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
   - Fuzzy Logic
   - 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

1. duration: length of the normal or attack activity in seconds.
2. src_bytes: number of bytes sent from source to destination.
3. num_failed_logins: number of failed login attempts.
4. root_shell: returns 1 if root shell is obtained, else returns 0.
5. num_access_files: number of operations on access control files.
6. srv_count: number of connections to the same service as the current connection in the past two seconds.
7. serror_rate: percentage of connections that have “SYN" errors.
8. same_srv_rate: percentage of connections to the same service.

RLD09

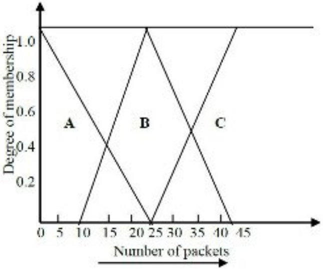
- RLD09 was created because KDD99 is 14 years old.
- Data was captured from a university in Bangkok, Thailand.
- The data has 10 million data packets.
- 17 different types of attacks - divided into denial of service attacks, probe attacks. It also has normal activity.
- 12 features, which include the number of packets, source ports, and destination ports.

Rules

- Elements of one set are separated into different sets in order to differentiate between normal connections and attacks.
- If <*condition*> then <*action*>.
- Specify the details of a packet such as the IP address or port number.
- If a packet matches any of the rules in the intrusion detection system, the system will take appropriate action, which may include stopping the connection or logging off the system.

Fuzzy Logic

- Used to detect patterns that have a behavior that is between normal and unusual.
- If <*condition*> then <*consequence*>.
  - *condition* is a fuzzy variable and *consequence* is a fuzzy set
- If the number of packets with the same destination address is 20, and *a*=10, *b*=25, *c*=45, then the degree=.6 and the region=B so the number of packets=medium.



```
if x is between $a$ and $b$ then
    degree = $(x - a)/(b - a)$
else if x is between $b$ and $c$ then
    degree = $(c - x)/(c - b)$
else
    degree = 0.0
end if
```

Genetic Algorithms

Determining the Accuracy of an Algorithm

Explain training and test set, false positive, false negative, true positive, true negative, detection rate.

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
- Main Points of Research
- Fuzzy Algorithm
- Algorithm Overview
- Experimental Design and Results

4 Conclusions

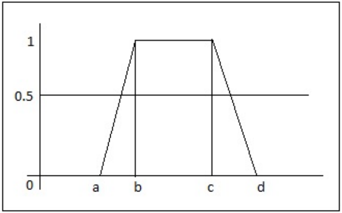
Main Points of Research

- Detecting new or unknown types of attacks in a network.
- The intrusion detection system used is able to identify normal network activity as well as attacks using a fuzzy genetic algorithm.

- Ran experiments using only RLD09, and experiments using KDD99 and RLD09 together.

Measuring the Probability of a Record Being an Attack

- Trapezoidal shape



- The parameters are the values of a feature.

```
if data value is between $b$ and $c$ then
    prob = 1.0
else if data value is between $a$ and $b$ then
    prob = $(\mathrm{data} - a)/(b - a)$
else if data value is between $c$ and $d$ then
    prob = $(d - \mathrm{data})/(d - c)$
else
    prob = 0.0
end if
```

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

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