# Research of Detection and Prevention of Phishing and Proposal of Phishing Detector Solution

Chin Xin Yi
*School of Computing*
*Asia Pacific University of Technology*
*and innovation (APU)*
Kuala Lumpur, Malaysia
Tp059119@mail.apu.edu.my

Dr. Intan Farahana Binti Kamsin
*School of Computing*
*Asia Pacific University of Technology*
*and innovation (APU)*
Kuala Lumpur, Malaysia
intan.farahana@staffemail.apu.edu.my

***Abstract*- In these last few years, the cases of online phishing scams rose swiftly, making hundreds of thousands of victims lost their sensitive information which could expose them in danger. To solve this problem, this research aims to build an improved detective and preventive anti-phishing software that can more effectively decrease the phishing scams worldwide. To further extending the previously solutions and develop an upgraded software by providing a more comprehensive, secured and easy way in order to efficiently prevent and detect phishing scams. To carry out the data collection session, the respondent is decided to be the staff and students in Asia Pacific University and the expected number of respondents is 100. As for the process of sampling the data gathered, the systematic sampling, which is categorized under probability sampling, is selected in this paper. Moreover, quantitative research will be decided to be the technique to collect data in this project, as one of the types of quantitative research, survey is comparatively suitable than the other methods. Undoubtedly, phishing attacks has become one of the most serious cybercriminals, causing loss of personal information and financial. Therefore, the proposed solution should be spread and applied in order to mitigate the cases of phishing attacks, achieving a secured internet environment for the users.**

***Keywords— Phishing, Detect, Prevent, Anti-phishing***

## I. INTRODUCTION

In these last few years, the cases of online phishing scams rose swiftly, making hundreds of thousands of victims lost their sensitive information which could expose them in danger. To solve this problem, this research aims to build an improved detective and preventive anti-phishing software that can more effectively decrease the phishing scams worldwide. Phishing is a social engineering method which intends to attack a system bug at the user's end. For instance, a system may be technologically secured to avoid password theft, but if an attacker delivers a phished update password request, an unsuspecting user's password may be exposed. By phishing, attackers aim to illegally get personal information from customers. Moreover, it is believed that phishers use a variety of techniques to deceive the victims. For example, email messages, instant chats, forum comments, phone calls, and social networking information (Philip, Varghese, & Thomas, 2021). When the enters their credentials on a malicious website given by the attacker, the latter gains the victims login credentials that can be used for malicious purpose. According to the report revealed by the Aite Group, U.S. Identity Theft: The Stark Reality, discovered that losses from identity theft cases estimated $502.5 billion in 2019 and are set to escalate 42 percent to $712.4 billion in 2020. Not only that, the losses are projected to soar to $721.3 billion (Insurance Information Institute, 2020).

## II. LITERATURE REVIEW (RESEARCH DOMAINS)

### *A. Phishing*

Phishing is identified as the deception of obtaining important information such as usernames, passwords, bank account numbers, and credit card numbers for illegal attempts. Phishing schemes are now considered to be the most typical type of cybercrime. Phishing attacks can be found in diverse of places, including the online payment industry, webmail, financial institutions, file hosting and cloud storage, so on and so forth (Pujara & Chaudhari, 2018). According to the article by scholar from Canada, the main reasons making users being trapped in the phishing scams are lack of knowledge regarding cyberattacks, duplicity of visual and attention sufficiency, respectively. (Dhamija, Tygar, & Hearst, 2006) Most of the users do not expose to the information technology deeply and frequently, so that they have no enough knowledge about computer system and security and security indicators, resulting in easily to be frauded. Next, to visually deceit the users, phishers use a variety of measures including deceptive text, images with underlying text masked, images that mimics windows, windows with underlying windows masked, and deceptive appearance and feeling. (Dhamija, Tygar, & Hearst, 2006). The type of phishing includes Deceptive Phishing, Spear Phishing, Clone Phishing, Whaling, Link Manipulation, Voice Phishing (Bhavsar, Kadlak, & Sharma, 2018), Malware-Based Phishing, Web Trojans, Hosts File Poisoning, System Reconfiguration Attacks, DNS-Based Phishing aka Pharming, Content-Injection Phishing, Search Engine Phishing (Gaurav, Mishra, & Jain, 2012). A weird tone or greeting, grammatical and spelling flaws, abnormalities in email addresses, URLs, and domain names, threats or a sense of danger; suspicious attachments; peculiar requests, particularly for credentials; the communication was no longer initiated by the recipient; information about payments or other personal information (Cofense, 2021).

### *B. Phishing Detection*

It uses specialized technology, to reveal something that is partially hidden or ambiguous. Phishing detection systems that detect a phishing website by using a web browser on the client environment or professional software on the server, or phishing detection and user training courses. This is because they depend on inexperienced Internet users only infrequently. When a website is determined as a phishing or suspected phishing website, it is either blocked or the user is warned that the website may be illegitimated. This solution required very little user training and did not require any adjustments to a website's current authentication mechanisms. The following measures are used to evaluate the detection methods' accuracy.

### C. *Phishing Prevention*

It is to avoid something from occurring or someone from acting up in a particular manner. Phishing prevention systems use two-factor authentication and two-way authentication to create an additional protection to authentication protocols and user interaction platforms, preventing phishing threats. This minimises the chances of a user being misled by a phishing website created by an attacker. Watermarking-based, RFID-based, external authentication devices-based, image password-based, dynamic security skin-based, smart card-based, and QR Code-based techniques, for instance, are all types of phishing prevention strategies. (Varshney, Misra, & Atrey, 2016) Most phishing attacks may be prevented using these approaches, but they need improvements and assistance on the part of the website, as well as cooperation and understanding on the part of the user.

### D. *Anti-phishing*

Anti-phishing is a term used to describe measures to prevent phishing attempts. Anti-phishing works in helping users to distinguish and filter different classifications of phishing attacks. There are several anti-phishing techniques such as Content Filtering, Black Listing, Symptom-Based Prevention and Domain Binding provide support on emails, web sites as well as addresses, web page and browser, respectively. (Gaurav, Mishra, & Jain, 2012)

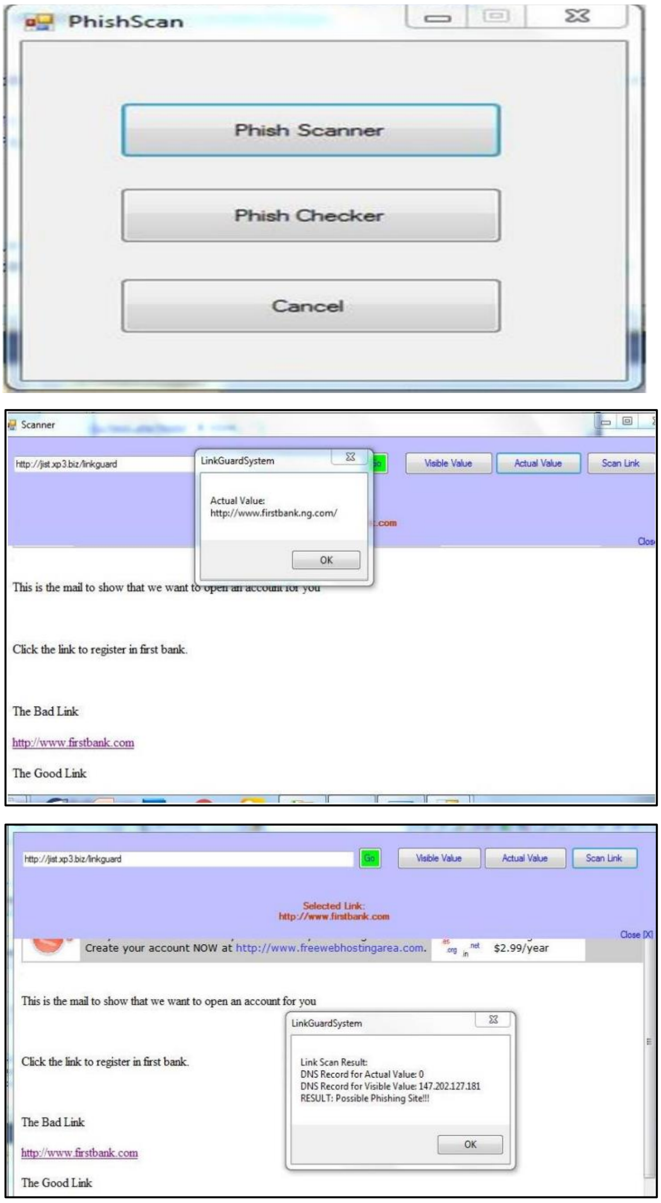
TABLE I. PHISHING, PHISHING-DETECTION, PHISHING PREVENTION, ANTI-PHISHING

| | Phishing | Phishing-Detection | Phishing Prevention | Anti-phishing |
|---|---|---|---|---|
| Juan Chen & Chuanxiong Guo, 2006, China. | Yes | Yes | Yes | Yes |
| Asha Joseph & K. John Singh, 2018, India. | Yes | Yes | Yes | Yes |
| Michael Stepp, 2005, United States. | Yes | Yes | Yes | No |
| Edona Fasllija & Hasan Ferit Enişer & Bernd Prünster, 2019, Austria & Turkey. | Yes | Yes | No | No |
| Jason E. Thomas, 2018, United States. | Yes | Yes | Yes | Yes |
| Achu Thomas Philip & Bibin Varghese & Dr. Smita C Thomas, 2021, India. | Yes | Yes | Yes | Yes |

## III. SIMILAR SYSTEM

### A. *LinkGuard*

Due to its scalability and capacity to execute many languages, the proposed system will be constructed with the C# programming language programmed in Visual Studio Platform, as well as a MySQL database which will function as a storage for the links. The system will be a software program that functions similarly to a web browser (Akinyede, Adelakun, & Olatunde, 2018). Phishers frequently use unlawfully obtained information to encourage target to access the links attached in the phishing e-mail (Singh, Sakshi, & Jayant, 2019). The LinkGuard Algorithm works through the comparison between the visual connection and actual link. The URL's likeness to a known trustworthy site is also calculated using this approach. The implementation of LinkGuard is seen in the following screenshots.



http://jist.xp3.biz/linkguard

LinkGuardSystem

Actual Value:
http://www.firstbank.ng.com/

This is the mail to show that we want to open an account for you

Click the link to register in first bank.

The Bad Link

http://www.firstbank.com

The Good Link

http://jist.xp3.biz/linkguard

Selected Link:
http://www.firstbank.com

Create your account NOW at http://www.freewebhostingarea.com. $2.99/year

This is the mail to show that we want to open an account for you

Click the link to register in first bank.

LinkGuardSystem

Link Scan Result:
DNS Record for Actual Value: 0
DNS Record for Visible Value: 147.202.127.181
RESULT: Possible Phishing Site!!!

The Bad Link

http://www.firstbank.com

The Good Link

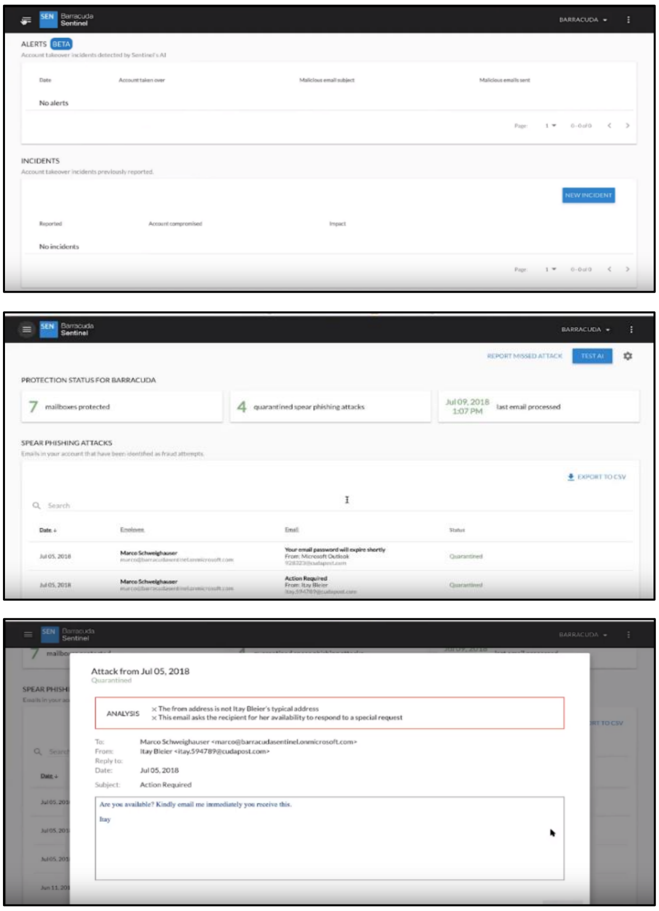
Fig. 1. Implementation of LinkGuard (Akinyede, Adelakun, & Olatunde, 2018)

### B. *Cantina*

CANTINA analyses the content of a webpage to determine the validation, as opposed to other methodologies that look at the external features of a website, such as the URL and domain name. The TF-IDF (term frequency/inverse document frequency) information retrieval algorithm is used by CANTINA. CANTINA is good at detecting phishing websites in users' legitimate email when it is used in connection with algorithms, and its most common error is mislabeling spam-related URLs as phishing URLs (Nadar, Thakur, & Junankar, 2020).

### C. *Barracuda Sentinel*

Barracuda Sentinel is a comprehensive cloud-based tool that supports artificial intelligence, interconnection with Microsoft Office 365, and brand protection to secure against company email corruption, impersonation attack, phishing, and other cybercrimes (Network). The following screenshots are the implementation of Barracuda Sentinel.



ALERTS

Account takeover incidents detected by Sentinel's AI

| Date | Account taken over | Malicious email subject | Malicious emails sent |
|---|---|---|---|
| No alerts | | | |

INCIDENTS

Account takeover incidents previously reported.

| Reported | Account compromised | Impact |
|---|---|---|
| No incidents | | |

PROTECTION STATUS FOR BARRACUDA

7 mailboxes protected

4 quarantined spear phishing attacks

Jul 09, 2018 1:07 PM last email processed

SPEAR PHISHING ATTACKS

Emails in your account that have been identified as fraud attempts.

| Date ↓ | Employee | Email | Status |
|---|---|---|---|
| Jul 05, 2018 | Marco Schweighauser<br>marco@barracudasentinel.onmicrosoft.com | Your email password will expire shortly<br>From: Microsoft Outlook<br>928323@cudapost.com | Quarantined |
| Jul 05, 2018 | Marco Schweighauser<br>marco@barracudasentinel.onmicrosoft.com | Action Required<br>From: Itay Bleier<br>itay.594789@cudapost.com | Quarantined |

Attack from Jul 05, 2018

Quarantined

ANALYSIS
× The from address is not Itay Bleier's typical address
× This email asks the recipient for her availability to respond to a special request

To: Marco Schweighauser <marco@barracudasentinel.onmicrosoft.com>
From: Itay Bleier <itay.594789@cudapost.com>
Reply to:
Date: Jul 05, 2018
Subject: Action Required

Are you available? Kindly email me immediately you receive this.

Itay

Fig. 2. Implementation of Barracuda Sentinel (Advice, 2022)

TABLE II. PHISHING COMPARE AND CONTRAST TABLE

| | LinkGuard | Cantina | Barracuda Sentinel |
|---|---|---|---|
| To enhance the ability of detection especially in form of image and videos. | No | No | No |
| To be able to handle enhanced attack, Cross-Site Scripting (XSS). | No | No | No |
| To be independent from other techniques to assist in operating. | No | No | Yes |
| To offer a large number of training materials and options to users. | No | No | No |

The proposed system is aimed to achieve the goals above including enhancement of the ability of detection especially in form of image and videos, the other type of enhanced attack such as XSS, not replying on the other techniques to assist in detecting, as well as allowing a large amount of training materials and options to users (Joesph & Singh, 2018) (Xiang, Hong, Rose, & Cranor, 2011). The majority of the stated abilities have yet to be fulfilled by the systems indicated. As a result, the proposed system is believed to be necessarily established and upgraded to strengthen anti-phishing efficiency.

## IV. PROBLEM STATEMENT, AIMS AND OBJECTIVES

With the rapidly grown of popularity of internet all around the world, internet has become part of many people's daily basis. However, it could lead to cybercrimes as some skilful people cheats the others in order to earn benefits such as money through improper way. For example, phishing scams. Phishing scams are getting more common in recent years, becoming the most domain danger to worldwide. FBI's Internet Crime Complaint Center (IC3) reveals that phishing scams is claimed to be have the most victims in 2020, with more than 240,000 victims reporting about $50 million in losses (Brennan & Smith, 2021). Cybercriminals intend to steal user's personal and financial information by using malicious software or social engineering (Gupta, Tewari, Jain, & Agrawal, 2016). Although there are several anti-phishing approaches such as LinkGuard, PhishHook and Barracuda Sentinel which can detect and prevent the phishing, they are unable to root out the issues of phishing as they have some limitations. Phishing scams could result in identity fraud, leading to financial loss and the victims may face the issues of reputation, credit and others (IRS, 2018). In short, it is necessary to have an improved anti-phishing solution in order to more efficiently solve the phishing issue globally.

The aim of this research is to further extending the previously solutions and develop an upgraded software by providing a more comprehensive, secured and easy way in order to efficiently prevent and detect phishing scams. While the objectives of the research is:

- To enhance the ability of detection especially in form of image and videos.
- To be able to handle enhanced attack, Cross-Site Scripting (XSS).
- To be independent from other techniques to assist in operating.
- To offer a large number of training materials and options to users.

In the world of internet, everyone may become the victims of cyber-attacks especially phishing. According to the Internet Crime Report 2020 released by FBI, phishing was the most popular among cybercrime, having 114,702 cases in 2019 and 241,324 occurrences in 2020. Not only that, it is said that there were more than 11 times as many phishing claims as in 2016 (Federal Bureau of Investigation, 2020). This issue can be traceable back to the lack of awareness and knowledge of the users. There is a report carried on in order to investigate how many users are able to identify a phishing scam by testing 19,500 people from 144 countries. The result indicates that only 3% correctly identified all examples given and 80% misidentified at least one of the phishing emails, which may lead to information loss such as credentials or credit card information (Rossi, 2015). Furthermore, the tricks of phishing have become more and sophisticated. Cofense, a leading company that aims to stop phishing attacks, reveals the increasing of the types of phishing attacks. Bad actors rapidly attack users through the trusted platform such as SharePoint and OneDrive by using cloud filesharing services. In 12 months, there are more than 5,200 Sharepoint phishing emails were claimed, also about 2,000 OneDrive attacks. In addition, some trusted institutions including Amazon AWS, Google and Adobe has been reported that faced phishing issues in 2021 (Cook, 2021). Therefore, it is vital to find out a better and more efficient solution which can largely mitigate the phishing problem as well as contribute to the reduction of identity theft and fraud globally, making the environment in internet secured and reliable.

## V. Methodology

To carry out the data collection session, the respondent is decided to be the staff and students in Asia Pacific University and the expected number of respondents is 100. The data gathering process will be ethically conducted including making sure the respondents are willing to answer the survey and ensuring the integrity as well as confidentiality of responses.

As for the process of sampling the data gathered, the probability sampling, which can be also called random sampling, every element in the universe is allowed to have equal chance to be selected by using this technique. (Ebeto, 2017) In short, this approach emphasizes the fairness and equality of the options. In this paper, systematic sampling is selected to conduct among the types of probability sampling.

Systematic sampling picks respondents from a broader population according to a random starting point but with a specified, periodic interval. This interval is also known as the sampling interval is computed by deriving the total population by the sample size required. (HAYES, 2022) After collecting the data from respondents, the population list will be arranged according to the ascending order of answering the questions, so that the population order is random and it could provide a representative sample which can illustrate a general conclusion.

The reason of choosing the systematic sampling because it brings out many benefits including the easiness of execution and comprehension. Besides that, systematic sampling is comparatively structured than other method, keeping the task of selecting a sample group simpler and less difficult. Not only that, by using systematic sampling, the chance of bias is reduced since the sample frame's list is arranged in a random order. (Voxco, 2021)

Quantitative research will be decided to be the technique to collect data in this project, it numerically represents information which can be computed using mathematics, allowing the researchers to analyse and comprehend it through data analysis. (Chipeta, 2020) Generally, it concentrates the closed-ended questions such as "yes or no" and "what". The typical examples of quantitative research are surveys, case studies and questionnaires. Usually, the quantitative data illustrates the preferences of respondents instead of the reason of making the choices. The reasons choosing quantitative techniques are less expensive during implementing as only the paper or URL is required, the risk of bias is mitigated since the quantitative data is statistical so the own thinking of researcher does not impact much on the result, as well as easier to analyse because the answer provided is simple and certain. (kwiksurveys, 2021).

Fig 3. illustrates how the proposed system will work. First and foremost, it will get the current URL from the user. Next, the analyser which acts as a similarity comparison will start working, if the URL is determined as negative, it will be stored in the database. Otherwise, it will be considered as phishing site and the alerter will be triggered in order to inform the user.

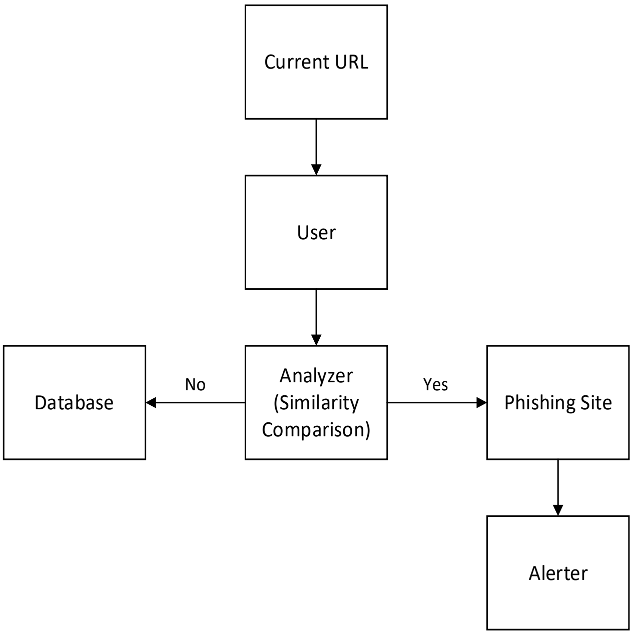


Fig. 3. Proposed System Overview

## VI. Conclusion

Undoubtedly, phishing attacks has become one of the most serious cybercriminals, causing loss of personal information and financial. Hence, to solve the mentioned issue, an anti-phishing solution that can detect and prevent the malicious websites has been proposed by integrating the strengths and improving the weaknesses of the others similar system proposed previously. The measures mentioned in this paper should be spread and applied in order to mitigate the cases of phishing attacks, achieving a secured internet environment for the users.

## References

Editor. (2020). *What is a hex editor?* SweetScape Software Inc - 010 Editor - Pro Text/Hex Editor | Edit 200+ Formats | Reverse Engineering. https://www.sweetscape.com/articles/hex_editor.html

Achu Thomas Philip, Bibin Varghese, & Smifa C Thomas. (2021). Anti-Phishing Technology Using Neuro-Fuzzy Approach On Fog. *Introduction*, 1.

ADAM HAYES. (2022, January 8). *Systematic Sampling*. Investopedia. https://www.investopedia.com/terms/s/systematic-sampling.asp#:~:text=Systematic%20sampling%20is%20a%20type,by%20the%20desired%20sample%20size

Akinyede, R. O., & Adelakun, J. A. (2018). Detection and prevention of phishing attack using Linkguard algorithm. *Journal of Information*, *4*(1), 10-23. https://doi.org/10.18488/journal.104.2018.41.10.23

Barracuda Network. (n.d.). Barracuda Sentinel - MSP. Barracuda MSP Solutions | Security, Backup and Recovery for MSPs.

https://barracudamsp.com/resources/pdf/data-sheets/DS_Sentinel_MSP_Final.pdf

Ben Rossi. (2019, March 29). *Think you can spot a scam? 97% of people wouldn't know a phishing email if it hooked them*. Information Age. https://www.information-age.com/think-you-can-spot-scam-97-people-wouldnt-know-phishing-email-if-it-hooked-them-123459514/

Brooke Brennan, & Kevin Smith. (2021, June 22). *FBI tech Tuesday: Protecting yourself from spoofing and phishing scams*. Federal Bureau of Investigation. https://www.fbi.gov/contact-us/field-offices/phoenix/news/press-releases/fbi-tech-tuesday-protecting-yourself-from-spoofing-and-phishing-scams

Carlo Ebeto. (2017). Biometrics & Biostatistics International Journal Sampling and Sampling Methods. *Sampling*, 2.

Chipeta, C. (2020, June 15). *Best data collection methods for quantitative research*. Tools and support for product and pricing research - Conjoint.ly. https://conjointly.com/blog/data-collection-quantitative-research/

Cofense. (2020, March 23). *10 most common signs of a phishing email: Is your business protected?* https://cofense.com/knowledge-center/signs-of-a-phishing-email/

Federal Bureau of Investigation. (2020). *Internet Crime Report 2020*. FBI. https://www.ic3.gov/Media/PDF/AnnualReport/2020_IC3Report.pdf

FreeOnlineSurveys. (2020, March 16). *Quantitative data collection and analysis*. https://freeonlinesurveys.com/survey-research/quantitative-data-collection

Gaurav, Madhuresh Mishra, & Anurag Jain. (2012). Anti-Phishing Techniques: A Review. *Classification of phishing attacks*, 350-355.

Gupta, B. B., Tewari, A., Jain, A. K., & Agrawal, D. P. (2016). Fighting against phishing attacks: State of the art and future challenges. *Neural Computing and Applications*, *28*(12), 3629-3654. https://doi.org/10.1007/s00521-016-2275-y

Insurance Information Institute. (2020). *Facts + statistics: Identity theft and cybercrime*. III | We are the trusted source of unique, data-driven insights on insurance to inform and empower consumers. https://www.iii.org/fact-statistic/facts-statistics-identity-theft-and-cybercrime

IRS. (2021, November 23). *Phishing, identity theft and scams*. Internal Revenue Service | An official website of the United States government. https://www.irs.gov/newsroom/phishing-identity-theft-and-scams

Joseph, A., & Singh, K. J. (2018). Real time detection of phishing attacks using a variant of Linkguard algorithm. *Journal of Computational and Theoretical Nanoscience*, *15*(11), 3303-3307. https://doi.org/10.1166/jctn.2018.7615

Km Sakshi Gupta, Mr. Pratik Singh, & Dr. KP Jayant. (n.d.). A Comprehensive Study on Phishing Attack Taxonomy and LinkGuard Algorithm. *Abstract*, 1.

Nadar, L. P. (2020). Phish-hook: Phishing site detection using URL features. *International Journal for Research in Applied Science and Engineering Technology*, *8*(7), 1304-1305. https://doi.org/10.22214/ijraset.2020.30421

Purvi Pujara, & M. B. Chaudhari. (2018). *Phishing Website Detection using Machine Learning : A Review*. ResearchGate | Find and share research. https://www.researchgate.net/profile/Er-Purvi-Pujara/publication/331198983_Phishing_Website_Detection_using_Machine_Learning_A_Review/links/5c6bd4ae4585156b5706e727/Phishing-Website-Detection-using-Machine-Learning-A-Review.pdf

Rachna Dhamija, J. D. Tygar, & Marti Hearst. (2006). Analysis of a Phishing Database. *Why Phishing Works*, 10.

S., D. (2019, December 31). Barracuda Sentinel. SoftwareAdvice. https://www.softwareadvice.com.au/software/92413/barracuda-sentinel

Sam Cook. (2021, May 17). *Phishing statistics and facts for 2019–2021*. Comparitech. https://www.comparitech.com/blog/vpn-privacy/phishing-statistics-facts/

Syed Muhammad Sajjad Kabir. (2016). METHODS OF DATA COLLECTION. *Types of Data*, 201-275. https://www.researchgate.net/publication/325846997_METHODS_OF_DATA_COLLECTION

Vaishnavi Bhavsar, Aditya Kadlak, & Shabnam Sharma. (2018). Study on Phishing Attacks. *TYPES OF PHISHING ATTACK*, 28. https://www.researchgate.net/publication/329716781_Study_on_Phishing_Attacks

Varshney, G., Misra, M., & Atrey, P. K. (2016). A survey and classification of web phishing detection schemes. *Security and Communication Networks*, *9*(18), 6266-6284. https://doi.org/10.1002/sec.1674

Voxco. (2021, October 21). Systematic sampling. *Voxco*. https://www.voxco.com/blog/systematic-sampling/

Xiang, G., Hong, J., Rose, C. P., & Cranor, L. (2011). CANTINA+. *ACM Transactions on Information and System Security*, *14*(2), 1-28. https://doi.org/10.1145/2019599.2019606