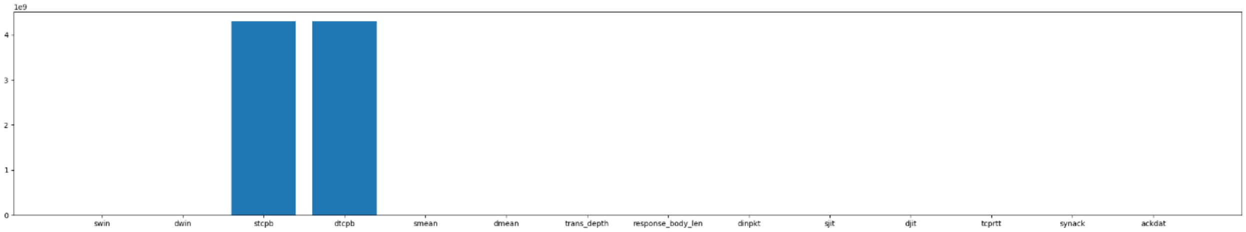
CS-5110/6110

Code Significance Report

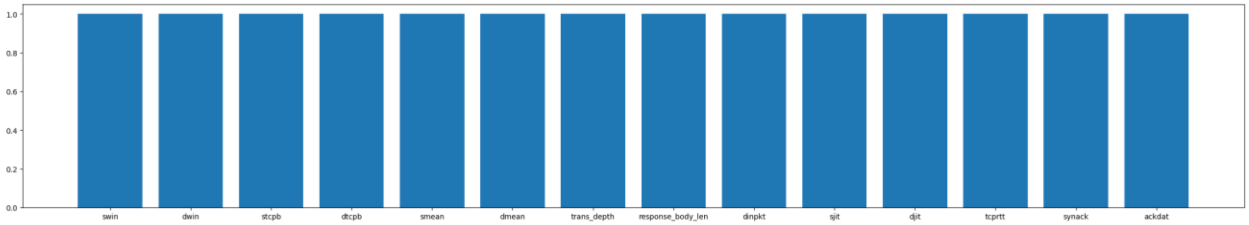
Noah Call, Michael Childs, Justin Roylance

For our project, we referred to a publication dedicated to the research and investigation of utilizing Game Theory in the application of detecting network intrusions by malicious actors. Commonly, intrusion detection is completed using complex machine learning models and neural networks. However, through the usage of Game Theory, it has been proposed that comparative results can be achieved with only a fraction of the cost of resources and time. Thus, we attempted to replicate the results found in the original work to validate their findings and explore our own methods and potential improvements.

Our project consisted of three major sections: Data Preprocessing, Game Theory function implementation, and intrusion simulation/ evaluation. Firstly, for preprocessing, the UNSW_NB15 dataset was retrieved and then preprocessing began. The various techniques for preprocessing directly followed the original paper's techniques. However, the dataset seems to have been updated since it was originally used (for several reasons). First, the dataset we found was already split into Training and Testing sets, unlike in the original paper. Next, there were found to be no NaN values in the dataset—unlike in the original paper where extensive techniques are required to fill in values. With this data integrity validation completed, we proceeded with Feature Coding. As in the original paper, Feature Coding was not performed, and so we also omitted this step. Next, we moved to Feature Scaling and Normalization. The original features appeared as such:



Following a min-max normalization, the features appeared as such:



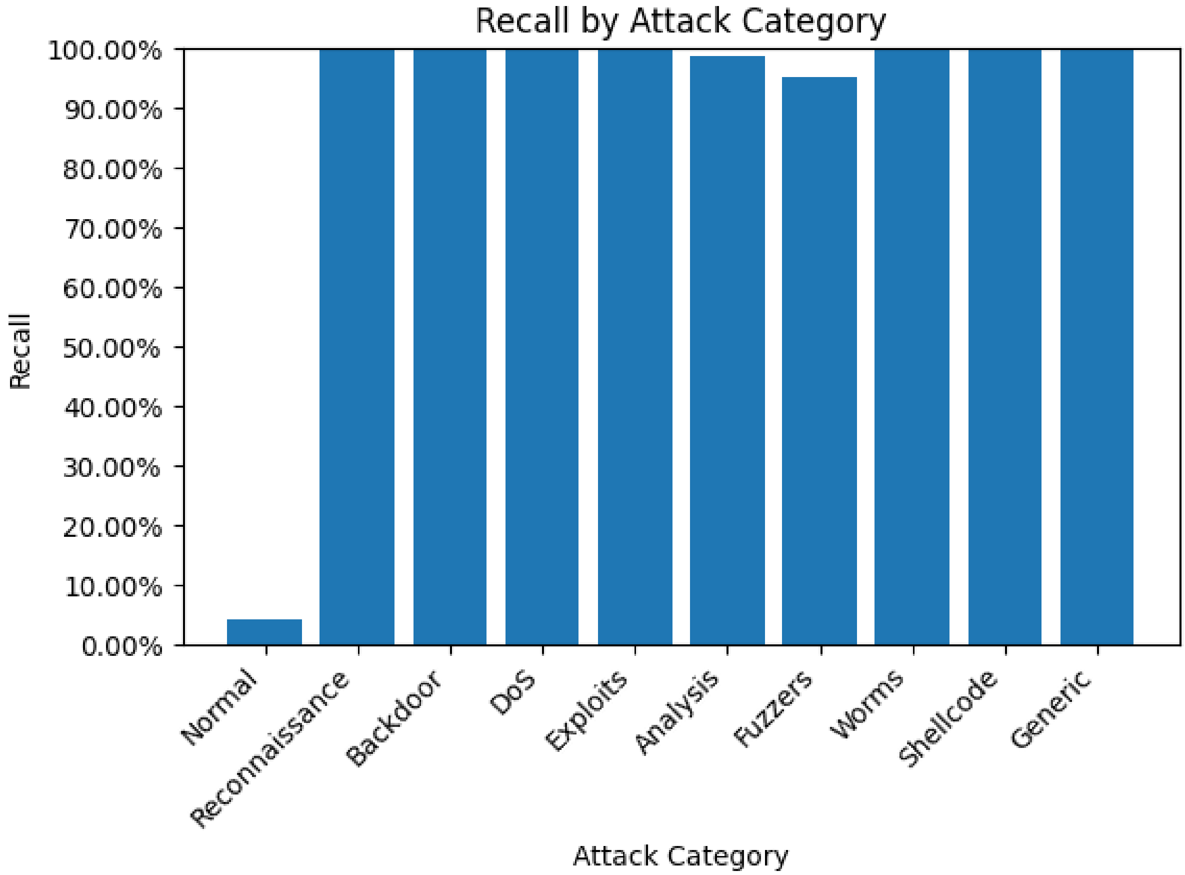
This was an essential step, as otherwise the dataset would have been entirely dominated by the stcpb and dtcpb values. This would have heavily skewed the dataset and the ability to identify attacks. With this step completed, Preprocessing concluded, and the data was output to new CSV files for future use.

Following Preprocessing, we proceeded to implement the Game Theory functionality for calculating Utility and Nash Equilibrium values. For this, we followed the paper's decision to make Player 1 the intruder and Player 2 the intrusion detection system. By using these functions, we were able to identify the best strategies to use in Nash Equilibrium as well as to find the minimum solution strategies for each respective player.

With the Game Theory implementation done, we could finally move on to evaluating Game Theory in the general intrusion detection system. This section proved to be the most difficult and obfuscated section to replicate from the original paper. This is due to the original paper 1.) using a paper which, is unfortunately, not available online (despite being referenced in the paper; see reference 7 of the paper) and 2.) The implementation completed by the original paper was very vague in how it was completed and lacking any code examples, and thus we were left to our own devices to be able to complete a similar model.

Unfortunately, this is the greatest divergence of our project from the original work, as it was nearly impossible to closely replicate due to the heavy complexity of processing the dataset. As such, we were able to provide similar recall results for most categories (thus generally validating the original claim that Game Theory could be effectively used for Intrusion Detection), but we were unable to find an implementation that didn't provide poor recall results for the Normal network traffic. This has proved to be a strong limitation to an otherwise excellent project implementation. However, we believe that with additional time and sources, we would be able to solve this issue.

Recall results are shown below:



This concludes the Network Intrusion Detection using Game Theory project for our group. Overall, we are very satisfied with the results and findings that we've produced, as well as the general validation of using Game Theory in an otherwise machine learning dominated subject.