Analysis of the US-CERT DAC

Josh McNutt <jmcnutt@cert.org>

FloCon: Netflow Analysis Workshop
July 21, 2004

CERT® Network Situational Awareness Group
Software Engineering Institute
Carnegie Mellon University
Pittsburgh, PA 15213-3890

The CERT Network Situational Awareness Group is part of the Software Engineering Institute. The Software Engineering Institute is sponsored by the U.S. Department of Defense.

© 2004 by Carnegie Mellon University
Outline

• Data
• Graphical Displays
• Detecting Trends
• Anomaly Detection
• Roadmap
Data

• **Snort**
  - Signature-based alerts
  - Pre-processor alerts

• **Origin**
  - Multiple networks of varying size

• **Volume**
  - ~30-50 million alerts per month

• **Ancillary Information**
  - Country code
  - Netblock
IDS Data: challenges

• No new attacks
  – Only matches known signatures

• Lack of context
  – Don’t know what we are not seeing

• Non-standardized signature rule sets
  – No administrative control

• Missing Data
  – Uncertainty: Sensor failure vs. no intrusion attempts
TCP Destination Port Changes

Comparison of port activity across organizations shows monthly trends.
Share of New Source IP Addresses

Share of new daily source IP addresses stays fairly consistent.
Signature Class Transition

Transition probabilities highlight sequential patterns in data.

- **Current State**
  - Source IP records alert on Destination IP
- **Transition probability**
  - Percent chance for next class of alert recorded
- **Most source/dest combos** involve only one signature class
- **Small transition probability** for
  - Privilege Escalation

[Graph showing transition probabilities for different states]

© 2004 by Carnegie Mellon University
Daily Transition Probabilities

Transition probabilities can be monitored over time to identify consistent sequences.
Trend Detection

• Current month vs. previous month
  – Across organizations
  – % changes

• Time Series
  – Fit trend line
    – Arbitrary time period
  – Seasonal Components
  – Regression with ARMA errors
Anomaly Detection

- Goal: Identify data points which deviate from overall pattern of data
- Our current implementation (Figure of Merit)
  - Evaluate hours
  - Record # alerts, # source IP addresses, # destination IP addresses, # signatures
- For each hour, we want measure of how deviant it was.
Mahalanobis distance: 2D case

- Compute distance metric between each hour and the **average** hour
- When measuring **Euclidean** (Mahalanobis) Distance, all points along **circle** (ellipse) are same distance from the center
  - Points on larger circle/ellipse are greater distance from center
- Shape of the ellipse
  - Function of correlation between variables
- Generalizes to n dimensions (Ellipsoid)
Analysis Roadmap

- Incorporate flow data
- Automating trend detection
  - Time series analysis
- Clustering
  - Group sources by similar activity patterns
    - Temporal correlation
    - Targeting similarities
    - Signature usage
  - Look for evidence of possible coordination