NavPro: Network Analysis and Visualization using Provenance Data

Chris Botaish and Mike Calder

Advised by:
Jeff Diewald (58), Tom Moyer (58), Nabil Schear (58)
George Heineman (WPI)

15 October 2014

This work is sponsored by the Assistant Secretary of Defense for Research and Engineering under Air Force Contract FA8721-05-C-0002. Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the United States Government. Approved for public release; distribution is unlimited.
Provenance

- History of ownership to guide authenticity
Data Provenance

• Detailed record of the origin and evolution of data in a system

• Series of events that together can provide full history of activity

• Linux Provenance Modules (LPM)
  – Developed by researchers at Lincoln Laboratory, University of Florida, and University of Oregon
  – Places hooks after security checks in SELinux
  – Captures history as low-level actions

Why Is It Useful?

• Fills knowledge gaps that other types of forensic data have
• Can be used to verify the integrity and confidentiality of data
• Allows a cyber analyst to re-create a crime scene
• Can give insight into the mission impact of a cyber attack
Why Is It Not Useful?

- Each user-level event translates into many low-level provenance messages, which are useless alone.
- The data is extremely dense and cannot be manually parsed by a human.
- Storing this data can be difficult because thousands of messages produced by one system every second.
Raw Provenance Data

843382 664a5d06-6b30-4da2-b29b-6cd00cd4a913 3851 3 {udevadm,control,--
property=UDEV_BIOSDEVNAME=,HOME=/,KEYBOARDTYPE=pc,rd_LVM_LV=vg_lion226/lv_root,TERM=linux,PATH=/sbin:/bin:/usr/sbin:/usr/bin,SYSFONT=latarcyrheb-sun16,LANG=en_US.UTF-8,PWD=/,KEYTABLE=us}

843491 664a5d06-6b30-4da2-b29b-6cd00cd4a913 4152 3 {rm,-f,/etc/udev/rules.d/71-
biosdevname.rules,HOME=/,KEYBOARDTYPE=pc,rd_LVM_LV=vg_lion226/
lv_root,TERM=linux,PATH=/sbin:/bin:/usr/sbin:/usr/bin,SYSFONT=latarcyrheb-
sun16,LANG=en_US.UTF-8,PWD=/,KEYTABLE=us}

843659 664a5d06-6b30-4da2-b29b-6cd00cd4a913 3851 3 {udevadm,trigger,--
action=add,HOME=/,KEYBOARDTYPE=pc,rd_LVM_LV=vg_lion226/
lv_root,TERM=linux,PATH=/sbin:/bin:/usr/sbin:/usr/bin,SYSFONT=latarcyrheb-
sun16,LANG=en_US.UTF-8,PWD=/,KEYTABLE=us}

848772 664a5d06-6b30-4da2-b29b-6cd00cd4a913 3853 3 {/sbin/modprobe,-
b,acpi:LNXSYSYST:,UDEV_LOG=3,ACTION=add,DEVPATH=/devices/LNXSYSTM:00,SUBSYSTEM=acpi,MODALIAS=acpi:LNXSYSYST:,SEQNUM=748}

849237 664a5d06-6b30-4da2-b29b-6cd00cd4a913 3853 3 {/sbin/modprobe,-
b,acpi:LNXSYBUS:,UDEV_LOG=3,ACTION=add,DEVPATH=/devices/LNXSYSTM:00/
LNXSYBUS:00,SUBSYSTEM=acpi,MODALIAS=acpi:LNXSYBUS:,SEQNUM=752}
NavPro

• Forensic tool to analyze and visualize provenance data in a network
  – Present data in an interactive view
  – Complement data with context and interpretation
  – Allow for investigation on specific entities
  – Leverage cloud to solve computational/storage challenges

• Extensible framework to interpret provenance data
  – Support extensions for different data sources
  – Allow customization of interpretation logic
Architecture

- Data Source
- Classifier
- Database
- Query API
- Web Client

Parser Plugin Pool
Normalizer Plugin Pool
Classification

Server
Client
Interpretation
User-Level Events

• Create File
• Delete File
• Access File
  – Read
  – Write
  – Execute
• Change File Permissions/Ownership

Every event has a timestamp and description, and is associated with a host, user, process, and file.
User Interface
Bookmarks

- Sensitive File Access
- Suspicious Process Behavior

Add Bookmark
Alerts (cont.)
Export

- Bookmarks
- Alerts 1
- Export
- Mike

- Raw Data
- Visualization

Graph showing data over time from 08:30 to 10:30.
Scenario 1

• Sally downloaded a file (vacationPhoto) from an address she recognized (mrSally@hotmail.com)
• She double-clicked on the file, and it suddenly disappeared
• She checked the email again and noticed that she had misread the address (mrSa11y@hotmail.com)
• She called her IT department to explain the situation
Phishing
Scenario 2

- The shellshock bug has been discovered and unveiled to the public
- ACME hosts a website on apache web server that stores credit card information
- Carl, ACME’s network administrator, rushes to upgrade his server’s bash version
- It is Carl’s responsibility to make sure that the company's data was not exposed during time before the patch was deployed
Shellshock
Evaluation

• User study with ISD and SSD
  – Noted that NavPro bridges the gap between network- and host-based forensic tools
  – Were impressed by the power of the tool and the context it provided with every action

• Requirements
  – Completed all baseline and realistic outcomes
  – Completed 5 additional features outside original scope

• Maintained scrum development process
• Identified realistic scenarios to solve
Acknowledgements

• George Heineman – WPI
• Jeff Diewald – Group 58
• Tom Moyer – Group 58
• Nabil Schear – Group 58
• Sean Cavanaugh – ISD
• Paul Burke – SSD
• Ted Clancy – WPI
• MIT Lincoln Laboratory
• Worcester Polytechnic Institute
Questions?
Baseline Requirements

- NavPro will be accessible through a web browser.
- NavPro will allow a user to input Linux Provenance Modules (LPM) HiFi data.
- NavPro will allow a user to view activity performed on (or by) an entity (user, process, or file).
- NavPro will simplify provenance data from system calls to readable actions.
- NavPro will allow a user to search for activity based on different types of entities.
- NavPro will be deployable in an automated way.
Realistic Outcomes

- NavPro will have the capability to accept provenance data from different data sources, operating systems, etc. through an extensible plugin system.
- NavPro will allow a user to monitor a network of computers from the web application.
- NavPro will allow a user to view activity performed by a specific host in a network.
- NavPro will allow backward and forward navigation through visualizations.
- NavPro will allow a user to export the raw data from a table or chart visualization based on the current filter set.
- NavPro will allow a user to set alerts for entities (hosts, users, processes, files) and be alerted when an event occurs involving that entity.
Beyond Initial Scope

• NavPro will display condensed high-level versions of “readable” actions.

• NavPro will quantify relationships between entities through search result filters.

• NavPro will live update data in view of the currently visualized entity.
  – Note: This feature was implemented, but later removed.

• NavPro will allow a user to save visualizations as bookmarks so they can be revisited later in the session.

• NavPro will track provenance events that cross multiple host machines.