ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 9, September 2017

A Survey on Datamining Based Intrusion Detection System

B.Gayathri, Dr. Antony Selvadoss Thanamani

Research Scholar, Department of Computer Science, NGM College, Pollachi, India

Head of Department, Associate Professor, Department of Computer Science, NGM College, Pollachi, India

ABSTRACT: Database mining can be characterized as the way toward mining for understood, once unidentified, and conceivably fundamental data from outrageously tremendous databases by proficient learning revelation procedures. The privacy and security of client data have turned out to be noteworthy open arrangement tensions and these nerves are accepting expanded enthusiasm by the both open and government legislator and controller, security advocates, and the media. In this survey we discusses about architecture of data mining, preparation in data mining and intrusion detection in data mining. Latest advancements in data innovation have empowered gathering and preparing of huge measure of individual information insurance laws and view point for protection and security enactment.

KEYWORDS: Database mining, Database security, Data Privacy, Inferences, Intrusion Detection, Law.

I. INTRODUCTION

Security and Privacy protection in data mining and data engineering have been a public policy concern for decades. However, rapid technological changes, the rapid growth of the internet and electronic commerce, and the development of more sophisticated methods of collecting, analyzing, and using personal information have made privacy a major public and government issues. The field of data mining is gaining significance recognition to the availability of large amounts of data, easily collected and stored via computer systems. Recently, the large amount of data, gathered from various channels, contains much personal information. When personal and sensitive data are published and/or analyzed, one important question to take into account is whether the analysis violates the privacy of individuals whose data is referred to. The importance of information that can be used to increase revenue cuts costs or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data privacy is growing constantly. For this reason, many research works have focused on privacy-preserving data mining, proposing novel techniques that allow extracting knowledge while trying to protect the privacy of users. Some of these approaches aim at individual privacy while others aim at corporate privacy. Data mining, popularly known as Knowledge Discovery in Databases (KDD), it is the nontrivial extraction of implicit, previously unknown and potentially useful information from data in databases. Knowledge discovery is needed to make sense and use of data. Though, data databases (or KDD) are frequently treated as synonyms, data mining is actually part of the knowledge discovery process. [1,2,3]. Usually, data mining e.g. data or knowledge discovery is the process of analyzing data from different perspectives and summarizing it into useful information from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases. [4] Although data mining is a comparatively new term but the technology is not. Companies have used powerful computers to filter through volumes of superstore scanner data and analyze market research reports for many years. However, continuous innovations in computer processing power, disk storage, and statistical software are dramatically increasing the accuracy of analysis while driving down the cost. [5] Data mining, the discovery of new and interesting patterns in large datasets, is an exploding field. One aspect is the use of data mining to improve security, e.g., for intrusion detection. A second aspect is the potential security hazards posed when an adversary has data mining capabilities. Privacy issues have attracted the attention of the media, politicians, government agencies, businesses, and privacy advocates.

II. ARCHITECTURE OF DATA MINING

Data mining is described as a process of discover or extracting interesting knowledge from large amounts of data stored in multiple data sources such as file systems, databases, data warehouse and etc. This knowledge contributes a lot of benefits to business strategies, scientific, medical research, governments and individual. The architecture contains modules for secure safe-thread communication, database connectivity, organized data management and efficient data analysis for generating global mining model.

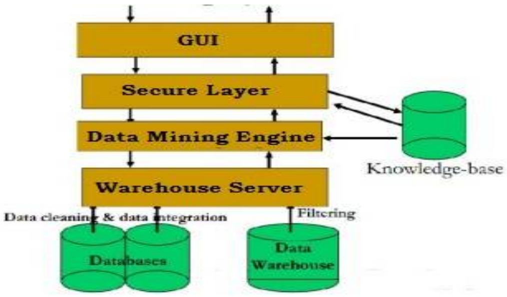

Fig: Architecture of Data Mining

III. PREPERATION IN DATA MINING

As data mining initiatives continue to evolve, there are several issues Congress may decide to consider related to implementation and oversight. These issues include, but are not limited to, data quality, interoperability, mission creep, and privacy, [7] as with other aspects of data mining, while technological capabilities are important, other factors also influence the success of a project's outcome. We generate an enormous amount of data as a by-product of our everyday transactions (purchasing goods, enrolling for courses, etc.), visits to Web sites and interactions with government (taxes, census, car registration, voter registration, etc.). Not only is the number of records we generate increasing, but the amount of data gathered for each type of record is increasing.

A. Data Quality

Data quality is a multifaceted issue that represents one of the biggest challenges for data mining. Data quality refers to the accuracy and completeness of the data. Data quality can also be affected by the structure and consistency of the data being analyzed.The presence of duplicate records, the lack of data standards, the timeliness of updates, and human error can significantly impact the effectiveness of the more complex data mining techniques, which are sensitive to subtle differences that may exist in the data.To improve data quality, it is sometimes necessary to "clean" the data, which can involve the removal of duplicate records, normalizing the values used to represent information in the database.

B. Data Application Areas Mining

There are many areas of data mining application in most popular are Science (astronomy, bioinformatics, drug discovery), Business (advertising, customer relationship management, investment, manufacturing, entertainment, telecom, e-commerce, banking, marketing, health), web (serach engines, bots), government (law enforcement, proofing tax chater, anti-terror).

C. Interoperability

Related to data quality, is the issue of interoperability of different databases and data mining software. Interoperability refers to the ability of a computer system and/or data to work with other systems or data using common standards or processes. Interoperability is a critical part of the larger efforts to improve interagency collaboration and information sharing through e-government and homeland security initiatives. For data mining, interoperability of databases and software is important to enable the search and analysis of multiple databases simultaneously, and to help ensure the compatibility of data mining activities of different agencies. Data mining projects that are trying to take advantage of existing legacy databases or that are initiating first-time collaborative efforts with other agencies or levels of government may experience interoperability problems. Similarly, as agencies move forward with the creation of new databases and information sharing efforts, they will need to address interoperability issues during their planning stages to better ensure the effectiveness of their data mining projects.

D. Privacy

As additional information sharing and data mining initiatives have been announced, increased attention has focused on the implications for privacy. Concerns about privacy focus both on actual projects proposed, as well as concerns about the potential for data mining applications to be expanded beyond their original purposes. For example, some experts suggest that anti-terrorism data mining applications might also be useful for combating other types of crime as well.[6] So far there has been little consensus about how data mining should be carried out, with several competing points of view being debated. Some observers contend that tradeoffs may need to be made regarding privacy to ensure security. Others observers suggest that existing laws and regulations regarding privacy protections are adequate, and that these initiatives do not pose any threats to privacy. Still other observers argue that not enough is known about how data mining projects will be carried out, and that greater oversight is needed. There is also some disagreement over how privacy concerns should be addressed. Some observers suggest that technical solutions are adequate initiatives. Data mining has attracted significant interest especially in the past decade with its vast domain of applications. From the security perspective, data mining has been shown to be beneficial in confronting various types of attacks to computer systems. However, the same technology can be used to create potential security hazards. In addition to that, data collection and analysis efforts by government agencies and businesses raised fears about privacy, which motivated the privacy preserving data mining research. One aspect of privacy preserving data mining is that, we should be able to apply data mining algorithms without observing the confidential data values. [6,7] This challenging task is still being investigated. Another aspect is that, using data mining technology an adversary could access confidential information that could not be reached through querying tools jeopardizing the privacy of individuals. Some initial research results in privacy preserving data mining have been published. However, there are still many issues that need further investigation in the context of data mining from both privacy and security perspectives. This workshop aims to provide a meeting place for academicians to identify problems related to all aspects of privacy and security issues in data mining together with possible solutions. Researchers and practitioners working in data mining, databases, data security, and statistics are invited to submit their experience, and/or research results.

IV. INTRUSION DETECTION IN DATA MINING

Intrusion Detection System (IDS) is meant to be a software application which monitors the network or system activities and finds if any malicious operations occur. Tremendous growth and usage of internet raises concerns about how to protect and communicate the digital information in a safe manner. Nowadays, hackers use different types of attacks for getting the valuable information. Many intrusion detection techniques, methods and algorithms help to detect these attacks. This main objective of this paper is to provide a complete study about the definition of intrusion detection, history, life cycle, types of intrusion detection methods, types of attacks, different tools and techniques, research needs, challenges and applications.

V. DATA BASED IDS

Data based IDS is one common type of IDS that analyzes network traffic at all layers of the Open Systems Interconnection (OSI) model and makes decisions about the purpose of the traffic, analyzing for suspicious activity. Most NIDSs are easy to deploy on a network and can often view traffic from many systems at once. A term becoming more widely used by vendors is “Wireless Intrusion Prevention System” (WIPS) to describe a network device that monitors and analyzes the wireless radio spectrum in a network for intrusions and performs countermeasures which

DOI: 10.15680/IJIRCCE.2017. 0509017

monitors network traffic for particular network segments or devices and analyzes the network and application protocol activity to identify suspicious activity. It can ident different types of events of interest. It is most commonly deployed.

The NIDS are also called passive IDS since this kind of systems inform the administrator system that an attack has or had taken place, and it takes the adequate measures to assure the security of the system. The aim is to inform about an intrusion in order to look for the IDS capable to react in the post. Report of the damages is not sufficient. It is necessary that the IDS react and to be able to block the detected doubtful traffics. These reaction techniques imply the active IDS. The comparison between firewall and IDS is follows

VI. COMPARSION: FIREWALLVS IDS

Criteria	**Firewall**	**IDS**
Data Security	Moderate	High
Data Integrity	Moderate	Very High
Acknowledgement	Medium and Need more improvements	Exact data acknowledgement
Warning	When system on	Any time. Even in offline

Table1: Comparison between Firewall and IDS

VII. GLIMPSE OF LITERATUTRE

SNO	TITLE	AUTHOR	ISBN&YEAR	TECHNIQUES	ADVANTAGE	DISADVANTAGE
1	Introduction to datamining&knowledgediscovery	Two crows	Year-1999		Data mining tools need to be guided by users who understand the business, the data, and the general nature of the analytical methods involved.	-
2	Data mining: Introductory and Advanced topics	Dunham M.H Sridhar s	ISBN:978-81-931039-1=3.Year: june2016	Privacy preserving	Security assurances	

DOI: 10.15680/IJIRCCE.2017. 0509017

3	Discovering knowledge in data: An Introduction to Data mining	Larose,D.T	ISBN:0-471-66657-2 Year:2006	complex algorithm(Kohonen networks)	Privacy	
4	link mining a survey	L.Getoor,c.p Diehi	Vol-7,pp3-12 Year-2005	Data mining methods with example		Un certainty
5	From data mining to KDD	Fayyad u.m Piatetsky-Shapiro.G Smyth p	Year:1996	Data mining step of the KDD process (methods &algorithms)	Efficient data analysis	
6	Defining privacy for data mining	Clifton C M.kanmura J.vaidya	Year:2002	Individual privacy Control privacy Corporate privacy	Protection and security enactment	
7	Challenges with legacy data: knowing your data enemy is the first step in overcoming it.	Scott W. Ambler	Year:july-2001		Client on security	Different levels of detail. Different modes of operation. Varying timeliness of data

VIII. CONCLUSION

In this article, survey has been done for detecting fraud, assessing risk and product retailing data mining involves the use of data analysis tools to discover previously unknown, valid patterns and relationship in large data sets. This is also study of data mining privacy preserving techniques gives security assurance and individual privacy control gives the protection and security enactment.

REFERENCES

1 Introduction to Data Mining and Knowledge Discovery, Third Edition ISBN: 1-892095-02- 5, Two Crows Corporation, 10500 Falls Road, Potomac, MD 20854 (U.S.A.), 1999.

2 Dunham, M. H., Sridhar S., "Data Mining: Introductory and Advanced Topics", Pearson Education, New Delhi, ISBN: 81-7758-785-4, 1st Edition, 2006

3 Larose, D. T., "Discovering Knowledge in Data: An Introduction to Data Mining", ISBN 0-471-66657-2, ohn Wiley & Sons, Inc, 2005.

4 L. Getoor, C. P. Diehl. "Link mining: a survey", ACM SIGKDD Explorations, vol. 7, pp. 3-12, 2005.

5 Fayyad U.M., Piatetsky-Shapiro G., Smyth P. "From Data Mining to KDD: An Overview", AAAI/MIT Press, 1996.

6 Clifton, C., M. Kantarcioglu and J. Vaidya, "Defining Privacy for Data Mining," Purdue University, 2002 (see also Next Generation Data Mining Workshop, Baltimore, MD, November 2002.

7 Scott W. Ambler, "Challenges with legacy data: Knowing your data enemy is the first step in overcoming it", Practice Leader, Agile. Development, Rational Methods Group, IBM, 01 Jul 2001.