Approaches and Data Processing Techniques for Intrusion Detection Systems

Pakkurthi Srinivasu
Associate Professor, Dept. of CSE,
Anil Neerukonda Institute of Technology and Sciences,
Sangivalasa,Vishakapatnam Dist, Andhra Pradesh, India

P.S. Avadhani
Professor, Dept. of CS & SE,
Andhra University,
Visakhapatnam Dist, Andhra Pradesh, India

Vishal Korimilli
Final Year, Dept. of CSE
Anil Neerukonda Institute of Technology and Sciences,
Sangivalasa,Vishakapatnam Dist,
Andhra Pradesh, India

Prudhvi Ravipati
Final Year, Dept. of CSE
Anil Neerukonda Institute of Technology and Sciences,
Sangivalasa,Vishakapatnam Dist,
Andhra Pradesh, India

Abstract
The process of monitoring the events occurring in a computer system or network and analyzing them for sign of intrusions is known as intrusion detection systems (IDS).In this paper an overview of types of attacks, IDS components, and classifications of IDS is briefly given. Two approaches from the classification of IDS are also presented. They are anomaly-based detection and misuse-based detection. Anomaly-based detection approach is extremely powerful and novel tool which identifies anomalies as deviations from "normal" behavior and automatically detects any deviation from it. Misuse-based detection approach explicitly defines the attack behavior and classifies all events matching these specifications as attack. Data processing techniques for intrusion detection systems are also presented.

Key words:
Intrusion Detection System, Type of attacks, Anomaly based intrusion detection, Misuse based intrusion detection, Network Security, Data Processing Techniques.

1. INTRODUCTION

Information has become an organizations most precious asset. Organizations have become increasingly dependent on it since more information is being stored and processed on network-based systems. The wide spread of e-commerce has increased the necessity of protecting the system to a very high extend. Confidentiality, integrity and availability of information are major concerns in the development and exploitation of network based computers systems. Intrusion detections Systems can detect, prevent and react to the attacks.

The concept of intrusion detection was first introduced by Anderson to complement conventional computer security approaches in 1980[1]. Anderson defined an intrusion attempt or a threat to be a deliberate unauthorized attempt to

- access information,
- manipulate information, or
- render a system unreliable or unusable.

Since then, several techniques for detecting intrusions have been studied[2] Intrusions are actions that attempt to bypass security mechanisms of computer systems and they are any set of actions that threatens the confidentiality, integrity and availability of network resource. The meaning of the set of properties is:

Confidentiality: Information is not made available or disclosed to unauthorized individuals, entities or processes.
Integrity: Data has not been altered or destroyed in an unauthorized manner.
Availability: A system or system resource that ensures that it is accessible and usable upon demand by an authorized system user. It is one of the core characteristics of a secure system.

2. TYPES OF ATTACK

DARPA Intrusion Detection Evaluation program by MIT Lincoln Labs (MIT Lincoln Laboratory) acquired nine weeks of raw TCP dump data. The data set contains 24 attack types. These attacks fall into four main categories:
Denial-of-Service (DoS) attacks
In this type of attacks an attacker makes some computing or a memory resource too busy or too full to handle legitimate request, or denies legitimate users access to a

Manuscript received December 5, 2009
Manuscript revised December 20, 2009

machine. Examples are Apache2, Back, Land, SYN fold, Ping of Death, Smurf etc.,

Remote to Local (R2L) attacks
In this type of attacks an attacker who does not have an account on a remote machine sends packets to that machine over a network and exploits some vulnerability to gain local access as a user of that machine. Examples are Ftp_write, Guest, IMap, Named, Sendmail, etc,

User to Root (U2R) attacks
In this type of attacks an attacker starts out with access to a normal user account on the system and is able to exploit system vulnerabilities to gain root access to the system. Examples are Eject, Loadmodule, ps, Xterm, Perl, Fdformat.

Probing attacks
In this type of attacks an attacker scans a network of computers to gather information to find known vulnerabilities. An attacker with a map of machines and services that are available on a network can use this information to look for exploits. Examples are Ipsweep, Mscan , Saint, Satan, Nmap

3. INTRUSION DETECTION SYTEMS

3.1 Terminology of IDS

Intrusion detection is a growing field. This section describes the meaning of some of the important ID concepts as they may appear in this paper [11].

Attack
An action conducted by one adversary, the intruder, against another adversary, the victim. The intruder carries out an attack with a specific objective in mind. From this perspective an administrator is responsible for maintaining a system, an attack is a set of one or more events that may have one or more security consequences. From the perspective of an intruder, an attack is a mechanism to fulfill an objective.

Intrusion
It is a common synonym for the word “attack”; more precisely, a successful attack. In this paper, we often use the term intrusion to include attack.

Intruder
Intruder is a person who carries out an attack. Attacker is a common synonym for intruder. The words attacker and intruder apply only after an attack has occurred. A potential intruder may be referred to as an adversary. Since the label of intruder is assigned by the victim of the intrusion and is therefore contingent on the victim's definition of encroachment, there can be no ubiquitous categorization of actions as being intrusive or not.

Vulnerability
A feature or a combination of features of a system that allows an adversary to place the system in a state that is contrary to the desires of the people responsible for the system and increases the probability or magnitude of undesirable behavior in or of the system.

Incident
It is a collection of data representing one or more related attacks. Attacks may be related by attacker, type of attack, objectives, sites, or timing.

Exploit
It is the process of using a vulnerability to violate a security policy. A tool or defined method that could be used to violate a security policy is often referred to as an exploit script.

False negative
An event that the IDS fail to identify as an intrusion when one has in fact occurred.

False positive
It is an event, incorrectly identified by the IDS as being an intrusion when none has occurred.

3.2 ID System Components

The functionality of IDS can be logically distributed into three components: sensors, analyzers, and a user interface.

Sensors
Sensors are responsible for collecting data. The input for a sensor may be any part of a system that could contain evidence of an intrusion. Example types of input to a sensor are network packets, log files, and system call traces. Sensors collect and forward this information to the analyzer.

Analyzers
Analyzers receive input from one or more sensors or from other analyzers. The analyzer is responsible for determining if an intrusion has occurred. The output of this component is an indication that an intrusion has occurred. The output may include evidence supporting the conclusion that an intrusion occurred. The analyzer may provide guidance about the actions to be taken to counter the attempt or intrusion.

User interface
The user interface to IDS enables a user to view output

from the system or control the behavior of the system. In some systems, the user interface may equate to a "manager," "director," or "console" component.

3.3 Classification of ID Systems

Although every IDS can be conceptually viewed as having a sensor, an analyzer, and a user interface, the types of data examined and the types of data generated by a particular IDS may vary significantly. ID systems can be classified into one of the following categories based on the types of data they examine:

a) Network-based.

Network-based IDS detects intrusions by reading all the incoming packets and trying to find suspicious patterns. It monitors multiple hosts and investigates network traffic by connecting with network devices like hub, switch, and router. It also inspects the outgoing or local network traffic as well to identify the internal intruders (individuals misusing their privileges)

b) Protocol-based.

Protocol-based IDS is installed on a server and it analyzes the protocol that is used by the server. It sits at the front end of the server, monitoring and analyzing the dynamic behavior and state of the communication protocol between a connected device (a user/PC or system) and server.

c) Application protocol-based.

Application protocol-based IDS normally sit between groups of services/processes monitors and analyze the behavior and state of application protocol in use by the system between two connected devices.

d) Anomaly-based.

Anomaly-based IDS detects computer intrusions by monitoring system activity and classifying it as either normal or anomalous. It attempts to characterize normal operation, and tries to detect any deviation from normal behavior. The advantage of anomaly based detection is a relatively high detection rate for new types of intrusions.

e) Misuse-based

It is also called as Signature based IDS. It performs the simple process of matching patterns corresponding to a known attack type. It is also known as pattern-based IDS. The main advantage of signature-based IDS is that it has a relatively low rate of false alarms, which means it has relatively high precision. And the main disadvantage of these IDS is that the detection rate of attacks is relatively low, because attacker will try to modify the basic attack signature in such a way that it will not match the known signatures of that attack and it cannot detect a new attack for which a signature is not yet installed in the database.

f) Host-based.

Host based IDS identifies intrusions by analyzing system calls, application logs, file system modifications, and other host activities. It is installed on individual systems to examine the internal interfaces. It monitors the operating system and writes data to log files and/or triggers alarms.

g) Hybrid-based.

Hybrid-based IDS combines one or more approaches. Host agent data is combined with network information to form a comprehensive view of the network.

3.4 Functions of IDS

Intrusion Detection Systems performs a variety of functions [3]:

- Monitoring and analysis of user and system activity
- Auditing of system configurations and vulnerabilities
- Assessing the integrity of critical system and data files
- Recognition of activity patterns reflecting known attacks
- Statistical analysis for abnormal activity patterns
- Operating System audit trail management, with recognition of user activity reflecting policy violations.

4. IDS ANALYSIS APPROACHES

Intrusion detection systems must be capable of distinguishing between normal and abnormal user activities, to discover malicious attempts in time. The activity of the user in IDS may fall into any one of the behaviors specified in figure 1.

Figure 1: Behavior of the user in the system [4]

In order to classify actions, intrusion detection systems uses analysis approach.An analysis approach is a method used by IDS to determine whether or not an intrusion has occurred. There are two major categories of analysis approaches:

4.1 Anomaly-based Detection Approach

It identifies any unacceptable deviation from expected behavior. Expected behavior is defined, in advance, by a profile developed manually or automatically. An automatically developed profile is created by software that collects and processes characteristics of system behavior over time and forms a statistically valid sample of such behavior. Some of these deviations do not require further examination while some do. An anomaly might include

- users logging in at strange hours
- unexplained reboots or changes to system clocks
- unusual error messages from mailers, daemons, or other servers
- multiple, failed login attempts with bad passwords
- unauthorized use of the 'su' command to gain UNIX root access
- users logging in from unfamiliar sites on the network

Normal behavior patterns are useful in predicting both user and system behavior. Anomaly detectors construct profiles that represent normal usage and then use current behavior data to detect a possible mismatch between profiles and recognize possible attack attempts.
In order to match event profiles, the system is required to produce initial user profiles to train the system with regard to legitimate user behaviors. There is a problem associated with profiling: when the system is allowed to "learn" on its own, experienced intruders (or users) can train the system to the point where previously intrusive behavior becomes normal behavior. An inappropriate profile will be able to detect all possible intrusive activities. Furthermore, there is an obvious need for profile updating and system training which is a difficult and time consuming task.
Given a set of normal behavior profiles, everything that does not match the stored profile is considered to be a suspicious action. Hence, these systems are characterized by very high detection efficiency (they are able to recognize many attacks that are new to the system), but their tendency to generate false alarms is generally a problem.
Advantages of this anomaly detection method are: possibility of detection of novel attacks as intrusions; anomalies are recognized without getting inside their causes and characteristics; less dependence of IDSs on operating environment (as compared with misuse/ signature-based systems); ability to detect abuse of user privileges.
The biggest disadvantage of this method is a substantial false alarm rate as the system usage is not monitored during the profile construction and training phases. Hence, all user activities skipped during these phases will be illegitimate.

4.2 Misuse-based detection Approach

Misuse-based detection approach identifies patterns corresponding to known attacks. This includes passive protocol analysis which is the use of sniffers in promiscuous. It also includes signature analysis which is the interpretation of a series of packets (or a piece of data contained in those packets) that are determined, in advance, to represent a known pattern of attack.
Signature-based IDS performs simple pattern matching process. It is also known as pattern-based IDS. The main advantage of signature-based IDS is that it has a relatively low rate of false alarms, which means is has relatively high precision.
The main disadvantage of this method is, it is not capable of detecting unknown, novel attacks to over come this problem the system has to maintain continuous update of the attack signature database.

5. DATA PROCESSING TECHNIQUES

Depending on the type of approach followed in intrusion detection, various techniques are employed for data processing. Some techniques are described briefly:

5.1 Statistical analysis approach:

This approach is a frequently used method for data processing. The user or system behavior (set of attributes) is expressed as number of variables over time. These variables include user login, logout, number of files accessed in a period of time, usage of disk space, memory, CPU etc. The frequency with which these variables are updated varies from a few minutes to one much higher time (for example month). The system stores mean values for each variable. Executing predefined threshold values used for detecting intrusions. These statistical methods are often used in

implementations of normal user behavior profile-based Intrusion Detection Systems.

5.2 Neural Networks:

The essence of neural networks is to learn the behavior of actors in the system (E.g. Users, daemons etc). Neural networks use its learning algorithms to learn about the relationship between input and output vectors and to generalize them to extract new input/output relationships. The advantage of using neural networks over statistics is its ease in expressing and learning nonlinear relationships between variables. Experiments using neural networks resulted that the behavior of super user (roots) are predictable.

5.3 Fuzzy Logic

The essence of Fuzzy logic is to handle variations in degree of truth ness or falsity of attribute values observed during the analysis of data. This involves designing of Fuzzy sets and Fuzzy logic to handle such data. Advantage of Fuzzy sets over Crisp sets is that the former permits the gradual assessment of the membership of elements in a set while the latter allows only two membership function values (0 or 1) in the interval [0, 1]. This approach enhances the classification computational efficiency of Intrusion detection systems.

5.4 Machine learning:

This is an artificial intelligence technique. This approach stores the user-input stream of commands in a vectorial form and is uses those as reference for normal user behavior profiles. These Profiles are then grouped in a library of user commands that share common characteristics [6]. It allows computers to learn based on data, such as from sensor data or databases. A major focus of machine learning research is to automatically learn to recognize complex patterns and make intelligent decisions based on data. Hence, machine learning is closely related to fields such as statistics, probability theory, data mining, pattern recognition, artificial intelligence, adaptive control, and theoretical computer science.

5.5 Data mining:

It refers to a set of techniques that use the process of extracting previously unknown but potentially useful data from large stores of data. Data mining methods helps in processing large system logs (audit data). Decision Trees [5] is one of the fundamental data mining technique used in intrusion detection. Decision tree models allow one to detect anomalies in large databases. Another technique refers to segmentation, allowing extraction of patterns of unknown attacks [7]. This is done by matching patterns extracted from a simple audit set with those referred to warehoused unknown attacks [8]. A typical data mining technique is associated with finding association rules that allows one to extract previously unknown knowledge on new attacks [9] or builds on normal behavior patterns. Data mining methods helps in reducing the rate of false alarms generated often by anomaly detection by correlating data related to alarms with mined audit data [10]. Data mining commonly involves four classes of task:

Classification – It arranges the data into predefined groups. Common algorithms include Nearest Neighbour, Naive Bayes Classifier and Neural Networks.

Clustering – It is like classification but the groups are not predefined, so the algorithm tries to group similar items together.

Regression - Attempts to find a function which models the data with the least error. A common method is to use Genetic Programming.

Association rule learning - Searches for relationships between variables.

5.6 Expert systems:

These systems work on a previously defined set of rules describing an attack. All security related events incorporated in an log files (audit data) are translated in terms of if-then-else rules.

5.7 Colored Petri Nets:

This approach is often used to generalize attacks from expert knowledge bases and to represent attacks graphically.

5.8 State-transition analysis:

It describes an attack with a set of goals and transitions that must be achieved by an intruder to compromise a system.

6. CONCLUSION AND FUTURE WORK

Intrusion Detection is one of the major concerns in any computer networks environment. Many techniques are in

use presently. IDS tools are working in conjunction with other information security tools, such as firewalls and allow for the complete supervision of all network activity. It is very likely that IDS capabilities will become core capabilities of network infrastructure such as routers, bridges and switches and operating systems. In future we would like to find out how data mining, fuzzy logic and genetic algorithms can help to improve intrusion detection and most of all anomaly detection. For that purpose we have presented two approaches for identification of intrusion in network. Data mining , fuzzy logic, genetic algorithms along with neural networks will aid an analyst to distinguish attack activity from common every day traffic on the network.

References

[1] Anderson P, "Computer Security Threat Monitoring and Surveillance". Technical report, James P Anderson Co., Fort Washington, Pennsylvania, April 1980.

[2] Aurobindo Sundaram, "An Introduction to Intrusion Detection ",

[3] http://www.cc.gatech.edu/~wenke/ids-readings/ Intrusion-Detection-Intro.ps.gz "An Introduction to Intrusion Detection & ASSESSMENT", ICSA, Inc. http://www.icsalabs.com/icsa/ docs/html/ communities /ids/whitepaper/ Intrusion1.pdf

[4] Anitha K. Jones, Robert S. Sielken, "Computer System Intrusion Detection: ASurvey". http://www.cs.virginia.edu /~jones/IDS-research/Documents/jones-sielken-survey-v1 1.pdf

[5] Wei Fan, Mathew Miller, Salvotare J, Wenke Lee, Philip K. Chan, **"**Using artificial anomalies to detect unknown and known network intrusions". In Proceedings of the First IEEE International Conference on Data Mining, San Jose, CA, November 2001.

[6] Jack Marin, Daniel Ragsdale, John Surdu, "A Hybrid Approach to the Profile Creation and Intrusion Detection", DARPA Information Survivability Conference and Exposition (DISCEX II'01), Volume-I, June, 2001.

[7] Wenke Lee ,Salvatore J. Stolfo , Kui W. Mok, "Adaptive Intrusion Detection: a Data Mining Approach", December 2000, http://www.cc.gatech.edu ~wenke/ papers /ai_review.ps

[8] Wenke Lee, Rahul A. Nimbalkar, Kam K. Yee, Sunil B. Patil, Pragneshkumar H. Desai, Thuan T. Tran, Salvatore J. Stolfo "A data mining and CIDF based approach for detecting novel and distributed intrusions" . Lecture Notes In Computer Science; Vol.1907. Proceedings of the Third International Workshop on Recent Advances in Intrusion Detection**.** Springer,2000,pp49-65.

[9] Tim Bass, "Intrusion Detection Systems and Multisensor Data Fusion" Communication of the ACM, Vol. 43, April 2000, pp. 99-105.

[10] Stefanos Manganaris, Marvin Christensen, Dan Zerkle, and Keith Hermiz, "A data mining analysis of RTID alarms". Computer Networks, Volume 34, October 2000, pp. 571-577.

[11] Julia Allen, Alan Christie, William Fithen, John McHugh ,John McHugh, Jed Pickel, Ed Stoner, "State of the Practice of Intrusion Detection Technologies" January 2000 Technical Report Networked Systems Survivability program, Carnegie Mellon Software Engineering Institute, Pittsburgh.

Prudhvi Ravipati, pursuing his Final Year Computer Science and Engg., in ANITS, Visakhapatnam , AP, INDIA. His interests include Data Mining, Soft Computing, Neural Networks and Fuzzy Logic. He is also a Student Member of IEEE. He is very active in fundamental research activities in the CSE Dept. of ANITS.

Pakkurthi Srinivasu received his M.Tech(CST) from Andhra University , Visakhapatnam, Andhra Pradesh, India. Presently he is working as Associate Professor in Computer Science and Engineering in Anil Neerukonda Institute of Technology and Sciences, Visakhapatnam Dist, AP, India. His research area includes Intrusion Detection, Network Security, Neural network, Data mining and fuzzy logic.

Prof. P.S.Avadhani did his Masters Degree and PhD from IIT, Kanpur.
He is presently working as Professor in Dept. of Computer Science and Systems Engineering in Andhra University college of Engg., in Visakhapatnam. He has more than 50 papers published in various National / International journals and conferences. His research areas include Cryptography, Data Security, Algorithms, and Computer Graphics.

Vishal Korimilli, pursuing his Final Year Computer Science and Engg., in ANITS, Visakhapatnam , AP, INDIA. His interests include Data Mining, Soft Computing and Intrusion Detection applications. He is also a Student Member of IEEE. He is very active in fundamental research activities in the CSE Dept. of ANITS.