypted the data into the cipher text, this cipher text is the encrypted data. Firstly, take the dataset as input then uses the Hill cipher to encrypt information to image hill cipher represent as $C = E(K, P) \text{ mod } 26$, as shown in figure 3.4, after that store the encrypted data into the database in form of image. Then mine the information from the image by using the data mining algorithm such as iterative algorithm and process the data (or information) and generates the knowledge from it. This knowledge information is very important information for decision making process in an organization.

VI. RESULT

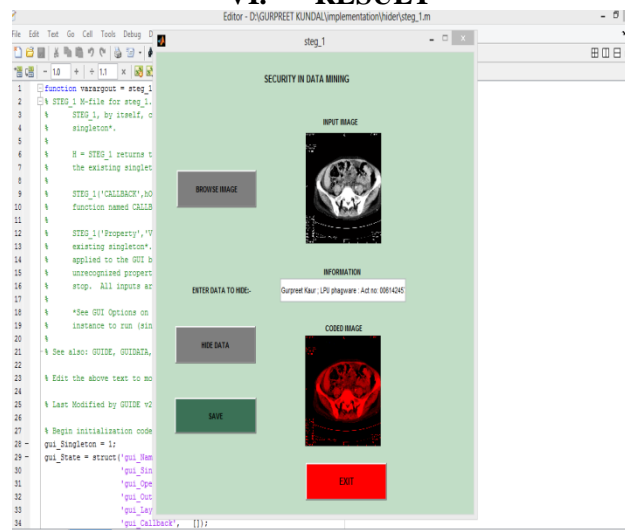


Fig 5: Security in data mining

In this case, we upload an image in our database. We write some information in the text field. Which we want to encode in the image. When we click on hide data an decoded image is shown. . This coded image is developed by merging image and information. Inside the coded image some information is stored. When information is stored in an image, its color is changed due to changes in the binary values.

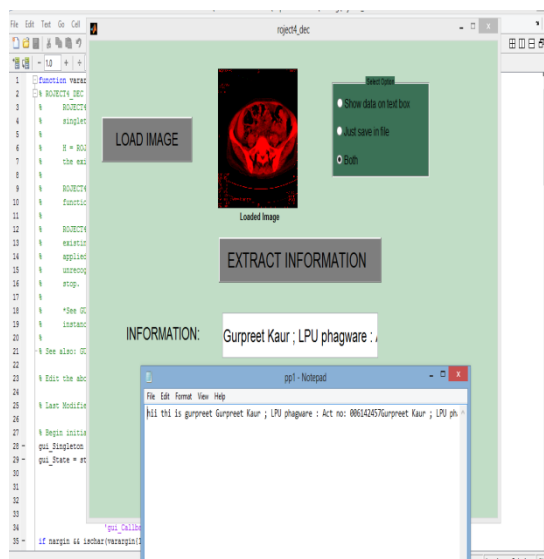


Fig 6: Data Extraction

Here we can extract the information from the image. In this case, embedded image is browse by the user. Hence it give three options. In the first case, when we click on show data on text box, it will extract the information from the iamge and show in the text field. If we click on both, it will also show the information in text field as well as it will save the data in the file.

REFERENCES

- [1] Han, J., and Kamber, M. (2006) *Data Mining: Concepts and Techniques*, Morgan Kaufmann, India, p. 6-10, 33-34.
- [2] <http://blogs.dlt.com/cybersecurity-wednesdays-reduce-cyberpain-restrict-data-access/>
- [3] Dansana Jayanti, DeyDebadutta and Kumar Raghvendra (2013) "A Novel Approach: CART Algorithm for Vertically Partitioned Database in Multi-Party Environment", Proc. of IEEE Conference on Information and Communication Technologies (ICT), pp. 829-834.
- [4] D. Karthikeswarant, V.M. Sudha , V.M. Suresh and A. Javed Sultan (2012) "A Pattern Based Framework For Privacy Preservation Through Association Rule Mining", IEEE-International Conference On Advances In Engineering, Science And Management (ICAESM) , pp.816-821.
- [5] Kaur Gurtaptish, MalhotraSheenam (2013) "A Hybrid Approach for Data Hiding using Cryptography Schemes", International Journal of Computer Trends and Technology (IJCTT).4(8),pp. 2917-2923..
- [6] HeMiao, VittalVijayandZhangJunshan (2013) "Online Dynamic Security Assessment With Missing PMU Measurements: A Data Mining Approach", Proc. of IEEE Transaction On Power System.28 (2), pp. 1969-1977
- [7] Gurpreet Kaundal, Sheveta Vashisht, Disquisition of a Novel Approach to Enhance Security in Data Mining, www.ijaret.org Vol. 1, Issue X, Nov.2013 ISSN 2320-6802