**SURVEY ON PAPERS RELATED TO WEB PHISING DETECTION**

**NAME:** Detection of E-mail phising attacks using Machine Learning

**AUTHOR**: Dhruv Rathee, Sumanmann.

**PUBLISHED YEAR:** 2021

## BASIC DESCRIPTION:

## Phishing is the most prominent cyber-crime that uses camouflaged e-mail as a weapon. It is defined as the strategy adopted by fraudsters in-order-to get private details from persons by professing to be from well- known channels like offices, bank, or a government organization. In this era of modernization, electronic mails are accustomed globally as communiqué channel for both private and professional purposes. The particulars exchanged over e- mails are often confidential and sensitive for example info of bank statements, payment bills, debit-credit reports, and authentication data. This makes e-mails precious for hackers because they can exploit these details for maleficent intends. The main goal of the attackers is to acquire personal details by deceiving the e-mail recipient to click noxious link or download the attachment under false pretences.

## HIGHLIGHTS OF THE PROJECT:

From this project report we understood that the following results,

1. Planning—this involves identifying the targets, the information sought, and creating/identifying the tools and techniques that will be used in the attack (such as emails with malicious links and the spoof sites these links direct to).

2. Phishing—the stage during which the identified targets are phished using the resources created in Stage 1.

3. Infiltration—depending on the method used, this stage will vary but it essentially consists of the response from the target and gaining access to the personal information sought.

4. Data collection and exploitation—this is the stage at which the phisher extracts the information sought and utilizes it to achieve the ends established during the planning phase. This often involves fraud whereby the attackers impersonate the victims to access their accounts, etc. Another common occurrence is the selling of this personal data on the online black market.

5. Exfiltration—finally, the phisher attempts to remove as much evidence of their attempt as possible (such as the deletion of fake sites). There may also be some analysis on the success of the attack and assessment of future attacks.

**Overview of the project:**

The person who wants to escape from the fake websites will be informed about the phising websites to the user immediately. Phishing e-mails have transformed and modified into numerous categories, some are well-crafted to appear legitimate **.** The phishing e-mails are sent from authorized accounts, so that the mails seems to appear legit. To accomplish their mission, attackers often keep track of the user personal details. They always send mails from only those accounts that belong to a friend or a previous business colleague by tricking the victim with fake e-mail IDs. The criminals often work very hard to make e-mails appear genuine by entailing believable-wordings, graphic-interface, and logos. Although, phished e-mails do not appear similar as actual phishing emails are custom-built for their anticipatedobjectives.

Machine learning-based approaches help in detecting phishing attacks more efficiently by giving lower false-positive rates and high accuracy in comparison with other methods.

Nine features were extracted from the e-mail and the last few features were obtained from the WHOIS query. They used larger datasets of about 7000 normal e-mails and 860 phishing e-mails for training and testing the classifier. They focused specifically on URL properties which might not be the appropriate technique because identification of phished emails depends on various factors. Also, criminals could use tools to obfuscate URLs such as tiny URLs (https://tiny.qe/) and design them to appear legal. Their filter scores 97.6% measure, false-positive rate of 0.13%, and a false-negative rate of 3.6%.

## RESULT:

From this survey a number of anti-phishing software techniques. Some of the important aspects in measuring phish-ing solutions are:

• Detection accuracy with regards to zero-hour phishing attacks. This is due to the fact that phishing websites are mostly short-lived and detection at hour zero is critical.

• Low false positives. A system with high false positives might cause more harm than good. Moreover, end-users will get into the habit of ignoring security warnings if the

classifier is often mistaken.

Generally, software detection solutions are:

• Blacklists.

• Rule-based heuristics.

• Visual similarity.

• Machine Learning-based classifiers.

**CONCLUSION:**

The paper presents a survey analysis of actual phishing email identification works from various perspectives. This survey is unique in the sense that it relates works to their openly available tools and resources. Many ML methods have been adopted to identify phishing emails, but these cannot effectively detect new phishing scams, which needs significant manual feature engineering. Anti-phishing technology developed on the source code features is quite slow in terms of the classification of phishing emails given its dependence on third-party services and scraping of the email content. The analysis of the presented works revealed that not much work had been performed on phishing email detection using natural level Natural Language Processing (NLP) techniques. Therefore, many open issues are associated with this phishing email detection.