Following the Attack Chain

Our presentation covers the Cybersecurity dataset for the UMD Info Challenge 2022. The questions that we sought to answer were: could we build a classifier to differentiate attack packets from normal packets, predict what kind of internet traffic a packet represents, and how to understand using data analysis in order to better defend against cyber attacks.

Our dataset consisted of network packets with columns of various packet details you might find in a pcap such as a Base64 encoded payload, destination and source IPs and ports, and TCP flags. In order to classify between hostile and benign network traffic, we built a neural network based on the TabTransformer architecture proposed in Huang et. al. 2020. This allowed us to project the packet details into a higher dimensional space to learn how correlated values connect within the context of predicting cyber attacks.

To understand our network data, we sought to analyze the time details, location data, and protocols associated with the packets in our network. We found that HTTP and SSH were two of the most commonly used ports as well as several other commonly exploited ports such as Samba, FTP, and default Metasploit ports. We were able to create heatmaps for the location data and pinpoint where external IP addresses were coming from down to the street level. We were also able to create a chart showing the attack flow for the network penetration, identifying one IP address connecting with all the others and probably serving as the initial point of entry through which the attackers conducted lateral movement. Lastly, we were able to identify specific attack chains such as TCP attacks that led to the creation of a secure webshell on one of the victim machines.

We were able to build a classifier that predicted attacks with around 98% success and were able to confirm that some of the most simple cybersecurity mitigations can be taken to stop attacks such as these. Beyond the data, we recommend the following to those trying to defend their networks against attacks such as these. The most common ports for attack should be closed if not necessary and monitored closely if they are. Implementing a network to analyze these packets will also allow defenders to shut down attack chains by identifying a behavior and blocking packets from that IP address.