Appeared in IEEE S&P (Oakland) '20

Combating Dependence Explosion in Forensic Analysis Using Alternative Tag Propagation Semantics

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

Stony Brook University
Computer Science

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*

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

Local simplifications and visualization

Subject & Data Tags

Subject Tag

Suspicious: Process may have been compromised.

Benign: Believed to be benign; may contain vulnerabilities.

(Data) Integrity Tag

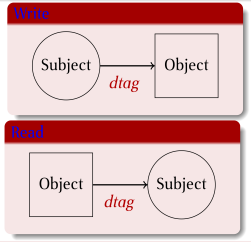
- Low: $[0.0, 0.5)$
- High: $[0.5, 1.0]$

(Data) Confidentiality Tag

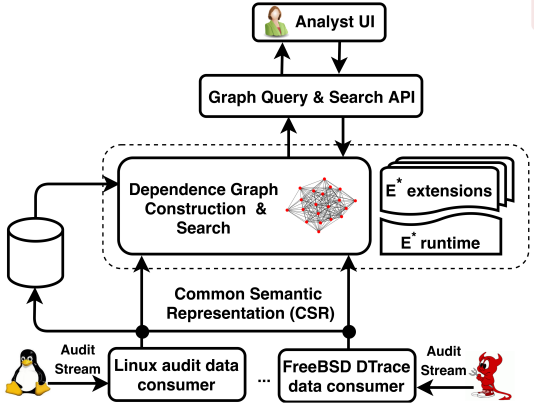
- High: $[0.0, 0.5)$
- Low: $[0.5, 1.0]$

Default Tag Propagation

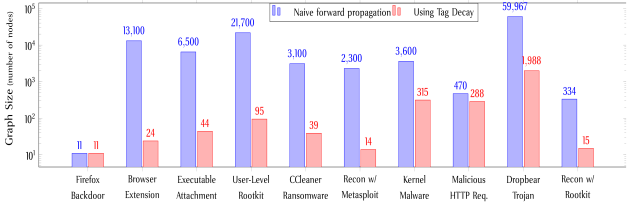
Normally, tags propagate in the direction of *information flow*



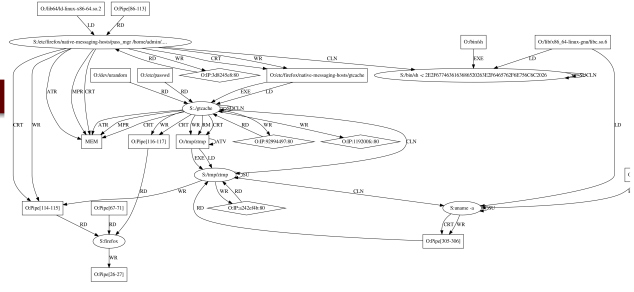
Default propagation causes *dependence explosion*, which leads to *massive (unreadable) scenario graphs*



Graph Size Reduction



Scenario Graph for Vulnerable Browser Extension



Tag Attenuation

- Key intuition: *Objects* are lousy intermediaries for propagating attacks
- Key Idea: *Attenuate* tag propagation from *benign* subjects to objects
- Implemented by adding a small constant a to data tag value:

$$object.dtag = subject.dtag + a$$

Tag Decay

- Key intuition: If a benign process is compromised by suspicious input, this will happen soon after consuming it
- Key Idea: *Gradually lift* tags of *benign* processes to *quiescent* value T_q
- By decaying data tag d exponentially at rate r

$$d = max(d_0,\ d_0 * r^t + (1 - r^t) * T_q)$$

Performance

Data set	Total events	Memory Usage (GB)	Graph generation time/attack (sec.)
L3	714 M	0.49	0.043
L4	36.5 M	0.11	0.053
F3	21 M	0.19	0.030
F4	37.2 M	0.11	0.220