Flexible visualization of a 3rd party Intrusion Prevention (Security) tool: A use case with the ELK stack

M D Poat[1], J Lauret[1], D Fedele[1],
[1] Brookhaven National Laboratory, P.O Box 5000, Upton, New York 11973-5000, USA

mpoat@bnl.gov, jlauret@bnl.gov, fedele@bnl.gov

Abstract. A difficult aspect of cyber security is the ability to achieve automated real time intrusion prevention across various sets of systems. To this extent, several companies are offering comprehensive solutions that leverage an "accuracy of scale" and moving much of the intelligence and detection on the Cloud, relying on an ever-growing set of data and analytics to increase decision accuracy. Often, they provide tools to visualize the decision workflows in attack prevention (as well as tune the algorithm) but those solutions are not always practical as companies see the problem as "global" that is, from a unified Cyber-security standpoint. However, a key to a successful Cyber-security program is transparency and trust: from an experimental team viewpoint, this specifically means having the ability to immediately see what and from where, who has been blocked and being able to inform the community in case of a revoked access without the need for filing a "ticket" (that may eventually be answered) – in other words, rapid response to their user-base is essential but solutions targeting "sub-groups" in an organization are not often available.

We have come up with a versatile solution leveraging the ELK stack (Elasticsearch, Logstash, & Kibana) and an IPS (Intrusion Prevention System) based WAF (Web Application Firewall) from Signal Sciences. Signal Science allows the streaming of detailed logs in a Logstash format suitable for custom solutions for visualization. By combining these two tools, we have strengthened our security posture and enabled individual experiments to monitor their own traffic. Specifically, the IPS WAF provides unique data such as country of origin, protocol, response code, source IP, and paths accessed.

In this contribution, we will show how we engineered a visualization solution so experiment groups could access a dashboard with predefined graphs but also, where they can create individual customizable dashboards used to display blocked traffic and troubleshoot latency issues. We will discuss the details and procedures for developing and configuring these tools and how it benefits cyber security postures across our scientific based environment.

1. Introduction

A perfect cyber security tool is one all can agree on. If both the cyber teams and the system administrators (whose systems are monitored and managed by cyber) agree with the tools being used, security and productivity should thrive. One of the challenging aspects of cyber security is the ability to achieve automated real time intrusion prevention across various sets of systems. Many tools offered by

large companies have moved their tools to the cloud offering scalability with the intelligence and detection updated in real time one cannot match with a home-based only solution. Typically, cyber teams view issues as global (whole organization) which can lead to closing off details to the groups they monitor, hence leaving groups in the dark to what is happening on their systems. From a scientific experiment viewpoint, lacking the ability to see what cyber tools are doing to our systems in real time becomes a detriment when troubleshooting issues relating to network blocks, user access prevention, and malicious attacks especially when 100% uptime is needed. By working closely with our site cyber team, we have come up with a flexible solution leveraging the ELK Stack (Elasticsearch, Logstash, Kibana) and an IPS (Intrusion Prevention System) based WAF (Web Application Firewall) from Signal Sciences. By combining these two tools we have enhanced our security posture and enabled individual experiments and groups to visualize and monitor their own web traffic

2. IDS and IPS

In cyber security there are two types of intrusion tools. Intrusion Detection Systems (IDS) and Intrusion Prevention Systems. IDS are only for detection and monitoring; it will not act on its own and requires human intervention to review the results. An IPS on the other hand will act on its own and will automatically accept/reject traffic based on its ruleset [1]. An IPS does rely on a database that is continuously updated with new threats. Both tools analyze network traffic and compare the contents to their database of know threats. Cyber groups benefit from deploying both an IDS & IPS.

3. The ELK Stack

The ELK stack is a group of opensource products from Elastic that is both distributed and scalable. ELK stands for Elasticsearch, Logstash, & Kibana are the three core components of the ELK stack. There is an additional component known as Beats which are data shipper agents used to send specific data to your ELK stack. For our work with the IPS WAF & ELK, we only require the three core components. Elasticsearch is the distributed search and analytics engine at the heart of the Elastic Stack [2] Elasticsearch will store and index diverse datasets, provide real-time search, and analytics. Elasticsearch goes far beyond simple data retrieval and formation with machine learning and other data analytic tools built in. Logstash is the data collection engine with real-time pipelining capabilities [3]. Logstash can take input data from multiple sources, format the data, and normalize it into a unified data set. The data is then shipped to Elasticsearch where it will be indexed. Lastly, Kibana is the visualization tool for ELK enabling users to search, analyse, and visualize their data [4]. You can view your logs in a formatted fashion, create charts, graphs, and leverage the power of visualization to provide the best overview of your data.

4. Signal Sciences WAF

Signal Sciences offers a cloud based WAF (Web Application Firewall) IPS (Intrusion Prevention System). Typically, a WAF will require extensive manual tuning which may not always be effective and can produce false positives. The Signal Sciences WAF is unique as it is not limited to the common legacy detection and blocking techniques. Signal Sciences uses their proprietary software tools called SmartParse [5] and Power Rules [5] to protect web servers against intrusions and attacks such as SQL Injection, Cross Site Scripting (XSS) attacks, Directory Traversal attacks, and other OWASP (Open Web Application Security Project) Top 10 injection attacks [6]. SmartParse makes highly accurate detections by analysing requests parameters to determine whether the code is actually executable and tokenizes the results [5]. With Power Rules, privileged users can setup thresholds, automatic blocking, and alert triggers specific to their web application and business logic within the Signal Sciences Console [5].

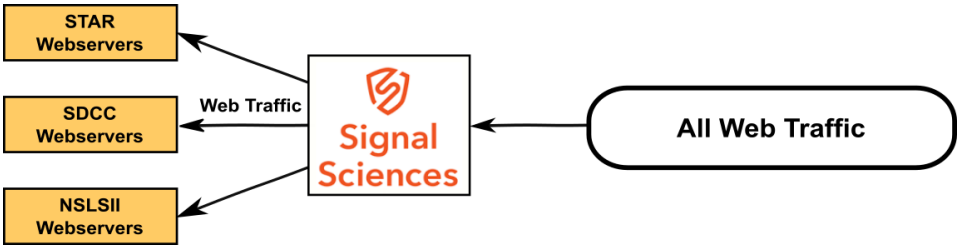

Figure 1. Diagram illustrating the web traffic flows through the Signal Sciences WAF before it reaches the individual webservers.

As shown in Figure 1, all traffic including malicious traffic must be filtered through the WAF before it reaches the webservers. All the web traffic that is filtered through the WAF, both good and bad, generates traffic log output. Signal Sciences offers a Dashboard to view the traffic logs, but it can only show all the WAF data for all webservers instead of individual webserver or domain specific data. This is an issue as we want to share WAF data to specific users and groups for webservers that pertain to them only. Fortunately, the Signal Sciences WAF can export the web traffic logs to Logstash which we are then able to inject into our ELK stack. The ELK stack will provide users an authorized mechanism to view their specific data.

5. WAF & ELK Stack Workflow

By combining the WAF IPS & the ELK stack we can enhance our security posture and visualize all web traffic in real time with historical lookback. If the user is an attacker and attempts an SQL Injection, Cross Site Scripting attack, etc. the traffic will be rejected with a Signal Sciences response code of "406". The 406 response code (similar to an HTTP 406 Not Acceptable Request) [7] is very useful as it provides a defining metric to determine the majority of good and bad traffic.

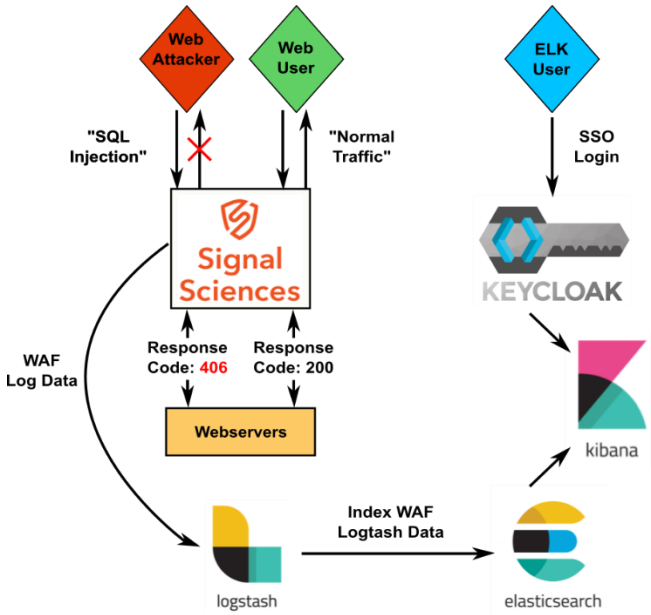

Figure 2. Diagram illustrating the entire workflow from initial web traffic all the way to the ELK dashboard user.

As shown in Figure 2, A user or attacker will access the webservers while unknowingly have its traffic filtered through the WAF. All the traffic logs are then forwarded to Logstash, formatted in JSON, then

forwarded to Elasticsearch. When we receive the initial log payload from at Elasticsearch, it needs to be indexed. Indexing ensures individual data points are assigned the correct data type, whether it's a date, text, IP address, or number. With a non-standard Elasticsearch data stream, such as the WAF data, it may not import properly without special filtering. Initially, the timestamp in our data was incorrect which led us to writing a custom Java Kibana scripted field to correct this. While were able to resolve this on the Logstash side in the end, it was a useful practice as one could customize a hard input data stream at the Index Pattern level. Once the data was indexed, the data stream could be seen in the Kibana web interface. We then setup a Kibana dashboard where we create our visualizations.

6. Kibana Dashboard for Visualizing the WAF

Kibana Dashboards allow us to create custom visualizations to see any aspect of our data. Kibana does provide many built in visualizations for ELK provided plugins, but with custom data you will likely need to create your own. When designing a Dashboard for both real-time and historical cyber intrusions, we found the most useful visualizations are such that immediately show what is happening on the systems along with charts that can provide historical lookback to see general trends.

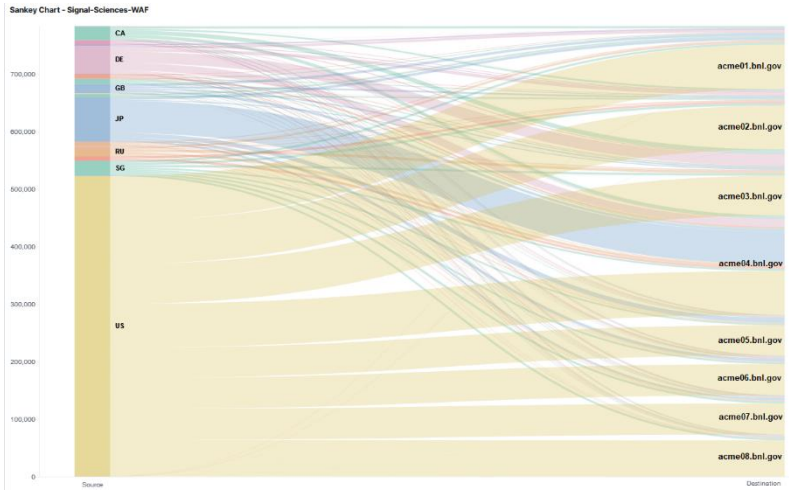

Figure 3. Sankey Chart showing which countries are generating what amount of traffic to each specific webserver.

As shown in Figure 3, the Sankey Chart is showing the flow rate of traffic from individual countries to each webserver. As is it shown, two thirds of all traffic come from within the US while the remaining one third comes from all other accessing countries combined. The time frame of the Sankey Chart can be changed in real time, therefore during attacks or troubleshooting we can see which country is the origin of traffic.

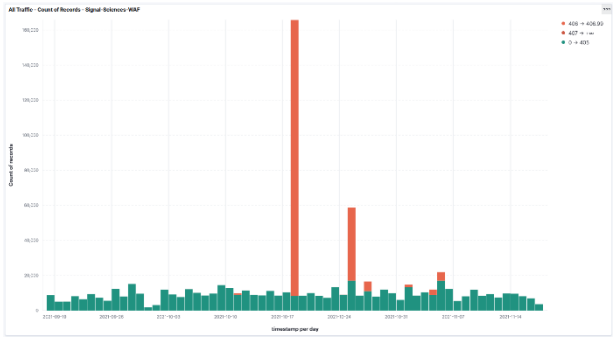

Figure 4. All Traffic bar chart showing all traffic color coded by green (normal network traffic) compared to red (attack traffic).

As shown in Figure 4, the "All traffic" bar chart provides a quick overview of all the web traffic divided by color for good and bad traffic. With this simple interface, users can click on the individual bars in the chart, which will automatically reframe the Kibana dashboard to the specific time frame. This helps pinpoint where the attacks are coming from, the number of attacks and will reframe more detailed charts providing the actual details of the attacks. As shown in the chart, the large red spikes are attacks that typically come in large waves from specific IP addresses.

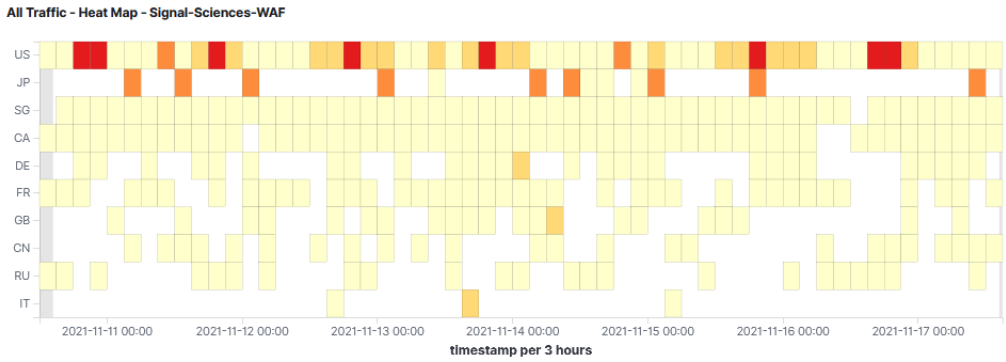

Figure 5. Heat map showing which countries are generating X amount of traffic in each time frame

As shown in Figure 5, the heat map shows us which countries are generating X amount of traffic in a given time frame. The red color is heavier traffic while the lighter colors is less traffic. Users can click on different colored boxes which will "zoom in" on the chosen timeframe.

7. Conclusion

Combining the ELK stack to visualize Signal Sciences WAF data has provided a useful cyber security tool to our user community. This tool has empowered users with insights to their traffic patterns as well as network blocks. The requirements to deploy an ELK stack for the Signal Sciences WAF is quite light. Around 6GB of data / year / experiment, 16-32GB RAM nodes, and SSDs are recommended. Rather than a black box, the end-users now feel more at ease, lowering the bar for WAF adoption within their ecosystem. Experimental groups are more open to new security tool as well as they have first-hand insight to the data and further enabling Cyber teams to enhance the overall security posture. A Cyber tool where both the Cyber teams and the end-users agree to deploy is a win-win.

Acknowledgements

This work was supported by the Office of Nuclear Physics within the U.S. Department of Energy's Office of Science.

References

[1] IDS vs. IPS, Jeff Peters 2020 - https://www.varonis.com/blog/ids-vs-ips
[2] Elasticsearch - https://www.elastic.co/guide/en/elasticsearch/reference/current/elasticsearch-intro.html
[3] Logstash - https://www.elastic.co/guide/en/logstash/current/introduction.html
[4] Kibana - https://www.elastic.co/guide/en/kibana/current/index.html
[5] Signal Sciences "Detection and Blocking" - https://info.signalsciences.com/detection-and-blocking
[6] OWASP Top Ten - https://owasp.org/www-project-top-ten/
[7] HTTP Not Acceptable - https://developer.mozilla.org/en-US/docs/Web/HTTP/Status/406