Volume 2, Issue 5, May 2012 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Learning of Intrusion Detector in Conceptual Approach of Fuzzy Towards Intrusion Methodology

R.Shanmugavadivu
Assisatant Professor
PSG college of Arts&Science
Coimbatore.
Shan_vadivu@yahoo.com

Dr.N.Nagarajan
Principal,Coimbatore Institute of Engineering and Technology
Coimbatore.

Abstract-This paper proposes the Information system security is the Integrity and safety of its resources and activities, In the cyber world, it can be almost impossible to trace sophisticated attacks to their trace source. Almost all business organizations have is that use integrated technologies such as the networks of computer's company intranets. Under these circumstances threats from outside the organization must be addressed, because the damages from non-secured information systems can result in catastrophic consequences for the organization. To reduce this dependence, various preprocessing techniques such as data mining, neural networks, Petri nets, state transition diagram, genetic algorithms and fuzzy based logics are used. In this Proposed System, We have designed fuzzy logic based system for identifying the intrusion activities within a network. The proposed fuzzy logic-based system can be able to detect an intrusion behavior of the networks since the rule base contains a better set of rules. The Experiments and evaluations of the proposed intrusion detection system are performed with the KDD Cup 99 intrusion detection dataset. The Experimental results shows identifying whether records are normal or attack.

Keywords: Intrusion Detection System (IDS), Anomaly based intrusion detection, Fuzzy logic, Rule learning, KDD Cup 99 dataset.

I. Introduction

The network based threats we mean that in order to become effective potential attackers require network access to corporate computer systems or to networks used by corporate computer systems. Intruders attempt to attack networks to gain hold of information resources on the network. The attacks can be classified as, a) Interruption- Denying of service to authorized users. b) Interception- Unauthorized user obtaining access to a service. c) Modification-Unauthorized access and tampering of data. d) Fabrication-Counterfeit data. By Installing IDS within the corporate network, one can protect the information without the need of secure gateway. IDS have the ability to perform specific actions when an event takes place. The actions such as monitors and analysis the real time or the near real time warning of events that are identified by examining the network vulnerabilities scans. There are two popular approaches available currently to intrusion detection methodology .1.Knowledge Based IDS (referred as Signature based IDS) 2.Behaviour Based IDS(referred as Anomaly based IDS).Knowledge based IDS use a database of previous attacks and known system vulnerabilities to look for current attempts to exploit their vulnerabilities and trigger an alarm if a vulnerability is found. In Behavior based IDS it takes the dynamic approach in the sense that they detect deviations from the learned patterns of the user behavior. Generating rules is vital for IDS to differentiate standard behaviors from strange behavior by examining the dataset which is a list of tasks created by the operating system that are registered into a file in historical sorted order . In the proposed system, we have designed anomaly based intrusion detection using fuzzy logic. The input to the proposed system is KDD Cup 1999 dataset, which is divided into two subsets such as, training dataset and validating dataset. At first, the training dataset is classified into five subsets so that, four types of attacks (DoS (Denial of Service), R2L (Remote to Local), U2R (User to Root), Probe) and normal data are separated. we generate fuzzy rule in accordance with the definite rule by fuzzifying it in such a way, we obtain a set of fuzzy if-then rules with consequent parts that represent whether it is a normal data or an abnormal data. The Inference Engine collects and executes rules, Rules evaluate problem states as a group or one or more outcome variables, Variables contain the current set of problem-

specific data. These data are acquired by starting a separate backward chaining inference process to find the value of a variable that appears in the outcome of some other rules. When all the rules have been executed, the inference engine makes the values of the outcome variable available to the application that invoked the inference engine. These rules are given to the fuzzy rule base to effectively learn the fuzzy system. In the validating phase, the test data is matched with fuzzy rules to detect whether the test data is an abnormal data or a normal data.

II. LEARNING OF INTRUSION DETECTOR

The 1998 DARPA Intrusion Detection Evaluation Program was prepared and managed by MIT Lincoln Labs. The objective was to survey and evaluate research in intrusion detection. A standard set of data to be audited, which includes a wide variety of intrusions simulated in a military network environment, was provided. The 1999 KDD intrusion detection contest uses a version of this dataset. The raw training data was about four gigabytes of compressed binary TCP dump data from seven weeks of network traffic. This was processed into about five million connection records. Similarly, the two weeks of test data yielded around two million connection recordsA connection is a sequence of TCP packets starting and ending at some well defined times, between which data flows to and from a source IP address to a target IP address under some well defined protocol. Each connection is labeled as either normal, or as an attack, with exactly one specific attack type. Each connection record consists of about 100 bytes.

Attacks fall into four main categories:

- DOS: denial-of-service, e.g. syn flood;
- R2L: unauthorized access from a remote machine, e.g. guessing password;
- U2R: unauthorized access to local superuser (root) privileges, e.g., various ``buffer overflow'' attacks;
- Probing: surveillance and other probing, e.g., port scanning.

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. This makes 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. The datasets contain a total of 24 training attack types, with an additional 14 types in the test data only.

feature name	*description*	*type*
duration	length (number of seconds) of the connection	continuous
protocol_type	type of the protocol, e.g. tcp, udp, etc.	discrete
service	network service on the destination, e.g., http, telnet, etc.	discrete
src_bytes	number of data bytes from source to destination	continuous
dst_bytes	number of data bytes from destination to source	continuous
flag	normal or error status of the connection	discrete
land	1 if connection is from/to the same host/port; 0 otherwise	discrete
wrong_fragment	``wrong'' fragments	continuous
urgent	number of urgent packets	continuous

Table 1: Basic features of individual TCP connections.

feature name	*description*	*type*
hot	number of ``hot'' indicators	continuous
num_failed_logins	number of failed login attempts	continuous
logged_in	1 if successfully logged in; 0 otherwise	discrete
num_compromised	number of ``compromised'' conditions	continuous
root_shell	1 if root shell is obtained; 0 otherwise	discrete

su_attempted	1 if ``su root" command attempted; 0 otherwise	discrete
num_root	number of ``root" accesses	continuous
num_file_creations	number of file creation operations	continuous
num_shells	number of shell prompts	continuous

Table 2: Content features within a connection suggested by domain knowledge.

III.CONCEPTUAL APPROACH OF FUZZY TOWARDS INTRUSION METHODOLOGY

For intrusion detection, a wide variety of techniques have been applied specifically, data mining techniques, artificial intelligence technique and soft computing techniques. Most of the data mining techniques like association rule mining, clustering and classification have been applied on intrusion detection, where classification and pattern mining is an important technique. Similar way, AI techniques such as decision trees, neural networks and fuzzy logic are applied for detecting suspicious activities in a network, in which fuzzy based system provides significant advantages over other AI techniques.

Fuzzy logic is combined with some reasoning framework, such as the rule –based expert systems or decision trees, Fuzzy systems incorporate fuzzy logic in to their reasoning mechanism to gain a decided advantage over the crispness of Boolean logic for a wide range of problems. Thus the Rule based Expert systems cast their rules using fuzzy sets and fuzzy operators. The use of fuzzy logic provides a high degree of flexibility and robustness for a model. The different steps involved in the proposed system for the intrusion detection are described as follows:

1. Acquire and Organize the Data

2. Select Variables and Define properties

3. Discover Rules

4. Create the Fuzzy knowledgebase5. Validate the model

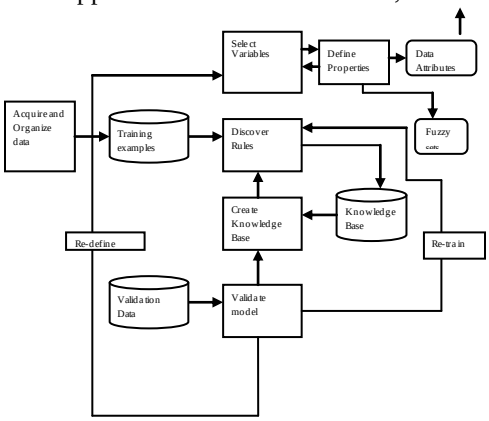

Figure.1- The Rule Induction and Model Development Methodology

IV FUZZY RULE INDUCTION ALGORITHM

It involves three processes 1. The description of the model variables. 2. The generation of candidate rules . 3. The Selection of the final Rule set.The set of input/output vectors in the data as follows

$(x^1_{v1}, x^1_{v2} \ldots .. x^1_{vn}, y^1_k)$ $(x^2_{v1}, x^2_{v2} \ldots .. x^2_{vn}, y^2_k)$

where x_{vi} are the input data values

And y is the desired outcome value. The purpose of the induction scheme is to discover the functional relationship.

$Y \leftarrow f(x_1, x_2, \ldots x_n)$

which in a fuzzy system is a form of close approximation rather than representation. so the model that obeys the relationship is in expression

$Y \approx f(x_1, x_2, \ldots x_n)$

The degree of granularity in the model as well as the clarity and depth of the patterns in the training file determines closely by the above expression.

V. CREATE THE FUZZY KNOWLEDGEBASE

A knowledge base represents the final results of the rule discovery process and consists of rules and variable definitions. A knowledge base becomes a self-contained

container of rules and variables that is opened and run by the fuzzy inference engine.

Executing a fuzzy model involves defining the methods of fuzzification, aggregation, and defuzzification.

Fuzzification and correlation involves finding the membership of x_{vi} in B01 and x_{v2} in S01 and using this to measure the degree of control exercised by the rule.This means that the control degree for a vector of input data points $x_{v1}, x_{v2} \ldots . x_{vi}$ is found by

$$d^{i}_{o} = \pi^{n}_{i=1} \mu A_i(x_i)$$

As an illustration, The controlled degree for the previous rule is computed as $d^{rl}_{O} = \mu BO1(x_{v1}) * \mu SO1(x_{v2})$

A Control degree for each rule forms the scaling factor that correlates the degree of truth in the antecedent with the degree of strength (control) in the consequent (outcome). Based on the Defuzzification strategy for the rule-induced model sums the scaled outcomes and divides by the sum of scaling.

$$X^{D}_{vo} = \sum^{n}_{i=1} d^{i}_{o} * \hat{x}_{vo}$$

$$\frac{}{\sum^{n}_{i=1} d^{i}_{o}}$$

where X^{D}_{vo} The defuzzified value of the outcome variable.

x_{vo} The center of the outcome fuzzy region.

VI. Validate the model

Once model's Knowledge base has been generated, the model prediction effectiveness must be tested. The predictive precision of a model depends on a variety of the factors. The amount and richness or the training data. The resolution granularity of the partitioning fuzzy sets. Validation measures how well the model performs by keeping track of the average standard error as it predicts an outcome value of the each incoming record. If the model does not perform well, we have two basic choices : retrain the model or redefine the model variables. The proposed system is implemented in MATLAB (7.8) with JAVA SDK 1.6 for the performance of the system is evaluated. For experimental evaluation, we have taken KDD cup 99 dataset , which is mostly used for evaluating the performance of the intrusion detection system. For evaluating the performance, it is very difficult to execute the proposed system on the KDD cup 99 dataset since it is a large scale. So we have used subset of 10% of KDD Cup 99 dataset for training and validating.

Training

Normal - 25000
DOS - 25000

VII. RESULTS & ANALYSIS

The Overall Accuracy is calculated using the F Test Calculator, F Test Calculator is an online statistics tool for data analysis programmed to determine whether two independent estimates of variance can be assumed to be estimates of the same variance. This calculator generates the F Test value according to the given inputs of standard deviation of first data set and standard deviation of second data sets. ***F Test*** is experiment based on the ratio of two variations. It is used to verify whether two independent estimates of variations can be assumed to be estimates of the same variations. If two means or treatments are considerably different, the variation in treatment will be greater than the variation due to random dissimilarity among individuals. The two variations which are estimates of σ^2 are calculated from sample means and by pooling the variations from the samples. No difference would be expected between the variations if estimates of F were calculated many times for a series of samples drawn from a population of normally distributed variations. Only the positive values are considered in F distribution therefore this is not a symmetrical distribution. The expected values of a quantitative variable within several pre-defined groups differ from each other.

F Test Formula

The one way ANOVA F Test statistic test in which the test statistic has an F-distribution under the null hypothesis can be calculated from the following formula

F = estimate of σ^2 from means

estimate of σ^2 from individuals

F = Variance between Treatments

Variance within Treatments

F = Variance of Treatments /Variance of Error

The collection of tools employs the study of methods and procedures used for gathering, organizing, and analyzing data to understand theory of Probability and Statistics. The analysis of the datasets can be done given in the following tables(4,5,6,7).

Probes - 4107
RLA - 77
URA - 42
Testing

Normal - 26000
DOS - 26000
Probes - 4107
RLA - 77
URA - 42

Table-4 (Training dataset)

Table-5(Testing dataset)

Training Data Testing

Testing Data Testing

Normal - 99.4160
DOS - 90.1440
Probes - 37.08
RLA - 15.58
URA - 19

Normal - 99.4385
DOS - 90.4154
Probes - 37.08
RLA - 15.02
URA - 18.98

Table-6 (Training data Testing)

Table-7(Testing data Testing)

As the Result, While running the Training Dataset and Testing Dataset, in MATLAB (7.8) with JAVA SDK 1.6 A message will Display and Shows that the data is normal or attack one. The evaluation metrics are computed for both training and validating dataset in the validating phase and the obtained result for all attacks and normal data .

VIII. CONCLUSION

We have developed an anomaly based intrusion detection system in detecting the intrusion behavior within a network. A fuzzy decision-making module was designed to build the system more accurate for attack detection, using the fuzzy inference approach. An effective set of fuzzy rules for inference approach were identified automatically by making use of the fuzzy rule learning strategy, which are more effective for detecting intrusion in a computer network. At first, the definite rules were for attack data as well as normal data. Then, fuzzy rules were identified by fuzzifying the definite rules and these rules were given to fuzzy system, which classify the test data. Executing a fuzzy model involves defining the methods of fuzzification, aggregation, and defuzzification. Here we have also used the F-test statistical methodology to find out the determine group of trial which differs significantly from an expected value. We have used KDD cup 99 dataset for evaluating the performance of the proposed system and experimentation results showed that the proposed method is effective in detecting various intrusions in computer networks.

IX REFERENCES

[1].Dubois, D,, Prade, H,, and Yager, R. R., (eds.), "Readings in Fuzzy Sets for Intelligent Systems, "Morgan Kaufmann Publishers Inc., 1993.
[2].Klir, George J. and Yuan, Bo, "Fuzzy Sets and Fuzzy Logic: Theory and Applicatons," Prentice Hall, May 1995.
[3] Kosko, Bart, "Fuzzy Engineering," Prentice Hall,
[4] James F. Kurose " BComputer Networking: A Top-Down Approach (5th Edition) " .
[5] Ankur Garg, "Information system Security" sixth edition
[6] Stephen Northcutt ,"Network Intrusion Detection (3rd Edition)" Sams;
[7] Michael Ryan and James Power.," Using Fuzzy Logic: Towards Intelligent Systems by Jun Yan, Prentice-Hall, 1995
[8] David F. Griffiths ,"Introduction to Matlab, Version 2.3, Department of Mathematics, University of Dundee, Copyright 1996.
[9] Klir & Yuan, "Fuzzy Sets and Fuzzy Logic: Theory and Application"
[10] Carter, Hogue ,"Intrusion Prevention Fundamentals" Cisco Press
[11] John Viega, Matt Messier, Pravir Chandra ,"Network Security with OpenSSL Cryptography for Secure Communications"Publisher: O'Reilly Media
edition (January 1, 2000)

