**Cyber Security Attack Detection**

**Introduction**

The key role of Cyber security attack detection is to detect the malicious traffic over the network. The Bot-IoT dataset were used for the study of this project. Bot-IoT is consist of 73 million of records with 46 features. Different models of ML and DL were trained using Bot-IoT dataset to predict the cyber security attack detection. Lastly, Generative adversarial Network (GAN) was used for generative samples. The ML and DL trained model were also tested with generative samples to check the robustness of the models.

**Methodology and Results**

The Bot-IoT dataset were downloaded, and top 10 features were selected. The dataset has the category class that is label with normal and malicious. And the category is further divided into five attack type labeled as Normal, Reconnaissance, DDoS, DoS, Information Theft. Firstly, Random Forest model of machine learning was trained for binary classification and got 99% accuracy. The two dense models with 5 hidden layers were also trained for binary and multiclass classification respectively. The ‘tanh’ activation function was used on all convolutional layers. The “sigmoid” and “softmax” was used as activation function on output layer of binary classification model and multiclass classification model respectively. The architecture of dense model is presented in fig 1 and fig 2.

Table

Description automatically generated

Figure 1: Structure of deep learning model for binary classification.

Table

Description automatically generated

Figure 2: Structure of deep learning model for multi class classification.

We got the 99% and 97% accuracy for both models. The confusion matrix of RF and accuracy, loss graph for binary and multi class deep learning model is represent in below figures.

Chart

Description automatically generated

Figure 3: Confusion matrix of random forest model.

Chart

Description automatically generated with low confidence

Figure 4:Accuracy plot of dense model for binary classification.

Chart, line chart

Description automatically generated

Figure 5: Loss plot of dense model for binary classification

Chart, line chart

Description automatically generated

Figure 6: Accuracy plot of dense model for multi class classification.

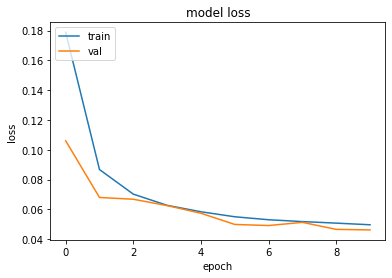


Figure 7: Loss plot of dense model for multi class classification.

Lastly, we develop the GAN model using the architecture of previously develop deep learning model with the loss of 0.214. The 750 samples were generated using the trained GAN model. After performing preprocessing steps, generated samples were predicted by both binary classification models (random forest and dense). The random forest catches the 675 sample as malicious from the generated samples and showed the 94% accuracy. The confusion matrix of RF on generated samples is represent in fig 6. Moreover, dense model showed 97% accuracy.

