Joycelyn Chow, Zachary Blumenfeld

16 August 2022

ASPIRE Project Abstract – Summer Session Year 2021-2022

Our goal during the 2021-2022 summer internship session was to build a neural network that can perform classification tasks related to cybersecurity, allowing for the detection of future compromises and attacks. There were specific requirements for our project that we had to follow. We were required to use supervised machine learning, the machine learning library TensorFlow, python to code our training model in VS Code, and the environment Anaconda to run python and manage other packages.

We began by researching how artificial intelligence (AI) was being used in cybersecurity and finding what kinds of datasets were already available online. Since we discovered that the majority of AI cyber applications involve fraud, malware, or other nefarious activities, we found two applicable datasets on website phishing links and IoT botnet attacks. Then, we built two separate algorithms: one to predict whether a URL is legitimate or not and the other to detect abnormal network traffic patterns that could have been triggered by a botnet. However, we encountered challenges with our datasets. The phishing dataset didn’t contain the actual phishing links and informative features while the botnet dataset wasn’t explicitly labeled and was split into different files. Therefore, we had to add a column to the botnet dataset to label malicious and benign data with a one or zero and then concatenate all the files into one. For the phishing dataset, we had to research an entirely new dataset with the desired characteristics. After we solved our issues, we ran our algorithms through the model and continued making adjustments to improve our neural network’s accuracy.