## Security Lab

## Real Time Filtering of Malicious URLs

Ishwarya S - 2017103537

Gokul K - 2017103526

Harshini T - 2017103532

**Abstract:**

The rapid development of communication technologies and broadband speeds has greatly democratized the internet. The number of internet users has grown exponentially in the past few years. This has caused a sharp increase in the global internet traffic. This has also opened up several avenues for online fraud. There are several malicious URLs on the internet that are used to scam and cheat people. In many cases privacy of the person is violated and this can also lead to money thefts.

Therefore it is of great importance to filter the malicious URLs on the internet. But this process is a bottleneck because of the high internet traffic. Hence it is essential to develop a lightweight filtering system to filter out the harmful URLs.

In this project we aim to develop a robust and scalable classifier that can perform URL filtering in real-time using lightweight features so that this reduces the processing pressure of the back-end malicious URL detection systems based on content analysis.

We will be exploring 3 different algorithms for this classifier and provide a comparative study on the performance of the algorithms using Precision, Recall and Accuracy metrics.

The different URL features like URL Length, number of special characters in the URL, origin server, origin country, charset, content length, registration date of the website, last modified date have been used for the classification process. This will be performed as a binary classification problem.

**Dataset:**

The Malicious and Benign websites dataset from Kaggle will be used for this project.

**Machine Learning Algorithms:**

Three Machine Learning Algorithms have been used for this project

1. Multi-Layer Perceptron
2. Random Forest Classifier
3. Support Vector Machine

**Language and Frameworks:**

The project will be built using the Python programming language. Frameworks like Scikit-Learn, Tensorflow, Keras, Numpy and Pandas will be used for this project.