CHALLENGES OF COLLECTING LARGE AMOUNTS OF LOG DATA:

Oftentimes, collecting, logging and monitoring large amounts of data can prove challenging for security analysts.

In Project 1 of my Cybersecurity Bootcamp, we collected and investigated log data. To begin, I created and downloaded a Filebeat and Metricbeat config file and playbook via Ansible. Utilizing the ELK server I created, Filebeat ships log files and plays the role of logging agent. Metricbeat periodically collects metrics from the operating system and from the ELK server. Both Filebeat and Metricbeat send the collected data to a specified output, and in this Project I investigated sample data that was imported to Kibana for visualization and investigative practice purposes.

This Project required me to search for and interpret data such as time frame, where website users and visitors are located, which machine operating system they were using, and what country produced traffic on the website. Furthermore, data such as HTTP error response codes, types of downloaded files, and the size (in bytes) of the downloaded files were analyzed for suspicious activity. In this case, it was found that on 2022-03-21 at 18:00, a user originating in India downloaded an RPM file that was 16,837 bytes. I was able to find the source IP address and the geo coordinates of the activity as well as the OS of the source machine being Windows 8. There was also cause for suspicion because I was able to find that the visitor’s traffic originated from Facebook.com.

In order to find this information on Kibana, I needed to create a security rule within my ELK server to allow port 5601 access. This was created in my Azure Portal ELK Virtual Machine. Within Kibana, I was able to find this information using the LOGS Dashboard as well as the Logs and Metrics tabs. I was able to filter by source country, interpret the charts “Unique Visitors vs. Average Bytes” (which found the largest download in the last 7 days) and “File Type Scatter Plot” (which helped me figure out that the file which was downloaded was an RPM file, and more - to find much of the latter information I described. The Logs page on Kibana provided the information on the source and destination countries, the source IP address, the geo coordinates, the URL that was accessed by the visitor, the website that the visitor’s traffic originated from, etc. These tools and graphics used on the Kibana webpage were useful in finding and interpreting the data and finding the necessary information.

While the sample data on Kibana yielded large amounts of Log Data, Kibana also provided resources which made it easier to filter and sort by my desired parameters. The individual charts and graphs were adjustable to the dates, times, regions, etc. that I wished to research. This made it easier to elucidate the specific timestamp and details that I needed for the Project.

In this Project, I just scratched the surface of Kibana and I did not have to inspect malicious or possible nefarious activity or data, nor did I have to look at advanced web traffic data. Having to access this additional data (as well as the multitude of other data available on Kibana) may have proven more difficult and time consuming - however, I believe that I would be able to tap into my knowledge of Kibana and how the filters assist in data collection and analysis. This knowledge is beneficial and can be translated into larger data sets and used to extract more information depending on the situation.