ISSN (Online): 2456-0448
International Journal Of Innovative Research In Management, Engineering And Technology
Vol. 2, Issue 7, July 2017

Survey of Network Inrusion Detection Techniques (Phase Wise Analysis Elucidation)

[1] Gritto.D, [2] Mohamed Suhail.M
[1]M.Phil. Research Scholar Department Of Computer Science,Vels University, Chennai.
[2]Assistant Professor, Department Of Bca & It,Vels University, Chennai.

Abstract: ***Escalation of the internet and the demand for magnifying the network security is increasing exponentially. The intrusion detection is an active network monitoring practice for finding the unauthorized access, policy violations, anomalous behaviors, malicious attacks, unconcerned packets detection etc. The present security fraternities like antivirus, cryptography and firewalls could not ensure complete protection for the systems linked with the internet. This paper reviews various data mining based intrusion detection techniques. Deep emphasis is given to observe the best among the Machine learning algorithm that records the optimal degree of attack detection and false alarm rate. The analysis is made using KDD cup '99 dataset. The algorithms like J48, Random tree, and Random forest are evaluated and equated to identify their detection ratio.***

Keywords- ***IDS-Intrusion Detection System; MBID ABID,HBID, NBID -(Misuse, Anomaly, Host, Network) Based Intrusion Detection; Packet Sniffing; Feature selection; Outlier detection; False Alarm; J48; RF-Random Forest; RT-Random Tree.***

I. INTRODUCTION

The rapid growth in the volume of sensitive data traversing over the network has increased the probability of security attacks proportionately. Endanger of attacks like DoS, U2R, R2L, Spoofing, Sniffing, and Probing etc are ever present. The organizations are configured with the excellent technologies for detecting and avoiding malicious attacks. The tools like antivirus, firewall and IDS strengthen the security infrastructure. **Antivirus** checks the programs, files or software that are already stored or installed in the system for any vulnerability. The **Firewall** is also called as IP filter analyzes the packet header and prohibits the traffic from the intruders IP address. The firewall stops the exchange of data abruptly without any intimation.

The **Intrusion** is defined as any policy violation that attempts to compromise the Confidentiality, Integrity, and Availability **(CIA)** of the security infrastructure within the organization. The **IDS** signals the alarm on sensing any unauthorized access, policy violations, anomalous behavior, malicious attacks, unconcerned packets detection or any other security breaches. The IDS analysis the whole packet i.e., the header and the payload for the availability of any attack patterns, if any positive signature is present then the alarm will be generated for alerting the security manager. The IDS provides dough security barrier among the other.

The IDS in general are classified based the method of detection or resource configured.

MIBD is based on pattern matching procedure. The attack signature database with known attack pattern profile is created. The instantly exploited payload pattern is compared with the learned attacks patterns in the database for detecting the intrusion. *The general misuse base detection approaches includes: Rule based system, State transition analysis, Data mining and Machine learning etc.* **ABID** is based on heuristics procedure. The anomaly detection identifies any abnormality in the network behavior for detecting the intrusion. The baseline database enumerating the set of accepted network behavior or benign activities is created. The current event is compared with the baseline database for finding any deviation and the event classified either as normal and abnormal event. *The general anomaly detection approaches includes: Artificial intelligence, Neural Network, Data mining etc.* **HIDS** identifies the anomalous behavior on a specific host. HIDS in general inspects the critical system files for any security threat. The deployment of HIDS will defense the local systems like server, router, gateway, DNS or any intersecting nodes.

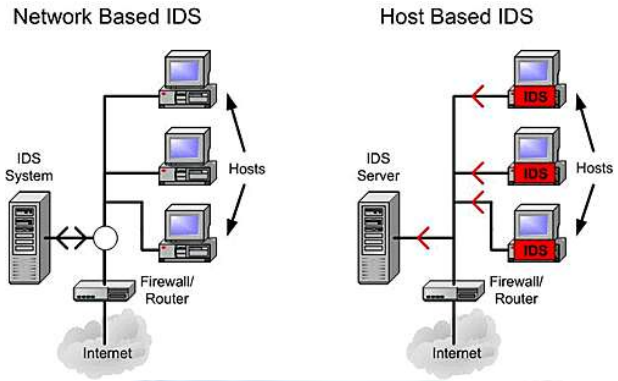
ISSN (Online): 2456-0448
International Journal Of Innovative Research In Management, Engineering And Technology
Vol. 2, Issue 6, June 2017

HIDS Tools: Tripwire, Cisco HIDS, Symantec ESM.

NIDS protects the entire network system to be safe and secure by inspecting the traffic for any attack signatures. NIDS notifies the attack by sending alerts to the administrator or kills the host connection by sending RST, or enforces strong firewall policies for blocking the connection.

NIDS Tools: Snort, Cisco NIDS, Net prowler.

Figure 1: NIDS Vs HIDS



The limitation with HIDS is that they are not capable of monitoring the traffic that is not directed to a particular host. They are inefficient for real time instantaneous attack detection since they rely on local system resource. They are also hard to integrate with gateway, DNS etc. The intruders can easily compromise is HIDS based system by cracking the host server through the control of C&C server. The NIDS in contrast, are tailored to detect serious intrusions like unauthorized access, DoS and bandwidth stealing etc. They are also suitable for real time based intrusion detection. In research work the various techniques that are related to network intrusion detection is reviewed and evaluated.

II. KDD Cup '99 Dataset

The analysis of this survey is done using KDD cup '99 dataset. The KDD cup '99 is the accumulation tcpdump network traffic segment of DARPA volumes 4GB. The packet dissemination of the dataset contains 41 features and 24 types of attacks. The attacks are classified into 4 types.

Denial of Service (DoS) is an attack event in which the perpetrators or intruders disrupts the legitimate user from accessing the system or network resources temporarily. The DoS is achieved by making the memory of the resources busy through traffic flooding. The traffic flooding is an act of generating huge mass of well-planned request with the objective of prohibiting the service and there by degrading the system performance.

Probing refers to the acquisition of vulnerable information about the objective network from the external network. On learning the susceptibility the intruders plot the attack plan to exploit the weakness. The attackers in general surfs for the host with open port by sending wisely designed packet to all destination port numbers once, if any open port is identified the stealthy action starts.

User to Root Attack (U2R) the attackers of this class gain the password of the normal user using packet sniffing technique. They masquerade like normal user and exploit the vulnerabilities to gain the root access and super user privileges.

Remote to Local Attack (R2L) the user with no access privileges tries to gain illegal access of the machine in the local network by sending packets and cracking the security infrastructure. The attacker gain access through the machine that is located outside the network through remote access.

Table 1: Attack types and categories

Category	Attack types
DoS	apache, back, land, mailbomb, neptune, pod
Probe	ipsweep, mscan, nmap, portsweep, saint, satan processtable, smurf, teardrop, udpstorm
U2R	buffer_overflow, loadmodule, perl, rootkit, ps sqlattack, xterm
R2L	ftp_write, guess_password, imap, multihop

The feature in KDD cup '99 dataset is classified into three categories.

Basic Features: Includes the attributes extracted from TCP/IP connection packets. These basic features are generally extracted from the packet header the typical attributes are src_bytes, dst_bytes, protocol etc.

Content Features: Includes the attributes extracted from TCP/IP connection packets. These basic features are generally extracted from the packet payload the typical attributes are number of failed login attempts, number of file creation operation etc.

Traffic Features: Includes the features that are computed using the window intervals and are divided into two categories.

Same host feature: Attributes of connection that has same destination address for long time.

Same Service Features: Attributes of connection has same service for long time.The typical attributes are dst host count; dst host srv count; dst host same srv rate; dst host diff srv rate; dst host same src port rate;

III. Data Acquisition

The intrusion detection involves online monitoring of system resources for any anomalous behavior. The customary method for misuse or attack detection involves inspecting the log files, event statistics user connectivity etc. But these techniques are infeasible for complex network monitoring so packet sniffing technique is used. The packet sniffing is an act of capturing the data stream packets over the network. The captured packet is intercepted to analyze the network activities for detection of any intrusion, troubleshooting and forensics etc. Several packet sniffing tools like tcpdump, wireshark, ngrep, network miner and snoop can be used for capturing the packets. The packet loss is one of the noticeable issues in the packet sniffing. The packet lost must be prevented for improving the reliability and the rate of attack detection, so more than one tool can be configured for traffic capturing to avoid packets loss. Compare the packets captured by each tool and find any missing packets from either of the tool. The new packets thus derived are merged with the existing packet group.

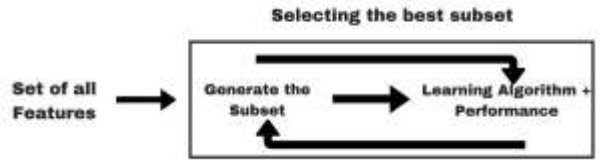
IV. Data Preprocessing

Data preprocessing is a technique of transforming the inconsistent raw fact to complete information by avoiding any irregular and susceptible features like noise, missing values etc. The essential task is to eliminate the redundant data and unconcerned features to avoid confusions and generation of inaccurate knowledge. The most common task in data preprocessing is the feature selection and outlier detection.

Feature Selection

Feature selection or attribute selection involves extraction of attributes that are most relevant to the predictive model. Feature selection is the process of selecting a subset of relevant features. There are three general classes of feature selection algorithms: **Filtering method** is an independent statistical method that assigns a score for each attribute. The attributes are either selected or rejected at later stages. **Wrapper Method** selects set of features with different combinations. Each combination is evaluated and compared to other combinations for assigning cross combination score. The best combination is chosen. **Embedded Method** selects the best features that contribute to the accuracy of the concern model.

Figure 2: Feature Selection



Features of KDD cup '99 dataset includes: duration; protocol type; service; flag ; source bytes; destination bytes; land; wrong; fragment; urgent; hot; failed logins; logged in; #compromised; root shell; su attempted; #root; #file; creations; #shells;#access files; #outbound cmds; is hot login; is guest login; srv count; serror rate; srv serror rate; rerror rate; srv rerror rate; same srv rate; diff srv rate; srv diff host rate; dst host count; dst host srv count; dst host same srv rate; dst host diff srv rate; dst host same src port rate; dst host srv diff host rate; dst host serror rate; dst host srv serror rate; dst host rerror rate; dst host srv rerror rate.

Outlier Detection

The outlier in intrusion detection depicts any suspicious observations that are highly deviated from other observations. The genres of outlier detection algorithm include: Naive Bayes, Decision Tree, K-mean, Genetic, SVM , EM, AdaBoost, Page ranking algorithm etc. The review of former three algorithms is done in this work.

Naive Bayes Algorithm: Naive Bayes algorithm is based on probability model. The Naive Bayes classifier calculates the set of probabilities by counting the frequency and combinations of values in a given data set. The algorithm uses Bayes theorem and assumes all attributes to be independent.

K-means Algorithm: Groups the objects based on their feature values into K disjoint clusters. Objects that are classified into the same cluster have similar feature values. K is a positive integer number specifying the number of clusters, and has to be given in advance. Here are the four steps of the K-means clustering algorithm:

[1] Define the number of clusters K.

[2] Initialize the K cluster centroids. This can be done by arbitrarily dividing all objects into K clusters, computing their centroids, and verifying that all centroids are different from each other. Alternatively, the centroids can be initialized to K arbitrarily chosen, different objects.

[3] Iterate over all objects and compute the distances to the centroids of all clusters. Assign each object to the cluster with the nearest centroid.

[4] Recalculate the centroids of both modified clusters.

[5] Repeat step 3 until the centroids do not change any more.

Genetic Algorithm: This algorithm uses Genetic algorithms are very much suitable for outlier detection. GA based outlier detection techniques in pseudo –

Input: A dataset for outlier detection

Output: Outlier with lowest fitness value

1. Generate random population of N individuals.
2. Fitness function f(x) for each chromosome is evaluated.
3. [New Population] Repeat the following steps to create new population

i) [Selection] Select two parents from the population according to their fitness.

ii) [Crossover] with the crossover probability crossover the parents to form new offspring. If no crossover is performed the offspring is resulted as parents.

iii) [Mutation] with the mutation probability mutate the offspring at each locus.

iv) [Accept] Place new offspring in the population.

4. [Replace] Use new generated population for the next iteration.
5. [Test] If the termination condition is satisfied, return the best solution.

6. [Result] Sort the fitness value in descending order, the lower value is identified as outliers.
7. [Loop] Go to step 2 for next

V. Evaluation of Intrusion

The evaluation for intrusion detection is done by examining the packet header or the payload or the both. The preliminary detection can be done by classifying the packets based on various features like source, destination IP address, port number, type of service, number of packets and bytes sent, time of data transfer and different source-destination pairs etc. The well-known port numbers are susceptible for easy attacks. The most malicious attacks can be detected by finding the protocol type of the packets (Web/HTTP, TCP, UDP, and ICMP). Some anomaly analysis can be done by accessing the time intervals, when the time of data transfer exceeds the definition of the protocol. The other manual technique of detection includes the source and destination IP address detection. If any intruders ID address is traced then alarm will be triggered.

VI. Evaluation Classifiers

In this work the efficiency of J48, Random tree and Random forest algorithm is compared. **J48 classifier** is a simple C4.5 decision tree for classification. It creates a binary tree. The decision tree approach is most useful in classification problem. With this technique, a tree is constructed to model the classification process.

Random tree usually refer to randomly built trees used as classifier. **Random forest** is a learning method for classification. This operates by constructing decision trees at training time and outputting the class of same mode**.**

VII. Evaluation Metrics

There are many characteristics to estimate the IDS. The accuracy of detection and false alarm rate tops the scale in the survey. The selection of appropriate feature and classification tactics results in optimal detection with minimized false alarm rate. The prediction should categorize the accurately as normal or attack. The various metrics used in the evaluation and estimation is listed below.

Table 2: Confusion matrix

Traffic	**Classified as**	
	Normal	**Attack**
Normal	T_N	F_P
Attack	F_N	T_P

$$TruePositive(T_P) = \frac{T_P}{T_P + F_N}(AaA)$$

$$TrueNegative(T_N) = \frac{T_N}{T_N + F_P}(NaN)$$

$$FalsePositive(F_P) = \frac{F_P}{F_P + T_N}(NaA)$$

$$FalseNegative(F_N) = \frac{F_N}{F_N + T_P}(AaN)$$

$$DetectionRate = \frac{T_P}{T_P + F_N}$$

$$Falsealaramrate = \frac{F_P}{F_P + T_N}$$

$$Accuracy = \frac{T_P + T_N}{T_P + T_N + F_P + F_N}$$

$$Precision = \frac{T_P}{T_P + F_P}$$

where T_P,F_P: True positive, False Posiive;

T_N, F_N: True negative and False Negative;
AaA: Attack as Attack (Detection Rate, True Positive);
AaN: Attack as Normal (False Negative);
NaN: Normal as Normal (True Negative);
NaA: Normal as Attack (False alaram, False Positive)

Table: 3 Instance Table

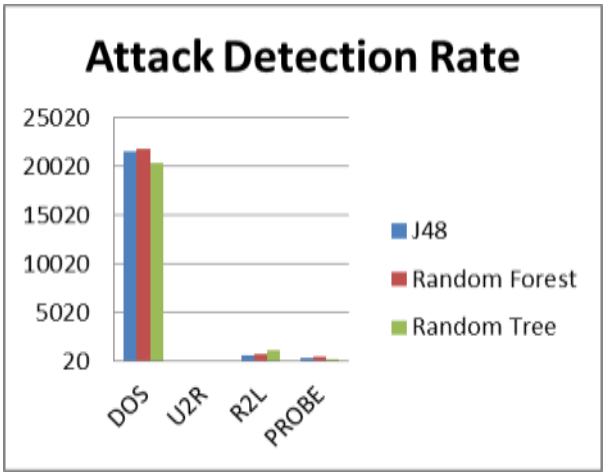
Attack class	**Training Data Set**	**Test Data Set**
DoS	43736	26814
U2R	12	29
R2L	136	1889
Probe	455	485
Normal	12449	7067
Total	56788	36284

Table: 4 Categorization Table

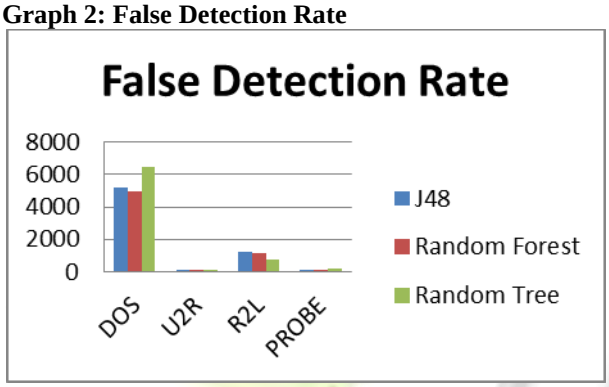
Classifier	**DoS**		**U2R**	
	CORRECT	**FALSE**	**CORRECT**	**FALSE**
J48	21612	5202	8	21
Random Forest	21856	4958	3	26
Random Tree	20368	6446	14	15
Classifiers	**R2L**		**PROBE**	
	CORRECT	**FALSE**	**CORRECT**	**FALSE**
J48	649	1240	364	121
Random Forest	705	1184	468	17
Random Tree	1161	728	260	225

The results show that the J48 and Random forest result in optimal detection rate for all five categories of traffic. The J48 classifier detection accuracy is higher comparing to Random forest. The Random forest result is higher than the Random tree. The result estimation may also vary based on the test dataset soon for evaluation.

Graph 1: Attack detection Rate



Graph 2: False Detection Rate



VIII. Conclusion

This paper reviews various intrusion detection techniques and estimates that the best among the machine learning techniques are J48, RF and RT. The experimental result reveals RF and RT produce optimal detection rate and false attack detection rate. The J48 produces fair results when comparing with the other two. The accuracy rate can also be increased by combining of two algorithms. Huge challenges are involved in detecting the intrusion in real time environment like cloud based infrastructure. The indexing of known attack patterns in the profile database is a tedious task. The future work involves the devising intrusion detection model that implements the optimal indexing features.

References

[1] Kalpana Jaswal, Pravween Kumar and Seema Rawat, "Design and Development of a Prototype Application for Intrusion Detection using Data Mining," in UP, 978-1-4673-7321-2/15/$31.00 IEEE, 2015.
[2] B. Raju and B. Srinivas, "Network Instruction Detection System Using KMP Pattern Matching Algorithm," in Warangal, India, IJCST vol. 3, pp. 33-36, January 2012.
[3] Zhou Chunyue, Liu Yun and Zhang Hongke, "A Pattern Matching Based Network Intrusion Detection System," in Beijing, China, 1-4244-0342-1/06/$20.00 IEEE, 2006.
[4] L. Vokorokos and A. Balaz, "Host – Based Intrusion Detection System," in Kosice, 978-1-4244-7652-7/10/$26.00 ©2010 IEEE, 2010.
[5] Alaoui- Adib Saad, Chougdakli Khalid and Jedra Mohamed, "Network Intrusion Detection System Based on Direct LDA," in Rabat, 978-1-4673-9669-1/15/$31.00 IEEE, 2015.
[6] Anuradha and Anita Singhrova, "A Host Based Intrusion Detection System for DDOS Attack in WLAN," in Murthal Sonepat, India, 978-1-4577-1386-611$26.00 IEEE, 201I.

[7] F. Lydia Catherine, Ravi Pathak and V. Vaidehi, "Efficient Host Based Intrusion Detection System Using Partial Decision Tree and Correlation Feature Selection Algorithm," in Chennai, India, 978-1-4799-4989-2/14/$31.00 IEEE, 2014.
[8] Mahbod Tavallaee, Ebrahim Bagheri, Wei Lu and Ali A. Ghorbani, "A Detailed Analysis of the KDD CUP 99 Data Set," 978-1-4244-3764-1/09/$25.00 IEEE, 2009.
[9] Robert Moskovitch, Shay Pluderman, Ido Gus, Dima Stopel, Clint Feher, Yisrael Parmet, Yuval Shahar and Yuval Elovici, "Host Based Intrusion Detection Using Machine Learning" in Israel, 1-4244-1330-3/07/$25.00 IEEE, 2007.
[10] Ed' Wilson Tavares Ferreira, Ailton Akira Shinoda, Ruy De Oliveira, Valtemir Emerencio Nasciment and Nelcileno Virgilio De Souza Araujo, "A Methodology for building a Dataset to Assess Intrusion Detection Systems in Wireless Networks," in Brazil, E-ISSN: 2224 - 2864, vol. 14, pp. 113-119, 2015.
[11] Firkhan Ali Bin Hamid Ali and Yee Yong Len, "Development of Host Based Intrusion Detection System for Log Files," in Langkawi, Malaysia, 978-1-4577-1549-5/11/$26.00 IEEE, 2011.
[12] Lata and Kashyap Indu, "Novel Algorithm for Intrusion Detection System," in Haryana, India, IJARCCE, vol. 2, Issue 5, May 2013.
[13] S. Sobinsoniya and S. Maria Celestin Vigila, "Intrusion Detection System: Classification and Techniques," in Tamilnadu, India, 978-1-5090-1277-0/16/$31.00 IEEE, 2016.
[14] Yanjie Zhao, "Network Intrusion Detection System Model Based on Data Mining," in Weifang. China, 978-1-5090-2239-7/16/$31.00 IEEE, 2016.