Hybrid Anomaly Intrusion Detection System Using K-Means and SVM Gaussian kernel

Prof. P. M. Pondkule, Prof. O.C. Nilakhe, Dr. A. D. Bhagwat, Prof. A. D. Homkar

Assistant Professor, Dnyanshree Institute of Engineering & Technology, Sajjangad, Satara, Maharashtra, India

Abstract **Intrusions detection systems (IDSs) are systems that try to detect attacks whenever they occur. IDSs assembles network traffic information from some point on the network or computer system, which is used to protect the network from intruder. Intrusion Detection Systems are either Misuse-Detection based or Anomaly Detection Based. Misuse-Detection based IDSs the IDS analyzes the information it gathers and compares it to large databases of attack signature whereas anomaly detection based IDSs can also detect new attacks by using heuristic methods. In this paper, a hybrid anomaly intrusion detection system is proposed by using K-means for clustering of packet and SVM Gaussian Kernel for the cost efficient classification of packets.**

***Keywords* —IDS, SIEM, NIDS, HIDS.**

I. Introduction

Nowadays with the excess use of Internet and online procedures requesting a secure channel, it has become an basic requirement to secure the network. There are various hazardous sources including software virus mostly as the operating systems and software used becomes more functional[1]. Intruders who do not have rights to access these data can lift important and private information belonging to network users. Firewalls are hardware or software systems placed in between two or more computer networks to stop the attacks, by dividing these networks using the rules determined for them. It is very clear that firewalls are not enough to secure a network completely because the attacks committed from intruders from outside the network are stopped whereas inside attacks are complicated to detect[1].

An Intrusion Detection System (IDS) is a security software tool that detects unwanted activities coming from various internet sources like spamming, spoofing, filtering etc. and gives an alert message on any doubtful activity is observed. It scans the network traffic or a system for harmful activities or policy attempts. Any malicious venture or violation is often reported to either to administrator or collected centrally employing a security information and event management (SIEM) system. A SIEM system consolidates outputs from multiple sources and uses alarm filtering techniques to differentiate malicious activities from these collected data.

Although intrusion detection systems monitor networks for probably malicious activity, they are additionally disposed to false alarms. Hence, organizations got to fine-tune their IDS product once they initial install them. It suggests that properly fixing the intrusion detection systems to acknowledge what traditional traffic on the network seems like as compared to malicious activity.

Intrusion bar systems additionally monitor network packets incoming the system to see the malicious activities concerned in it and directly sends the warning notifications.

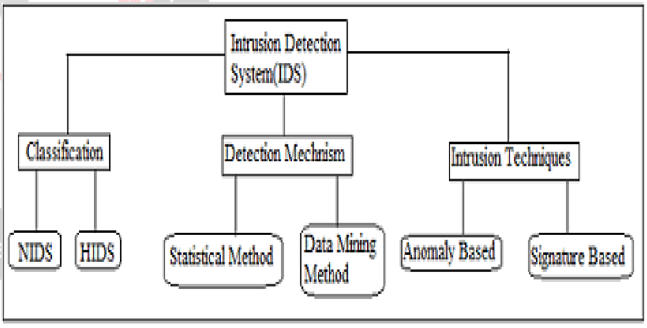

Fig.1.Intrusion Detection System- Classification, Detection Mechanism & Intrusion Techniques.

II. Types of Intrusion Detection System

Different types of Intrusion Detection System:

1. Network Intrusion Detection System (NIDS):

Network intrusion detection systems (NIDS) detect malicious activity like denial of services (DoS) attacks, port scans by monitoring network traffic. It investigate all the packet passing on the entire network and matches the traffic that is passed on the subnets to the collection of known attacks. Once an attack is identified or abnormal activity is observed, the alert message is sent to the administrator. An example of an NIDS is installing it on

DOI : 10.35291/2454-9150.2020.0722

the subnet where firewalls are located in order to see if someone is trying compromise the firewall.

2. **Host Intrusion Detection System (HIDS):** Host intrusion detection systems (HIDS) run on independent hosts or a devices on the network. A HIDS monitors the incoming and outgoing packets from the device and will alert the administrator if suspicious or malicious activity is detected. The principle of operation of HIDS depend on the fact that the successful intruders or botmaster will generally leave the trace of their activities such as keystroke logging, identity theft spamming, botnet activity, spyware-usage. This system is an instance based system where it takes a snapshot of existing system files and compares it with the previous snapshot. If the analytical system files were compromised then an alert is sent to the administrator to investigate the same. An example of HIDS usage can be seen on mission critical machines, which are not expected to change their existing system.

INTRUSION TECHNIQUES OF IDS:

1. **Signature- based Method:** Signature-based IDS detects the attacks on the basis of the some specific patterns such as number of bytes or number of 1's or number of 0's in the network traffic. It also detects on the basis of the already known malicious instruction sequence that is used by the intruders. In IDS detected pattern are known as signatures.

Some of the attacks which are already present in the system can be easily detected by Signature-based IDS, but there can be new malware attacks which can't be detected as their patterns are new to the system.

2. **Anomaly- based Method:** Anomaly-based IDS was introduced to detect the unknown malware attacks as new malware are developed so drastically. In anomaly-based IDS by using machine learning to create a trustful activity model and anything coming is compared with that model and it is declared suspicious if it is not found in model. Machine learning based method has a better comprehensive property in comparison to signature-based IDS as these models can be trained according to the applications and hardware configurations.

III. LITERATURE REVIEW

The Botnet population is growing rapidly and they represent some computer armies and have become a huge threat on the Internet[1].

A hybrid IDS combines K-means and two classifiers: K-Nearest Neighbor and Naïve Bayes for anomaly detection. Entropy based feature selection algorithm which selects the essential attributes and removes the redundant attributes. This algorithm operates on the KDD-99 Data set; this data set is used worldwide for evaluating the performance of different intrusion detection systems. The next step is clustering phase using k-Means [2].

Hybrid IDS combine the two approaches in one system. The hybrid IDS is come by combining packet header anomaly detection (PHAD) and network traffic anomaly detection (NETAD) which are anomaly-based IDSs with the misuse-based IDS Snort which is an open-source project. [3]

KDD99 (knowledge Discovery and Data Mining) intrusion detection[5] contest is specified. This system can detect the intrusions and further classify them into four categories: Denial of Service (DoS), U2R (User to Root), R2L (Remote to Local), and probe. The main goal is to reduce the false alarm rate of IDS1.[4]

IV. PROPOSED SYSTEM

In this proposed architecture, we are capturing a packets from network using JPCAP and WINPCAP. Data preprocessing is a technique which transforms the raw data into clear format. In machine learning feature reduction is the process of reducing the number of random variables under consideration come by a set of principal variables. It is divided into feature selection and feature extraction. In feature selection a set of relevant features are selected from set of features.

Feature selection:

F={X1,X2,X3,……Xn}

F'⊆F={X1',X2',…..,Xm'}

Feature extraction transforms or projects the original set of features into a new subspace which has smaller no of dimensions. Projection to M<N dimensions, In feature extraction and selection we can improve and maintain the classification accuracy and simplify classification complexity. In training we train the model to identify the suspicious packets received from packet capture module. k-means is an unsupervised learning algorithm which is used to solve many clustering problem. In proposed module it creates the clusters of similar packets.

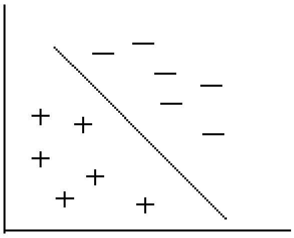

Fig. K-means Clustering example

SVM algorithm is used to classify the packets and generate the results.

DOI : 10.35291/2454-9150.2020.0722

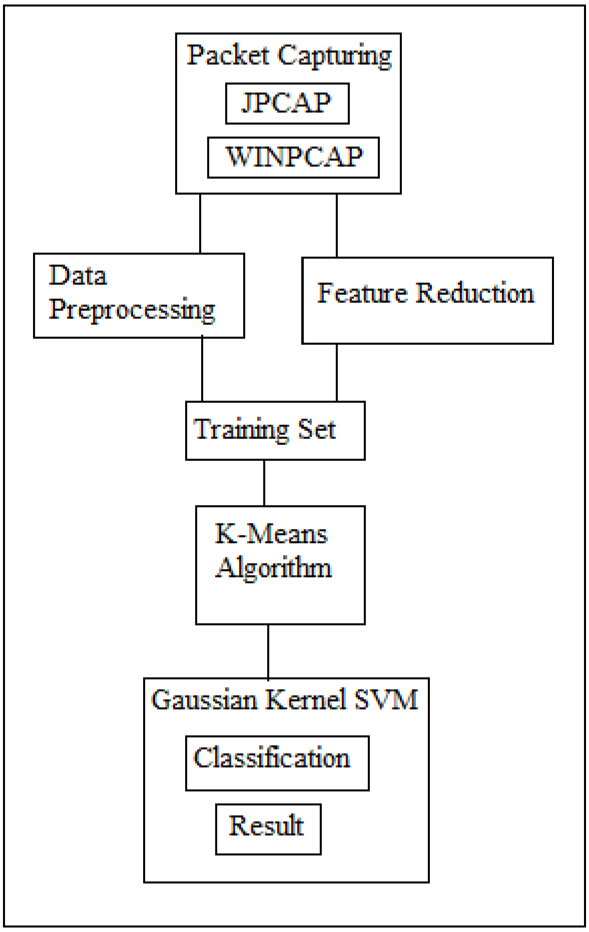

Fig.2. Proposed Architecture of Hybrid IDS

3.2. K Means Algorithm for Intrusion Detection:

1. Choose randomly five data records as initial Clusters Mean (cluster centre).

2. Evaluate the new centriod for the dataset, for each data record x from D,

3. Find the Euclidean distance between data record x and each cluster mean.

4. Allocate data record x to the closest cluster.

5. Re-calculate the mean for current cluster collections.

6. Perform the procedure until we get stable clusters.

7. Use these centroid classification of anomaly and normal traffic

The objective function is [7]:

$J=\sum^{k}_{i=1}\sum^{n}_{j=1}\ dij(Xj,Ci)$

Where dij(Xj,Ci) is a chosen distance measure (Euclidean distance) between a data point xj and the cluster center ci, is an indicator of the distance of the data points from their respective cluster centers.

3.3. SVM Classifier

SVM classifiers are used as it produces better results for binary classification when compared to other classifiers. But use of Linear SVM has some disadvantages of getting less accuracy result, over fitting result and robust to noise. These short comings are effectively overcome by the use of Gaussian Kernel SVM where nonlinear kernel functions are used.

$K(x,y) = \exp(-\|x-y\|^{2)}/(2\sigma^{2})$

The Gaussian Kernel function is used in proposed system mainly because it has less mathematical calculations. The kernel value of Kernel function always lies between zero and one. Thus, it is used as default kernel function for SVM classifier.

V. PROPOSED EXPERIMENTAL EVALUATION

For evaluation no of evaluation matrix are used for error metric; to know what are the errors on assuming H. for accuracy: precision, recall, etc. Suppose we want to make prediction of a value for a target feature on example X. Y is the observed value of target feature on X, Y' is the predicted value of target feature on example X.

Y'=h(X).

If Y' and Y are same then no error & else error is present. Absolute error on single training example=|h(x-y)| Absolute error on no of training example=1/n∑| h(x-y)|

Sum of squares errors=1/n∑δ(h(x)-y)^{2}

δ = 1 if h(x), y are different.

δ = 0 if h(x), y are same.

Accuracy=(TP+TN)/(P+N)

Precision=TP/(TP+FP)

Recall=TP/P

Where,

P=Positive

N=Negative

TP=True Positive

FP=False Positive

TN=True Negative

In proposed hybrid architecture we will get a good accuracy level for the detection of intrusion and reduce the misclassification of packets.

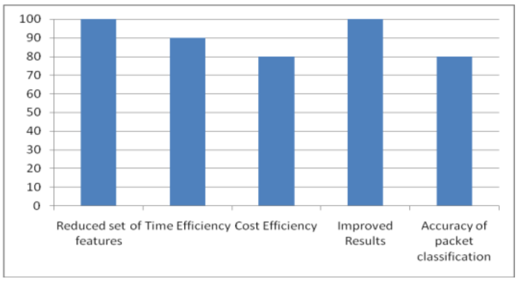

Fig.3. Performance and result analysis of proposed system

Above fig.3 shows the performance and result analysis of proposed system. Here we consider performance measure as reduced feature set, time efficiency, cost efficiency, improved results and accuracy of packet classification.

VI. Conclusion

Signature-based systems can only detect attacks that are present in system before whereas anomaly based systems are able to detect unknown attacks. Anomaly-based IDSs make it possible to detect attacks whose signatures are not included in rule files. In proposed architecture we get a good level of accuracy of packet classification by using k-means algorithm and SVM Gaussian Kernel. SVM Gaussian Kernel is mostly responsible for cost reduction of proposed system.

References

[1] Ms.Pooja M. Pondkule ,Mrs. B. Padmavathi BotShark - Detection and Prevention of Peer-to-Peer Botnets by Tracking Conversation using CART", International Conference on Electronics, Communication and Aerospace Technology,ICECA 2017.

[2] Debra Anderson, Thane Frivold, and Alfonso Valdes, "NIDES Next-generation Intrusion Detection Expert System (NIDES)", A Summary, Computer Science Laboratory, SRI-CSL-95-07, May 1995 5.Te-Shun Chou and Tsung-Nan Chou, "Hybrid Classified Systems for Intrusion Detection," Seventh Annual Communications Networks and Services Research Conference, pp. 286-291, 2009.

[3] M. Ali Aydın A. Halim Zaim K. Gökhan Ceylan "A hybrid intrusion detection system design for computer network security" Volume 35, Issue 3, May 2009, Elesvier Pages 517-526

[4] Dr.V.Suganthi,*1 , P. K. Manoj Kumar 2 "Intrusion Detection System – A Literature Survey",

[5] S. Chen, B. Mulgrew, and P. M. Grant, "A clustering technique for digital communications channel equalization using radial basis function networks," *IJREAM Trans. Neural Networks*, vol. 4, pp. 570–578, Jul. 1993.

[6] S.Peddabachigari,A.Abraham,c.Grosan,J.Thomas, "modeling intrusion detection system using hybrid intelligent system",Journal of Network and Computer Applications 30 114-132, 2007.

[7] C. Xiang, P.C. Yong, L.S. Meng, "Design of multiple level hybrid classifierfor intrusion detection system using bayesian clustering and decision tree", Pattern Recognition Letters 29 918-924,2008.

[8] Sanoop Mallissery , Jeevan Prabhu,and Raghavendra Ganiga, "Survey on intrusion detection method",2011.

[9] Deepthy K Denatious & 2Anita John, "Survey on Data Mining Techniques to Enhance Intrusion Detection",2012.

[10] Changxin Song, Ke Ma Institute of Computer Information & Technology of Qinghai Normal University Network Center of Qinghai Normal University Qinghai, China., "Design of Intrusion Detection System Based on Data Mining Algorithm",2009.

[11] Li Tian1, Wang Jianwen1 Department of Computer Science, North China Electric Power University (NCEPU), Baoding 071003, China, "Research on Network Intrusion Detection System Based on Improved K-means Clustering Algorithm",2009

[12] Li Xue-yong, Gao Guo- "A New Intrusion Detection Method Based on Improved DBSCAN",2010.

[13] Z. Muda, W. Yassin, M.N. Sulaiman and N.I.Udzir, "Intrusion Detection based on K-Means Clustering and One R Classification",2011.

DOI : 10.35291/2454-9150.2020.0722