Abstract

Given network traffic is it possible to find patterns that are indicative of a cyber attack and then use those patterns to detect future attacks? The dataset given is simulated traffic of a simulated network that has labeled communications indicating whether or not individual packets are benign or malicious. Each packet also includes any information that would be recorded by an intrusion detection system (IDS) such as the data in the packet, IPs, and port numbers. First, we complete a cursory analysis of the dataset, including the most popular ports and protocols, the most common times when malicious packets are sent, and geographic locations of attack versus normal IPs. We will then zero in on individual attacks, using data visualization to show the physical progression of an attack over time between machines both outside and in the network. Which is then mapped directly onto google maps using the latitude and longitude acquired form the public IP addresses. Next, we will track from the start of an attack and analyze the traffic within that attack time period to identify common protocols, data, and any other information that indicates an attack is occurring. By understanding the sequence of packet communications that compose the attack chain, this provides vital information for system and network defense. Using this information, we will build a classifier that can identify whether or not a particular packet is malicious or benign.