Phishing Attack Prevention: How to Identify and Detect Phishing Attacks

Literature Review

by

**Suraj Pratap Samirlal Poddar**

Harrisburg University of Science and Technology

Spring 2021

Literature Review

[Introduction 3](#_Toc65175113)

[Problems and Challenges 3](#_Toc65175114)

[Prevent Phishing 4](#_Toc65175115)

[Detect Phishing 4](#_Toc65175116)

[Providing Stake holder training 5](#_Toc65175117)

[Different Types of Phishing Attacks 5](#_Toc65175118)

[Malware Phishing 5](#_Toc65175119)

[Pharming 5](#_Toc65175120)

[Vishing 5](#_Toc65175121)

[Deceptive Phishing 6](#_Toc65175122)

[Spear Phishing 6](#_Toc65175123)

[Content Injection Phishing 6](#_Toc65175124)

[Host File Poisoning 6](#_Toc65175125)

[Learning to Detect Phishing Emails 6](#_Toc65175126)

[Hybrid Learning based Phishing Detection 8](#_Toc65175127)

[CRI approach to address phishing 8](#_Toc65175128)

[Taxonomy of phishing Attack 9](#_Toc65175129)

[Anti-phishing Technique Modus Operandi 9](#_Toc65175130)

[Lifecycle of Phishing Attack 9](#_Toc65175131)

[Visual Similarity Based Phishing Detection and Filtering Approach 10](#_Toc65175132)

[Taxonomy of Phishing Detection and Filtering Based on Visual Similarity 11](#_Toc65175133)

[Visual Similarity Assessment 11](#_Toc65175134)

[References 12](#_Toc65175135)

# Introduction

Over the years, given the fast adoption of latest technological advancements, security exploitation/attacks have significantly on a rise topping the human vulnerability. Phishing has become one of the threats to cyber world creating a massive damage of worth billion dollars every year. Phishing can be performed in different methods involving web, email, messages etc. Over time there has many articles describing the techniques and new attacks based on phishing however, they have failed to outline all the risk factors involved and provide and efficient/effective solution to mitigate this form of a threat. In this section we will be comparing different data from various sources to perform a literature review. Emails has been always one of the main online application which is used extensively by lots of users, business, Govt and different organization for communicating with one another and sharing data and therefore phishing emails has always shown us a serious threat to digital commerce as they are used to scam both individual and financial institutions

# Problems and Challenges

Hacker/phishers have become advanced and more talented/skilled at recreating/forging sites to appear identical including recreating logos and graphical presentation. No doubt Phishing is evolving to tackle the defenses and bypass the detection. In these circumstances I think a sequential approach is needed which will help mitigate these kind of attacks

1. Prevent Phishing
2. Detect Phishing
3. Provide Stake holder training

Diagram

Description automatically generated

Figure 1 Step by Step process

### Prevent Phishing

Phishing can be prevented by taking extra precautions and blocking it before it reaches the users, this includes blocking or blacklisting the phishing site by filtering emails. The first procedure is to identify the URL either manually or by ML, although this can help catch few sites, it cannot catch them all

### Detect Phishing

This method is deemed to be one of the effective methods as this can identify the link to phishing sites through the spam filters which are majorly used by the email servers. Phishing filters are constructed by Machine Learning techniques. Most of the browsers already have an extension installed on their browser which has a passive or an active indicator warning the users and alerting them at the same time

### Providing Stake holder training

By providing training to all the users can help avoiding the users clicking random sites and falling into Phishing scams. Most of the organizations create internal training program to combat against advanced phishing attacks which is often referred as anti-phishing techniques/methods. One of the famous anti-phishing game is Anti-Phishing Phil which is a micro game that helps a user identify suspicious URLs and spread awareness also provide a detail view of the consequences

# Different Types of Phishing Attacks

### Malware Phishing

A Malware which runs on the user’s machine, these are populated either via wicked softwared or by sending malicious files to the user via email as an attachment which upon downloaded can record the keystrokes

### Pharming

Pharming is a technique which is used to install a malicious code on users machine to manipulate and direct them to a fraudulent website without the users knowledge

### Vishing

Vishing is a technique which happens through a voice communication system in this case by making phone calls or by leaving a voicemail to the user acting to be a real company to lure a user to give away their personal identifiable data such as their back A/c Details, SSN, or Credit card details

### Deceptive Phishing

Deceptive Phishing is known as a very common type of phishing scam. In this technique a scammer can fake invoices or any attachments in deceiving users into opening an email or a text which is related to user’s day to day activity. For example, users W2 form or users tax statement. Malicious actors/scammers can send you a text/email with a company name which users trust and respond to

### Spear Phishing

Spear Phishing is a technique where email is sent to a well-researched target which claiming to be a trusted user, the entire intent is to lure the user to give away their personal information or infect devices with malware.

### Content Injection Phishing

Content injection phishing is a type of phishing where a illicit actor creates the contents of a legitimate sites with fake content and makes it look like a trusted site by luring the victim to give away their personal/confidential information to a hacker

### Host File Poisoning

Host File Poisoning converts a FQDN/URL into an IP address before it is routed over the Internet

# Learning to Detect Phishing Emails

Even though there are few advantages to filter a phishing attack delivered in the form of an email , there are not that many methods tailored to concentrate on phishing emails as opposed to spams in general. The closely related option is to create policies on the smtp server based on the classification of what we call it as a good email vs a phishing email. As the attacks improve, we expect there will be new features that will be added which will help in identifying and combining the information form internal as well as external resources. This approach is used to detect phishing websites, or the content sent via email to lure a user. While most of the email gets filtered it is very hard to accurately differentiate, although 90% of the phishing emails are blocked there is still a 10% chance where the user can always fall into the trap

The prosaic approaches for phishing detections have always been at a lowest accuracy, machine learning provides us a better outcome for phishing detection, however, are time consuming and not a match at scale. There are various methods that can help us in Phishing detection

Machine Learning Based Phishing Detection

Diagram

Description automatically generated

Figure 2 Machine Learning

To tame ML and make it effective in detecting phishing attacks is to collect data at hand and they must have features related to legit websites and phishing websites. Figure 2 Machine Learning shows the working model of ML for phishing attack detection. A stream of data is provided as input to train ML model to help predict the phishing attacks and legit traffic.

Author in El Asssal et al. introduced a conventional structure which is named after phish Bench, which enables us to not only analyze but also assess the current detection techniques by understanding the framework and performance measurements. The tests show us that the classification dropped when the portion dropped among the authentic and phishing decreasing from 1 to 10

### Hybrid Learning based Phishing Detection

The technique compares real and phishing with counter strategies like tokenization, word/text parsing and by stopping word evacuation, this type of approach has been found to be accurate against existing techniques

### CRI approach to address phishing

This type of approach focuses on a solution resulting in a holistic anti phishing review. All the approaches primarily focus on how to prevent the attack without actually thinking about the crime and the respective problem associated with it, the following literature shows that providing solution without knowing/exploring the actual issue is not the right way to tackle the threat. As a reason to which a CRI approach was implemented to explore the crime factor, review mitigation techniques and investigate the missing pieces. The major goal is to provide a future researcher with the information at the same time strengthen the literature review. Crime has increased with the expansion of IT, IoT and any form of digital devices that can talk over the Internet where cyber criminals have got their hands on to exploit in the real world

### Taxonomy of phishing Attack

Attackers use different techniques for phishing. In social way they try to attract by sending emails to provide bank information, credit card and convenience them to respond. They might also include links to open which can directly hack your system and attackers can also remote to victim’s computer to collect all the data.

### Anti-phishing Technique Modus Operandi

After receiving phishing emails anti-phishing techniques are also implemented either by redirecting them to SPAM folder or by showing a warning when user tries to open the link.

### Lifecycle of Phishing Attack

* Attackers try to create a fake copy of organization and sends URL to all Internet users using email id & social networking sites.
* In case of fake emails there are different solutions to block them based on structure features of mails so that fake email ids are not blocked.
* If fake email id works and user tries to open link, there will be some browser techniques based on client end.
* If it by passes through all the blockages, then attacker can steal all the data from user.

### Visual Similarity Based Phishing Detection and Filtering Approach

A user will not understand phishing attack as it will look similar to legitimate sites by seeing high standard URL’s layout, Images, fonts. Fake and original sites look exactly same except the URL’s are not same, but user may not check URL’s before clicking the link. If attacker doesn’t match with the original website, then there are chances of users using it are very less.

An attacker can fool user by following

1. Visual Appreance: Phishing website will exactly look like original website as attackers will copy the HTML source code to look same.
2. Address Bar: They can also cover the address bar by using any image so that users will not understand.
3. Embedded Objects: Use of different images, text to hide the content from phishing detection techniques.
4. Favicon: Every website will have a unique favicon image and attacker will try to copy same image if not user can easily identify fake website.

Based on different surveys 90% of users couldn’t identify fake websites by seeing the visual appearance and content. Even experienced users cannot understand and most of them doesn’t even look at the address bar because of this most of the users are trapped by attackers.

### Taxonomy of Phishing Detection and Filtering Based on Visual Similarity

Different approaches have been proposed to detect phishing attack based on user’s education and software. An educated user should always check address URL before clicking any link and software-based approach like black-list will have all phishing URL’s but fails as it’s difficult to update new websites that are created every day.

### Visual Similarity Assessment

Based on approach there are two modules in which first module will detect the suspicious URL’s and other module will detect the layout, content. In Block level each block is compared with original website and then it shows matching blocks. Layout level is dividing the total number of matching blocks to all blocks in original site. Overall style is calculated by style feature and similarity if threshold is more than legitimate website then it is considered as phishing webpage

# References

A. N. Shaikh, A. M. Shabut and M. A. Hossain, "A literature review on phishing crime, prevention review and investigation of gaps," *2016 10th International Conference on Software, Knowledge, Information Management & Applications (SKIMA)*, Chengdu, 2016, pp. 9-15, doi: 10.1109/SKIMA.2016.7916190.

Khonji, Mahmoud & Iraqi, Youssef & Jones, Andy. (2013). Phishing Detection: A Literature Survey. IEEE Communications Surveys &amp Tutorials. PP. 1-31. 10.1109/SURV.2013.032213.00009.

Purkait, S. (2012), "Phishing counter measures and their effectiveness – literature review", *Information Management & Computer Security*, Vol. 20 No. 5, pp. 382-420. https://doi.org/10.1108/09685221211286548

Basit, A., Zafar, M., Liu, X. *et al.* A comprehensive survey of AI-enabled phishing attacks detection techniques. *Telecommun Syst* **76,** 139–154 (2021). https://doi.org/10.1007/s11235-020-00733-2

El Aassal, A., Baki, S., Das, A., & Verma, R. M. (2020). An in-depth benchmarking and evaluation of phishing detection research for security needs. *IEEE Access*, *8*, 22170–22192.

Vayansky, Ike & Kumar, Sathish. (2018). Phishing – challenges and solutions. Computer Fraud & Security. 2018. 15-20. 10.1016/S1361-3723(18)30007-1

M. Khonji, Y. Iraqi, and A. Jones, “Phishing detection: a literature survey,” *IEEE Communications Surveys & Tutorials*, vol. 15, no. 4, pp. 2091–2121, 2013.

R. Islam and J. Abawajy, “A multi-tier phishing detection and filtering approach,” *Journal of Network and Computer Applications*, vol. 36, no. 1, pp. 324–335, 2013.

A. K. Jain and B. B. Gupta, “Comparative analysis of features-based machine learning approaches for phishing detection,” in *Proceedings of the 10th INDIA-COM*, New Delhi, India, 2016.

M. Chandrasekaran, et al., ”Phishing email detection based on structural properties”, in New York State Cyber Security Conference (NYS) , Albany, NY ,” 2006