ABSTRACT

The advancements of network technology pushed everything to its limits, boosting productivity in almost all sector of industry that exist. However, the benefits also comes with risks, with how interconnected the world is, silent attacks againts networks appears causing a breach of sensitive information leading to loss of valuable objects. In order to combat these attacks, a network must be fitted with a plethora of protection, one such protection resides within an intrusion detection system (IDS). An IDS passively protects a network by comparing incoming packets with known signatures of all kinds of attacks, where it would alarm a network engineer to let them know an anomaly has entered the network. The traditional system of large database consisting signature of attacks has proved its success, but with the ever-changing shape of attacks, an additional implementation of artificial intelligence (AI) is needed to help detecting new and undocumented attacks.

AI powered IDS has been around for some time, with studies covering from Machine Learning (ML) methods to Deep Learning (DL) methods. However, more advanced DL models thats capable of handling large complex datasets like convolutional neural network (CNN) is conceptually flawed to handle tabular datasets, which are almost all form of intrusion datasets. Not only that, many studies shruged the idea of generalization across different environments, which not only needed but crucial to an applicable IDS solution.

Enter TabNet, a DL architecture capable to support tabular datasets natively. Using TabNet, this study created a model to generalize across 2 types of datasets, an 'impact' dataset, and a 'lifecycle' dataset. The 'impact' dataset consisted of CIC-IDS-2017 and CIC-IDS-2018, while the 'lifecycle' dataset consisted of CREMEv1 and CREMEv2. With all dataset tested against each other, the results was astonishing, achieving a generally high range of F1-Score performance from 82% up to 99% compared to XGBoost with more variable range from 22% up to 99% across the board for both binary and multiclass generalization problems.

Keywords: Deep Learning, Intrusion Detection System, Cybersecurity, TabNet