BATCH NUMBER: B10-4A6E

PROJECT TITLE: Web Phishing Detection

TEAM MEMBERS:


1. Ilakiya B
2. Mohitavarshini A S
3. Nithilasri T
4. Yuvaharini A

# LITERATURE SURVEY

## PAPER 1

**TITLE:** ANTI-PHISHING BROWSER EXTENSIONS

**YEAR:** 2015

**AUTHORS:** Oinam Bhopen Singh and Dr. Hitesh Tahbildar

## DESCRIPTION:

With the advancement of the Internet and the World Wide Web, our world has transformed into a global village, allowing us to interact, educate, and buy and sell goods and services with the click of a mouse. However, we have seen an increase in illegal activities such as phishing, hacking, and so on, so cyber crime is now recognised as a major international problem, with phishing in particular becoming a serious issue for all organisations dealing with online e-commerce.

Phishing attacks on the World Wide Web have increased significantly over the last decade. Phishing is a criminal mechanism that uses both social engineering and technology to steal consumers' personal identity information and financial account credentials. Spoofed e-mails purporting to be from legitimate business organisations and agencies are used in social engineering to direct users to fake websites that trick recipients into disclosing personal information such as usernames and passwords. The technical element entails creating websites that appear to be legitimate websites of banks, financial institutions, and so on, by corrupting local navigational infrastructures and redirecting users to fake websites.

It also instals malware on PCs in order to directly steal users' personal information by setting up systems to intercept users' online account names and passwords. According to the Anti-Phishing Working Group's (APWG) June 2014 phishing trend report, the number of phishing websites increased by 10.7% over the fourth quarter of 2013, with 125,215 unique phishing websites. Payment services remain the most targeted industry sector for phishing. Every year, hundreds of millions of dollars are lost due to unsuspecting users entering personal information into Fake web sites. To combat these threats, software vendors and companies with a vested interest in phishing attack prevention have leased a variety of anti-phishing toolbars. In this paper, we conducted a literature review on anti-phishing browser extensions. The benefits and drawbacks of each browser extension, as well as the methods used, are listed. Finally, we compare the various browser extensions based on the technique used, cost, user friendliness, and security algorithm.

## PAPER 2

**TITLE:** Detecting Phishing Website Using Machine Learning

**YEAR:** 2020

**AUTHORS:** Abdul Razaque, Fathi Amsaad , Mohamed Ben Haj Frej

## DESCRIPTION:

Phishing is defined as impersonating a legitimate site in order to defraud users by stealing their personal information such as usernames, passwords, account numbers, national insurance numbers, and so on. Phishing scams may be the most common type of cybercrime today. There are numerous domains where phishing attacks can occur, such as the online payment sector, webmail and financial institutions, file hosting or cloud storage, and many others. Phishing was more prevalent in the webmail and online payment sectors than in any other industry sector. Phishing can be accomplished through email phishing scams and spear phishing; therefore, users should be aware of the risks and should not place their complete trust in common security applications.

Machine Learning is one of the most effective techniques for detecting phishing because it eliminates the shortcomings of existing approaches. The most important goal of the proposed project is to validate the website's validity by capturing blacklisted URLs. To notify the user on a blacklisted website via pop-up while they are attempting to access, as well as to notify the user via email while they are attempting to access. This proposed project will enable administrators to add blacklisted URLs to alert users during their investigation.

The two project scopes are well known as user scope and system scope. The user bears some responsibility for the system. The system includes a few standards and policies that must be followed in order for the system to function properly. If a blacklisted website is accessed, the user can be notified. The administrator can capture the blacklisted URLs and alert the user.

## PAPER 3

**TITLE:** Phishing Web Site

**YEAR:** 2019

**AUTHORS:** Pratiksha Yewale, Prajkata Jadhav, Prajkta Zende, Dhanashree Nikalje

## DESCRIPTION:

The Anti-Phishing Working Group (APWG) defines phishing as a criminal scheme that uses both social engineering and technical deception to steal consumers' personal identification data, including financial account credentials. Typically, phishers deceive users by sending spoofed e-mails that appear to be from a reliable source, such as a bank or a reputable commerce agency. Malware-based phishing and deceptive phishing are the two main types of phishing attacks. Malware-based phishing methods exploit security flaws in the user's system to install malicious software. This software then records sensitive and confidential data and sends it to the phisher. In deceptive phishing, an attacker sends out emails that appear to be from trusted sources.

• These e-mails invite recipients to click on a link that takes them to a bogus website designed to trick recipients into disclosing sensitive information. In this type of attack, the phisher employs several techniques to deceive the user, and Berthold et al. Social engineering refers to all methods and scenarios devised by phishers in order to create a convincing context.

• Imitation, which consists in creating websites that appear to be legitimate.

• E-mail spoofing, which allows a phisher to spoof an e-source mail's address. URL concealment, which allows phishers to conceal the URL to which a user is redirected

## PAPER 4

**TITLE:** Phishing Detection using Map-reduce and PART Algorithm

**YEAR:** 2016

**AUTHORS:** Mr. Rakshith Raj K.R., Prof. Prabhakara B.K

## DESCRIPTION:

Phishers create bogus web pages in order to steal personal information from users without their knowledge. He will unknowingly provide his information to the phishers. According to the Anti-Phishing Working Group [15], there were nearly 40,000 phishing attacks between January 2014 and March 31, 2014. Almost 95% of users are unaware of phishing and how to avoid becoming a victim. There are no anti-phishing techniques that can completely stop phishing activity. In recent years, there has been an increase in phishing activity. According to recent statistics, phishing is a major threat to personnel data and financial loss as a result of unknowingly becoming a victim of phishing. Hadoop evolved from Google's MapReduce, a software framework that divides a large set of data into numerous small parts. Map converts one set of data into another, where individual elements are broken down into key/value pairs. Second, there is the reduce task, which takes the output of a map as input and combines those data tuples into a smaller set of Key/Value pairs or tuples. These tuples are executable on any node in the cluster. The Hadoop Distributed File System (HDFS) is intended to store very large data sets reliably and to stream those data sets to user applications at high bandwidth. Thousands of hosts in a large cluster are directly connected to storage and run user application tasks.

The system is faster in detecting phished webpages in this proposed approach than in existing anti-phishing techniques. In the proposed system, MapReduce is used to collect URL information from various nodes in a distributed environment, which is faster. It is also an open-source software programme for storing and processing large datasets on community hardware clusters. Based on the data provided by Hadoop MapReduce, the PART algorithm will predict the authenticity of the URL.