Provenance Graphs

Thank you so much for participating in our study! This survey is about interpreting provenance graphs. Provenance Graphs are directed graphs where:

- Nodes represent system objects (e.g., files, processes, sockets).
- Edges denote causal relationships between these objects.

We use data provenance graphs in auditing and intrusion detection for cyber security. These graphs describe a system's processes and execution. They help gather information regarding a data's origin, its present state, and who acted upon it.

The survey consists of three sections, each containing 3 questions. Below are the instructions for each section:

Section 1:
You will be shown an image of a provenance graph that was detected as a potential attack. Please analyze the graph and decide whether it represents a false positive (Benign) or an Attack.
Tip: You may use any external resources, such as websites or references, to help interpret the graph.

Section 2:
You will be shown an image of a provenance graph accompanied by a text description below it. Please consider both the image and the text information to decide whether the graph represents a false positive (Benign) or an Attack.

Section 3:
You will be provided only a text description of a provenance graph. Based on the description, please decide whether it represents a false positive (Benign) or an Attack.

Don’t stress about your performance—just do the best you can! We are **NOT** collecting your data so it cannot be traced back to you.
Your participation is incredibly valuable, and we sincerely thank you for contributing to our project.
We’re grateful for your time and effort in helping us better understand these systems!

Benign or Attack?

Nodes represent system objects (e.g., files, processes, sockets). Edges denote causal relationships between these objects.

1. Is this benign or an attack? *

- ◯ Benign
- ◯ Attack

1b. Can you explain briefly (1-2 sentences) why you think it is benign/an attack?

Benign or Attack?

Nodes represent system objects (e.g., files, processes, sockets). Edges denote causal relationships between these objects.

2. Is this benign or an attack? *

- ○ Benign
- ○ Attack

2b. Can you explain briefly (1-2 sentences) why you think it is benign/an attack?

Benign or Attack?

Nodes represent system objects (e.g., files, processes, sockets). Edges denote causal relationships between these objects.

3. Is this benign or an attack? *

- ○ Benign
- ○ Attack

3b. Can you explain briefly (1-2 sentences) why you think it is benign/an attack?

Benign or Attack?

Nodes represent system objects (e.g., files, processes, sockets). Edges denote causal relationships between these objects.

The graph shows typical behavior of syslogd writing to system log files such as /var/log/maillog and /var/log/auth.log. However, the write operation to /home/frank/Archives.lock, a file in a user-specific directory, is anomalous and potentially indicates misuse of the syslogd process to create or manipulate unauthorized files.

4. Is this benign or an attack? *

- ◯ Benign
- ◯ Attack

4b. Can you explain briefly (1-2 sentences) why you think it is benign/an attack?

Benign or Attack?

Nodes represent system objects (e.g., files, processes, sockets). Edges denote causal relationships between these objects.

5. Is this benign or an attack? *

- ◯ Benign
- ◯ Attack

5b. Can you explain briefly (1-2 sentences) why you think it is benign/an attack?

Benign or Attack?

Nodes represent system objects (e.g., files, processes, sockets). Edges denote causal relationships between these objects.

The graph reveals typical email service activities by imapd, such as opening user mail files. However, it also shows highly suspicious actions, including accessing sensitive password files (/etc/pwd.db and /etc/spwd.db), writing to temporary files (/tmp/.31.29ada), and communicating with external IPs (128.55.12.118:46790 and 193.40.5.73:80). The presence of wget indicates potential exfiltration or unauthorized file downloading.

6. Is this benign or an attack? *

- ◯ Benign
- ◯ Attack

6b. Can you explain briefly (1-2 sentences) why you think it is benign/an attack?

Benign or Attack?
The graph shows newsyslog performing a standard operation on its configuration file /etc/newsyslog.conf.d/amd.conf. There are no anomalies or suspicious activities.

7. Is this benign or an attack? *

- ◯ Benign
- ◯ Attack

7b. Can you explain briefly (1-2 sentences) why you think it is benign/an attack?

Benign or Attack?
The graph illustrates normal SaltStack operations, such as logging to /var/log/salt/minion and reading module files from the SaltStack directory. However, the write operation to /proc/* is highly suspicious, as this directory is typically read-only and used for system metadata, with writes often associated with potentially malicious activities.

8. Is this benign or an attack? *

- ◯ Benign
- ◯ Attack

8b. Can you explain briefly (1-2 sentences) why you think it is benign/an attack?

Benign or Attack?
The graph mostly depicts benign system operations such as inspecting or interacting with system files (/var/run/ld-elf.so.hints) and terminal devices (/dev/tty). However, the process vUgefal accesses /var/run/ld-elf.so.hints, which is suspicious due to its unusual and non-standard naming. This anomaly raises concerns about potentially unauthorized access.

9. Is this benign or an attack? *

- ◯ Benign
- ◯ Attack

9b. Can you explain briefly (1-2 sentences) why you think it is benign/an attack?

General Questions

Which presentation did you prefer?

- ◯ Graph Only
- ◯ Text Only
- ◯ Graph and Text

Did you prefer Text Only or Graph Only?

- ◯ Graph Only
- ◯ Text Only

Did you find that the text descriptions were pretty accurate and helpful?

- ◯ Accurate and helpful
- ◯ Not accurate, but helpful for context

Please give us any feedback on the graph descriptions or provenance graphs that you might think will be helpful for future development: