Phishing Websites & Counter Measures

Sai Teja Veeramalla[1], 4 th Mohammed Raziuddin[1], Tathagata Bhattacharya[1], 5 th Mohammed Akram Ahmed[1], and Jaideep Nutakki[1]

[1]Auburn University at Montgomery

January 24, 2023

Abstract

Today's internet technology is so prevalent that it has improved the quality of people's lives in a variety of ways, such as online social networking and banking. Security risks to systems and networks are continually evolving because of the development of Internet technologies. Phishing is one such major issue, in which attackers try to steal the user's credentials by using bogus emails, websites, or both. The creation of solutions to counter phishing attacks is a top priority for both businesses and academics. Therefore, it is crucial for businesses to focus on end-user education while preventing phishing threats. Consequently, our paper's goal is demonstrate the different approaches to clone the webpage. We will start by going into great depth into the background of phishing attacks and the motivation of the perpetrators. Then, we will give a taxonomy of the many sorts of phishing.

*Note: how hackers create phishing websites.

1st Sai Teja Veeramalla

Computer Information System *Cybersecurity*

Auburn University at Montgomery

MONTGOMERY, USA

tbhatta1@aum.edu

4thMohammed Raziuddin Computer Science

Auburn University at Montgomery

MONTGOMERY, USA

mohamme@aum.edu

2nd Dr. Tathagata Bhattacharya Computer Information System *Cybersecurity*

Auburn University at Montgomery

MONTGOMERY, USA

tbhatta1@aum.edu

5th Mohammed Akram Ahmed Computer Science

Auburn University at Montgomery

MONTGOMERY, USA

Mlnu28@gmail.com

3rd Jaideep Nutakki

Computer Information System

cybersecurity

Auburn University at Montgomery

MONTGOMERY, USA

jnutakki@aum.edu

Abstract—

Today's internet technology is so prevalent that it has improved the quality of people's lives in a variety of ways, such as online social networking and banking. Security risks to systems and networks are continually evolving because of the development of Internet technologies. Phishing is one such major issue, in which attackers try to steal the user's credentials by using bogus emails, websites, or both. The creation of solutions to counter phishing attacks is a top priority for both businesses and academics. Therefore, it is crucial for businesses to focus on end-user education while preventing phishing threats. Consequently, our paper's goal is demonstrate the different approaches to clone the webpage. We will start by going into great depth into the background of phishing attacks and the motivation of the perpetrators. Then, we will give a taxonomy of the many sorts of phishing.

Keywords: Web Spoofing, URL Spoofing, Domain Spoofing, Click-Jacking, Spam Links.

I. INTRODUCTION

The term 'phishing' is coined in the mid-1990s and is from the term 'fishing' because it involves trying to outwit someone into a trap [11] Phishing is when attackers send malicious emails, suspicious URL designed to trick people into falling for a scam. Typically, the intent is to get users to reveal financial information, system credentials or other sensitive data.

Phishing is metaphorically like fishing in the water [12], but instead of trying to catch a fish, attackers try to steal consumer's personal information When a user opens a fake webpage and enters the username and protected password, the credentials of the user are acquired by the attacker which can be used for malicious purposes Phishing websites look very similar appearance to their corresponding legitimate websites to attract a large number of Internet users.

1.1 Social Engineering

A collection of techniques that scam artists use to manipulate human psychology. Social engineering techniques include forgery, misdirection, and lying all of which can play a part in phishing attacks. On a basic level, phishing emails use social engineering to encourage users to act without thinking things through. Estimates suggest that 59% of all email is spam Lexically analysing the URLs can enhance the performance and help to differentiate between the original email and the phishing URL [9] is a fraudulent attempt to obtain sensitive information such as usernames, passwords, and credit card details (and money) by disguising as trustworthy in an electronic communication [6] Phishing attacks use a combination of social engineering is the art of getting users to compromise information systems. Instead of technical attacks on systems, social engineers target humans with access to information, manipulating them into divulging confidential information or even into carrying out their malicious attack through influence and to persuade [10] users into giving away sensitive information that the attacker can used to make financial profit. Normally phishers hijack a banks web pages and send emails to the victim in order to trick the victim to visit the malicious site in order to collect the victim bank account information and card number [1] Instead of tricking the users in emails

to give out their passwords e.g., for online-banking and eBay, the attacker redirects the users to its servers, which imitate the original websites. For a normal user it's hard to tell if he is connected to the original site, as the right address is shown in the browser, and he relies on the correct name resolution. Even in case of additional security mechanisms, such as SSL, many users tend to discard warning messages or do not check for secured connection at all. Albeit there is no known case of pharming yet, it seems likely that future

1.1 DNS Cache Poisoning

Domain Name System uses different techniques to introduce false IP addresses to client servers. The use of cache memory is the most seen type, which is known as DNS cache poisoning [32] DNS cache poisoning exploits will be used in that way. Additionally, an incorrect DNS resolution has impact on protocols besides http that rely on the name resolution, such as ftp, pop3, Imap and smtp. Due to DNS Spoofing an attacker can redirect and record the login attempts and gather the data of the user's accounts [8]. Also, the sniffing of outgoing mails by redirection over a prepared server is possible

1.2.1 Types of Phishing Attacks

There are three types of phishing attacks through social engineering, which are web spoofing, phone phishing, spear phishing and clone phishing

1.2.2 Web Spoofing

allows an attacker to create a "shadow copy" of the entire World Wide Web. Accesses to the shadow Web are funnelled through the attacker's machine, allowing the attacker to monitor all the victim's activities. Attacks can be carried out on today's systems, endangering users of the most common Web browsers. Web spoofing allows an attacker to create a "shadow copy" of the entire World Wide Web [7]. Accesses to the shadow Web are funnelled through the attacker's machine, allowing the attacker to monitor all the victim's activities. A spoofing attack is like a con game: the attacker sets up a false but convincing world [3] around the victim Phone Phishing

1.2.3 Phone Phishing

Phone phishing is criminal activity [32] using the social engineering service often the use of telephone or mobile phone to accrue the sensitive or private information to make phishing financial profit, over 4000 cases of voice phishing are committed per year [4] and the cost per victim is over US$1000. The preparation for phishing includes getting ready for crime, recruiting telemarketers, and creating scripts. The next step involves randomly making international and Internet calls to many people.

Clone Phishing

In this case, the attacker is attempting to clone the online portal that often requests login credentials by imitating actual websites. He will also attempt to send the victim junk links via phishing emails. When the victim opens the phishing email and clicks on the spam link made by the attacker, it will redirect to a fake page made by the attacker when the victim is supposed to enter sensitive information like a user ID and password. This will allow the attacker to steal and save the credentials entered by the victim in a text file and database record on the attacker server, after which we will redirect the victim to the legitimate websites as an authenticated user.

1.2.5 Spear Phishing

A spear-phishing attack targeting a specific user may leverage information [33] such as his/her username and email address to craft an email that is personalized to the user. This spear phishing technique will certainly improve the success rate of the attack and techniques that can be leveraged by an attacker to find contextual information.

1.2.6 Whaling

This mainly targets high-profile employees of big organizations to excess highly confidential information [34]. It is also called CEO fraud, here hackers use social engineering to phish users to give away their bank credentials employee data, etc. These attacks are even difficult to detect as they do not use malware or fake websites

Impacts of Phishing Attacks

According to a study by Gartner, 51 million US Internet users have identified the receipt of e-mail linked to phishing scams and about 2 million of them are estimated to have been tricked into giving away sensitive information [31] Throughout the world, phishing attacks continue to evolve and gain momentum. In 2012, total phishing attacks increased by 160% over 2011, signifying a record year in phishing volumes. [14] In June 2018, the Anti-Phishing Working Group (APWG) reported as many as 51,401 unique phishing websites, another report by RSA estimated that global organizations suffered losses amounting to $10 billion due to phishing incidents in 2016 [13], These statistics have proven that the existing anti-phishing solutions and efforts are not truly effective. The most widely deployed anti- phishing solution is the blacklist warning system, found in conventional web browsers such as Google Chrome and Mozilla Firefox. The blacklist system queries a central database of already-known phishing URLs; thus, it is unable to detect newly launched phishing websites. Hillary Clinton presidential campaign chairman, John Podesta's Google email account was "hacked" in March 2016 prior to the US election [35]. The hacker simply sent a phishing email to Protester's Gmail account and lured him to disclose his login credentials. In the phishing email, Podesta had been invited to click on a link (i.e., Unified Resource Locator, so called "URL") warning him to change his password immediately. However, the URL did not link to a secure Google web page, instead directing the user blindly via bit.ly, which is a service used to shorten URLs. Podesta hack didn't require much technical skills. Instead, the hacker merely used social engineering techniques to make the attack successful the attack and techniques that can be leveraged by an attacker to find contextual information.

1.4 Contributions

First, we chose to work on a project about phishing scams. Throughout this journey, we for all intents and purposes have done a lot of research, which really is significant. We essentially have discovered that several phishing assaults basically take place generally worldwide in a big way. We generally have kind of decided to basically duplicate a website which basically looks similarly to the pretty original website, which kind of is significant. To begin with, we used the HTML and PHP languages to clone the website like Aum Student Login Portal and Initially, we mostly decided to for the most part create phishing page of Aum Student Login Portal using the source code of the for all intents and purposes original website of Aum portal to harvest the student login credentials.

Literature Survey

With the advent of Phishing webpages, researchers have investigated supervised and unsupervised learning models for detecting phishing webpages for instance, Moghimi, Mahmood, and Ali Yazdani Varjani [14]. support vector machine (SVM) algorithm to classify webpages. their experiments indicate that the proposed

model can detect phishing pages in internet banking with accuracy of 99.14% true positive and only 0.86% false negative alarm Afroz and Green Stadt [15] developed Phish Zoo technique this technique constructs a website profile using a fuzzy hashing approach in which the website is represented by several criteria that differentiate one website from another including images, HTML source code, URL, and SSL certificate. A. Desai, J. Jatakia, R. Naik, and N. Raul [16] created an extension to Google Chrome to detect phishing websites content with the help of machine learning algorithms,S. Parekh, D. Parikh, S. Kotak, and P. S. Sankhe [17] proposed a model with answer for recognizing phishing sites by utilizing URL identification strategy utilizing Random Forest algorithm, X. Zhang, Y. Zeng, X. Jin, Z. Yan, and G. Geng [18] proposed a phishing detection model to detect the phishing performance effectively by using mining the semantic features of word embedding, semantic feature and multi-scale statistical features in Chinese web pages, y Ma et al. [19], Zhang et al wrote Python scripts to automatically download confirmed phishing websites" URLs from PhishTank. PhishTank is a collaborative clearing house for data and information about phishing on the Internet Jeeva and Raj Singh [20] extracted features related to transport layer security together with URL based features such as length, number of slashes, number and positions of dots in URL and subdomain names. Rule mining was used to establish detection rules using the apriorist algorithm on the extracted features. Experimental results showed that 93% of phishing URLs were detected. Jain and Gupta [21] presents an anti- phishing approach, which uses machine learning by extracting 19 features in the client side to distinguish phishing websites from legitimate ones, Peng, Harris, and Sawa [22], NLP is applied to detect phishing emails. It performs a semantic analysis of the content of emails (as simple text) to detect malicious intent. Prakash, Kumar, Kompella, & Gupta, 2010 [23], These systems use an approximate matching algorithm to check whether the suspicious URL exists in the blacklist or not S. Aonzo, A. Merlo, G. Tavella, and Y. Frat Antonio, [24] represented the Multifactor Authentication technique uses two or more authentications to login into the accounts/systems. One is password and other is code generated by an app through SMS, phone calls or emails. By this method only authenticated person can login into his accounts Tech5(Machine Learning Approach, 60%) was identified as one of the most effective anti-phishing techniques, one of the early developed whitelist was proposed by Chen and Guo [25], which was based on users' browsing trusted websites. The whitelist monitors the user's login attempts and if a repeated login was successfully executed this method prompts the user to insert that website into the whitelist. One clear limitation of Chen and Guo's method is that it assumes that users are dealing with trustful websites, which unfortunately is not always the case. Zhang H, Liu G, Chow TWS, Liu W [26] presented a new framework for content-based phishing detection using a Bayesian approach. Selection Lee and Kim [27] proposed a suspicious URL detection system called WARNINGBIRD for Twitter. Li et al. [28] proposed a combination of linear/nonlinear domain conversion methods to represent the core problem more clearly and to improve the performance of classifiers in identifying malicious URLs Yang L, Zhang J, Wang X, Li Z, Li Z [29] presented a new approach to phishing detection based on an inverted matrix online sequential over-learning machine that takes into account three types of features to characterize a website. They used the Sherman Morrison Woodbury equation to reduce matrix inversion. They introduced the online queue extreme learning machine to update the training model. De La Torre Parra et al. [30] proposed a cloud-based distributed deep learning framework for phishing attack detection. Wu, et al [36] empirically investigated three simulated anti-phishing toolbars to determine how they were effective at securing participants from visiting fraudulent websites Bait Alarm [37] is comparatively more efficient as VSBPD compares the text and their style in two websites, Visual Similarity Based Phishing Detection (VSBPD) [38] gives a warning to the user whenever he tries gives his credentials to an untrusted website Google Safe Browsing API [39] allows the client side applications to check if a URL is blacklisted from a list which is continuously updated by Google, Juan Chen, and Chuanxiong Guo designed and developed Link Guard algorithm [40] to detect Spoofed hyperlinks in the phishing mails

Methodology:

There are numerous methods that have been used in the past to duplicate various websites, such as Facebook, Instagram, GitHub, etc., but all of these methods were only effective on websites that lacked form validation and were weak security. In this project, however, we implemented a new method that can replicate all types

of websites that have form validations and anti-click jackings.

3.1 Cloning the Original Website

The attacker set up everything necessary to create a replica of the original website using a fake webpage. He will then create a phishing email that contains a link to the fake webpage, so that when the victim enters data using the link, the data is immediately posted to the fake website rather than the legitimate one, and the attacker's database will be updated with the phishing information.

Fig.1 Attackers cloning the original website

Downloading the Source Code

Here in our experiment, we mimic the MY AUM portal to show how a Website Spoofing attack works. At the very beginning of our endeavor, we used the HTTrack software tool to retrieve the My AUM website's whole source code.

Fig.2 Win HTTrack website copier software

Creating a new PHP file:

We developed a POST.PHP file and added a few lines of PHP code to it in the second step to harvest the credentials from the victims that visited our cloned MY AUM website.

PHP Algorithm

START

STEP 1 : *opening a file called creds_log.txt*

STEP 2: *we use each variable and value submitted by users via POST requests made on our web page's form input fields*

STEP 3: *Creates an open handle for a text file called creds_log.txt*

STEP 4: *The function takes two parameters first*

It takes in a string and second, it takes to write out text **STEP 5:** *Exit*

END

Modifying The Html:

The action URL was changed to POST.PHP once the POST.PHP file was generated, and the amended index.html page was saved.

Fig 3. Modifying action URL to POST.PHP

HOSTING THE FILES:

Presently here is the practical part, where we facilitated all three records Index.html, POST.PHP, and AumUsers_Cred.txt to form our fake My AUM site online so other individuals can browse.

Fig.4 Hosting the Files

Results and Discussion:

We provide the victim with our URL, and each time the victim tries to log in using his username and password and hits the login button, the data is sent to our domain server rather than to MY AUM server.

Fig.5 Hosting Success

4.1 Storing Credentials of Victims:

The victims' login information was successfully collected and saved in the already-created empty file (Amusers_cred.txt).

Fig:6 Storing the Credentials

Errors Handling:

After conducting extensive research and conducting numerous experiments to get it successful, we managed to clone another web page, such as GitHub, using our techniques. Once we hosted the GitHub web page on our domain and server, we were able to achieve a success rate of "15%." Initially, when we try to clone the aum web page, the success rate is "0%" due to security features embedded in the aum web page that prevent us from cloning. We concluded that our method for effectively cloning a webpage only worked on other web pages rather than the original webpage.

The success rate in achieving the Aum Web Page is depicted in Fig. 7's graph.

5.1 Creating Duplicate Input Fields:

The login button initially does not trigger when we try to enter our ID and password in the input fields of a replica of the M: y AUM portal's homepage; nevertheless, when we try to enter our ID and password without the login button activating, it does. To solve the issue of not being able to trigger the login button when entering the user id and password, we decided to duplicate input fields underneath the original input fields of the user id and password. Unless the login button is triggered, we cannot post the user id and password details to our domain.

Fig. 8 Illustrating the Multiple Input Fields

Html Style Elements:

We chose to resize the original input fields in the HTML file to make it more realistic, like the original My aum web page, thus we added a few lines of style elements to the original input fields.

Figure 9: Exhibits Adding Additional Style Elements Figure

Hidden Input Fields:

We were able to correctly conceal the empty fields after successfully adding the extra style components to the original input fields.

Successfully concealing the empty input fields is shown in Fig. 10.

6 Conclusion

Attacks using phishing are still one of the biggest risks to people and businesses today. This is mostly driven by human engagement in the phishing cycle, as was mentioned in the article. Phishers often prey on human weaknesses in addition to promoting favorable technology settings (i.e., technical vulnerabilities). Age, gender, internet addiction, user stress, and many other characteristics have been found to affect a person's vulnerability to phishing. Along with more established phishing channels (like email and the web), newer phishing mediums like phone and SMS phishing are becoming more popular. Along with the expansion of social media, phishing on social media has also become increasingly prevalent. Concurrently, phishing has evolved beyond financial crimes and collecting sensitive information to include cyberterrorism, hacktivism,

REFERENCES

1. Steinhoff, U., A. Wiesmaier, and R. Araújo. "The state of the art in DNS spoofing." Proc. 4th Intl. Conf. Applied Cryptography and Network Security (ACNS). 2006.
2. Ramzan, Zulfikar, and Candid Wüest. "Phishing Attacks: Analyzing Trends in 2006." CEAS. 2007.
3. Felten, Edward W., et al. "Web spoofing: An internet con game."

Software World 28.2 (1997): 6-8

1. Choi, Kwan, Ju-lak Lee, and Yong-tae Chun. "Voice phishing fraud and its modus operandi." Security Journal 30.2 (2017): 454-466.
2. Bhavsar, Vaishnavi, Aditya Kadlak, and Shabnam Sharma. "Study on phishing attacks." Int. J. Comput. Appl 182 (2018): 27-29
3. Banu, M. Nazreen, and S. Munawara Banu. "A comprehensive study of phishing attacks." International Journal of Computer Science and Information Technologies 4.6 (2013): 783-786
4. Yadav, Surendra, and Brahmdutt Bohra. "A review on recent phishing attacks on the Internet." 2015 International Conference on Green Computing and Internet of Things (ICGCIoT). IEEE, 201
5. Peltier, Thomas R. "Social engineering: Concepts and solutions. "Information Security Journal 15.5 (2006):
6. Chiang Yung-chen. Social engineering and the social sciences in China, 1919-1949. Cambridge University Press, 2001
7. Krombholz, Katharina, et al. "Advanced social engineering attacks." Journal of Information Security and applications 22 (2015): 113-122
8. Vijayalakshmi, M., et al. "Web phishing detection techniques: a survey on the state-of-the-art, taxonomy and future directions." Iet Networks 9.5 (2020): 235-246.
9. Ankit Kumar, and Brij B. Gupta. "Phishing detection: analysis of visual similarity-based approaches." Security and Communication Networks 2017 (2017)
10. Chiew, Kang Leng, et al. "A new hybrid ensemble feature selection framework for machine learning-based phishing detection system." Information Sciences 484 (2019): 153-16
11. Moghimi, Mahmood, and Ali Yazdani Varjani. "New rule-based phishing detection method." Expert systems with applications 53 (2016): 231 -242.
12. Qabajeh, Issa, Fadi Thabtah, and Francisco Chiclana. "A recent review of conventional vs. automated cybersecurity anti-phishing techniques." Computer Science Review 29 (2018): 44-55.
13. A. Desai, J. Jatakia, R. Naik, and N. Raul, "Malicious web content detection using machine leaning," RTEICT 2017 - 2nd IEEE Int. Conf. Recent Trends Electron. Inf. Commun. Technol. Proc., vol. 2018–Janua, pp. 1432–1436, 2018.
14. S. Parekh, D. Parikh, S. Kotak, and P. S. Sankhe, "A New Method for Detection of Phishing Websites: URL Detection," in 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), 2018, vol. 0, no. Icicct, pp. 949–952
15. X. Zhang, Y. Zeng, X. Jin, Z. Yan, and G. Geng, "Boosting the Phishing Detection Performance by Semantic Analysis," 2017.
16. J. Ma, L.K. Saul, S. Savage, G.M. Voelker, beyond block lists: Learning to detect malicious web sites from suspicious URLs, In: Proc.15th ACM SIGKDD Int.Conf. Knowledge Discovery and Data Mining, Paris, France, 2009, pp. 1245-1254.
17. S.C. Jeeva, E.B. Raj Singh Intelligent phishing URL detection using association rule mining Human-centric Computing and Information Sciences, 6 (1) (2016)
18. A.K. Jain, B.B. Gupta towards detection of phishing websites on client-side using machine learning based approach telecommunication Systems, 68 (4) (2018), p. 687
19. T. Peng, I. Harris, Y. Sawa detecting phishing attacks using natural language processing and machine learning IEEE 12th international conference on semantic computing (ICSC), Laguna Hills, CA (2018),

pp. 300-301

1. P. Prakash, M. Kumar, R.R. Kompella, M. Gupta Phish net: Predictive block-listing to detect phishing attacks 2010 Proceedings IEEE INFOCOM (2010), pp. 1-5
2. S. Aonzo, A. Merlo, G. Tavella, and Y. Fratantonio, "Phishing attacks on modern android," Proc. ACM Conf. Comput. Commun. Secur., pp. 1788–1801, 2018, Doi: 10.1145/3243734.3243778.
3. Chen J., Guo C. Online detection, and prevention of phishing attacks (Invited Paper) First International Conference on Communications and Networking in China, ChinaCom '06, IEEE, Beijing (2006)
4. Zhang H, Liu G, Chow TWS, Liu W (2011) Textual and visual content-based anti-phishing: an ap-

proach. IEEE Trans Neural Newt 22(10):1532–1546
5. Lee S, Kim J (2013) Warningbird: a near real-time detection
6. system for suspicious URLs in twitter stream. IEEE Trans Depend Secure Comput 10(3):183–195
7. Li T, Kou G, Peng Y (2020b) Improving malicious URLs detection via feature engineering: Linear and nonlinear space transformation methods. Inf Syst 91:101494
8. Yang L, Zhang J, Wang X, Li Z, Li Z, He Y (2021) An improved elmbased and data preprocessing integrated approach for phishing detection
9. De La Torre Parra G, Rad P, Choo KKR, Beebe N (2020) Detecting internet of things attacks using distributed deep learning. J Netw Comput Appl 163:102662
10. Kirda, Engin, and Christopher Kruegel. "Protecting users against phishing attacks with antiphish." *29th Annual International Computer Software and Applications Conference (COMPSAC'05)* . Vol. 1. IEEE, 2005.
11. Kumar, Gaurav. "Best plan to protect against phone phishing attack." *American Journal of Computer Science and Information Technology*3.5 (2015): 167-172.
12. Chuenchujit, T. 2016. A taxonomy of phishing research, University of Illinoi
13. J. Hong, "The State of Phishing Attacks
14. Krieg, G. And Kopan, T. 2016. CNN News, is this the email that
15. Herzberg, Amir, and Ahmad Garba, "Trust bar: Protecting (Even naive) web users from spoofing and phishing attacks." Computer Science Department Bar Ilan University, PP 1-28, Jul 2004.
16. Jian Mao, Pei Li, Kun Li, Tao Wei, and Zhenkai Liang, "Bait AlarmDetecting Phishing Sites Using Similarity in Fundamental Visual Features," Intelligent Networking and Collaborative Systems (INCoS), 2013 5th International Conference on Intelligent Networking and Collaborative System, PP. 790-795, 2013, Xi'an.
17. Eric Medvet, Engin Kirda, and Christopher Kruegel, "Visual- Similarity-Based Phishing Detection," SecureComm '08 Proceedings of the fourth international conference on Security and privacy in communication networks, Article no 2, PP. 1-11, 2008.
18. Safe Browsing API – Google Developer, available at: https://developers.google.com/safe–browsing/
19. Juan Chen, and Chuanxiong Guo, "Online Detection and Prevention of Phishing Attacks," Communications a