Anomaly Intrusion Detection in Computer Network using FCM clustering

Abhilasha. A. Sayar [1] , Sunil. N. Pawar[2] , Vrushali Mane[3]
[1,2,3.] Electronics Dept. & BAMU,
Jawaharlan Nehru Engg. College,
Aurangabad,Maharashtra

Abstract: **- Internet is World Wide Web, which is covering each and every part of world. With tremendous growth in internet world is coming close to an individual identity but at the same time creates a threat of being pirated. Number of software applications are being developed and brought in market. Along with this development there is a risk in computer network system of being robed. Government sectors, military, companies, institutions and many more sectors working over network have daily transaction and sharing of data. If due to some abnormal activities data which is to be transferred is pirated then it may result in serious matter. Thus securing network is at most priority in today world. For securing network system a tool came into picture, and that is Intrusion Detection System.**

Keywords: - ***Anomaly detection; fuzzy logic; clustering.***

I. INTRODUCTION

Network security has been an issue since computers have been networked together. The Evolution of the Internet has increased the need for security systems. With the growth of the Internet and its potential, there has been subsequent change in business model of corporate sector across the world. More and more people are getting connected to the Internet every day to take advantage of e-Business. Internet connectivity at work has therefore become a very critical aspect of today's e-business. Through internet we can have things to be done in fraction of second; files can be uploaded or downloaded at great speed. Thus internet which is giving all good things can also have nuisance of being hacked. Data which is vital must be protected from these attacks. An intrusion can be defined as ‘‘an act of a person of proxy attempting to break into or misuse a system in violation of an established policy resulting into compromise of target system’’ [1]. So to protect systems from intruders, intrusion detection system is needed. IDS is software and/or hardware System for monitoring and detecting data traffic or user behavior to identify attempts of illegitimate accessing system manipulation through a network by malware and/or intruders (crackers, or disgruntled employees). Hackers and nasty users can get access to company’s internal systems in various reasons. These are,

- Software bugs called vulnerabilities
- Lapse in administration
- Leaving systems to default configuration

The malicious users use different techniques like Password cracking, sniffing unencrypted or clear text traffic etc. to exploit the system vulnerabilities Therefore, there needs to be some kind of security to the organization’s private resources from the Internet as well as from inside users.

Different organizations across the world deploy firewalls to protect their private network from the Public network. But, when it comes to securing a Private network from the Internet using firewalls, no network can be hundred percent secured. This is because; the business requires some kind of access to be granted on the internal systems to Internet users.

II. INTRUSION DETECTION SYSTEM

Intrusion Detection concept was introduced by **James Anderson** in 1980**,** defined an “Intrusion attempt or threat to be potential possibility of a deliberate unauthorized attempt to access information, manipulate information, or render a system unreliable or unusable”[2].A general IDS architecture is based on the consideration of four types of functional modules (Fig. 1):

- E blocks (‘‘Event-boxes’’): This kind of block is composed of sensor elements that monitor the target system, thus acquiring information events to be analyzed by other blocks.
- D blocks (‘‘Database-boxes’’): These are elements intended to store information from E blocks for subsequent processing by A and R boxes.
- A blocks (‘‘Analysis-boxes’’): Processing modules for analyzing events and detecting potential hostile behavior, so that some kind of alarm will be generated if necessary
- R blocks (‘‘Response-boxes’’): The main function of this type of block is the execution, if any intrusion occurs, of a response to thwart the detected menace. [3]

  In other words intrusion detection system consists of information source, analysis engine and decision model. [4]

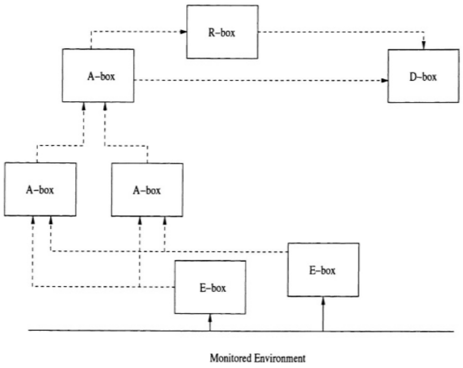


Fig.1. Architecture of IDS

Intrusion detection can be classified as anomaly intrusion detection and misuse detection based on the approach that is selected. Misuse based detection system are having database of information of attacks that occurred, it compares the audit data against stored data, if there is no match then intrusion occurred and if matched then a legal access has been occurred. [5]

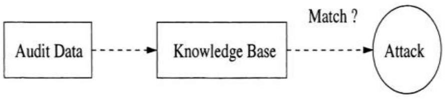


Fig. 2 Block Diagram of a misuse IDS

On the other hand anomaly intrusion detection detects all attack that are totally unknown to the system. Here the system first makes the system model and then checks for abnormal activities.[5]

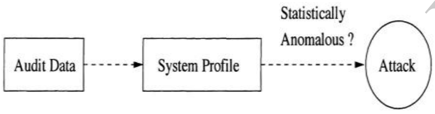


Fig.3. Block Diagram of a anomaly IDS

III. TECHNIQUES USED IN IDS

As attacks are every so often increasing with growth in internet many techniques are in practice to detect intrusion. These techniques mainly differ in way of implementation, algorithm, their working and many more factors. Techniques are designed on basis of certain rules, computational approach, statistical methods. [6]

Generally company's or even individual deal with bulky volume of data, and thus working with this data, classifying it in labels become a tedious job. This detection falls under category of misuse detection which works only on known attacks. With progress of internet data can be transmitted to and fro at high-speed, along with this data there might be possibility of abnormal activities. New types of attacks are launched now and then. So with misuse detection it is very intricate to get rid of these attacks. In order to come up with these difficulties we need to develop a system that deals with unlabelled data and is able to detect unknown attacks. Thus a system that handles such unknown attacks or unusual activities is nothing but anomaly intrusion detection system. Here unknown attack simply means the data which is deviating from normal behavior.

Under such assumption we built an anomaly intrusion detection system that detects malicious activities which are totally unknown by using fuzzy clustering algorithm.

IV. FUZZY LOGIC BASED IDS

Many methods are designed to detect suspicious activities, most of these method use neural network, data mining, genetic algorithm tools etc. Beside this there is machine learning algorithm that lay down a path to detect intrusions. Machine learning algorithm is basically of two types; they are supervised method and unsupervised methods. Main goal of supervised detection is to build a model that differentiates incoming labeled data. Some of supervised method use techniques of neural network, genetic algorithm, support vector machine among others. [6]

On other hand unsupervised method also called as data clustering take a different approach by grouping unlabelled data into clusters based on similarities. Data clustering helps one to find similarities in data and putting similar data into groups [7]. Denial of Service attacks (DoS), Distributed DoS (DDoS), network/host scans, and spreading worms or viruses are examples of the different attacks that daily threaten the integrity and normal operation of the network. The principal challenge in automatically detecting and analyzing network attacks is that these are a moving and ever-growing target. Unsupervised learning is a recent approach in knowledge exploration. It is widely used on/with unlabeled data, such as extracting relevance that exists in records.

A. FCM CLUSTERING

Clustering is the primary task of data mining. As years passed data mining field progressed and number of clustering methods aroused. Clustering is basically of two types hard i.e. crisp clustering and soft i.e. fuzzy clustering. A characteristic of the crisp clustering method is that the boundary between clusters is fully defined. But in many cases, the boundaries between clusters cannot be clearly defined. Some patterns may belong to more than one Cluster. In such cases, the fuzzy clustering method provides a better and more useful method to classify these patterns where the data may either belong to one cluster or other. [8]

The FCM algorithm is one of the most widely used fuzzy clustering algorithms which allows one piece of data to belong to two or more clusters. This technique was originally introduced by Professor Jim Bezdek in 1981. The FCM algorithm attempts to partition a finite collection of elements $X=\{x_1, x_2, ..., x_n\}$ into a collection of c fuzzy clusters with respect to some given criterion. It is based on minimization of the following objective function [9]:

$$J_m = \sum_{i=1}^{N}\sum_{j=1}^{C} u_{ij}^{m}\left\|x_i - c_j\right\|^2 \quad , \quad 1 \le m < \infty$$

where *m* is any real number greater than 1, u_{ij} is the degree of membership of x_i in the cluster *j*, x_i is the *i*th of d-dimensional measured data, c_j is the d-dimension center of the cluster. Fuzzy partitioning is carried out through an iterative optimization of the objective function shown above, with the update of membership u_{ij} and the cluster centers c_j by:

$$u_{ij} = \frac{1}{\sum_{k=1}^{C}\left(\frac{\left\|x_i - c_j\right\|}{\left\|x_i - c_k\right\|}\right)^{\frac{2}{m-1}}} \quad ,$$

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^{m} \cdot x_i}{\sum_{i=1}^{N} u_{ij}^{m}}$$

B. STEPS FOLLOWED IN PROPOSED SYSTEM

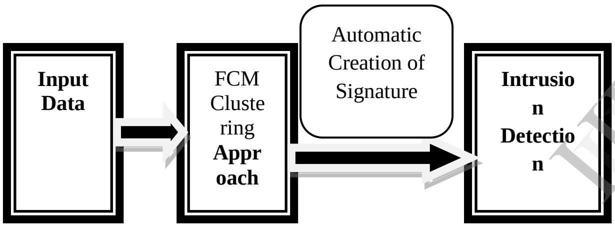


Fig: 4 Organization of the system

The steps to develop proposed system are as follow:-[10]

1. Create log file.

2. Find out maximum data flow in current log file.

3. Apply sliding time windowing scheme for every 1sec.

4. Aggregation process for traffic flow

5. Creation of feature space matrix by using following formula:-

X(1)=[sip dip sp dp nsip/ndip y(1)/ndip]

II^{ly} we have to create feature space matrices for all time windows data set.

i.e., X= € (x1, x2…….xn). And then apply clustering algorithm and declare smallest group of cluster as outlier

6. Detect outlier using outlier detection algorithm.

7. Trace back outlier in feature space matrix, aggregation and log file.

8. Use trace data to Create signature for anomalous flow.

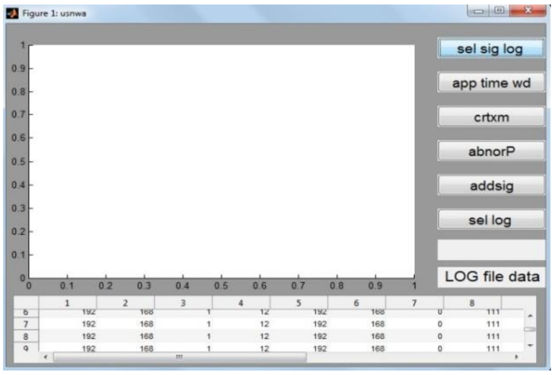
9. Signature will be logged and updated the signature table.

10. Signature table can be use for online detection anomalous flow.

V. OBSERVATIONS

The GUI for the system is as shown below.

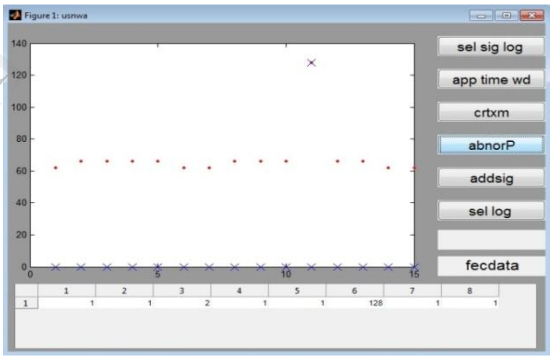
Step 1:-



	1	2	3	4	5	6	7	8
6	192	168	1	12	192	168	0	111
7	192	168	1	12	192	168	0	111
8	192	168	1	12	192	168	0	111
9	192	168	1	12	192	168	0	111

Fig: 5 Snapshot 1

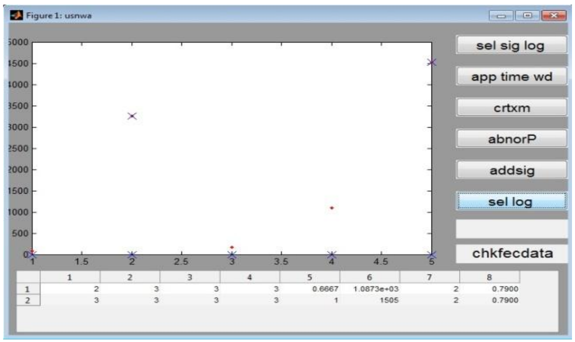
Step: 5 In this step, the abnorP button is click to find out the abnormalities.



	1	2	3	4	5	6	7	8
1	1	1	2	1	1	128	1	1

Fig: 6 Snapshot2

Step: 10



	1	2	3	4	5	6	7	8
1	2	3	3	3	0.6667	1.0873e+03	2	0.7900
2	3	3	3	3	1	1505	2	0.7900

Fig: 7 Snapshot 3

A. RESULTS OF ITERATIONS

Iteration count = 1, obj. fcn = 907261.574555

Iteration count = 2, obj. fcn = 283224.819661

Iteration count = 3, obj. fcn = 221392.863361

Iteration count = 4, obj. fcn = 215563.589325

Iteration count = 5, obj. fcn = 214587.370868

Iteration count = 6, obj. fcn = 214404.017594

Iteration count = 7, obj. fcn = 214368.316248

Iteration count = 8, obj. fcn = 214361.263493

Iteration count = 9, obj. fcn = 214359.861555

Iteration count = 10, obj. fcn = 214359.582117

Thus with observations obtained by running FCM algorithm, it is seen that Objective function is reduced.

VI. CONCLUSION

The Unsupervised Network Anomaly Detection Algorithm that we have proposed presents many interesting advantages w. r. t. previous proposals in the field unsupervised anomaly detection. It uses exclusively unlabeled data to detect traffic anomalies, without assuming any particular model or any canonical data distribution, and without using signatures of anomalies or training. We have verified the effectiveness of UNADA to detect real single source-destination and distributed network attacks in real traffic traces from different networks, all in a completely blind fashion, without assuming any particular traffic model, clustering parameters, or even clusters structure beyond a basic definition of what an anomaly is.

VII . REFERENCES

[1] Khattab M. Alheeti Alanbar University Iraq, “Intrusion Detection System and Artificial Intelligent”.

[2] Shaik Akbar Assoc. Profr, Dept. of C.S.E, SVIET, Dr.K.Nageswara Rao Prof & H.O.D, Dept. of C.S.E P.V.P.S.I.T, India, Dr.J.A.Chandulal Prof, Dept. of C.S.E GITAM University, “Intrusion Detection System Methodologies Based on Data Analysis” International Journal of Computer Applications (0975 – 8887) Volume 5– No.2, August 2010.

[3] P. Garcı´a-Teodoro [a],*, J. Dı´az-Verdejo [a], G. Macia´-Ferna´ndez [a], E. Va´zquezb, “ Anomaly-based network intrusion detection: Techniques, systems and challenges”.

[4] J.T. Yao S.L. Zhao L. V. Saxton Department of Computer Science University of Regina, “A study on fuzzy intrusion detection”.

[5] Shingo Mabu, *Member, IEEE*, Ci Chen, Nannan Lu, Kaoru Shimada, and Kotaro Hirasawa, *Member, IEEE*, “ Technical Correspondence: An Intrusion-Detection Model Based on Fuzzy Class-Association-Rule Mining Using Genetic Network Programming”.

[6] Shaik Akbar Assoc. Profr, Dept. of C.S.E, SVIET, Dr.K.Nageswara Rao Prof & H.O.D, Dept. of C.S.E P.V.P.S.I.T, India, Dr.J.A.Chandulal Prof, Dept. of C.S.E GITAM University, “Intrusion Detection System Methodologies Based on Data Analysis” International Journal of Computer Applications (0975 – 8887) Volume 5– No.2, August 2010.

[7] Veronica.S.Moertini, “Introduction to five data clustering algorithms”.

[8] Mitchell D’silva, Deepali Vora M.E. Information Technology, Vidyalankar Institute of Technology, Mumbai, India. “Comparative Study of Data Mining Techniques to Enhance Intrusion Detection”.

[9] Linquan Xie1, Ying Wang1,2, Liping Chen2, and Guangxue Yue, Jiangxi University of Science and Technology, Jiaxing, China, “An Anomaly Detection Method Based on Fuzzy C-means Clustering Algorithm”. Proceedings of the Second International Symposium on Networking and Network Security (ISNNS ’10) Jinggangshan, P. R. China, 2-4, April. 2010, pp. 089-092.

[10] Pedro Casas, Johan Mazel and Philippe Owezarski CNRS; LAAS; 7 avenue du colonel Roche, F-31077 Toulouse, France “UNADA: Unsupervised Network Anomaly Detection using Sub-Space Outliers Ranking”.