ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

Network Level Anomaly Detection System with Principal Component Analysis

Ranjita Patil, Dr. V. R. Ghorpade

Student, Dept. of CSE, D Y Patil College of Engineering and Technology, Maharashtra, India

Professor, Dept. of CSE, D Y Patil College of Engineering and Technology, Maharashtra, India

ABSTRACT: The anomaly based approaches are not dependent on an existing knowledge base. These techniques detect the malicious behavior and identify them based on the network traffic patterns. Possible network threats are identified by finding deviations from normal behavior of network. This detection does not end the task but it's the prerequisite to mitigate the impact of malicious activities. The localization of identified malicious activity is important to take the control actions. However localization of malicious behavior in anomaly detection approach is yet at the starting phase of research. In the developed system, network traffic based anomaly detection is done to achieve homeostasis. Homeostasis is biological metaphor in which system is able to sustain the stable condition by taking control measures upon the observed changes in the environment. To maintain the stability of network, detection and localization of malicious activities in network traffic is important. The Principal Component Analysis (PCA) is best known anomaly detection method used in the system. PCA uses the principal components score to detect the anomalies.

KEYWORDS: Anomaly, PCA, Homeostasis, Anomaly localization, Network Traffic Features

I. INTRODUCTION

Network security system is of paramount importance in the present communication environment. The computers may be located at same location or at different distant locations which forms the net like structure, which is base for communication and many more services. While communicating every computer in network has an increasing number of security threats. The information and services available on network is an intellectual asset. These intellectual assets need to be safe and made available whenever authorized user request for it. With new types of attacks appearing continually, developing manageable and suitable network security approaches (IDS-Intrusion Detection System) is a severe challenge. Based on the way by which analysis is carried out, intrusion detection systems can be said either as signature-based system or anomaly-based system. Signature and anomaly-based systems are comparable in terms of abstract operation and composition. Signature-based concepts are widely used in variety of security software for example anti-virus software. These approaches fail when any new activities, not having known signature, takes place at network traffic.

Anomaly-based system detects any new activity which shows deviation from normal known behavior. Developed system uses one of the techniques of anomaly based intrusion detection system and localize the anomaly. The system is inspired from the biological course of action Homeostasis. *Homeostasis* is a biological metaphor able to retain a stable condition of organism by taking control actions on the incessant changes in their internal and external environments. Using this concept the network will be able to detect the malicious anomalies (change in external environment) and to generate the response (such as traffic rate limiting, node quarantine) to maintain the stable status of system. Anomaly based detection technology has the high rate of false alarms. An attempt is being made to enhance the ability of developed anomaly-based technology by integrating the score based classification with it. This detection does not end the task but it's a start of taking steps towards the mitigation of impact of malicious anomaly and to generate the signature for future use.

In computer networking a connection provides connectivity between computer and the Internet, a network, or another computer. Under some well defined protocol data flows to and from one node to another node in computer

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

network. Each connection is observed as to label it as normal or as an anomaly. Thus to observe the connections computer traffic must be monitored. There are many other benefits of network traffic monitoring:

- Helps to avoid bandwidth and server performance bottlenecks.
- Helps to deliver better quality of service to user.

The network traffic is observed and certain control actions are taken. Where, network control is the procedure of administration, controlling or reducing the network traffic. It is used by network administrator to diminish overcrowding, latency and packet loss. The control actions in network are carried with the help of connecting devices as switch and router.

II. RELATED WORK

For the network security process which can detect the anomalies, it is required to keep watch on the network traffic of system. This can be achieved by packet capturing. Keita Fujii in [1]; [2][3] gives the details of JPCAP. The JPCAP is a free JAVA library used to capture and send network packets. Using JPCAP, developer can develop application to capture packets from network interface and examine them in java to gain different patterns of the network traffic. Using Jpcap, developers develop an application for computer network to capture network packets at selected network interface. In data packets capture facet, Jpcap is a cross-platform, without any charge, wonderful packet capture structure. Using Jpcap, developer may program an application that have the function of data packet capture and protocol analysis. [4]

In [5],[6], PCA algorithm is used for pre-treatment of the statistical network traffic features which is able to refine the results of traffic identification; PCA algorithm is imported into the pretreatment process, some isolated secondary factors in statistical features are removed. The identification of the network traffic is important in the network management of Internet Service Providers. This technology can analyze the flow in the network especially in the backbone to optimize the network structure and can block unwanted traffic packets for some security reason. This work can be extended to detect the anomaly in network by investigating the relationship between the network features.

In [7] two approaches of the anomaly detection are stated. It considers an outlier as an observation that is numerically distant from the rest of the data. The anomaly detection model trained using unlabeled data that consist of both normal as well as attack traffic and the model trained using only normal data and a profile of normal activity is created.

Step wise anomaly detection technique is proposed by George Nychis [10]. The first step is to preprocess the input. In the second step, statistical analysis and data transforms are implemented to part network behaviors either as normal behavior or anomalous behaviors and noise. A vast range of techniques can be applied for data analysis, e.g., Wavelet Analysis, Covariance Matrix analysis, and Principal Component Analysis. In the final step, decision theories such as Generalized Likelihood Ratio (GLR) are suggested to check the deviation within network behavior.

Anomaly detection technique detects abnormal behavior that has significant deviations from a pre-established normal profile. The advantage of anomaly detection techniques is that they do not require known attack signature and can thus detect novel attack. Principal component analysis (PCA) is a powerful technique for analyzing and identifying patterns in data. It finds the most important axis to express the scattering of data. By using PCA, the first principal component (PC) is calculated, which reflects the approximate distribution of data. Thus Mohammad Ahmadi Livani and Mahdi Abadli [6] presented PCA-based centralized approach, called PCACID.

III. SYSTEM ARCHITECTURE

A necessity of secure computer network is increasing rapidly. Hence secure network intrusion detection systems are required to be designed. The developed system is anomaly based intrusion detection system. In this section spotlight is on the architecture of system. System Architecture is the conceptual model which defines the structure and behavior of the system. It is a formal description and representation of a system. Let's focus on the system components and their relationships. The developed system focuses on capturing and analyzing the network traffic.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

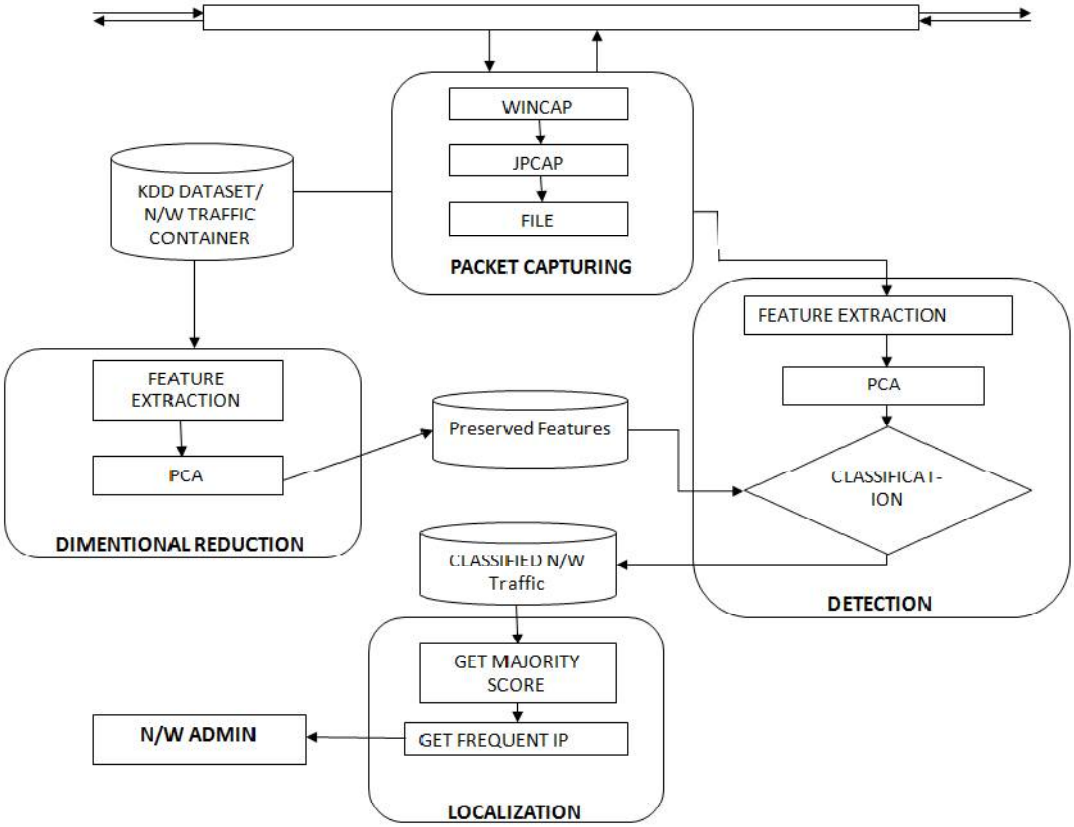

Fig. 1. System Architecture

The input dataset for the system comprise of the KDD dataset or file of captured network traffic. The outline of the developed system is as follows:-

1. Input Dataset is obtained from the KDD dataset (KDD is free dataset available for researchers to work on network security) or system can capture the own network traffic.
2. Data cleansing can be done manually form KDD dataset
3. Network traffic capturing is carried out by WINPCAP and JPCAP.
4. WINPCAP accept the network traffic at lower level i.e. at network adaptor
5. JPCAP java library used for network traffic analysis.
6. 41 Network features such as source IP, destination IP, source port, destination port etc. are available in KDD dataset.
7. Principal component algorithm (PCA) is used to find the K principal components out of 41 features.
8. Score for anomaly is based on K principal components. This score is preserved.
9. Score of captured traffic is calculated (Based on K principal components).
10. With reference to preserved data, the classification of network traffic is done.
11. To localize the anomaly in network, frequent source IP for anomaly is detected.

IV. NETWORK LEVEL ANOMALY DETECTION

Anomaly detection helps to detect the new abnormal behavior in the network traffic. Unknown activities at network are detected by noting the deviation of network traffic. Any abnormal behavior results in changes in the network traffic patterns. Thus to analyze the network traffic, packet capturing is first step.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

A. Packet Capturing

First step required for packet capturing is to acquire the record of network interfaces on the machine. To do so, JPCAP provides JpcapCapator.getDeviceList() method. It returns an array of NetworkInteface objects. A NetworkInterface object holds some information of corresponding network interface. Network interface provides information such as name, description, IP and MAC addresses, and datatlink name and description. Once user select the network interface from the obtained list, then the JpcapCaptor.openDevice() method is used to open the interface. To capture the packets WinPcap and JPcap is required by the system. As WinPcap is base for JPcap to capture the packets at network, the well configured WinPcap is to be installed by the system.

B. Feature Extraction

Feature extraction is done by using the jpcap and jpcap.packet library in JAVA. It has number of class files useful for packet handling, in JAVA program. The *jpcap* package is used to find the network connections available at system. It has class files, some of these are as follows:

a) JpcapCaptor.class
b) NetworkInterface.class ,etc.

The package jpcap.packet is used for extracting the features of packet. It has number of class files as follows:

a) ARPPacket.class
b) DataLinkPacket.class
c) EthernetPacket.class
d) ICMPPacket.class
e) IPPacket.class

All the packet details extracted using above mentioned classes are stored in the log files for further work. Each log file contains details of the network traffic. This log file is source for analyzing module to generate the statistical features and network connection features.

C. Anomaly Detection and localization

The components that show large variation and explain the major cumulative proportion of the total sample are known as principal components. These principal components have a tendency to be sturdily relating the features that have relatively large variances and covariances. Consequently the observations that show deviation based on the first few components are marked as outliers. PCA has been applied to the intrusion detection problem as a data reduction technique. Anomalies are qualitatively not like the normal instances i.e. Anomalies show great deviation from the recognized normal patterns and thus can be marked as attacks. No effort is made to make a distinction of different types of attacks. To set up a detection algorithm system perform PCA on correlation matrix of the normal group. The correlation matrix is used because each network feature is measured in diverse scales. The training data must not contain any outliers when it is used for determining the detection criterion.

The network traffic data, based on which the anomaly detection system works is the in and out flow within network nodes. Thus due to the vastness of network features, it is difficult to identify the source of anomalous activity. In this case localization of IP means to find the IP of node which is source for the anomalous activity. In earlier work system has detected the anomaly and each network packet is labeled as normal/anomaly. The present module localizes the IP with the help of heavy hitter algorithm. In heavy hitter algorithm the frequency by which the specific IP is source for malicious packet is measured. If the frequency of IP in malicious packets exceeds than that of the prefixed threshold then that IP is considered as the source for malicious activity.

V. RESULT ANALYSIS

An experiment was conducted with different size datasets. For result analysis when actual results are compared with the expected results, it is observed that as the data set size increases the actual results move close to expected results. In Fig. 2 this picture can be seen clearly. In TABLE 1; TP, TN, FP, and FN values are shown.

These records are used to measure the sensitivity, specificity and accuracy of the system

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

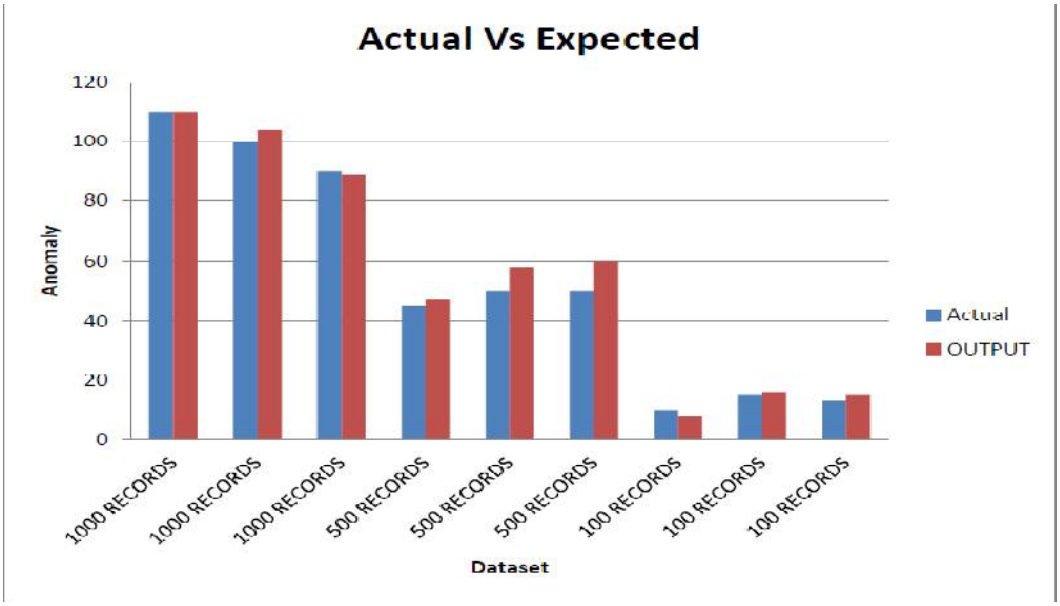

Fig. 2. Graph for results of experiment

In order to calculate the sensitivity, specificity and accuracy of the system it is necessary to evaluate the following four values in accordance with anomaly detection system.

True Positive (TP): There is a possibility of anomaly and system identifies anomaly correctly.
False Positive (FP): There is no any possibility of the anomaly and system identifies an anomaly.
True Negative (TN): There is no any possibility of anomaly and the system does not identify anomaly.
False Negative (FN): There is a possibility of anomaly and the system does not identify anomaly.

TABLE 1 : TP, TN, FP and FN values obtained form system results

Sr. no.	DATASET	(TP)	(FP)	(FN)	(TN)
1	1000 RECORDS	104	6	6	884
2	1000 RECORDS	97	7	3	893
3	1000 RECORDS	87	2	3	908
4	500 RECORDS	43	5	2	450
5	500 RECORDS	48	10	2	440
6	500 RECORDS	47	13	3	437
7	100 RECORDS	7	1	3	89
8	100 RECORDS	14	2	1	83
9	100 RECORDS	11	4	2	83

$$Sensitivity = \frac{Number\ of\ True\ Positives}{Number\ of\ True\ Positives + Number\ of\ False\ Negatives}$$

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

$$Specificity = \frac{Number\ of\ True\ Negatives}{Number\ of\ True\ Negatives + Number\ of\ False\ Positives}$$

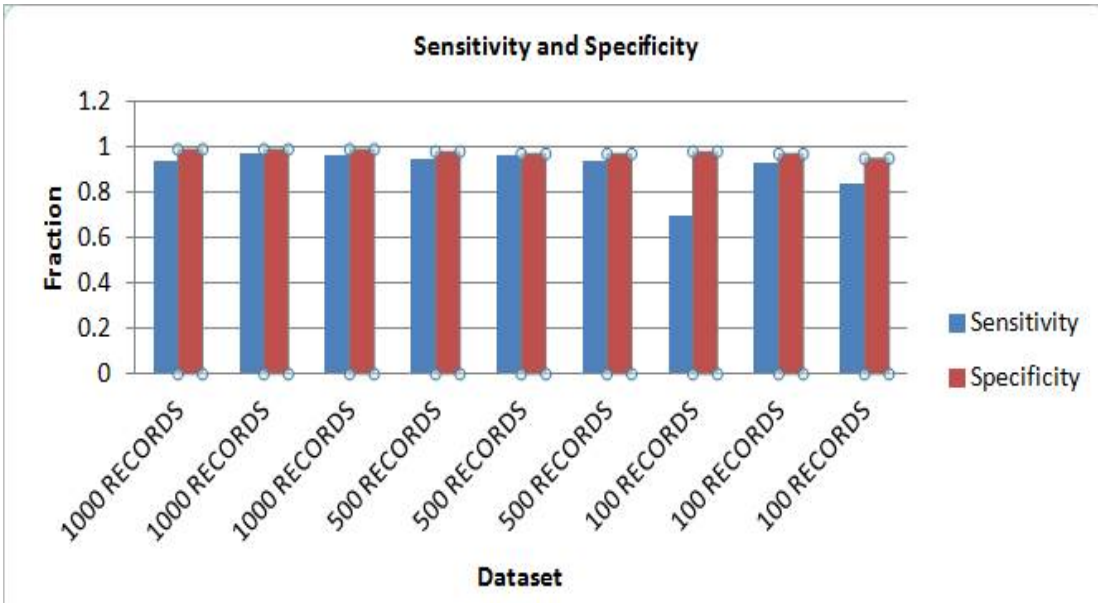

Fig. 3. Graph shows sensitivity and specificity

Similarly Accuracy of the system is defined by the following formula:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

.

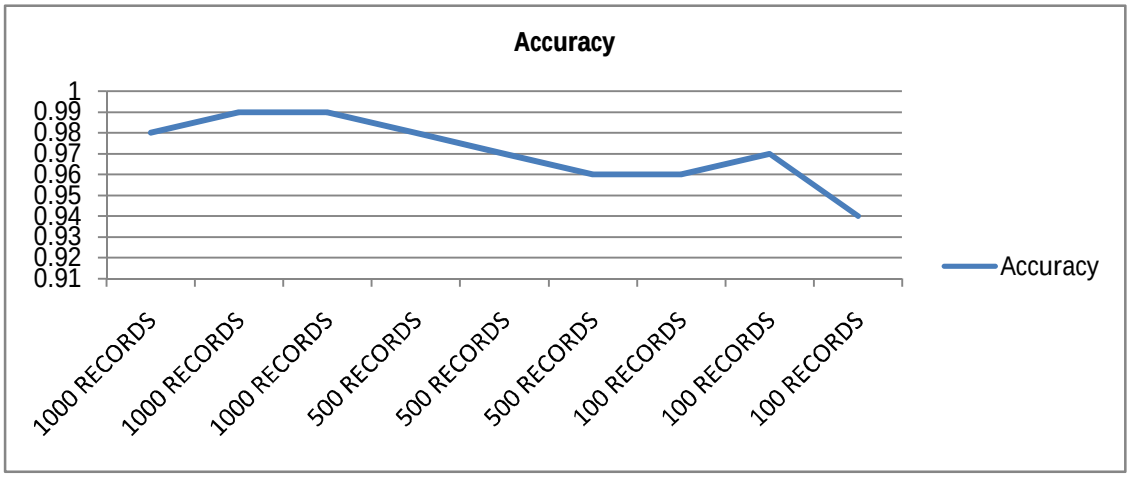

Fig. 4. Accuracy

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

It has been observed that the accuracy of system is affected by the size of dataset in Fig. 4. Large dataset helps to increase the system accuracy. This relation between accuracy and size of dataset is easily observed in *Accuracy graph*

VI. CONCLUSION

The rapid increase in the network attacks is a major concern. The loss is estimated to be in disruption of services to an authorized user. Novel techniques of anomaly based intrusion detection are helping to detect malicious activities at network. But need is to localize the source of these malicious activities. The developed system proves to be effectual in monitoring the network traffic. The system effectively reduces the high dimensional data to low dimensional data. Dimension reduction helps to decrease the complexity of network traffic data. System performs the score based detection of anomaly with moderate accuracy and localizes the anomaly. The developed system efficiently monitors the network traffic and aid the localization of source of malicious activities. But the ever changing nature of the network traffic patterns suggests many challenges in this field. However localization helps network admin to take mitigation actions. In future there is a scope to add robustness to network devices so that these devices will be capable to take mitigation actions without human intervention.

REFERENCES

[1] Ruoyi Jiang, Hongliang Fei and Jun Huan, "A Family of Joint Sparse PCA Algorithm for Anomaly Localization in Network Data Streams", *IEEE transactions on Knowledge and data Engineering*, vol 25, November 2013.
[2] Adetunmbi A.Olusola., Adeola S.Oladele. and Daramola O.Abosede, "Network Data Packet Capture and Protocol Analysis on Jpcap-Based", *Proceedings of the World Congress on Engineering and Computer Science 2010 Vol I WCECS 2010*, October 20-22, 2010, San Francisco, USA.
[3] Navneet Kaur Dhillon and Mrs. Uzma Ansari, "Enterprise Network Traffic Monitoring, Analysis, and Reporting Using WINPCAP Tool With JPCAP API", *International Journal of Advanced Research in Computer Science and Software Engineering 2 (11)*, pp. 95-101, November- 2012.
[4] Padmini Rathore, and Nitin Jain, "JPCAP, WINPCAP Approach For Intrusion Detection System", *International Journal of Engineering Research & Technology (IJERT)*, Vol. 2 Issue 7, July – 2013.
[5] Yuh-Jye Lee, Yi-Ren Yeh and Yu Chiang Frank Wang, "Anomaly Detection via Online Oversampling Principal Component Analysis", *IEEE transactions on Knowledge and data Engineering*, vol 25, July 2013.
[6] Mohammad Ahmadi Livani and Mahdi Abadi, "A PCA-based Distribution Approach for Inrusion Detection in Wireless Sensor Network",*IEEE International Symposium on Computer Networks and Distributed Systems Feb 23-24 2011*.
[7] http://math.nist.gov/javanumerics/jama/.
[8] http://www.jpcap.sourceforge.net/javadoc.
[9] http://kdd.ics.uci.edu/database/kddcup99/.