WEB PHISHING DETECTION USING MACHINE LEARNING

Problem statement:

Phishing is a non-ethical method comprising both social engineering and technical tricks to capture user’s information and sensitive credentials like financial credentials. Some of the social engineering techniques use spam mails, pretending as a legitimate company or organization, that are specially designed to forefront users to knock-off websites that manoeuvre recipients to fall into the trap which steal financial credentials like user-ids and passwords. Technical intrigue methods install malicious software onto the systems, to capture the data directly, often using systems to intercept users online account user names and passwords. There are multiple websites who ask user to provide sensitive data such as username, password or credit card details etc. often for malicious reasons. Over the years Internet has dominated the world by dragging half of the world’s population exponentially into the cyber world. With the booming of internet transactions, cybercrimes rapidly increased and with anonymity presented by the internet, Hackers attempt to trap the end-users through various forms such as phishing, SQL injection, malware, man-in-the-middle, domain name system tunnelling, ransomware, web trojan, and so on. Among all these attacks, phishing reports to be the most deceiving attack.

To overcome the drawbacks of blacklist and heuristics based method, many security researchers now focused on machine learning techniques. Machine learning technology consists of a many algorithms which requires past data to make a decision or prediction on future data. Using this technique, algorithm will analyze various blacklisted and legitimateURLs and their features to accurately detect the phishing websites including zero- hour phishing websites.