Security Log Analysis

Kay Avila

Cal Poly Pomona/CTSC Scholarship for Service Workshop
October 14, 2017
Background

- Thank you
  - Warren Raquel and Vlad Grigorescu (NCSA)
  - Mark Krenz (IU CACR)
- About me
Why log analysis?

- Logs are ubiquitous
- Logs are underappreciated
- Logs will make you an IT hero
Defining “logs” and “log analysis”

- Log
  - Any system- or device-generated record (including intelligence feeds)

- Log analysis
  - Collection
  - Event management
  - Analysis
  - Response
Log collection sources

Rough categories of locally produced logs:

- **System logs**
  - Examples: syslog, Windows event log

- **Network logs**
  - Examples: traffic metadata, network services

- **Application logs**
  - Examples: web server access logs, Shibboleth
Log collection sources, cont.

- Intelligence feeds complement the logs
- Internal or external, closed or open
- External examples:
  - REN-ISAC Security Event System
  - Critical Stack
  - Team Cymru Malware Hash Registry
Event Management - Questions

- How and where will the logs be sent?
- What format are they in (and can it be converted)?
- What rate of incoming data must be supported?
- How will the logs be stored and later retrieved?
- How long will the logs be stored? Should only some be preserved, or all?
- Is it a big deal if we miss some?
Event Management - Guidelines

- Filter where possible
- Centralize
- Give access on a need-to-know basis
- Consider a log management tool
  - Examples:
    - ElasticStack (ElasticSearch + Logstash + Kibana)
    - Splunk
Analysis and Response

- Methods to gain insight
  - Manual investigation
  - Customized alerts
  - Automated intelligence
  - Deep dives
- Response is what you do with the insight
Getting Started

- Why?
  - Something *is* happening
  - Something *has* happened
  - Something *will* happen

- Utilize the scientific method
Blue Waters Supercomputer
https://bluewaters.ncsa.illinois.edu/

- Cray XE/XK Hybrid
- AMD 6276 “Interlagos” processors
- NVIDIA GK110 (K20X) “Kepler” accelerators
- Cray Gemini torus interconnect
- Peak performance >13 PF
- ~27,000 compute nodes
- 1.6 PB of memory
- 25 PB of online disk storage
- 500 PB of disk storage
Apache Log Format

22.64.26.43 - - [19/Sep/2017:00:16:21 -0500] "GET /getting-started HTTP/1.1" 200 11820 "https://www.google.com/" "Mozilla/5.0 (Windows NT 10.0; Win64; x64; Trident/7.0; rv:11.0) like Gecko"

Requestor’s IP
Date/time of request
Path requested
Server response code

Size of response
Referer
User-agent
Basic Command Pattern

cat apache_access.log | <command> | <command> ...

One Last Note

- We all come from different backgrounds - which is great!
- This exercise will be too easy for some people and too hard for others.
  - Please ask (and/or answer) questions!
- The code provided emphasizes teachability over elegance.
  - Thank you, yes, I’m aware there’s UUOC. 😊
We thank the National Science Foundation (grant 1547272) for supporting our work.

The views and conclusions contained herein are those of the author and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the NSF.