Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 4, Issue. 6, June 2015, pg.170 – 176

RESEARCH ARTICLE

Artificial Immune System Based Statistical Model for Intrusion Identification

Garima Rathee[1]
Student, PDM College of Engineering, Bahadurgarh, Haryana
Parveen Bano[2]
Assistant Professor, PDM College of Engineering, Bahadurgarh, Haryana
Sugandha Singh[3]
Associate Professor & HOD(CSE Deptt.), PDM College of Engineering, Bahadurgarh, Haryana

Abstract: Intrusion Detection is one of the major security applications that can be applied in real time or using statistical analysis. In this paper, an AIS based statistical model is presented for intrusion identification. The work is here defined as a feature adaptive model to identify different kind of DOS attacks over the network. The paper has presented the work model for intrusion detection. The results show that the model has provided the identification of various network attacks.
Keywords: Intrusion, Attacks, AIS, Statistical

I. INTRODUCTION

Data mining offers great promise in helping organizations uncover patterns hidden in their data that can be used to predict the behavior of customers products and processes. However, data mining tools need to be guided by users who understand the business, the data, and the general nature of the analytical methods involved. Realistic expectations can yield rewarding results across a wide range of applications, from improving revenues to reducing costs. Building models is only one step in knowledge discovery. It's vital to properly collect and prepare the data, and to check the models against the real world. The "best" model is often found after building models of several different types, or by trying different technologies or algorithms. Choosing the right data mining products means finding a tool with good basic capabilities, an interface that matches the skill level of the people who'll be using it, and features relevant to your specific business problems.

Data mining is a process that uses a variety of data analysis tools to discover patterns and relationships in data that may be used to make valid predictions. The first and simplest analytical step in data mining is to describe the data — summarize its statistical attributes (such as means and standard deviations), visually review it using charts and graphs, and look for potentially meaningful links among variables (such as values that often occur together). As we know collecting, exploring and selecting the right data are critically important.

But data description alone cannot provide an action plan. We must build a predictive model based on patterns determined from known results, then test that model on results outside the original sample. A good model should

© 2015, IJCSMC All Rights Reserved

never be confused with reality (you know a road map isn't a perfect representation of the actual road), but it can be a useful guide to understanding your business.

The final step is to empirically verify the model. For example, from a database of customers who have already responded to a particular offer, you've built a model predicting which prospects are likeliest to respond to the same offer. Can you rely on this prediction? Send a mailing to a portion of the new list and see what results you get.

Data mining based IDS can efficiently identify these data of user interest and also predicts the results that can be utilized in the future. Data mining or knowledge discovery in databases has gained a great deal of attention in IT industry as well as in the society. Data mining has been involved to analyze the useful information from large volumes of data that are noisy, fuzzy and dynamic. Fig. 1 illustrates the overall architecture of IDS. It has been placed centrally to capture all the incoming packets that are transmitted over the network. Data are collected and send for pre-processing to remove the noise; irrelevant and missing attributes are replaced. Then the pre-processed data are analyzed and classified according to their severity measures. If the record is normal, then it does not require any more change or else it send for report generation to raise alarms. Based on the state of the data, alarms are raised to make the administrator to handle the situation in advance. The attack is modeled so as to enable the classification of network data. All the above process continues as soon as the transmission starts.

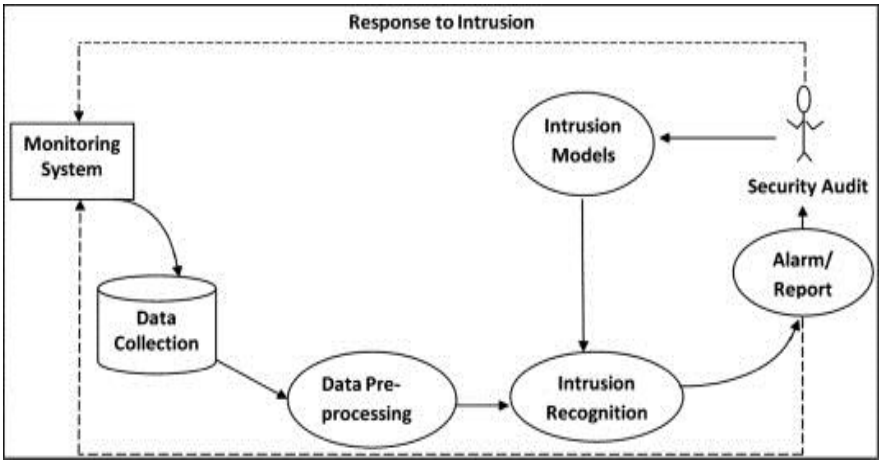

Figure1: Structure of Intrusion Detection System

II. LITERATURE SURVEY

In Year 2012, Justin M. Beaver performed a work, "A Learning System for Discriminating Variants of Malicious Network Traffic". This work describes a system that leverages machine learning to provide a network intrusion detection capability that analyzes behaviors in channels of communication between individual computers. With this approach, zero day detection is possible by focusing on similarity to known traffic types rather than mining for specific bit patterns or conditions. This also reduces the burden on organizations to account for all possible attack variant combinations through signatures.

In Year 2007, Vera Marinova-Boncheva performed a work, "Applying a Data Mining Method for Intrusion Detection". Widespread use of networked computers has made computer security a serious issue. Every networked computer, to varying degrees, is vulnerable to malicious computer attacks that can result in a range of security violations, such as, unauthorized user access to a system or the disruption of system services. In this paper Author would like to show how a data mining tool like XLMiner™ can also contribute to intrusion detection by extracting classification rules.

In Year 2010, Varun Chandola performed a work, "A Reference Based Analysis Framework for Analyzing System Call Traces". Reference based analysis (RBA) is a novel data mining tool for exploring a test data set with respect to a reference data set. The power of RBA lies in it ability to transform any complex data type, such as symbolic sequences and multivariate categorical data instances, into a multivariate continuous representation. Author demonstrates the application of the RBA framework in analyzing system call traces.

© 2015, IJCSMC All Rights Reserved

In Year 2010, Athira.M.Nambia performed a work, "Wireless Intrusion Detection Based on Different Clustering Approaches". In this paper, Author are finding optimital set of features from collected WLAN data using a Ranking Algorithm technique. Then with the aid of different data mining techniques such as K-Means, self organizing map and decision tree, these features are analyzed and the performance comparison is carried out.

In Year 2010, Rajeshwar Katipally performed a work, "Multistage Attack Detection System for Network Administrators Using Data Mining". In this paper, Author presents a method to discover, visualize, and predict behavior pattern of attackers in a network based system. Author proposed a system that is able to discover temporal pattern of intrusion which reveal behaviors of attackers using alerts generated by Intrusion Detection System (IDS). Author use data mining techniques to find the patterns of generated alerts by generating Association rules. Presented system is able to stream real time Snort alerts and predict intrusions based on Presented learned rules.

In Year 2010, Joseph R. Erskine performed a work, "Developing Cyberspace Data Understanding: Using CRISP-DM for Host-based IDS Feature Mining". This paper applies the Cross Industry Standard Process for Data Mining (CRISP-DM) to develop an understanding of a host environment under attack. This method of searching for hidden forensic evidence relationships enhances understanding of novel attacks and vulnerabilities, bolstering ones ability to defend the cyberspace domain. The methodology presented can be used to further host-based intrusion detection research.

In Year 2007, Tsong Song Hwang performed a work, "A Three-tier IDS via Data Mining Approach". Author introduced a three-tier architecture of intrusion detection system which consists of a blacklist, a whitelist and a multi-class support vector machine classifier. The first tier is the blacklist that will filter out the known attacks from the traffic and the whitelist identifies the normal traffics. Presented system has 94.71% intrusion detection rate and 93.52% diagnosis rate. Presented three-tier architecture design also provides the flexibility for the practical usage.

In Year 2006, Carrie Gates performed a work, "Challenging the Anomaly Detection Paradigm A provocative discussion". In this paper Author question the application of Denning's work to network based anomaly detection, along with other assumptions commonly made in network-based detection research. Author examine the assumptions underlying selected studies of network anomaly detection and discuss these assumptions in the context of the results from studies of network traffic patterns.

In Year 2011, ungsuk Song performed a work, "Statistical Analysis of Honeypot Data and Building of Kyoto 2006+ Dataset for NIDS Evaluation". In this paper, Author present a new evaluation dataset, called Kyoto 2006+, built on the 3 years of real traffic which are obtained from diverse types of honeypots. Kyoto 2006+ dataset will greatly contribute to IDS researchers in obtaining more practical, useful and accurate evaluation results. Author provide detailed analysis results of honeypot data and share Presented experiences so that security researchers are able to get insights into the trends of latest cyber attacks and the Internet situations.

In Year 2010, C.I. Ezeife performed a work, "NeuDetect: A Neural Network Data Mining Wireless Network Intrusion Detection System". This paper proposes NeuDetect, which applies a classification rule mining Neural Network technique to wireless network packets captured through hardware sensors for purposes of real time detection of anomalous packets. The proposed system, NeuDetect, solution approach is to find normal and anomalous patterns on pre-processed wireless packet records by comparing them with training data using Back-propagation algorithm.

In Year 2012, A.S. Aneetha performed a work, "Hybrid Network Intrusion Detection System Using Expert Rule Based Approach". In this paper Author have proposed a new frame work based on a hybrid intrusion detection system for known and unknown attacks in an efficient way. This frame work has the ability to detect intrusion in real time environment from the link layer. The detection rate of the hybrid system has been found to increase as the unknown attack percentage increases whereas in misuse, detection rate is found to decrease and in anomaly detection rate remains constant.

In Year 2008, Urko Zurutuza performed a work, "A Data Mining Approach for Analysis of Worm Activity Through Automatic Signature Generation". This paper proposes a novel framework to automatically discover and analyze traffic generated by computer worms and other anomalous behaviors that interact with a non-solicited traffic monitoring system. Network packets are analyzed by an Intrusion Detection System (IDS), and new signatures are generated clustering those which remain unknown for the IDS.

In Year 1999, Wenke Lee performed a work," Mining in a Data-flow Environment: Experience in Network Intrusion Detection". Author discuss the KDD process in "data-flow" environments, where unstructured and time dependent data can be processed into various levels of structured and semantically rich forms for analysis tasks. Author present procedures for analyzing frequent patterns from lower level data and constructing appropriate features to formulate higher level data. Author have applied Presented tools to the problem of building network intrusion detection models.

© 2015, IJCSMC All Rights Reserved

LTC Bruce D. Caulkins performed a work," A Dynamic Data Mining Technique for Intrusion Detection Systems". Author report the findings of Presented research in the area of anomaly-based intrusion detection systems using data-mining techniques described in section 3.3 to create a decision tree model of Presented network using the 1999 DARPA Intrusion Detection Evaluation data set. After the model was created, Author gathered more data from Presented local campus network and ran the new data through the model.

III. PROPOSED WORK

In this present work, a human cell based protected cell generation approach is defined to identify the intrusion detection or attack in public network. The work is about to perform the effective and reliable network communication in feature adaptive network. The work is here defined to identify the intrusion over the network and to provide the secure communication over the network. The network is here defined as the centralized system and the statistics is collected on server side. The server adaptive analysis is here performed to identify intrusion in the network.

A) Problem Definition

Security is always the major issue for any network defined in public or private domain. This kind of network suffers from intrusion attack applied by some internal or external nodes. There are number of approaches applied to analyze the network. These approaches are either applied on each network node on a centralized system. In this work, a statistical analysis based global approach is defined that can be applied on any centralized system to identify the intrusion probability. This predictive approach can be applied on a wireless network, web server, mobile server or cloud system. In this work, an intelligent artificial immune system based approach will be defined to identify the attack probability over the network. The work will not only identify the attack as well as classify it under symptom analysis such as DOS attack, man in middle attack etc. In this work, a danger theory based analysis approach is defined. This approach will work in same way as the tissue cells in human body work. According to this, cells generates a control body mechanism to analyze different molecules that comes in contact to the cell surface. These cells analyze the molecule under safe and danger module under integral sensing. The same kind of analysis will be here will be performed by the centralized system to identify the safe communication and the danger adaptive communication. Based on the analysis, the false positive communication that will be recognize as the intruder communication will be identified. The criticality ratio will also be formed under danger theory based on symptom analysis. The work will be applied on KDD network communication dataset. The work will be implemented in weka integrated java environment.

B) Research Methodology

The presented work is about to present a hybrid model to perform the intrusion detection on the network dataset. The presented model is defined in three layers given as

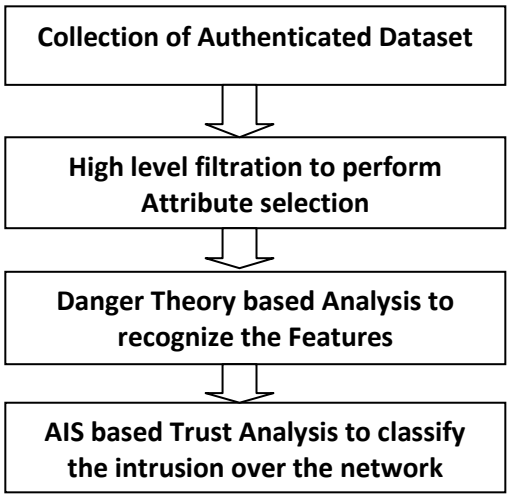

Figure2: Research Model

© 2015, IJCSMC All Rights Reserved

Module 1 : Collection of dataset
To process on the proposed mining operation we need some authenticated dataset so that the analysis can be performed. The analysis will be performed respective to the detection ratio. The work is about to obtain the high recognition ratio from the system. One of such dataset is KDD dataset
The KDD99 dataset is now the benchmark for training, testing and evaluating learning IDSs , so it is basic for IDS developers.

Module 2 : Filteration
From original database is converted to binary format by admin. Extension database is created by extending these binary format data. Finally original binary format and extended data are integrated to form integrated database. Fuzzy logic will be used for selecting items to identify the most appropriate attributes that can be used identify the intrusion over the network.

Module 3 : Danger Theory
Danger Theory is basically a filtration algorithm that will perform the feature based analysis on dataset. It perform the grouping of communicating data and will group them respective to the anomaly coefficient analysis. It is defined as the prediction model where the probabilistic and analytical decision will be taken regarding the group creation.

C) Cell Formation

The cell formation is here defined based on the statistical parameter analysis and the applying the threshold to categorize the relative values. This kind of cell formation based on the value level parameteric analysis. The statistical operations integrated with the work include mean value, standard deviation, MSE value analysis. Based on this, the validity of the dataset is obtained. If the obtained structural analysis returns the positive acceptance, the safe cell specification is obtained. If the cell structure specification is negative, the danger cell be considered.
The attribute adaptive weighted analysis is here considered to generate the relational analysis based on the attribute specification and generation. This kind of analysis also defined specific to the probabilistic and determistic analysis. This analysis is based on the constraints analysis integrated with statistical measures so that the attribute level generation is performed. This analysis is applied with threshold values to obtain the relative acceptance to the decision vector. Based on these vectors, the attack probability can be measured. The pattern set based cell formation is done for probabilistic analysis.

D) Bayesian Network

BNs are probabilistic graphical models that encode probabilistic dependence relations among variables. A Bayesian network, Bayes network, belief network or directed acyclic graphical model is a probabilistic graphical model that represents a set of random variables and their conditional dependencies via DAG. This classifier learns from training data the conditional probability of each attribute Ai given the class label C. Classification is then done by applying Bayes rule to compute the probability of C given the particular instances of A1…..An and then predicting the class with the highest posterior probability. The goal of classification is to correctly predict the value of a designated discrete class variable given a vector of predictors or attributes. The Bayesian network structure S is a directed acyclic graph (DAG) and the nodes in S are in one-to-one correspondence with the features X. The arcs represent casual influences among the features while the lack of possible arcs in S encodes conditional independencies.

IV. RESULT

The work is here defined in an easy and effective way. To present the work effectively, the graphical interface is designed. This interface is to process the training and testing dataset along with specification of work stages.
Here figure 3 is showing the instance based analysis obtained from the work. As the figure shows that the work has improved the recognition rate. The number of instances correctly recognized are higher than existing approach.

© 2015, IJCSMC All Rights Reserved

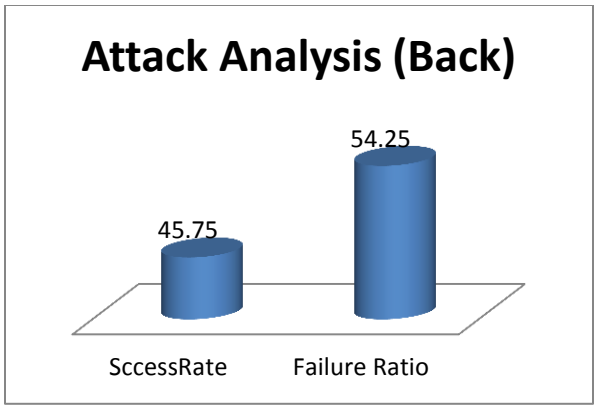

Figure3 : Instance Based Analysis (Overall)

Here figure 4 is showing the success ratio analysis obtained from the work. As the figure shows that the work has improved the recognition rate. The figure is showing results for Back attack. Here figure shows that the half of this attack are identified.

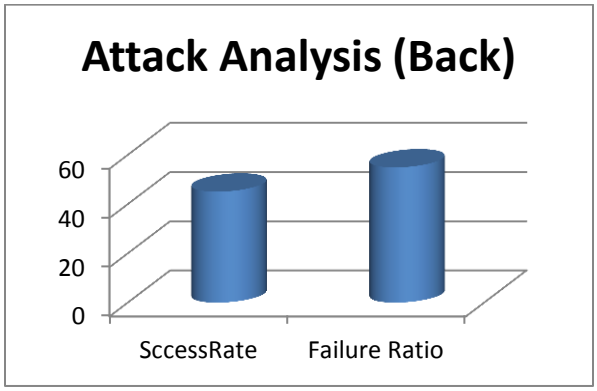

Figure 4: Attack Analysis (Back)

V. CONCLUSION

In this paper, an immune system based work model is presented to identify different kind of DOS attacks over the network. The work is here defined to process the network statistics to identify the intrusion. The implementation results show that the work has provided the clear identification of various kind of associated attacks.

REFERENCES

[1] Justin M. Beaver," A Learning System for Discriminating Variants of Malicious Network Traffic", Workshop, October 30–November 2, 2012, Oak Ridge, Tennessee, USA. ACM 1-58113-000-0/00/0010

[2] Vera Marinova-Boncheva," Applying a Data Mining Method for Intrusion Detection", International Conference on Computer Systems and Technologies - CompSysTech'07

[3] Varun Chandola, "A Reference Based Analysis Framework for Analyzing System Call Traces", CSIIRW '10, April 21-23, Oak Ridge, Tennessee, USA ACM 978-1-4503-0017-9

[4] Athira.M.Nambia, "Wireless Intrusion Detection Based on Different Clustering Approaches", A2CWiC 2010, September 16-17, 2010, India 978-1-4503-0194-7/10/0009

[5] Rajeshwar Katipally, "Multistage Attack Detection System for Network Administrators Using Data Mining", CSIIRW '10, April 21-23, Oak Ridge, Tennessee, USA ACM 978-1-4503-0017-9

[6] Joseph R. Erskine, "Developing Cyberspace Data Understanding: Using CRISP-DM for Host-based IDS Feature Mining", CSIIRW '10, April 21-23, Oak Ridge, Tennessee, USA, ACM 978-1-4503-0017-9 .

© 2015, IJCSMC All Rights Reserved

[7] Tsong Song Hwang, "A Three-tier IDS via Data Mining Approach", MineNet'07, June 12, 2007, San Diego, California, USA. ACM 918-1-59593-792-6/07/0006

[8] Carrie Gates, "Challenging the Anomaly Detection Paradigm A provocative discussion", NSPW 2006, September 19-22, 2006, Schloss Dagstuhl, Germany. ACM 978-1-59593-857-2/07/0007

[9] Jungsuk Song, "Statistical Analysis of Honeypot Data and Building of Kyoto 2006+ Dataset for NIDS Evaluation", BADGERS '11 April 10-13, 2011, Salzburg.

[10] C.I. Ezeife, "NeuDetect: A Neural Network Data Mining Wireless Network Intrusion Detection System", IDEAS10 2010, August 16-18, Montreal, QC [Canada]; ACM 978-1-60558-900-8/10/08

[11] A.S. Aneetha, "Hybrid Network Intrusion Detection System Using Expert Rule Based Approach", CCSEIT-12, October 26-28, 2012, Coimbatore [Tamil nadu, India] ACM 978-1-4503-1310-0/12/10

[12] Urko Zurutuza, "A Data Mining Approach for Analysis of Worm Activity Through Automatic Signature Generation", AISec'08, October 27, 2008, Alexandria, Virginia, USA. ACM 978-1-60558-291-7/08/10

[13] Wenke Lee," Mining in a Data-flow Environment: Experience in Network Intrusion Detection", KDD-99 San Diego CA USA ACM 1999 l-581 13-143-7/99/08

[14] LTC Bruce D. Caulkins," A Dynamic Data Mining Techniquefor Intrusion Detection Systems".

© 2015, IJCSMC All Rights Reserved