World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:17, No:10, 2023

Phishing Attacks Facilitated by Open-Source Intelligence

Urva Maryam

***Abstract*—**Private data are more often breached by clever social engineering rather than exploiting technical vulnerabilities in the systems. Complete information security requires good data safety practices to go along with technical solutions. Hackers often begin their operation by simply sending spoofed emails or fraudulent URLs to their targets and trick them into providing sensitive information such as passwords or bank account details. This technique is called phishing. Phishing attacks can be launched on email addresses, open ports and unsecured web browsers. This study uses quantitative method of research to execute phishing experiments on the participants to test their response to the phishing emails. These experiments were run on Kali Linux distribution which came bundled with multiple open-source intelligence (OSINT) tools that were used in the study. The aim of this research is to see how successful phishing attacks can be launched using OSINT and to test the response of people to spoofed emails.

***Keywords*—**OSINT, phishing, spear phishing, email spoofing, theHarvester, Maltego.

I. Introduction

THE world has revolutionized into an era of technological advancements. With new technology comes greater risk. As information has become an important asset to the world, there have been developments to exploit it and use it for nefarious purposes. Open-source intelligence has become a touchstone of breaching process. OSINT is a hacker's first asset in targeted attacks. It comes under social engineering where the attacker uses free platform to extract useful information about the target. There are many tools and search engines which make such information available. These tools and search engines perform a deep search on targets. People have social media accounts such as LinkedIn that contain their information i.e., email addresses (work and personal). Social media tends to play a major role in enquiring about an individual or organization. Also, social media users openly share their information publicly which makes it easier for the attacker to find the entering points. Data show that 66% of out-clicks on an individual's search engine leads to social media profiles [1]. Studies related to web searches reveal that between 4% [2] to 10% [3] of searches contain the name of a person indicating that there is a personal interest inhibited in the searches. Information discovered from these resources may be used in illegitimate activities such as stalking, stealing, identity theft, financial and identity frauds. On the contrary, OSINT can be helpful in diagnosing and detecting weak ends of a network to safeguard existing vulnerabilities. The activities involved in gathering and correlating such information through the use of publicly available tools is called OSINT. OSINT uses information "openly available to all" [4]. The work of collecting and analyzing information is not a recent phenomenon, it has been in progress for years, some more definitions of OSINT can be given to widen the understanding of this term:

As defined by NATO, "OSINT is intelligence derived from publicly available information, as well as other unclassified information that has limited public distribution or access." [5]

> "Unclassified information that has been deliberately discovered, discriminated, distilled and disseminated to select audience to address a specific question" [6].
>
> "Open-source intelligence (OSINT) is intelligence that is produced from publicly available information and is collected, exploited, and disseminated in a timely manner to an appropriate audience for the purpose of addressing a specific intelligence requirement" US-DoD (Department of Defense) [7].
>
> "Collect, Process and correlate public information from open data sources such as the media, social networks, forums and blogs, public government data, publications or commercial data" [8].

A potential attacker uses social engineering techniques for network intrusion. This is done by gathering intelligence from a diverse spectrum of freely available online software and sources and this can be processed to launch a phishing attack. Even government and military intelligence agencies use OSINT to launch clandestine phishing attacks. Phishing has been a major concern for the cyber industry.

Attackers adopt new methodologies to exploit a system or a network. With the help of latest techniques, it is convenient for attackers to conduct fraudulent activities which could result in immense financial losses and damage the reputation of the target organization. Phishing is a cybercrime in which attackers send fraudulent emails or text messages, or set up fake websites, in an attempt to steal sensitive information such as login credentials, financial information, and other personal data. These attacks often use social engineering techniques to trick victims into believing that the message or website is legitimate. The term "phishing" is derived from the fact that the attackers are essentially "fishing" for sensitive information, using bait in the form of fake emails, websites, or messages. The word "phishing" is often spelled using leetspeak, which is a form of written slang that replaces letters with numbers or special

Urva Maryam is with School of Electrical Engineering and Computer Sciences, Islamabad, Pakistan (e-mail: urvam95@gmail.com).

characters [9]. As of 2018, Symantec found that the rate of email-based phishing attacks has decreased since 2018, but it is important to note that phishing is still a major threat and continues to evolve [10], [11]. Of all the phishing attacks, spear phishing is the most common infection vector used in distribution of malware, used by 71% of groups in 2018 and 65% of groups in 2019 [12]. Despite the presence of Anti-Phishing Working Groups (APWG), the presence of phishing attacks has increased dramatically. The purpose of this research is to produce quantitative evidence of work performed to indicate how OSINT facilitates phishing attacks.

The main contributions of this paper are as follows:

- A set of indicators are proposed to support and justify that the research has been used for awareness and educational purposes.
- The research was conducted with a complete documentation and useful description of the tools used, ensuring transparency and reproducibility of the findings.
- A primary objective of this research is to educate users about the susceptibility of phishing attacks which can be launched against them using OSINT tools.

Briefly, this study contributes to use the OSINT ecosystem to produce details about individuals using intelligence tools and target them in a phishing attack precisely in email spoofing. This research applies social engineering by employing a phishing technique applied on individuals with the help of information retrieved from multiple OSINT tools. This will increase the understanding of the general public regarding their publicly available information and the consequences that publicly exposed information inherits.

For this study, quantitative research methodology was opted using a hands-on experience. All aspects of this study were conducted in an ethical manner. The consent of the participants was taken before conducting the study. Section II consists of the related work to our study. Section III comprises of the study methodology being used in the research, the equipment and software, and the detailed documentation of the experiment. Section IV consists of the results obtained from the study and the analysis built from it. Lastly, all the references have been mentioned.

II. Related Work

OSINT resource has advanced past its peripherals in recent years. It has become an important part of social engineering. OSINT collects publicly available data with tools such as Maltego, Google Dorks, Shodan, theHarvester, Nmap and many others. A significant amount of research work has been conducted in this area, as discussed in the Subsections *A, B,* and *C,* comprising quantitative, qualitative and mixed methods. In recent years, phishing attacks have also turned lethal in cyber-world producing malignant content to breach security of multifarious platforms. These attacks are either directed using emails, contracted from OSINT tools, embedded with viruses, malwares or spywares or use phishing URLs which prompt their targets to enter confidential information. There have been several works done in the field of social engineering, a few of the studies are highlighted below:

A. Visual Perspective Analysis

The study, Visual Perspective Analysis" [13], conducted by Francined et al. focuses on quantitative method of research. It analyzed two OSINT material sources: the research dissemination databases and educational resources repositories. The study focuses on the final stages of the OSINT process, emphasizing the need to effectively utilize the information gathered for assertive purposes, given that OSINT represents a significant advancement in security and defense. Ubuntu Linux was used to perform web scraping to extract data from web pages. The data were classified into three categories: the first represented the sources that did not have published OSINT resources, the second represented the sources that publish OSINT without metadata, and the third category represented the sources that published OSINT with metadata. Based on these three categories, OSINT subareas were identified and analyses was made [13].

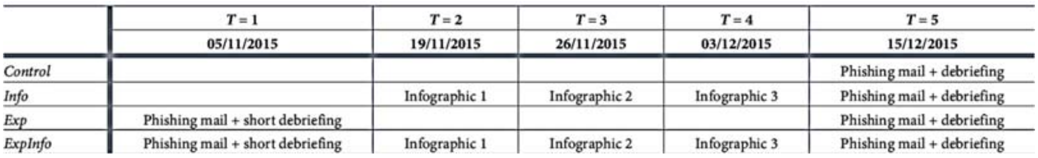
B. Designing and Conducting Phishing Experiments

Phishing experiments in the study "Designing Ethical Phishing Experiments" were designed and conducted by Finn and Jakobson [14]. They conducted three phishing experiments: Social phishing, a study of eBay query features, and Man in the Middle attacks.

In the first experiment, email addresses of users were collected and later used to send spoof emails which appeared to be sent by their friend but rather it was sent by the researchers. Users were sent a phishing URL in the spoof email, when they clicked the link, they were redirected to a phishing page which prompt them to enter their credentials. All the experiments were conducted in a legal and ethical way, and later the subjects were debriefed about the experiment. The second experiment was conducted on eBay users to assess the vulnerability of HTML markups. Spoof emails were sent to users which appeared to have been sent by eBay. The success rate of this experiment was 7% to 11% (depending on the conditions). The last experiment was conducted to study the vulnerabilities of MITM (Man-in the-middle) attacks. In this attack, users received spoof emails appearing to have been sent by eBay which was embedded with a URL. Users who clicked on URL were linked to a site acting as MITM attacker. The researchers analyzed the study to conclude that individuals who had meagre knowledge of security could not comprehend non-secure channels [14].

C. Informing, Simulating Experience, or Both: A Field Experiment on Phishing Risks

In *Informing, Simulating Experience, or Both: A Field Experiment on Phishing Risks* [15], Baillon et al. have adopted quantitative approach to test phishing risks. The study was conducted on the members of the Dutch Ministry. Subjects were debriefed after the experiment was conducted to ensure the legality of this study in the boundary of ethics. Subjects were sent phishing emails in different patterns. The experiment was conducted in 5 days on different groups. The groups were formed based on the emails that were sent to them. Fig. 1 describes the grouping:



	T = 1	*T* = 2	*T* = 3	*T* = 4	*T* = 5
	05/11/2015	19/11/2015	26/11/2015	03/12/2015	15/12/2015
Control					Phishing mail + debriefing
Info		Infographic 1	Infographic 2	Infographic 3	Phishing mail + debriefing
Exp	Phishing mail + short debriefing				Phishing mail + debriefing
ExpInfo	Phishing mail + short debriefing	Infographic 1	Infographic 2	Infographic 3	Phishing mail + debriefing

Fig. 1 Experimental Timeline [15]

The analysis was done based on three responses on users' end: Visit, Fill, Fill|Visit. The essence of collecting these responses is rooted in the potential danger posed to users, even if they simply click on a phishing link [15].

III. Research Methodology

A. Quantitative Research

The objective of performing a quantitative research is finding a relationship between independent variables and other dependent outcome variables. The information is gathered through an experiment or a survey. Then, the result is analyzed to test the strength of the hypothesis given. Quantitative research focuses on numbers and logic to determine a relation between a theory and empiricism. This study focuses on experimental analysis of phishing techniques which are conducted with the help of information collected from OSINT tools.

B. Software and Environment Setup

The entire research and testing were performed in an isolated virtual environment. This helps in isolating bugs if there are any and enables the control of all variables in the investigation being conducted. The virtual environment was created using VMware 16.0 which was run on a Windows 10 PC. The Linux distro Kali 2020.4 64-bit was the OS of choice for this research. It is a Debian based Linux which came bundled with many penetration testing tools that were invaluable during this research.

C. Gathering Emails

theHarvester was the tool used to gather basic publicly available information. It is an open-source tool that comes bundled with Kali as a part of its penetration testing suite. LinkedIn was the source to test this tool and it provided the emails it found which was convenient.

The next step was to use *Maltego* to map all of these emails to domains other than LinkedIn to see how much information can be gathered just by having email addresses of a few subjects. Maltego is a very powerful OSINT tool that helps put the entire first sweep of theHarvester into a broader perspective and can answer questions such as how many domains does a subject use with the same email address and one can go on and do a very thorough search for an individual subject using this. Even questions related to what percentage of LinkedIn users use a professional address separate from the email addresses that they use on other social media accounts, or if they keep them separate entirely. The first step of gathering just the emails was done again using Maltego on other social media websites and the results returned were much more than just an email: it had emails, pictures among other personal information from these websites.

Now we have our dataset of emails and other information available to conduct the next part of the experiment.

D. Spoof Email Phishing

The email experiment was divided into two parts between 20 subjects (10 per part): the first where the subject is simply redirected to another link to simulate a phished website and the second in which the user actually downloads a file.

For the first one, a webhook was created using *SendInBlue*, which is a marketing service for online advertisement campaigns that proved helpful for this experiment. The webhook provided feedback on how many people actually opened the link and went to the linked URL.

The second case was a bit more complex to test, a Google Drive plugin *OrangeDox***,** when enabled, provided insights into how many people downloaded a file, plus provided even more insights such as time of download, the browser used and more.

For the phished link part of the test, an open-source software the *EmBomber* was used. It is a Python software that can send mass emails, to bypass traditional email spam protections only five emails per subject were sent containing the spoofed webhook URL.

emkei.cz was the other web tool used to send the simulated malicious file link. It is a web-based tool that can be used to quickly send emails using a spoofed email address. *Browse Sec VPN* was used to access this website as such websites operate secretly and do not want legitimate Public IPs to access them. The link to the downloadable simulated malicious file was sent using this.

IV. Results

There was a total of 20 subjects to whom these emails were sent and they were divided into two groups depending on the type of phished link they were sent. Half were sent a webhook link and half were sent a link to a downloadable file that was being tracked to gather data.

For the first group, three out of 10 subjects clicked the link in the email to go to the simulated phished URL, and in the other group, two out of the 10 participants actually downloaded the file linked in the email.

The main goal of this research was to see how much information can be gathered using OSINT and if that can be used to perform a successful phishing attack. It was demonstrated in this experiment that by using simple tools, a lot of information can be gathered and it can be used to perform a successful phishing attack.

V. Conclusion

There is disseminated publicly available information out there online that can easily be scraped using OSINT tools and then can be easily be exploited for a variety of email-based attacks or even more. It has been demonstrated in this experiment that phishing attacks can be performed without any hassle while spoofing the source and people can easily fall prey to these tactics.

If a hacker was inclined to, he or she may go to more elaborate lengths to have higher chances of success by tracking across the target across social media domains and pandering to their interests in the email. The target would likely be even more likely to fall prey to such a phished email attack especially if it is also spoofed to look like a legitimate source.

Acknowledgment

The author would like to acknowledge the participants who accommodated them in performing the experiment and their professors for motivating and educating them to conduct a study and write a research paper.

References

[1] Maria-Hendrike Peetz, Edgar Meij, Maarten de Rijke, and Wouter Weerkamp. 2012. “Adaptive temporal query modeling”. *In Proceedings of the 34th European conference on Advances in Information Retrieval (ECIR'12).* Springer-Verlag, Berlin, Heidelberg, 455–458. https://doi.org/10.1007/978-3-642-28997-2_40

[2] A. Spink, B. J. Jansen, and J. Pedersen. “Searching for people on web search engines”. J*ournal of Documentation, 60(3):266–278, 2004.* doi: 10.1108/00220410410534176

[3] Guha, R. "Disambiguating people in search." *In The Thirteenth International World Wide Web Conference, WWW2004.* 2004.

[4] R. A. Norton, “Guide to Open Source Intelligence”. *Intell. J. US Intell.* Stud. 2011, 18, 65–67.

[5] NATO. “Open Source Intelligence Handbook”; *North Atlantic Treaty Organization*: Brussels, Belgium, 2001.

[6] Korkisch, F. NATO “Gets Better Intelligence; Center for Foreign and Defense Policy”: Vienna, Austria, 2010.

[7] U.S. Government Publishing Office, “Responsibilities of Secretary of Defense pertaining to National Intelligence Program”, 2006,

[8] Pastor-Galindo, J.; Nespoli, P.; Gomez Marmol, F.; Martinez Perez, G. “The not yet exploited goldmine of OSINT: Opportunities, open challenges and future trends”. *IEEE Access 2020, 8, 10282–10304.*

[9] Alabdan, R. “Phishing Attacks Survey: Types, Vectors, and Technical Approaches”. *Future Internet* 2020, *12*, 168. https://doi.org/10.3390/fi12100168

[10] Symantec. “ISTR Internet Security Threat Report 2019”. *Symantec 2019*, 24, 61.

[11] Symantec. “ISTR Internet Security Threat Report 2015”. *Symantec 2015*, 20.

[12] Symantec. “ISTR Internet Security Threat Report 2018” Volume 23. *2018.*

[13] J. F. Herrera-Cubides, P. A. Gaona-García, and S. Sánchez-Alonso, “Open-Source Intelligence Educational Resources: A Visual Perspective Analysis,” *Applied Sciences*, vol. 10, no. 21, p. 7617, Oct. 2020, doi: 10.3390/app10217617

[14] P. Finn and M. Jakobsson, "Designing ethical phishing experiments," in *IEEE Technology and Society Magazine*, vol. 26, no. 1, pp. 46-58, Spring 2007, doi: 10.1109/MTAS.2007.335565.

[15] Baillon A, de Bruin J, Emirmahmutoglu A, van de Veer E, van Dijk B. “Informing, simulating experience, or both: A field experiment on phishing risks”. *PLoS One*. 2019;14(12):e0224216. Published 2019 Dec 18. doi:10.1371/journal.pone.0224216