Intrusion Detection with Genetic Algorithms and Fuzzy Logic

Emma Ireland

Division of Science and Mathematics
University of Minnesota, Morris
Morris, Minnesota, USA

December 2013
UMM CSci Senior Seminar Conference

The Big Picture

●

●

●

●

●

Outline

1. Background
2. Genetic Algorithm Implementation
3. Fuzzy Genetic Algorithm Implementation
4. Conclusions

Outline

1. Background
   - Types of Networking Attacks
   - Detection Methodologies
   - Data Sets - KDD99 and RLD09
   - Rules
   - Genetic Algorithms
   - Determining the Accuracy of an Algorithm
2. Genetic Algorithm Implementation
3. Fuzzy Genetic Algorithm Implementation
4. Conclusions

Types of Networking Attacks

Explain DoS, remote to user, user to root, probe

Detection Methodologies

Explain signature-based and anomaly-based detection

KDD99

- Generated by simulating a military network environment in 1999.
- Has long been a standard data set for intrusion detection.
- Data in the set is classified as normal or attack activity.
- KDD99 uses 41 features.
    - *Features* are properties of a *record*, (either an attack or normal activity), that are used to describe the activity.

Some Features of KDD99

- duration: length of the normal or attack activity in seconds.
- num_failed_logins: number of failed login attempts.
- root_shell: returns 1 if root shell is obtained, else returns 0.
- serror_rate: percentage of connections that have “SYN" errors.

RLD09

- RLD09 was created because KDD99 is 14 years old.
- Data was captured from a university in Bangkok, Thailand.
- The data has 10 million data packets, 17 different types of attacks (divided into denial of service and probe attacks), and 12 features.

Rules

- Elements of one set are separated into different sets in order to differentiate between normal connections and attacks.
- If-Then format
  - If the length of the activity is 4 seconds, then the probability of it being an attack is 100%.

Genetic Algorithms

Determining the Accuracy of an Algorithm

- False positive (FP): intrusion detection system incorrectly identifies normal activity as being an attack.
- False negative (FN): intrusion detection system fails to identify harmful activity.
- True positive (TP): intrusion detection system correctly identifies activities to be attacks.
- True negative (TN): intrusion detection system correctly identifies activities to be normal.
- Detection rate (DR): the number of true positives divided by the total number of intrusions that happen.

Outline

1 Background

2 Genetic Algorithm Implementation
- Algorithm Overview
- Experimental Design and Results

3 Fuzzy Genetic Algorithm Implementation

4 Conclusions

Algorithm Overview

Experimental Design

Results

Outline

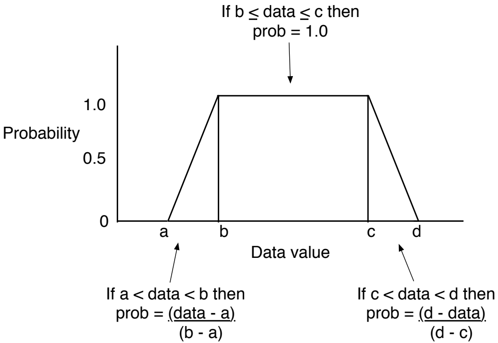
1. Background

2. Genetic Algorithm Implementation

3. Fuzzy Genetic Algorithm Implementation
   - Fuzzy Algorithm
   - Algorithm Overview
   - Experimental Design and Results

4. Conclusions

Measuring the Probability of a Record Being an Attack



Example:

- Feature: duration (length of the activity in seconds).
- a=1, b=3, c=5, d=7
- The length of the activity is 6 seconds (between c and d).
- prob = $\frac{d - \mathrm{data}}{d - c} = \frac{7-6}{7-5} =$ 0.5

Encoding of Features and Rules

- The four parameters are encoded into blocks.
- Each block is a feature with values between 0.0 and 7.0.

010	011	100	101
a=2	b=3	c=4	d=5

- A rule has 12 blocks of features, at the end is the type of attack.

010	011	100	101	......	010	011	101	111	DoS
a=2	b=3	c=4	d=5	......	a=2	b=3	c=5	d=7	
Block 1					Block 12				Type

Algorithm Overview

```
for each record do
    for each rule do
        for each feature do
            prob = fuzzy(); // Trapezoidal
            fuzzy rule shape
            totalprob = totalprob + prob;
        end for
        if totalprob > threshold then
            class is attack;
        end if
    end for
    find $A$, $B$, $\alpha$, and $\beta$
end for
calculate fitness
crossover(), mutation()
```

Fitness function:

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

A: # of attack records.
B: # of normal records.
α: # of attack records correctly identified as attack.
β: # of normal records incorrectly classified as attack.

Experiments

- A variety of experiments were run. Two experiments used just RLD09, and three experiments used KDD99 and RLD09 together.
- The experiments used a total of 16,000 records of normal activity and 10,500 records of attack activity. Of the attack records, 4,000 were denial of service attacks and 6,500 were probe attacks.

Experiments Using Only RLD09

Experiment 1

Experiments Using Only RLD09

Experiment 2

Experiments Using Both RLD09 and KDD99

Experiment 1

Experiments Using Both RLD09 and KDD99

Experiment 2

Experiments Using Both RLD09 and KDD99

Experiment 3

Outline

1 Background

2 Genetic Algorithm Implementation

3 Fuzzy Genetic Algorithm Implementation

4 Conclusions

Conclusions

Thanks!

Thank you for your time and attention!

Questions?

References