Combating Dependence Explosion in Forensic Analysis Using Alternative Tag Propagation Semantics

Md Nahid Hossain, Sanaz Sheikhi and R. Sekar
Secure Systems Lab

**Drawbacks of Existing Approaches**

- “Needle in a haystack”: hard to distinguish real attacks within a flood of false alarms
- “Connecting the dots”: no help in understanding the overall campaign

**Solution:** Use provenance information

**Issue:** Dependency Explosion

**Subject & Data Tags**

**Subject Tag**
- Suspicious: Process may have been compromised.
- Benign: Believed to be benign; may contain vulnerabilities.

**Default Tag Propagation**
- Normally, tags propagate in the direction of information flow
- Default propagation causes **dependence explosion**, which leads to massive (unreadable) scenario graphs

**Our Approach**

- Policy-based attack detection and root cause identification
- Modulate dependency flow using **subject tags**
- Conservative dependence propagation for **suspicious** processes
- Selective dependence propagation for **benign** processes

**Scenario Graph Generation**

- Provenance-based alarm clustering:
  - Attribute an alarm to an ancestor that also triggered alarms.
- Entry point identification:
  - Trace back from largest clusters to a source node (e.g., network connection).
- Re-propagating tags:
  - Assign suspicious Subject tag to entry point, repropagate tags as needed.
- Forward search:
  - Run depth-first search and prune away nodes with high data integrity

**Graph Size Reduction**

- Data set: Linux audit data consumer
- FreeBSD DTrace data consumer

**Tag Attenuation**

- Key intuition: Objects are noisy intermediaries for propagating attacks
- Key idea: Attenuate tag propagation from **benign** subjects to objects
- Implemented by adding a small constant \( \alpha \) to data tag value:
  \[
  object_{tag} = subject_{tag} + \alpha
  \]

**Performance**

<table>
<thead>
<tr>
<th>Data set</th>
<th>Total events</th>
<th>Memory Usage (GB)</th>
<th>Graph generation time/attack (sec.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>L3</td>
<td>74 M</td>
<td>0.49</td>
<td>0.043</td>
</tr>
<tr>
<td>L4</td>
<td>36.5 M</td>
<td>0.11</td>
<td>0.053</td>
</tr>
<tr>
<td>F3</td>
<td>21 M</td>
<td>0.29</td>
<td>0.030</td>
</tr>
<tr>
<td>F4</td>
<td>37.2 M</td>
<td>0.11</td>
<td>0.220</td>
</tr>
</tbody>
</table>