Data Mining: A Necessity For Information Security

Vishal Bhatnagar, Sanur Sharma, Ambedkar Institute of Advanced Communication Technologies and Research, Delhi, India

ABSTRACT:

In today's digital age security of information is of utmost importance. With petabytes of data floating around, it becomes essential that only the authorized users have access to it. With the increasing amount of information, Data Mining has become a significant resource for information security. Data mining provides the best applicable technique, according to the different information security areas, for achieving a desired level of privacy. In this paper we have identified the different emerging areas of information security and the various data mining techniques that can be applied in those areas to mitigate the rising security risks and threats.

Keywords**:** *Data mining, Information security*

1. Introduction

Information security is an area that deals with protecting data from intrusions, malwares, frauds and any criminal activities that are surfacing in digital media with a very fast rate and to maintain non-repudiation, confidentiality and integrity of data. Information security is an essential part for securing information of systems and critical infrastructures. With an increased use of computer applications and internet applications, no matter how much the systems and data are secured there are always some vulnerabilities that arise due to the proliferated use of these applications. More recently with the advancements in the field of information security, data mining techniques have found their place in this area.

Data mining is extraction of hidden, useful and precious data from large and high dimensional databases. It was introduced with an aim to support large databases that are used in various business applications for predicting future trends, analysing data and making proactive decisions. Data mining has emerged as tool that provides its users to identify the vulnerabilities and helps in providing a defensive mechanism against various threats to the information systems. There are various applications of data mining in the field of information security. The most common and largely discussed one such area in the field of information security is intrusion detection where the threats to the system are identified and prevented. A lot of work has been done in this area by the researchers and various data mining techniques have been applied for detection and protection of the systems. With the advancements in the area of information security, the applications of data mining have also increased immensely to various other areas of information security and are not restricted to just intrusion detection and prevention systems.

This paper addresses the various information security areas and the application of various data mining techniques in those areas, which will help researchers to identify and apply various data mining techniques to deal with the security issues that arise in the information security areas. The paper focuses on providing a comprehensive study that identifies the various applications of data mining in terms of security be it the information contained in the systems or the information about the critical infrastructures or establishment of various security policies or security in communication of various systems.

The paper is organised as follows: Section 2 presents the literature survey. Section 3 defines the research methodology used. Section 4 outlines an introduction to information security and its dimensions. Section 5 provides an overview about data mining and its various techniques. Section 6 discusses the role of data mining in information security. Section 7 presents the application of data mining in information security. Section 8 discusses the implications of research and Section 9 concludes by presenting some directions for future research.

2. Literature Survey And Motivation

Research on information security issues requires multi-disciplinary studies spanning different areas like intrusion detection, application security, forensics, access control, web service security etc. To properly understand the use of data mining in these information security areas it is necessary to understand the threats and vulnerabilities that arise in the systems. Data mining techniques apart from detecting threats and vulnerabilities can also provide security using various privacy preserving data mining techniques. There has been an immense growth in the area of data mining and its other forms like stream data mining, web mining, distributed data mining and real time data mining which brought about the different aspects of data mining which motivated us to understand its applications in the most imperative and a requisite in today’s computer and cyber world that is the information security.

A lot of research has been done in the area of information security where plenty of data mining techniques have been applied. The most common and one of the vital areas is intrusion detection and prevention system where the threats and anomalies are identified and the system is protected against any threats and attacks. Data mining has been a major contribution to such systems by providing an implementation of various techniques like classification, clustering, association rule mining, and link analysis to achieve a desired level of security of information systems. Lee and Stofolo (1998) proposed an agent based architecture for intrusion detection systems where learning agents continuously compute and update the models for detection. Lee et al., (2001) focused on issues related to deployment of data mining in real time environments. Nguyen and Choi (2008); Dartigue et al., (2009) have proposed their works in the area of network intrusion detection by using classifier models. Next is the application of data mining in malwares, Komashinskiy and Kotenko (2010) in their works have presented heuristic techniques that can be used for malware detection and is based on static potentially dependent features. Prasad and Ramakrishna (2010a, 2010b) have used data mining techniques in the area of secure software engineering to extract security requirements and finding vulnerabilities and bugs in the software.

Another application of data mining in information security is in the field of forensics. Chen et al., (2004) have discussed crime data mining techniques to identify patterns from both structured and unstructured data for detecting cybercrime. Thongtae and Srisuk (2008) have discussed the application of data mining tools to predict future trends and behaviours of criminals by using the information from large crime databases. Guo and Li (2008) have proposed a combined probabilistic and neuro-adaptive approach to prevent credit card frauds from taking place. For access control authentication and authorization can be done using biometrics and role based access control. Thuraisingham and Khan (2005) have proposed works in the area of biometrics by using techniques like LDA and PCA. In the area of web service security a lot of work has been done in securing virtual communities like social networks and e-commerce and e-business. Zhou et al., (2008) have discussed various anonymization techniques for social network of which one such technique discussed is clustering. Pattanaik and Ghosh (2010) in their work defined the role of data mining in e-payment systems. Kwapisz et al., (2010); Weiss and Lockhart (2011) have their works in the area of mobile phone security and they have used accelerometer data from mobile phones to identify and authenticate users by using data mining techniques. Tang et al., (2010) have presented a mobile user authentication scheme that applies data mining on cell phone's application history and GPS information to identify the user.

For secure transmission of information, data mining techniques are also used in the field of cryptography and steganography for secure multi party computations. Pinkas (2002); Sharma and Ojha (2010) have worked in the area of privacy preserving data mining for cryptography. Hussein et al., (2008) and Pejas (2005); Liu et al., (2008) have applied data mining techniques in the area of steganography and steganalysis respectively. For data base security data mining techniques can also be applied. Chen and Chu (2006) have proposed a Semantic Inference Model (SIM) which uses data mining techniques like bayesian networks for evaluating inference probability of sensitive attributes.

Weaver et al., (2011) have used data mining to control, manage and define security policies and their evolution. Data mining techniques have also been used for security testing. Techniques like fuzzing uses data mining to identify vulnerabilities from the software (Wu et al., 2009). Apart from this data mining techniques have also been used for security auditing by organizations for their complete systems. Sirikulvadhana (2002) proposed the use of data mining tools in place of general audit tools to improve audit performance. Denker (2002) discussed the applications of data mining in the field of auditing and how it helps uncover hidden knowledge

Another area in information security is that of risk management and assessment. Koyuncugil and Ozgulbas (2009) have used data mining for detecting operational risks in financial profiling. A critical infrastructure is another area where data mining is being used to protect sensitive information of various critical infrastructures. Hooper (2010) has implemented data mining techniques which are used to trace back and find anomalies that can harm a critical infrastructure.

One of the very important application areas of data mining is the surveillance systems. It includes systems like wireless sensor networks and satellite tracking systems. Li et al., (2009); Sa et al., (2011) have proposed works on wireless sensor networks using data mining techniques like clustering to perform aggregation of data or clustering of sensor nodes and finding anomalous activities in the networks using agent based models respectively. Kamppi et al., (2009) have identified various threats and risks that occur in satellite tracking systems.

3. Research Methodology

The authors have adopted an empirical research methodology where in the study was done to understand the sensitive nature of information and the various data mining techniques that can be adopted to achieve security in different information security areas. A qualitative research was done across various journals and conference data bases to gather the relevant material in the field of information security. Accordingly different information security areas were identified and the role of data mining was found in terms of security. The study provides a good comprehensive base for understanding the importance of data mining in the emergent and profound field of information security.

4. Information Security

The Committee on National Security Systems (CNSS) has defined information security as the protection of information and its critical elements, including the systems and hardware that use, store, and transmit that information. An important industry standard in information security is the CIA Triangle. The three dimensions that CIA defines are Confidentiality, Integrity and Availability. Parker (2002) proposed an alternative model to CIA that was called as the six elements of information security. These are confidentiality, possession, integrity, authenticity, availability and utility.

Some important characteristics of information are:

- *Confidentiality*: Confidentiality means prevention of information from disclosure by any unauthorized user. To maintain confidentiality of data various cryptographic techniques can be used. Cryptographic techniques use various encryption decryption algorithms by which the information is converted in to meaning less data so that any unauthorized user cannot understand or derive the data.

- *Integrity*: Integrity means that the information or data while in transit should not be corrupted, deleted, changed or modified and should remain intact, correct and complete when it reaches the destination. It means that the information should be free from any corruption.

- *Availability*: Availability means that the information should be in its desired format and should be made available to the authorized users only without any impediment.

- *Utility*: The information at receivers end should be useful in the sense that it should be in a meaningful format. If the information or data in not in a format which is understandable to the user than it is useless. Hence utility means the usefulness of information.

- *Authenticity*: The state of being genuine is known as authenticity. The information should be authentic that is it should not be contrived and should be original. The information received by end user should be in same state as it was when transmitted.

- *Possession*: Possession is the state of procuring information. It means the right of having information. Information is said to be in possession when it is being procured independent of any characteristics.

5. An Introduction To Data Mining

Data Mining can be defined as extracting or mining knowledge from large amount of data (Han and Kamber, 2001). It is considered as one of the steps in knowledge discovery process where valuable information is extracted from large data bases. It can also be defined as a process of extracting valid, previously unknown, non-trivial and useful information from large databases (Rao, 2003). Shaw et al (2001) have classified data mining tasks into five categories as Dependency analysis, class identification, concept description, deviation detection and data visualization summarized below.

I. Dependency Analysis: It can be defined as a process of finding associations between entities and finding relationships among them like market basket analysis.

II. Class Identification: It can be defined as grouping various entities into classes. It includes two types of identification tasks –mathematical taxonomy and concept clustering (Shaw et al., 2001).

III. Concept Description: In concept description groups are made on the basis of the domain knowledge and databases, without any compulsory descriptions. It includes tasks like data summarization and data comparison.

IV. Deviation Detection: It includes tasks like finding changes in data and anomaly detection that is finding actions that are different form the benign actions.

V. Data Visualization: It includes finding and analysing different patterns which are complex in nature. It can be used to explore the databases and can be used alone or in combination with any of the above mentioned tasks.

There are various Data mining techniques that have been proposed in literature so far. Some of them are mentioned below:

- *Classification:* It is a technique which is based on identification and formation of classes based on certain criteria and is predefined. It is used to predict future actions using various techniques like decision trees, neural networks and memory based reasoning. The two most common techniques are discussed below:

- *Neural Networks:* Neural Networks follow predictive model which are based on biological modelling capability and predicts data by a learning process.

- *Decision Trees:* It is a tree like structure where the leaf node represents or predicts the decision and the non-leaf node represents the various possible conditions that can occur. It includes algorithms like CART and CHAID.

- *Clustering:* Clustering is used to group data items into clusters which are not predefined. It is basically of two types- Hierarchical and Non Hierarchical. It includes algorithms like K-means Clustering and DBSCAN.

- *Association:* Association is aimed at finding relationships among data sets and entities that are present in the data bases. A classic example of association technique is market basket analysis and includes algorithms like Apriori and dynamic item set counting.

- *Genetic Algorithms:* Genetic Algorithms are based on the concept of evolution of genes that is carrying the features from one stage to other and follow optimization techniques for the same.

6. Role Of Data Mining In Information Security

With an increased use of computers there has been a rise in security issues. While transmitting information electronically the major issue that arises is of its security. The information should be secure while in transit. With the increased use of internet, there has been a tremendous rise in computer attacks. Internet as a medium has become a major weapon for an attacker to launch attacks in virtually no time The goal of information security is to prevent data from intrusions, frauds and malwares like viruses, worms and more and from any criminal activities. Due to the large amounts of data in databases which requires security due to its sensitive nature, there is various privacy preserving data mining techniques that have been proposed in literature. Bertino et al., (2005) has presented taxonomy of Privacy Preserving Data Mining (PPDM) techniques based on the classification given by Verykios et al., (2004).

Verykios et al., (2004) has given five dimensions which can be used for classifying various PPDM algorithms. These are data distribution, data modification, data mining algorithm, data or rule hiding and privacy preservation. Based on these dimensions, Bertino et al., (2005) has presented taxonomy of various PPDM algorithms; it includes Heuristic based techniques, reconstruction based and cryptography based techniques:

- *Heuristic-Based Techniques*: These techniques are mainly used for centralized data where the raw and aggregated data is hidden using techniques like perturbation, aggregation, blocking, swapping, sampling and generalization.

- *Reconstruction Based Techniques*: These techniques are also used for centralized data and they aim at hiding the raw sensitive data by applying techniques that are based on probability distributions.

- *Cryptography Based Techniques*: These techniques are used for distributed scenario and are aimed at using encryption techniques. Cryptographic techniques generally use secure multiparty computations.

Some other privacy preserving data mining techniques are pseudonomyzation or naïve anonymization and anonymization techniques which are defined below:

- *Pseudonomyzation*: It is a privacy preservation technique by which the sensitive information or the identifiers are replaced with false names in order to hide the sensitive information but this technique is not very secure and is susceptible to re-identification attack.

- *Anonymization*: The objective of anonymization is to hide the sensitive data in such a way that an unauthorized party cannot infer anything from the published data while the authorized party can analyse the data to get the desired results.

All these above mentioned approaches use association rule mining, classification or clustering techniques to achieve security.

7. Application of Data Mining in Information security

The role of data mining in information security defines its practical and persistent use in the field of information security. We as authors have recognised the various areas in information security where data mining has been used. The figure no.01 represents the various information security areas and the different data mining techniques that can be applied on them to achieve and maintain better security.

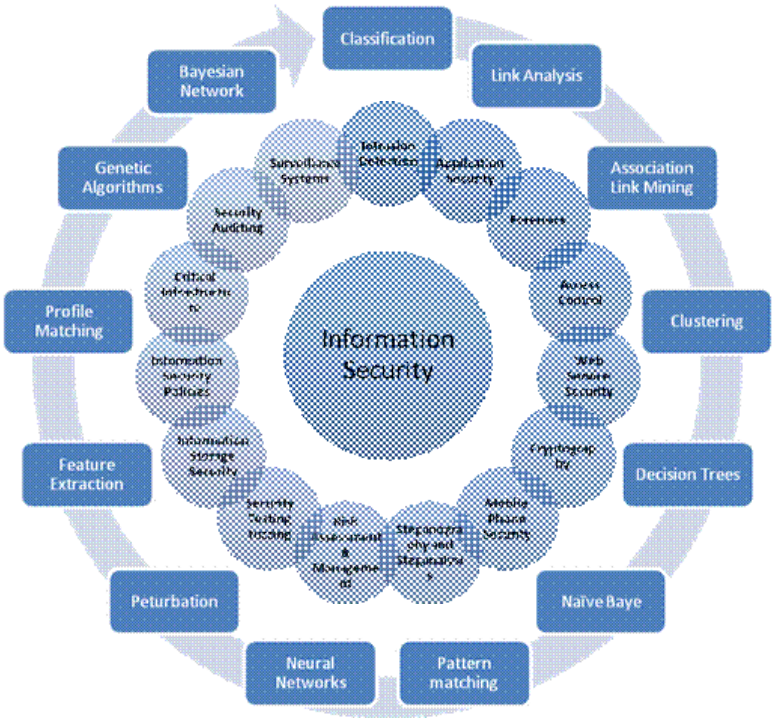

Figure 1: Data Mining Techniques And Their Application In Information Security

There are various applications of data mining in information security. Data mining techniques like classification, clustering association rule mining, link analysis, decision trees, neural networks, genetic algorithm, pattern analysis and many more can be applied to different information security areas. They can be used to:

- Identify various anomalies and malwares like viruses, worms and trojans in the system by classifying the benign and anomalous activities in to different groups and comparing each incoming data with them.
- Extract various security requirements, perform fuzzing techniques to identify vulnerabilities, define and find audit trails and establish security policies.
- Detect various cyber-crimes like credit card fraud, money laundering frauds and other financial crimes and identifying classes of criminals.
- Secure databases, authenticate and authorize users to provide access control according to their roles.
- Distinguish and identify anomalies that can harm critical infrastructure.
- Perform surveillance using wireless sensor networks and satellite tracking systems and securing them.
- Perform financial profiling and detecting operational risks.

Based on these applications Table: 1 presents the various information security areas, the data mining techniques that can be applied in those areas and there applications with advantages and disadvantages.

S. No.	Information Security Areas	Data Mining Technique(s)	Application	Advantages of applying DM	Disadvantages of applying DM
1.	• **Intrusion Detection**	• Classification: -Decision Trees	• Anomaly Detection	• For huge and multidimensional	• Association Rule mining is Slow

			-SVM -Naive Bayes • Link Analysis • Association Rule mining • Clustering	• Misuse detection	data clustering is considered better.	
2.	**Application Security**	• **Malwares**	• Decision Tree • Naïve Bayes	• To detect malware using data available without executing the analysed computer program or by processing dynamic (behavioural) data obtained during the program's execution.	• Flexible control systems • Phased penetration scheme • Modularity of updates • Presence of efficient concealment mechanisms • Using the aggressive ways of counteracting against antivirus means and also by • insufficiency of methods signature- and rule-based detection methods	• Low decision making speed • High level of resources' usage • Existence of some specific loopholes
		• **Operating System Security**	• Clustering • Decision trees • Classification	• To isolate the cause of a system crash/failure. • Any change in program resource consumption can be used to identify an unwanted activity.	• Existing solutions can be harnessed to create type-cause relationship. • Help maintain the health of the system by detecting and terminating unwanted programs.	• Mining data in extensive crash logs can be time consuming • A false alarm may result in termination of a genuine activity
		• **Secure Software Engineering**	• Classification • Clustering • Pattern Matching • Pattern Mining	• Data mining can be used in gathering and extracting security requirements • Extracting algorithms and business rules from code • Mining applications for requirements and business rules for new projects. • Finding security vulnerabilities and bugs.	• Improves software productivity • Software quality • Reliability • Maintenance	• Generally offline mining is done where data is already collected and stored. • Due to offline mining the mining process becomes a bit time consuming. • For efficient working stream mining should be adopted
3.	**Forensics**	• **Fraud Detection**	• Link Analysis	• Link analysis supports the	• Discovers frequently	• Association rule mining requires

			• Association rule mining • Neural Networks	detection and analysis of money laundering and other financial crimes • Association rule mining can be used to detect potential future network attacks	occurring item sets in a database and presents the patterns as rules • Neural Network Results in the creation of an automatic learning algorithm	rich and highly structured data • Neural Networks have a high runtime complexity.
		• **Crime Investigation**	• Clustering • Classification	• Clustering can help in identification of classes of criminals • Maximizing the diagnosis of Credit Card frauds by minimizing both the number of false alarms and the number of fraud transactions not recognized.	• Maximize or minimize intra-class similarity • Classification can reduce the time required to identify crime entities	• High computational intensity typically required • Classification requires reasonably complete training and testing for accurate results
4.	**Access Control**	• **Authentication (using Biometrics)**	• Artificial Neural Network(ANN) • Linear Discriminant Analysis(LDA) and Principal Component Analysis(PCA) along with LDA	• Pattern Recognition • Attack Detection	• ANN is a powerful technique which can predict not only for the seen data, but also for the unseen data.	• LDA, (PCA+LDA) is a powerful technique for predicting seen data, but cannot predict unseen data. • ANN predictive power may decrease when dealing with too many classes
		• **Authorization (using role based model)**	• Clustering • Pattern Recognition	• To identify the right roles by grouping users with similar set of permissions to form one role	• Effectively use existing user permissions to create RBAC • Help create the right number of roles and hierarchy required in the enterprise security model	• Supervised learning can be difficult to implement due to non-availability of training data.
5.	**Web Service Security**	• **Virtual Communities**	• Perturbation • Clustering • Permutation	• Hiding Sensitive Data. Anonymization of data • Analysing Social Network Data	• Provides desired level of privacy to multidimensional and dynamic evolution of data.	• Repeated anonymization of data can lead to • More information loss. • Inaccuracy

						• Decrease in utility.
		• **E-commerce and e-business**	• Classification • Clustering • Association Rule Mining	• E- payment systems • Customer profiling • recommender systems • securing customers and transaction data	• Maintains a certain level of abstraction • maintaining logs and auditing of data • provide desired security	• Owing to non-linear scalability and due to large amounts of data, most data mining algorithms are time consuming • Generated models are complex and not easily comprehendible
6.	• **Cryptography**		• Classification • Clustering • Association Rule Mining • Profile Matching	• Algorithms like ID3 can be used to implement secure multi party protocols	• Yao's two-party protocol is pretty efficient, as long as the sizes of the inputs, and the size of the circuit computing the function, are reasonable.	• Yao's two party protocol is costly. • Algorithms Like ID3 poses some problems like • Increase in computational overhead due to large database sizes. • Due to large circuit representations the computations involves evaluating polynomials and therefore requires computing multiplications and exponentiations. • x round of id3 depends on the x-1 previous rounds. So the implementation may require repetitive encoding of one step corresponding to specific results of previous round.
7.	• **Mobile Phone Security**		• Classification	• Smart phone authentication using rule based classifiers. • Cell phone based biometric identification and authentication using	• Efficient and require low computing power • User friendly and require no additional authentication hardware • Provide accurate identification	• It is not so efficient in the way that it needs a certain period of time to collect samples before triggering the authentication phase. • The incompleteness of data

			accelerometer data.	and authentication in realistic settings	collected due to poor signal receptions can prevent rules from getting stable.
8.	• **Steganography and Steganalysis**	• Classification Naïve Bayes SVM • Association Rule Mining • Feature Extraction • Pattern Recognition	• Implementation of Steganography based secure multi party protocols • Used for privacy preservation of distributed and heterogeneous data • In Steganalysis techniques like feature extraction and pattern recognition are used to find out the hidden data.	• Faster performance in case of very large database. • Optimized Computation. • Acceptable level of privacy	• Reduction in privacy level
9.	• **Risk Assessment and Management**	• Classification (decision trees- CHAID)	For financial profiling and detecting operational risks.	• Help prevent financial crisis by detecting it at an early stage • Help manage resources effectively for better financial performance • Help monitor the risk factor and determine strategies to mitigate them	• In order to produce accurate results it requires a substantial training set
10.	• **Security Testing fuzzing**	• Genetic Algorithm (Genetic Algorithm Mutation Operator- GA Mutator)	• To Identify Software vulnerabilities using Fuzzing Technique. • Multidimensional fuzzing uses genetic algorithms to identify large number of vulnerabilities in the software	Directed Extended Fuzzing which is a multidimensional vulnerability fuzzing is • Easy • Effective • Efficient and • Find more vulnerabilities than single dimensional fuzzing	• DX Fuzzing is hard to apply in practically large applications • It cannot solve program's strong program checks.

11.	• **Information Storage Security**	• Classification • Bayesian Network	• Securing Databases via Semantic Inference model. Bayesian network is used for evaluating inference probability. If the value of sensitive attributes can be inferred with a probability greater than their pre-specified thresholds then the access will be denied.	• Use of Bayesian networks reduces the computation complexity. • It can handle inferences effectively during query processing. • It can detect collaborative inference attacks.	• Due to large amounts of data in database, scalability issues may arise.
12.	• **Information Security Policies**	• Classification • Decision Trees • Hierarchal Change Mining	• Data mining techniques can be used to deal with evolving nature of security policies.	• Data Mining provides a building block to deal with multiple levels of evolving security policies. • Data mining helps in reduction of errors while updating security policies. • It detects changes and pattern of changes across multiple layers of security policies.	• Management of security policies is complex due to insufficient changes and multi-layered nature.
13.	• **Allied/ Critical Infrastructure. - SCADA**	• Classification • Clustering • Association Rule Mining • Genetic algorithms	• Data mining techniques can be used to trace back or distinguish anomalies that can harm a critical infrastructure. • It can be used to design a class or group at each intermediate point of critical infrastructure to identify security levels. • Vulnerability assessment	• Provides countermeasures against interception of sensitive data. • Effective traceability	• The data set should be organised and ready for mining. Pre-processing may be required to get it ready for mining
14	• **Security Auditing**	• Clustering • Association Rule Mining	• To find breaches in business policy • To define Audit trails	• Provides the ease of automatically manipulation complex data	• Incomplete or limited data may give inconclusive results

			• Decision trees • Classification and Prediction	• Finding the right audit approach based on client data • To train new auditors to test new systems	• Reduce the number of professional staff requirement	• Cannot replace professional staff totally
15.	**Surveillance Systems**	• **Wireless Sensor Networks**	• Classification Naïve Bayes • Clustering	• Data mining techniques can be used to cluster the sensor nodes into different groups. • It can be used to detect abnormal events with in clusters that are present in a network and provide security to the data. • The anomalous events can be detected by various classification techniques.	• It provides high detection rate. • There are less false positives. • High Accuracy • Large network can be divided into clusters for easy operability	• A substantial training set is required to get accurate results
		• **Satellite Tracking Systems**	• Classification • Clustering • Association rule mining	• Data mining techniques can be used for securing the data processing segment of the satellite tracking systems. • It can be used to identify and classify anomalies when the end user communicates with the data centre. • Security at both ends that is at the data processing segment and at end user segment can be achieved.	• Privacy of the data can be preserved without compromising the utility • Outlier detection can be easily achieved	• Substantial amount of data should be present before any conclusive results can be obtained

Table 1: Table Presenting Applications Of Data Mining In Information Security

8. Implications Of Research

The study carried by us helped to identify and found new and emerging areas where data mining can help to provide the security of information which is the need of the current era. The findings of the authors will be helpful for new researchers who would like to do deeper research in the various applications of data mining in information security area. The widespread application of data mining in various fields are justified by the research of the authors which

paved way for more intensive usage of data mining in information security area. The researchers who are already working in this area will come to know about the latest trends and direction of research for application of data mining in information security area. The in-depth study of such literature also helps to gain a better understanding as what should be next step towards improving the effectiveness of privacy preserving/anonymization of data.

9. Limitations Of The Study

The key efficacy of the study is to progress in the area of information security by applying various data mining techniques to achieve security. A number of important limitations that need to be addressed are:

- The study is limited to the data mining techniques that can be applied in information security area and are not tested. Some of them are just proof of concepts.
- The current study was unable to include any real analysis.
- There could be more data mining techniques that are possibly applied than those discussed.
- The research provides an insight in to this area and can be further evaluated.

10. Conclusions And Future Research

The authors tried to present the application of the Data mining techniques in information security area through the survey of the literature. Organization should try to incorporate the latest DM techniques to remove the shortening in their privacy preservation methodology by adopting data mining in their setup. We have seen through this paper that successful implementation of Data Mining in the organization can improve the performance of the companies which is the need of the present business environment for privacy preservation. The final outcome will be improving the security of the data of the organization which play a vital role in company's growth.

11. Acknowledgement

We are thankful to Ambedkar Institute of Advanced Communication Technologies and Research, Delhi (India) for providing the necessary help in carrying out the research.

12. References

Bertino, E., Fovino, I. N. and Provenza, L. P. (2005), 'A Framework for Evaluating Privacy Preserving Data Mining Algorithms', Data Mining and Knowledge Discovery, vol. 11(2), pp. 121-154.

Chen, H., Chung, W., Xu, J.J., Wang, G., Qin, Y. and Chau, M. (2004), 'Crime Data Mining: A General Framework and Some Examples', Computer: Security and Privacy in an Online World, vol. 37 (4), pp.50 – 56.

Chen, Y. and Chu, W. (2006), 'Database Security Protection via Inference Detection', IEEE International Conference on Intelligence and Security Informatics.

Dartigue, C., Jang, H.I. and Zeng, W. (2009), 'A New Data-Mining Based Approach for Network Intrusion Detection' , Seventh Annual Communication Networks and Services Research Conference, pp.372 – 377

Denker, B. (2002), 'Data Mining and the Auditor's Responsibility', for Information Systems Audit and Control Association, InfoBytes.

Guo, T. and Li, G. (2008), 'Neural data mining for credit card fraud detection', 2008 International Conference on Machine Learning and Cybernetics, pp.3630-3634.

Hooper, E. (2010), 'Intelligent Strategies for Smart Grid and Cyber Security', International Journal of Intelligent Computing Research (IJICR), vol. 1 (3), pp.124–129.

Hussein, M., El Sisi, A. and Ismail, N (2008), 'Performance Tuning of Steganography Algorithm for Privacy Preserving Association Rule Mining in Heterogeneous Data Base', New Technologies, Mobility and Security, NTMS '08, pp.1-6.

Kämppi, P., Rajamäki, J. and Guinness, R. (2009), 'Information security risks for satellite tracking', International Journal of Computers and Communications, vol. 3 (1), pp.9-16.

Komashinskiy, D. and Kotenko, I. (2010), 'Malware Detection by Data Mining Techniques Based on Positionally Dependent Features', 18th Euromicro International Conference on Parallel, Distributed and Network-Based

Processing (PDP), pp.617 – 623.

Koyuncugil, A.S. and Ozgulbas, N. (2009), 'Financial Profiling for Detecting Operational Risk by Data Mining', Academic and Business Research Institute Conference, Orlando 2009, available at http://www.aabri.com/OC09manuscripts/OC09117.pdf (accessed on 4th December, 2011).

Kwapisz, J.R., Weiss, G.M. and Moore, S.A. (2010), 'Cell phone-based biometric identification', Fourth IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS), pp.1-7.

Han, J., Kamber ,M. (2001) 'Data mining: Concepts and Techniques',The Morgan Kaufmann Series in Data Management Systems.

Lee,W. ,Stolfo, S.J. , Chan, P.K., Eskin, E., Fan, W. , Miller, M., Hershkop, S. and Zhang, J. (2001), 'Real time data mining-based intrusion detection', Proceedings on DARPA Information Survivability Conference & Exposition II, vol. 1, pp.89 – 100.

Lee, W. and Stolfo, S.J. (1998), 'Data Mining Approaches for Intrusion Detection', In Proceedings of the 7th USENIX Security Symposium.

Li, N. , Zhang, N. , Das, S.K. and Thuraisingham, B. (2009), 'Privacy preservation in wireless sensor networks: A state-of-the-art survey', Ad Hoc Networks 7 (ELSEVIER), pp.1501-1514.

Liu, Q., Sung, A.H., Ribeiro, B., Wei, M., Chen, Z. and Xu, X. (2008), 'Image complexity and feature mining for steganalysis of least significant bit matching steganography', Information Sciences: An International Journal, vol. 178 (1), pp.21-36.

Nguyen, H.A. , Choi, D. (2008), 'Application of Data Mining to Network Intrusion Detection: Classifier Selection Model', APNOMS '08 Proceedings of the 11th Asia-Pacific Symposium on Network Operations and Management: Challenges for Next Generation Network Operations and Service Management , pp.399 – 408.

Parker, B.D (2002), 'Motivating the Workforce to Support Security Objectives: A Long-Term View' available at http://www.iwar.org.uk/comsec/resources/sa-tools/Motivation-for-Information- Security.pdf (accessed on 1st December, 2011).

Pattanaik, S. and Ghosh, P.P. (2010), 'Role of Data Mining in E-Payment systems', (IJCSIS) International Journal of Computer Science and Information Security, vol. 7 (2), pp.262-266.

Pejas, M. (2005), 'Aggregated Hypothesis Testing for Steganalysis', Transactions on Enformatika, International Academy of Sciences, Systems Sciences and Engineering, pp.265-269.

Pinkas, B. (2002), 'Cryptographic techniques for privacy-preserving data mining', ACM SIGKDD Explorations Newsletter, vol. 4 (2), pp.12-19.

Prasad, A.V.K. and Ramakrishna, S. (2010a), 'Software Architectures Design Patterns Mining for Security Engineering', International Journal of Computer Science and Information Technologies, vol. 1 (5), pp.408-413.

Prasad, A.V.K. and Ramakrishna, S. (2010b), 'Data Mining for Secure Software Engineering – Source Code Management Tool Case Study', International Journal of Engineering Science and Technology, vol. 2 (7), pp.2667-2677.

Rao, I. K. R. (2003), "Data Mining and Clustering Techniques", DRTC Workshop on Semantic Web, December.

Sa, M., Nayak, M.R. and Rath,A.K. (2011), 'A Simple Agent Based Model for Detecting Abnormal Event Patterns in a Distributed Wireless Sensor Networks', ICCCS '11 Proceedings of the 2011 International Conference on Communication, Computing & Security, pp.67-70.

Sharma, A. and Ojha, V. (2010), 'Implementation of Cryptography for Privacy Preserving Data Mining', International Journal of Database Management Systems (IJDMS), August 2010, vol. 2 (3), pp.57-65.

Shaw, M.J., Subramaniam, C. , Tan, M.J. , Welge, M.E. (2001), "Knowledge management and data mining for marketing", Decision Support Systems , Elsevier Science Publishers B. V. Amsterdam, The Netherlands, pp.127-137

Sirikulvadhana, S. (2002), 'Data Mining As A Financial Auditing Tool', M.Sc. Thesis in Accounting, Swedish School of Economics and Business Administration available at http://www.pafis.shh.fi/graduates/supsir01.pdf (accessed on 1st December, 2011).

Tang,Y. , Hidenori, N. and Urano, Y. (2010), 'User authentication on smart phones using a data mining method', International Conference on Information Society (i-Society), pp.173-178.

Thongtae, P. and Srisuk, S. (2008), 'An Analysis of Data Mining Applications in Crime Domain', IEEE 8th International Conference on Computer and Information Technology Workshops, CIT Workshops 2008, pp.122 – 126.

Thuraisingham, B. and Khan, L. (2005), "Data mining applications in biometrics", Technical report for Department of Computer Science at The University of Texas, Dallas.

Vaidya, J. and Clifton, C (2002), "Privacy preserving association rule mining in vertically partitioned data", In 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM Press, pp.639-644.

Verykios, V.S., Bertino, E., Nai Fovino, I., Parasiliti, L., Saygin, Y., and Theodoridis, Y. (2004), "State-of-the-art in privacy preserving data mining", SIGMOD Record, 33(1), pp.50-57.

Weaver, G.A. , Foti, N. , Bratus, S. , Rockmore, D. and Smith, S.W. (2011), 'Using hierarchal change mining to manage network security policy evolution', Hot-ICE'11 Proceedings of the 11th USENIX conference on Hot topics in management of internet, cloud, and enterprise networks and services, pp.8-8.

Weiss, G.M. and Lockhart, J. W. (2011), 'Identifying User Traits by Mining Smart Phone Accelerometer Data', Proceedings of the Fifth International Workshop on Knowledge Discovery from Sensor Data, pp.61-69.

Wu, Z. , Atwood, J.W. and Zhu, X. (2009), 'A New Fuzzing Technique for Software Vulnerability Mining', Proceedings of CONSEG-09: International Conference on Software Engineering, pp.59-66.

Zhou, B., Pei, J. and Luk, W.S. (2008), "A brief survey on anonymization techniques for privacy preserving publishing of social network data", SIGKDD Explorations, NY, USA, vol. 10 (2), pp.12-22.

About the Authors:

Dr. Vishal Bhatnagar received the B-Tech degree in Computer-Science and Engineering from Nagpur University in Nagpur, India in 1999 and the M-Tech in Information-Technology from Punjabi University, Patiala, India in 2005 and PhD from Shobhit University in 2010.Vishal Bhatnagar is Associate Professor in Computer-Science and Engineering department at Ambedkar Institute of Advanced Communication Technologies and Research (Govt. of Delhi), GGSIPU, Delhi, India. His research interests include Database, Advance Database, Data warehouse and Data-mining. He has been in teaching for more than eight years. He has guided under-graduate and post-graduate students in various research projects of databases and data mining. He can be reached by email at vishalbhatnagar@yahoo.com.

Sanur Sharma received her B-Tech in Computer Science and Engineering from GGSIPU, Delhi, India in 2010 and is currently pursuing M-Tech in Information Security from GGSIPU, Delhi, India. Her research interests include database, data warehouse, data mining, and social network analysis. She can be reached by email at sanursharma@yahoo.co.in

---