Cray EX40 (Chicoma) Intrusion Detection Project

Daniel Wild | University of New Mexico | HPC-SYS

Project Goals

GUIDING QUESTION:

To what extent can HPC identify and alert on anomalous cluster activity through performing network traffic analysis without impacting cluster resources?

PROBLEM:

There is a potential for misuse or abuse of cluster resources. Does HPC have the appropriate tools to identify such activity in a timely manner?

METHOD:

Developed Ansible roles for Suricata and Zeek to monitor and alert on cluster traffic from data forwarded through a mirrored port to a security intrusion detection server. Forwarded alert data to Splunk using rsyslog. Designed Splunk dashboards for Suricata and Zeek to analyze data.

Security Host Architecture

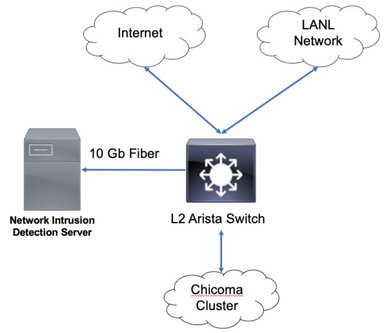

Chart 1. Security IDS Architecture

Results

SSH FREQUENT AUTHENTICATION FAILURES

Frequent SSH authentication failures from the same IP address (1610 attempts in 14 hours) which was the result of client misconfiguration by users.

- Utilization of a file transfer client which cached credentials, resulting in authentication failures due to consistent timeout of credentials
- Connection failure resulting in a Git command failing over to askpass that was not setup correctly
- Script failures

These issues were brought to the users' attention and resolved.

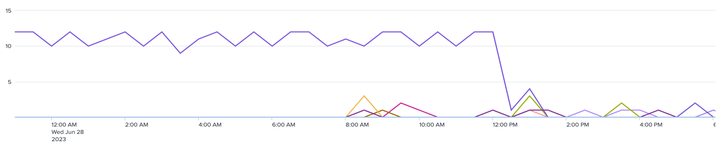

Chart 2. SSH Authentication - Resolved

HTTP TRAFFIC WITH APIPA ADDRESSES

HTTP traffic was seen to be frequently (~37% of all HTTP traffic) rejected by the LANL Web Proxy. Much of that traffic was addressed with Automatic Private IP Addressing (APIPA) addresses. That traffic was probably intended for the localhost but due to misconfiguration was forwarded to the LANL Web Proxy which rejected the traffic.

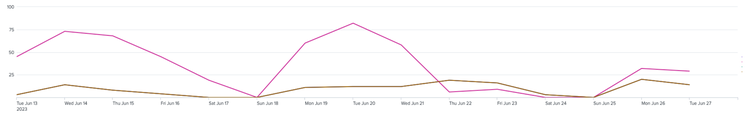

Chart 3. HTTP Forbidden -- APIPA

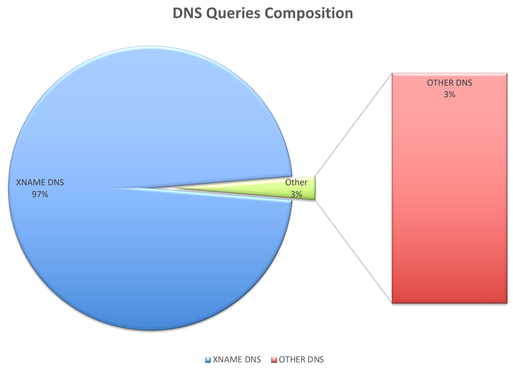

Chart 4. DNS Query Composition

DNS TRAFFIC ANOMALIES

Anomalous DNS requests were detected due to switch and node misconfiguration.

- DNS requests (approximately 17k/hr) directed to Google DNS service (8.8.8.8).
- DNS requests (approximately 40k/hr) directed to the institutional DNS server for xname (Cray specific node name schema) hosts that do not exist.

The Google DNS requests were generated because of a default configuration on 9 Aruba switches which enabled Aruba Central Cloud. This configuration was corrected, and these DNS requests were eliminated on June 7, 2023.

Chart 5. Google DNS Resolved

The cause for non-existent xname (a Cray-specific node name convention) DNS queries is still being investigated via support ticket #227610. 489 non-existent hosts are typically queried 90-100 times every hour comprising approximately 97% of all DNS traffic on the cluster. These lookups are not resolvable and therefore forwarded to the institutional DNS, causing a larger impact than a local resource issue.

- Why is a node performing xname lookups for hosts that do not exist?
- Are any user workflows impacted by failed requests?
- Impact on institutional DNS resources and other DNS clients

INVALID CERTIFICATES DETECTED IN STREAM

Four certificates were identified as invalid by Zeek, which uses the Mozilla certificate store to validate the certificate authority chain. Institutional certificates were converted into Zeek's format of hex 1-byte chunks delineated by "\" . The certificate issues persisted after adding the institutional certificates to Zeek because most of the "invalid" certificates were intentionally self-signed.

Chart 6. Certificates – "Invalid"

SMTP TRAFFIC

SMTP traffic on the network was generally related to slurm jobs which include the job status (completed, failed, or timeout). This added a metric of network health specifically regarding slurm jobs.

Issues and Challenges

DETERMINING BASELINE FOR NETWORK TRAFFIC

1. A significant challenge was packet loss due to singular elephant flows that overwhelmed the buffers.
   - An eBPF filter used to bypass larger flows.
2. It was challenging to determine what was "normal" traffic in the cluster compared with a conventional network.
   - Anomalous TCP stack issues such as Split routing to/from the backend.
   - Successful anonymous LDAP bindings.
3. Baseline was difficult to establish because metrics such as network traffic bytes to cluster utilization had no direct correlation.

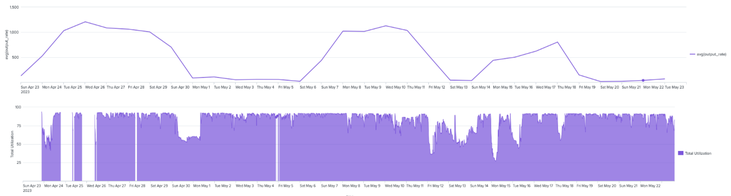

Chart 7. Corelate Cluster Utilization and Network Traffic

Added Functionality

ANSIBLE ROLES

The security intrusion detection system build included creating Suricata and Zeek Ansible roles for future production deployment. The project can be found at: Project Gitlab repository

ZEEK CUSTOMIZATION

A Zeek specific script was created to identify when files are downloaded specifically using curl, wget, or aria2 user agents. Downloaded files are logged by Zeek along with their md5 hash in a custom log file. A Python script then performs an API call to VirusTotal to generate a report for that file which can identify malicious files.

Mentors: Skip McGee & Thomas Areba

Los Alamos NATIONAL LABORATORY
NNSA National Nuclear Security Administration
Managed by Triad National Security, LLC, for the U.S. Department of Energy's NNSA.
LA-UR-23-28604