AN EFFICIENT MECHANISM FOR INTRUSION DETECTION AND PREVENTION SYSTEM

S. Venkateswaralu1, Sasi Bhanu J 2, Grace Priscilla D3, Bala Krishna Kamesh D4

1 Professor, CSE Dept, KLUniversity, Andhra Pradesh, India, somu23@yahoo.com
2 Asst Professor, CSE Dept, KLUniversity, Andhra Pradesh, India, sweetsasi2002@yahoo.co.in
3 M.Tech Student, CSE Dept, KLUniversity, Andhra Pradesh, India, prici.diyya@gmail.com
4 Asst Professor, F.E.D, KL University, Andhra Pradesh, India, kameshdbk@kluniversity.in

Abstract

Intrusion detection systems (IDSs) are used to detect attacks. Intrusion detection systems gather network traffic information on computer system and then use this information to secure the network. Intrusion Detection systems are many types. Misuse –based and anomaly-based are the types of Intrusion detection systems. Misuse –based IDS only detect known attacks, anomaly based IDS detect the new attacks. Now days data mining methods have gained importance in addressing network security issues, including network intrusion detection. This paper proposes a new method to detect intrusions. The proposed method uses the super vector machine and rough set theory to detect the intrusions. Rough Set Theory (RST) can be used to pre process the data and reduce the attributes. Next, the attributes were selected by RST will be sent to the Super Vector Machine(SVM). The incremental SVM and rough set theory can be combined to improve the performance and decrease the sample set.

Index Terms: Intrusion detection, Data mining, Network security, Super Vector Machine, Rough set theory.

1. INTRODUCTION

Security is becoming an important issue because the Internet applications are growing. The present security technologies focusing on firewall, access control, encryption etc, but system security also important. This can be enhanced by Intrusion detection. Intrusion detection is the process of checking the events in network or system and detects the intrusions. The number of attacks also increased, to reduce the attacks Intrusion detection systems can be used. There are several numbers of attacks like DDoS attacks.

DDoS attack contains two stages, the first stage is Zombie, and Zombie is a computer used by the intruder in the Internet and install attack tools in these comprised systems. The software is to send dummy traffic designated toward the victim. Many number of web sites like Yahoo, CNN are victims of DDoS Attacks.

Generally there are two types of approach taken toward intrusion detection: Anomaly-based detection And misuse-based detection. Misuse-based detection analyzes the system activities, next it finds the matches between these activities. The advantage of this method is very efficient and quickly identifies the intrusions. But the drawback is it can only detect the known attacks. Second one anomaly–based detector [1] detects the behaviors on a system network that is not normal. Advantage of this method is, it detect the attacks even detail information of attack does not known. But the drawback is it requires a large set of training data.

Our proposed method uses super vector machine with rough set theory to detect intrusions. The proposed method is very efficient method.

Section 2 presents the related work. In section 3, Intrusion detection Techniques. In section 4, Approaches in IDS. In section 5, The Methodology. In section 6, Conclusion.

2. RELATED WORK

Misuse-based detection
It analyzes the system activities, next it finds the matches between these activities. The advantage of this method is very efficient and quickly identifies the intrusions. But the drawback is it can only detect the known attacks.

Anomaly-based detection
It detects the behavior s on a system network that is not normal. The advantage of this method is k it will detect the attacks even detailed information does not know. But the
drawback is it requires large set of training data. NIDS and HIDS [3] also have some drawbacks.

3. INTRUSION DETECTION TECHNIQUES

3.1 Statistical models
These techniques contain mean and standard deviation.

3.2 Markov process model
It analyzes the state transitions for each system call.

3.3 Rule-based algorithm
In rule based algorithm RIPPER [5, 6] algorithm uses the Classifications by creating a list of rules from a set of training data.

3.4 Data Mining Techniques
Data mining techniques uses the large data sets collected by a system. There are many data mining techniques in intrusion detection system.

3.5 Immune system approach
In this approach applications are modeled in terms of call sequences [3], [7].

4. APPROACHES IN IDS

4.1 Artificial neural network
ANN is composed of simple processing nodes and connections between them. Here a sub set of nodes will act as Input nodes and another subset of nodes act as output nodes. Back-propagation and feed-forward algorithms are used for intrusion detection [9].

4.2 Super vector Machine
SVM method solves the problems of learning, Classification and regression. Support vectors are subset of training data. SVM uses mapping of the input data into a high-dimensional feature space to support Vector Machines defined by a kernel function.

4.3 Evaluation Datasets
The datasets can be used for testing and evaluating different anomaly detection methods. The four broad classes of attacks are listed below.

4.3.1 Distributed Denial of service (DDoS)

Some Types of DDoS Attacks

Spoofing Attack
In spoofing attack, the attacker complicates the network by creating routing loop, attracting or replaying the routing information.

Selective Forwarding Attack
In selective forwarding attack, once the attacker is included, the attacker can randomly choose to forward certain packets and drop others a variation of this attack is when the malicious node only drops packets coming from a specific source node. Such attacks are most difficult to detect.

Flooding Attack
The primary aim of flooding attacks is to cause exhaustion of resources on victim system. This process is analogous to TCP SYN attacks where, attacker sends many connection establishment requests, forcing the victim to store state of each connection request.

De-synchronization Attack
In this attack Modifications can be occur in control flags or sequence numbers are usually made.

TCP Reset Attack
Here the attacker sends a fake TCP RESET packet to the victim. Then it causes the victim to unknowingly terminate its TCP connection.

4.3.2. Probe attack
In Probe attack attacker scans a network to collect information or finds the weakness of the machines.

Types of probe Attacks:
1. IPsweep
   Identifies active machines.

4.3.3. User to Root attack
In User to root attack the attacker access a normal account so that the attacker can perform unauthorized actions on the network.

Types of U2R
1. Fdformat
   It sends the overflow data and it affects the root shell.

4.3.4. Remote to Local
In this attack the attacker sending packets to a machine over a network, and then harm the machines. The attacker can access a normal account.

Types of R2L
1. Dictionary
   It gains user access and it services on Telnet, rlogin.

5. THE METHODOLOGY
In the below flowchart is intrusion detection method First, data pre-processing and data discretion are utilized for data arrangement. The RST can be used to find useful features.

![Flowchart](image)

**Figure 1: Work flow of the system in training phase**

5.1 Pre Processing
In Pre Processing phase the data can be cleaned and some useless data will be deleted.

5.2 Rough Set Theory
Using rough set theory [4], [5] the number of attributes can be reduced. It is one of the data mining application which can be used to reduce the features.

5.2.1 Basic concepts of Rough Set Theory

**Information System**
The information system IS can be shown as IS = (U, A)

Where U is the Universe (a dataset of process) U={x1, x2, x3, x4, x5, x6, ...xm} and A are the attributes of a process for instances. A is the set of attributes (features, attributes). Each attribute a∈A is a mapping from U into a value set V_a. Subsets of U are concepts.

The definition of an information function is fa: U→ Va, Va is the set of values of the attributes.

U={x1, x2, x3, x4, x5, x6, ...x10}

A= {a1, a2, a3}
The domains of the particular attributes are:
V 1={1,2,3}
V 2={1,2}
V 3={1,2,3,4}
i.e., the domain of each attribute is the set of values of this attribute.

![Diagram](image)

**Figure 2: The Incremental training approach**

**Indiscernibility relation**
For every set of attributes B B B ⊆ A A A, an indiscernibility relation Ind (B) defined in the following way: two objects, x_p and x_q are indiscernible by the set of attributes B in A, then if B(x_p)=B(x_q) for every system B⊆A. Information system A(B)=(U, B) and define the B-indiscernibility IND(B). For a set B of attributes, [x]_B the equivalence class of x∈U.

**Lower and upper approximations**
The rough set theory depends on two basic concepts, that are lower and the upper approximation of a set.

Let X denote the subset of elements of the universe U(x⊆U), the lower approximation of X in B(B⊆A), denote as BX.

BX = { x∈U | [x]_B ⊆ X }.

The lower approximation of the set X is a set of objects x_p. The upper approximation of the set X, denoted as BX, is the union of these elementary sets.

BX = { x∈U | [x]_B ⊆ X | X ≠ ∅ }.

For any object X_p of the lower approximation of X(x_p∈BX), for any object X_p of the lower approximation of X(x_p∈BX), x_p belongs to X

BNQ=BX-BX
If the lower and upper approximation are identical i.e., \( B_X = \overline{B_X} \), then set \( X \) is definable, otherwise, set \( X \) is undefinable in \( U \). There are four types of undefinable sets in \( U \):

1. If \( B_X \neq \emptyset \) and \( B_X \neq U \), \( X \) is called roughly definable in \( U \);  
2. If \( B_X \neq \emptyset \) and \( B_X = U \), \( X \) is called externally undefinable in \( U \).  
3. If \( B_X = \emptyset \) and \( B_X \neq U \), \( X \) is called internally undefinable in \( U \).  
4. If \( B_X = \emptyset \) and \( B_X \neq U \), \( X \) is called totally definable in \( U \).

RST can be used to combine the similar attributes and to reduce the number of attributes. So it can increase the processing speed and rise the detection rate.

**Functional dependence.** For given \( A = (U, A) \), \( C, D \subseteq A \), by \( C \rightarrow D \) is denoted the functional dependence of \( D \) on \( C \) in \( A \) that holds if \( \text{IND}(C) \subseteq \text{IND}(D) \).

### 5.3 Incremental Super Vector Machine

Super vector machine can be used to learn with large amount of data. SVM analyze the data and recognize patterns. The phases in svm are training, testing and incremental learning. Incremental learning is better than non incrementally learning. SVM uses mapping of the input data into a high-dimensional feature space to Support Vector Machines defined by a kernel function. The learning process will takes place in the feature space, and the data points only appear inside dot products with other points. SVM used in classification and regression techniques. There are several batches in each batch the data can be trained. Next svm is trained on new data in each new batch. After each training step the results will be produced. So time and memory consumption are reduced.

Incremental SVM learning is particularly interest in an online system, and for active learning. In an online system, the data is frequently collected continuously in time. Incremental SVM[10] provides the solution for extract online learning.

In Fig 2 the incremental process is show. The original data come from other network security organizations. There are two ways to update classification result, first one is the service provider will update the classification model next system can be updated.

### 5.4 Intrusion Detection

In SVM the data is separated by a hyper plane defined by a number of support vectors. The Support vectors are subset of training data used to define the boundary between two classes; it solves the problem by mapping input data into high-dimensional feature space using a kernel function.

In high dimension space the hyper planes can be created that allows linear inseparability.

\[
( p_1, q_1 ), \ldots, ( p_n, q_n ), p \in R^m, q \in \{+1,-1\}
\]

Where \( ( p_1, q_1 ), \ldots, ( p_n, q_n ) \) are a train data, \( n \) is the numbers of samples, \( m \) is the inputs vector,

\[
(w.p) + b = 0
\]

The category formula is:

- \((w.p) + b \) if \( q_i = +1 \)
- \((w.p) + b \) if \( q_i = -1 \)

SVM has kernel function that will be used to solve many problems. SVM can deal with linear inseparable problems.

This method will reduce the training and prediction time.

**Topological clustering**

The topological clustering can be used in the phase of training and testing phase of IDS in order to observe the results and to check whether the same patterns fall into the same cluster.

**Unsupervised Data**

SVM has the ability to handle linear data set. It is unsupervised method and it will detect intrusions without knowing it.

![Figure 3: The hyper plane of SVM](http://www.ijesat.org)
6. CONCLUSION

The proposed method uses the super vector machine and rough set theory to detect the intrusions. Rough Set Theory can be used to pre process the data and reduce the attributes. Next, the attributes were selected by RST will be sent to the Super Vector Machine. The incremental SVM and rough set theory can be combined to improve the performance and decrease the sample set.

REFERENCES


