CPE106L – SOFTWARE DESIGN LABORATORY

3rd Quarter SY 2019-2020

**P.I.R.A.N.H.A**

BY

**Arnett Abarquez**

**Regina Bianca A. Alde**

**Mary Lee D. Apelo**

**Joshua Ron G. Garcia**

**Joshua C. Gonzalez**

**Omar Mukhtar Y. Julkipli**

**Dominic Narciso T. Ronquillo**

**Engr. Dionis A. Padilla**

Professor

**MAPUA UNIVERSITY**

**April 2020**

**ACKNOWLEDGEMENT**

The group would be pleased to acknowledge and express our gratitude for the support and assistance on this project. The group would like to thank our professor, **Engr. Dionis A. Padilla,** his expertise in the field of programming served vital in the formulation of this program.

The group would like to acknowledge the hands that help us, from our classmates from our subject to our org-mates in our organization that guided and assisted us. They were a big help to the contribution because of their support.

And finally, we would like to thank the Almighty for the source of all knowledge, understanding, and wisdom. For we had the patience and will to create our project.

**TABLE OF CONTENTS**

TITLE PAGE i.

ACKNOWLEDGEMENT ii.

TABLE OF CONTENTS iii.

ABSTRACT iv.

Machine Problem Background and Introduction 1

**Abstract**

The program focuses on Detecting Phishing Websites using Machine Learning to accurately detect the Phishing Websites. Phishing Websites targets the human vulnerabilities rather than the software vulnerabilities. It is the process of attracting online users in order to acquire the user’s sensitive information such as password and usernames. The program was developed using the Python programming language, the students programmed a machine learning algorithm in order to detect the phishing websites. The users of this program can use it to become aware on these phishing websites and protect their sensitive information by detecting phishing websites.

**Keywords: Phishing, Phishing Websites, Detection, Machine Learning.**

**Background and Introduction**

The internet is the most rapidly changing technology in our time, it plays an important role in most individuals life. The internet has become this most convenient and valuable mechanism for supporting transactions such as e-banking and e-commerce transactions. It gained the trust of users to provide their private information on online platforms, because of its convenience and how the different institutions are transferring their data in the internet.

As a result, thieves want to get their hands on that private information, it has been a major security problem since then. One of the tools that security thieves use is Phishing Websites, which are fake websites that rely on human vulnerabilities rather than software vulnerabilities to get information from the user.

Detecting Phishing Websites using machine learning is one way to prevent users to be deceived by a Phishing Website. To Detect Phishing Websites using machine learning, before a user opens a website the user copies the link of the website to the detecting program, then the program will show whether the website is phishing or real website.

**Overview**

The students created a program that detects phishing websites using machine learning algorithm that checks whether the website is Phishing or a real website. The students used their knowledge that they obtained in the subject CPE106L and created a project that detects phishing websites with the use of machine learning algorithm, phishing websites are then stored in a file so that the program can comprehend and compare the differences between a phishing website and a non-phishing website.

**Statement of the Problem**

As we go on our daily lives, we learn how to adopt and excel to the technology that we are experiencing right now. This mechanism that gave us leisure and convenience taught us how to have an easy life as well as providing our personal information online. People tend to experience the danger of giving it to fake websites that needs some online transactions or online payments. In this way, it can harm their funds and their trust in doing online settlements.

**Objectives**

The main objective of this project is to effectively detect and predict phishing websites, including, but not limited to e-commerce through the use of an algorithm.

* + To create a program that allows the admin to add phishing website URL into the system wherein the system could access and scan the phishing website and by using algorithm, it will add new suspicious keywords into the database.
  + To create a program that allows users to register themselves to the website to login into the system.
  + To create a data base for the blacklist that stores malicious and/or phishing websites.
  + To create a program that allows the users to send a feedback to the admin with regards to the website inputted.
  + To create a program that allows the admin to change his/her registered password for security purposes.
  + To apply the lessons that the students have learned in Software Design.
  + To create the programs using Python.

**Significance of the Study**

Phishing has been a common attack on vulnerable people by making them disclose their personal information using counterfeit websites. The aim of such websites is to steal personal information such as username, password, banking information, and many more. Phishers use the websites that are very similar in terms of visual and semantics to the actual websites. As technology continues to evolve, phishing strategies have begun to advance rapidly and to avoid this, there is a need to use anti-phishing tools to detect phishing. Machine learning is an effective instrument in combating phishing attacks.

The generalization of this project would be a great contribution to the vast knowledge in relation to phishing. As mentioned, it can be used by a large number of e-commerce websites to have a good consumer-customer relationship. It also gives the users ease by allowing them to make online payment securely. Through the use of an algorithm, it provides better performance and results as compared to other traditional classification algorithms.

**Scope and Delimitations**

The project focuses on detecting and predicting phishing websites, most especially, those of e-commerce websites. It also allows the users to register themselves and input malicious websites, which would then be detected by the system if the website is good—safe to use, or bad—phishes for confidential information such as banking details, personal details, and many more. Moreover, it lets the users to send a feedback to the admin regarding the website inputted. Lastly, it allows the admin to change his/her registered password for security purposes.

The project limits the students to focus on the use of Python programming language.

**Review of Related Literature**

1. **Phishing**

Phishing is a cybercrime that targets the human vulnerabilities to obtain sensitive information from a human. Phishing can be done through emails, texts, calls, and websites, by impersonating a legitimate institution. Once they get the sensitive information from the person, it can result in Identity theft and Financial Loss.

1. **Types of Phishing**
   1. **Vishing** 
      * Vishing is a portmanteau of the words voice and phishing. Vishing is done over calls; they impersonate an institution or person when they call.
   2. **Smishing** 
      * SMS Phishing or Smishing is done over text messaging. They can ask sensitive details from you or send a link that can lead to a Phishing Website.
   3. **Search Engine Phishing** 
      * Search Engine Phishing is done by targeting specific words in a search engine, then when the person clicks the link that looks like a legitimate website, the link goes to a Phishing website that impersonates the legitimate website.
   4. **Spear Phishing** 
      * This type of Phishing is designed to target a group of users. This kind of Phishing has greater risks of stealing the user's sensitive information, because the phishers have already researched their target users.
   5. **Whaling** 
      * Whaling is like Spear Phishing; the difference is that Whaling targets the big bosses of the companies like CEO, CFO, and COO. Those positions are considered big players in the information that they know, and they are called “whales”.
2. **Characteristics of Phishing Domains**
3. **Features Used for Phishing Domain Detection**
   1. **URL-Based Features**
   2. **Domain-Based Features**
   3. **Page-Based Features**
   4. **Content-Based Features**
4. **Detection Process**
5. **Sample Test Run with Screenshots and Discussion**

**BIBLIOGRAPHY**

**Alswailem, A., Alabdullah, B., Alrumayh, N., & Alsedrani, A.** (2019). *Detecting phishing websites using machine learning*. 2019 2nd International Conference on Computer Applications & Information Security (ICCAIS). Retrieved on April 7, 2020 from [https://doi.org/10.1109/cais.2019.8769571](https://doi.org10.1109/cais.2019.8769571)

**Kiruthiga, R., & Akila, D.** (2019). *Phishing Websites Detection Using Machine Learning.* International Journal of Recent Technology and Engineering (IJRTE). Retrieved on April 9, 2020 from <https://www.ijrte.org/wp-content/uploads/papers/v8i2S11/B10180982S1119.pdf>

**Büber, E.** (2018). *Phishing URL Detection with ML.* Towards Data Science. Retrieved on April 9, 2020 from <https://towardsdatascience.com/phishing-domain-detection-with-ml-5be9c99293e5>

**Yi, P., et al.** (2018). *Web Phishing Detection Using a Deep Learning Framework.* Hindawi: Wireless Communications and Mobile Computing. Retrieved on April 9, 2020 from <https://doi.org/10.1155/2018/4678746>

**Mahajan, R., & Siddavatam, I.** (2018). *Phishing Website Detection using Machine Learning Algorithms.* International Journal of Computer Applications. Retrieved on April 9, 2020 from <https://www.researchgate.net/publication/328541785>

**KnowBe4.** (n.d.). *What is phishing?* Phishing | General Phishing Information and Prevention Tips. Retrieved on April 9,2020 from

<https://www.phishing.org/what-is-phishing>

***14 types of phishing attacks that IT administrators should watch for*.** (2020, March 27). SysCloud Blog. Retrieved on April 9,2020 from <https://blog.syscloud.com/types-of-phishing/>