LA-UR-19-27103

sFlow Monitoring for Security and Reliability

Xava Grooms
University of Kentucky
xavagrooms@uky.edu

Robert Rollins
Michigan Technological University
rvrollin@mtu.edu

Collin Rumpca
Dakota State University
collin.rumpca@trojans.dsu.edu

MOTIVATION

- Through the utilization of sFlow, this project aims to improve security and reliability by identifying network abnormalities
- Los Alamos National Laboratory (LANL) possesses switches capable of using sFlow, which are not currently being utilized for monitoring
- Practical and scalable monitoring solutions are imperative to HPC improvements
- As High Performance Computing (HPC) systems move towards exascale, monitoring remains an afterthought of design

sFlow

sFlow is an open-source network protocol that collects network traffic on a switch using a packet sampling method. Packet sampling involves taking a portion of the data collected for analysis, instead of processing all packets. This allows for large data flows to be processed and for the data to be analyzed more efficiently. In this research, network data was forwarded to Splunk, a data visualization software, from the sFlow enabled Arista Switch.

SYSTEM SETUP

Dell PowerEdge Servers	Switches
9-Node Compute Cluster	Dell Ethernet
Master Node	Arista Ethernet
External Server	Mellanox Infiniband

Each node ran the CentOS 7 operating system. The master node and external server contained hard disks for stateful boot, while the compute nodes were diskless for stateless boot. ***Figure 2*** is a dashboard created in Splunk that depicts the cluster topology.

How can HPC monitoring be improved?

sFlow enabled switches provide an **open-source** solution for better **HPC cluster monitoring**.

Splunk dashboards created from **sFlow data** can improve the **security** and **reliability** of an HPC cluster through **real-time anomaly detection**.

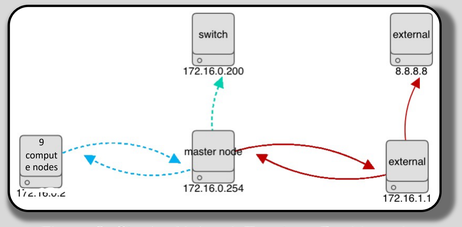

Figure 2. Cluster Network Topology Dashboard

RESULTS

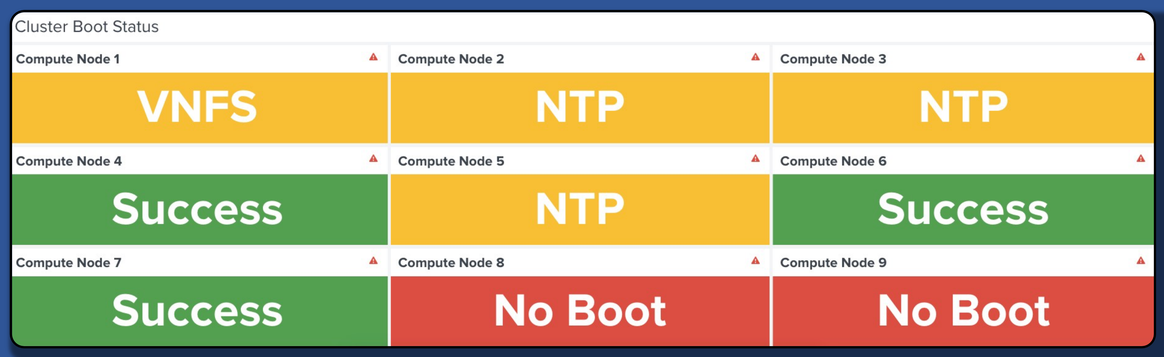

Figure 1. Cluster Boot Dashboard with service status indicators

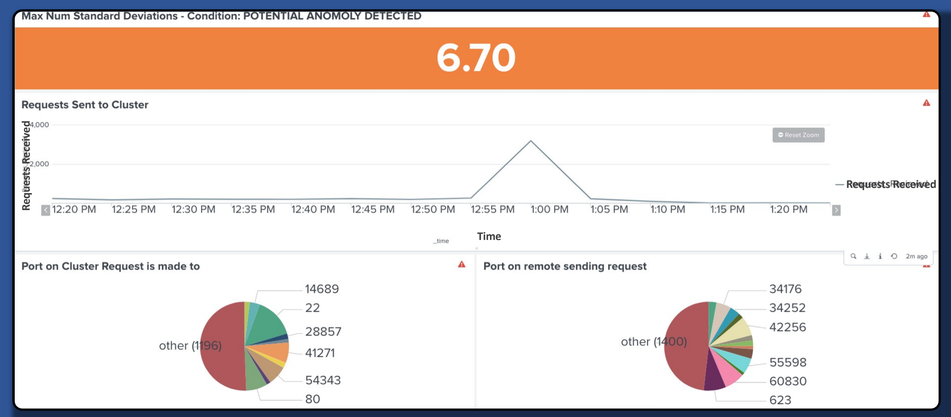

Figure 3. Port Traffic Dashboard with standard deviations from average, requests sent to cluster, and ports used.

ANOMALY TESTING

- Various network services were mimicked or exploited during the abnormality testing stage of this research
- **Standard Network Traffic:** First, a baseline of known good traffic was created. The protocols that were simulated was Hypertext Transfer Protocol (HTTP), Secure Shell (SSH), and File Transfer Protocol (FTP)
- **SSH Enumeration:** SSH Enumeration is an attack on a server that attempts to find valid usernames with a brute force method
- **DNS Tunneling:** DNS tunneling is an exploit that circumvents system firewalls and can go undetected if a trusted HTTP domain server is used
- **Mass HTTP Data:** Mass HTTP Data was used to create an asymmetric data flow from the cluster

CONCLUSION

Splunk dashboards were created from data captured through sFlow and have proven to aid in HPC monitoring for security and reliability. The ***port traffic dashboard*** shown in ***Figure 3*** allows for the identification of anomalies in network traffic for improved security. The status of the boot process of the cluster can be determined by the ***cluster boot dashboard*** shown in ***Figure 1***, which enhances the reliability of the cluster. Through real-time alerts within the Splunk dashboards, anomalies can be easily detected via data captured with sFlow.

ACKNOWLEDGMENTS

- Mentors: Michael Mason, Marc Santoro, Nicholas Jones
- CSCNSI Leads: Catherine Hinton, Reid Priedhorsky
- CSCNSI Instructor: J. Lowell Wofford
- HPC Division at LANL