MODERNIZED INTRUSION DETECTION USING ENHANCED APRIORI ALGORITHM

Lalli[1] and Palanisamy[2]

[1]Department of Computer Science, Bharathidasan University, Trichy
Lalli_bdu@yahoo.co.in
[2]Department of Computer Science & Engineering, Alagappa University, Karaikudi
vpazhanisamy@yahoo.co.in

ABSTRACT

Communication networks are essential and it will create many crucial issues today. Nowadays, we consider that the firewalls are the first line of defense but that policies cannot meet the particular requirements of needed process to achieve security. Most of the research has been done in this area but we are lagging to achieve security needs. Already many models such as ADAM, DHP, LERAD and ENTROPHY are proposed to resolve security problems but we need an efficient model to detect new types of various intrusions within the entire network. In this paper, we proposed to design a modernized intrusion detection system which consist of two methods such as anomaly and misuse detection. Both are integrated and also used to detect novel attacks. Our system proposed to discover temporal pattern of attacker behaviors, which is profiled using an algorithm EAA (Enhanced Apriori Algorithm). This is experimented with a simple interface to display the behaviors of attacks effectively.

KEYWORDS

Intrusion Detection System, Data Mining, Security, Intrusions, Apriori Algorithm

1. INTRODUCTION

Intrusion detection is a device that monitors activity to identify malicious or suspicious events and also produce reports to a management station. Many new intrusions or attackers are coming out every day to abusing the computing system privileges. An intrusion detection system is a sensor, like smoke detector, that raises an alarm if any specific things occur. According to the alarms, the security analyst will be detected as positive and negative ones separately. If it is a smaller network, the intrusion detection systems perform simple pattern matching and report situations that match a pattern corresponding to a known attack type. But the detection of new attacks is not possible in a complex network. The quantity of traffic makes it obligatory for network administrators and network security experts to use specific tools, called intrusion detection systems (IDS), to prune down the monitoring activity they require to do. The IDS are evaluated on the basis of its accuracy, efficiency and usability. The characteristics used to evaluate the accuracy of the IDS are as follows. 1. Detection Rate: The number of attacks to be detected by the system. 2. False Positive Rate: It is the percentage of normal data that the system in-correctly determines to be intrusive. The accuracy of the different IDS can be compared with the detection rate against false positive rate. The comparison should be represented a plot named as Receiver Operating Characteristic (ROC) curve. The accuracy gives a measure of the ability of the IDS to handle noisy, discrete data and to adapt.

In general, detection mechanism used by IDS can be classified into two major categories. 1)Signature based detection: Models built from well known attack types, that is already known attack patterns. 2)Anomaly based detection: Modeled using normal traffic and deviation from this profile is considered anomalous. 3) Hybrid of both anomaly and misuse detection: Many systems are used anomaly detection and misuse detection to detect various types of intrusions

DOI : 10.5121/ijwmn.2013.5206

effectively. To build powerful IDS, it is necessary to enumerate the desirable characteristics such as fault tolerance, imperviousness to subversion, scalability, adaptability, minimal overhead, configurability, graceful degradation of service.

The rest of the paper is organized as follows. In section 2, we briefly describe about Data Mining and their related work for IDS. We put forth a detailed description of our algorithm in section 3. We discuss about the experimental results in section 4. In section 5, we summarize the paper.

2. DATA MINING TECHNOLOGY

Data Mining is knowledge discovery in databases. Data Mining is knowledge discovery in databases. Data mining techniques are used to find out the hidden patterns automatically in an entire database. Data warehouse is an integration of multiple databases, data cubes and files. We have mined all the data, to predict it according to their relationship. Data mining is the search for the relationships and global patterns that exist in large databases but are hidden among vast amounts of data. It is also denoted as automated statistical analysis. The most commonly used techniques in data mining are: Neural networks, Genetic algorithms, Association rule mining, Clustering and Classification. The association rule mining technique is applied for our proposed work.

2.1. Related Work

Ye Changguo [1] has described about Network Intrusion Detection System (IDS), as the main security defending technique, are second guard for a network after firewall. Ye Changguo [1] has described about Network Intrusion Detection System (IDS), as the main security defending technique, are second guard for a network after firewall. The author described that the huge database log can be mined with the help of fuzzy rule mining algorithm. It gives the better results compared to Apriori algorithm.

Aly Ei-Senary et al[2] has integrated Apriori and Kuok's algorithms to capture the features of network traffic using fuzzy logic rules. Amir Azimi alasti et al[3]says that the reduction of false positive alerts and correlation between alerts are done by formalized SOM. The accuracy can be improved by Alert Filtering and Cluster Merging algorithms. Anderson J. P.[4] using data mining techniques and signature based tools for MINDS. The model can be developed by Denning D.E et al[5] to monitor the abnormal activities. Dewan M et al[6] has used ISABA for alert classification and also to reduce the number of false positives. Each and every user behavior can be identified by Neural network by Jake Ryan et al[7].

Jin-Ling Zhao[8] proposed to achieve the high detection rate and low false alarm rate through clustering and genetic optimizing stages. Norouzian M.R et al[9] described the neural network concepts to detect various attacks and also to classify six groups of two hidden layers in Multi-Layer Perceptron (MLP). All the system calls to be traced out by Host based Intrusion Detection System. OSwais. S et al[10] proposed the membership function can be tuned by genetic algorithm, it is performed by using approach based IDS and also implemented by Gas. Alan Bivens et al[11] has proposed the NIDS model to classify self organizing maps for data clustering. The MLP neural network is one of the best ways to creating uniform, grouping of inputs and they are detected effectively when a dynamic number of inputs are presented. S. Sathyabama et al[12] has described about clustering techniques and outliers for grouping the user's behaviors and also detecting different behaviors easily.

Teng.H.S [13] applied sequential rules to capture a user's behavior over time. According to the User activity patterns, the specific rules are used to find out the specific activities. High quality sequential patterns are automatically generated using inductive generalization and lower quality patterns are eliminated. An automated strategy for generation of fuzzy rules obtained from definite rules using frequent items. Taeshik Shon et al [14] proposed an enhanced SVM approach framework for detecting and classifying the novel attacks in network traffic such as packet profiling using SOFM, packet filtering using PTF, field selection using Genetic Algorithm and packet flow-based data pre-processing. SOFM clustering was used for normal profiling. The SVM approach provides false positive rate similar to that of real NIDSs.

Stephen F. Ownes, R. R. Levary[15] has stated that the expert system techniques are used for intrusion detection systems. The fuzzy sets are used to find out various types of attacks. Depends up on the types of threat it is easy to implement the expert system technology.

In that specific problem, we wish to tackle through our IDS is, to generate a modernized intrusion detection which will effectively detects both anomaly and misuse detections using EAA (Enhanced Apriori Algorithm) and also will reduce false positives effectively.

3. Architecture framework and Proposed Enhanced Apriori Algorithm

It is necessary to detect the potential attacks for anomalous behavior from the existing patterns because we predict the future actions of the attacker's. Sequential pattern mining techniques are used to analyze various IDS alerts. In this paper, we propose a model modernized intrusion detection system consist of two methods, anomaly detection and misuse detection to detect intrusions. We have integrated both anomaly and misuse detection due to its ability to detect novel attacks. Our system is used to find out the behaviors of various attacks using EAA.

In this paper we proposed to design and develop an intelligent data mining intrusion detection system and its core part a composite detection engine with anomaly detection and misuse detection features and the two detection engines work serially to detect the user's activity in turn.

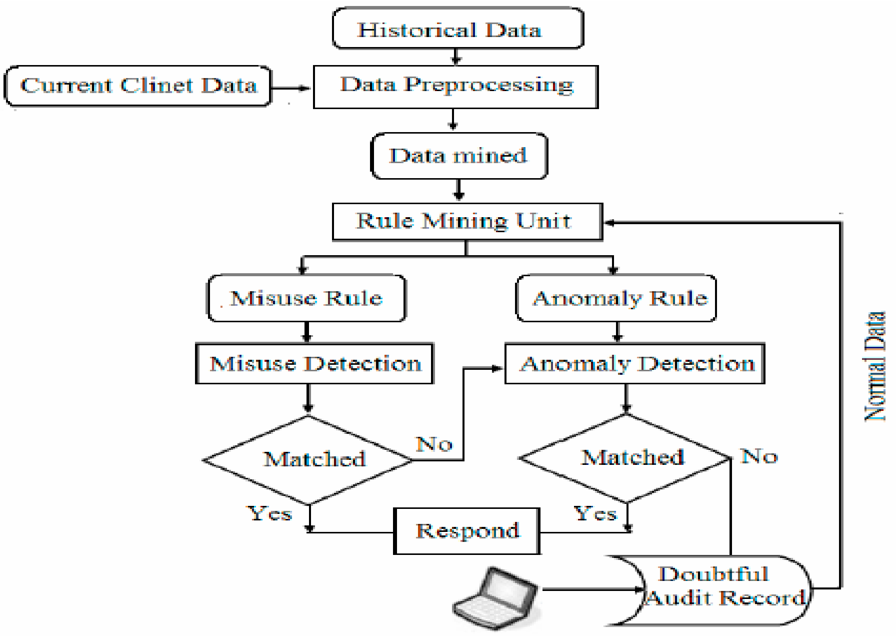


Figure 1. Design Architecture

The real time data's are collected from the database audit system. We have analyze the audit data and to find out its normal and abnormal behaviors. As a result, identify the abnormal activity through statistical analysis.

The design of architecture framework is mainly composed of the following parts: data collecting and pre-processing module, association rule mining module and intrusion detection analysis module, etc as shown in Figure 1.

3.1 The Proposed Algorithm - EAA

In this section, we proposed the Enhanced Apriori Algorithm(EAA). It is an extension of basic Apriori Algorithm. Several rule sets are framed as follows.

Rule 1 : If both anomaly and misuse module detected an attack then the specific attack can be classified by misuse module.

Rule 2 : If misuse module can detect an attack then the specific attack can be classified by misuse module.

Rule 3 : If anomaly module can detect an attack then the specific attack can be declared as unclassified attack.

```
L1= {frequent candidates};
**for** (k= 2; $L_{k-1}$ !=∅; k++) **do begin**
$C_k$= candidates are generated from $L_{k-1}$ (cartesian product $L_{k-1}$ x $L_{k-1}$ and
eliminating any k-1 size itemset is not a frequent one);
While $L_{K-1}$≠ϕ
$C_k$ <- Generate($L_{K-1}$)
For transactions t € T
Ct, <- Subset($C_k$t)
For candidates c € $C_t$
count[c] <- count[c] + 1
$L_K$ <- { c € $C_k$ | count [c] >= € }
K <- K+1
Return µ $L_k$
    **for each** transaction t in database **do**
        increment the count of all candidates in
         $C_k$ that are contained in t
    $L_k$ = candidates in $C_k$ with min_sup
    **end**
Return µ $L_k$;
    Let's define:
        $C_k$  candidate itemset of size k
        $L_k$  frequent itemset of size k


Main steps of iteration are:
    Find frequent set $L_{k-1}$
    Join step: $C_k$ is generated by joining $L_{k-1}$ with itself
              ( cartesian product $L_{k-1}$ x $L_{k-1}$ )
     Prune step (apriori property):
       Any (k – 1) size itemset is not a frequent subset
       of k size itemset, then it should be removed
    Frequent set $L_k$ has been achieved
```

Figure 2. EAA Algorithm

Association Mining aims to extract interesting correlation, frequent patterns, and associations among sets of items or objects in transaction databases. Association rules satisfy both a minimum support threshold (min_sup) and minimum confidence threshold (min_conf). Audit data can be obtained from various databases, it consist of number of rows (audit record) and columns (system features). Association rules are used to capture the consistent behaviors and then it is implemented through aggregate rule sets. Our proposed algorithm is in pseudo code format.

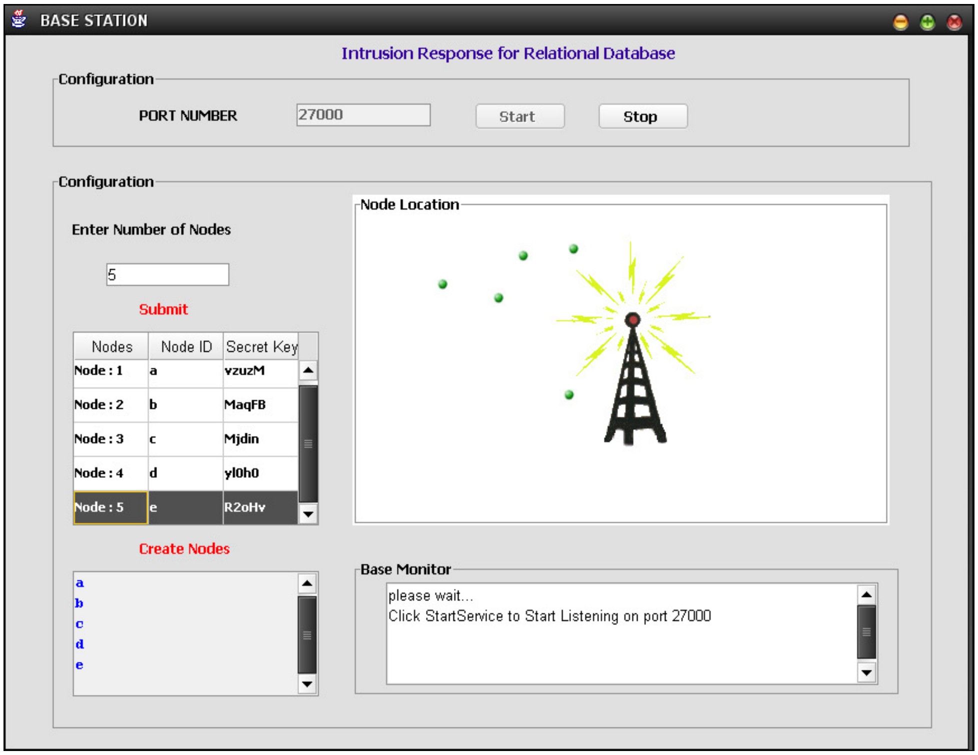
4. Experimental Results

The proposed EAA is experimented to find out the efficiency of detecting false positives and false negatives that are occurred in IDS. The new type of attacks can be detected effectively by EAA and then automatically knowledge base can be updated. The experimental result shows, that it was better to find normal packet performance, faster in terms of execution time and also detecting the effectiveness and accuracy of false positives and false negatives.

The input parameters are data size (The size of the each audit record in database) and Data Length (Processing time of each audit record measured in seconds). Our proposed technique EAA approach gives better performance in evaluating large size audit data when compared to the existing techniques.

False Positive Rate = Number of false positives / Total number of abnormal labels.

False Negative Rate = Number of false negatives / Total number of normal labels.



PORT NUMBER 27000

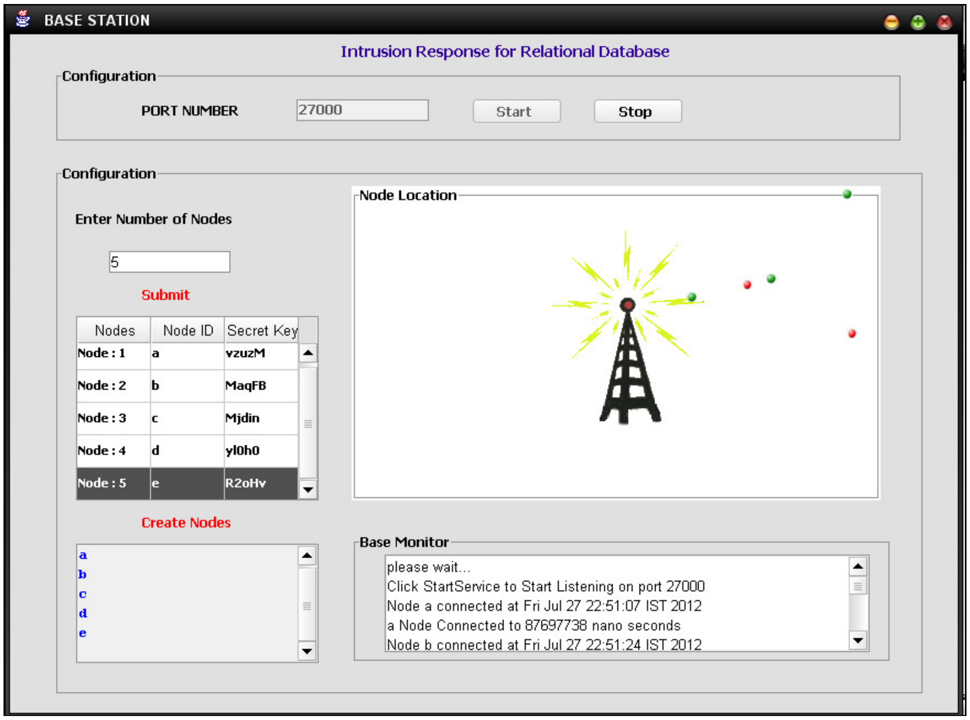
Enter Number of Nodes 5

Nodes	Node ID	Secret Key
Node : 1	a	vzuzM
Node : 2	b	MaqFB
Node : 3	c	Mjdin
Node : 4	d	yl0h0
Node : 5	e	R2oHv

a
b
c
d
e

please wait...
Click StartService to Start Listening on port 27000

Figure 3. Nodes without intrusion



PORT NUMBER 27000

Enter Number of Nodes 5

Nodes	Node ID	Secret Key
Node : 1	a	vzuzM
Node : 2	b	MaqFB
Node : 3	c	Mjdin
Node : 4	d	yl0h0
Node : 5	e	R2oHv

a
b
c
d
e

```
please wait...
Click StartService to Start Listening on port 27000
Node a connected at Fri Jul 27 22:51:07 IST 2012
a Node Connected to 87697738 nano seconds
Node b connected at Fri Jul 27 22:51:24 IST 2012
```

Figure 4. Nodes with intrusion

Table 1. Evaluation between existing and proposed Approach

Data Size	Data Length (Existing)	Data Length (Proposed)
1876	34	10
876	56	13
4526	87	20
5435	35	27
987	65	11
1234	27	16
675	74	08
2345	55	22
5432	33	25
5261	23	23
2176	22	19
165	11	02
765	65	09
763	23	09
343	25	10
7845	64	33

This paper improves the detecting speed and accuracy as a goal, and proposed a more efficient EAA (Enhanced Apriori Algorithm) to abnormal detecting experiment to be based on network.

In this approach we shows that new type of attack can be detected effectively in the system and the knowledge base can be updated automatically.

Finally the experimental results are compared and analyzed. Our proposed algorithm is effectively identify malicious attacks and also increase the efficiency. This section will shows the result based on the inputs as, Data Size: The size of the each audit record in database and Data Length: Processing time of each audit record measured in seconds. From the Figure 5. We can conclude that our proposed technique called EAA approach gives better performance in evaluating large size of audit data when compared to the existing techniques.

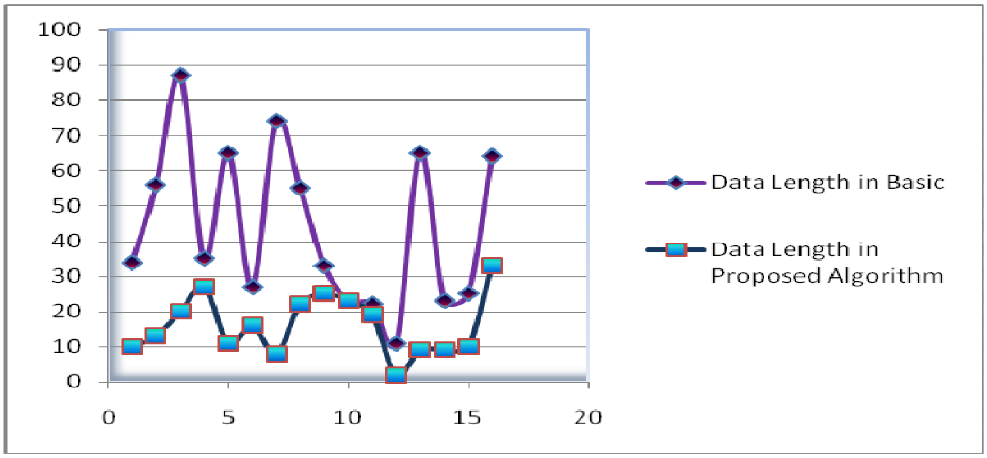


Figure 5. Performance Analysis

The IDS are evaluated on the basis of its accuracy, efficiency and usability. The characteristics used to evaluate the accuracy of the IDS are, *Detection Rate*: It is the percentage of attacks that a system detects and *False Positive Rate*: It is the percentage of normal data that the system incorrectly determines to be intrusive.

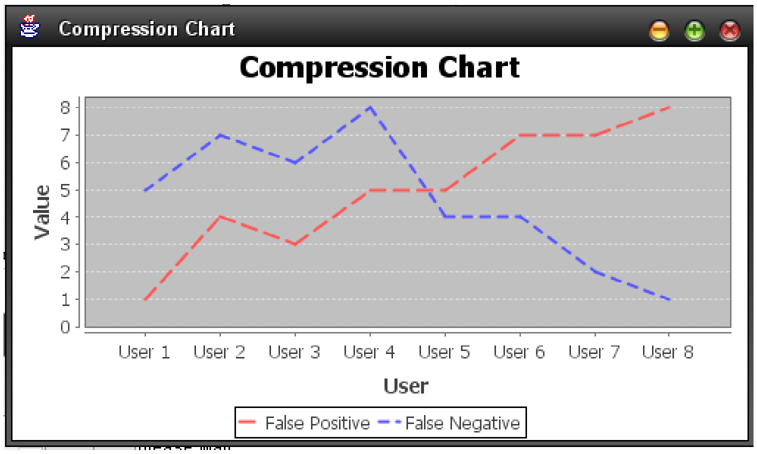


Figure 6. Compression chart for false positives and false negatives

Different types of IDS can be analyzed according to the detection rate and it is plotted as a Receiver Operating Characteristic (ROC) curve. The accuracy gives a measure of the ability of the IDS to handle noisy, discrete data and to adapt. The compression range of false positives and false negatives are shown in Figure 6.

The experimental results shows that the security range between existing and proposed technique based on the ratio (59:67) is in Figure 7.

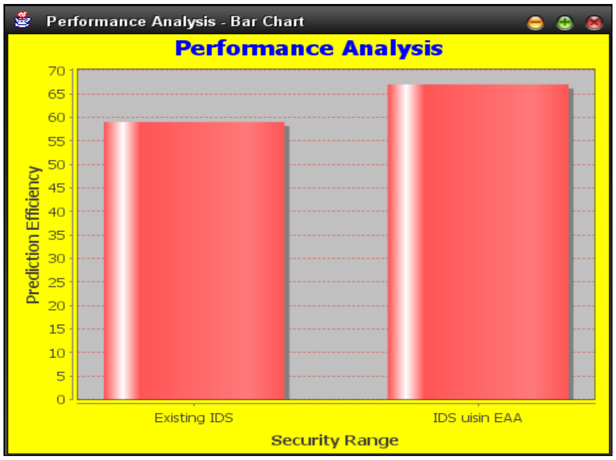


Figure 7. Performance Analysis for Existing and Proposed system

An existing research, it is based on simple and basic association algorithm to enhance the efficiency of association rule mining, which is not sufficient one to generate frequent item set. Another problem is in fuzzy set theory, it is often requires several analysis before the number of frequent item set produce. It can be very sensitive to the choice of initial analysis and it does not yield the same result with each run, since the resulting frequent item set depends on the initial random assignments. So it is not sufficient for runtime intrusion detection efficiency.

Reasons for Supremacy over other algorithms:-

- Proposed Design will be better than existing to find normal packet performance.
- Proposed Design will be faster than existing in terms of execution time.
- Proposed Design will be smaller than existing and easy to understand and implement.
- It will do not contain complex structure, control flow will be well defined and looping structure will be minimized. Due to the above facts it will take very less time for execution.

It is a modern Intrusion Detection with Intelligent Analysis that meets the challenges that traditional intrusion detection systems failed to meet:

- need for accuracy
- need for active responses
- need for speed and reliability
- need for usability

5. CONCLUSION & FUTURE WORK

Using advantages of characteristic extraction of data mining theory in dealing with large amounts of data, the new model MIDS which combines both misuse detecting and anomaly detection is proposed in this paper. In this model we applying Enhanced Apriori (EAA) algorithm to the training data set that containing a large number of intrusions to establish and update the 'normal' and 'abnormal' behavior rules. It is, therefore, concluded that the given methods can be easily adapted to the configuration of MIDS to reduce the manual efforts used to configure them, reduce alert generation and hence improve their effectiveness. And the results of simulation manifested that the new model has a trait of high detection rate and low false positive rate. Besides, unknown intrusion can be detected at the same time effectively.

In Future Enhancement, we investigate the methods and benefits of combining multiple simple detection models. In Future Enhancement, we investigate the methods and benefits of combining multiple simple detection models. We enhance this model by using multiple audit data sets from multiple data streams to improve the accuracy and efficiency.

REFERENCES

[1] Ye Changguo , "The Research on the Application of Association Rules Mining Algorithm in Network Intrusion Detection" Transactions on Software Engineering, IEEE Communication Magazine.

[2] Aly Ei-Semary, Janica Edmonds, Jesus Gonzalez-Pino, Mauricio Papa, "Applying Data Mining of Fuzzy Association Rules to Network Intrusion Detection", in the Proceedings of Workshop on Information Assurance United States Military Academy 2006, IEEE Communication Magazine, West Point, Y,DOI:10.1109/IAW.2006/652083.

[3] Amir Azimi, Alasti, Ahrabi, Ahmad Habibizad Navin, Hadi Bahrbegi, "A New System for Clustering & Classification of Intrusion Detection System Alerts Using SOM", International Journal of Computer Science & Security, Vol: 4, Issue: 6, pp-589-597, 2011.

[4] Anderson.J.P, "Computer Security Threat Monitoring & Surveilance", Technical Report, James P Anderson co., Fort Washington, Pennsylvania, 1980.

[5] Denning .D.E, "An Intrusion Detection Model", Transactions on Software Engineering, IEEE Communication Magazine, 1987,SE-13, PP-222-232,DOI:10.1109/TSE.1987.232894.

[6] Dewan Md, Farid, Mohammed Zahidur Rahman, "Anomaly Network Intrusion Detection Based on Improved Self Adaptive Bayesian Algorithm", Journal of Computers, Vol 5, pp-23-31, Jan 2010, DOI:10.4.304/jcp 5.1.

[7] Jake Ryan, Meng - Jang Lin, Risto Miikkulainen, "Intrusion Detection With Neural Networks", Advances in Neural Information Processing System 10, Cambridge, MA:MIT Press,1998,DOI:10.1.1.31.3570.

[8] Jin-Ling Zhao, Jiu-fen Zhao ,Jian-Jun Li, "Intrusion Detection Based on Clustering Genetic Algorithm", in Proceedings of International Conference on Machine Learning & Cybernetics (ICML),2005,IEEECommunicationMagazine,ISBN:0-7803-9091-DOI:10.1109/ICML.2005.1527621.

[9] Norouzian.M.R, Merati.S, "Classifying Attacks in a Network Intrusion Detection System Based on Artificial Neural Networks", in the Proceedings of 13th International Conference on Advanced Communication Technology(ICACT), 2011,ISBN:978-1-4244-8830-8,pp-868-873.

[10] Oswais.S, Snasel.V, Kromer.P, Abraham. A, "Survey: Using Genetic Algorithm Approach in Intrusion Detection Systems Techniques", in the Proceedings of 7th International Conference on Computer Information & Industrial Management Applications (CISIM), 2008, International Journal of Computer Applications (0975 – 8887) Volume 35– No.8, December 2011 56 IEEE Communication Magazine,pp-300-307,ISBN:978-0-7695-318-7,DOI:10.1109/CISM.2008-49.

[11] Alan Bivens, Chandrika Palagiri, Rasheda Smith, Boleslaw Szymanski, "Network-Based Intrusion Detection Using Neural Networks", in Proceedings of the Intelligent Engineering Systems Through Artificial Neural Networks, St.Louis, ANNIE-2002, and Vol: 12, pp- 579-584, ASME Press, New York.

[12] Sathyabama.S, Irfan Ahmed.M.S, Saravanan.A,"Network Intrusion Detection Using Clustering: A Data Mining Approach", International Journal of Computer Application (0975-8887), Sep-2011, Vol: 30, No: 4, ISBN: 978-93-80864-87-5, DOI: 10.5120/3670-5071.

[13] Teng.H.S, Chen.K and Lu.S.C, "Adaptive Real-Time Anomaly Detection using Inductively Generated Sequential Patterns, in the Proceedings of Symposium on research in Computer Security & Privacy, IEEE Communication Magazine,1990, pp-278-284.

[14] Taeshik Shon, Jong Sub Moon, "A Hybrid Machine Learning Approach to Network Anomaly Detection", Information Sciences 2007, Vol: 177, Issue: 18, Publisher: USENIX Association, pp- 3799-3821, ISSN:00200255,DOI:10.1016/j.ins-2007.03.025.

[15] Stephen F. Owens, Reuven R. Levary, "An adaptive expert system approach for intrusion detection", International Journal of Security and Networks, Vol: 1, No: 3/4, pp: 206-217, 2006.

[16] Tcptrace software tool, www.tcptrace.org.

[17] SNORT Intrusion Detection System. www.snort.org.

[18] Sans Institute. The SANS Security Policy Project.
http://www.sans.org/newlook/resources/policies/policies.htm

Authors

Author 1 : Lalli. M is working as a Assistant Professor in the Department of Computer Science & Engg., Bharathidasan University, Trichy, Tamil Nadu, India. She has 10 Years of experience in teaching. Her area of interest is Manets. Other areas of on interest include Information Security and Computer Networks and she is pursuing Ph.D in mobile Ad-hoc network security under the guidance of Dr. V. Palanisamy.

Author 2. Palanisamy.V is working as a Head of the Department of Computer Science & Engg Alagappa university. Karaikudi, Tamil Nadu, India. He has 20 Years of Teaching and 15 Years of experience in research. His area of Specialization is Wireless networks and network security. Other areas of an interest is include Algorithms.