Time Series Data Mining for Context-Aware Event Analysis

Mona Lange

Characterization of the field of research

- IT security difficult to maintain / plethora of IDS/IPS/FW events  
  What is the problem?
- Event fusion, filtering, prioritization / detecting important activities  
  How do I address it.
- Mission-criticality tradeoff handled appropriately
- No human in the loop  
  USP
Context: Critical Infrastructures – ACEA
Automatically Acquired: Vulnerabilities

PANOPTESEC Network Topology overview

The Network Topology overview offers to the operator a view of the reachability of exploitable devices.
Automatically Acquired: Network Topology

PANOPTESEC Network Topology overview

The Network Topology overview offers to the operator a view of the reachability of exploitable devices. The operator can open the network view and observe vulnerabilities...
Attacks: Reactive and Proactive View
Objective of this Research

• **Online: Context-Aware Event Analysis**
  • Normalize heterogeneous events from multiple sources
  • Filter and fuse events
  • Prioritization by operational impact assessment based on important activities ("workflows")

• **Offline: Time Series Data Mining**
  • Learn to identify workflows based on mining network traffic
  • Formally represent workflows as stochastic processes
  • Mission Oriented Network Analysis (MONA)
Context-Aware Event Correlation

Security Sensor Events → Syslog (Normalization) → Correlation (multi-threaded, window-based) → Correlated prioritized event

Monitored System

Network Traffic → MONA → Formally modeled workflows → Workflows involving mission-critical systems → Network and Vulnerability Inventory

Analyzer ID | Time
---|---
CEDET01IDS | 2016-01-24 1:02:31.20

Source IP Address | Destination IP Address
20 85.1.1.8 | 132.8.1.5

CVE ID | Tag
CVE-2016-00034 | VULNVERIFIED

IP | Vulnerability identifier
---|---
132.8.1.5 | CVE-2016-0034
Support for Other Modules

• Enables other modules to work at all (normalization)
• Reduces load due to fusion and filtering
• Prioritization allows subsequent modules to focus on mission-critical events such that...
  • ... attacks can be matched and ...
  • ... relevant response plans can be generated ...
  • ... in realtime
**Network Service Dependency**

**Direct Dependency:** A -> B, if A requires B to satisfy certain requests from its clients [Chen, Xu, al.]

**Indirect Dependency:** A -> B; A -> C, if request A -> B and A -> C are caused by the same activity

Detecting Dependencies

Normalized Cross-Correlation

\[
q_{r,s}(\tau) = \frac{1}{\text{bins}} \sum_{t=0}^{\text{bins}} (r_t - \mu_r)(s_{t+\tau} - \mu_s) \frac{\rho_r \rho_s}{},
\]

\[
t_{\text{delay}} = \arg\max_{\tau=\{0, \ldots, (t_{\text{max}}-t_{\text{min}})\} \subseteq \mathbb{N}} q_{r,s}(\tau).
\]

\[
q_{r,s}(t_{\text{delay}}) \geq \theta
\]
HMM for Workflow Modeling

(a) \text{Client} \rightarrow \text{DNS Server}

(b) \text{Client} \rightarrow \text{Load balancing server}
Context-Aware Event Correlation

Security Sensor Events

Syslog (Normalization)

Correlation (multi-threaded, window-based)

Correlated prioritized event

Analyzer ID | Time
---|---
CEDET01IDS | 2016-01-24 1:02:31.20

Source IP Address | Destination IP Address
---|---
20 85.1.1.8 | 132.8.1.5

CVE ID | Tag
---|---
CVE-2016-00034 | VULNVERIFIED

Network and Vulnerability Inventory

IP | Vulnerability identifier
---|---
132.8.1.5 | CVE-2016-0034
Workflows for Event Prioritization

Using Workflows for Event Prioritization

Using a list of **mission-critical network devices**, workflows can be used to identify whether mission-critical network devices are affected.

---

**Event Prioritization**
LLC – Scalability Tests

- Production environment 19.7.16 for about 7 hours
- LLC successfully deployed
- Overall >6M Syslog messages were received
- Due to the criticality of the production environment, IPS sensors and FWs block unexpected attempts of communication (white listing).
  - Therefore, as was expected, no LLC alerts were produced
  - Only events were processed
- LLC is able to perform within an operational environment
- Reduce the overall number of reported events by at least a factor of 2
LLC – Scalability Tests

Number of incoming Syslog messages

Consecutive time windows with a 1 second tumble
LLC – Functionality and Performance Tests

• Emulation environment
• Functionality
  – Provides input for both HOC implementations
  – Used in operational workshop w/o any problems
• Performance
  – 10,000 events/sec 2CPUs
  – 100,000 events/sec 4CPUs
  – 1,000,000 events/10sec 4CPUs

MONA: Performance Analysis

ACEA-Network + Synthetic networks

Precision = \( \frac{TP}{TP + FP} \)

Recall = \( \frac{FP}{TP + FN} \)

\( F - measure = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \)

True Positives (TP),
False Positives (FP),
False Negatives (FN)

<table>
<thead>
<tr>
<th>Flows per communication between indirectly dependent network services:</th>
</tr>
</thead>
<tbody>
<tr>
<td>MONA</td>
</tr>
<tr>
<td>Sherlock</td>
</tr>
<tr>
<td>Orion</td>
</tr>
</tbody>
</table>
Summary

• **Online: Context-Aware Event Analysis**
  ✓ Normalize heterogeneous events from multiple sources
  ✓ Filter and fuse events
  ✓ Prioritization by operational impact assessment based on important activities ("workflows")

• **Offline: Time Series Data Mining**
  ✓ Learn to identify workflows based on mining network traffic
  ✓ Formally represent workflows as stochastic processes
  ✓ Mission Oriented Network Analysis (MONA)
Bibliography


