Near Real-Time Anomaly Detection

Project Team: Gavin Cutchin, Gabriella Rivera, Jared Dunn
Computer Science
Project Number CS 23-316

Faculty Advisor(s): Changqing Luo, Ph.D.
Sponsor: Capital One
Mentor(s): Geeta Shyamala, Jake Jeffers, Ben Polk, Andrew Kepley

Machine learning models are a powerful tool for use in adaptive decision making in any given application. Here we assess the effectiveness of machine learning models against cyber security threats to a network of virtual machines where network traffic and attacks are simulated to check both the response time and robustness in decision making from the machine learning model. The model is hosted inside of an Amazon Web Services SageMaker endpoint where each piece of network traffic and corresponding process call(s) gets sent for inference against the endpoint to decide if the network traffic is malicious with a corresponding percent threat value. With an average inference response time of 64 milliseconds, SageMaker allows as close to real-time inference and response without the need to host the model locally. Across a sample of 2000 labelled testing logs we found that we could accurately discern the difference between a log that represented a threat and otherwise with a difference in threat average of 9.7% (fig. 1). Across a sorted random sampling of 100 labelled testing logs (fig. 2) no sus labelled log tested below a 24% threat value and non sus/evil log tested below a 13% value. The model performs as expected with data coming directly from a labeled and source. These results can be used to further enhance cyber security models as well as implement real-time network security solutions.

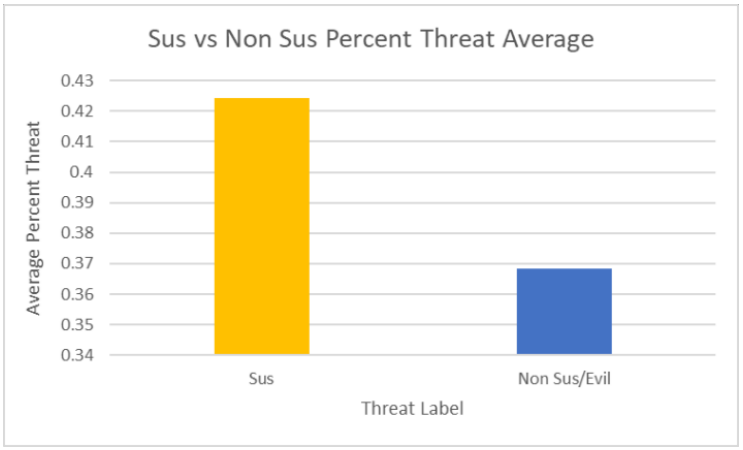

Figure 1 - Average Threat Analysis (n = 2000)

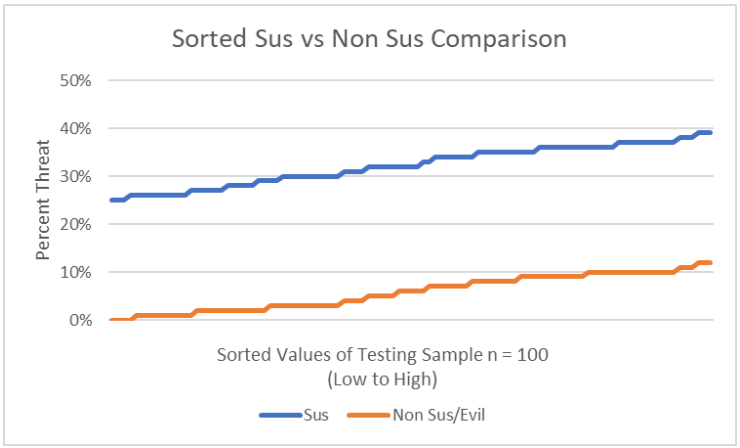

Figure 2 - Sorted Sample Comparison (n = 100)