**Phish Catcher: Client-Side Defence Against Web Spoofing Attacks Using Machine Learning**

**Abstract**

Cyber security confronts a tremendous challenge of maintaining the confidentiality and integrity of user’s private information such as password and PIN code. Billions of users are exposed daily to fake login pages requesting secret information. There are many ways to trick a user to visit a web page such as, phishing mails, tempting advertisements, click-jacking, malware, SQL injection, session hijacking, man-in-the-middle, denial of service and cross-site scripting attacks. Web spoofing or phishing is an electronic trick in which the attacker constructs a malicious copy of a legitimate web page and request users’ private information such as password. To counter such exploits, researchers have proposed several security strategies but they face latency and accuracy issues. To overcome such issues, we propose and develop client-side defence mechanism based on machine learning techniques to detect spoofed web pages and protect users from phishing attacks. As a proof of concept, a Google Chrome extension dubbed as Phish Catcher, is developed that implements our machine learning algorithm that classifies a URL as suspicious or trustful. The algorithm takes four different types of web features as input and then random forest classifier decides whether a login web page is spoofed or not. To assess the accuracy and precision of the extension, multiple experiments were carried on real web applications. The experimental results show remarkable accuracy of 98.5% and precision as 98.5% from the trials performed on 400 classified phished and 400 legitimate URLs. Furthermore, to measure the latency of our tool, we performed experiments over forty phished URLs. The average recorded response time of Phish Catcher was just 62.5 milliseconds.

**Existing system**

Now-a-days we are heavily dependent on online data such as Online news, Email Messages, Online Reviews, Online Post and many more. This online content access open doors for attackers to allure normal users by sending enticing messages of jackpot wining with fake phishing URL or spoofing websites. Whenever user click on such URL or navigate to spoofing website then they will ask user to enter login details and then attackers will use those login details to gain access to banking or any other financial websites and grab or steal all user money or any other secret information.

In existing system, machine learning and signatures-based algorithms were introduced but their detection rate are not accurate

**Disadvantages:**

1. Less Accuracy
2. More time taking process

**Proposed system**

In proposed system, employing Random Forest algorithm to detect phishing URLS. Random Forest algorithm has inbuilt support for features optimizations and selection which help in enhancing prediction accuracy. Random forest will apply group of trees on dataset to filter and remove irrelevant data and then select only optimized features.

**Advantages:**

1. High Accuracy
2. Takes less time

**HARDWARE & SOFTWARE REQUIREMENTS:**

**HARDWARE REQUIRMENTS:**

* processor :   intel i3(min)
* Hard Disk  :   500 GB.
* RAM :   4GB.

**SOFTWARE REQUIRMENTS:**

* Operating system : Windows 10 (min)
* Coding Language  : python