Accessing LinkedIn and Google E-mail Databases Using Kali Linux and TheHarvester

Zornitsa Terneva[1], Ivaylo Vladimirov[2] and Desislava Nikolova[3]

Abstract – **In this scientific paper, a test is made on how reliable and secure databases are. The simulation shows how easily hackers can access a database and use its contents maliciously. It is performed on a virtual machine Kali Linux using software program - TheHarvester. People's personal email addresses have been used and shown for educational purposes only. The aim is to analyse the accessibility of data collections.**

Keywords – **Databases, Kali Linux, TheHarvester, Phishing attacks, e-mails.**

I. Introduction

Phishing is a sort of social engineering attack that is frequently used to obtain personal information, such as login credentials and banking details (credit card information). [1]

Phishing attempts often begin with an email attempting to collect sensitive information by user engagement, such as clicking on a malicious link or downloading an infected file. Email fraud has increased by more than 400% in the last few years. The growth and success of email phishing have also led to the branching of the method.[2]

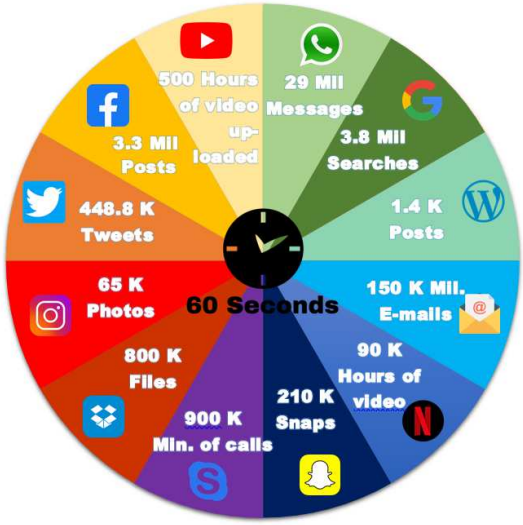

Fig. 1 What happens on the Internet in 60 seconds.

[1]Zornitsa Terneva is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria. E-mail: zterneva@tu-sofia.bg

[2]Ivaylo Vladimirov is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: ivladimirov@tu-sofia.bg.

[3]Desislava Nikolova is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria. E-mail: dnikolova@tu-sofia.bg

The risk of cyber-attacks is extremely high and even one infected file can create huge headaches in the cyber world today. Nowadays, the Internet is constantly used and millions of files are sent, processed and stored in just one minute.[3] In Figure 1 you can see the amount of information that is being transferred in just 60 seconds.

II. The Problem

The phrase "social engineering" refers to a wide range of harmful behaviors carried out via interpersonal behavior. It employs psychological manipulation and deception to lure users into committing security errors or disclosing sensitive confidential information. The fact that social engineering depends on human error rather than software and operating system flaws makes it incredibly dangerous and problematic. [4]

The phishing attack is one of the most common forms of social engineering cyber-attacks. This type of cyber-attack usually is contained in a fraudulent emails, sent by hackers and received by users/victims, consisting of text messages that instill in their recipients a feeling of anxiousness, curiosity, nervousness, excitement, or even terror and panic. The victim is then requested and encouraged to provide private information by clicking on links that lead to harmful and unsecured websites, or downloading and opening attachments containing malware software.[1]

Today, phishing attacks usually use generalized as "lures". Usually, hackers aim to gain access to credentials that give access to sites containing credit card or bank information. The most used method is by impersonating the financial institution they use or different subscription services like PayPal, Google Play Store, Apple App Store, Netflix, Facebook, Twitch, OnlyFans and many more. [2]

In a Gartner survey [5], about 19% of all respondents reported that they have clicked on a hyperlink in an e-mail, containing phishing cyber-attacks and slightly more than 3% admitted to disclosing personal and business information such as passwords. However, none of the available studies provides a comparison between the baseline success rate for different types of phishing attacks. [6]

As an example of such an attack, you can see in Figure 2. It is an e-mail sent to a private e-mail of mine: z*******@abv.bg consisting of a phishing link. The attacker/hacker impersonalises himself as the support/help center of the online financial service PayPal by masking his e-mail address and copying the visual style of an original e-mail. He is also using a generic style of text with the combination of a worrying message to make his potential

978-1-6654-2887-3/21/$31.00 ©2021 European Union

victim feel anxious and to think about what the problem with his account is and not question the authenticity of the received e-mail.

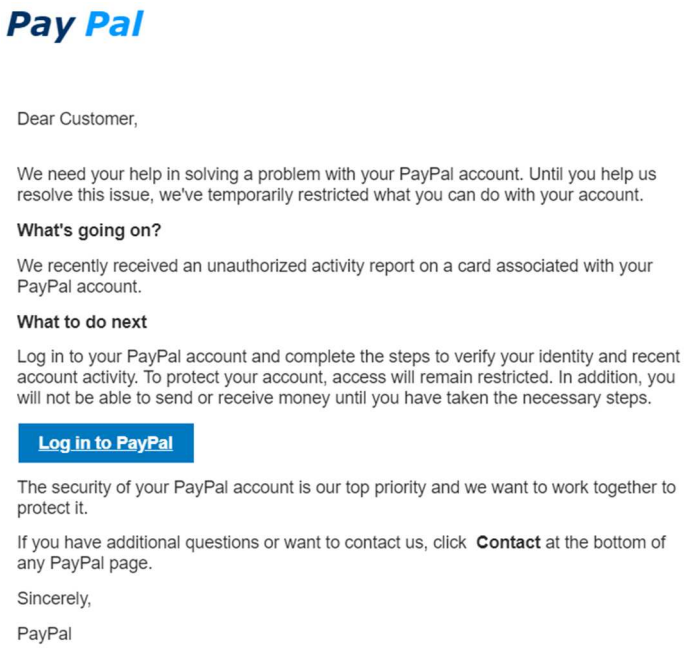

Fig 2. Example message containing a phishing attack link

Almost all legitimate emails from businesses to their clients contain information that phishing hackers cannot simply access. Some corporations, such as PayPal, always address their clients personally in emails by using their username. Therefore if you receive an email in which the recipient is addressed as "Dear Customer", "Dear Sir/Madam" , etc., it probably is an attempt of a phishing attack.[7]

III. The Simulation

For the purposes of this study, a simulation is made. The aim of this simulation is to show the ease at which an e-mail database can be collected. For this test a personal computer - ASUS K550JX-DM014D - with the following characteristics has been used:

- Processor: INTEL CORE i7-4720HQ;
- RAM: 12 GB;
- Graphics card: NVIDIA GEFORCE GTX 950M;
- Operational System: Windows 10;
- Type: Laptop

In order to perform the simulation, the installation of a virtual machine is required. The software used for the test is called "TheHarvester". In Figure 3 the interface of "TheHarvester" can be seen.

It is a simple open source tool for accessing email addresses, domains and subdomains, URLs and IP addresses, intended for use in the early stages of penetration testing. The program can be installed on the Kali Linux OS. [8]

This tool has two work modes:

- Passive Mode – In this mode, the source from which the information will be collected can be specified.
- Active Mode – When the mode is active, it allows a DNS call-back to be made for all open ranges (-n), DNS brute force for the domain name (-c) or DNS TLD extension detection (-t). For example, it is used to find information about users on different social media, (for example LinkedIn, Facebook, Twitter), to access e-mail databases, etc. [9]

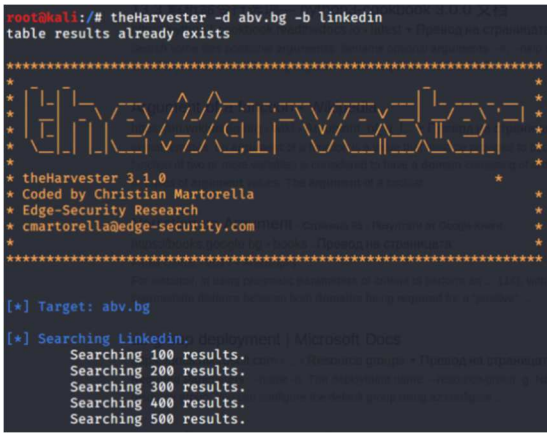

Fig 3. Access to the Linkedin database

TheHarvester increases the high percentage of robotic internet traffic.[10] This software program is a part of the OSINT group of tools, which are meant to be used in a Linux virtual machine. The hackers can use random OSINT tools depending on their target.[11] The reasons for using TheHarvester software may be different. The hacker may want to access (steal) a specific account (bank account, Netflix account, etc.), revenge, access to the corporate network (see. Figure 5), etc. Such social engineering attacks require only the victim's email address and sometimes a mobile phone. [12]

An example of a complete batch terminal command is:

- theharvester -d polito.it -l 600 -b google -f google-result.html

With this command, hackers can collect 600 Google email addresses in just a few minutes, which can then be specifically targeted for later attacks.

The simulation has the following steps:

- Step ONE: Preparation

Make sure that the used PC has the Kali Linux OS. Hackers use virtual OS in order to mask their traces easier. The VMWare workstation is a good alternative. It can be downloaded from their official site:

  - https://www.vmware.com/products/workstation-player/workstation-player-evaluation.html

After that Kali Linux is needed to be downloaded and installed from the official site:

  - https://www.kali.org/downloads/

- Step TWO: TheHarvester installation

To download and install "TheHarvester" type in the following command in the Linux basch terminal. This will install software and other packages and tools on which it depends.

  - sudo apt-get install theharvester

- Step THREE: Collection of data

To collect all available data from a social media or database (in this case LinkedIn) ending in a certain string (In this case abv.bg – a widely used Bulgarian e-mail service provider) this command line has been used:

- theharvester -d abv.bg -b linkedin

TheHarvester program provides easy access to many databases containing sensitive user information, such as e-mails, usernames, etc. "TheHarvester" is a great Python script tool to use for reconnaissance, it is simple but highly effective and its developers make regular updates to make sure the software is up to date with all search engines.[13]

The result of this simulation can be seen in Figure 4, which contains some of the extruded e-mail addresses. This is an easy and fast way to collect personal data. It took only a few minutes for the while extraction.

Fig 4. Result of the extraction simulation

IV. COUNTERMEASURES

Hackers attack people from a particular company to gain access to the company's database. The graph in Figure 5 presents the assets of a company and the path a hacker can try to penetrate to get to the most wanted data stored in the company's database servers.[14] The most vulnerable employees of the chain are:

- New employees.
- Staff not from the IT department or engineering department (accountants, human resources, etc.);
- Dissatisfied employees;
- CEO.

The choice to concentrate their cyber-attacks on new employees is not random. They are the most vulnerable unit in any company. Periodic training is extremely important so that hackers do not gain access to the company's resources.[15]

Most businesses strive to educate their employees/users/customers on how to spot fake emails in order to keep them from being easily cyber hacked. They also encourage people to forward questionable and problematic emails to specific domains that deal with similar issues. PayPal, for example, has a dedicated e-mail address - "spoof@PayPal.com." The department's, behind it, job is to investigate reported cases and alert customers of existing threats.[16]

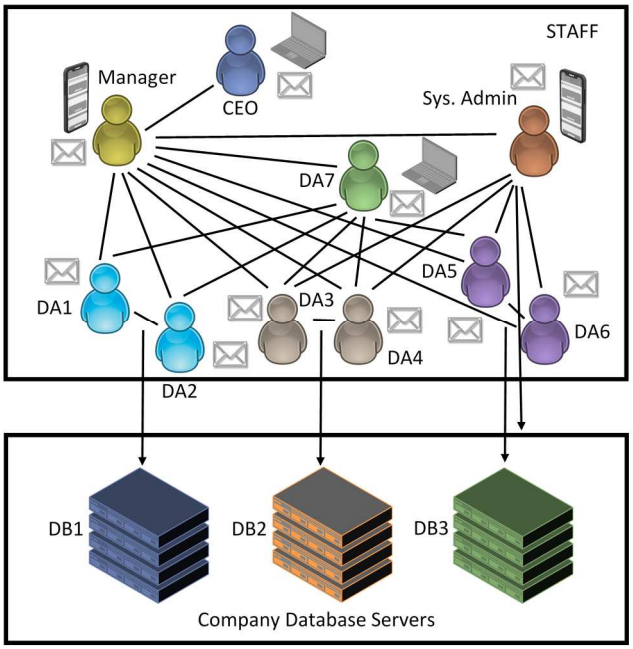

Fig 5. Structure of the communication in a company

However, it is dangerous to presume that the presence of personal user information alone guarantees that the correspondence is authentic. According to certain research, the appearance of personal customer details has no substantial effect on the probability of success of phishing assails implying that most users do not pay close attention to these kinds of elements. [17]

To reduce security threats, any organization must include a security policy. Authentication is critical in security policy. Each user has individual access privileges to various database items. The management of access privileges is handled by authentication techniques. This is the most used method for protecting objects of data and databases. One of the most important services that every data management system must provide is access control. Unauthorised read and write actions on the data must be prevented. Controlling access privileges can also assist to limit the dangers that might compromise the database's security on the primary servers.[18]

Access control systems include:

- File permissions – They allow a user/software to create, view, modify or remove a file from a server.
- Permissions for programs – They give software the ability to run a process on an application server.
- Data rights - They allow a user/software the ability to extract or re-enter/modify information in a data server.

In order to access a data server, the specific user must have the needed authorisation and must identify himself first. The OS or the network service can provide external verification

for users. The Secure Socket Layer can also be used to implement access control.[12][18]

Another way to countermeasure these attack is by using some kind of encryption, where the transferred information is converted into a cipher - code. Thus, the information cannot be accessed by anyone except for individuals who have the cipher-key to decode the text. This is one of the main methods of protecting data from unauthorized users (hackers).

V. Conclusion

Social engineering attacks are common and it is very easy for a user to be successfully misled and thus to have sensitive information stolen from them, lose money, etc. Many attacks are successful because the victims are not well enough informed.

As you can see in this article, accessing email databases is extremely easy. Linkedin has been specially selected for the purposes of this simulation. In recent years, it has become extremely popular around the world and is one of the main targets for accessing user emails among hackers themselves.

Acknowledgement

This scientific document is conducted under the grant of the national program "Young scientists and postdoctoral students" 2020-2021, Ministry of Education and Science, Bulgaria.

References

[1] Z.Ramzan, “Phishing Attacks and Countermeasures” In: P.Stavroulakis, M.Stamp (eds), “Handbook of Information and Communication Security”, Springer, 2010;

[2] Z.Terneva, Master thesis: “Design and development of a solution for malware protection”, 02.2020;

[3] M.Velev, “Penetration testing and information security”, Pragmatic lectures, 2019;

[4] Z.Wang, L.Sun and H.Zhu, "Defining Social Engineering in Cybersecurity," in IEEE Access, vol. 8, pp. 85094-85115, 2020;

[5] B.B.Gupta, N.A.G.Arachchilage, K.E.Psannis, “Defending against phishing attacks: taxonomy of methods, current issues and future directions.” Telecommun Syst 67, 247–267, 2018;

[6] T. N. Jagatic, N. A. Johnson, M. Jakobsson, F. Menczer, “Social Phishing” Communications of the ACM, Vol. 50, Pages 94-100, October 2007;

[7] M.Till-Rogers, "Protecting Yourself from Consumer Fraud and Scams.", LawNow 38, 2013;

[8] M. Prince, L. Holloway, E. Langheinrich, B. Dahl, A. Keller, „Understanding How Spammers Steal Your E-Mail Address: An Analysis of the First Six Months of Data from Project Honey Pot”, International Conference on Innovations in Computer Science and Engineering, 2005;

[9] D.Brewer. "A Link Obfuscation Service to Detect Webbots", 2010 IEEE International Conference on Services Computing, 07/2010;

[10] K.Uehara, K.Mukaiyama, M.Fujita, H.Nishikawa, T.Yamamoto, K.Kawauchi, M.Nishigaki, „Basic Study on Targeted E-mail Attack Method Using OSINT“ , International Conference on Advanced Information Networking and Applications, 2019;

[11] N.Yu, Z.Tuttle, C.Jake Thurnau, E.Mireku. "AI-Powered GUI Attack and Its Defensive Methods", Proceedings of the 2020 ACM Southeast Conference, 2020;

[12] C.Routh, B.DeCrescenzo, S.Roy, „Attacks and Vulnerability Analysis of E-Mail as a Password Reset Point", Fourth International Conference on Mobile and Secure Services , 2018;

[13] P.Engebretson, “The Basics of Hacking and Penetration Testing: Ethical Hacking and Penetration Testing Made Easy”, Syngress Publishing, 2011;

[14] W.Wang, J.Bickford, I.Murynets, R.Subbaraman, A. Forte, G.Singaraju, „Catching the Wily Hacker: A Multilayer Deception System", 35th IEEE Sarnoff Symposium, 2012;

[15] N.Klimburg-Witjes, A.Wentland, "Hacking humans? Social Engineering and the construction of the “deficient user” in cybersecurity discourses." Science, Technology, & Human Values, 2021;

[16] L.Zeltser, "Phishing Messages May Include Highly-Personalized Information", The SANS Institute, March 2006;

[17] M. Jakobsson J. Ratkiewicz, „ Designing Ethical Phishing Experiments: A study of (ROT13) rOnl query features ", 5th International Conference on World Wide Web, May 2006;

[18] M.Malik, T.Patel, „Database Security - Attacks and Control Methods”, International Journal of Information Sciences andTechniques, March 2016.