An Effectual and Secure Approach for the Detection and Efficient Searching of Network Intrusion Detection System (NIDS)

Lekhraj Mehra
M. Tech. Scholar
Department of Computer Engineering,
Swami Keshwanand Institute of Technology, India

Mukesh Kumar Gupta
Reader,
Department of Computer Engineering,
Swami Keswanand Institute of Technology, India

Monika Bhatt Guruji
Asst. Professor,
Department of Computer Engineering,
Geetanjali Institute of Technical Studies, Dabok
Udaipur, Rajasthan, India

ABSTRACT

The concept behind this particular aspect lies on the fact to determine and customize the simplicity and the most basic scenario. The basicity lies on the fact that we have been using the concept of Data Mining and even the algorithms are included that merely includes the efficiency of NIDS that is Network Intrusion Detection System. We have seen a lot of aspects and different concepts being used till this time with different methodologies and functionalities and even used and worked on different technologies. In this we have the major consideration that revolves over and around the algorithm and an emphasis on the technology of Data Mining with software concept of Java eclipse and have tried to improve the working functionality more efficient.

The problem statement comprises of two objectives one is to improve the detection rate and false alarm rate of NIDS uses classification And ensemble technique and the second objective is to improve search efficiency of a NIDS by using association rule mining technique.

Keywords

Data Mining, NIDS, Apriori Algorithm.

1. INTRODUCTION

The basic requirements for the detection, improvement & also for the classification technique in network attacks. Present context with enormous data in a digital world that needs to be saved and a method which can be implemented as “Intrusion Detection (IND)”, these could be defined as either hardware or software that mainly analyze & measures activities of computer systems or network system.

As the world grows with large amount of digital data, which needs to be protected in some way; one way to protect this data is used Intrusion detection systems (IDSs). IDSs are software or hardware systems that monitor and analyze the activities occurring on a computer system or on the network for the sign of intrusion and for malicious activities. The intrusions are defined as efforts to compromise the confidentiality, integrity, and availability, or without checking the security mechanisms of a computer or network. Intrusions can be done in different forms, some of them are by accessing others system from internet by getting unnoticed or authorized user misses using his or her privileges or authorized user try to gain more privileges.

There are basically two types of IDSs one is a host based IDS and other one is network IDS. Host based IDS analyses the stand alone system logs and the packets which come to its network interface card and monitors system activities for malicious activities by users, whereas network IDS monitors analyses the packets which will affect multiple hosts on to the network. There are two ways to analyze or to detect attacks: misuse detection and anomaly detection. Misuse detection, use patterns of well-known attacks which is called as signature to match and identify known intrusions, drawback of this is it fails to detect unknown attacks, but the advantage of this detecting attacks without generating an overwhelming number of false alarms. Anomaly detection, identify abnormal behavior on the system or network. It creates a profile based on the normal behavior of the legitimate user which is collected periodically. Afterwards Analyzer collects the event occurring and use a variety of measures to check whether the activities or behavior deviates. The drawback of this is it generate a large number of false alarms, but the advantage of this is it detects unusual behavior. The most widely used approach is signature based Network IDS.

2. DATA MINING

Data mining refers to extracting information from large amounts of data. Data mining has involved a countless new researchers in the information industry, due to an interdisciplinary field and extensive availability of huge amounts of data and the forthcoming need for turning such data into useful information. The information gained can be used for applications extending from market analysis, scam detection, and customer retention, to invention control and science exploration [1].

Database systems and Internet-based global information systems have also emerged and play an energetic role in the information industry, there are various fields coming together by using data mining-

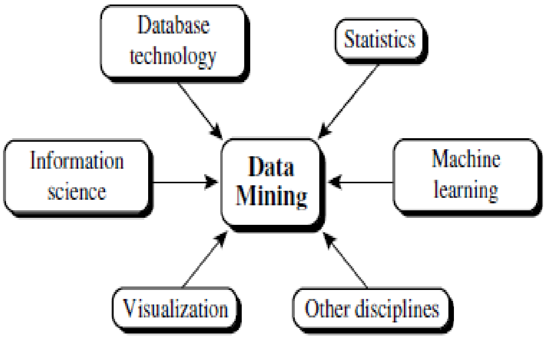

Figure 1. Data mining as multiple discipline [1]

The data mining may also integrate techniques from spatial data analysis, information retrieval, pattern recognition, image analysis, signal processing, computer graphics, Web technology, economics, business, bioinformatics, or psychology [1].

Data mining as a new appreciated method to the IDS to improve data reduction, detection abilities covering hidden patterns, deviances of known attacks or new ones; and maximizing the cost-benefit relationship for an ID utilization [1].

3. NIDS

3.1 Concept of NIDS

Now a day's the world is growing with a large amount of digital data, which needs to accurate detection of intrusion in the network within minimum time in network intrusion detection system. This can be done in various way; one way to protect this data is to use Intrusion detection systems (IDSs). IDSs are software or hardware systems that monitor and analyze the activities occurring on a computer system or on the network. The intrusion can be defined as effort to compromise the confidentiality, integrity, availability, or without checking the security mechanisms of a computer or network. Intrusions can be done in different forms, some of them are by accessing others system from internet by getting unnoticed or authorized user misses using his or her privileges or authorized user try to gain more privileges [2].

3.2 Networking IDS

There are basically three types of IDSs

1. Host based IDS: Host based IDS is used to monitor encrypted traffic data to a specific host. It works on information collected from within an individual computer system. This approach is based on statistics and probability theory and all attacks are taken as a sample space. Then, the set is broken down using statistics that is based on mutually high-class sets. The sample space are used to construct intrusion detection algorithm by the generated subsets. A host based intrusion detection system (HIDS) consists of several IDS over a large network that communicate with each other, or with a central server that enables network monitoring [3].

2. Network IDS: Network based IDS detect malicious activity by monitoring the entire network traffic. IDS systems are installed in general by placing the network interface card in promiscuous mode to capture all networks traffic segments. These networks traffic is checked by IDS to find attacks. To detect known attacks, the logged packets are analyzed and compared with known signature. This approach can block the DOS attack in a virtualized environment. But it cannot detect all types of attacks. It detects only known intrusions coming from external network [3].

3. Hybrid-based IDS: Hybrid systems as the name suggests, is a system that is developed usually using a combination of both host-based and network-based systems, apart from this still many of the IDSs are considered stronger in one field or the other [4].

3.3 Differences in HIDS and NIDS

There are some differences in network based intrusion detection system and host based intrusion detectable that are shown in table.

Table 1. Table Shows Difference between HIDS and NIDS

Types	Advantages	Drawbacks
HIDS	1. It can detects the attacks that do not involve the network, can analyze what an application is doing.	1. They are insulated from network activities, must be installed on every host.
NIDS	1. It can monitor multiple hosts at a time. 2. It can correlate attacks against multiple hosts. 3. It does not affect the host performance. 4. It can detect attacks that are not visible from single hosts.	1. It must be able to keep up with the network speed. 2. It may have problem with encrypted channels.

3.4 Classification of Attacks

Generally there are mainly four categories of attacks but here we are discussing six categories [5].

1. **Denial of Service (DoS):** DoS is the common attack in which hacker makes a computing class on memory resources which are either engaged or full to serve harmful networking requests and thus, denying user's access to a machine [5].

2. **Remote to User Attacks (R2L): In** this type of attack the user sends packets of messages to another through the internet, which they does not have access to expose the machines vulnerabilities and exploit privileges which a local user would have [5].

3. **User to Root Attacks (U2R):** The exploitations in which the hacker starts the system with a normal user account and attempts to abuse vulnerabilities in order to gain super user advantage [5].

4. **Probing:** Probing is an attack in which the hacker scans a networking device in order to determine weaknesses and injuries that may later the system crashes. This technique is commonly used in data mining [6].

5. **Eavesdropping attack:** it is a network layer attack. It capturing massage packets transferred from one computer to another and reading the security information like passwords and any kind of confidential information [6].

6. **Man-In-The-Middle Attack:** In this the attacker makes independent connections with the victims and conveys messages between them and make them believe that they are talking directly on private network. But the aspect lies about that attacker knows all information [6].

4. EXISTING APPROACHES FOR IDS

IDS can be designed using different kind of soft computing techniques as below-

4.1 Artificial Neural Network (ANN)

Neural networks are an approach to computing that involves developing mathematical structures with the ability to learn. The methods are the result of academic investigations to model nervous system learning. Neural networks have the remarkable ability to derive meaning from complicated or imprecise data and can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. Artificial Neural Network (ANN) in intrusion detection is used to generalize data from incomplete data and able to classify data as being normal or intrusive.

An Artificial Neural Network (ANN) is an information-processing paradigm that is inspired by the biological nervous systems, like the brain and even other information. The main aspect of this system is information processing. This paradigm is the novel structure of the information processing system. It consists of neurons that are interconnected elements (neurons) working in unison to solve specific problems. Each neuron is linked to certain of its neighbors with coefficients of connectivity that represent the strengths of Connecting, also the process to learn is to be completed by adjusting the criteria that works to make the overall network to output appropriate results [7].

4.2 Support Vector Machine (SVM)

SVM is used to detect intrusions based on limited sample data, where its dimensions will not affect the accuracy. It deals with multi-class classification problem. It transforms data into a feature space "F" which has a huge dimension. Its generalization depends on the geometrical characteristics of the learning data not on the dimensions of the input space. Training a support vector machine leads to a quadratic optimization problem with bound constraints and one linear equality constraint [8]

4.2 Genetic Algorithm (GA)

Genetic algorithms (GAs) are used to select network features or to determine optimal parameters which can be used in other techniques for achieving good results and improving accuracy of IDS. The process usually begins with randomly generated population of chromosomes, which signify all optimal solution of a problem that is considered as candidate solutions. From each and every chromosome are changed over as bits according to different positions. Characters or numbers are referred to as genes. A valuation functions are used to calculate the goodness of each chromosome according to the desired solution. For survival and combination the selection of chromosomes is biased towards the fittest chromosomes [9]

4.3 Fuzzy Logic

It a form of reasoning that is derived from fuzzy set theory. This provides a powerful mechanism for representing obscure concepts. Data mining methods are used to automatically learn patterns from large quantities of data. The integration of fuzzy logic with data mining methods will help to creates more summarized patterns at a higher level than at the data level. Patterns that are more summarized are user dependent on data will be helpful for the instruction detection. Simple Fuzzy rules allow us to construct if-then rules [9].

4.4 Decision Trees

A decision tree is the most popular classification techniques used for detection of intrusion, anomalies and web based attacks. Decision trees are tree like structure in which the nodes are labelled with attribute names, the edges are labelled with possible values for this attribute and the leaves labelled with different classes. Decision trees generates the association rules where internal nodes represent a test on attributes, branch represents an outcome of the test and leaf node represents a class label. Decision Trees algorithms are used in machine learning and pattern recognition; there are various decision tree learning algorithms like ID3, C4.5 and CART [6].

4.5 Ripper Rule

Ripper stands for (Repeated Incremental Pruning to Produce Error Reduction) the rule which is efficient rule based learning algorithm that process the various noisy dataset. Its target that take on more than two unique values. It consist of two stages which are used to minimize the amount of error [10].

1. First initialize the rule condition.
2. Usages rule optimization technique.

4.6 Bayesian Network

A Bayesian Network involves both a graphical model and a probabilistic model that represents a random variable and condition through a Direct Acyclic Graph (DAG) which represent a random variable that may be discrete or continuous. It provides the high accuracy and high speed for handling large database. In the graph nodes represent random variables and conditional dependencies are represented by edges. Thus, Nodes which are not connected represent variables which are conditionally independent to each other. For each variable classifier maintain one conditional probability table (CPT) that require higher computational efforts [6].

4.7 K-nearest Neighbour

K-Nearest Neighbour (k-NN) is a type of Lazy learning, that stores a given training tuple and waits until it is given a test tuple. It is a case based learner that classifies the objects based on closet training examples in the feature space. For a given unknown tuple, a k-Nearest Neighbour looks the pattern space for the k-training tuples that are closest to the unidentified tuple. It is the simplest algorithm among all the machine learning algorithms. [6].

5. INTRUSION DETECTION BEFORE DATA MINING

While performing intrusion detection on network, generally focus on fundamental issues rather than data mining. As the data enter the sensor tunes, incident investigation and system performance seeks attention. The analyst team grew to handle the load, and training and team coordination were the issues of the day. But the level of exploration and attack on the internet was constantly increasing, along with the amount of data that was collected and puts in front of the analysts by us.

6. DESIGNING OF FUNCTIONAL MODEL

The following section describes procedure to extend network intrusion detection system (NIDS) and improve to make searching of signature easier in NIDS.

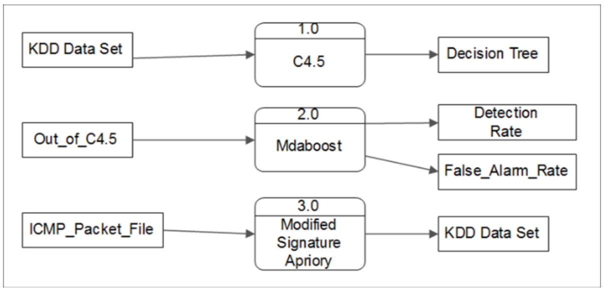

Figure 2. Functional Model

6.1 Solving Approach

In this paper to solve the problem data mining approach is used as follows:

1. Dataset which contains attacks are passed to C4.5 algorithm and the output of this algorithm is a text file with classification of attacks and normal packets.
2. This classified file is given as an input to Mdaboost ensemble algorithm to improve the accuracy of classification.
3. Dissertation uses open source NIDS (Snort) which, takes a packet from network and compares it with the signatures stored in database this process is time consuming, in order to improve search efficiency we use modified signature Apriori which is based on Apriori algorithm of data mining.

6.2 An Architecture for IDS

The major problem of using data mining approaches in intrusion detection is that it requires a large amount of audit data that will compute the profile rule sets. That one need to compute a detection model for each resource in a target system makes the data mining task daunting. Moreover, this learning (mining) process is an integral and continuous part of an intrusion detection system because the rule sets used by the detection module may not be static over a long period of time. Proposed system architecture, includes proposed NIDS model.

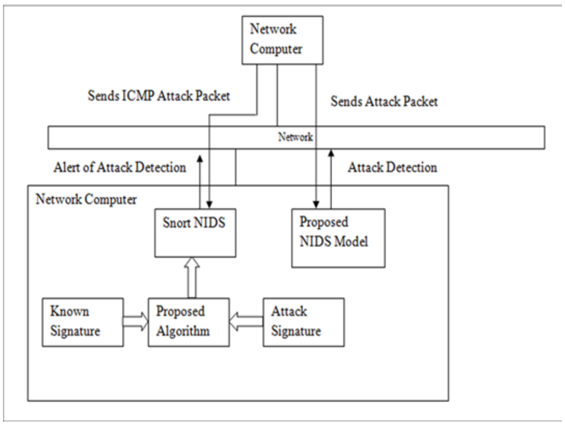

Figure 3. Architecture for Intrusion Detection System

6.3 Efficient Searching

In the second half of this paper, we use modified signature Apriori and mathematics associated with it to make searching of signature easier in source NIDS (Snort) which, takes a packet from network and compares it with the signatures stored in database this process is time consuming, in order to improve search efficiency.

In order to improve search efficiency, we keep the minimum support at a very high level and uses scan reduction technique, Scan the database D to calculate the support of an attack in all transactions T. This will generate L_1 frequent attack set L_1 = {s_1, s_2, s_3....... Sn} here Sn is the support count of each attack which is above minimum support. It is increased by one when same attack is encountered again.

1. Add knw_sign + attack into Cn
2. Ln = {c∈Cn | support (c) ≥ Minimum_ Support}
3. Final candidate set is formed like bellow
4. For each attack in L1 add sig + attack into Ck
5. Lk = {c∈Cn | support (c) ≥ Minimum_Support}

We know that C_i+1 is generated from Li* Li. Now, a Ci generated from Ci * Ci, instead of from Li* Li, will of larger size than | Ci | where Ci is generated from Li* Li. However, if | Ci'| is not much greater than | Ci |, we may save one round of database scan. With this technique we can reduce the database scan and hence the searching for the signature of an attack becomes faster and the search efficiency gets improved.

7. RESULT ANALYSIS

Outcomes of the paper is a comparison of the results of the proposed algorithm with Genetic programming along C4.5 algorithm and improved search efficiency for signatures using a modified signature algorithm in NIDS

Table 2. Comparison Table

Parameter	Proposed System	Previous System
Detection Rate	89.56%	82.43%
False Alarm Rate	0.1%	0.1%

For the efficient searching, the ping ICMP packet to a PC which has IP address 192.168.1.244 where the NIDS is running. It shows that, packets were sent and same were received back with 0.0% loss.

8. CONCLUSION

Some papers used data mining approach in IDS, but they focused on anomaly detection IDS and concentrated on finding association among attributes and not considered content of traffic for mining and some focused on signature based NIDS but these papers have not considered both detection rate and search efficiency parameters of NIDS.

In this paper proposed work it considered both detection rate and search efficiency parameters of NIDS in order to improve the performance of NIDS. The proposed system is compared with the previous system, on the basis of parameter, detection rate, it results in a better system with 89.56 % in compared 82.43% accuracy. But, false alarm rate results 0.1 % for both proposed and previous system. It established that the proposed method to improve the detection rate and false alarm rate is one of the alternatives to build NIDS model and also it is proved that search for signature using modified signature Apriori is also easy and faster.

9. FUTURE SCOPE

- It uses the different dataset on this dissertation to check how this model works.
- Use these models to build commercial NIDS.
- Use different technique of data mining to create NIDS model.

10. ACKNOWLEDGMENTS

A journey is easier when you travel together. Interdependence is more valuable than independence. I would like to express my gratitude to all those who gave me the possibility to complete my paper.

I take immense pleasure to express my deep and sincere gratitude to my esteemed guide, Mr. Mukesh Kumar Gupta for his valuable guidance, and spending of his precious hours for my work.

If anyone is acknowledged first, they are my family members and my lovely wife and my son Nimish & daughter Diyanshi.

11. REFERENCES

[1] Jiawei Han, Micheline Kamber, Jian Pei, “Data Mining: Concepts and Techniques: Concepts and Techniques”, ISBN 978-0-12-381479-1, Elsevier, 2011.

[2] Rafeeq Ur Rehman, “Intrusion Detection Systems with Snort: Advanced IDS Techniques Using Snort, Apache, Prentice Hall Professional, 2003.

[3] Manthira Moorthy S and Rajeshwari M, “Virtual Host based Intrusion Detection System for Cloud”, International Journal of Engineering and Technology, Vol 5 No 6, ISSN 0975-4024, 2014

[4] S. Shidore and V. K. Bhusari, “Evasion of Network Intrusion Detection System Using functional Framework”, International Journal of Application or Innovation in Engineering & Management, Volume 3, Issue 6, ISSN 2319 – 4847, 2014.

[5] Mostaque Md. Morshedur Hassan, “Network Intrusion Detection System Using Genetic Algorithm and Fuzzy Logic”, International Journal of Innovative Research in Computer and Communication Engineering, Vol. 1, Issue 7, September 2013.

[6] Subaira.A.S, Anitha.P, “A Survey: Network Intrusion Detection System based on Data Mining Techniques”, International Journal of Computer Science and Mobile Computing Vol.2 Issue. 10, pg. 145-153, 2013.

[7] Akansha Parashar and Praneet Saurabh, “A Novel Approach for Intrusion Detection to improve the detection rate using Artificial Immune system and Neural Network Technique”, People’s Journal of Science & Technology, Vol.2 (1),ISSN 2249 5487, 2012

[8] Muamer N. Mohammad, Norrozila Sulaiman and Emad T. Khalaf, “A Novel Local Network Intrusion Detection System Based on Support Vector Machine”, Journal of Computer Science, 7 (10) 1560-1564, ISSN 1549-3636, 2011.

[9] Mohammad Sazzadul Hoque, Md. Abdul Mukit and Md. Abu Naser Bikas, “An Implementation of Intrusion Detection System Using Genetic Algorithm”, International Journal of Network Security & Its Applications, Vol.4, No.2, 2012.

[10] S. K. Wagh, M. S. Bhalerao, A. M. Pathan, J. S. Bhole and S. L. Pingale, “To Implement an Effective Real-Time Network Intrusion Detection System to Reduce False Alarm Rate and Improve Efficiency, Using Machine Learning Techniques”, International Journal of Researchers, Scientists and Developers Vol. 1 No. 1 , ISSN 2347-3649, 2013.