An Efficient and Accurate Intrusion Detection System to detect the Network Attack Groups using the Layer wise Individual Feature Set

Prof.S.S.Manivannan[1], Dr.E.Sathiyamoorthy[2]
[1&2] School of Information Technology and Engineering
VIT University, Vellore, Tamil Nadu , India
[1]manivannan.ss@vit.ac.in, esathiyamoorthy@vit.ac.in[2]

Abstract :

In the field of Network Security, Intrusion is the severe threat for various Networks. So an efficient Intrusion Detection System is required to detect the intrusions that are spread through the Network. The main idea of this paper is to reduce the average control path latency incurred between request and response of the system as well as the increasing the detection rate of network attack groups. This paper proposes two approaches to design the efficient and accurate Intrusion Detection System. The proposed system make use of Individual Feature Set and Layer wise approach to achieve the efficiency and accuracy in detecting the network attack groups. The proposed system categorizes the network attacks in to four groups such as Denial of Service attacks, User to Root attacks, Remote to Local attacks and Probe attacks. Experimental results shows that the attack detection rate of the proposed method is high when compared to the other methods such as Support Vector Machine, C4.5 algorithm, Decision Tree with Principle Component Analysis, K means clustering and Multi Classifier for detecting the network attacks.

Keywords:

Intrusion Detection, Network attacks, Individual Feature Set, Layered approach, Latency

1. Introduction

Networks such as Internet and Intranet plays a vital role in Information Technology, business, education and other fields. The rapid usage of internet and its applications are increasing day by day. The possibilities of undetected intrusions from Networks are also increasing. So the efficient and accurate Intrusion Detection System is essential. An Intrusion Detection System (IDS) is used to monitor network traffic, check for suspicious activities and notifies the network administrator or the system. In some instances, the IDS might also react to malicious or anomalous traffic and will take action such as barring the user or perhaps the IP address source from accessing the system. IDS are a device or software application that monitors network behavior for malicious activities and security policy violations. IDS typically record information about the malicious activities and inform to security administrators of observed events and produce the reports.

Intrusion may attack the systems either manually or via software expert systems that operate on logs. An intrusion is a deliberate, unauthorized attempt to access or manipulate information or system. When suspicious activity is from the internal network it is classified as misuse intrusion detection. IDS are dedicated assistants that are used to monitor the rest of the security infrastructure. Today’s security infrastructure are becoming extremely complex, it includes firewalls, identification and authentication systems, access control list, virtual private networks, encryption products, virus scanners, and more. Failure of one of the above component will result in the flaw in the security policies. Different Intrusion Detection Systems methods are:

(i) Network Intrusion Detection Systems (NIDS)
(ii) System Integrity Verifiers (SIV)
(iii) Log File Monitors (LFM)
(iv) Deception Systems (Honey Pots)

A Network Intrusion Detection System (NIDS) is an intrusion detection system that attempts to discover unauthorized access to a computer network by analyzing traffic on the network for signs of malicious activity. In NIDS, the individual packets flowing through a network are analyzed. The NIDS can detect malicious packets that are designed to be overlooked by a firewall

2. Literature Survey

A number of Intrusion Detection Systems were developed to detect the Network attacks. Boughaci, D. Drias, H. Bendib and A.Bouznit have presented the Distributed Intrusion Detection Framework based on Autonomous and Mobile agent [1]. The system uses five types of agents such as administrator agents, analyzer agents, connection agents, crisis agents and update behavior agents. These five agents interact with each other to perform the detection task. Mrutyunjaya Panda and Manas Ranjan Patra have proposed efficient data mining algorithm called naive bayes [2] for anomaly based network intrusion detection. Here Naive Bayes technique performs better in terms of false positive rate, cost and computational time when applied to KDD'99 data sets compared to a back propagation neural network. FuHau Hsu, Fanglu Guo and Tzicker Chiueh have presented a network-based buffer overflow attack detection system called Nebula [3] that detects the known buffer overflow attacks based on the packets observed. Nebula is built on a centralized TCP/IP architecture that effectively defeats all existing Network Intrusion Detection System techniques. Nebula incorporates a payload type identification mechanism that reduces the false positive rate. Charles Sutton and Andrew McCallum have presented an introduction to conditional random fields for relational learning. They have presented an example of applying a general Conditional Random Fields (CRF) [4] to a practical relational learning problem. Conditional Random Fields can capture long distance dependencies between labels. To represent these long-distance dependencies a skip-chain CRF is proposed, a model that jointly performs segmentation and collective labeling of extracted mentions.

Christina Warrender, Stephanie Forrest and Barak Pearlmutter have proposed a method to detect the intrusions using System Calls. Using system-call data sets [5] generated by several different programs, the ability of different data modeling methods were compared to represent normal behavior accurately and to recognize the intrusions. The factors affecting the performance of each method were analyzed. The weaker methods than Hidden Markov Models (HMMs) are likely to be sufficient to detect the intrusions. H.Debar, M.Becker and Siboni have presented a neural network component [6] for an Intrusion Detection System. The model introduces the use of a neural network component for modeling user's behavior as a component to detect the intrusions. The approach based on the IDES (Intruder Detection Expert System) which has two components, an expert system and a statistical model. The model learns the habits of a user when he works with the computer and raises warnings when the current behavior is not consistent with the previously learned patterns.

E. Tombini, H. Debar, L. Me and M. Ducasse have presented a serial combination of anomaly and misuse intrusion detection techniques applied to HTTP traffic [7]. Combining an anomaly and misuse intrusion detection techniques offers the advantage of separating the monitored events between normal and intrusive. A serial architecture provides the operator with better detection results. Meera Gandhi and S.K.Srivatsa have proposed a system to detect and combat some common attacks on network systems. A signature based Intrusion Detection System [8] will monitor packets on the network and compare them against a database of signatures or attributes from known malicious threats. The system displays the list of attacks in the log and informs the administrator for evasive action. Paul Dokas, V. Kumar and Jaideep Srivastava have proposed a method based on Protecting against Cyber Threats in Networked Information Systems [9] . Traditional signature based techniques for detecting cyber attacks can only detect previously known intrusions and are useless against novel attacks and emerging threats. Research at the University of Minnesota is focused on developing data mining techniques to automatically detect attacks against computer networks and systems. Experimental results on live network traffic at the University of Minnesota show that the new techniques show great promise in detecting novel intrusions.

Hyunsang Choi, Heejo Lee and Hyogon Kim have proposed a Parallel Coordinate Attack Visualization (PCAV) method for detecting unknown large-scale Internet attacks. PCAV [10] displays network traffic on the plane of parallel coordinates using the flow information such as the source IP address, destination IP address, destination port and the average packet length in a flow. The parameters are used to draw each flow as a connected line on the plane, where a group of polygonal lines form a particular shape in case of an attack. Wei Wang, Xiao H. Guan and Xiang L. Zhang have proposed a new efficient intrusion detection method based on hidden Markov models (HMMs) . HMMs are applied to model the normal program behaviors using traces of system calls issued by processes [11]. The output probability of a sequence of system calls is calculated by the normal model built. If the probability of a sequence in a trace is below a certain threshold, the sequence is flagged as a mismatch. If the ratio between the mismatches and all the sequences in a trace exceeds another threshold, the trace is then considered as a possible intrusion.

Wenke Lee, Salvatore Stolfo and Kui W. Mok have proposed a method called data mining audit data to build intrusion detection models [12] . The idea is to mine system audit data for consistent and useful patterns of program and user behavior, and use the set of relevant system features presented in the patterns to compute classifiers that can recognize anomalies and known intrusions. Wenke Lee and Salvatore J. Stolfo have proposed Data mining approaches for Intrusion Detection. Data mining techniques [13] are used to discover consistent and useful patterns of system features that describe the program and user behavior. The association rule algorithm and the frequent episodes algorithm are used to compute the patterns, which are used to describe the program or user behavior. An agent-based intrusion detection system is designed to compute and provide the detection models to detect the intrusions.Y.S. Wu, B. Foo, Y. Mei, and S. Bagchi have proposed a Collaborative Intrusion Detection System (CIDS) A Framework for Accurate and Efficient IDS . CIDS [14] employs Snort, a network level IDS, Libsafe, an application level IDS, and a new kernel level IDS called Sysmon. The system has a manager to which the detectors communicate their alarms using a secure message queue. The manager has a graph-based and a Bayesian network based aggregation method for combining the alarms to finally come up with a decision about the intrusion. Z. Zhang, J. Li, C.N. Manikopoulos, J. Jorgenson and J. Ucles have presented a Hierarchical Network Intrusion Detection System. Hierarchical Intrusion Detection (HIDE) system [15], detects network-based attacks as anomalies using statistical preprocessing and neural network classification. Five different types of neural network classifiers such as Perceptron, Backpropagation (BP), Perceptron-backpropagation-hybrid (PBH), Fuzzy ARTMAP and Radial-based Function were tested. The results showed that HIDE can reliably detect UDP flooding attacks with attack intensity as low as five to ten percent of background traffic.

3. PROPOSED SYSTEM

A. Proposed Methodology

The proposed system concentrates on increasing the accuracy of the network attacks and efficiency in detecting the attacks by using individual Feature Set and Layer wise method. Individual Feature Set (IFS) is used to give the high attack detection accuracy and layer wise method is used to give high efficiency.

B. Connection Establishment

The connection is established between client and server in order to enable file transfer over the network through the concerned layers. A connection is a sequence of TCP packets in which data flows between source IP address and a target IP address under a well defined protocol.

C. Connection Features

Salvatore J. Stolfo has defined higher-level connection features that help in distinguishing normal connections from attacks. These connection features are called as Generic Feature Set.

D. Generic Feature Set

Massachusetts Institute of Technology (MIT) Lincoln Labs , USA has conducted the “1998 DARPA Intrusion Detection Evaluation Program”. The generic features set uses the original data set managed with DARPA Intrusion Detection Program. Three types of generic feature set are :

i) Basic features of TCP connections
ii) Content features within a connection
iii) Computed traffic features

The various features to be considered are described below.

Table. 1 Basic Features of TCP connections

Sl. No	Feature Name	Description
1	duration	length of the connection
2	protocol_type	type of the protocol
3	service	network service on the destination, e.g., http, telnet
4	src_bytes	number of data bytes from source to destination
5	dst_bytes	number of data bytes from destination to source
6	flag	normal or error status of the connection
7	land	1 if connection is from/to the same host/port; 0 otherwise
8	wrong_fragment	number of wrong fragments
9	urgent	number of urgent packets

Table.2 Content Features within a connection

No	Feature Name	Description
1	hot	number of ``hot" indicators
2	num_failed_ logins	number of failed login attempts
3	logged_in	1 if successfully logged in; 0 otherwise
4	num_compromised	number of compromised conditions
5	root_shell	1 if root shell is obtained; 0 otherwise
6	su_attempted	1 if su root command attempted; 0 otherwise
7	num_root	number of root accesses
8	num_file_creations	number of file creation operations
9	num_shells	number of shell prompts
10	num_access_files	number of operations on access control files
11	num_outbound_cmds	number of outbound commands in an ftp session
12	is_hot_login	1 if the login belongs to the hot list; 0 otherwise
13	is_guest_login	1 if the login is a guest login; 0 otherwise

Table. 3 Computed Traffic Features

Sl. No	Feature Name	Description
1	count	number of connections to the same host
2	serror_rate	% of connections that have SYN errors to same host
3	rerror_rate	% of connections that have REJ errors to same host
4	same_srv_rate	% of connections to the same service
5	diff_srv_rate	% of connections to different services
6	srv_count	number of connections to the same service
7	srv_serror_rate	% of connections that have SYN errors to same service
8	srv_rerror_rate	% of connections that have REJ errors to same service
9	srv_diff_host_rate	% of connections to different hosts

The above Tables 2,3 and 4 show the list of 31 features under generic feature set.

E. Network Attack Groups

All network attacks can be grouped under 4 main types of attacks. They are :

- Denial of Service (DoS) attack
- Unauthorized access to Root (U2R) attack
- Remote to Local (R2L) attack
- Probe attack

F. Individual Feature set for each attack group

The four network attack groups such as DoS, U2R, R2L and Probe attacks are executed in different ways. So it is necessary to fix the individual feature set for each network attack group. Therefore, features are selected for each layer based on the attacks executed in each layer.

Individual Feature Set for Probe attacks:

Probe attack is an attempt to gain access to a computer and its files through a known or probable weak point in the computer system. So basic TCP connection features like duration of connection and source bytes are important while other features like number of files created and number of files accessed are not considered for detecting probes.

Example probe attacks: ipsweep, nmap, portsweep

We have selected 4 features :

- duration
- protocol_type
- src_bytes
- service

Individual Feature Set for DoS attacks:

A denial-of-service attack (DoS attack) is an attempt to make a computer or network resource unavailable to its intended users. One common method of attack involves saturating the target machine with external communications requests, such that it cannot respond to legitimate traffic. So traffic features are selected for DoS attacks.

Example DoS attacks: pod, smurf attacks

We have selected 8 features :

- duration
- protocol_type
- src_bytes
- count
- dst_host_same_servicerate
- dst_host_serror_rate
- dst_host_srv_serror_rate
- dst_host_reerror_rate

Individual Feature Set for R2L attacks:

R2L attacks are unauthorized access from a remote machine. As R2L attacks deal with network level and host level , features related to both network and host levels are considered for R2L attacks.

Example R2L attacks : guess_passwd, imap, phf, spy attacks

We have selected 10 features :

- duration
- protocol_type
- src_bytes
- num_failed_logins
- num_compromised
- num_file_creations
- num_shells
- mum_access_filr
- is_host_login
- is_guest_login

Individual Feature Set for U2R attacks :

U2R attacks are unauthorized access to local root privileges. As U2R attacks involve with the content of the connection, content features within connection are considered for U2R attacks.

Example U2R attacks: buffer overflow and root kit attacks.

We have selected 6 features :

- num_compromised
- root_shell
- num_root
- mum_file_creations
- num_access_files
- is_host_login

G. Each attack group as a Layer

Individual feature set is selected for each attack group based on its attack method. So each attack group is treated as a layer. The four main attack groups such as DoS, U2R, R2L and Probe layers are trained separately by the set of features selected for the concerned layer. The below Fig.1 shows the overall System Architecture for the Intrusion Detection System.

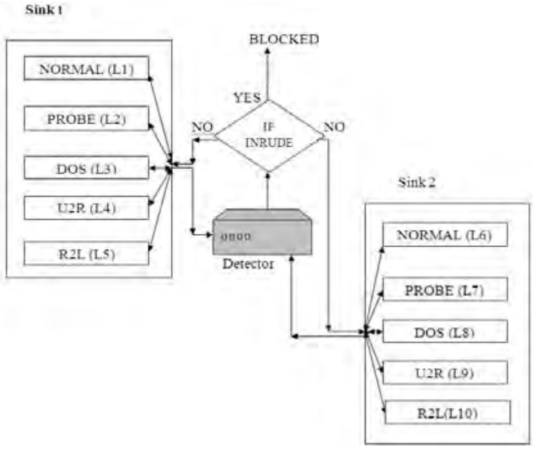

Fig.1 System Architecture with detector

The proposed architecture is designed in such a way that each attack group is working as a layer. When the file is transferred over the network, the file is given to the probe layer. The individual feature set selected for probe layer will monitor the packets and allow the packets to pass though the DoS layer if the packets are found to be legitimate else the packets are blocked in probe layer itself. The process is repeated for next layers such as U2R and R2L attacks The incoming packets are analyzed at each layer based on the features selected at the concerned layer. The packets are blocked in the current layer if it is detected as intrusion, otherwise the packets are allowed to pass through the next layer if it is detected as legitimate packets.

H. Applying Individual Feature Set to Layers

Algorithm :

Step 1: Construct the Network with DoS, U2R, R2L and probe layers

Step 2: Create the packets

Step 3: Forward the packets to centralized scheduler

Step 4 : Find the desired destination

Step 5: Perform feature selection for DoS, U2R, R2L and probe layers

Step 6: Obtain the Individual Feature Set for each layer

Step 7 : Apply the Individual Feature Set to the 4 layers

Step 8: Test the incoming packets and find whether the packets are legitimate or coming from the attacker

Step 9: If the packets are detected as attack, block it in the current layer itself otherwise allow the packets to pass through the next layer

Step 10: Repeat the above steps for the remaining layers.

4. RESULTS & DISCUSSION

Implementation Setup: The proposed system was implemented in Java Environment. The main form is designed using Java Swing class. Browse option is used to select a source file in a Network. The layer is selected in which the source file is to be transferred. The Fig.4 below shows the intrusions that are coming from various port types. The file was transferred through the particular layer. The intruder ports are detected by the system if any. The Fig.2 shows the detection of probe layer attacks.

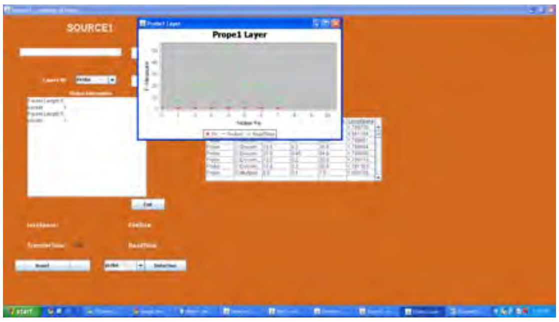

Fig. 2 Detection of Probe layer

The features are selected for DoS Layer, U2R layer, R2L layer and probe layer. During the transmission of the file from client to server, attacks are detected if any and the graph is drawn for each and every layer.

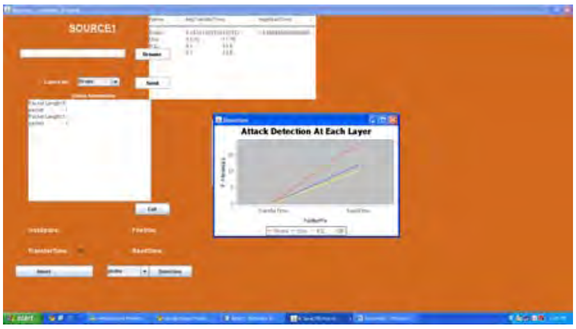

Fig. 3 Attack detection for each layer

Table .4 Comparison of Detection Rate in Percentage

Approach / Method	DoS	U2R	R2L	Probe
Individual Feature Set applied to Layers	**97.8**	**88.2**	**30.2**	**98.6**
K- Means Clustering	96.8	9.83	12.6	87.6
C 4.5 Algorithm	96.0	1.80	08.5	85.5
Decision Tree with PCA	95.0	8.02	06.5	35.0
KNN classifier	92.6	13.0	23.0	38.0
Support Vector Machine	91.6	12.0	22.0	36.7

The network attacks listed in the DoS, U2R, R2L and probe attacks are tested and detected by the Individual Feature Set applied to layer mechanism. The experimental results show that the proposed system achieves high detection rate of network attacks when compared to the other methods such as K- Means Clustering, C 4.5 algorithm, KNN classifier and Support Vector Machine.

5. CONCLUSIOS AND FUTURE WORK

The infrastructure of current networks is inefficient against powerful network attack types such as DoS, U2R , R2L and Probe attacks. So an efficient and accurate intrusion detection system is required to detect these network attack groups. Individual Feature Set is selected for DoS layer, U2R layer, R2L layer and probe layer. The selected features are applied to layers to detect the network attacks. Experimental results show that the proposed system is effective and accurate against various network attack groups. The future model will aim at implementing the same system in ad hoc wireless networks. This kind of system will improve the detection rate of network attacks.

6. REFERENCES

[1] Boughaci, D. Drias, H. Bendib, A. Bouznit, Y.Benhamou, "Distributed Intrusion Detection Framework based on Autonomous and Mobile Agents". International Conference on Dependability of Computer Systems, pp 248 – 255. 2006.

[2] Mrutyunjaya Panda and Manas Ranjan Patra, "Network Intrusion Detection using Naive Bayes", International Journal of Computer Science and Network Security, Vol.7 No.12, pp 258-263, 2007

[3] FuHau Hsu, Fanglu Guo, Tzicker Chiueh, "Scalable Network based Buffer Overflow Attack Detection", Proceedings of ACM/IEEE symposium on Architecture for Networking and Communications Systems, 2006.

[4] C. Sutton and A. McCallum, "An Introduction to Conditional Random Fields for Relational Learning," Introduction to Statistical Relational Learning, University of Massachusetts, USA, 2006.

[5] C. Warrender, S. Forrest, and B. Pearlmutter, "Detecting Intrusions Using System Calls: Alternative Data Models", Proceedings of IEEE Symposium of Security and Privacy (SP '99), pp. 133-145, 1999.

[6] Debar, H., Becker, M, Siboni, "A neural network component for an intrusion detection system", Proceedings of IEEE Computer Society Symposium on Research in Security and Privacy, pp: 240 – 250, 1992.

[7] E. Tombini, H. Debar, L. Me, and M. Ducasse, "A Serial Combination of Anomaly and Misuse IDSes Applied to HTTP Traffic", Proceedings of 20th Annual Computer Security Applications Conference (ACSAC '04), pp. 428-437, 2004.

[8] Meera Gandhi and S.K.Srivatsa, "Detecting and preventing attacks using network intrusion detection systems", International Journal of Computer Science and Security, Vol.2 No.1, 2008.

[9] Paul Dokas V. Kumar and Jaideep Srivastava, "Protecting Against Cyber Threats in Networked Information Systems", Army High Performance Computing Research Center, University of Minnesota, Minneapolis, 2003.

[10] Hyunsang Choi, Heejo Lee and Hyogon Kim, "Fast detection and visualization of network attacks on parallel coordinates", Elsevier Journal of Computers and Security, pp 276-288, 2009.

[11] W. Wang, X.H. Guan and X.L. Zhang, "Modeling Program Behaviors by Hidden Markov Models for Intrusion Detection, Proceedings of International Conference on Machine Learning and Cybernetics (ICMLC'04),vol. 5, pp. 2830-2835, 2004.

[12] Wenke Lee , Salvatore J. Stolfo and Kui W. Mok, "Mining Audit Data to Build Intrusion Detection Models", Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining, New York, 1998

[13] Wenke Lee Salvatore J. Stolfo, "Data mining approaches for Intrusion Detection", Proceedings of the 7th USENIX Security Symposium, San Antonio, Texas, 1998.

[14] Y.S.Wu, B. Foo, Y. Mei, and S. Bagchi, "Collaborative Intrusion Detection System(CIDS): A Framework for Accurate and Efficient IDS," Proc. 19th Ann. Computer Security Applications Conf. (ACSAC '03), pp. 234-244, 2003.

[15] Z. Zhang, J. Li, C.N. Manikopoulos, J. Jorgenson and J. Ucles, "HIDE: A Hierarchical Network Intrusion Detection System Using Statistical Preprocessing and Neural Network Classification", Proc. IEEE Workshop Information Assurance and Security (IAW '01),pp. 85-90, 2001