Volume 5, Issue 8, August 2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Fuzzy Logic and Genetic Based Intrusion Detection System

R. Ravinder Reddy
Asst. Professor
CSED-CBIT
Hyderabad, India

Dr. Y Ramadevi
Professor & Head
CSE-CBIT
Hyderabad, India

Dr. K. V. N Sunitha
Principal
BVRIT
Hyderabad, India

Abstract: Analyzing the intrusion behavior is very important in the e-world, every second huge amount of packets are traveling through the network among these packets which of them are normal or anomalous packets. To analyze such behavior so many techniques are used so far, among them many are detecting the packets properly but few of them are false positives or false negatives. To reduce these false alarms so much of research is going on in this area. Even though finding such behavior accurately is very complicated task. Here we are proposing a fuzzy logic with a data mining method which is a class-association rule mining method based on genetic algorithm. Due to the use of fuzzy logic ,this system can deal with mixed types of attributes and also avoid the sharp boundary problem. Genetic algorithm is used to extract many rules which are required for anomaly detection systems. An association rule mining method is used to extract sufficient number of important rules for the user's purpose rather than to extract a sufficient number of important rules for user's purpose rather than to extract all the rules meeting the criteria which are useful for misuse detection. this approach is finding the intrusive behavior and tried to reduce the false alarm rate.

Keyword: Intrusion detection, Fuzzy, classification techniques

I. INTRODUCTION

With the enlargement of technology in the computer networks, the necessitate of security is also grow. Computer security is Significant in almost any technology, Industry which operates on computer systems. Now a day's no any area is there where the security is not applied. Security is useful in the Banks, in Data mining, for cloud computing or other significant areas. So the One aspect is make use of data mining to get better security e.g. for intrusion detection. Data mining is widely used in business indemnity, banking, retail, science research astronomy, medication, and government security detection of criminals and terrorists which is the paramount equipment for finding the well-informed patterns.

The effectuality of an Intrusion detection system is measured using its probability of giving a signal upon an intrusion i.e. attack detection rate and the ratio of false alarms in them. The great concern in relevance of data mining techniques for attack detection and identification purposes has been evaluated. The dilemma of attack detection can be reduced to a data mining task of classifying data. Precisely, given a set of data points belonging to dissimilar classes normal activity and attacks of different types one has to separate them as accurately as possible by means of a model. There are two different approaches used to recognize attacks-

- Misuse detection
- Anomaly Detection

Misuse detection where attacks are detected by means of their known signatures but do not detect the unknown attacks. In Anomaly detection, firstly normal system behaviour is created and any variation from that as defined profile marked as anomaly.

misuse detection, the normal-pattern rules and intrusion-pattern rules are extracted from the training dataset. Classifiers are built up according to these extracted rules. While, for anomaly detection, we focus on extracting as many normal-pattern rules as possible. Extracted normal-pattern rules are used to detect novel or unknown intrusions by evaluating the deviation from the normal behavior.

II. PRELIMANARIES

DESIGN

Fuzzy Class-association-rule mining based on GA method for intrusion detection system overcomes many problems like sharp boundary problem, deals with a mixed database, and increases rule pool size. Therefore, extraction of many rules as compared to other method is possible. Support and fitness factors are calculated for each rule. Fitness function contributes to mining more rules with higher accuracy.

Proposed system objectives are as follows:

- Avoiding the sharp boundary problem by using fuzzy set theory.
- Use of mixed database, increases the detection rate and increases accuracy.
- Increases the size of rule pool by using the genetic operators.

- Flexibly applied to both misuse and anomaly intrusion detection.

The features of the proposed method are summarized as follows.

- Genetic-based fuzzy class-association-rule mining can deal with both discrete and continuous attributes in the database, which is practically useful for real network-related databases.
- Sub-attribute utilization considers all discrete and continuous attribute values as information, which contributes to avoid data loss and effective rule mining in genetic.
- The proposed fitness function contributes to mining more new rules with higher accuracy.
- The proposed framework for intrusion detection can be flexibly applied to both misuse and anomaly detection with specific designed classifiers.
- Experienced knowledge on intrusion patterns is not required before the training.
- High detection rates (DRs) are obtained in both misuse detection and anomaly detection.

III. METHODOLOGY

An IDS (Intrusion Detection System) is a system for detecting intrusions and reporting to the proper authority. Therefore, intrusion detection systems have attracted attention, as it has an ability to detect intrusion accesses effectively. These systems identify attacks and react by generating alerts or by blocking the unwanted data/traffic. The proposed system includes fuzzy logic with a data mining method which is a class-association rule mining method based on genetic algorithm. Due to the use of fuzzy logic, the proposed system can deal with mixed type of attributes and also avoid the sharp boundary problem. Genetic algorithm is used to extract many rules which are required for anomaly detection systems. An association-rule- mining method is used to extract a sufficient number of important rules for the user's purpose rather than to extract all the rules meeting the criteria which are useful for misuse detection.

Here, the concept of Genetic-based fuzzy class-association-rule mining is introduced in detail. The fuzzy membership values are used for fuzzy rule extraction, and sub-attribute-utilization mechanism is proposed to avoid the information loss. Meanwhile, a new genetic structure for association-rule mining is built up so as to conduct the rule extraction step. In addition, a new fitness function that provides the flexibility of mining more new rules and mining rules with higher accuracy is given in order to adapt to different kinds of detection. After the extraction of class-association rules, these rules are used for classification. In this paper, two kinds of classifiers are built up for misuse detection and anomaly detection, respectively, in order to classify new data correctly.

For extracting the rules with attributes of continuous value, fuzzy set theory is combined with association rule mining algorithm. Fuzzy Class-association-rule mining based on GA method for intrusion detection system overcomes many problems like sharp boundary problem, deals with a mixed database, and increases rule pool size. Therefore, extraction of many rules as compared to other method is possible. Support and fitness factors are calculated for each rule. Fitness function contributes to mining more rules with higher accuracy.

Proposed system objectives are as follows:

- Avoiding the sharp boundary problem by using fuzzy set theory.
- Use of mixed database, increases the detection rate and increases accuracy.
- Increases the size of rule pool by using the genetic operators.
- Flexibly applied to both misuse and anomaly intrusion detection.

The proposed GA-based intrusion detection using fuzzy data mining approach contains two modules where each works in a different stage. In the training stage, using the GA and fuzzy-association rule mining algorithm, a set of classification rules are generated from KDD dataset. In the intrusion detection stage, the generated rules are used to classify incoming data from a test file. Once the rules are generated, the intrusion detection is simple and efficient.

3.1 ALGORITHMS

3.1.1. Data Pre-processing

Intrusion detection techniques are misuse intrusion detection and anomaly intrusion detection. For detecting intrusion, rules are required (i.e. rules for normal and attack data). For these, sorting of normal records and attack records from KDD training dataset is required. Inputs of some important features of this sorted dataset are given to the pre-processor. The same pre-processing steps are required for both datasets (i.e. normal and attack dataset). Steps for pre-processing of attributes/features are shown in the following algorithm:

Algorithm: Classify KDD dataset, Pre-processing the data.

1. Select KDD dataset
2. Classify whole dataset into "normal" and "attack" class
3. Transform attributes to numeric value
4. Select important attribute/features

In above algorithm, classification method data mining is used for classifying the whole dataset into two classes i.e. "normal" and "attack". Feature selection is necessary because the use all available features are computationally infeasible.

3.1.2. Genetic Algorithm

Data pre-processing algorithm generates rules which are stored in the rule pool i.e. normal rule pool contains normal records and attack rule pool contains records for intrusion. The following algorithm is common for both i.e. normal and attack rule pool and explains about the genetic algorithm and its operators.

Algorithm: Generating chromosomes and calculating the fitness.

1. Initialize the population
2. MutationRate = 0.35
3. While number of generation is not reached
4. For each chromosome in the population
5. For each precalculated chromosome
6. Find fitness
7. End for
8. Assign optimal fitness as the fitness of that chromosome
9. End for
10. Remove some chromosomes with worse fitness
11. Apply crossover to the selected pair of chromosomes of the population
12. Apply mutation to each chromosome of the population
13. End while.

In the above algorithm, each rule is referred to as a chromosome or individual. In each generation, apply crossover and mutation to increase the number of rules. For Crossover a pair of individuals is determined by first selecting two individuals from the rule pool. A single point crossover is used to reproduce more individuals. In a single point crossover, exchange of genes (attributes value) between two individuals with respect to some point is carried out.

Range of fitness value is [-1, 1], so threshold fitness is 0 in this approach. Once the individuals are selected for making a pair, avoid repeated selection of individuals to make other pairs. The above procedure is then repeated until no individuals for making pairs are remaining. At the end of this algorithm, a large number of rules will be available for further processing. For Anomaly detection, the quantity of rules matters more than quality, whereas for misuse detection quality rules are required. So for both detection systems this algorithm is best suited[5].

3.1.3. Fuzzy Logic

After applying a genetic algorithm on normal and intrusion rule pool, all possible combinations of rules will be reproduced. On a large dataset, now apply fuzzy logic to avoid the sharp boundary problem. In this module, types of attributes i.e. discrete and continuous are used. For continuous attributes like duration, source bytes, destination bytes, find the maximum values for each attributes and then divide these values into LOW, MEDIUM and HIGH ranges, and find the fuzzy membership value for each attribute. For discrete attributes, numbers of columns are fixed on the basis of types of values for that attribute that is protocol attribute is divided into TCP, UDP and ICMP. The following algorithm shows fuzzy logic implementation for the rule pool[6].

Algorithm: Fuzzy Rule Extraction.

1. β = average value of attribute Ai; γ = the largest value of attribute Ai in the dataset;
2. Select record from the rule pool
3. Process all selected attribute
4. Set fuzzy membership value for each continuous attribute

$\alpha + \gamma = 2\beta$

5. Calculate fuzzy membership value for each continuous attribute
6. Store all fuzzy rules in fuzzy rule pool
7. Repeat step 5 until all selected columns are covered
8. Repeat step 2 until all records in the rule pool are considered.

A predefined membership function is assigned to each continuous attribute and the linguistic terms can be expressed by the membership function. The parameters α, β, and γ in a fuzzy membership function for attribute Ai is set as follows:

β is average value of attribute Ai in the database

γ is the largest value of attribute Ai in the database

3.1.4. Class-Association-Rule mining (CARM)

After fuzzy implementation, the fuzzy rule pool will be generated and this rule pool is given as an input to association rule mining. Association Rule Mining is a two-step process:

1. Find all frequent itemsets using Apriori algorithm.
2. Generate strong association rules from the frequent itemsets

For rule generation, antecedent part is generated by using apriori algorithm and for consequent; classification method is used in which the whole KDD dataset is distributed into two classes, that is normal and attack class on the basis of labels provided in the dataset. The following algorithm is used for finding the frequent itemsets from the dataset that is Apriori algorithm[7]:

Algorithm: Apriori algorithm for finding frequent itemsets.

Pass 1

1. Generate the candidate itemsets in C_1
2. Save the frequent itemsets in L_1

Pass k

1. Generate the candidate itemsets in C_k from the frequent itemsets in L_{k-1}
   1. Join L_{k-1} p with L_{k-1}q, as follows:
      insert into C_k
      select $p.\text{item}_1, p.\text{item}_2, \ldots, p.\text{item}_{k-1}, q.\text{item}_{k-1}$
      from L_{k-1} p, L_{k-1}q
      where $p.\text{item}_1 = q.\text{item}_1, \ldots p.\text{item}_{k-2} = q.\text{item}_{k-2}, p.\text{item}_{k-1} < q.\text{item}_{k-1}$
   2. Generate all (k-1)-subsets from the candidate itemsets in C_k
   3. Prune all candidate itemsets from C_k where some (k-1)-subset of the candidate itemset is not in the frequent itemset L_{k-1}
2. Scan the transaction database to determine the support for each candidate itemset in C_k
3. Save the frequent itemsets in L_k

At the end of above algorithm, the rule pool contains rules which are used for testing of the system. For misuse detection, train the system by giving attack data as an input and form the rules for attack data. For anomaly detection, train the system by giving normal data as an input and form the rules for normal data.

3.1.5. Performance evaluation

Detection of attack/intrusion can be measured by following metrics:

False positive (FP): Corresponds to the number of detected attacks but it is in fact normal.

False negative (FN): Corresponds to the number of detected normal instances but it is actually as attack, in other words these attacks are the target of intrusion detection systems.

True positive (TP): Corresponds to the number of detected attacks and it is in fact as attack.

True negative (TN): Corresponds to the number of detected normal instances and it is actually normal.

The accuracy of an intrusion detection system is measured with respect to detection rate and false alarm rate.

A. Detection rate (DR)

Detection rate refers to the percentage of detected attack among all input test data, and is defined as follows:

$$\text{Detection Rate} = \frac{TP}{TP + TN} * 100$$

B. False Positive Rate (FPR)

False positive rate refers to the percentage of normal data which is wrongly recognized as an attack, and is defined as follows:

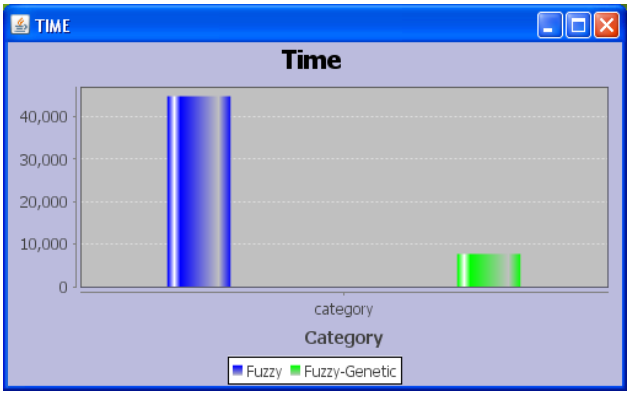
$$\text{False Positive Rate} = \frac{FP}{FP + TN} * 100$$

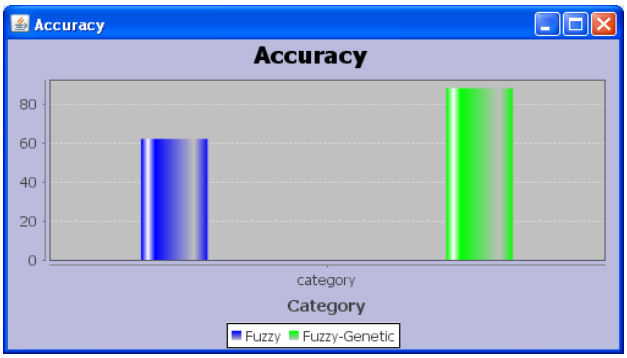
C. False Negative Rate (FNR)

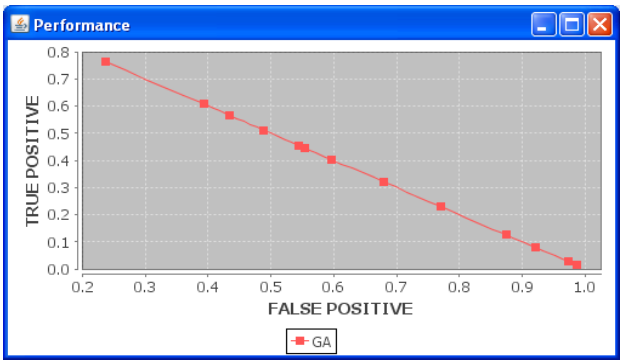
False negative rate refers to the percentage of attack data which is wrongly recognized as normal, and is defined as follows:

$$\text{False Negative Rate} = \frac{FN}{FN + TP} * 100$$

IV. EXPERIMENTS AND RESULTS







V. CONCLUSION AND FUTURE WORK

Conclusion and Future Enhancements

One preliminary IDS concept consisted of a set of tools intended to help administrators review audit trails. User access logs, file access logs, and system event logs are examples of audit trails. Fred Cohen noted in 1984 that it is impossible to detect an intrusion in every case, and that the resources needed to detect intrusions grow with the amount of usage.

Dorothy E. Denning, assisted by Peter G. Neumann, published a model of an IDS in 1986 that formed the basis for many systems today.Her model used statistics for anomaly detection, and resulted in an early IDS at SRI International named the Intrusion Detection Expert System (IDES), which ran on Sun workstations and could consider both user and network level data. IDES had a dual approach with a rule-based Expert System to detect known types of intrusions plus a statistical anomaly detection component based on profiles of users, host systems, and target systems. Lunt proposed adding an Artificial neural network as a third component. She said all three components could then report to a resolver. SRI followed IDES in 1993 with the Next-generation Intrusion Detection Expert System (NIDES)[10].

The Multics intrusion detection and alerting system (MIDAS), an expert system using P-BEST and Lisp, was developed in 1988 based on the work of Denning and Neumann. Haystack was also developed this year using statistics to reduce audit trails. Wisdom & Sense (W&S) was a statistics-based anomaly detector developed in 1989 at the Los Alamos National Laboratory. W&S created rules based on statistical analysis, and then used those rules for anomaly detection.

In 1990, the Time-based Inductive Machine (TIM) did anomaly detection using inductive learning of sequential user patterns in Common Lisp on a VAX 3500 computer. The Network Security Monitor (NSM) performed masking on access matrices for anomaly detection on a Sun-3/50 workstation. The Information Security Officer's Assistant (ISOA) was a 1990 prototype that considered a variety of strategies including statistics, a profile checker, and an expert system. ComputerWatch at AT&T Bell Labs used statistics and rules for audit data reduction and intrusion detection.

Then, in 1991, researchers at the University of California, Davis created a prototype Distributed Intrusion Detection System (DIDS), which was also an expert system. The Network Anomaly Detection and Intrusion Reporter (NADIR), also in 1991, was a prototype IDS developed at the Los Alamos National Laboratory's Integrated Computing Network (ICN), and was heavily influenced by the work of Denning and Lunt. NADIR used a statistics-based anomaly detector and an expert system[11].

The Lawrence Berkeley National Laboratory announced Bro in 1998, which used its own rule language for packet analysis from libpcapdata. Network Flight Recorder (NFR) in 1999 also used libpcap. APE was developed as a packet sniffer, also using libpcap, in November, 1998, and was renamed Snort one month later. APE has since become the world's largest used IDS/IPS system with over 300,000 active users.The Audit Data Analysis and Mining (ADAM) IDS in 2001 used tcpdump to build profiles of rules for classifications. In 2003, Dr. Yongguang Zhang and Dr. Wenke Lee argue for the importance of IDS in networks with mobile nodes.

The goal of intrusion detection is to monitor network assets to detect anomalous behavior and misuse. This concept has been around for nearly twenty years but only recently has it seen a dramatic rise in popularity and incorporation into the overall information security infrastructure. Beginning in 1980, with James Anderson's paper, Computer Security Threat Monitoring and Surveillance, the notion of intrusion detection was born. Since then, several pivotal events in IDS technology have advanced intrusion detection to its current state.

Currently, market statistics show that IDS is amidst the top selling security vendor technologies and should continue to rise. Furthermore, government initiatives, such as the Federal Intrusion Detection Network, (FIDNet) are also adding impetus to the evolution of IDS. Advancements in IDS will ultimately push security technology into a whole new arena of automated security intelligence[12].

Finally, it is expected that the future IDS product will be a grocery shelf of choices that will fit a broad range of needs. It won't be a single product, but an integrated system of products from multiple vendors. It will include network-based sensors, host-based sensors, and a centralized anomaly detection system that analyzes logs sent to it by the sensors. The anomaly detection system will take predetermined actions depending on the nature and severity of the detected threat.

REFERENCES

[1] Mabu S., Chen C., Shimada K., "An Intrusion-Detection Model Based on Fuzzy Class-Association-Rule Mining Using Genetic Network Programming," IEEE Transactions Systems, Man, Cybernetics C, Application and Reviews, volume 41, number 1, pp. 130–139, January 2011.

[2] http://en.wikipedia.org/wiki/Intrusion_detection_system#Development

[3] http://www.symantec.com/connect/articles/evolution-intrusion-detection-systems

[4] http://en.wikipedia.org/wiki/Intrusion_detection_system#Development

[5] http://en.wikipedia.org/wiki/Genetic_algorithm

[6] http://en.wikipedia.org/wiki/Fuzzy_logic

[7] Agrawal R. and Srikant R., "Fast algorithms for mining association rules," in Proceeding 20th VLDB Conference, Santiago, Chile, pp. 487–499, 2008.

[8] http://whatis.techtarget.com/definition/fuzzy-logic

[9] Ektefa M., Memar S., "Intrusion Detection Using Data Mining Techniques," IEEE Trans., 2010.

[10] Helm B., "Fuzzy Association Rules: An Implementation in R," Master's Thesis, Vienna University of Economics and Business Administration Vienna, 2007.

[11] Gong R., Zulkernine M., Abolmaesumi P., "A Software Implementation of a Genetic Algorithm Based Approach to Network Intrusion Detection," Proceedings of the Sixth International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Dis tributed Computing and First ACIS International Workshop on Self-Assembling Wireless Networks, IEEE, 2005.

[12] Naidu N. and Dharaskar R., "An Effective Approach to Network Intrusion Detection System using Genetic Algorithm", International Journal of Computer Applications (0975 - 8887) volume 1 No.2, 2010.

