Hybrid Intrusion Detection System Design for Advanced Threats

Thet Thet Htwe, Dr. Nang Saing Moon Kham
University of Computer Studies, Yangon
moonkhamucsy@gmail.com, hetthethtwe@ucsy.edu.mm

Abstract

Due to the increase in the use of internet and rise in number of network attacks, intrusion detection has become significant research issue. The objective of this work is to propose a hybrid intrusion detection system design to solve the security issue of the today advanced threats. In this research work, both anomaly and misuse detection systems are used so that not only the attacks that is already known but also new form of threats can be detected. Agent technology is used to implement in this system. Each detection engine is autonomous and feedback mechanism is provided so that the system will be modified if new form of threats is detected. This proposed system will give high accuracy in detection rate and reduce the false positive rate.

1. Introduction

Nowadays internet comes in useful for those who want to share information and communicate with each other. As the use of internet keeps growing under its own momentum, the security becomes major issue that needs to be dealt with for every organization. Most of the organizations try to deploy the multiple layer defensive tools within their infrastructure, such as firewalls, intrusion detection system and virus scanner as a defended mechanisms to protect from unauthorized access of their asset. Yet there is no perfect security and network administrators have to handle the newly emerged attacks time after time. Intrusion detection system is one of the mechanisms that can detect activities which violate the security policy of one organization. The workload of the network administrator can be lightened considerably if the detection system becomes effectives.

IDS are usually divided into two groups based on data source used and data analysis methods. According to the information resources used, IDS can be classified into host-based IDS or network-based IDS. According to different data analysis methods, IDSs can be classified as misused or anomaly based IDS.

Different data sources used: Different data such as system log, registry changes, driver loading, etc collected from system are used by host-based IDS. Data collected from network stream and packets are used by network-based IDS.

Different analysis method used: User activities are compared with known signature patterns of attackers in misuse detection. These signatures are stored in database and if matches found labeled as intrusive activities and system generates the alert. Most of today's commercial IDSs are based on misuse detection which can only detect the known attacks. An anomaly detector analyzes a set of characteristics of the monitored system (or network) and identifies activities that deviate from the normal behavior. Any observable behavior of the system can be used to build a model of the normal operation of the system. Anomaly detection has the potential of detecting novel attacks. However anomaly detection suffers from the basic difficulty of defining what is normal. When the system or user behavior varies widely, methods based on anomaly detection tend to produce many false alarms [1].

In this paper, we will intend to propose a system design that will use to solve the security issue of the today advanced threats. In this proposed work, network information resources will be use to identify the malicious activity by deploying hybrid IDS which is combination of both misuse and anomaly. Agent technology is used to implement in this system. Each detection

engine is autonomous and feedback mechanism is provided so that the system will modify if new form of threats is detected. The rest of the paper is organized as follows.

Section 2 will be literature review and study of the intrusion dataset will be mentioned in section 3. Section 4 is brief about the intelligence agent technology. In section 5 proposed system will described. Conclusion and reference will be presented in section 6.

2. Literature review

In this section some of the most relevant researches done on KDD dataset using hybrid network intrusion detection system through years have been highlighted.

As a review on HIDSUR: A Hybrid Intrusion Detection System Based on Real-time User Recognition proposed by Alexandr Seleznyov and Seppo Puuronen in 2000, they try to eliminate the shortcoming of both misuse and anomaly detection approaches through the cooperation of these two detection approaches. As a presentation technique, they used temporal-probabilistic network approach to catch the temporal aspects of user's behavior. In their work features of artificial intelligence in intrusion detection are also introduced. The use of many databases in their design in order to classify and control, management of these database may become difficult while concurrent update occurs [5].

Abdelhamid BELMEKKI and Abdellatif MEZRIOUI (2005) proposed using active agent for intrusion detection and management architecture based on agent technology. In their approach, a set of agents is applied by communicating and cooperating together to improve the enterprise security. The collection and analysis of data is performed by different agents. Decision for countermeasures can be taken either by agent itself or by the core agent. Their architecture combined NIDS and HIDS to minimize the some limitation of both IDSs. Moreover their architecture can improve the efficiency of intrusion detection. They make use of information on resources and organizational structure in the process of intrusion detection management. Distributed architecture based on autonomous and cooperative agents improved efficiency for their system [6].

In 2007, Baojun Zhang and Jiebing Wang proposed hybrid intrusion detection system for complicated network based on autonomous agents. It combines the advantages of misuse and anomaly detection. It uses SVM to construct the anomaly classifier and uses multi-attributes method to abstract attributes form attacks. Their experiment results demonstrate that proposed system has high performance and good practicability [7].

Kiran Dhangar, Deepak Kulhare and Arif Khan in 2013, proposed a hybrid intrusion detection system. Either host-based or network-based IDS capture packets and pass the captured packets to rule matching process with the help of detector agent. The rule matching process checks attacks criteria from database and produce alarm if any type of attack find in the captured packet. This system will suffer increase in positive alarm rate because there is no additional method to reduce alarm rate. Another weak point is that new attack cannot be detected if there are no attack criteria for this new attack in database [8].

3. Study on Intrusion Dataset

3.1. Overall description on NSL-KDD dataset

NSL-KDD is an improved version of KDD99 which is used to survey and evaluate research in intrusion detection. MIT Lincoln Labs has managed a dataset called KDD99, a version of 1998 DARPA dataset. KDD99 consists of large number of redundant and duplicate records. These duplicate records can negative effect on training process of a classifier. In NSL-KDD, these redundant records are eliminated. Thus the NSL-KDD dataset that attract much attention to researchers is its reduced version of KDD dataset. In NSL-KDD both train-set and test-set are provided.

The NSL-KDD dataset consists of 41 features and one class attribute which defines the connection attacks or normal. The overall true classes of the NSL-KDD data set [9], both training and testing are shown in Table 1.

Table 1. NSL-KDD dataset

	Normal	Attack	Total
KDDTrain+	67343	58630	125973
KDDTest+	9711	12833	22544

3.2. Attack types in NSL-KDD

Attacks in NSL-KDD fall into four main categories: DoS, R2L, U2R, and probing.

i. Denial of Service Attack (DOS): is an attack in which the attacker makes some computing or memory resource too busy or too full to handle legitimate requests, or denies legitimate users access to a machine.
ii. User to Root Attack (U2R): is a class of exploit in which the attacker starts out with access to normal user account on the system and is able to exploit some vulnerability to gain root access to the system.
iii. Remote to Local Attack (R2L): occurs when an attacker who has the ability to send packets to machine over a network but who doesn't have an account on that machine exploits some vulnerability to gain local access as a user of that machine.
iv. Probing Attack: is an attempt to gather information about a network of computers for the apparent purpose of circumventing its security controls.

The dataset contain a total number of 24 training attack types, with additional 14 attacks types in the test data only [2].

List of attacks and attacks distribution in NSL-LDD are described in Table 2 and Table 3.

Table 2. List of attacks in KDD Train+ and Test+

	KDDTrain+	KDDTest+
DOS	neptune, teardrop, land, smurf, pod, back	neptune, teardrop, land, smurf, pod, back, processtable apache2, mailbomb, udpstrom
PROBE	satan, nmap, ipsweep portsweep, mscan	satan, nmap, ipsweep portsweep, mscan, saint
U2R	buffer_overflow, rootkit, perl, Loadmodule	buffer_overflow, rootkit, perl, loadmodule, ps, sqlattack,xterm
R2L	guess_passwd, imap, warezmaster, phf, warezclient, spy, sendmail, multihop, ftp_write	guess_passwd, imap, warezmaster, phf, warezclient, spy, sendmail, multihop, ftp_write xlock, snmpguess, xsnoop, httptunnel, Snmpgetattack, worm

Table 3. Attack distribution

KDDTrain+		KDDTest+	
Normal	67343	Normal	9711
Dos	46538	Dos	7387
Probe	11656	Probe	2421
U2R	52	U2R	69
R2L	384	R2L	2867
Total	125973	Total	22544

The majority of attack patterns are from Dos attack group and very few U2R attack contain in dataset. It is important to note that

the test data is not from the same probability distribution as the training data and it includes specific attack types not in the training data which make the task more realistic. Some intrusion experts believe that most novel attacks are variants of known attacks and the signature of known attacks can be sufficient to catch novel variants.

3.3 NSL-KDD features

NSL-KDD consists of same features as KDD 99 but it does not include redundant records in the train set and there are no duplicate records in the test sets. There are total numbers of 41 features and one class attribute. Features can be classified into three groups: basic features, traffic features and content features.

i. Basic features: these features are taken out from individual TCP connection.

ii. Traffic features: these features are derived based on connections in the past two seconds.

   a) The "same host" features examine only the connections in the past two seconds that have the same destination host as the current connection, and calculate statistics related to protocol behavior, services, etc.

   b) The "same service" features examine only he connections in the past two seconds of that have the same service as the current connection.

The features "same host" and "same service" are together called time-based traffic features. Some probing attacks scan the host's port more than two seconds, for example once per minute. Therefore connection records were also sorted by destination host, and features were constructed using a window of 100 connections to the same host instead of a time window. This set of features is called host-based traffic features.

iii. Content features: Unlike most of the DOS and probing attacks, there appear to be no sequential patterns that are frequent in records of R2L and U2R attacks. This is because the DOS and probing attacks involve many connections to some host in a very short period of time, but the R2L and U2R attacks are embedded in the data portions of packets, and normally involve only single connection. To look for suspicious behavior in the data portions, such as failed login attempts, some features are added by domain knowledge. These features are called content features [3].

In Table 4, 41 features which consist in dataset are listed and some of the example derived features are categorized in Table 5.

Table 4. Features in NSL-KDD

1.	duration	22.	is_guest_login
2.	protocol	23.	count
3.	service	24.	srv_count
4.	falg	25	serror_rate
5.	src_byte	26.	srv_ serror_rate
6.	dst_byte	27.	rerror_rate
7.	land	28.	srv_ rerror_rate
8.	wrong_fragment	29.	same_srv_rate
9.	urgent	30.	diff_srv_rate
10.	hot	31.	Srv_diff_host_rate
11.	num_failed_login	32.	dst_host_count
12.	logged_in	33.	dst_host_srv_count
13.	num_compromised	34.	dst_host_samesrv_rate
14.	root_shells	35.	dst_host_diff_srv_rate
15.	su_attempted	36.	dst_host_same_src_port_rate
16.	num_root	37.	dst_host_srv_diff_port_rate
17.	num_file_creation	38.	dst_host_serror_rate
18.	num_shells	39.	dst_host_srv_serrateror_rate
19.	num_access_file	40.	dst_host_rerror_rate
20.	num_outbound_cmds	41.	dst_host_srv_rerror_rate
21.	is_host_login		

Table 6. Categories of derived features

Features Categories	Feature Name	Description	Type
Basic Features	duration	Length of the connection	C
	protocol	Type of protocol	D
	service	Network service on the destination	D
Content Features	num_failed_logins	Number of failed login attempts	C
	num_compromised	Number of compromised condition	C
	su_attempted	1 if "su root" command attempt; 0 otherwise	D
Traffic features	serror_rate	%of connection that have "syn"	C
	rerror_rate	% of connection that have "REJ"	C
	srv_diff_host_rate	%of connections to different hosts	C

D – Discrete, C – Continuous

3. Agents Technology overview

AI is the part of computer science concerned with designing intelligent computer systems, that is, computer systems that exhibit the characteristics associated with intelligence in human behavior such as understanding language, learning, reasoning and solving problems [4]. Intelligent agent technology becomes great interested area of the research and new application development nowadays.

In agent based IDS there is no central station, therefore no central point of failure. Overcoming the deficiency of centralized structure is the major reason for using agents in the intrusion detection field. The agents usefulness includes also reduction of network load overcoming of the network latency and support for disconnected operation. Agent typically possesses several or all of the following characteristics which are very attractive for network.

Most desirable characteristic of agents are:

i. Autonomy: is the ability of an agent to operate without direct intervention of humans or other agents and to have some kind of control based on its internal state and/or external environment.
ii. Socialability: is the capability of an agent to integrate itself in a large environment populated by a society of agents with which the agent has to exchange messages to achieve purposeful actions. This property is satisfied even when system have to share their knowledge and mental attitudes (beliefs, goals, desires, etc.)
iii. Proactivity: is the ability of an agent to anticipate situations and change its course of action. It is a relevant property which occurs in the network and system management in order to avoid disastrous effects on global performance. Indeed, proactive agents are capable of exhibiting goal-direct behaviors by taking some initiatives.
iv. Reactivity: this kind of behavior means that the agent reacts in real-time to changes that occur in its environment.
v. Adaptability: is the ability of an agent to modify its behavior over time to fulfill its problem-solving goals.
vi. Intelligence: the term intelligent means that the agent is able to exhibit a certain level of intelligence priority, ranging from predefined actions (planning) up to self learning (define new actions) [10].

Communication between agents permits to collect information which aid what behavior they should exhibit when attacks occurs. The task of administration becomes easier if we give the more autonomy to agent in the control of overall intrusion detection system.

4. Proposed system design

The architecture of the hybrid intrusion detection system is proposed here. This system

is composed of information gathering (sniffer) agent, anomaly agent, misuse agent, monitor agent and control agent. Intelligent agent concept will be applied for information collection for input stage. The use of misuse detection alone cannot achieve the satisfactory result that we want because it cannot detect new threats. So we make use of both types of IDS, anomaly and misuse in order to eliminate the weakness of each other. Each detection system is autonomous and will produce the alerts.

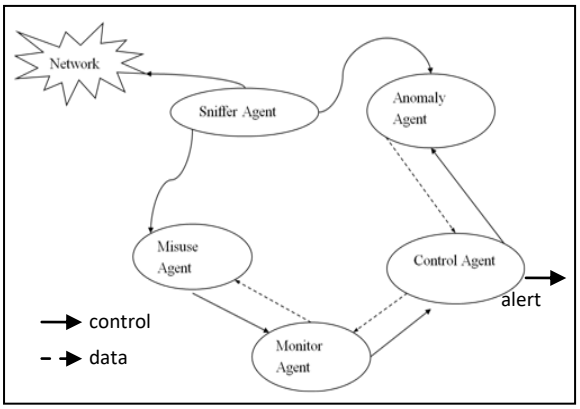

Figure 1. Hybrid Intrusion Detection System Design for Advanced Threats

Sniffer agent: Sniffer agent will collect the network traffic and send collected data to anomaly and misuse agents.

Anomaly agent: Anomaly agent will perform preprocessing, attribute selection and decision making for detection of threats upon incoming network traffic. Alerts will be generated and sent to control agent when the system detect the threats.

Misuse agent: Misuse agent will analyze the data received from sniffer agent and will produce alerts if any match upon its rules. It is also make a rule update depend on available information from monitor agent.

Monitor agent: The task of monitor agent is to listen the information from the control agent for rules update and control information, on behalf of misuse agent.

Control agent: The control agent will listen the alerts generated by the misuse and anomaly agents. Then it will produce alerts, rules and control information in respect of agent necessary for detection of new threats.

The research area for intrusion detection continues to active because it is difficult to detect unknown attacks make the system to generate so many false positive and great amount of alert volume that cause decrease in performance and accuracy of entire system. If the system is intelligence enough, we can lessen the workload of the network administrator. Moreover we can detect newly threats if the system can thought and has the intelligence like human being. This proposed architecture of hybrid IDS will try to diminish these problems.

5. Conclusion

In this paper, we introduce the hybrid intrusion detection system design based on intelligent agent technology to detect advanced threats. In order to get the system intelligent detection, agent terminology will be deployed and appropriate feature selection process will be performed to detect advanced threats. As a future work we will implement this proposed system and then evaluation will be made to this system. Detection rate and false positive rate will be used as performance criteria. Deploying both types of network IDS and agent technology, as an expected result, our proposed system design will give the better overall accuracy improvement and also reduce in false positive rate.

References

[1] S.Dua and X.Du, Data Mining and Machine learning in Cyber security, Auebach Publications, 6000 Broken Sound Parkway NW, Suite 300 Boca Raton, FL 33487-2742, 2011.

[2] S.Bahl and S.K.Sharma, "Improving Classification Accuracy of Intrusion Detection System using Feature Subset Selection", 2015 IEEE, DOI 10.1109.

[3] https://kdd.ics.uci.edu/database/kddcup99/task.html.

[4] M.Wooldridge, "An Introduction to MultiAgent Systems", Second Edition, John Wiley & Sons Ltd, The Atrium,

Southern Gate Chichester, West Sussex, PO 19 8SQ, United Kingdom, 2009.

[5] A.Seleznyov and S.Puuronen, "HIDSUR: A Hybrid Intrusion Detection System Based on Real-Time User Recognition", Proceeding of the 11th International Workshop on Database and Expert Systems Applications (DEXA'00), 2000 IEEE.

[6] A.BELMEKKI and A.MEZRIOUI, "Using Active Agent for Intrusion Detection and Management", International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce, 2005 IEEE.

[7] B.Zhang X.Pan, and J.Wang, "Hybrid Intrusion Detection System for Complicated Netwrok", Fourth International Conference on Fuzzy Systems and Knowledge Discovery, 2007 IEEE.

[8] K.Dhangar, D.Kulhare, and A.Khan, "A Proposed Intrusion Detection System", Journal International Journal of Computer Applications, March 2013, Volume 65-No.23.

[9] https://iscxdownloads.cs.unb.ca/is cxdownloads

[10] K.Boudaoud, H.Labiod, R.Boutaba, and Z.Guessoum," Network Security Management with Intelligent Agents", Sixth International Conference on Intelligence in Networks Intelligence, Vienna, Austria, 2000.

[11] M.Tavallaee, E.Bagheri, W.Lu and Ali A. Ghorbani, "A Detail Analysis of the KDD CUP 99 Data Set", IEEE Symposium on Computational Intelligence in Security and Defense Applications, 2009.

[12] S.Choudhury and A.Bhowal,"Comparative Analysis of Machine Learning Algorithms along with Classifiers for Network Intrusion Detection", International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), India, May 2015. pp.89-95.