Intrusion Detection with Genetic Algorithms and Fuzzy Logic

Emma Ireland
Division of Science and Mathematics
University of Minnesota, Morris
Morris, Minnesota, USA 56267
irela065@morris.umn.edu

ABSTRACT

Keywords

Intrusion detection, genetic algorithm, fuzzy logic

1. INTRODUCTION

2. BACKGROUND

2.1 Types of Intrusion Detection Systems

2.2 Types of Networking Attacks

2.3 Detection Methodologies

2.4 Genetic Algorithm

2.5 Fuzzy Logic

2.6 KDD99 Data Set

3. FUZZY GENETIC ALGORITHM IMPLEMENTATION

The focus of the research is on detecting new or unknown types of attacks in a network. The intrusion detection system is able to identify normal network activity as well as classify different attack types using a fuzzy genetic algorithm. This kind of algorithm is able to learn new attacks, and has a high detection rate. The system used is evaluated in terms of detection speed, detection rate, and false alarm rate.

3.1 RLD09 Data Set

The KDD99 dataset is 14 years old, and newer attack types are not included in it because of its age. Because of this, the authors of [1, 2] created their own data set, RLD09. To create the data set, the authors captured network data from the Computer Engineering Department at King Mongkut's University of Technology Thonburi, in Bangkok, Thailand. The data has around ten million preprocessed data packets. It has 17 different types of attacks, as well as normal network activity. The attacks can be divided into denial of service attacks and probe attacks. A packet sniffer was used to get information about TCP, UDP, and ICMP headers from protocol packets. Then this information is processed into 12 features by counting connections between a source IP and destination IP. Some example features include the number of: packets, source ports, and destination ports. The following are examples of data records, where each one has 12 feature values.

- 21,21,15,0,21,0,0,0,0,0,0,0, attack
- 102,2,2,0,0,1,102,0,0,0,0,0, normal

3.2 Algorithm Overview

The algorithm that is used in [1, 2] first randomly finds a rule. Then the rule is improved in the training phase. After that, the rules are used to classify the data into classes in the testing phase. The pseudo code below describes the fuzzy genetic algorithm that is used.

Ask Nic how to put in algorithm.

The fitness function to be maximized is:

$$fitnessfunction = \frac{\alpha}{A} - \frac{\beta}{B}$$

In the implementation, a population size of 10 was used for each generation. An individual in the population represents a possible detection rule. The two best individuals from a present generation are preserved for the next generation. The other individuals in the new generation come from single-point crossover.

3.3 Fuzzy Algorithm

In order to measure the probability of an attack, a trapezoidal shape was used in the algorithm. The trapezoidal shape has four parameters: a, b, c, and d. The following algorithm calculates the probability of being attacked:

Ask Nic how to put in algorithm.

Put in trapezoidal shape figure.

3.4 Features and Rules

The KDD99 dataset is composed of 41 features. The authors of [1, 2] used the following eight features in their system: duration, src_bytes, num_failed_logins, root_shell, num_access_files, srv_count, serror_rate, same_srv_rate.

UMM CSci Senior Seminar Conference, December 2013 Morris, MN.

Figure 1: Fuzzy encoding for a feature

010	011	100	101
a	b	c	d

The value of each feature is normalized to be a number between 0.0 and 7.0, and then is encoded into blocks of binary strings. See Figure 1 for an example of a block. A rule has 12 blocks of features, and at the end of the string is the type of attack.

One record is passed into a rule. Each feature in a record is matched to one block of the rule. The parameters of each block measure the probability of an attack using the trapezoidal fuzzy rule shape. The probabilities of each block are then combined, and the average of the probabilities is compared with a threshold to determine if the record is an attack class or normal class. Then the predicted result is compared with the actual result.

3.5 Experimental Design and Results

4. GENETIC ALGORITHM IMPLEMENTATION

4.1 Algorithm Overview

4.2 Experimental Design and Results

5. CONCLUSIONS

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] P. Jongsuebsuk, N. Wattanapongsakorn, and C. Charnsripinyo. Network intrusion detection with fuzzy genetic algorithm for unknown attacks. In *Information Networking (ICOIN), 2013 International Conference on*, pages 1–5, 2013. *This is one of my main sources, and talks about using IDS with fuzzy logic and genetic algorithms. It uses a newer data set (RLD09).*

[2] P. Jongsuebsuk, N. Wattanapongsakorn, and C. Charnsripinyo. Real-time intrusion detection with fuzzy genetic algorithm. In *Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 2013 10th International Conference on*, pages 1–6, 2013.