International Journal of Innovative Research in Information Security (IJIRIS) ISSN: 2349-7017(O)
Issue 1, Volume 2 (January 2015) ISSN: 2349-7009(P)
www.ijiris.com

Study on Exploring Data Mining Techniques for Network Intrusion Detection

Anthony Raj.A
Department of computer Science,
Sri Bhagawan Mahaveer Jain College,KGF
Research Scholar / CS
PRIST University, Thanjavur, INDIA

Dr. A. Arul Lawrence Selvakumar
Professor & Head, Dept of CSE,
Rajiv Gandhi Institute of Technology,
Bangalore, Karnataka
INDIA

ABSTRACT -- Network intrusion detection is very important mechanism for detecting intrusions in networks. Data mining techniques play very important role in detecting intrusions in networks. Intrusions cause damage to the data and compromise integrity and confidentiality and availability of the data. Though many intrusions preventing software's are developed and installed in network stations and network operating systems still finds vulnerabilities of the systems using network hacking techniques and tries to break the security walls of the system and enters despite the intrusion preventive mechanism built in. Hence network administrators and data management team feels the intrusion detection techniques are essential in-order to know that intrusion occurred or not and track them in and out entries in the network, so that steps can be taken for further for preventing intrusions or to block the intrusions that keep coming in. though In Corporate world many intrusion detection software's and detections techniques of different solutions are being developed still the people are not satisfied with the performance of the IDS. So many researches are still carrying on in this area to seek efficient techniques which are used for Intrusion detection. In this research papers gives the idea of what is network intrusion detection system and what it supposed to do and what are the problems with this technology and finally focus on Data Mining Techniques & data mining process to build more effective intrusion detection systems.

Keywords — Network Intrusion Detection System, link analysis, sequence analysis, association rule, data set, classifier

1. INTRODUCTION

In the field of networking the area of security comprise policies adapted by the network administrator to prevent and monitor unauthorized access. Network security involves authorization of network access which is controlled by the network administrator. Internet is network of networks and not network of host. In information security the Network Intrusions are the activities that violate the security norms of the network system. Network Intrusion Detection system is Mechanism used to identify, monitor intrusions that travel through on a network wire and analyze the traffic packets on a network for intrusion detection. The main goal of the network IDS is to identify the attacks and security threats as and when happen by providing the real time network monitoring and second provide attack information to the network administrator and third fix the attacks by taking preventing measures and finally store attack events in the disk for analysis to identify which is normal and intrusion. [1]

We need intrusion detection techniques looking at a perspective of building a secure network. Secure a network first analyzing the vulnerability of the network environment. In second line of defense is intrusion prevention system. If the prevention fails the intrusion detection system comes in to the picture. If Intrusion detection system do fails (for example denial of service) then we want to rely on intrusion respond/tolerance systems. In this paper mainly focused on Intrusion detection system and their techniques. Intrusion detection to work First assumption is that system activities can be or are observable, second assumption is that normal and intrusive activities must have distinct evidence. The goal of intrusion detection is to analyze audit data and find out the evidence of Intrusion. Main techniques used are misuse detection & Anomaly detection. Misuse detection which is based on patterns finding of well known attack and anomaly detection which based on deviation from normal pattern usage of system. This paper gives new ideas and insights on the intrusion detection development process using data mining and focuses mainly on intrusion detection techniques. [16]

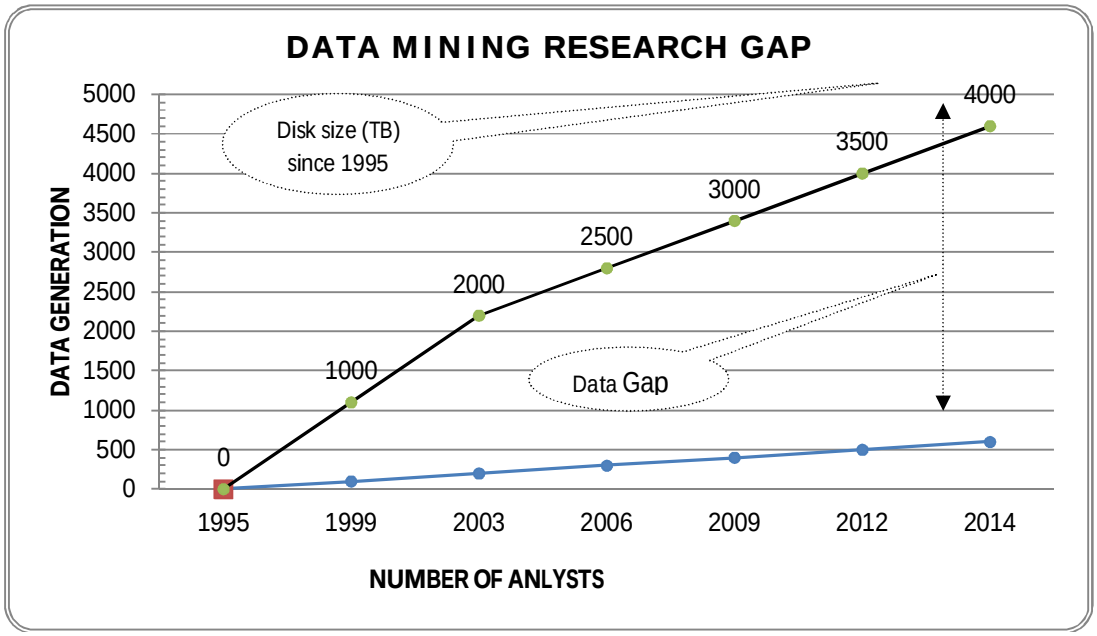
2. DRAWBACKS OF CURRENT INTRUSION DETECTION TECHNIQUES

The reasons for drawbacks of current state of intrusion detection techniques mainly due to Poor theoretical foundations and development methodology, development process to develop intrusion detection systems. So most of the IDS pure based knowledge on software engineering techniques involves studying particular network configuration, operating system environment and application software and possible attack methods that can be launched, so based on that knowledge IDS developed and hoping that will work. But the networking environment really too complicated. So just going through software engineering process is very slow and very expensive process. [2]

© 2015, IJIRIS- All Rights Reserved

3. GROWTH OF DATA MINING RESEARCH

In order understand what data mining is, it would helpful to get a feel of the amount of data been generated now a days As per the quote from the Eric Schmid who was CEO of Google over a decade at one of the conferences he said: "Every two days now, we create as much information as we did from the dawn of civilization up until 2003." i.e. if we combine the amount of recorded information from the dawn of civilization up until 2003, we are able to create that much information in just two days now. This analogy conveys a better feel that how much data there is being generated today. So there is massive amount of data created daily basis even today i.e. we are extremely in data- rich situation and posed too many information security problems. But the problem is that much of the data that has been generated never analyzed at all. The plot shown below represents that the curve represents how much data had been generated across time starting from 1995 to till 2014. The data that was there represented by the curve increasing exponential way. This trend continues as we could imagine even today. But if we look at number of data analysts that are available to analyze the data gone up only insignificant amount. If we look at below curve pretty much flat curve which represents gap between number of data that's been generated and the number of people to analyze the data i.e. there is huge research gap between the generation of data and data mining and research work. [3]



To bridge this gap requires the solution of fundamentally new research problems, which can be grouped into the following broad challenges: (a) developing algorithms and systems to mine large, massive and high dimensional data sets, (b) developing algorithms and systems to mine new types of data, (c) developing algorithms, protocols and other infrastructure to mine distributed data,(d) improving the ease of use of data mining systems, and (e) developing appropriate privacy and security models for data mining . In order to respond to these challenges, we require applied, multidisciplinary and interdisciplinary research in data mining and knowledge discovery [5]

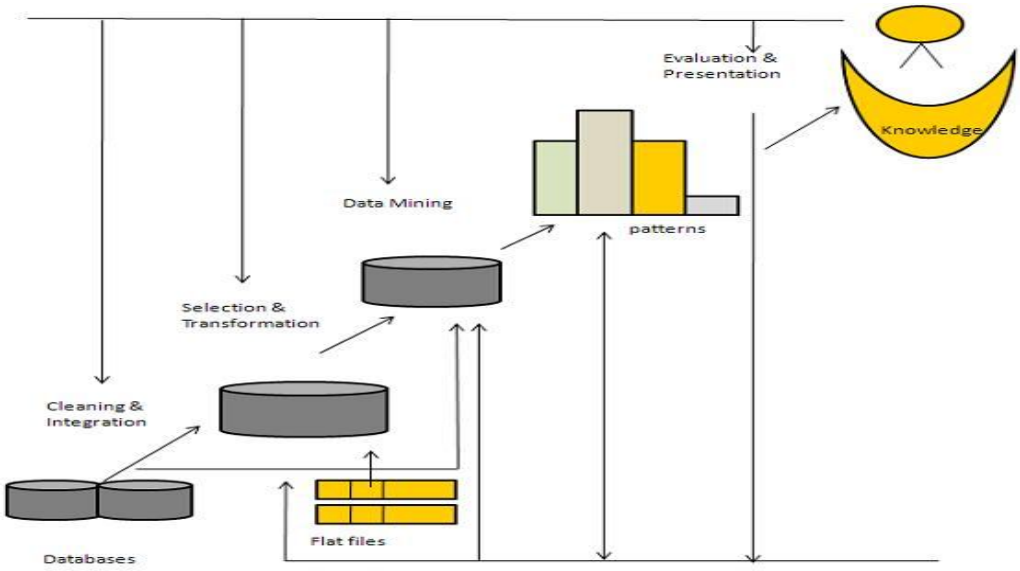
3.1 Data mining method for developing ID models

This paper research work tries to build some enhanced & efficient evidence of intrusion detecting tools or rules that can be used widely to get some statistical network traffic pattern background from the network system of audit data. The advantage of Intrusion detecting tools or rules guides the model developers and gives some starting point and that tells when intrusion is launched, the nature of the normal traffic background of the network is revealed. The resulted unique patterns from these set of rules or tools directs the system developers or engineers quickly to come up with better intrusion detection rules or models , instead of going purely from the new start, and that's the main motivation of this research work.[2]

The research approach is basically three main steps the first step is that to compute statistical pattern of system activity from the audit data and second step identify what are the intrusion patterns that are unique and different from the normal and then based on the analysis third step construct number of features or rules or signatures that can uniquely identify the intrusion that are separate intrusion from normal and based on those features we can easily come up with intrusion signatures or rules. Finally methods and results have been valuated to the greater degree. [3] [16]

© 2015, IJIRIS- All Rights Reserved

3.2 Data mining process of building ID models



The process of applying data mining techniques to build intrusion detection model is iterative in nature. First step start with low audit data from the system since it is binary in nature need to be processed and converted in to ASCII data and in the next step analyze the packets and the events from the audit data and then again we summaries the packets into connection data or summaries events in to session or connection. Session or connection are defined according to schema that consist of number of features that would be very useful to separate normal connection from intrusion ones. It is possible to redefine those features by computing the patters from the connection records and then based on the unique intrusion patterns that separate intrusion from normal ones then with those features build very predictive models and thus given connection record can be identified as intrusion or normal. The performance of the model could be enhanced by just going back and redefining more features then computing more patterns. [5] [16]

3.3 DATA MINING ALGORITHMS FOR INTRUSION DETECTION

Data mining is the process of uncovering or discovering or unveiling hidden potentially useful information from the database or using data warehouse. Data mining process categorized in to two broad categories. These are descriptive data Ming information and predictive data mining information. Descriptive data mining information basically finding pattern that are human interpretable. For example of descriptive tasks are clustering, association rule discovery, and sequential pattern discovery. Predictive data mining information is finding the value of attribute using the value of other attribute. For example predictive tasks are classification and regression and deviation detection. [3]

In this paper we discuss data mining algorithms which are useful to build an efficient intrusion detection models. There are many algorithms in data mining field but only few are very relevant to intrusion detection system. First one is classification algorithms: in classification the task is to map in to predefine categories according to that we can identify the given network connection data normal or intrusion Classification analysis help us to provide a better understanding of large data. Classifier is used for classification. Here the test data is used to estimate the accuracy of classification rules. The classification rules can be applied to the new data tuples if the accuracy is considered acceptable. Basic technique used is rule learner which simpler than complicated mining algorithms. Here rules are intuitive which are obtained through intuition rather than from reasoning or observation and useful for human engineers to interpret and process normal and intrusion data. [5][16]

Another technique most often used was link analysis which determines relation between frequent system features. Example Association rules are mostly used. Sequence analysis used to analyze system and network activities and used to find the sequential pattern. Example Frequent episodes algorithms are mostly used. In most of the system and network activities are temporal context (Of or relating to or limited by time) for example denial of service relies on flooding the targets. They are very strong temporal statistical pattern. For example lot of packets coming to the victim in a very short period of time can be seen as sequential information which is very useful for intrusion detection purposes.

© 2015, IJIRIS- All Rights Reserved

4. CONCLUSION

As the data network systems keeps on increasing every day we need to have efficient and effective techniques for IDS to monitor the networks and to identify, if there is any security violation breached. Data mining techniques for IDS are capable of extracting patterns automatically and adaptively from a large dataset. Various methods related to intrusion detection system are studied briefly. Paper states the methods and techniques of data mining to aid the process of Intrusion Detection in data network environment. The concept intercepting and implementing in real time data networks gives more scope for the research community to work and to build new IDS along with the network expansion.

5. ACKNOWLEDGMENT

I Give thanks to God And wish to acknowledge Professor Wenke lee Computer science Department, Florida Institute of Technology, Melbourne & Dr.A.Arul L.S., my Advisor and other known & unknown research scholars for their support and sharing ideas, views in their respective papers which really gave me an inspiration to complete this paper.

REFRENCE

[1]. http://www.combofix.org/what-it-is-network-intrusion-detection-system.php
[2] www.cerias.purdue.edu CS 590E Security seminar Oct 11 2000
[3] http://www.techcrunch.com/2010/08/04/schmidt-data
[4] K.P.Soman, Shyam DiwakarV. Ajay "Insight into Data Mining Theory and Practice
[5] http://www.tutorialspoint.com/data_mining/dm_quick_guide.htm
[6] R.Venkatesan "A survey on Wireless Intrusion Detection using Data Mining Techniques " in International Journal of Innovative Research in Advanced Engineering(IJIRAE) March 2014
[7] Yi,Mon Aye, Phyu, Thandar " Layering Based Network Intrusion Detection System to Enhance Network Attacks Detection" in International Journal of Science and Research. September 2013
[8] Maher Salem and Ulrich Buehler "Mining Techniques in Networks Security To Enhance Intrusion Detection Systems" in International Journal of Network Security & Its Application. November 2012
[9] Poonam Dabas, Rashmi Chaudhary "Survey of Network Intrusion Detection Using K-Mean Algorithm" in International Journal of Advanced Research in Computer Science and Software Engineering
[10]Pavel Nevlud, Miroslav Bures, Lukas KAPICAK, Jaroslav ZDRALEK "Anomaly-Based network Intrusion Detection Methods" in Information and Communication Technologies and services. December 2013
[11] http://www.slideshare.net/Tommy96/data-mining-techniques-for-network-intrusion-detection-systems.
[12] Robert Mitchell and Ing-ray chen "A survey of Intrusion Detection in Wireless Network Applications. Department of Computer Science. Virginia Tech.
[13]Eric Bloedorn, Alan D.Christiansen, William Hill, Clement Skorupka, Lisa M.Talbot "Datamining for Network Intrusion Detection: How to Get Started" The MITRE Corporation Mclean.
[14]Vikas Markam, Lect. Shirish Mohan Dubey "A General Study of Associations rule mining in Intrusion Detecion System" in International Journal of Emerging Technology and Advanced Engineering January 2012
[15]Huy Anh Nguyen and Deokjai Choi"Application of Data Mining to Network Intrusion Detection: Classifer Selection Model" Chonnam National University, Computer Science Department. Gwangju. Korea.
[16]Wenke Lee, Salvatore J.Stolfo, Philip K.Chan " Real Time Data Mining-Based Intrusion Detection" Computer science Department, Florida Institute of Technology, Melbourne
[17]paul Dokas, Levent Ertoz,Vipin kumar, Aleksandar Lazarevic,Jaideep Srivastava, Pang-Nig Tan " Data Mining for Network Intrusion Detection" University of Minnesota, Minneapolis. USA
[18]Theodoros Lappas and Konstantinos Pelechrinis "Data Mining for (Network) Intrusion Detection system" Department of Computer Science and Engineering. Riverside.
[19]Subaira A.S. Anitha P. "An Efficient Classification Mechanism for Network Intrusion Detection System Based on Data Mining Techniques: A survey" in International Journal of Computer Science and Business Informatics. Coimbatore, October 2013 India.

© 2015, IJIRIS- All Rights Reserved