Phishing Attack Prevention: How to Identify and Detect Phishing Attacks

Methods

by

**Suraj Pratap Samirlal Poddar**

Harrisburg University of Science and Technology

Spring 2021

Table of Contents

[Phishing Methods and Techniques 3](#_Toc68817124)

[Phishing Attacks and Defenses 11](#_Toc68817125)

[Related Work 14](#_Toc68817126)

[Methodology and Approach 15](#_Toc68817127)

[Email Image only phishing 17](#_Toc68817128)

[Email Field Manipulation 17](#_Toc68817129)

[Email Redirection 20](#_Toc68817130)

[Pop up window 21](#_Toc68817131)

[Countermeasure against phishing Scams 21](#_Toc68817132)

[Development Methodology 22](#_Toc68817133)

[Requirements Gathering and Analysis 23](#_Toc68817134)

[Data Collection 24](#_Toc68817135)

[System Analysis & Design 24](#_Toc68817136)

[System Analysis 24](#_Toc68817137)

[System Design 24](#_Toc68817138)

[Research Design 26](#_Toc68817139)

[Implementation 26](#_Toc68817140)

[System Testing 26](#_Toc68817141)

[User Evaluation 28](#_Toc68817142)

[Deliver to Use 28](#_Toc68817143)

[References 29](#_Toc68817144)

# Phishing Methods and Techniques

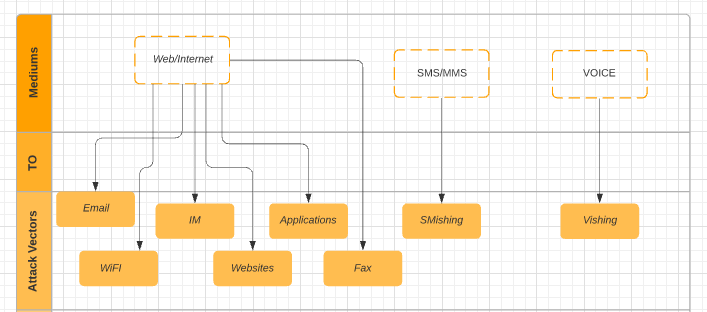
Phishing attacks can be more than drawing money from a person or an organization, the intent is to steal something valuable in the form of a data. The attack can be broken down into stages – Plan (Bait), Setup (hook) and Attack (Catch). These components are interlinked, however are applicable partially, for example some attack vectors are suitable to specific media and different approach while on the other hand the techniques are the various approaches that are utilized during an attack, mostly used in conjunction with Social Engineering to increase the attacker’s chance of success. The paper focusses on mainly 3 wide types of Media i.e, Web, Voice and SMS

With the evolution of Internet and inception of numerous websites, The Internet is now a new method of communication which is used globally to connect and exchange information which ranges from email to social apps providing means to the phishers to hook potential victims

Voice is one of the oldest and longest effective methods that exists and are one of the common ways to convey/exchange information. This can be used to deceive a ‘user’ into revealing their private/personal information and use the information for their benefit

Short Messaging service / Texting via Carriers mobile Network involves short messages providing a convenient way for the scammers/phishers to interact with a ‘user’ or an individual in an attempt to steal their personal information

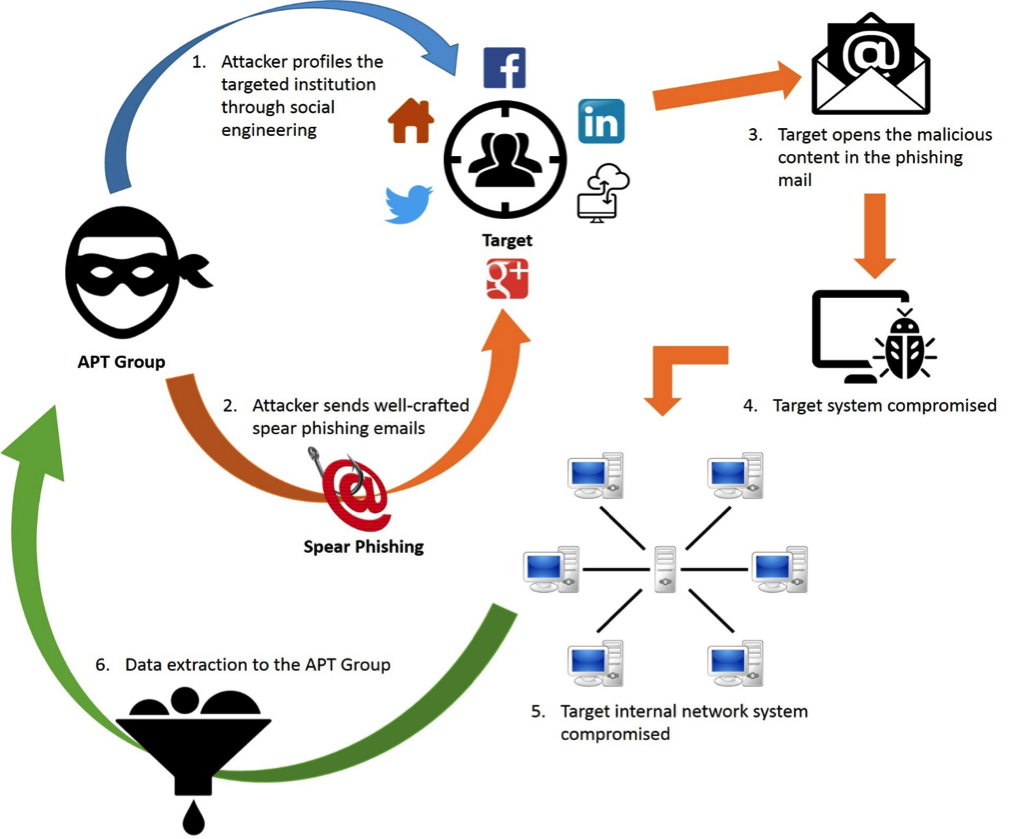
Therefore, the overall phishing can be categorized into Medium, Vectors and Technical approach taken to carry out the method. The below figure explains how Vectors are interlinked to Medium and how a Phishing attack is implemented



Medium to Vector Mappings

Email

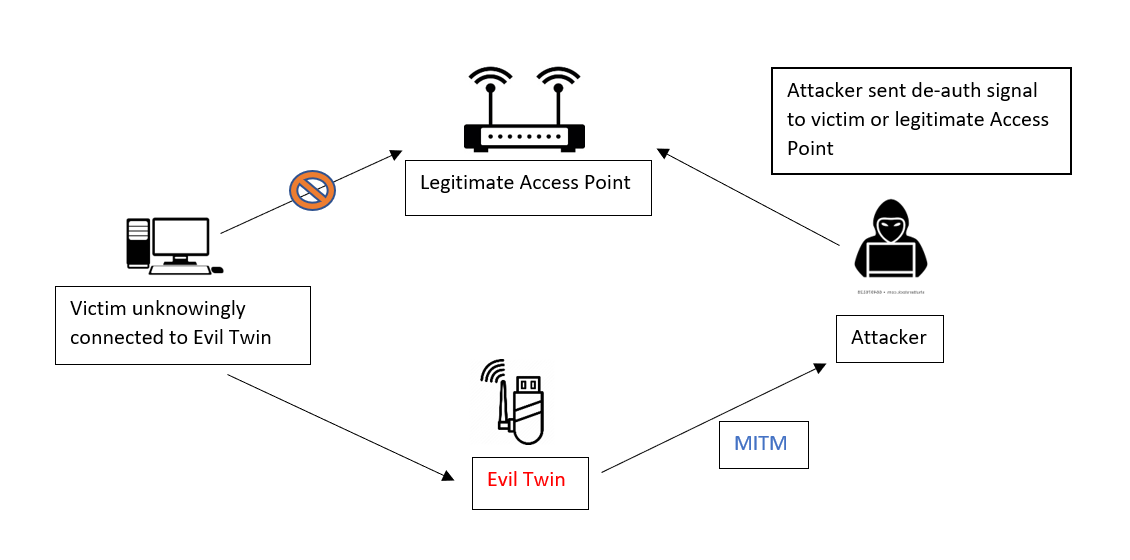
Medium which comprises of wide range of attack vectors is Web/Internet. The famous of all and the first one to consider is electronic mail. In this type of attack the crook usually has a fake domain name registered to closely resemble or mimic a legitimate organization and sends out hundreds of thousands of emails by targeting a wide range of users, these emails are usually crafted in a general form to make it look like original and carry out or to perform actions and lure victim into giving away their personal information to the attacker. This Vector is heavily used and can be easily distributed to a huge number of users/recipients and can hide the geographical location of Attacker



Email Phishing Cycle

WIFI

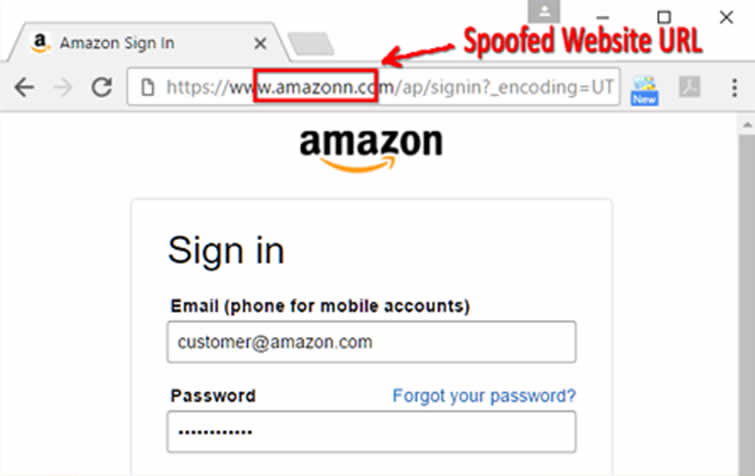
Wifi phishing is carried out in a public hotspot location, the attacker creates a duplicate Network by mimicking a legitimate SSID to harvest victim, often known as ‘Evil Twin’ attack, this makes the user connect to the fake SSID and use this medium in a form of spear phishing , which can install a malware on the Users machine or to intercept traffic on this network by stealing personal information being transmitted by people using the hotspot



Wifi Evil Twin Attack

Websites

Fake or Fraudulent websites are another source for scammers. These websites are made to look legitimate and reap victims PII when a user tries to login. As the users are more inclined towards believing the phishing attacks are most commonly through email or via other messaging services, they tend to pay less attention while visiting websites making them vulnerable to these kinds of attacks. One such example is listed below, the user is trying to visit Amazon.com, however he is unaware there is and additional ‘n’ added to the URL at the end of the URL



Website spoof attack

FAX

Fake eFAX emails are now carrying Trojan or RMS RAT viruses to compromise user’s machine and ultimately to record user’s personal information. This type of vector is similar to a traditional fax, however without the need for a fax machine. The major advantage of this feature is to receive fax in the form emails, thus removing the dependencies on traditional printers.

Text, application

Description automatically generated

eFax Attack

IM

Instant Messaging is an older form of online communication to exchange information. With the dawn of internet relay chat, there have been subsequent amount of IM tools like Yahoo Messenger, google messenger and have evolved into various social media streams and IM applications such as Whatsapp, Signal, telegram etc. Messages are not just text based, they have clearly evolved and now include photos, gifs, hyperlinks etc.., the IM apps are now equipped with Video or Voice calling also making it an idea medium for phishers. This type of real-time chat can engage an individual or a victim in revealing their personal information by threating or informing their A/c has been locked and have to enter their credentials to unlock it accordingly

Graphical user interface, text, application

Description automatically generated

IM Phishing attack

Smishing

Short messaging system or multi-media messaging system is responsible for the following medium of Smishing. While there are multiple ways to lure a victim in this the most famous of all is to send a hyperlink to an users that either directs to a website or downloads the malware, which upon success the phisher can continue with their attacks by creating botnets or through accessing codes for login. Another way of Smishing is to send a text message containing a vital message for example, user or the recipient receives a text from Fedex informing the user their package with tracking code containing a hyperlink as shown in the picture below, the victim is then redirected to a website to enter their login information resulting in an identity theft

Graphical user interface, text, application, chat or text message

Description automatically generated

Smishing attack

Vishing

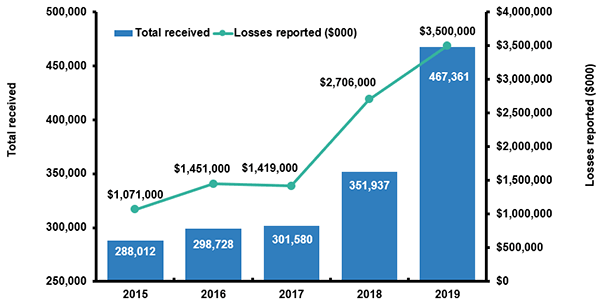
Vishing is an old medium of phishing scams, the evolution of VOIP i.e, voice over IP has boosted this approach in the current generation. Vishing uses the form of a spoofed number which appears as a legitimate caller on the Users Screen. VoIP technology is used to overcome the limitation of a physical location where the call originates from and thus making them look alike to legitimate calls. There are numerous numbers of reasons as to why this form of phishing is often successful

Example, Call centers – the most common way strangers are approached for asking their personal information. Second most common is acceptance of automated telephone systems.

### Phishing Attacks and Defenses

While there are many types of research work that has already been done in Phishing, we are going to pick the most common technique and describe the technical aspects of typical phishing attack and what can be done to avoid these type of scams

((2013), April 2013)Below is a graph which shows the losses due to overall phishing attacks from Krebon Security, the yearly graph shows the rise of loses exponentially every year



(Krebson Security, n.d.) Yearly loss

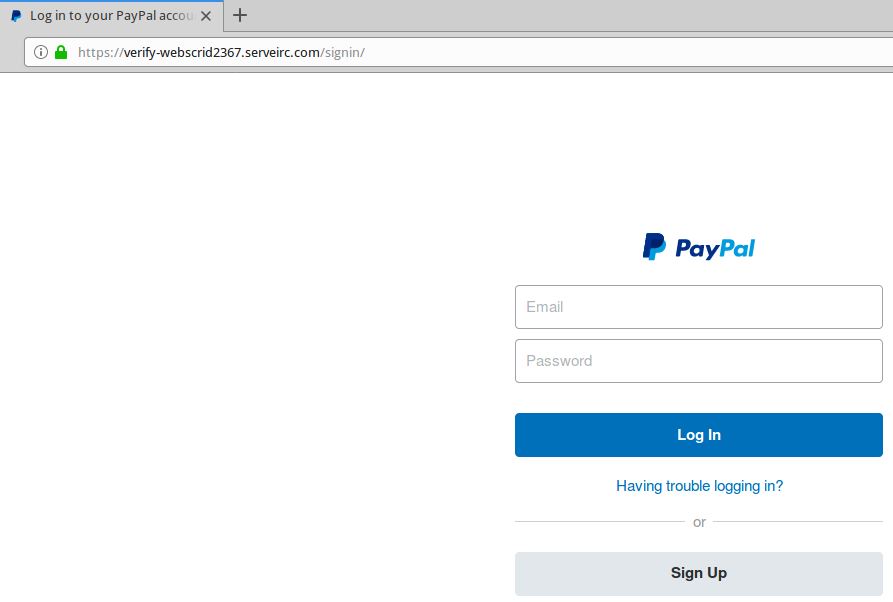
As shown in the below figure, the Global Internet portal has been the industry most favorable to phishers which has accounted for 32% of phish attack, while the second top most popular sector was Banking sector at 18% and payment sector at 10% . The Social Network and blogs accounted for 6% of attacks overall

Chart, pie chart

Description automatically generated

Industry favorable to Phishers

Most of the phishing techniques uses some form of a technical approach in a form of a spoofed email or website, this includes the use of a subdomain through a misspelled url intentionally created by phishers. Below are few phrases that a phishing site may contain



Phishing page

The above page looks like a legitimate PayPal page, however it’s not, it’s a phishing website, as soon as the user puts their credentials, it gets recorded in customers Database which can cause serious harm to an individual as this will compromise their financial account. The methodology and respective approached of phishing attacks are discussed in the below sections

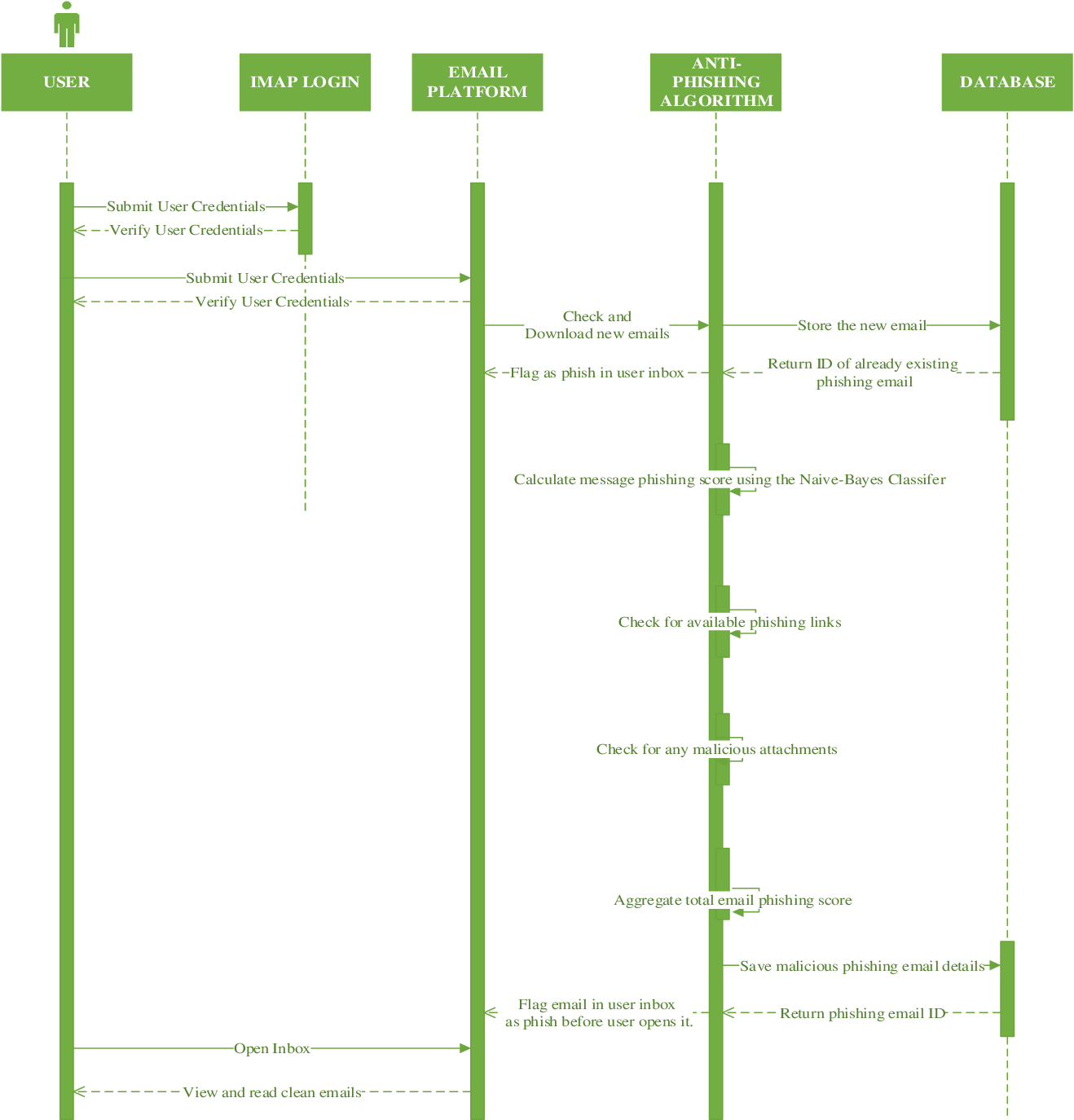
## Related Work

In a whitepaper named SOPHOS published in 2005, Phish Guru provides anti phishing training system that teaches an individual from falling for a phishing attack by sending users simulated phishing emails. The training emails is a part of their day-to-day routine, which is just like a phishing email, urging people to go to a website or to download a file. If a user falls for the training phishing email they are provided with a instruction message and outlines the risk associated with the phishing attempt and offer tips they can follow to protect themselves. By using techniques like above and monitoring users behavior we can use a paradigm analysis of an individual to detect a potential phishing website. This has been the most successful method so far in preventing phishing attacks

(Chuanxiong, 2005) proposed a new anti-phishing algorithm called Link Guard by making use of the general hyperlinks in phishing attacks. Since these are built on generic characteristics Link Guard can not only identify and detect the known attacks but also the unknown attacks. As per (Tom Jagatic, 2005) who have listed context aware phishing shows how an attacker can gain trust of an user or an individual by knowing their shopping preference or their banking history which can be easily revealed from the browser history or cookies or to identify any personal information like their maiden name which is often collected from the data required by the law

# Methodology and Approach

Below are the stages to prevent email phishing attacks, the sequence diagram shows how a user submits his login credential which is run against an Anti-Phishing Algorithm before downloading an email and how can the email be flagged as phishing



Sequence Diagram for preventing a phishing attack

Based on the data collected email phishing can be further categorized based on four common techniques, email with image, email field manipulation, redirection and a pop-up window attach. The first two are categorized under 1 to 3 while the last two are categorized at 1 to 6 types

## Email Image only phishing

This is a common type of attack where the email is enclosed with binary images, gifs etc. to supply a garbage stream of words in order to bypass spam filters. To counter this scenario, we can allow only the plain-text email by disabling the html-based code (Beardsley, 2005) However, lot of users now a days expect a graphical appearance in their email and therefore some creative prevention needs to be implemented in order to protect users from falling for it. One of the options can be to provide ability to choose between a plain text and HTML, this will allow user to switch back to plain text if they find anything suspicious

## Email Field Manipulation

Another form of email phishing is to change headers. Most of the email client software depends on the message ‘From’ field in order to identify a sender, the ‘From’ field can easily be changed with minimal effort on respective SMTP server that allows the use of it without any form of authentication. This attack has evolved, and the recent phishing emails have the ability to not only alter ‘From’ field but also ‘Received’ fields (Consumer Fraud and Identity Theft Complaint Data, 2005)

Email Headers contains the information on the path that an email takes, now you may find conflicting information between a traversed path and a Return-path often containing misleading content. The return path is usually validated through SPF Sender Policy Framework, from the below screenshot the name of the email is obviously not validated by design. (Email Forensics; 2. Headers and Body, n.d.)

Text

Description automatically generated

Email Headers

The above was carried out on a Hotmail SMTP server to mimic the attack above and I was able to configure it on my Outlook without much effort as it doesn’t not need any sort of authentication, this can help the phisher disguise/hide himself to send an email to a victim easily. For example, if the Phisher tries to send an email as an executive body of the company, it can be a CEO, CFO or CTO asking confidential information or personal information from an employee then this would be considered as one of the terrible phishing attacks as this will lead in every sincere employee giving away their information as they would ultimately think it is going to their boss. This can be exposed by checking the Headers, however not many people know about email headers.

The following image below shows how a phishing scan is carried out. There is a spoofed email sent to myuniversity.edu to a wide range of users. The emails is crafted to look exactly like a university email, the email claims that the user password is about to expire and that they must follow the instructions in the email to go to myuniversity.edu/renewal in order to renew their password within 24 hours (Imperva Phishing Attacks, n.d.)

Graphical user interface, text, application, email

Description automatically generated

Phishing email example

The users here are redirected to a phishing page which is designed similar to an original page which requests for both the new and the old password. The Phisher hijacks the password and activates a malicious script in the background to steal user’s cookie, this results in XSS attack providing privileged access to university Network. Similar kind of techniques are applied to make the users fall for an attack, the attacker can threaten the user about their Account expiration and that they have to take action immediately, this creates an unwanted pressure on the user and makes them less conscious and likely to error. If you look at the URL closely, you will see how a misspelled domain name or subdomain name can look similar making the user less aware of an attack

Graphical user interface, application

Description automatically generated

These types of attacks can be avoided by enabling 2FA or any type of Multi Factor Authentication and by adhering to strict password policies, Alternatively the organization can also look to integrate their portal through some sort of a single sign on which will force an user to authenticate against the legitimate server preventing the attack (Microsoft Phishing Filter: A New Approach to Building Trust in E-Commerce Content, Microsoft, 2005)

## Email Redirection

The redirection is considered as one of the creative techniques where a user is redirected to fake / fraudulent websites and the easiest way to do so is to craft random links that can be generated automatically. The URL can easily be obfuscated to make it smaller intentionally instead of pasting a lengthy URL or to change it something like ‘click here to proceed’ which redirects the user to a fraudulent website. There have two other services that have gained traction and are used to perform identity thefts by the phishers since they are free, they are html forms and DNS Redirectors. In this the attacker has a fair chance to resolve IP address to a selected DNS record transparent to email user and can be used to redirect victim to a fraudulent links. On the other hand, HTML forms are integrated within the phishing email with call to action that requires the user to verify their account activity

These types of attacks can be avoided through anti phishing techniques by configuring no hex encoded ASCII characters in domain name and by avoiding use of http link more than one time without nesting. Although there are ways to prevent it, redirection can become very challenging due to the fact redirections are not often associated with phishing

## Pop up window

These types of logins are used in banking sectors to perform login at the same time to track customer functions. This may require the user to disable pop-up blocker in order for them to login to their personal accounts. This might create a challenge as there can be two windows that pop up, one for the user login and the other which might ask identity information but to a fraudulent website. The best approach to prevent this is to educate users on what to expect

# Countermeasure against phishing Scams

* Avoid downloading free software from an unknown site, this will protect you from downloading a malicious file
* Always enable host-based firewall on machines, patch your OS with latest updates on regular basis, also keep the browser and applications always up to date to avoid any zero day
* Educate and know about the fraudulent activities on the Internet to identify phishing scams
* Look for a ssl certificate when you visit a website which usually consists of a padlock symbol on the status bar of your browser
* Never click on a hyperlink embedded in the email without properly analyzing it, the user can make use of website like virus total to identify whether it’s a real site and that there are clean

There are many more detailed anti-phishing steps which an organization can take to protect their users against various level of phishing attacks

# Development Methodology

The following paper will explain the research methodology and justification to why they are preferable, the methodology has six phases which consists of Requirement Gathering, System Design, Implementation, System Testing, User Evaluation and Delivery

Diagram

Description automatically generated

## Requirements Gathering and Analysis

The following was covered in the above two sections while illustrating outlined objective and literature review. This has helped in clear definition of analysis and problem to achieve these objectives

## Data Collection

The Data Collection involved collection of phishing email s for testing , the phishing email were taken from online organizations that have reported phishing attacks , for example (Berkeley Information Security and Policy , 2017)

## System Analysis & Design

### System Analysis

Gathering all the information and their respective key process while identifying emails. This helped design an algorithm from various findings of wide range of phishing emails by ensuring the algorithm met the criteria required to prevent phishing emails

## System Design

System Design was created as per the analysis and in relation to sequence and class diagram shown in the use case section below

Diagram

Description automatically generated

## Research Design

To design an adaptive anti-phishing algorithm to prevent users from falling for phishing attacks. The research takes a Quantitative research technique and helped in incremental learning by making use of Bayes Theorem to implement a classifier (Experimental Research and Design. Retrieved from Okstate.edu, n.d.)

## Implementation

Python is used as the programming language to design an algorithm based on a simplistic approach by (Prasad, 2016), This was also been used to potentially develop a machine learning program in the past providing standard information and learning resources (Richer, 2017). The implementation uses both python and sql lite and are open source utilities

## System Testing

This phase includes deploying research involved and determining whether the system is providing with accurate results and whether or not easy to use

For testing the below email is sent to an user where the algorithm has access to respective IMAP account. The Phishing email (Berkeley Information Security and Policy , 2017) is a brief example on how the credentials can be compromised

Graphical user interface, text, application

Description automatically generated

The algorithm has the below body which shows an extracted email source

Text, letter

Description automatically generated

The below results are returned by running it through a classifier categorizing the information

Text, letter

Description automatically generated

The below screenshot shows how the sample emails were analyzed and the details were saved into the threat table

Graphical user interface, text, application, email

Description automatically generated

## User Evaluation

This will involve random users to test the algorithm and provide their respective feedback

## Deliver to Use

This will be the final phase to test algorithm’s performance built during this research

# References

(n.d.). Retrieved from Krebson Security: https://krebsonsecurity.com/2010/03/fbi-online-fraud-costs-skyrocketed-in-2009/

(n.d.). Retrieved from Experimental Research and Design. Retrieved from Okstate.edu: https://www.okstate.edu/ag/agedm4h/academic/aged5980a/5980/newpage2.html

(2013), A. A. (April 2013). Preventing Phishing Attacks using One Time Password and User Machine Identification. *International Journal of Computer Applications*, Volume 68–No.3, pp. 7-11 .

(2005). Retrieved from Consumer Fraud and Identity Theft Complaint Data: http://www.consumer.gov/sentinel/pubs/Top10Fraud2005.pdf

(2005). Retrieved from Microsoft Phishing Filter: A New Approach to Building Trust in E-Commerce Content, Microsoft: https://www.globaltrust.it/documents/press/phishing/MicrosoftPhshingFilter.pdf

(2017, April 24). Retrieved from Berkeley Information Security and Policy : https:?/security.berkeley.edu/resources/phishing/phishing-examples-archive

A. Alswailem, B. A. (2019). Detecting Phishing Websites Using Machine Learning. *2nd International Conference on Computer Applications & Information Security (ICCAIS), Riyadh, Saudi Arabia*, pp. 1-6, doi: 10.1109.

A. N. Shaikh, A. M. (2016). A literature review on phishing crime, prevention review and investigation of gaps. *10th International Conference on Software, Knowledge, Information Management & Applications (SKIMA), Chengdu*.

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

*APWG*. (n.d.). Retrieved from APWG: https://www.antiphishing.org

*APWG*. (2003). Retrieved from APWG: https://www.antiphishing.org

Bakhshi, T. P. (2008). *A Practical Assessment of Social Engineering Vulnerabilities”, Proceedings of the 2nd International Conference on Human Aspects of Information Security and Assurance.*

Basit, A. Z. (n.d.). *A comprehensive survey of AI-enabled phishing attacks detection techniques*. Retrieved from Telecommun Syst 76, 139–154 (2021). : https://doi.org/10.1007/s11235-020-00733-2

Beardsley, T. (2005). *Phishing Detection and Prvention.* Practial counter-fraud Solutions White Paper, 3Co, Corporation.

Carella A, K. M. (2017). *Impact of security awareness training on phishing click-through rates. In: 2017 IEEE international conference on Big Data (Big Data).*

Chhikara, J. (2013). Phishing & Anti-Phishing Techniques. *Case Study International Journal of Advanced Research in Computer Science and Software Engineering*.

Chuanxiong, J. C. (2005). IEEE Communications and Networking, NSFC, 2005 . *Online Detection and Prevention of Phishing Attacks*.

Easttorn, C. &. (2011). *Computer Crime, Investigation, and the Law. Boston, MA: Course Technology.*

El Aassal, A. B. (2020). An in-depth benchmarking and evaluation of phishing detection research for security needs. *IEEE Access, 8, 22170–22192*.

*Email Forensics; 2. Headers and Body*. (n.d.). Retrieved from https://medium.com/@p.matkovski/email-forensics-2-headers-and-body-3e6280820983

Gürel, M. B. (2013). Detection of phishing attacks. *2018 6th International Symposium on Digital Forensic and Security (ISDFS), Antalya, 2018, pp. 1-5, doi: 10.1109/ISDFS.2018.8355389 Chhikara, Jyoti. (2013). Phishing & Anti-Phishing Techniques*.

Gürel, M. B. (2018). Detection of phishing attacks. *6th International Symposium on Digital Forensic and Security (ISDFS), Antalya*, pp. 1-5, doi: 10.1109/ISDFS.2018.8355389.

Gupta, A. K. (2016). Comparative analysis of features-based machine learning approaches for phishing detection. *10th INDIA-COM, New Delhi, India*.

*Imperva Phishing Attacks*. (n.d.). Retrieved from https://www.imperva.com/learn/application-security/phishing-attack-scam/

J, S. (2017). *Persistent training. In: Advanced persistent training, Apress, Berkele.*

Jakobsson, M. &. (2007). *Phishing and Countermeasures:Understanding the Increasing Problem of Electronic Identiry Theft.* Hoboken, New Jersey.

Jampen, D. G. (2020). *A comparative literature review. Hum. Cent. Comput. Inf. Sci. 10, 33* . Retrieved from https://doi.org/10.1186/s13673-020-00237-7

Khonji, M. &. (2013). Phishing Detection: A Literature Survey. *IEEE Communications Surveys &amp Tutorials*, PP. 1-31. 10.1109/SURV.2013.032213.00009.

Kumaraguru P, R. Y. (2007). *Getting users to pay attention to anti-phishing education: evaluation of retention and transfer. In: Proceedings of the anti-phishing working groups 2nd annual eCrime researchers.* Pittsburgh, PA.

M. Adil, R. K. (2020). Preventive Techniques of Phishing Attacks in Networks. *3rd International Conference on Advancements in Computational Sciences (ICACS), Lahore, Pakistan, 2020*, pp. 1-8, doi: 10.1109/ICACS47775.2020.9055943.

Moore, T. &. (2007). *Examining the Impact of Website Take-down on Phishing. Proceedings of the Anti-Phishing Working Group 2nd Annual e-Crime Researchers Summit.* Pittsburgh.

Neupane A, R. M. (2015). *A multi-modal neuro-physiological study of phishing detection and malware warnings. In: Proceedings of the 22nd ACM SIGSAC conference on computer and communications security—CCS ’15. ACM Press,.* Denver, Colrado.

Orunsolu AA, S. A. (2017). *An empirical evaluation of security tips in phishing prevention: a case study of Nigerian banks. Int J Electron Inf Eng.*

Prasad. (2016). Retrieved from Top 20 Python Machine Learning Open Source Projects, updated. Retrieved from KD nuggets:: http://www.kdnuggets.com/2016/11/top-20-python-machine-learning-open- source-updated.html

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

Richer, &. C. (2017). Retrieved from Saed. Naive Bayesian. Terieved from saedsayad.com: http://www.saedsayad.com/naive\_bayesian.html

Tom Jagatic, N. J. (2005). Social Phishing. *Communications of ACM*.

Vahid Shahivari, M. M. (2020). *Phishing Detection Using Machine Learning Techniques.*

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

White, M. &. (2008). *Assessing Our Knowledge of Identity Theft: The Challenges to Effective Prevention and Control Efforts. Criminal Justice Policy Review.*