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

- Computer lab gets large numbers of login attempts that are attempts at intrusion.
  - Trying to gain root access to the system.
  - Delete files of other users and change user passwords.
- With an intrusion detection system (IDS) -> classify those attempts.
- Intrusion detection systems provide one way of detecting attacks by monitoring network activities for malicious or abnormal behaviors and then producing reports, alerts, and actions.
- Training an IDS: use a fuzzy genetic algorithm.

Outline

Outline

Types of Networking Attacks

- Denial of Service (DoS):
  makes machine inaccessible to user by making it too busy to serve legitimate requests.
  - Systems lock out user from account after certain number of failed login attempts. Attacker would be able to use this to prevent users from logging in, by failing to log in enough times to lock the account.

- Probe:
  examines machine in order to collect information about weaknesses that in the future could be used to compromise the system.
  - Attacker could be trying to determine what version of a software is being run on that machine, and if that version has a known issue then that allows them to attempt to attack that.

Detection Methodologies

- Signature-based detection: compares well-known patterns of attacks that are already known to the IDS against captured events in order to identify a possible attack.
  - Simple and effective way to detect known attacks, ineffective against new kinds of unknown attacks.
- Anomaly-based detection: looks for patterns of activity that are rare and uncommon.
  - Harder to do than signature-based detection, can be an effective way to detect new, unknown attacks.

KDD99

- Generated by simulating a military network environment in 1999.
- Has long been a standard data set for intrusion detection.
- Data was processed into 5 million *records*.
  - A record is a sequence of TCP packets, between which data flows to and from a source IP address to a target IP address.
- Each record is classified as either normal or attack activity.

Features of KDD99

- KDD99 uses 41 *features*, which are properties of a record that are used to describe the activity and help to distinguish normal connections from attacks.

- duration: length of the record in seconds.
- num_failed_logins: number of failed login attempts.
- root_shell: returns 1 if root shell is obtained, else returns 0.

RLD09

- RLD09 was created because KDD99 is 14 years old, newer attack types are not in KDD99 because of its age.
- Data was captured from a university in Bangkok, Thailand.
- As well as normal network activity, has 17 different types of attacks (divided into denial of service and probe attacks).

Rules

- A commonly used approach for detecting intrusions is to use rules.
- If-Then format: If (*condition*) then (*consequence*).
  - The condition is composed of one or more features, and the consequence says if it is an intrusion or not.
  - If *duration = 4* then *intrusion*.

Training and Testing Sets

- The authors divided the data set into two subsets, a *training set* and a *test set*.
- The given algorithm is then trained on the training set to look for patterns.
- These patterns are then verified using the test set.

Determining the Accuracy of an Algorithm

	Predicted	
Actual	Not Attack	Attack
Not Attack	True Negative (TN)	False Positive (FP)
Attack	False Negative (FN)	True Positive (TP)

Detection rate (DR): percentage of normal and attack activity correctly classified from the total number of data records.

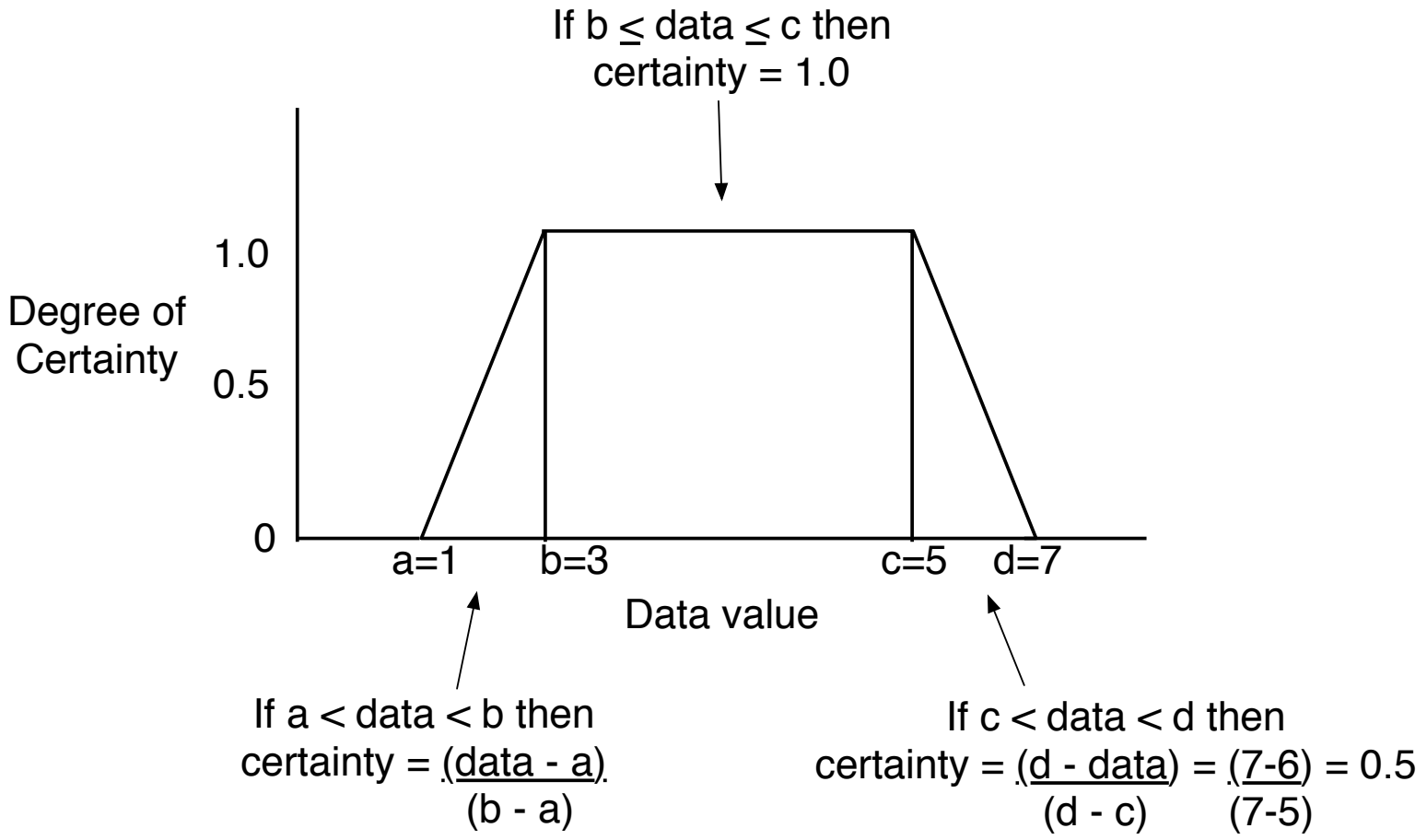
Outline

Fuzzy Logic

- Fuzzy logic is used in intrusion detection systems to find the degree of certainty of a record being an attack.
- Fuzzy logic rules are similar to the rules described before, except that consequence is a certainty factor.
  - If (*duration* = 6) then (*the degree of certainty of the record being an attack is 0.5*).

Finding the Degree of Certainty of a Record Being an Attack

Suppose that the feature is duration, and it is 6 seconds. Then data=6.



Encoding of Features and Rules

- The four parameters are encoded into blocks. Each block is a feature with values between 0 and 7.
- A rule has one block for each of 12 features followed at the end by a marker indicating the type of attack.

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

Figure : Based on [Jongsuebsuk *et al.*, 2013]

- The degree of certainty is computed for each of the 12 blocks, and if the sum of those is greater than a threshold, then it will be declared as an attack.

Outline

Genetic Algorithms

- GAs: search technique used to find solutions to problems.
- Possible solutions to problems can be represented in a variety of problem dependent ways, such as bit strings.
  - IDS rules can be represented as bit strings.
- First, randomly generated population of potential solutions is created. Mutation, crossover, selection are applied to each generation until acceptable solution is found or some time limit is exceeded.

Mutation and Crossover

- Mutation: random bits in an individual, or possible solution, are randomly changed.
  - Takes bits of rule and changes them to form slightly different rule.

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

$\rightarrow$

010	011	100	**111**
a=2	b=3	c=4	**d=7**

- Crossover: two individuals swap sequences of bits to form two new individuals.
  - Take 2 rules and creates new rules by swapping bits of old rules.

001	011	101	111
a=1	b=3	c=5	d=7

010	100	110	111
a=2	b=4	c=6	d=7

Selection and Fitness

- Selection: individuals that have better fitness are chosen to be parents.
- The fitness of an individual is specified by the fitness function, which determines the quality of a particular individual.

- In an IDS: fitness measures how well a rule classifies records as either attacks or normal activity. Selection combined with fitness function directs search towards effective solution.

Fitness function

The fitness function in the algorithm is:

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

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

Outline

Experiments Using Only RLD09

Experiment 1

- Fuzzy GA was used to create DoS and probe detection rules.
  Both rules were then used together in testing process to identify attacks from testing data set.

- Training set: 10,000 records.
  Test set: 26,500 records.

- If the record is either a DoS rule or a Probe rule, it is classified as an attack; else it is normal.

Experiments Using Only RLD09

Experiment 1 Results

	Attack	Normal	Total	FP(%)	FN(%)	DR(%)
DoS Training	1499	8501	10000	1.46	47.50	91.64
Probe Training	2496	7504	10000	1.83	15.38	94.79
Testing	10500	16000	26500	1.13	4.10	97.92

Experiments Using Only RLD09

Experiment 2

- Attacks pulled out of training set and kept for unknown data testing, to test that fuzzy GA could detect unknown attacks.
- Used fuzzy GA and a decision tree algorithm, which is another common algorithm for classification problems.
- 7 tests were run.
  For each test case there were 13 attack types plus normal activity that were in the training data set.
  3 attack types were used for the unknown testing data set.

Experiments Using Only RLD09

Experiment 2 Results (7 tests were run in total, 5 are shown here.)

Test Case	Unknown Attacks	Decision Tree DR (%)	Fuzzy Genetic DR (%)
1	Adv Port Scan (Probe) Ack Scan (Probe) Xmas Tree (Probe)	Avg = 98.33	Avg = 100
2	UDP Flood (DoS) Host Scan (Probe) UDP Scan (Probe)	Avg = 46.65	Avg = 99.80
3	Jping (DoS) Syn Scan (Probe) Fin Scan (Probe)	Avg = 99.70	Avg = 98.75
4	UDP Flood (DoS) RCP Scan (Probe) Fin Scan (Probe)	Avg = 70.35	Avg = 98.15
5	Http Flood (DoS) RCP Scan (Probe) Fin Scan (Probe)	Avg = 99.94	Avg = 97.50

Experiments Using Both RLD09 and KDD99

Three experiments used both RLD09 and KDD99.

Experiment 1 - Used fuzzy GA to classify normal activity and attacks from KDD99 and RLD09.

Data set	Attack	Normal	FP (%)	FN (%)	DR (%)
KDD99	160,117	39,337	0.13	1.55	98.72
RLD09	10,500	16,000	1.14	3.39	97.97

Experiments Using Both RLD09 and KDD99

Experiment 2

- Used the fuzzy GA to classify types of attacks in KDD99.
- 10 tests were run in total, 5 are shown here.

Test	Attack	Type	FP (%)	FN (%)	DR (%)
1	Back	DoS	85.33	0.00	16.56
2	PoD	DoS	84.66	0.00	15.58
3	Smurf	DoS	0.76	0.10	99.73
4	Portsweep	Probe	6.40	0.00	93.66
5	Satan	Probe	0.74	3.75	99.22

- 8 test cases had DR greater than 93%. Only 2 cases had low DR, (cases 1 and 2).

Experiments Using Both RLD09 and KDD99

Experiment 3

- Used the fuzzy GA to classify types of attacks in RLD09.
- 17 tests were run in total, 6 are shown here.

Test	Attack	Type	FP (%)	FN (%)	DR (%)
1	HTTP Flood	DoS	0.36	3.5	99.46
2	Smurf	DoS	0.02	0	99.98
3	UDP Flood	DoS	11.06	0	89.59
4	Fin Scan	Probe	2.58	0	97.50
5	IP Scan	Probe	13.01	16.4	86.89
6	Syn Scan	Probe	0.65	4.2	99.24

- 15 cases had DR greater than 97%. 2 cases had low DR, (cases 3 and 5).

Outline

Conclusions

- The fuzzy genetic algorithm had a higher detection rate than a decision tree algorithm in most cases.
- Fuzzy genetic algorithms are good at detecting unknown attacks.
- The use of fuzzy genetic algorithms in intrusion detection is an effective way of detecting attacks.

Thanks!

Thank you for your time and attention!

Questions?

Probe Attacks

- Advance Port Scan
- Ack Scan
- Xmas Tree
- Host Scan
- UDP Scan
- Syn Scan
- Fin Scan
- RCP Scan
- Portsweep
- Satan
- IP Scan

DoS Attacks

- UDP Flood
- Jping
- HTTP Flood
- Back
- PoD
- Smurf

References

Jongsuebsuk, P. and Wattanapongsakorn, N. and Charnsripinyo, C.
Network intrusion detection with Fuzzy Genetic Algorithm for unknown attacks.
In *2013 International Conference on Information Networking (ICOIN)*, pages 1-5, 2013.

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

See my Senior Seminar paper for additional references.