Malicious URL detection based on lexical traits using machine learning techniques.

Team members:

1. 123015128 Repala Srinivas Yaswanth.
2. 123003283 Annam Venkata Sai.
3. 123003284 Chevuru Venkata Rajesh.

Name of the Guide: Dr. Priya S

Abstract:

Cyber criminals' most advanced cyber-attack approach is to create and propagate malicious domain names or malicious URLs via email, messaging, popups, and other means. The main goal of the attack is to steal the victim's personal information, user credentials, or install malware on their computer. Researchers propose a variety of strategies, but machine learning-based detection outperforms them all. This study proposes a light-weighted technique that simply considers the URL's lexical properties. As a result, the Random Forest classifier performs better than the other classifiers in terms of accuracy.

The proposed method consists of four phases i.e., Feature Extraction, Feature reduction, Train the model and Testing. For machine learning algorithms, feature extraction is the first step. Not every feature in an extract is acceptable for classification. As a result, a feature reduction approach is required to determine the fitness of the features. Features are removed if there is no significant correction is present. Then, using the proper data set, train the model and test it with various URLs. Thus, malicious URL's can be identified.

Keywords**:** *Malicious URL Detection*, *Machine learning*, *Feature extraction*, *Feature reduction.*

Base paper details:

Lexical features based malicious URL detection using machine learning techniques, Saleem Raja A, Vinodini R, Kavitha A, Materials Today: proceedings, 2021.

Signature of the Guide