Distributive Network Anomaly Detection
A Simulation Method

Swarna Reddy

Problem Statement and Definitions

- How to identify malicious traffic in the network capture flow data in a distributive network
- Definition of potential malicious traffic patterns entering or leaving on a certain organizations network

Potential Attacks and Definitions

Attack Types:

DoS, Port Scans, Spamming, Worms, Beaconing and Command and Control

Distributive Denial of Service (DDoS):

Normally the goal of DoS attack is exhaust the resources of single host (computer or server) but in Distributive attack spread this attack to the whole network or part of the network at a period of time.

Types od DDoS attacks:

SYN flood,
PING of death,
SMURF

Port Scans:

Scanning for open ports for the purpose of listening

- General Scan
  - Scan a few destination IP addresses or ports may have time signature in sent SYN packets
- Vertical Scan
  - Scan a single destination IP address with a multiple destination ports
- Horizontal Scan
  - Scan for multiple destination IP addresses with a single or few destination ports

.

Spamming:

Sending high volumes of e-mail in a short period of time.

Worms:

copy themselves to other hosts and network shares without user interaction.

Beaconing:

is signaling method a host indicates its presence. The infected hosts signal their presence to the master.

Command and Control:

is command script from worms author to the worm to perform certain tasks.

Detection Approaches:

SYN flood: Monitoring number of SYN flags set to 1.

Port Scan:

Similar to SYN flood but watch for relatively with less frequency from one are very few IP addresses are scanning for number of ports.

Spamming:

Narrow down the IP addresses that connects to SMTP multiple servers

Worms Spreading and Beaconing:

Worms are spread through different approaches, one of them is through back doors on high ports.

Categories of NID Systems

Pattern-based detection

Anomaly-based detection

Network vs. Host based Intrusions

Protection Systems:
Fire wall vs. Intrusion Detection System(IDS)

known signature based attacks vs. unknown attacks.

Filtering techniques can be applied in the host computers with virus protection

At network level Supervised learning algorithms can be applied to classify the known attack signatures, algorithms that are based on anomaly detection techniques known to be better to detect unknown signatures or zero-day attacks.

Significance of DDoS Attacks

DDoS attacks exhibit many important features:

(i) DDoS attacks can involve thousands of computers. An attacker can compromise a set of intermediate hosts (zombies) that will launch the attack process. Obviously, increasing the number of zombies multiply the efficiency of the DDoS attack and early detection is harder. In most cases.

(ii) Unlike other attacks, the network flow used to carry out a DDoS attack slightly differs from a normal traffic. When analyzing packet headers and payloads, the Intrusion Detection System (IDS) can hardly realize that an intrusion is in progress. Therefore anomaly based detection strategies are more suitable to detect DDoS attacks.

(iii) The detection system can be bypassed by the attacker when the attack flow can not be processed by the IDS. This means that the DDoS detection approach should be possible but comes with a numerical cost.

Network Anomaly Detection A simulation Method

Network Simulation

Attack simulator: The simulation module that generates the test data for the purpose of Evaluating the anomaly detection method for distribution networks.

A user specified number of attacks can be randomly inserted in to the stream

Data simulator: Data assumed to be preprocessed for the purpose of specified network intrusion detection.

n1			**n2**		
Pkt.no	Pkt.size	Src.IP	Pkt.no	Pkt.size	Src.IP

Data generated for user specified number of nodes with specified features.

Attack Simulator: Generates the user specified number of attacks.

Severity of attack: Several tuning constants are used to control the number of attacked nodes and the number of times the attack can take place at given starting and ending points.

These attacks are injected in to normal traffic.
Matrix of attck:

an2

an1					
Pkt.no	Pkt.size	Src.IP	Pkt.no	Pkt.size	Src.IP

Method of simulation:

Normal packet size is generated from a binomial random variable with a mean following a uniform distribution.

Inter-arrival times are generated according to a Poisson distribution with a mean that is generated from a chi-square distribution.

Source IP addresses are not generated; instead, occurrences of new IP addresses are generated according to a Bernoulli distribution with a uniformly distributed mean.

Anomaly detection algorithm (Exponential Smoothing.):

The algorithm processes simulated network traffic.

Each packet variable is predicted from past observations.

The prediction of observation at time t is a weighted average of the past observations.

The largest weights are assigned to observations that are nearest in time.

The weights diminish as the distance in time between present observation and the past observation increase.

This prediction method sometimes referred to as exponential smoothing.

Attack Scenario:

The absolute deviations determine whether an attack has occurred.

A system attack at a given time will induce either unusually large, or small values in some or all of the responses observed at that time.

Node level:

The node level behavior is monitored by computing the Mahalanobis distance between the observed and predicted packet variables

System level:

System behavior is monitored for anomalous behavior by summing the Mahalanobis norms across nodes.

Analytics

The performance of the anomaly detection system is evaluated at the node and system level.

The single most important summary statistic is the sensitivity.

Sensitivity is the proportion of attacks that were detected as anomalous.

At the node level, sensitivity is computed by noting whether a packet was generated as an attack packet and noting whether the packet was identified as anomalous.

At the system level, data from all nodes is combined and the system is said to be under attack is any packet (across all nodes) was generated as an attack.

summary statistics:

True positive:

The fraction of detected attack packets that are truly attack packets

False positive:

The fraction of detected attack packets that are truly not attack packets

True negative:

The fraction of packets identified as normal that are truly normal (not anomalous)

False negative:

The fraction of packets identified as normal but in fact were generated as attack packets

Node attack detection

Method	True Pos	True Neg	Sensitivity
Z-score	0.889	0.793	0.000
Mahalanobis	0.970	0.795	0.013

System attack detection:

Method	True Pos	True Neg	Sensitivity
Z-score	0.000	0.793	0.000
Mahalanobis	0.890	0.809	0.099

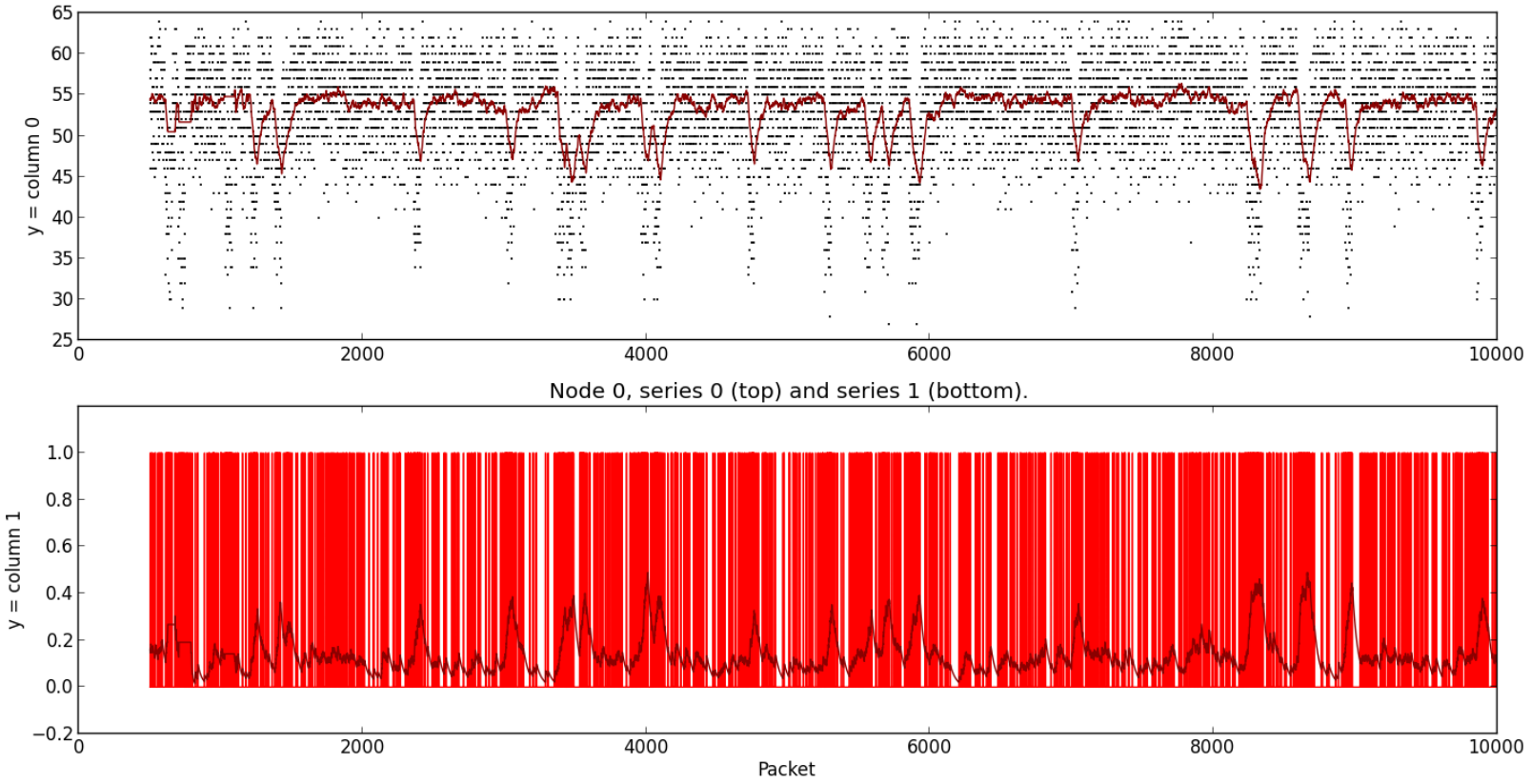
One Particular case

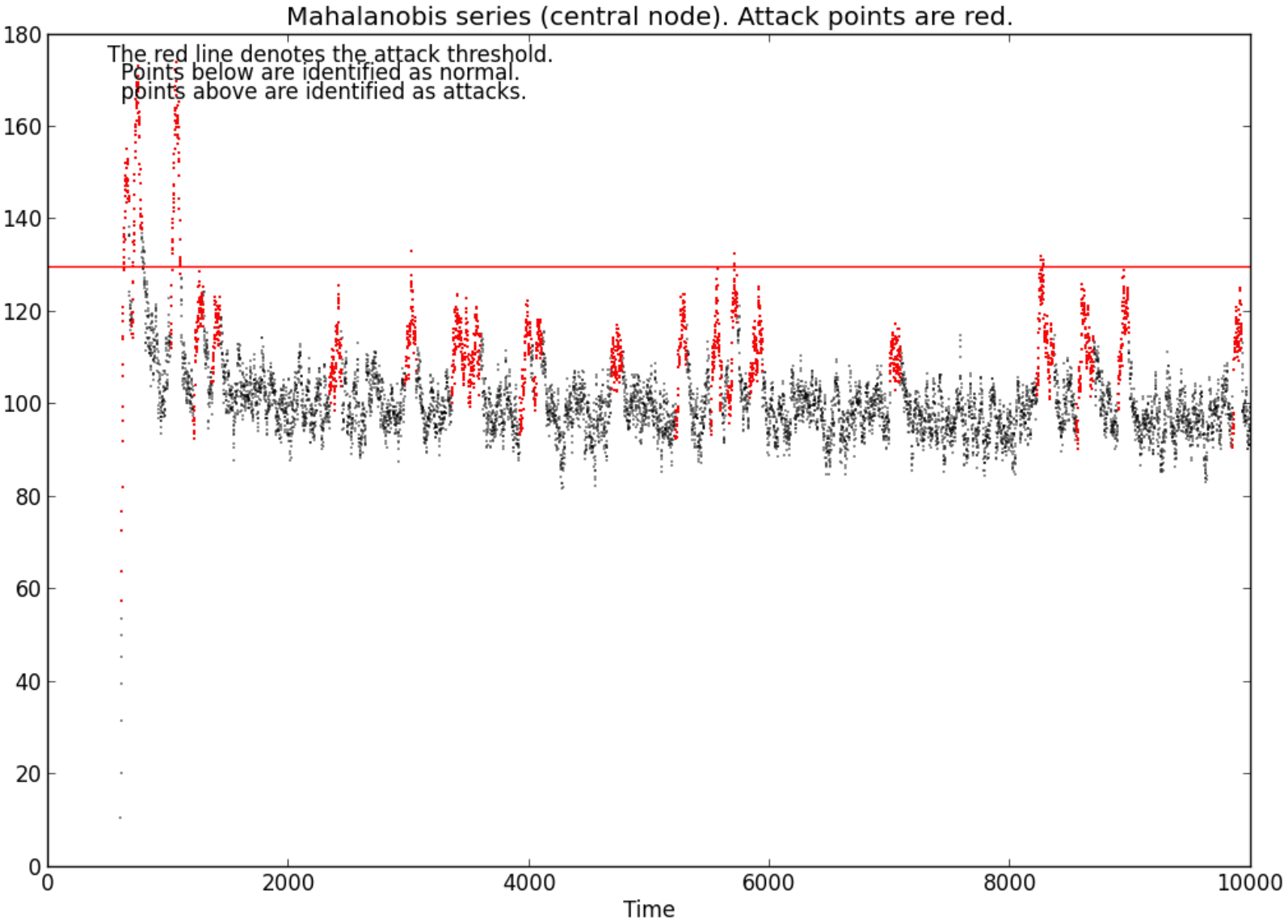
Number of correctly identified system attacks = 194
True positive rate (system attacks) = 0.89

Number of incorrectly identified system attacks = 24
False positive rate = 0.11

Number of correctly identified normal activity time steps = 7510
True negative rate = 0.809

Number of incorrectly identified normal activity time steps = 1771
False negative rate = 0.191


y = column 0
65
60
55
50
45
40
35
30
25
0
2000
4000
6000
8000
10000
Node 0, series 0 (top) and series 1 (bottom).
y = column 1
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0
2000
4000
6000
8000
10000
Packet


Mahalanobis series (central node). Attack points are red.
The red line denotes the attack threshold.
Points below are identified as normal.
points above are identified as attacks.
180
160
140
120
100
80
60
40
20
0
0
2000
4000
6000
8000
10000
Time

References:

Guillaume Delawaele et all, Extracting Hidden Anomalies using Sketch and Non Gaussian Multiresolution Statistical Detection Precedures
ACM 978-159593-785-8/07/2008

Romain Fontugne etall MAWILIB: Combinig Diverse Anomaly Detectors for Automated Anomaly Labeling and Performance Bechmarking ACM 1-4503 -4503 -0448 1/10/11