DNS Exfiltration Real-Time Detection

Mohamed Amr Mohamed Nagib Alansary
m.alansary@uottawa.ca
uOttawa ID: 300273142

Overview

Exfiltration of data over DNS and maintaining tunneled command and control communications for malware is one of the critical attacks exploited by cyber-attackers against enterprise networks to fetch valuable and sensitive data from their networks since DNS traffic is allowed to pass through firewalls by default, attackers can encode valuable information in DNS queries without fear of being detected.

Solution

In this report, we introduce a real-time mechanism to detect exfiltration and tunneling of data over DNS through training a machine learning model that is capable of detecting anomalies in DNS queries.

Algorithms

We trained three machine learning models and the accuracy of all models was 82% and so we had to choose our champion model based on the confusion matrix results as follows:

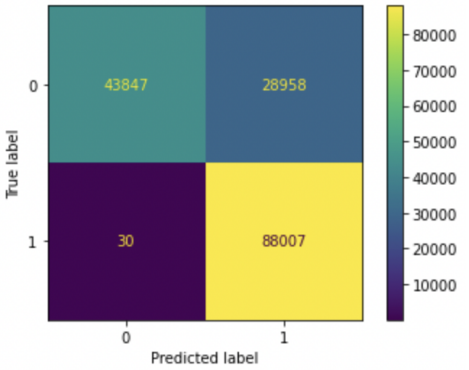
Random Forest

The classification report of Random Forest shows that the model accuracy is 82% and f1-score is 75% and 86% for classes 0 and 1 respectively.

```
Classification report:
              precision    recall  f1-score   support

           0       1.00      0.60      0.75     72805
           1       0.75      1.00      0.86     88037

    accuracy                           0.82    160842
   macro avg       0.88      0.80      0.81    160842
weighted avg       0.86      0.82      0.81    160842
```



The confusion matrix shows a very high number of false positives but low numbers of false negatives which we are interested in more.

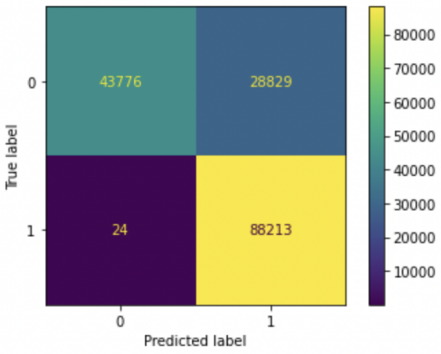
Random Forest is an ensemble learning algorithm that fits a number of decision tree classifiers and uses averaging to improve the accuracy.

https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html

Decision Tree

The classification report of Decision Tree shows that the model accuracy is 82% and f1-score is 75% and 86% for classes 0 and 1 respectively. The

Classification report:	precision	recall	f1-score	support
0	1.00	0.60	0.75	72605
1	0.75	1.00	0.86	88237
accuracy			0.82	160842
macro avg	0.88	0.80	0.81	160842
weighted avg	0.86	0.82	0.81	160842



confusion matrix shows a very high number of false positives but low numbers of false negatives which we are interested in more.

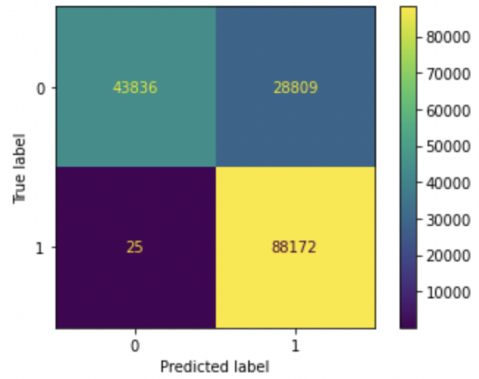
Decision tree is a non-parametric model used for both classification and regression.

https://scikit-learn.org/stable/modules/tree.html

XGBoost

The classification report of XGBoost shows that the model accuracy is 82% and f1-score is 75% and 86% for classes 0 and 1 respectively.

Classification report:	precision	recall	f1-score	support
0	1.00	0.60	0.75	72645
1	0.75	1.00	0.86	88197
accuracy			0.82	160842
macro avg	0.88	0.80	0.81	160842
weighted avg	0.86	0.82	0.81	160842



The confusion matrix shows a very high number of false positives but low numbers of false negatives which we are interested in more.

Gradient Boosting is an additive model in a forward stage-wise fashion; it allows for the optimization of arbitrary differentiable loss functions.

https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.GradientBoostingClassifier.html

Experiments

Model Evaluation

The three models have the same accuracy and f1-score for both classes. A reasonable way to select the champion model is based on the number and cost of false negatives and false positives. Our solution should put more weight on false negatives as the cost of classifying a malicious DNS query as benign is more than the cost of classifying a benign DNS query as malicious and therefore we will select the model that has reasonable false negatives and false positives. Based on this criteria our champion model is XGBoost.

Hyperparameter Tuning

We searched for the best hyperparameters using randomized search and obtained a little higher accuracy with the new parameters but also false negatives increased to 53 instead of 25 and so we used the default hyperparameter values instead.

```
 Best hyperparameters:
{'subsample': 1.0, 'min_child_weight': 5, 'max_depth': 5, 'gamma': 5, 'colsample_bytree': 1.0}
```

https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.RandomizedSearchCV.html

Results

The highest accuracy we could get is 82%, we can increase this accuracy using stateful features in addition to our stateless features. As we can see DNS exfiltration detection is achievable easily using machine learning algorithms having the advantage of real-time fast detection but as everything comes with a cost, we “must” tune our model regularly to keep its accuracy and evaluation metrics high and to enhance the model further.