Applying a Wide-Angle Lens for a New Take on Cybersecurity Use Cases

Introduction

Large cybersecurity data lakes enable deep-diving into data from multiple sources at a very specific point in time to elucidate unexpected or concerning activities, or performing specialized investigations of a particular user, device, or connection over a time range. However, an equally compelling set of use cases can be delivered by taking a broader view of the data that would not otherwise be possible without the availability of a large-scale data lake with substantial compute power.

This "wide-angle" view of the data enables the cybersecurity data scientist to achieve two perspectives. First, by encouraging the user to stand back and view the data holistically, important context can be observed that is not visible any other way. Second, the broader context supports the data scientist in the interpretation of findings.

Solution

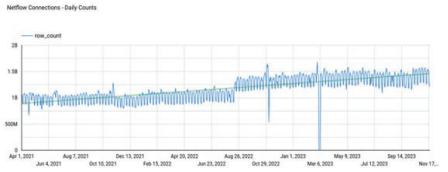
Carefully crafted queries were defined and executed based on their potential to provide network perspective at the highest levels. Connection volumes/frequencies, ports and protocols, byte volumes, destinations, and other views of the data were generated, and results evaluated for various time ranges. Learnings were extracted to support operational as well as threat detection objectives, providing a unique perspective only made available through the implementation of a large-scale data lake.

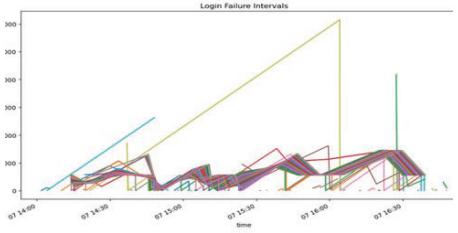
Informative Metrics Based on **Netflow Data**

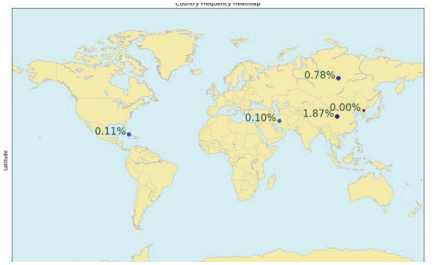
1. Calculate percentages of internal and external (to your network) IP addresses, within overall connection load (i.e., internal to internal, internal to external, etc.)
   - Evaluate any significant changes over time to determine if they are explainable
2. Calculate the percentage of each type of protocol (based on protocol ID/port) as part of the overall connection load
   - Evaluate any significant changes over time to determine if they are explainable
3. Calculate the number of connections where bytes are sent but none are received
   - This may indicate that the destination address is not responding as expected
4. Calculate the number of connections where no bytes are sent but bytes are received
   - Some may be explainable, but also may indicate an anomaly worth evaluating
5. Calculate the number of connections occurring on ports assigned to applications that are prohibited by policy at your organization (TeamViewer, AnyDesk, etc.)
   - Useful for policy monitoring and enforcement
6. Calculate the top n IP addresses that initiated connections in your network, comparing "internal" vs. "external" IP addresses. Recall that the IP address, if dynamic, may not be the same actual device or system generating all the connections occurring in the result set
   - Results may need to be filtered by resolving the historical accuracy of dynamic IP addresses to individual devices
   - Validate the activity observed and evaluate any significant changes over time to determine if they are explainable
7. Calculate the top n connections in terms of bytes sent. Consider the following nuances:
   - If the source of the connection is internal and the destination is external, could be exfiltration
   - If the source of the connection is external and the destination is internal, could be malware staging
8. Calculate the top n connections in terms of bytes received. Consider the following nuances:
   - If the source of the connection is internal and the destination is external, could be malware staging or downloads of suspicious content
   - If the source of the connection is external and the destination is internal, could be exfiltration
9. Calculate the top n connections in terms of connection duration (length)
   - Confirm connections between these endpoints should last as long as they do
10. If your organization has a policy regarding certain operating system configurations (e.g., use a specific set of systems for NTP, DNS, etc.), calculate the number of connections that are using the port(s) for those services but are not targeting the systems identified in the policy
    - Useful for policy monitoring and enforcement

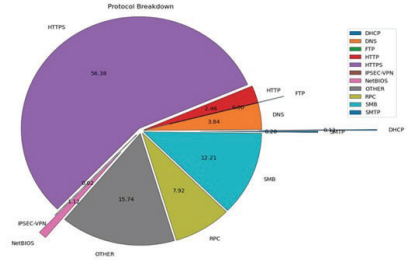
Conclusions

Sample learnings gleaned from taking a "wide-angle view" of our network data (see graphs):









Jennifer Marr, BA; Matthew Spitzer, PhD; and Rosalie Bakken, PhD
Mayo Clinic