CERIAS

The Center for Education and Research in Information Assurance and Security

Visualization of Network Traffic on Purdue High Performance Computing Resources

MaKayla McCartan[1], Akash Ravi[1], Erik Gough[2]

1. Computer and Information Technology
2. Purdue IT, Rosen Center for Advanced Computing (RCAC)

Overview

Purdue University is home to several high performance computing (HPC) resources, including campus computing clusters, storage systems and Anvil, a $10M NSF funded supercomputer. These HPC resources are connected to a “Science DMZ” (Figure 1) network designed to provide a friction-free path supporting low latency, high-speed data transfer. A Zeek-based intrusion detection system called PULSAR (Purdue Live Security Analyzer) is used for network monitoring of the Science DMZ. The IDS processes and stores JSON logs at a rate of **thousands of events per second**.

The scale and format of the IDS logs makes it almost impossible to manually investigate traffic trends without a Security and Information Event Management (SIEM) system. A SIEM was built to provide this functionality, where IDS logs are shipped using Filebeats and indexed in ElasticSearch via a log ingest pipeline that enriches the logs with GeoIP data. Kibana provides easy access for traffic filtering and visualization.

In this work, we use the SIEM to produce visualizations of network traffic on the Science DMZ, showing interesting traffic and attack trends for Purdue's HPC resources.

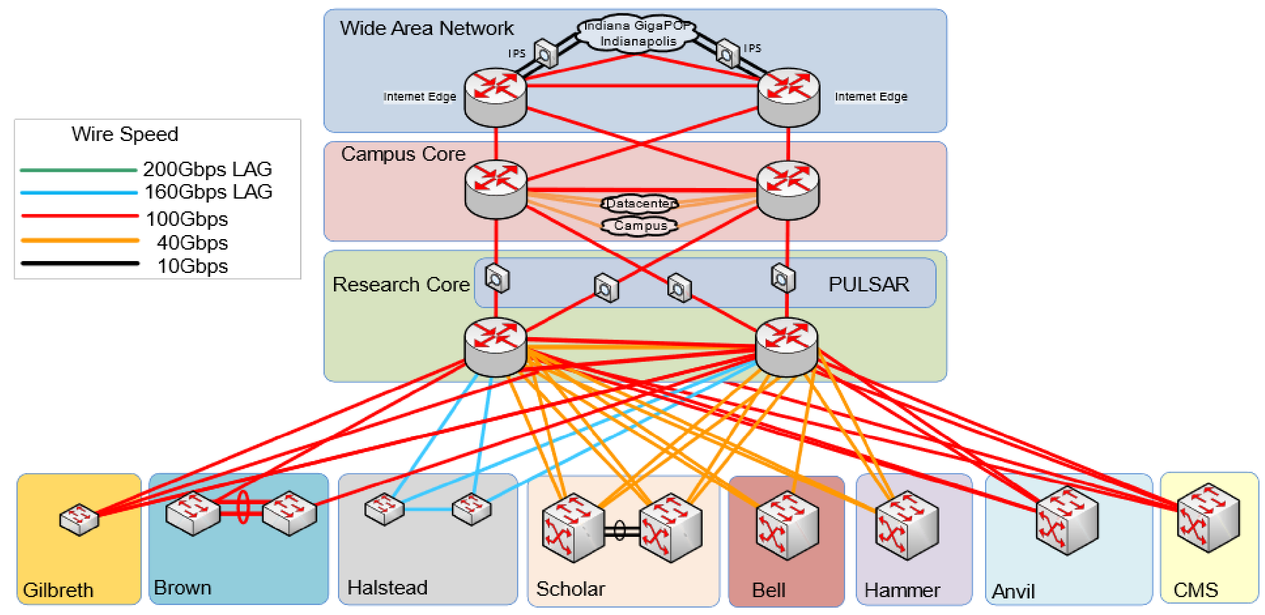

Figure 1. Purdue Research Network Diagram and Science DMZ

Visualizations

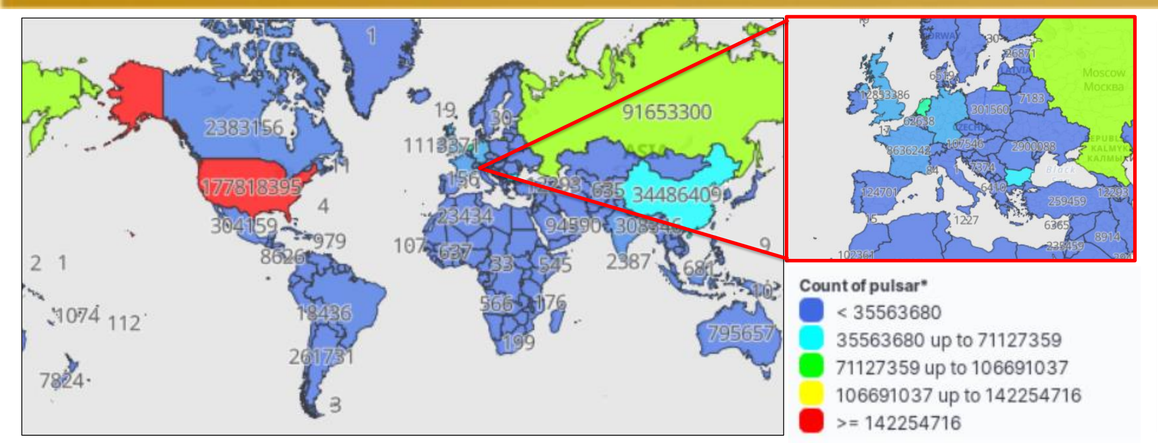

Figure 2. World Map of Port Scanning Origin

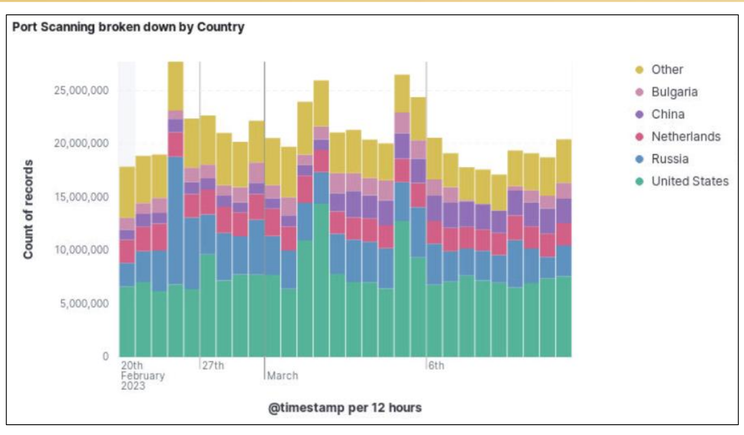

Figure 3. Top Countries for Port Scanning Origin

Figures 2 and 3 show port scanning from geographical origin of the scanning host. The data shows where the most port scanning attacks originate from. **Figure 3** specifically shows the countries with the most port scanning origins detected by the IDS.

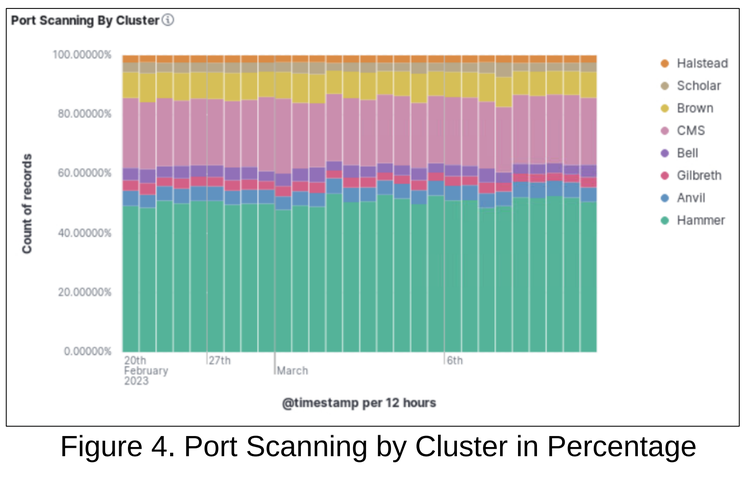

Figure 4. Port Scanning by Cluster in Percentage

Figure 5. Different Traffic Types in Percentages

Figure 4 shows the baseline percentage of port scanning each cluster receives on the Purdue research network.
Figure 5 shows the different types of network connections that are logged by the IDS. The combination of these graphs demonstrates most of the network connections on the Science DMZ are port scanning attempts.

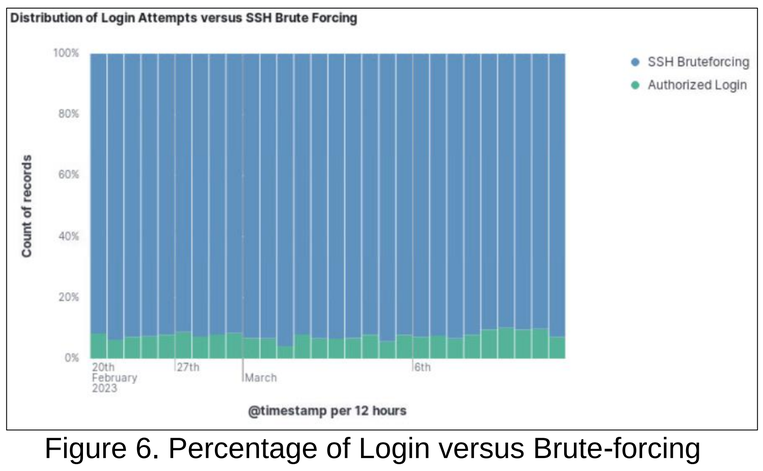

Figure 6. Percentage of Login versus Brute-forcing

Figure 6 shows the percentage of authorized login versus SSH brute-forcing attempts detected by the IDS. Implementation of BoilerKey two-factor authorization minimizes the threat of successful SSH brute-forcing attempts.

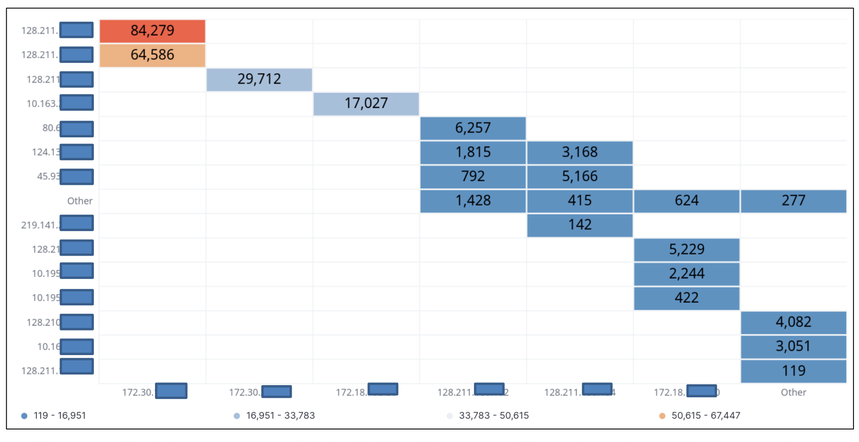

Figure 7. Heatmap of MySQL Source/Destination Connections

Figure 7 describes the number of connections from source and destination connections to different MySQL database servers on the Research Network. This shows at least two database services that are being accessed on the public Internet.

Future Work

We have shown the SIEM can be used to easily filter, create visualizations and gain insight into network traffic trends on the Purdue Science DMZ. In future work, we plan to implement additional SIEM-based alerting and automated threat mitigation through route filtering of traffic from identified attackers.

Acknowledgment

This work is supported by Research Experiences for Undergraduate (REU) funding for NSF OAC Project #2005632: Category I: Anvil - A National Composable Advanced Computational Resource for the Future of Science and Engineering

PURDUE UNIVERSITY

CERIAS