Literature survey

**WEB PHISHING DETECTION**

**Using Machine Learning**

**TEAM MEMBERS:**

* **PRIYADHARSHINI**
* **NOOR UL HUDAH**
* **SYED KARISHMA AFREEN**
* **MARYAM AYSHA**

**LITERATURE SURVEY ON WEB PHISHING DETECTION**

**ABSTRACT:**

Trying to gather personal information through deceptive ways is becoming more common nowadays. In order to assist the user to be aware of the access to such websites, the implemented system notifies the user through email and pop-up, when trying to access a phishing site. This paper proposes an approach of phishing detection system to detect blacklisted URL also known as phishing websites, so that individual can be alerted while browsing or accessing a particular website. Therefore, it can be utilized for identification and authentication and become a legitimate tool to prevent an individual from getting tricked.

**INTRODUCTION:**

Phishing can be defined as impersonating a valid site to trick users by stealing their personal data comprising usernames, passwords, accounts numbers, national insurance numbers, etc. Phishing frauds might be the most widespread cybercrime used today. There are countless domains where phishing attack can occur like online payment sector, webmail, and financial institution, file hosting or cloud storage and many others.

The objective which is the most vital thing in project is to verify the validity of the website by capturing the black listed URLs. To notify the user on blacklisted website through pop-up while they are trying to notify the user on blacklisted website through email while they are trying to access. This project will allow administrator to add blacklisted URL’s in order to alert user during inquiry.

**RELATED WORK:**

Phishing is a type of practice done on the Internet where individual data are obtained by illegal approaches. It supply of obtaining sensitive information, as an example, usernames, proposed project is to verity the validity of the website by passwords, and positive identification points of interest, often malignant reasons, by taking up the looks of an electronic correspondence. Phishing attack will be enforced in varied kind access and to notify the user on blacklisted website through like Email phishing, web site phishing, spear phishing Whaling, Tab off his guard, Evil twin phishing etc. Phishing is known as webpage violence. Phishing is often done by email spoofing or texting, and it typically guides user to enter points of interest at a fake web site which look and feel the same. It tries to handle the increasing range of phishing got to be met by clients in awareness and alternative efforts to ascertain protection numerous anti-phishing tools. A number of sites have currently created optional instruments for application. The primary key feature is to allow user to inquire whether visited websites is original or fake .This project proposes a security tool called web phishing detection using machine learning.

**LITERATURE REVIEW:**

The current situation that is majority of the population has been fooled into giving their personal details to hackers without noticing it. Many blacklisted website has been publish to appear as an original site in order to trap user by asking them to input their personal details. For example, password, bank account, email address and etc. Phishing activity in early 2016 was the highest ever recorded since it began monitoring in 2004.The total number of phishing attacks in 2016 was 1,220,523. This was a 65percent increase over 2015. In the fourth quarter of 2004, there were 1,609 phishing attacks per month. In the fourth quarter of 2016, there was an average of 92,564 phishing attacks per month, an increase of 5,753% over twelve years. According to the Anti-Phishing Working Group (APWG), there are at least 47, 324 phishing attacks and a top-ten American bank estimates that at least US$300 is lost for every hour that a phishing site remains up. Machine learning is that the science of obtaining computers to act while not being expressly programmed. Machine Learning was implement to develop this proposed system. Machine learning techniques identifies phishing URLs typically assess a URL based on some feature or set of features extracted from it. Thus, before coming to conclusion that this was the major problem, related products were examined and compared view their libation before progressing to the proposed project.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **TITLE** | YEAR | TECHNIQUE | FINDINGS | PROS AND CONS |
| PHISHTANK | 2006 | Phishtank was proposed to carry out the inspection once a link has been pasted on the section given. This allow user to keep on track of faked website. They can copy and paste the link in order to identify whether the site that they are going to access is safe or not safe | User can use the website search feature directly or they can use information from PhishTank through its API. A search engine displayed on PhishTank website is to be used as the first method. Using its API will be the second method. API service can be avail by software builder after registering themselves on PhishTank website | Pros : The purpose of API’s usage is for user who has basis information on software development.  Cons: There was no facility of displaying pop-up and email notification once user had access blacklisted website |
| PHISHZOO | 2009 | PhishZoo was proposed to evaluate a new method for web phishing detection based on profiles of complex sites’ appearance and content. | PhishZoo makes profiles of sites comprising of the website contents and images displayed. These profiles are kept in a local folder and are either synchronized against the newly loaded sites at the time of loading or against risky sites for instance, links in email offline | Pros: PhishZoo makes profiles of sites comprising of the website contents and images displayed.  Cons: There was no facility of displaying pop-up and email notification once user had access blacklisted website |
| GOLDPHISH | 2017 | GoldPhish was proposed to perceive and report phishing sites. This was done by using optical character recognition (OCR) to recite the text from an image of the page precisely from the company logo, grasping the top hierarchical areas from a search engine, and comparing them with the current web site | The forte of the tool lies in the user’s capability to recognize famous company logos. A phishing site cannot change a familiar company logo without the phishing target perceiving. | Pros: Being able to do phishing stimulations at any given time.  Cons: There was no facility of displaying pop-up and email notification once user had access blacklisted website |