UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO

MAURÍCIO ARIZA

Automated Social Engineering Attacks using ChatBots on Professional Social Networks

Thesis presented in partial fulfillment of the requirements for the degree of Master of Computer Science

Advisor: Prof. Dr. Jéferson C. Nobre

Porto Alegre
June 2023

CIP — CATALOGING-IN-PUBLICATION

Ariza, Maurício

Automated Social Engineering Attacks using ChatBots on Professional Social Networks / Maurício Ariza. – Porto Alegre: PPGC da UFRGS, 2023.

73 f.: il.

Thesis (Master) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Computação, Porto Alegre, BR–RS, 2023. Advisor: Jéferson C. Nobre.

1. Cybersecurity. 2. Social Engineering. 3. Automation. 4. Bots. I. Nobre, Jéferson C.. II. Título.

“Everyday a rascal and a fool leave their homes.
When they meet each other, a deal happens.”
— BRAZILIAN COMMON PHRASE

ACKNOWLEDGEMENTS

Always good to thanks God for all the blessings in my life. My deepest and sincere thank you to my wife Sabrina, that encouraged me to return to my academic journey, face together so many though scenarios and helped me to evolve as a human being. Our family it's my source of strength and stability. My parents, that always took the education of me and my brother as priorities, pushing us to study, read and grow, being always amazing examples. My advisor Prof. Dr. Jéferson Nobre, for being a symbol on how computer science and psychology can collaborate, for the patience with my insecurities and avoidances, and especially for believing in a new research topic. My co-advisor Prof. Dr. Lisandro Granville, that gave us the necessary confidence to start this research and the experience and background needed for so many decisions. My colleague and friend Dr. Antônio Azambuja, that not only heavily contributed but also filled the gaps on my experience over academical methodology. To PPGC, for offering the opportunity of a high quality strincto sensu formation also for working students. To UFRGS, for offering public, free, accessible and high-quality education.

ABSTRACT

The grow of the internet and social networks had intensified online human interactions, raising the risk of cyberattacks. Social Engineering uses the same baseline as scams and fraud, but using technology as a support to exploit natural human failures. These attacks have been causing high damage, from financial to social, to individuals and companies. Research has shown the capacity of Social Engineering attacks, however, there are few papers focusing on the evolution and trust on ChatBots and automation as a support for those attacks, achieving scalability without the need of more exposure from the malicious agent. This work presents an analysis of the capacity of a professional social network to detect and block automated Social Engineering threats to their users. This work develops a proof of concept structure to analyze the viability of an attack aiming to get access to personal or corporate sensitive data. The approach developed allowed us to observe the creation of a trust relationship between an user, the social network and a ChatBot, and the failures from social networks to identify and block this kind of behavior. To this end, an Automated Social Engineering bot was developed. It introduces itself as a recruiter, contacts and interacts with a group of social network users with predefined characteristics, and acquires data to demonstrate the weaknesses that allow automated Social Engineering attacks to happen without being detected or blocked. The analysis and discussion of the results allows demonstrating the security vulnerabilities present in professional networks and propose some mechanisms and controls to protect the users.

Keywords: Cybersecurity. Social Engineering. Automation. Bots.

Ataques Automatizados de Engenharia Social com o uso de *Bots* em Redes Sociais Profissionais

RESUMO

O crescimento da internet e das redes sociais tem intensificado as interações humanas, aumentando os riscos de ataques cibernéticos. A Engenharia Social utiliza a mesma base que golpes e fraudes, porém utilizando a tecnologia como suporte para explorar falhas naturais do ser humano. Esses ataques tem causado grandes danos, de financeiros a sociais, em indivíduos e empresas. Pesquisas tem demonstrado a capacidade dos ataques de Engenharia Social, porém existem poucos trabalhos focando na evolução e confiança no uso de ChatBots e automação como suporte a esses ataques, alcançando escalabilidade sem a necessidade de maior exposição do agente malicioso. Este trabalho apresenta uma análise da capacidade de redes sociais profissionais de detectar e bloquear ameaças automatizadas de Engenharia Social aos seus usuários. Este trabalho desenvolve uma estrutura de prova de conceito para analisar a viabilidade de um ataque visando o acesso a dados sensíveis pessoais ou corporativos. A abordagem desenvolvida permite observar a criação de uma relação de confiança entre um usuário, a rede social e um ChatBot, e as falhas das rede sociais em identificar e bloquear esses tipos de comportamento. Para esse objetivo, um bot automatizado de Engenharia Social foi desenvolvido. O mesmo se apresenta como um recrutador, contatata e interage com um grupo de usuários da rede social com características pré-definidas, e coleta dados para demonstrar as fraquezas que permitem que ataques automatizados de Engenharia Social aconteçam sem serem detectados ou bloqueados. A análise e discussão dos resultados permite demonstrar as vulnerabilidades de segurança presentes nas redes profissionais e propõe mecanismos e controles para proteger os usuários.

Palavras-chave: Segurança da Informação. Engenharia Social. Automação. Bots.

LIST OF ABBREVIATIONS AND ACRONYMS

AI	Artificial Intelligence
API	Application Programming Interface
ASE	Automated Social Engineering
FBI	Federal Bureau of Investigation
IDS	Intrusion Detection System
IPS	Intrusion Prevention System
IT	Information Technology
HR	Human Resources
PoC	Proof-of-Concept
SE	Social Engineering
SMS	Short Messaging Service
USA	United States of America

LIST OF FIGURES

LIST OF TABLES

CONTENTS

1 INTRODUCTION

The evolution and growth of technology changed the way people in society live and organize, making easier daily processes or even creating new paradigms. In the same way that new digital infrastructures offered improvements to human life, they also created a path for exploiting those benefits for malicious purposes. Tools like social networks, which have become, across the years, a virtual space for interaction and connection between individuals, are also a target of those cyber attacks (SHIRES, 2018), bringing new challenges and risks for the matter of cybersecurity (KLIMBURG-WITJES; WENTLAND, 2021).

Social networks offer online services and collect information on several aspects from individuals and businesses, creating a high-value database for profiling, which can be used in different ways as tools for attackers exploiting the trust relations between the users (CROSSLER; BÉLANGER, 2014), like the usage of virtual profiles to obtain sensitive information (PARADISE; SHABTAI; PUZIS, 2019).

Attackers make use of the connectivity of these social networks to expand their area of operation, a fact that increases the challenges of cybersecurity. The connectivity of social networks and the growth of the cognitive dimension of work make human resources one of the pillars of security (CULOT et al., 2019) (GREITZER et al., 2019).

Virtual attacks on social networks have exploited human interaction in conjunction with technological gaps, weakening the cybersecurity chain. Organizations have used defense solutions to face these risks, such as firewalls, Intrusion Detection Systems (IDS), Intrusion Prevention Systems (IPS), and antivirus. However, even though these defense mechanisms can be effective in protecting the infrastructure, they are not sufficient to entirely prevent threats targeting the users directly (SALAHDINE; KAABOUCH, 2019) (KLIMBURG-WITJES; WENTLAND, 2021).

The human factor is usually discussed as the weakest link in the Information Security chain, creating those methods of attack which use tools and services offered by the technology as a support to tricky victims into executing actions or giving information, which become known as Social Engineering (SE) (MITNICK; SIMON, 2003) (KLIMBURG-WITJES; WENTLAND, 2021).

Social Engineering, covering from phishing and virtual scams, to well-structured targeted attacks and extortion, cause huge impact on the victims, from financial losses to social and psychological damage and trauma. Attackers can have any person as their main target, or just using them as an step for bigger targets, like companies or high-profile individuals. It is really common also to attackers use Social Engineering as an entry door for other forms of attacks, like ransomware, malware and even identity theft. Reviewing the Internet Crime Report produced by FBI, the damage caused by direct Social Engineering attacks was over 4 billion dollars in USA only in 2022 (not considering the usage of SE for other cyber attacks), impacting around 370.000 victims (FBI, 2023).

With the rise of automation mechanisms, the usage of bots has become more common daily in online services. Bots are automated software that sometimes uses features such as artificial intelligence. Its functionality can execute operation and control commands to impersonate humans, simulating the activities of real users (SHAFAHI; KEMPERS; AFSARMANESH, 2016). They can be used for positive actions and purposes, such as helping users in their online experience.

As done with several other tools, bot capacity has also been used for malicious purposes. Social Engineers have sought to develop bots with intelligence, making automated interaction unnoticed by users. Bots developed for Automated Social Engineering (ASE) attacks make it possible for a single attacker to contact a large number of potential victims simultaneously due to their scalability capacity. With the increased usage of social networks to establish personal and professional relationships, a field is open for the actions of ASE bots (HUBER et al., 2009). The attacker aims to get the victim to reveal sensitive information, which can be used for data theft (HUBER et al., 2009) (DEWANGAN; KAUSHAL, 2016).

The literature does not include a huge amount of research on Automated Social Engineering using bots. Most work focuses on the social psychology perspective, discussing human behavior over Social Engineering actions (HUBER et al., 2009). Some authors also discuss the threads on social networks caused by Social Engineering, like using fake accounts, identity theft, and phishing (AL-CHARCHAFCHI; MANICKAM; ALQATTAN, 2019) (PIOVESAN et al., 2019). Another interesting discussion on some papers analyzes the intentional influence on public opinion, using bots to manipulate and convince users of social networks for intended purposes, such as inciting Twitter users to compromise network infrastructure (FREITAS; BENEVENUTO; VELOSO, 2014) (MESSIAS; BENEVENUTO; OLIVEIRA, 2018).

Social Engineering attacks using identity theft to create fake accounts on social networks targeting users' privacy and information security were also analyzed in some papers (AL-CHARCHAFCHI; MANICKAM; ALQATTAN, 2019). An interesting study of SE tasks automation through a bot on Facebook concludes that persuasion is an essential resource in the ASE process (HUBER et al., 2009). However, no papers were identified that present bots with intelligence to perform an automated human interaction to search for sensitive information without the perception of the user being targeted by the ASE technique.

After an extensive literature review, we can see cases of using bots and automation techniques to support Social Engineering attacks, but usually helping with specific tasks like data gathering or spreading information. But in the same way as the general concept idea for those technologies is to replace human activity/interaction under controlled circumstances, it also creates an opportunity for attackers to use bots in a deeply way, fully replacing the necessary interaction with the victim that happens on a SE attack, achieving the maximum scalability with less exposure for the malicious agent.

The development of research focused on attacks and offensive security face some ethical issues due to the potential damage on their evaluation, especially as the targets for Social Engineering are humans and not systems and machines. The ethical limitations are also discussed in this work as they will bring limitations to tests and results, but also raises important dilemmas not commonly discussed in the Computer Science field.

Considering the capacity offered by those malicious bots to an ASE attack, this work proposal is to analyze the current most popular professional social network, LinkedIn. Following the necessary ethical requirements for experiments, a proof of concept bot and basic infrastructure will be used to simulate an end-to-end attack on LinkedIn, using an attractive job offer as bait and including a selection of targets, data gathering, interaction, and exploitation. Besides validating the applicability of a full ASE attack, we evaluate the response and controls of the social network to those attacks, analyzing their capacity to block or mitigate real potential situations.

The general objective of this work is to evaluate LinkedIn capacity to detect and mitigate an ASE attack. Our specific objectives are the following: (i) develop and implement a proof of concept to validate the technical viability of an ASE attack; (ii) understand and evaluate the current ethical framework for SE attacks research; and (iii) discuss and suggest control improvements that can be applied by social networks to decrease the risks of ASE attacks to their users.

The present dissertation is organized as follows. In Chapter 2 the theoretical concepts that support the research are discussed, followed by the related work in Chapter 3. Next, Chapter 4 presents the methodological aspects, the proposal of the study and the identified limitations. Chapter 5 performs the simulations and tests done to evaluate the proposal and discuss the results. Finally, Chapter 6 brings the conclusion and an approach for future studies in this field.

2 BACKGROUND

This section introduces the background for the present work, discussing some related concepts. First, it starts with an overview of Social Engineering in general. Then it follows by an analysis of Bots, their theory, and usage. Finally, to summarize the field of this research, an overview of Automated Social Engineering as a result of the malicious goals of attackers and the tools and capabilities of the technology to support it.

2.1 Social Engineering

Social Engineering (SE) is defined as the technique of exploiting people aiming to access privileged or sensitive information using social and technological interactions (LIBICKI, 2018) (KLIMBURG-WITJES; WENTLAND, 2021). Users represent their real-world persona, which makes them targets for social engineers (DWYER; HILTZ; PASSERINI, 2007); in practice, the human factor is the weakest link in cybersecurity (MITNICK; SIMON, 2003).

SE put the malicious agent in a favored position in the information flow, taking advantage of a trust relationship established with the victim(MITNICK; SIMON, 2003). The development of this trust relationship uses psychological manipulation, influencing the individuals to realize specific actions. The gain of the trust of the victims is the main goal of Social Engineers.

The SE attacks can be many times detected, but they are not easily stopped (LIBICKI, 2018). Due to characteristics of human behavior exploited in those attacks, SE techniques drove the victims to answer requests and execute actions without a proper analysis of the situation and the level of sensitivity of the requested information. Different scenarios can be used by the attacker to attract the victim's attention. A friendly request for help, exploiting a tendency for kindness and courtesy, or situations exploiting greed or similar, offering an advantage for the victim, like a financial or sexual opportunity, are commonly used techniques (MITNICK; SIMON, 2003).

SE can take different forms, but usually follow the same basic approach, involving four steps (MITNICK; SIMON, 2003) (TIOH; MINA; JACOBSON, 2019):

i) Obtain an initial information about the victim to support the initial bait.

ii) Establish a trust relationship between the attacker and the victim.

iii) Exploit the information and connection to develop specific actions manipulating the victim.

iv) Execute the main strike, achieving attacker goals.

2.2 Bots

Bots are automated tools that realize actions based on pre-defined command and control operations. The term derives from the word Robot, and in the same way, it is a technological mechanism that tries to replicate human actions or behaviors. These mechanisms can have more simple script-like mechanisms in a request-response format or more advanced ones using Artificial Intelligence (AI) (FERRARA et al., 2016).

As a certain degree of intelligence is built into the tools to simulate human behavior, the capacity, and scalability for services increase. Besides the different classifications of bots and their goals, the relevant ones for this research are ChatBots, which simulate human conversations, and SocialBots, which operate on social networks (SHAFAHI; KEMPERS; AFSARMANESH, 2016). These tools, SocialBots and ChatBots, have been developed with the help of AI mechanisms that interact with users (FREITAS et al., 2015).

ChatBots are the integration of systems, tools, and scripts that promote instant messaging conversations with or without human participation (STOECKLI; UEBERNICKEL; BRENNER, 2018). They are developed to help human users in specific service situations, being some common usage customer service, communication service, and digital education service (GRIMME et al., 2017).

AI interaction devices have been evolving and a market of products, like personal assistants, have been growing, raising interest as a research field, especially for challenges like the usage of natural language (KHAN; DAS, 2018). With the increase of usage, ChatBots have been a constant target for attacks, that usually focus on the client module, the communication module, the response generation module, or the database (YE; LI, 2020).

SocialBots are developed for social media usage, being able to publish content or interact with users, for example, simulating regular human behavior in those networks (ROUSE, 2013). They need a technical infrastructure, with a combination of a social networking platform and technical requirements for automating the behavior of an account, using an Application Programming Interface (API) or proprietary mechanisms to interact with the platform (ASSENMACHER et al., 2020).

The main usage of SocialBots it's on automated social media accounts that impersonate an individual or character (HEPP, 2020), becoming each day more common to find these AI robots performing activities in the online environment (FREITAS et al., 2015). The degree of intelligence required in SocialBots to simulate human behavior sparks interest in research on the topic (FERRARA et al., 2016).

The capacity of the SocialBots to simulate human behavior is a powerful tool for malicious agents. They can be used for spreading false information (fake news) and manipulating public opinion, bypassing security mechanisms or sending spam and phishing (FREITAS et al., 2015). This allows ASE attacks using bots to bring Social Engineering to a new level of scalability (HUBER et al., 2009).

SocialBots are an effective tool to perform SE attacks aiming to gain access to sensitive information. They had the ability to compromise the structure of social networks, executing actions like (i) stealing identity; (ii) influencing users; (iii) increasing the number of followers; and (iv) inflating the popularity ratings of a particular profile account (BOSHMAF et al., 2011) (CAMISANI-CALZOLARI, 2012) (DEWANGAN; KAUSHAL, 2016).

2.3 Automated Social Engineering

Social Engineering attacks require time and resources to develop a trust relationship between the malicious actor and the victim. As human communication has been the base to the development of human-machine interfaces like bots, this capacity allows attackers to use them and automate the steps for the connection and rapport with the victim (GUZMAN; LEWIS, 2020).

Social Engineering attacks require time and resources to establish a trust relationship, which can be accomplished through automated mechanisms. ASE attacks require minimal human intervention, as an automated robot impersonate another human to to establish a connection with the victims (SHAFAHI; KEMPERS; AFSARMANESH, 2016) and can reach several targets simultaneously due their scalability capacity (MITNICK; SIMON, 2003) (HUBER et al., 2009). Automated attacks can be prepared using sensitive information and/or through influencing certain audiences (GALLEGOS-SEGOVIA et al., 2017).

Social networks facilitate communication, social interaction, and share of personal and corporate information, increasing their popularity in the cyber environment. These networks represent an attractive virtual space for attackers to exploit technical vulnerabilities and users' lack of knowledge and awareness of SE actions (AL-CHARCHAFCHI; MANICKAM; ALQATTAN, 2019). ASE attack using resources like SocialBots and phishing in those environments are each day more common, taking advantage of usage growth for personal and professional activities (KIMPE et al., 2020).

The connections formed in these virtual socialization environments allow a big exchange of information, reinforcing the role of networks as communicative structures for social relationships (CASTELLS, 2009). Cyberspace constitutes a promising scenario for the practice of all sorts of illicit acts without respecting geopolitical borders. The growth of social networks has enabled the creation of a large number of fake profiles, with the use of automated bots to support and scale the malicious activities (TIWARI, 2017).

3 RELATED WORK

The proposal to evaluate an ASE attack and develop a proof-of-concept (PoC) bot requires a structured literature analysis to support this research. This literature review allowed to identify that most related papers focus in the human behavior over actions done by SocialBots, phishing and/or SPAM.

Dewangan and Kaushal presents a model for detecting SocialBots used in political campaigns and marketing of products, having as input the behavior analysis. These actions bring with them security risks, considering the use of social networks for disseminating political positions and monitoring the consumption profile of users (DEWANGAN; KAUSHAL, 2016). This work brings a perspective on the capacity of bots to coexist on social networks and manipulate or influence the human users. The kind of influence analyzed by them was more on large audiences. Our work share similarities on the capacity of a bot to interact with humans without suspicious behavior, but focus on the manipulation of a single individual (1:1 relation) instead of large groups.

Aroyo et. al. discuss how SE exploits the trust relationship between users and bots. The work is based on the four stages of an SE attack proposed by Mitnick and Simon: (i) obtain the information about the victim; (ii) establish a trust relationship; (iii) exploit the information for the development of specific actions; and (iv) execute the attack to achieve their goal(s) (MITNICK; SIMON, 2003). A robot was developed to simulate this task. First the robot sought to obtain information with private questions. Then, it established a relationship of trust with the users, for a virtual and anonymous approach to the target (AROYO et al., 2018). With these actions, authors present in the research results that users have established a trust relationship with the tool. Among the requirements in the interaction with users the ethical aspects were considered, by these authors. The approach of gaining the user trust to later use it to get sensitive information brings several interesting points, especially as how people easily trust in the machine - a very important assumption for ASE attacks. For our own research the trust between the bot and the user happens more based on a bait and a storytelling around it, besides having all interactions happening through virtual environments instead of in-person interactions as done by Aroyo et. al. robot.

Al-Charchafchi, Manickam and Alqarran present a review of research on privacy and threats in social networks. For the authors, although the literature presents work on privacy, more effort is needed. The social networking environment is a rich source of personal data, making it an attraction for actions in social engineers, who exploit the users' lack of awareness and knowledge on security-related issues (AL-CHARCHAFCHI; MANICKAM; ALQATTAN, 2019). Being a literature review, the main goal of this work is to summarize research and concepts, it helps to build a solid baseline to create the assumptions for a research like our but the development and objectives are not the same.

The complexity of SE attacks is related to the combination of social strategies and techniques used to carry out a cybercrime (AL-CHARCHAFCHI; MANICKAM; ALQATTAN, 2019). In this context to mitigate the impacts of attacks, Piovesan et. al. claim that security policies can provide higher level of information security (PIOVESAN et al., 2019). However, they do not guarantee complete security.

Freitas, Benevenuto and Veloso present a discussion on the impact of the use of SocialBots on Twitter to characterize the behavior of the tool on a large database. In the results the authors highlight that the method they developed to characterize and detect SocialBots, had a 92% successful detection indicator (FREITAS; BENEVENUTO; VELOSO, 2014). Their work also focused in exploit a social network to introduce bots and use them to interact and manipulate users. The main difference besides the kind of social network analyzed is that their work focus on the strategies for bots to succeed and ways to identify them, while our research the bot is used as a tool to automatize the SE attack and target individuals, not general group manipulation.

Messias, Benevenuto and Oliveira, by analyzing the methods used to measure influence in social networks, evaluated the capacity of SocialBots to exploit the social network and increase their influence by manipulating users and algorithms. Researchers claim in their results that a simple bot can achieve high levels of influence on Twitter (MESSIAS; BENEVENUTO; OLIVEIRA, 2018). Besides the similarities on the need to create credibility to gain user trust, on our research we focused on the capacity of the bot to manipulate a single user and stay undetected by the social network controls.

Shafahi, Kempers and Afsarmanesh tested the capacity of bots to gain users' trust by interactions and influence on Twitter discussions, and them using this trust to send phishing to the users. The authors point to the need to raise the level of awareness about SocialBots phishing actions, as they become a threat to people and organizations (SHAFAHI; KEMPERS; AFSARMANESH, 2016). On their research the bots gain user trust by interactions and then apply a format of SE attack (phishing), as credibility is an important key to increase the chances of success in an attack. On our work we exploit the social network controls to create the credibility for the bait, and use the bot to keep the interest and attention through a storytelling (the promise of a job opportunity).

Paradise, Shabtai and Puzis analyze in the organizational context the strategies to monitor organizational social networks and detect SocialBots that aim to obtain data from the organization. The strategies were analyzed considering different levels of attacker knowledge using a simulation with real social network data (PARADISE; SHABTAI; PUZIS, 2019). The authors work on a research to evaluate the techniques to detect bots used for manipulation, which have similarities to our evaluation of LinkedIn enforcement of controls to detect and block bots and automation.

Huber et. al. present the cycle of an ASE attack using a Bot. The attack demonstrated how social networks can be used by social engineers to obtain information. To this end, two (2) experiments were conducted in the study. The first analyzed the ability of a bot to obtain information from social networks. The second performed the Turing test, which seeks to evaluate the ability of a machine to imitate a human being. Finally, for the authors, ASE with bots is scalable and requires fewer human resources. The tool was used in a proof of concept on Facebook. Their experiments allowed to ratify that it is possible to automate SE actions to obtain information and to demonstrate that the bot used was not identified by the security measures of the Facebook. The increasing number of users' social interactions on networks makes SE automation Bots an interesting tool for social engineers (HUBER et al., 2009). This was the work with more similarities to our, evaluate the capacity of a social network against ASE attacks. The main difference was our focus on professional social networks, where context helps to build the credibility and storytelling over the bait that increase the chances of success.

Work	Focus	What we add
Dewangan and Kaushal (2016)	Bot capacity to manipulate audiences	Manipulation over a single individual
Aroyo et. al. (2018)	Robot uses help as a way to gain trust	Bot gain trust using a bait and storytelling
Al-Charchafchi, Manickam and Alqarran (2019)	SE Survey/Literature review	SE attack based on the concepts
Freitas, Benevenuto and Veloso (2014)	Strategies for a bot to succeed in user manipulation	Automatize the manipulation of an user
Messias, Benevenuto and Oliveira (2018)	Manipulation of influence measurement	Influence an user by credibility and a bait
Shafahi, Kempers and Afsarmenesh (2016)	Using bots for phishing	Using bots for SE scams
Paradise, Shabtai and Puzis (2019)	Analysis of bot monitoring	Analyze the capacity of a social network to detect a bot attack
Huber et. al. (2009)	ASE attack on Facebook	ASE attack on LinkedIn

Table 3.1 – Summary of related work comparison.

4 METHODOLOGY

Social Engineering is defined as the technique of exploiting people aiming to access the data and information of potential targets of information systems using combinations with social and technological interactions (LIBICKI, 2018) (KLIMBURG-WITJES; WENTLAND, 2021). These attacks have four stages, namely: (i) obtaining the information about the victim for a first approach; (ii) establishing a trust relationship between the attacker and the victim; (iii) exploiting the information for the development of specific actions; and (iv) executing the attack to achieve their goal(s) (MITNICK; SIMON, 2003). As the users represent a real-world persona, they become targets for social engineers (DWYER; HILTZ; PASSERINI, 2007). It is already a common concept to classify the humans/users as the weakest link of the cybersecurity chain, as attacks exploiting their failures have the better success rates and sometimes requires less technical skills and risk for the attacker (DARWISH; ZARKA; ALOUL, 2012).

Social networks have the objective of creating interaction between humans. But beyond allowing a space where distance boundaries can be bypassed to enhance connections, they also bring to the virtual world many of the threads from the real world. But there is a difference as it is a space where people do not have the same awareness and capacity to recognize risks, which together to the stronger capacity of anonymity and impersonation become a perfect environment for SE (CROSSLER; BÉLANGER, 2014).

Comparing to other social networks like Facebook, Twitter and Instagram, professional social networks create a more corporate environment, focused on business connections and career growth. This business-like scenario creates a sense of trust and credibility, attracting headhunters looking for candidates as well companies looking for potential new customers. The relations formed in the professional networks are already exploited by social engineers, especially impersonating recruiters using attractive job opportunities as a bait to steal internal information or personal data of the victims.[1]

Currently, LinkedIn is the most popular professional social network, with more than 930 million members in more than 200 countries according to themselves (May 2023)[2]. The "LinkedIn User Agreement" defines on Section 8.2[3] the actions that are not allowed to users, highlighting forbidden use of false information or impersonation in the profile and usage of bots and automation to realize actions in the platform.

[1] https://www.ft.com/content/a8d262f4-5d52-4464-8714-e21a457aab33

[2] https://about.linkedin.com

[3] https://www.linkedin.com/legal/user-agreement#dos

Considering references of existent SE attacks on LinkedIn, we can found references of fake profiles or false information used for several reasons. Talking specifically about automation, if we search in the internet or code repositories we could find several bots and scripts specifically designed for LinkedIn. If the policies of the social network forbidden fake data and automation, but we can identify those happening, this indicates a potential lack or insufficient implementation of controls by the platform, or that they enforcement of policies is done based on complaints or reports done by the users. So besides all those observations, our goal is to do a real evaluation of LinkedIn ability to detect and/or block automation and fake data, as those are basic requirements to execute an ASE attack against the platform users. Once understanding the real risk level and LinkedIn response, we can also offer suggestions to improve their controls and enforcement, decreasing risk with no or minimum impact to usability.

4.1 Evaluation Proposal

We used a proof of concept scenario with 2 bots to evaluate the attack. The first one interacts with the social network to search and contact the victims with the bait - the Platform Bot. The second one is a ChatBot service that would act directly with the victims to execute the step of the job interview - the Recruiter Bot. Also, to support the execution of those actions, we created a fake LinkedIn profile impersonating a job recruiter.

For the Platform Bot role, we developed a Python code[4] to connect with LinkedIn. LinkedIn offers an API[5] for software interaction. However, regular human users do not navigate in a social network through an API, so we potentially would have different results if we follow this path. In order to better reproduce the same usage of a human, we used the Selenium library[6] to allow the bot to act in a request-response format through the browser.

For the Recruiter Bot, we had several options available that could execute the necessary actions without the need to ourselves develop custom code. Using a pre-defined set of job interview questions, plus information scrapped from the victim LinkedIn profile, the Recruiter Bot would basically conduct a fake job interview with the victim with the goal to collect sensitive information from current and past jobs. As it can execute both

[4] The tool is stored in a GitHub private repository. Access can be requested by contacting the authors.

[5] https://developer.linkedin.com/product-catalog

[6] https://www.selenium.dev/

HR-like interviews with more generic questions and a technical interview, the entire process can be executed by the same bot. Depending on the goal of the attacker the bot can include questions about specific companies or experiences the victim had - collected from the scrapped profile - looking to steal sensitive information about projects, customers, etc. Also, in the end of the interview, the attacker can "select" the victim for the position, stealing personal information through the signing of a fake work contract and requiring documents like a passport, for example, being the pre-attack for further identity theft.

Similar to the four-stages structure for SE attacks (MITNICK; SIMON, 2003), our proposal it's also organized in steps but follows a different structure: i) Authentication, ii) Search, iii) Approach and iv) Interview. This format allowed us to break the attack in stages and test each one individually, following the guidelines detailed in the Limitations (Section 4.2). The Figure 4.1 indicates the attack phases tested by each of the proposed proof of concept Bots.

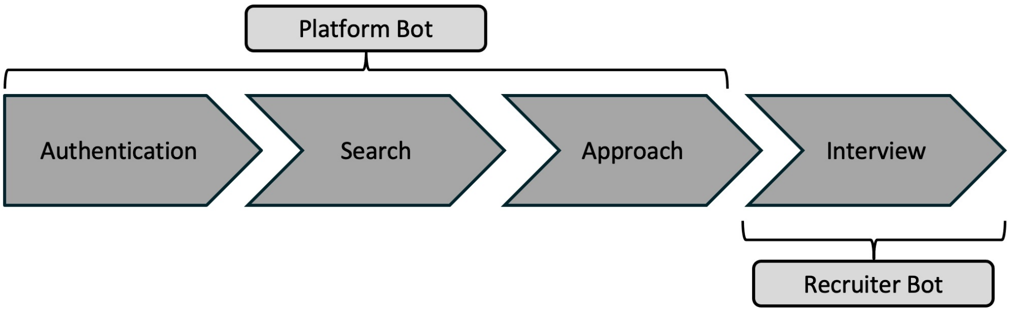


Figure 4.1 – Relation between the developed Bots and the attack phases.

1. Authentication: As illustrated in Figure 4.2, the objective of this step is to verify if the social network detects or have different behaviors when the user logon process is done using automation. For this evaluation, our Platform Bot opens LinkedIn website in the browser, maps the source code of the main page to identify the credential fields, fill them with the values received and then submit the credentials to server and conclude the authentication process, accessing then the main page of a logged user.

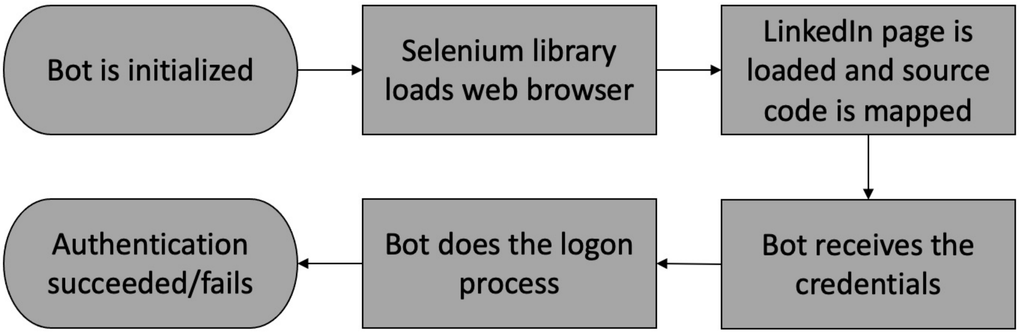

Bot is initialized
Selenium library loads web browser
LinkedIn page is loaded and source code is mapped
Bot receives the credentials
Bot does the logon process
Authentication succeeded/fails

Figure 4.2 – Attack authentication phase.

2. Search: This step aims to check the detection of automated searching of users. Similarly to the first step, the Platform Bot maps the page source code, identify the search field, run the search for the provided terms and them stores temporarily the returned profiles. Figure 4.3 also refers to this step.

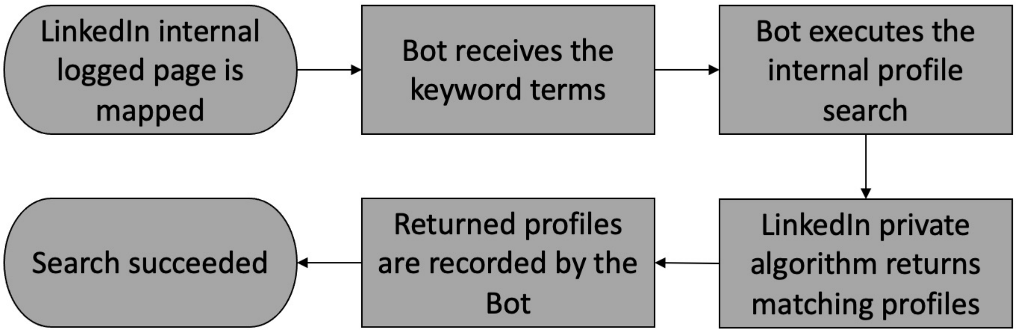

LinkedIn internal logged page is mapped
Bot receives the keyword terms
Bot executes the internal profile search
LinkedIn private algorithm returns matching profiles
Returned profiles are recorded by the Bot
Search succeeded

Figure 4.3 – Attack search phase.

3. Approach: The goal of this step is to start the interaction with the profiles of potential victims collected in the previous step. As also seen in Figure 4.4, using the stored profiles captured, the Platform Bot add them as contacts and send a custom message, which serves as the bait for interaction. This action happens to all profiles captured.

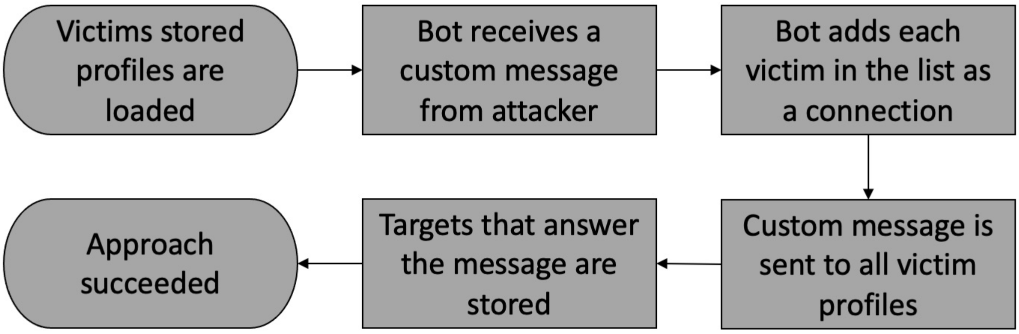

Victims stored profiles are loaded
Bot receives a custom message from attacker
Bot adds each victim in the list as a connection
Custom message is sent to all victim profiles
Targets that answer the message are stored
Approach succeeded

Figure 4.4 – Attack approach phase.

4. Interview: Based on the results of the bait sent in step 3, a script scrape data from the victim LinkedIn profile to feed the Recruiter Bot database, which them have enough information to execute a job interview with the victim. Figure 4.5 also illustrates the full cycle of this step. As in general even for real recruiter approaches the individual is contacted in LinkedIn and then the interview and other steps usually happens in different channels, it is expected that the bait would include an invitation to have the interview in a different channel than LinkedIn messenger itself.

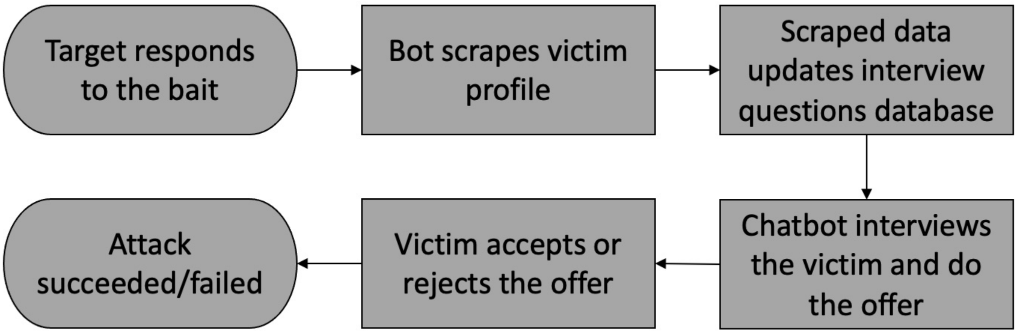

Target responds to the bait
Bot scrapes victim profile
Scraped data updates interview questions database
Chatbot interviews the victim and do the offer
Victim accepts or rejects the offer
Attack succeeded/failed

Figure 4.5 – Attack interview phase.

4.2 Limitations

SE born in the field of psychology, as besides the usage of technology as a support, to achieve their main goal attackers exploit human weaknesses and behavioural characteristics, research topics from Social Sciences. For a full understanding of the impact of a SE attack, we would need not to only validate technical aspects, but also to trick subjects and observe their behavior and actions. The field of social psychology research, especially due to the many scenarios involving human subjects, face several challenges on discussing the boundaries for ethical research. In history we have extreme examples like the famous Milgran's experiments on the 70's (RIECKEN, 1974), resulting in several impact and trauma for the subjects, situations that modern research understand as unethical.

Expose people to situations where they will be deceived, having their vulnerabilities exploited without their consent (or full understanding) violates ethical dilemmas. As results, later they can create frustration and stress due to expectations created, broken promises or the feeling of being fooled. Understanding and respect those boundaries was one of the drivers of this work, and even as it is not a technical topic is not possible to run research on ASE or any kind of cybersecurity attack without discussing and raising questions in the matter of research ethics.

The first challenge was on how to validate the proposal within keeping compliance to ethical policies. In a best effort to achieve that, we decided to break the attack life cycle in separated steps and evaluate each one individually. The results would offer a fair understanding of the application response, and crossing the data of the different steps should be possible to conclude the potential of an end-to-end attack.

The main difficulties happened on the Approach and Interview steps, as they would require at least some level of contact with subjects. Following the necessary requirements for ethical research, it would be necessary to invite subjects for the activities having their clear consent on being observed and analyzed, besides keeping a transparency on the goals, steps and general information about the activity. Considering the goals of an ASE attack, any information shared with the subjects would affect the results. Alternatives like sharing a different objective with participants (for example, help to evaluate an HR Interview Bot) can also be considered as misleading. Taking into account the effort involved for usage of real subjects and the minimal or nonexistent gain for the goals of this research, the decision was to not follow this path and instead focus on the technical response from the application.

Especially the Approach step happened in a thin line between keeping the assumptions and violating ethical barriers. As it would require to generate requests to real users in the platform, it was decided to set a limit to the minimum quantity of LinkedIn users receiving the request and the message. This minimum quantity of targets was defined as the profiles present in the first page of results (usually between 15 and 21 profiles), as approaching all them simultaneously or in a very small time window will only be achieved by using some kind of automation or actively doing SPAM or similar actions. As the goal was not to measure the users response to the bait and understanding the requirements to have human subjects participating, after sending the messages and validating that no blockers or similar happened on the platform, all the interactions were immediately cancelled/excluded to suppress the chances of them being viewed or replied by any real users.

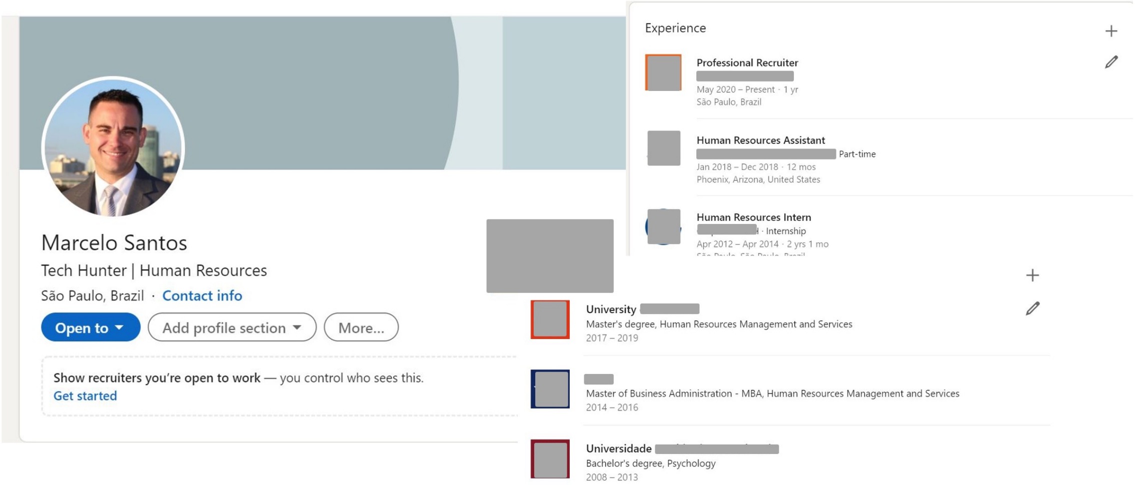
Understanding the difference between how many requests per second an user testing the platform could do versus an user just navigating through the regular usage allow us to identify behaviors that characterize automation without the need to identify the maximum threshold of the application or generate denial of service. Even as potentially some kind of control could be triggered after hundreds or thousands of requests are done, that will be more a control against denial of service or high throughput than a protection against automation. So huge numbers are not necessary to identify automation behavior.

For the Interview step, the evaluation focused on the main functionality of the Recruiter Chatbot: do a job interview. The only differences between a malicious and a real job interview will be the goals, as instead of trying to evaluate the capacity and skills of an individual for a certain role, the objective would be mainly steal data through the questions or by signing a contract and providing documents in a fake hiring. The usage of an already existent Recruiter Chatbot allows to only collect the data from the victim to feed the bot and observe if the proposed malicious questions are correctly distributed within the interview, without needing to validate the Recruiter Chatbot capacity itself.

Considering all those mechanisms implemented and decisions on how to test each step, would be fair to conclude that enough results could be achieved to validate the potential of the proposed attack without violating any associated ethical requirements. This precaution is a core topic for any kind of research and a deeper analysis of the impact of ethical research matters especially on the SE field it is an intriguing topic for further studies.

5 SIMULATION EXPERIMENTS

Our first step was to create the fake profile that will support the execution of the attack. We used a picture from a free image database and random data to include some previous work experience and educational background. We could associate the profile to real companies and universities with no checkings or verifications required. We not observe any impact due to the usage of false data, or due to the fact the since the creation of the profile all their interactions with LinkedIn platform were done using some kind of automation. Figure 5.1 shows some highlighted information of the fake profile created.



Marcelo Santos
Tech Hunter | Human Resources
São Paulo, Brazil · Contact info

Show recruiters you're open to work — you control who sees this.
Get started

Experience

Professional Recruiter
[redacted]
May 2020 – Present · 1 yr
São Paulo, Brazil

Human Resources Assistant
[redacted] Part-time
Jan 2018 – Dec 2018 · 12 mos
Phoenix, Arizona, United States

Human Resources Intern
[redacted] · Internship
Apr 2012 – Apr 2014 · 2 yrs 1 mo
São Paulo, São Paulo, Brazil

University [redacted]
Master's degree, Human Resources Management and Services
2017 – 2019

[redacted]
Master of Business Administration - MBA, Human Resources Management and Services
2014 – 2016

Universidade [redacted]
Bachelor's degree, Psychology
2008 – 2013

Figure 5.1 – Fake recruiter profile created.

The simulation experiments followed the steps of our proposed Attack Flow (Section 4.1). The Platform Bot was executed in a Windows machine running Python, the Selenium library and Google Chrome as a browser. For the Recruiter Bot, we used the SAP Conversational AI platform.

5.1 Testing Step 1 - Authentication:

The success criteria of this step is to execute authentication in the platform following different behaviors and observe if any of the simulations triggers controls or blockers in the application due to automation characteristics. For a comparison criteria, we defined three basic behavior patterns to be tested: (1) Do the logon process 10 times simultaneously, and (2) Do the logon process 10 times with a 5 seconds waiting time between each attempt, and (3) Do the logon process 10 times with a 10 seconds waiting time between each attempt. Those patterns try to replicate behaviors not expected from a real human user due to quantity or speed of attempts, especially as the execution is happening through web browser. For all the three patterns proposed, we also tested using the following variations to observe if they impact the results in any way:

- receive the credentials of the created fake profile in execution-time through script;
- read the credentials of the created fake profile from a file;
- use of wrong/invalid credentials;
- use of a public proxy to execute the logon from a random country, different of the one defined as the location of the user in the created fake profile.

As summarized in Table 5.1, in the results of our tests we do not observed any difference in the social network behavior, besides after some attempts with wrong credentials. We also executed the test of each pattern in different days to guarantee that the execution of one of them would not impact in the results of the others. As an alternative variation, we also executed 10 sequential logon attempts using the valid fake profile credentials and invalid ones, but manually (not using the bot), through web browser, which not demonstrate as well any difference on the results.

When testing using invalid credentials, both through bot or manually, after the 6th attempt LinkedIn start requiring a puzzle (similar to a Captcha verification) and/or additional validation like a code sent by email/SMS to proceed with login, indicating that brute force behaviors are identified and blocked, which not happens for all kind of automated access attempts.

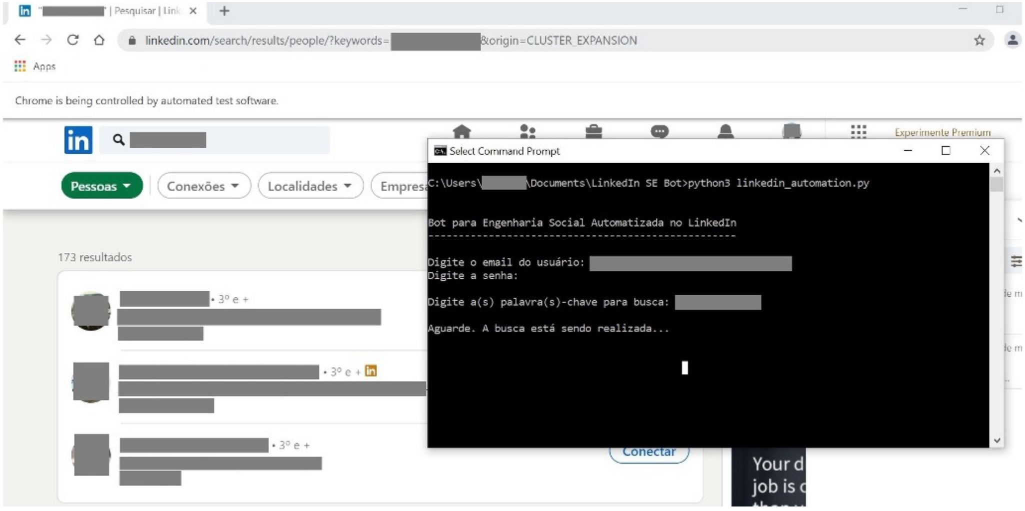
A potential point of discussion would be the quantity of requests done. Besides a similar sequence of login attempts under some circumstances could be reproduced by a human being, this behavior will only happen for conscious test purposes, not for regular usage. The standard process of login involve the input of credentials, logon and then navigation, with eventual typos causing the logon to fail a few times. In the same way that after some failed attempts (6, to be specific for LinkedIn) an additional control (the Captcha) it's enabled as this behavior is considered as suspicious, even that it can be reproduced by a human being. In the same way, we understand that the values around 10 login attempts simultaneously or in a very short time frame characterize a suspicious and potential automation behavior.

Variation	Simultaneous	5s Delay	10s Delay
Credentials in execution time	Success	Success	Success
Credentials from file	Success	Success	Success
Wrong/Invalid credentials	Fail/Captcha	Fail/Captcha	Fail/Captcha
Public proxy	Success	Success	Success

Table 5.1 – Login test results summary.

5.2 Testing Step 2 - Search:

On this step, we evaluated the capacity of the Platform Bot to execute queries in the social network without being detected. To run the queries, the authors defined a series of keywords to be used. It is important to highlight that the results of the search query, like the quantity of profiles returned, the order that they appear on the results and similar information are directly related to the LinkedIn search algorithm, and undertand or manipulate the results are not in the scope of this work. Keywords used were only a manner to evaluate the response for automated queries through web browser. Figure 5.2 demonstrates the Platform Bot executing the search phase.



linkedin.com/search/results/people/?keywords=[redacted]&origin=CLUSTER_EXPANSION

Chrome is being controlled by automated test software.

```
C:\Users\[redacted]\Documents\LinkedIn SE Bot>python3 linkedin_automation.py


Bot para Engenharia Social Automatizada no LinkedIn
--------------------------------------------------

Digite o email do usuário: [redacted]
Digite a senha:

Digite a(s) palavra(s)-chave para busca: [redacted]

Aguarde. A busca está sendo realizada...
```

Figure 5.2 – Search phase being executed by Platform Bot.

We enumerate for the test ten keywords, based on some IT skills associated with this project only for reference, which are: "test", "Social Engineering", "Bot", "Chatbots", "Social Networks", "Information Security", "Python", "Automation", "GitHub" and "API". Similar to what was done in Step 1, we used the following variations:

- Querying the same keyword 10 times simultaneously.
- Querying the same keyword 10 times with a 5 seconds waiting time between each.
- Querying 10 simultaneous sessions, each one using a different keyword.

- Querying the 10 different keywords in the same session, in sequence, with 5 seconds waiting time between each.

It was not a goal of this step to do stress/load testing or cause a denial of service in the application. The tested behaviors used a speed and/or quantity not expected to be executed by a human user, especially as they were executed through web browser. Besides the expected differences on the results (considering the different terms used and the LinkedIn algorithm, which are out of the scope of this paper), we not observe any differences in the variations, and all queries received correctly the results with a list of profiles associated with the keyword term. The Table 5.2 summarizes the results of the variations used in each keyword search query, where "Success" indicated that the query was run in that variation without trigger any controls of similar.

Keyword	Simultaneous	5s Delay	Simultaneous Sessions	Sequence
test	Success	Success	Success	Success
Social Engineering	Success	Success	Success	Success
Bot	Success	Success	Success	Success
Chatbots	Success	Success	Success	Success
Social Networks	Success	Success	Success	Success
Information Security	Success	Success	Success	Success
Python	Success	Success	Success	Success
Automation	Success	Success	Success	Success
GitHub	Success	Success	Success	Success
API	Success	Success	Success	Success

Table 5.2 – Search query test results summary.

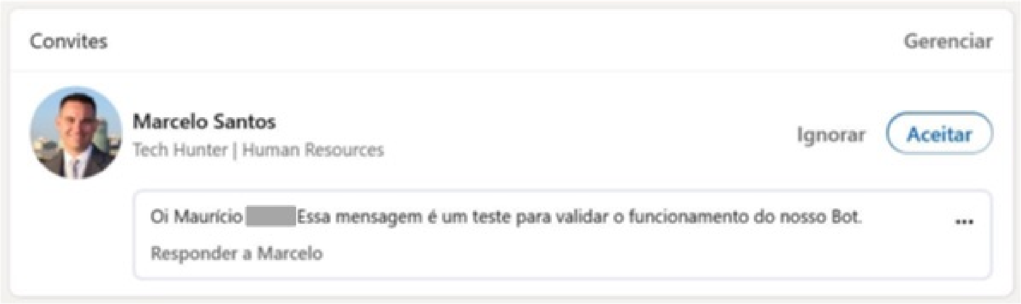
5.3 Testing Step 3 - Approach:

In the previous step, after running the queries, the Platform Bot keep a reference of the returned profiles to be used for this step. Here we had our main challenge on the already discussed ethical implications. Looking to have a limit on the impact of our research without prejudice to the results, we implemented the following controls to our bot:

- For each query, instead of the several pages of results, we keep only the ones in the first page, which were around 15-21 profiles per query.
- We only executed the approach 10 times, one per keyword, disregard variations on queries.

- After executing the approach, the bot keep a record of the success and them delete/cancel all their actions within the user/victim.

The approach happened with a connection request to the user and the send of a custom message, using tags to use the real user name instead of generic terms like ’dear user’. Again, once validated the request and the message, the request was cancelled and the message deleted for both sides, avoiding any further interaction with the users. Again no impact or actions from the side of social network were identified during any of the tests. Figure 5.3 shows an example of a custom message being tested to approach a target.



Marcelo Santos
Tech Hunter | Human Resources

Oi Maurício [redacted] Essa mensagem é um teste para validar o funcionamento do nosso Bot.

Figure 5.3 – Example of custom message.

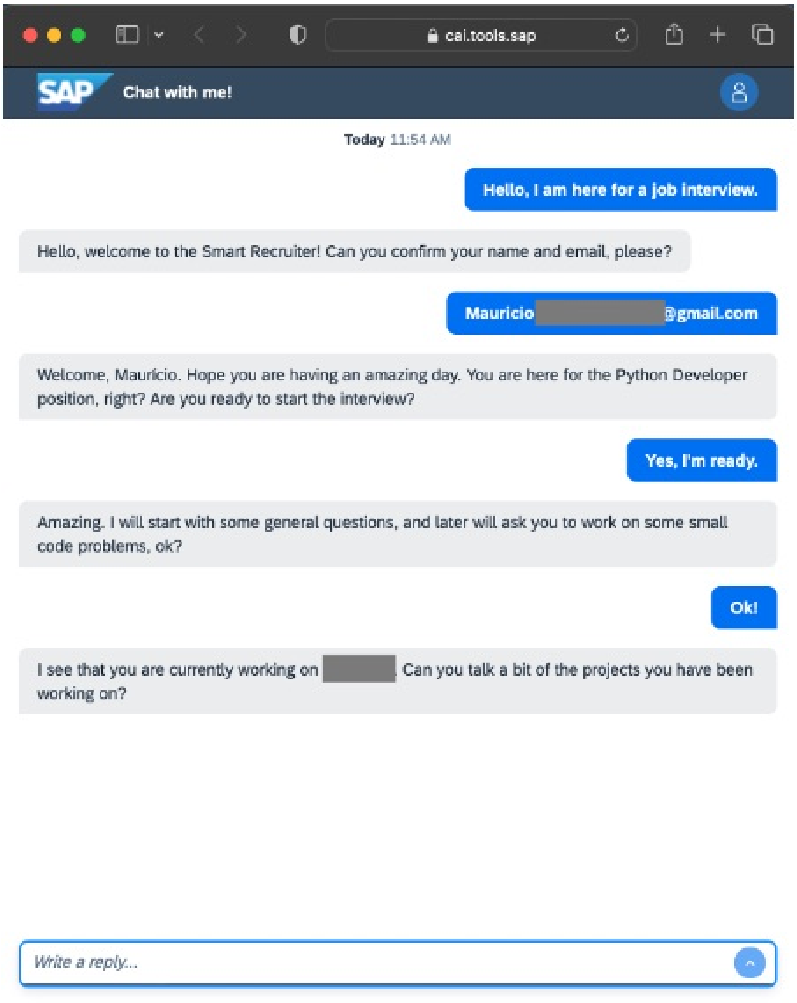
5.4 Testing Step 4 - Interview:

For this step our validation followed a different direction. We used the SAP Conversational IA[1], a platform for creation of chatbots. Instead of creating our own, we used an existent chatbot for job interviews available in the platform, the Smart Recruiter. Using this approach, besides some basic checks we did not need to validate the chatbot capacity to execute an interview, but only to provide our malicious input and observe the results. Based on an initial database of common questions for job interview already available in the chatbot, we used a script to scrape the data from the victim profile and use them as an input for additional questions, creating more specific questions like "How was your experience in Company X?", "Can you talk more about your skills on technology Y?" or even "Please mention some of the main customers and projects you had a key role on on company X".

Based on the observation of the chatbot interview for different profiles randomly selected from the previous step, it was capable to conduct a job interview without need for management or additional command/control. This result allow us to demonstrate their

[1] https://cai.tools.sap

capacity to be used for the proposed attack without having a real approach with subjects/victims, violating the already discussed ethical limitations. Figure 5.4 shows the beginning of an interaction with an user during an interview. Besides a regular chatbot instant messaging format was used during the tests, the platform offer tools to connect and execute the interview through different messaging channels using API or webhooks. It is even possible for an attacker to create a fake company website and embed the chatbot on it to create a more trustworthy scenario.



Hello, I am here for a job interview.

Hello, welcome to the Smart Recruiter! Can you confirm your name and email, please?

Mauricio [redacted]@gmail.com

Welcome, Mauricio. Hope you are having an amazing day. You are here for the Python Developer position, right? Are you ready to start the interview?

Yes, I'm ready.

Amazing. I will start with some general questions, and later will ask you to work on some small code problems, ok?

Ok!

I see that you are currently working on [redacted]. Can you talk a bit of the projects you have been working on?

Figure 5.4 – Recruiter Bot executing an interview.

5.5 Discussion of Results

The main contribution of this paper was a way to validate the hypothesis of lack or insufficient controls implemented by social networks. No matter this could be considered already known or expected in the technology field, we could not found official research or academic references as foundations to base any conclusions. It was not the goal to explore the human factor and the psychological matters associated to SE, but to observe if the technology channels allow or at least do not offer barriers to avoid those aspects being exploited in their users, especially on cases like our proposal were it is possible to achieve high scalability by the attacker.

Certainly, the main impact of implementing rigid controls in a social network will be on the user experience, as the result can made the usage less smoothly and easy, with potential migration of users to competing services and cause negative effects in success indicators of the platform. Although, it should be possible to find a balance and increase the security levels with no or minimum impact to the users.

Considering that the current LinkedIn's User Agreement already forbidden the usage of automation, some automation-detection controls can be applied. This will not only avoid ASE, but an entire behavior that it is already forbidden. However even this control can be implemented with some flexibility. A regular user connecting through web browser have some human limitations on their speed and quantity of requests - below a certain limit, not only you have a certain or potential automated behavior, you also have high chances of SPAM and other unsolicited interactions.

Based on our results, some examples of simple controls to detect automated behavior are:

- more than one simultaneous logon of the same user (with some variations, like if you consider an user logged on the laptop and the smartphone on the same time), or several successful logins in a short time period;
- several simultaneous and/or continuous requests (not only search queries, but for any action) in a quantity or time frame higher than the average capacity of an human being;
- send several contact requests and/or messages to different users simultaneously or in a short time frame (also potentially indicating SPAM).

The enforcement of controls on these proposed behaviors do not need to be the block or cancelling of the request. Requiring additional fields like a Captcha, similar to what is already used to avoid brute force attacks, can be an excellent way to avoid the automation, as they would be required only for certain scenarios that will not affect most of the regular users.

Going a bit beyond, imagining the need of certain users/scenarios where some automation can be useful or required - for real recruiters, for example, the enforcement of controls can be more rigid through web browser (where regular users, not automation, is expected) and more flexible through API, for example. This will allow a better monitoring and control by the platform, even being able from a business perspective to offer a certain quantity of free requests for minor customers (like independent small headhunters) or more robust professional services sold by the platform, like the already existing LinkedIn Recruiter.

The proposed controls can solve the automation issue, enforcing already existing policies with minimum impact to users. However, a different challenge is how easy is to create fake profiles, a problem not only for LinkedIn but to any social network in the current days. It is possible to build a base of interactions and contacts that can create a sense of legitimacy, organically through real users or even through a network of other false profiles.

Verification of users should involve complex validation processes. But the impact of fake profiles have been growing so fast that discussions over mandatory user validation are already happening on other social networks like Twitter[2]. We have been suggesting to implement this kind of functionality since the beginning of this research project, and in a very recent update, LinkedIn announced a verification program.[3]. It still have some limitations, like validation through company email for some registered companies or through your documents (available only in United States), but it is certainly an improvement. Considering Twitter goals as a professional social network, credibility and veracity should be a matter of interest for all their users. Potentially even without an enforcement many users will potentially look for this validation as a way to recognize their work and responsibility - or at least some groups like recruiters can be targetted. There are several opportunities, each with pros and cons, but certainly some kind of control in this direction it's necessary to make at least a bit harder for personification attacks.

[2] https://help.twitter.com/en/managing-your-account/about-twitter-verified-accounts

[3] https://www.linkedin.com/help/linkedin/answer/a1359065/verifications-on-your-linkedin-profile

6 CONCLUSION

Cyber attacks have been exposing the vulnerabilities of computer networks and applications. Especially the context of the social networks, becoming each day more important in people's lives, are a promising scenario for several malicious actions, and the current defense mechanisms are not efficient to mitigate or avoid them, highlighting the exploitation of trust using bots.

ASE bots offer great scalability with no need for more exposure of the attacker. This dissertation presented an analysis of LinkedIn over the perspective of an ASE attack, evaluating the platform capacity to detect, respond and mitigate those attacks, understading the risk level exposure of their users. In order to achieve that, we used the development of a bot-based approach to simulate ASE attacks using job proposals as a bait. Through a fake recruiter profile in LinkedIn, is it possible to identify and contact potential victims using automated mechanisms, looking for leakage of personal or corporate data.

The research developed in this dissertation faced several ethical limitations, that impacted not only the execution of the tests, but also the results as a consequence. Besides the approach testing each step individually helped to have a observation of the automation detection mechanisms, the limitation for testing with human subjects bring a lack to the understanding of the effectiveness of a similar attack.

Different from generic SE attacks like some types of phishing, that usually have a low success rate per message sent, the attack proposed in this work can be more categorized as a targeted attack. Generic attacks usually create a bait and send it to the maximum amount of targets, as the success will depends on the link, interest or similar between the target and the storytelling behind the bait. Targeted SE attacks depends on a previous information about the targets, which allow the attacker to use a bait designed specifically for that group, and the attack is sent only to that target group. Due to that, targeted attacks have a higher success rate per message, which require from attackers to better plan the targets and allow them to not need to hit millions of targets to have an average number of victims. Professional social networks as LinkedIn are spaces planned for business interactions and recruitment, so the entire context around it bring credibility to the used bait - a job opportunity. Also, as the Search phase looks for profiles based on certain keywords, the targets will have a connection to the proposed job, not for example offering a developer position to a salesperson, for example.

The complete absence of controls demonstrate the potential for similar SE actions in real world, which should raise awareness and important actions to address those threats in a Information Security strategy. User awareness continue to be an essential resource to protect against SE attacks, but considering the fast evolution of bot platforms and AI technologies, bringing capacity to create improved ways to convince or manipulate users like deepfakes, highlight that other measures must be taken. Platforms must improve their controls to detect automated systems and take actions like tagging or even blocking them, decreasing the number of malicious interactions that approach the final user. Without the technological support human awareness will not be enough against the constant number of improved attacks.

The main contributions of this research are the evaluation of the defense and response mechanisms of LinkedIn against an ASE attack, highlighting how a lack of proper controls allows the usage of automation for malicious purposes against the users. As the SE topic combines Computer Science and Psychology fields, this work also was a way to focus on the technological lacks instead of user behavior, an important discussion to evaluate the level of responsibility of the platforms and services on the discussions on how to protect against SE attacks. The discussion over limitations also bring attention to ethical matters, raising several questions due to the differences between general offensive security research and attacks focused in SE. Complete answers for ethical research requires a multidisciplinary work that go beyond the goals of this dissertation, but this work raised some important topics to drive discussions, and at least a review of the current framework should happen to allow more solid results that could help to better understand the threats and create improved solutions to the problem.

6.1 Future Work

This dissertation offers the base for several additional research topics. Based on the development and results during this work, and the discussions with other researchers during presentations, some themes especially have potential for not only complementary results for a more complete view and understanding of SE/ASE risk, but also guidelines for research Information Security risks in general.

Improvements to the PoC Platform Bot would allow to map and evaluate the limits for automated activities supported by the platform, increasing the quantity of requests done to observe any differences on the behavior and the potential limits for an ASE attack.

Understanding the limitations over testing the Recruiter Bot using real subjects, have an evaluation outside an attack scenario, analyzing the capacity of similar chatbots to interact, gain trust and convince/manipulate a real person, like creating a trust connection necessary to conclude a job interview, could offer a deeper analysis for the last step of this attack life cycle validation.

This work focused on professional social networks and used LinkedIn as their use case. An evaluation of other similar platforms, and comparison to other social networks would allow a more clear view on the current overall capacity of them to identify and detect automated behavior.

The recent development and capacity of Large Language Models (LLM) and commercial AI platforms like OpenAI ChatGPT[1] and Google Bard[2] bring an entirely new horizon to the capacity of automation. Not only those technologies bring a powerful and easier toolset to support attackers, we have already documented cases of those platforms lying to humans in order to achieve their tasks[3]. Any further study considering those models and platforms will be of extreme relevance for the topic.

A deeper discussion over the current ethical framework for SE and offensive security research would be not only important to offer clear guidelines for further work on the topic, but also to potentially bring improvements that could allow more solid results to help create improved protection and awareness against cyber threats.

[1] https://openai.com/blog/chatgpt

[2] https://bard.google.com

[3] https://www.washingtonpost.com/business/2023/03/19/chatgpt-can-lie-but-it-s-only-imitating-humans/814706ee-c650-11ed-9cc5-a58a4f6d84cd_story.html

REFERENCES

AL-CHARCHAFCHI, A.; MANICKAM, S.; ALQATTAN, Z. N. Threats against information privacy and security in social networks: A review. In: SPRINGER. **International Conference on Advances in Cyber Security**. [S.l.], 2019. p. 358–372.

AROYO, A. M. et al. Trust and social engineering in human robot interaction. **IEEE Robotics and Automation Letters**, IEEE, v. 3, n. 4, p. 3701–3708, 2018.

ASSENMACHER, D. et al. Demystifying social bots: On the intelligence of automated social media actors. **Social Media+ Society**, SAGE Publications Sage UK: London, England, v. 6, n. 3, p. 2056305120939264, 2020.

BOSHMAF, Y. et al. The socialbot network: when bots socialize for fame and money. In: **Proceedings of the 27th annual computer security applications conference**. [S.l.: s.n.], 2011. p. 93–102.

CAMISANI-CALZOLARI, M. Analysis of twitter followers of the us presidential election candidates: Barack obama and mitt romney. **Online). http://digitalevaluations.com**, 2012.

CASTELLS, M. Communication power. nueva york: oxford university press. 2009.

CROSSLER, R.; BÉLANGER, F. An extended perspective on individual security behaviors. **ACM SIGMIS Database**, ACM New York, NY, USA, v. 45, n. 4, p. 51–71, 2014.

CULOT, G. et al. Addressing industry 4.0 cybersecurity challenges. **IEEE Engineering Management Review**, v. 47, n. 3, p. 79–86, 2019.

DARWISH, A.; ZARKA, A. E.; ALOUL, F. Towards understanding phishing victims' profile. In: **2012 International Conference on Computer Systems and Industrial Informatics**. [S.l.: s.n.], 2012. p. 1–5.

DEWANGAN, M.; KAUSHAL, R. Socialbot: Behavioral analysis and detection. In: SPRINGER. **International Symposium on Security in Computing and Communication**. [S.l.], 2016. p. 450–460.

DWYER, C.; HILTZ, S.; PASSERINI, K. Trust and privacy concern within social networking sites: A comparison of facebook and myspace. **AMCIS 2007 proceedings**, p. 339, 2007.

FBI, I. C. C. C. **Federal Bureau of Investigation Internet Crime Report 2022**. 2023.

FERRARA, E. et al. The rise of social bots. **Communications of the ACM**, ACM New York, NY, USA, v. 59, n. 7, p. 96–104, 2016.

FREITAS, C. et al. Reverse engineering socialbot infiltration strategies in twitter. In: IEEE. **IEEE/ACM ASONAM 2015)**. [S.l.], 2015. p. 25–32.

FREITAS, C.; BENEVENUTO, F.; VELOSO, A. Socialbots: Implicações na segurança e na credibilidade de serviços baseados no twitter. **SBRC, Santa Catarina, Brasil**, p. 603–616, 2014.

GALLEGOS-SEGOVIA, P. L. et al. Social engineering as an attack vector for ransomware. In: IEEE. **CHILECON 2017**. [S.l.], 2017. p. 1–6.

GREITZER, F. L. et al. Positioning your organization to respond to insider threats. **IEEE Engineering Management Review**, IEEE, v. 47, n. 2, p. 75–83, 2019.

GRIMME, C. et al. Social bots: Human-like by means of human control? **Big data**, Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA, v. 5, n. 4, p. 279–293, 2017.

GUZMAN, A. L.; LEWIS, S. C. Artificial intelligence and communication: A human–machine communication research agenda. **New Media & Society**, SAGE Publications Sage UK: London, England, v. 22, n. 1, p. 70–86, 2020.

HEPP, A. Artificial companions, social bots and work bots. **Media, Culture & Society**, SAGE Publications Sage UK: London, England, v. 42, n. 7-8, p. 1410–1426, 2020.

HUBER, M. et al. Towards automating social engineering using social networking sites. In: IEEE. **2009 International Conference on Computational Science and Engineering**. [S.l.], 2009. v. 3, p. 117–124.

KHAN, R.; DAS, A. Build better chatbots. **A complete guide to getting started with chatbots**, Springer, 2018.

KIMPE, L. D. et al. Help, i need somebody: Examining the antecedents of social support seeking among cybercrime victims. **Computers in Human Behavior**, Elsevier, v. 108, p. 106310, 2020.

KLIMBURG-WITJES, N.; WENTLAND, A. Hacking humans? social engineering and the construction of the “deficient user” in cybersecurity discourses. **Science, Technology, & Human Values**, SAGE Publications Sage CA: Los Angeles, CA, v. 46, n. 6, p. 1316–1339, 2021.

LIBICKI, M. Could the issue of dprk hacking benefit from benign neglect? **Georgetown Journal of International Affairs**, Georgetown University Press, v. 19, p. 83–89, 2018.

MESSIAS, J.; BENEVENUTO, F.; OLIVEIRA, R. Bots sociais: Como robôs podem se tornar pessoas influentes no twitter? **Revista Eletrônica de Iniciação Científica em Computação**, v. 16, n. 1, 2018.

MITNICK, K. D.; SIMON, W. L. **The art of deception**. [S.l.]: John Wiley & Sons, 2003.

PARADISE, A.; SHABTAI, A.; PUZIS, R. Detecting organization-targeted socialbots by monitoring social network profiles. **Networks and Spatial Economics**, Springer, v. 19, n. 3, p. 731–761, 2019.

PIOVESAN, L. G. et al. Engenharia social: Uma abordagem sobre phishing. **REVISTA CIENTÍFICA DA FACULDADE DE BALSAS**, v. 10, n. 1, p. 45–59, 2019.

RIECKEN, H. W. Obedience to authority. an experimental view. stanley milgram. harper and row, new york, 1974. xx, 224 pp., illus. 10. **Science**, American Association for the Advancement of Science, v. 184, n. 4137, p. 667–669, 1974. ISSN 0036-8075. Disponível em: <https://science.sciencemag.org/content/184/4137/667>.

ROUSE, M. What is socialbot? **WhatIs.com**, 2013.

SALAHDINE, F.; KAABOUCH, N. Social engineering attacks: a survey. **Future Internet**, Multidisciplinary Digital Publishing Institute, v. 11, n. 4, p. 89, 2019.

SHAFAHI, M.; KEMPERS, L.; AFSARMANESH, H. Phishing through social bots on twitter. In: IEEE. **2016 IEEE International Conference on Big Data**. [S.l.], 2016. p. 3703–3712.

SHIRES, J. Enacting expertise: Ritual and risk in cybersecurity. **Politics and Governance**, v. 6, n. 2, p. 31–40, 2018.

STOECKLI, E.; UEBERNICKEL, F.; BRENNER, W. Exploring affordances of slack integrations and their actualization within enterprises-towards an understanding of how chatbots create value. In: **Proceedings of the 51st Hawaii International Conference on System Sciences**. [S.l.: s.n.], 2018.

TIOH, J.-N.; MINA, M.; JACOBSON, D. W. Cyber security social engineers an extensible teaching tool for social engineering education and awareness. In: IEEE. **2019 IEEE Frontiers in Education Conference (FIE)**. [S.l.], 2019. p. 1–5.

TIWARI, V. Analysis and detection of fake profile over social network. In: IEEE. **ICCCA 2017**. [S.l.], 2017. p. 175–179.

YE, W.; LI, Q. Chatbot security and privacy in the age of personal assistants. In: IEEE. **IEEE/ACM SEC 2020**. [S.l.], 2020. p. 388–393.

APPENDIX A — PUBLISHED PAPER - SBSEG 2022

In this appendix the paper entitled "Ataques Automatizados de Engenharia Social com o uso de *Bots* em Redes Sociais Profissionais" is presented. This was the first deliverable of the research presented in this dissertation which introduced the key concepts of this research, the theoretical baseline and an initial version of the proof of concept bot, with simplified tests and functionalities. It also raised the discussion over the ethical challenges faced by this research topic.

- **Title:** "Ataques Automatizados de Engenharia Social com o uso de *Bots* em Redes Sociais Profissionais"
- **Conference:** "XXII Simpósio Brasileiro em Segurança da Informação e de Sistemas Computacionais (SBSeg 2022)"
- **URL:** https://sol.sbc.org.br/index.php/sbseg/issue/view/984
- **Date:** September 12-15, 2022
- **Venue:** Santa Maria Federal University - Santa Maria, RS, Brazil
- **Digital Object Identifier (DOI):** https://doi.org/10.5753/sbseg.2022

Ataques Automatizados de Engenharia Social com o uso de *Bots* em Redes Sociais Profissionais

Maurício Ariza[1], **Antônio João G. de Azambuja**[1],
Jéferson C. Nobre[1], **Lisandro Z. Granville**[1]

[1]Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)
Porto Alegre – RS – Brazil

{mariza, antonio.azambuja, jcnobre, granville}@inf.ufrgs.br

Abstract. *Virtual human interactions have been intensified with the increasing use of the Internet and social networks, raising the risk of Social Engineering cyber threats. The usage of Bots in those attacks allow scalability in the exploitation of users trust, causing security risks. There are few papers focusing on automated Social Engineering actions using Bots. This paper presents an assessment of the controls used in a professional social network to identify and block automated attacks, using a Bot as a proof of concept. The analysis and discussion of the results allow demonstrating the security vulnerabilities present in professional networks that can be exploited to build a trust relationship between the user and a malicious Bot.*

Resumo. *As interações humanas virtuais têm sido ampliadas com o uso crescente da Internet e redes sociais, elevando os riscos de ameaças cibernéticas de Engenharia Social. O uso de Bots nesses ataques permite escalabilidade na exploração da confiança dos usuários, provocando riscos de segurança. Poucos são os trabalhos com foco nas ações automatizadas de Engenharia Social com o uso de Bots. Este artigo apresenta uma verificação dos controles de uma rede social profissional quanto à identificação e bloqueio desses ataques automatizados, utilizando um Bot de prova de conceito. A análise e discussão dos resultados permite demonstrar as vulnerabilidades de segurança presentes nas redes profissionais que podem ser exploradas para construção da relação de confiança do usuário com um Bot malicioso.*

1. Introdução

Ataques cibernéticos exploram as vulnerabilidades das estruturas de Tecnologia da Informação e Comunicação (TIC), incluindo as redes sociais, que se estabeleceram ao longo dos anos como ferramentas de interação humana [Shires 2018]. No entanto, essas redes sociais emergem desafios e riscos relacionados à Segurança Cibernética (SegCiber) [Klimburg-Witjes and Wentland 2021].

As redes sociais fornecem serviços *on-line*, coletam dados pessoais e corporativos formando uma base de dados de alto valor, sendo passível de ser utilizada como ferramentas para ataques cibernéticos que exploram as relações de confiança [Crossler and Bélanger 2014].

Essas relações de confiança no ambiente cibernético têm proporcionado um cenário para a prática de atos ilícitos, implicando em riscos de SegCiber. Os ataques têm explorado a interação humana em conjunto com as brechas tecnológicas, enfraquecendo a cadeia de segurança [Salahdine and Kaabouch 2019] [Klimburg-Witjes and Wentland 2021]. Na prática, o fator humano é o elo mais fraco na cadeia de SegCiber [Mitnick and Simon 2003]. A interconectividade das redes sociais e o crescimento da dimensão cognitiva do trabalho estão tornando os recursos humanos como um dos pilares da segurança [Culot et al. 2019] [Greitzer et al. 2019].

As organizações têm empregado soluções de defesa para enfrentar os ataques cibernéticos, tais como *Firewalls*, Sistema de Detecção de Intrusão (*Intrusion Detection System* - IDS), Sistema de Prevenção a Intrusão (*Intrusion Prevention System* - IPS) e Antivírus. No entanto, esses mecanismos de defesa não têm sido suficientes para impedir integralmente as ações de Engenharia Social (ES) no ambiente cibernético. Os atacantes que utilizam ES vêm adotando mecanismos automatizados para explorar as relações de confiança, tendo como objetivo obter dados e informações relevantes de potenciais alvos. Os ataques automatizados requerem pouca intervenção humana e são desenvolvidos visando simular o comportamento humano [Huber et al. 2009] [Shafahi et al. 2016].

O crescente aumento do uso das redes sociais para estabelecer relacionamentos pessoais e profissionais abre um campo para as ações de *Bots* de Engenharia Social Automatizada (ESA) [Huber et al. 2009]. Os *Bots* são *softwares* automatizados, que por vezes utilizam recursos como inteligência artificial (IA), e são capazes de executar comandos de operação e controle, sem a necessidade de participação humana. São ferramentas com a capacidade de se passar por seres humanos, imitando as atividades dos usuários reais [Shafahi et al. 2016].

Bots podem ser utilizados para ações positivas, como por exemplo, ajudar o usuário na sua experiência *on-line*. Para os autores [Dickerson et al. 2014] os humanos tendem a confiar nos *Bots*. Contudo, os *Bots* de ESA têm sido utilizados como uma ferramenta para ataques de ES, já que são escaláveis, permitindo que um único atacante contate um grande número de potenciais vítimas simultaneamente, na busca de informações confidenciais [Huber et al. 2009][Dewangan and Kaushal 2016].

Na literatura, poucos trabalhos apresentam análises sobre a ESA com o uso de *Bots*. A maioria dos trabalhos estuda a área da psicologia social, com foco no comportamento humano diante das ações de ES [Huber et al. 2009]. Os autores [Al-Charchafchi et al. 2019] e [Piovesan et al. 2019] abordam as ameaças à segurança nas redes sociais, decorrentes dos ataques de ES utilizando contas falsas, roubo de identidade e *phishing*. No sentido de influenciar os usuários nas redes sociais, há trabalhos que avaliam as vulnerabilidades das redes sociais com o uso de *SocialBots* para campanhas de convencimento nas redes. Os autores [Freitas et al. 2014] e [Messias et al. 2018] analisam o uso de *Bots* no *Twitter* para influenciar os usuários e comprometer a estrutura da rede. Já [Huber et al. 2009] propõem a automação das tarefas de ES por meio de um *Bot* no *Facebook*, concluindo que a persuasão é um recurso essencial no processo de ESA.

Nesse contexto onde o uso de *Bots* de ESA permite automação e escalabilidade dos ataques com menor exposição do agente malicioso em si, este trabalho se propõe a analisar a mais popular rede social profissional atualmente, o *LinkedIn*, e verificar se a mesma oferece controles que possam impedir ou dificultar a ação automatizada de ataques de Engenharia Social. As principais contribuições desse trabalho são: i) avaliar a capacidade de detecção e bloqueio de ataques automatizados por parte da rede social *LinkedIn*;

ii) implementar uma prova de conceito para validar a viabilidade técnica desses ataques; e iii) propor melhorias que possam ser utilizadas por essas redes a fim de diminuir os riscos de Engenharia Social Automatizada aos seus usuários.

O artigo inicialmente aborda, na Seção 2, os conceitos relacionados com a teoria para embasar a pesquisa. Na Seção 3, analisa os trabalhos relacionados com o tema da pesquisa. A seguir, na Seção 4, discorre sobre a apresentação do problema e o método de ataque. Na Seção 5, apresenta o protótipo, experimento e discussão dos resultados. A seguir, na Seção 6 as limitações da pesquisa são mencionadas. Por fim, apresenta-se na Seção 7 a conclusão e uma abordagem para trabalhos futuros.

2. Referencial Teórico

Embora sejam limitados os trabalhos no tema do uso de ataques automatizados de Engenharia Social com o uso de *Bots*, existem publicações relacionadas ao tema que ajudam a embasar teoricamente as premissas utilizadas para condução do trabalho e proposta.

2.1. Engenharia Social Automatizada

A ES refere-se à exploração do comportamento humano no tocante ao uso dos sistemas de informações para obtenção de dados e informações relevantes de potenciais alvos. Os ataques de ES colocam o atacante em uma posição favorecida no fluxo de informações, tirando proveito de uma relação de confiança [Mitnick and Simon 2003]. O desenvolvimento de uma relação de confiança faz uso da manipulação psicológica induzindo as pessoas realizarem ações específicas. Conquistar a confiança das vítimas é um objetivo dos engenheiros sociais.

Esses ataques podem ser detectados, no entanto, não são facilmente interrompidos [Libicki 2018]. As técnicas de ES direcionam os usuários a responderem solicitações sem uma análise adequada as informações disponibilizadas, seguindo 4 (quatro) estágios, a saber: i) obter as informações sobre a vítima para uma primeira abordagem; ii) estabelecer uma relação de confiança entre o atacante e a vítima; iii) explorar as informações para o desenvolvimento de ações específicas; e iv) executar o ataque para alcançar o(s) seu(s) objetivo(s) [Mitnick and Simon 2003] [Tioh et al. 2019].

Os ataques de ES demandam tempo e recursos para estabelecer um relacionamento de confiança. No entanto, o desenvolvimento de uma *interface* homem-máquina permite que tais relacionamentos sejam automatizados [Guzman and Lewis 2020]. Os engenheiros sociais podem utilizar a automação para desenvolver ferramentas pré-programadas para realizar tarefas sem a intervenção humana, possibilitando a escalabilidade dos ataques de ES [Huber et al. 2009] [Shafahi et al. 2016].

Os ataques automatizados podem ser preparados utilizando informações de valor e/ou influenciando determinados grupos nas redes sociais [Mitnick and Simon 2003] [Gallegos-Segovia et al. 2017]. Essas redes representam um espaço virtual atrativo para os atacantes explorarem as vulnerabilidades técnicas e a falta de conhecimento e conscientização dos usuários sobre ações de ES [Al-Charchafchi et al. 2019]. Uma das vulnerabilidades que são encontradas em redes sociais é a criação de perfis falsos, os quais constituem um percentual significativo dos usuários dessas redes [Tiwari 2017]. O Relatório de Investigação de Violação de Dados, publicado em 2021, descreve que 40 %

dos casos de violação dos dados tem relação com as ações de ES [1].

2.2. *Bots*

Bot é o termo resumido da palavra da língua inglesa *Robot*, que na tradução livre significa Robô. É uma ferramenta automatizada que realiza uma série de funções pré-programadas de operação e controle. Os *Bots* podem ser autênticos, que têm como objetivo realizar atividades úteis para os usuários, por outro lado também existem *Bots* de cunho malicioso, que podem realizar ataques para obter informações relevantes ou manter o controle do dispositivo acessado. *Bots* podem ser utilizados para ações de disseminação de informações falsas (*fake news*), *spam* e *phishing* [Freitas et al. 2015].

No contexto de ESA os cibercriminosos usam os *Bots* maliciosos para simular o comportamento humano, burlando os mecanismos de segurança. Com o crescimento das redes sociais e o grande volume de dados no ciberespaço, os engenheiros sociais passaram a espalhar *Bots* com comportamento semelhante ao do ser humano para um grande número de usuários. Esses *Bots* simulam conversas humanas, conhecidos como *ChatBots* e, os que atuam nas redes sociais, os *SocialBots* [Shafahi et al. 2016].

ChatBot é a integração de sistemas, ferramentas e roteiros que promovem conversas por mensagens instantâneas com ou sem a participação de humanos [Stoeckli et al. 2018]. São desenvolvidos para ajudar usuários humanos em situações de serviços específicos, não sendo exaustivo. Por exemplo: atendimento ao cliente, atendimento por telefone e serviço de educação digital [Grimme et al. 2017]. O uso da linguagem natural nos *ChatBots* é um desafio a ser superado para o desenvolvimento dessa ferramenta [Khan and Das 2018].

SocialBot é uma ferramenta de *software* que simula o comportamento humano para realizar interações automatizadas nas redes sociais [Rouse 2013]. Os *SocialBots* têm a capacidade de comprometer a estrutura das redes sociais, influenciando os usuários e aumentando o número de seguidores, para inflar os índices de popularidade de uma determinada conta de perfil [Camisani-Calzolari 2012]. Essa ferramenta é eficaz para ataques de ES, utilizando-se de informações sensíveis de possíveis vítimas, como o roubo de identidade [Dewangan and Kaushal 2016].

Essas ferramentas têm sido desenvolvidas com a ajuda de mecanismos de IA que interagem com os usuários [Freitas et al. 2015]. A IA é similar a inteligência humana, desenvolvida com a automatização conforme a necessidade da aplicação [Ferrara et al. 2016]. Na medida que um certo grau de inteligência é incorporado nas ferramentas para simular o comportamento humano, aumenta a capacidade e escalabilidade dos ataques.

3. Trabalhos relacionados

A proposta de desenvolvimento de um *Bot* automatizado para ataques de ES demanda uma análise estruturada da literatura para apoiar este estudo. A análise permitiu identificar que parte dos estudos de ES têm foco no comportamento humano diante das ações de *SocialBots*, *phishing* e *spam*.

[1] https://enterprise.verizon.com/resources/reports/2021/2021-data-breach-investigations-report.pdf

Os autores [Dewangan and Kaushal 2016] abordam o uso de *SocialBots* em campanhas políticas e *marketing* de produtos. No trabalho os autores apontam os riscos de segurança atrelados a essa prática, no que diz respeito ao acesso às informações pessoais dos usuários. Diante dos riscos o trabalho menciona a necessidade de identificação desses *SocialBots*. Esse procedimento permite assegurar a reputação de uma rede social que está sendo objeto do ataque. Para tanto, os autores desenvolveram um modelo de detecção desses *SocialBots* considerando a análise de comportamento.

O trabalho dos autores [Aroyo et al. 2018], discorre como a ES explora a relação de confiança entre os usuários e *Bots*. Com base no modelo de [Mitnick and Simon 2003] foi desenvolvido um *Bot* para simular um ataque de ES. Inicialmente o *Bot* buscou obter informações com perguntas de cunho privado. Na sequência estabeleceu uma relação de confiança com os participantes, por meio dos *Bots*, para uma aproximação anônima com o alvo. Nos resultados do estudo os participantes, na sua maioria (62%), demonstraram confiança na ferramenta mencionando o comportamento ético, tendo em vista que foi desenvolvida considerando questões éticas.

O artigo *Threats Against Information Privacy and Security in Social Networks: A Review* [Al-Charchafchi et al. 2019], apresenta uma revisão das pesquisas sobre privacidade e ameaças à segurança nas redes sociais. Para os autores, no que pese a literatura apresentar trabalhos sobre privacidade, mais esforços são necessários. O ambiente das redes sociais é uma rica fonte de dados pessoais, tornando-se um atrativo para ações dos engenheiros sociais, que exploram a falta de conscientização e conhecimento dos usuários nas questões relacionadas com a segurança.

Para [Al-Charchafchi et al. 2019] a complexidade dos ataques de ES está alinhada com a possibilidade da combinação das estratégias sociais e técnicas para realizar um crime cibernético. Nessa linha os autores [Piovesan et al. 2019] afirmam que políticas de segurança podem oferecer maior nível da segurança da informação, no entanto não garantem proteção completa.

Os trabalhos dos autores [Freitas et al. 2014], [Messias et al. 2018] e [Shafahi et al. 2016], discutem o impacto do uso dos *SocialBots* no *Twitter* para caracterizar o comportamento da ferramenta em uma grande base de dados, medir a capacidade da ferramenta influenciar os usuários na rede, detectar automaticamente esses *Bots* e analisar os riscos de segurança decorrentes do uso de *phishing*, por meio de *SocialBots* no *Twitter*.

Nos resultados os autores [Freitas et al. 2014], destacam que o método por eles desenvolvido para caracterizar e detectar os *SocialBots*, teve um indicador de detecção com 92% de sucesso. Os autores [Messias et al. 2018], afirmam que um simples *Bot* pode alcançar altos níveis de influência no *Twitter*. Já [Shafahi et al. 2016], apontam a necessidade de aumentar o nível de conscientização sobre as ações de *phishing* que utilizam *SocialBots*. Os autores afirmam que essas ações constituem uma ameaça para as organizações.

Por fim, o trabalho dos autores [Huber et al. 2009], apresenta o ciclo de um ataque de ESA, com o uso de um *Bot*. O ataque demonstrou como as redes sociais podem ser utilizadas pelos engenheiros sociais para obter informações. Para tanto, no trabalho foram realizados 2 (dois) experimentos. O primeiro analisou a capacidade do *Bot* em obter informações nas redes sociais. O segundo realizou o *Turing Test* [Turing 2009], que busca avaliar a capacidade de uma máquina imitar um ser humano.

Para os autores a ESA com *Bots* é escalável e requer menos recursos humanos. A ferramenta foi utilizada em uma prova de conceito no *Facebook*. Os 2 (dois) experimentos permitiram ratificar que é possível automatizar ações de ES para obter informações e demonstrar que o *Bot* utilizado não foi identificado pelas medidas de segurança do *Facebook*. O número crescente das interações sociais dos usuários nas redes, torna os *Bots* de automação de ES uma ferramenta interessante para os engenheiros sociais.

4. Solução Proposta

Ataques de ES em redes sociais já são conhecidos e documentados, pois são espaços de interação cujas características despertam grande interesse para agentes maliciosos. Em especial, a capacidade de personificar com facilidade algum personagem que possa ganhar a confiança da vítima [Crossler and Bélanger 2014].

Diferentemente de outras redes sociais como *Facebook*, *Twitter* e *Instagram*, as redes sociais profissionais promovem uma atmosfera de ambiente corporativo, focada em conexões e relacionamentos para crescimento na carreira. Essas redes despertam o interesse de recrutadores e empresas na busca por candidatos para suas vagas e perfis de clientes em potencial. Nesse contexto, elas apresentam um cenário que inspira maior credibilidade e confiança entre os seus usuários, tornando esse grupo alvos em potencial para ataques direcionados e complexos.

Embora não referenciada em trabalhos acadêmicos, uma forma de ataque de ES existente nessas redes se caracteriza pela criação de um perfil falso na rede profissional por um atacante, entrando então em contato com potenciais vítimas se apresentando como recrutador para uma oportunidade de trabalho[2]. A partir do interesse da vítima, o atacante realiza entrevistas no intuito de roubar informações. Algumas formas comuns são descobrir informações confidenciais de empresas ou projetos onde a vítima tenha trabalhado, ou, ao final do processo, oferecer um falso contrato de trabalho, solicitando dados pessoais e uma cópia do passaporte ou documento similar, informações que podem ser utilizadas para roubo de identidade. Todo o processo envolvido e o contato com a vítima são feitos de forma manual pelo atacante.

A ES tem sua questão central no campo da psicologia, tendo a computação como uma ferramenta para viabilizar o trabalho do atacante. O nosso trabalho busca implementar uma prova de conceito para validar a viabilidade técnica de ataques de ESA pela ausência ou insuficiência de controles de segurança nas redes sociais profissionais, fato que abre brechas para o trabalho de agentes maliciosos.

A Seção 8.2 da Política de Uso do *LinkedIn* [3] especifica quais ações são permitidas ou proibidas na plataforma, como o uso de informações falsas no perfil ou o uso de *Bots* e *Scripts*. Porém, parte dessas regras são aplicadas apenas através de denúncias por parte de outros usuários, e não por controles técnicos.

A ausência de formas de controle ou validação permite um usuário identificar-se

[2] https://www.forbes.com/sites/reneemorad/2017/06/30/how-to-avoid-the-latest-linkedin-scam/?sh=13e1d13849c1

[3] https://www.linkedin.com/legal/user-agreement#dos

como funcionário de qualquer empresa, mencionar habilidades em diferentes campos de estudo ou construir um perfil que possa ser do seu interesse. Embora proibidos, códigos de automatização são amplamente utilizados, sendo possível encontrá-los em repositórios públicos como o *GitHub*. Levando em conta apenas essas duas questões, um atacante pode: (i) criar um perfil que atraia o interesse de seus alvos; e (ii) utilizar técnicas de automatização para aumentar a escalabilidade do seu ataque, sem a necessidade de burlar os mecanismos de controle da plataforma.

4.1. Método de Ataque

Tendo como referência o trabalho feito por [Huber et al. 2009], buscamos na pesquisa validar se os mecanismos e processos de controle da rede social *LinkedIn* são capazes de identificar e bloquear um ataque de ES automatizado, onde as ações do fluxo de ataque são realizados utilizando um *Bot* visando diversos alvos simultaneamente, sem a interação manual direta entre o atacante e a vítima.

O método para testar a proposta utiliza como base o ataque apresentado no início da Seção 4, onde o atacante apresenta-se como um recrutador. Para isso, foi criado um perfil falso e um *Bot* com o objetivo de: i) realizar a busca de perfis de potenciais vítimas; ii) adicionar de forma automatizada uma grande quantidade de vítimas como contatos; e iii) enviar de forma simultânea mensagens às vítimas oferecendo uma falsa oportunidade de trabalho como chamariz.

O *Bot* precisa, portanto, ser capaz de identificar um grande número de usuários simultaneamente a partir de palavras-chave de interesse do atacante e entrar em contato com todos eles, sem ser identificado e bloqueado pelos controles da rede social. Para tanto, é necessário uma infraestrutura técnica, com a combinação de uma plataforma de rede social e requisitos para automação do comportamento de uma conta, utilizando uma *Application Programming Interface* (API - Interface de Programação de Aplicativo, tradução livre) ou mecanismos proprietários para interagir com a plataforma [Assenmacher et al. 2020].

5. Avaliação

5.1. Protótipo

O *LinkedIn* oferece uma API bastante completa para interação com a plataforma, sendo, portanto o caminho natural para uma interação feita através de *software*. Por decisão dos pesquisadores optou-se por replicar o comportamento de um usuário padrão via navegador realizando diversas ações simultâneas, caracterizando claramente o uso de automatização e violação das políticas de uso da rede social.

Para verificar a viabilidade do ataque, os autores então desenvolveram uma aplicação de prova de conceito em linguagem *Python* para interagir com a rede social, utilizando a biblioteca *Selenium* a fim de realizar as requisições diretamente através de um navegador.[4].

Foi criado um perfil falso do atacante utilizando informações aleatórias, valendo-se da ausência de validação das informações pela rede social. Dados como graduações e

[4] Por questões éticas na publicação de uma ferramenta de ataque, os autores optaram por manter o código da mesma em um perfil privado do GitHub, podendo fornecer acesso sob requisição.

diplomas, nível de conhecimento e experiências profissionais atuais e prévias, associando o indivíduo diretamente à empresas e instituições verdadeiras, podem ser registrados pelo atacante sem dificuldade. Toda essa construção auxilia na demonstração da veracidade permitindo passar confiança às vítimas.

O protótipo considerou um modelo para os estágios do ataque de ES [Mitnick and Simon 2003], sendo estruturado em 3 (três) etapas: i) Autenticação: para o acesso à rede social; ii) Busca: para mapeamento dos alvos; e iii) Abordagem: para contatar as vítimas. As etapas são detalhadas nos parágrafos a seguir.

1ª. Etapa de Autenticação: Essa fase busca verificar se a rede social é capaz de detectar o processo de autenticação de um usuário sendo executado de forma automatizada. Para isso, o *Bot* desenvolvido inicia o navegador e abre a página do *LinkedIn*, sendo automaticamente direcionado à página de *login*. O código-fonte da página é então mapeado para que os campos de autenticação sejam identificados. São solicitados o usuário e a senha do perfil do atacante, os quais são introduzidos diretamente nos campos apropriados a fim de acessar o página principal de um usuário autenticado.

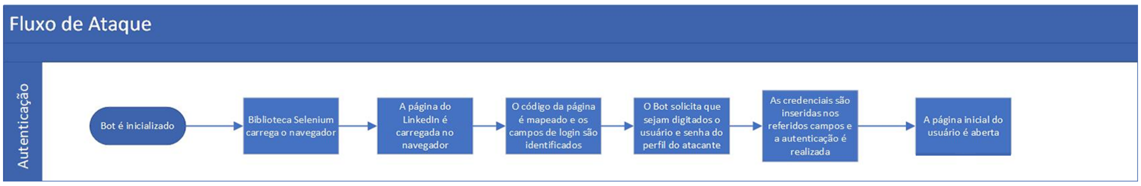

Fluxo de Ataque
Autenticação
Bot é inicializado
Biblioteca Selenium carrega o navegador
A página do LinkedIn é carregada no navegador
O código da página é mapeado e os campos de login são identificados
O Bot solicita que sejam digitados o usuário e senha do perfil do atacante
As credenciais são inseridas nos referidos campos e a autenticação é realizada
A página inicial do usuário é aberta

Figura 1. Fluxo de ataque - Etapa de Autenticação

2ª. Etapa de Busca: Essa fase busca verificar se a rede social é capaz de detectar a realização de diversas buscas realizadas simultaneamente de forma automatizada. Para isso, o *Bot*, já com o usuário do atacante autenticado, recebe palavra(s)-chave de busca. Novamente o código-fonte da página é mapeado, o campo de busca na *interface* do usuário é identificado e os termos são introduzidos no mesmo, fazendo com que o *LinkedIn* retorne uma lista de perfis baseada naqueles critérios. Embora exista uma relação entre o termo buscado e os resultados, a quantidade e ordem dos perfis exibidos como resultados da busca são definidos unicamente pelo algoritmo da própria rede social. Portanto para os objetivos desse trabalho os resultados da pesquisa em si são armazenados meramente para criação de um banco de dados de alvos pelo atacante, não fazendo parte desse escopo analisar o algoritmo da rede social e seu comportamento quanto ao retorno de resultados.

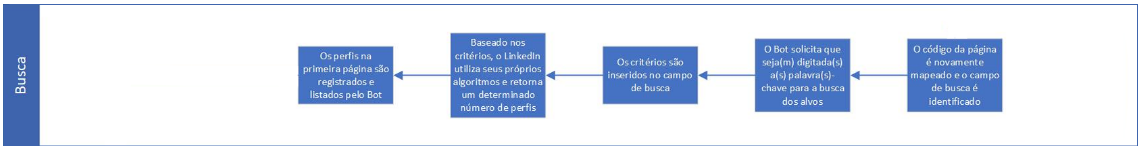

Busca
O código da página é novamente mapeado e o campo de busca é identificado
O Bot solicita que seja(m) digitada(s) a(s) palavra(s)-chave para a busca dos alvos
Os critérios são inseridos no campo de busca
Baseado nos critérios, o LinkedIn utiliza seus próprios algoritmos e retorna um determinado número de perfis
Os perfis na primeira página são registrados e listados pelo Bot

Figura 2. Fluxo de ataque - Etapa de Busca

3ª. Etapa de Abordagem: Essa fase busca verificar se a rede social é capaz de detectar o envio simultâneo de mensagens a diferentes usuários de forma automatizada. Para isso, o *Bot* recebe uma mensagem personalizada que será enviada para as vítimas. Mapeando o código-fonte da página onde se apresentam os resultados da pesquisa feita na etapa anterior, para cada uma das vítimas é identificado e acionado o botão para solicitar conexão com a mesma, a mensagem personalizada é introduzida no conteúdo da solicitação, utilizando parâmetros customizados para garantir que cada mensagem chame a vítima pelo próprio nome. Esse procedimento é repetido para todos os demais perfis listados.

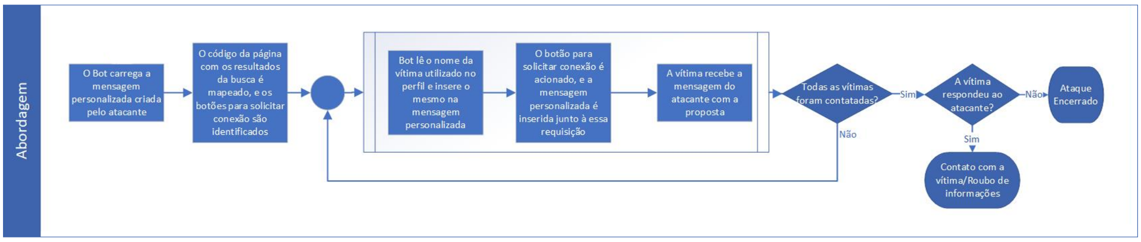

Abordagem
O Bot carrega a mensagem personalizada criada pelo atacante
O código da página com os resultados da busca é mapeado, e os botões para solicitar conexão são identificados
Bot lê o nome da vítima utilizado no perfil e insere o mesmo na mensagem personalizada
O botão para solicitar conexão é acionado, e a mensagem personalizada é inserida junto à essa requisição
A vítima recebe a mensagem do atacante com a proposta
Todas as vítimas foram contatadas?
Sim
Não
A vítima respondeu ao atacante?
Não
Sim
Ataque Encerrado
Contato com a vítima/Roubo de informações

Figura 3. Fluxo de ataque - Etapa de Abordagem

O uso de palavras-chave que identifiquem determinados tipos de profissionais - como um cargo ou habilidade técnica específica - pode ser uma forma de refinar os resultados da busca e definir um tipo de perfil específico de alvos. Porém, do ponto de vista do ataque a determinação de termos exatos a serem utilizados é de menor relevância, visto que o objetivo do atacante é mapear o maior número possível de vítimas.

5.2. Experimentos

Os testes foram realizados de acordo com as etapas de autenticação, busca e abordagem, detalhadas no Fluxo de Ataque. O código foi testado em uma máquina *Windows*, utilizando a versão 3.9 do *Python*, a versão 3.141.0 da biblioteca *Selenium* com o *driver* na versão 90.0.4430.24 do *driver* para o navegador *Google Chrome*, que por sua vez estava na versão 90.0.4430.72.

Testes da 1ª. Etapa - Autenticação: Sendo o objetivo dessa etapa verificar a capacidade da rede social identificar a autenticação de forma automatizada, configuramos o *Bot* de testes com as credenciais da conta criada para o atacante. O mesmo foi capaz de realizar a autenticação com sucesso mesmo com o processo sendo repetido diversas vezes em sequência ou de forma simultânea sem exibição de mensagens de erro, solicitação de controles do tipo *Captcha* ou quaisquer indícios que a rede tenha detectado aquele acesso como tendo sido realizado de forma automatizada, reforçando que pelo uso da biblioteca *Selenium* todas as ações são realizadas diretamente através do navegador. Para fins comparativos, os testes foram repetidos com variações como credenciais inseridas manualmente na execução, credenciais lidas em arquivo e uso de credenciais inválidas, não tendo sido observados diferenças de resultado além de sucesso quanto utilizadas credenciais válidas e erro quando utilizadas inválidas.

Testes da 2ª. Etapa - Busca: Para validar a resposta da rede social quanto à execução de buscas de forma automatizada, o *Bot* foi alimentado com diferentes palavras-chave, sendo executadas buscas de forma contínua e simultânea a fim de verificar alterações de comportamento por parte da mesma que pudessem indicar detecção de comportamento automatizado, não tendo sido observado nenhuma ação por parte dos pesquisadores. Não foram executados testes de estresse/carga pois entende-se que o objetivo de um atacante seja atingir o maior número possível de vítimas sem ser identificado, e não de causar negação de serviço na aplicação.

Foram realizados em menor escala alguns testes quanto à precisão do resultados, utilizando os termos ”teste”, ”Engenharia Social”, ”Bot”, ”Redes Sociais”, ”Segurança da Informação”e ”Python”, verificando uma amostra dos perfis retornados a fim de verificar a relevância e as diferenças quanto à quantidade de perfis retornados. É importante salientar porém que o objetivo dos pesquisadores foi identificar potencial mudança de comportamento por conta das buscas estarem sendo executadas através de um *Bot*, pois os resultados das buscas em si são resultado do próprio algoritmo do *LinkedIn*, não tendo influência direta por parte dos pesquisadores além do termo de pesquisa utilizado e sendo fora do escopo deste trabalho testar a capacidade de resposta do mesmo.

Testes da 3ª. Etapa - Abordagem: A terceira e última etapa seria verificar se ocorreria a detecção com a abordagem de indivíduos. Esta etapa era a mais diretamente afetada pelas questões éticas envolvidas nesse trabalho, que são discutidas em mais detalhes na Seção 6. Para execução dos testes, o *Bot* capturava os perfis retornados nas buscas da etapa anterior - a fim de limitar o número de alvos o código foi configurado para filtrar apenas os resultados presentes na primeira página de busca, sendo entre 15 e 21 perfis por palavra-chave. A cada rodada todos os alvos eram contatados simultaneamente, recebendo uma solicitação de conexão e uma mensagem personalizada, tendo sido verificado também a execução de duas rodadas simultaneamente. Em nenhum momento foi identificado novamente ações por parte da rede social. Após o envio, os pedidos de conexão e as mensagens eram automaticamente cancelados e excluídos antes da interação com os alvos ocorrer.

5.3. Discussão

Uma das contribuições deste trabalho foi a busca pela validação da hipótese de ausência de controles por parte das redes sociais, que embora seja conhecida no meio da tecnologia, não encontramos referências na academia.

Como mencionado na Seção 4, não é um dos objetivos desta pesquisa analisar as vulnerabilidades dos usuários em si e os aspectos psicológicos explorados pela Engenharia Social, e sim como a ausência ou ineficiência dos controles utilizados pelas redes facilitam os ataques e fornecem oportunidade de escalabilidade. Sendo assim foi possível realizar a prova de conceito com a aplicação do *Bot*, demonstrando o potencial de uso real para uma eventual atividade maliciosa.

No caso das redes sociais em geral, um dos principais argumentos contra o uso de controles mais rígidos é o impacto que os mesmos terão na usabilidade, podendo levar os usuários a migrarem para plataformas concorrentes. Porém é possível encontrar um balanço entre as necessidades, trazendo maior segurança aos usuários com um mínimo impacto.

A possibilidade de ataques de Engenharia Social Automatizada é um exemplo. Como já dito, atualmente a identificação e remoção desses usuários ocorre apenas a partir de denúncias. Partindo do princípio que a própria Política de Uso proíbe o uso de automação, quaisquer comportamentos que indicassem essas características poderiam ser bloqueados, por exemplo:

- Mais de um *login* simultâneo do usuário, ou diversos *logins* com sucesso seguidos;
- Quantidade de requisições simultâneas e contínuas acima da capacidade de serem produzidas por um ser humano utilizando a plataforma;e

- Adicionar como contatos ou enviar mensagens para grandes quantidades de usuários simultaneamente ou em uma janela curta de tempo (que também poderia indicar *SPAM*).

Considerando que certos serviços necessitam utilizar algumas ferramentas de automação - como recrutadores reais - esses controles poderiam ser mais rígidos nas conexões via navegador (onde o uso esperado é de ser feito por uma pessoa) e mais flexíveis via API (onde é possível inclusive ter um melhor monitoramento por parte da plataforma). Esse formato permitiria oferecer um determinado número de requisições sem custo para usuários menores e planos mais robustos com a contratação de serviços profissionais da plataforma, como o já existente *LinkedIn Recruiter*.

Controles que bloqueiem automação, como os sugeridos, terão pouco ou nenhum impacto no uso de usuários regulares. Além de diminuírem grandemente os riscos de ataques maliciosos automatizados, também reduziriam o número de *SPAMs* e demais serviços não solicitados que, embora violem as Políticas de Uso, ocorrem diariamente na plataforma.

Uma questão mais complexa porém é a facilidade de criação de perfis falsos, problema enfrentado pelas redes sociais no geral. Com a cultura de expansão das suas redes de contatos, não é difícil que um perfil novo tenha rapidamente conexões suficientes para demonstrar credibilidade, sem contar a possibilidade da criação de diversos perfis falsos que gerem credibilidade uns aos outros através de depoimentos e recomendações.

Mas como gerar essa credibilidade sem processos de validação complexos?

Levando em conta que discussões sobre obrigar a identificação dos usuários já esteja ocorrendo em outras redes como o *Twitter*[5], onde podemos dizer que as características de uso são um tanto diferentes de redes profissionais, em uma rede cuja missão é conectar os profissionais do mundo para torná-los mais produtivos e bem-sucedidos [6], não seria ainda mais importante o interesse na credibilidade dos usuários? Caso não seja possível aplicar para todos os usuários, um bom começo seria exigir a validação de usuários que atuem como recrutadores na plataforma, oferecendo tanto uma forma de maior reconhecimento para esses profissionais quanto tornando mais difícil a personificação desses papéis.

6. Limitações

Após casos famosos como os experimentos de Milgran nos anos 70, estudos que envolvam o engano de pessoas enfrentam fortes dilemas éticos em sua produção. A exposição de pessoas reais a situações onde as mesmas serão iludidas, tendo suas vulnerabilidades exploradas sem seu consentimento, potencialmente podem gerar frustração e estresse psicológico após sua realização. Trabalhos no campo da ES, embora normalmente utilizem a tecnologia como suporte, implicam nessas mesmas questões de estudos psicológicos, sendo portanto necessário uma forte atenção dos autores e algumas limitações aos experimentos práticos.

Foram analisadas diversas possibilidades de execução de testes, sendo muito claras as diversas limitações em quaisquer delas para que os aspectos éticos fossem respei-

[5] https://www.cnnbrasil.com.br/business/elon-musk-diz-que-quer-todos-os-humanos-reais-verificados-no-twitter/

[6] https://about.linkedin.com

tados. Por conta disso, optou-se por focar individualmente em cada uma das etapas e testá-las separadamente levando em conta o ponto de vista da ausência de controles da plataforma. Assim seria possível validar as condições necessárias para que um ataque de Engenharia Social Automatizada ocorresse, sem a necessidade de realização de um ataque de ponta a ponta, onde seria necessário que as vítimas do teste acreditassem na história personificada a fim de que os resultados pudessem ser realmente validados.

Em especial a etapa de Abordagem foi onde ocorreram os maiores desafios, visto que a única forma de validar a ausência ou insuficiência de controles seria realizando a abordagem em si. Limitar os pedidos de conexão e mensagens a um número mínimo que fosse suficiente para caracterizar um comportamento automatizado, mas permitir um rápido controle de danos a fim de evitar o contato real com usuários, foi a forma encontrada para validar essa etapa em uma linha bastante tênue entre os limites éticos para uma pesquisa deste tipo.

Desta forma, mesmo levando em conta todas as limitações apresentadas, acreditamos ter sido possível validar as características necessárias para provar a viabilidade de um ataque automatizado de Engenharia Social.

7. Conclusão

Os ataques cibernéticos estão expondo as vulnerabilidades das redes computacionais. Os mecanismos de defesa não têm sido eficientes para impedir os ataques que exploram relações de confiança com o uso de *Bots*. O espaço virtual, no contexto das redes sociais, constitui-se um promissor cenário para a prática de toda sorte de atos ilícitos.

Este artigo explorou a ausência ou insuficiência de controles por parte dessas plataformas para detecção ou bloqueio dessas ameaças, apresentando como prova de conceito um *Bot* para simular ataques de ESA tendo como atrativo para os usuários ofertas de emprego.

As principais contribuições deste trabalho foram: i) implementar uma prova de conceito para validar a viabilidade técnica desses ataques de forma automatizada; e ii) apresentar e avaliar as descobertas do experimento dos ataques automatizados de ES.

O experimento demonstrou a viabilidade técnica para *Bots* de ESA, visto que foi possível realizar ações de forma remota e simultânea sem que houvesse qualquer restrição ou bloqueio por parte da plataforma. Os resultados apresentam o potencial de ferramentas similares para ações de ES, fato que demanda a necessidade de enfrentar os desafios impostos pelas questões de SegCiber.

Como trabalhos futuros a realização de provas de conceito com um *Bot* mais robusto identificaria os limites máximos de ações automatizadas suportados pela plataforma e uma emulação mais realista de um ataque automatizado de ponta a ponta. A implementação de capacidade de *chatbot* também verificaria a capacidade de interação, personificação e convencimento da vítima para fechamento do ciclo de um ataque de Engenharia Social.

Referências

Al-Charchafchi, A., Manickam, S., and Alqattan, Z. N. (2019). Threats against information privacy and security in social networks: A review. In *International Conference on*

Advances in Cyber Security, pages 358–372. Springer.

Aroyo, A. M., Rea, F., Sandini, G., and Sciutti, A. (2018). Trust and social engineering in human robot interaction: Will a robot make you disclose sensitive information, conform to its recommendations or gamble? *IEEE Robotics and Automation Letters*, 3(4):3701–3708.

Assenmacher, D., Clever, L., Frischlich, L., Quandt, T., Trautmann, H., and Grimme, C. (2020). Demystifying social bots: On the intelligence of automated social media actors. *Social Media+ Society*, 6(3):2056305120939264.

Camisani-Calzolari, M. (2012). Analysis of twitter followers of the us presidential election candidates: Barack obama and mitt romney. *Online). http://digitalevaluations.com.*

Crossler, R. and Bélanger, F. (2014). An extended perspective on individual security behaviors: Protection motivation theory and a unified security practices (usp) instrument. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, 45(4):51–71.

Culot, G., Fattori, F., Podrecca, M., and Sartor, M. (2019). Addressing industry 4.0 cybersecurity challenges. *IEEE Engineering Management Review*, 47(3):79–86.

Dewangan, M. and Kaushal, R. (2016). Socialbot: Behavioral analysis and detection. In *International Symposium on Security in Computing and Communication*, pages 450–460. Springer.

Dickerson, J. P., Kagan, V., and Subrahmanian, V. (2014). Using sentiment to detect bots on twitter: Are humans more opinionated than bots? In *2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014)*, pages 620–627. IEEE.

Ferrara, E., Varol, O., Davis, C., Menczer, F., and Flammini, A. (2016). The rise of social bots. *Communications of the ACM*, 59(7):96–104.

Freitas, C., Benevenuto, F., Ghosh, S., and Veloso, A. (2015). Reverse engineering socialbot infiltration strategies in twitter. In *2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 25–32. IEEE.

Freitas, C., Benevenuto, F., and Veloso, A. (2014). Socialbots: Implicações na segurança e na credibilidade de serviços baseados no twitter. *SBRC, Santa Catarina, Brasil*, pages 603–616.

Gallegos-Segovia, P. L., Bravo-Torres, J. F., Larios-Rosillo, V. M., Vintimilla-Tapia, P. E., Yuquilima-Albarado, I. F., and Jara-Saltos, J. D. (2017). Social engineering as an attack vector for ransomware. In *2017 CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON)*, pages 1–6. IEEE.

Greitzer, F. L., Purl, J., Leong, Y. M., and Sticha, P. J. (2019). Positioning your organization to respond to insider threats. *IEEE Engineering Management Review*, 47(2):75–83.

Grimme, C., Preuss, M., Adam, L., and Trautmann, H. (2017). Social bots: Human-like by means of human control? *Big data*, 5(4):279–293.

Guzman, A. L. and Lewis, S. C. (2020). Artificial intelligence and communication: A human–machine communication research agenda. *New Media & Society*, 22(1):70–86.

Huber, M., Kowalski, S., Nohlberg, M., and Tjoa, S. (2009). Towards automating social engineering using social networking sites. In *2009 International Conference on Computational Science and Engineering*, volume 3, pages 117–124. IEEE.

Khan, R. and Das, A. (2018). Build better chatbots. *A complete guide to getting started with chatbots*.

Klimburg-Witjes, N. and Wentland, A. (2021). Hacking humans? social engineering and the construction of the "deficient user" in cybersecurity discourses. *Science, Technology, & Human Values*, page 0162243921992844.

Libicki, M. (2018). Could the issue of dprk hacking benefit from benign neglect? *Georgetown Journal of International Affairs*, 19:83–89.

Messias, J., Benevenuto, F., and Oliveira, R. (2018). Bots sociais: Como robôs podem se tornar pessoas influentes no twitter? *Revista Eletrônica de Iniciação Científica em Computação*, 16(1).

Mitnick, K. D. and Simon, W. L. (2003). *The art of deception: Controlling the human element of security*. John Wiley & Sons.

Piovesan, L. G., Silva, E. R. C., de Sousa, J. F., and Turibus, S. N. (2019). Engenharia social: Uma abordagem sobre phishing. *REVISTA CIENTÍFICA DA FACULDADE DE BALSAS*, 10(1):45–59.

Rouse, M. (2013). What is socialbot? *WhatIs.com*.

Salahdine, F. and Kaabouch, N. (2019). Social engineering attacks: a survey. *Future Internet*, 11(4):89.

Shafahi, M., Kempers, L., and Afsarmanesh, H. (2016). Phishing through social bots on twitter. In *2016 IEEE International Conference on Big Data (Big Data)*, pages 3703–3712. IEEE.

Shires, J. (2018). Enacting expertise: Ritual and risk in cybersecurity. *Politics and Governance*, 6(2):31–40.

Stoeckli, E., Uebernickel, F., and Brenner, W. (2018). Exploring affordances of slack integrations and their actualization within enterprises-towards an understanding of how chatbots create value. In *Proceedings of the 51st Hawaii International Conference on System Sciences*.

Tioh, J.-N., Mina, M., and Jacobson, D. W. (2019). Cyber security social engineers an extensible teaching tool for social engineering education and awareness. In *2019 IEEE Frontiers in Education Conference (FIE)*, pages 1–5. IEEE.

Tiwari, V. (2017). Analysis and detection of fake profile over social network. In *2017 International Conference on Computing, Communication and Automation (ICCCA)*, pages 175–179. IEEE.

Turing, A. M. (2009). Computing machinery and intelligence. In *Parsing the turing test*, pages 23–65. Springer.

APPENDIX B — PUBLISHED PAPER - WGRS 2023

This appendix presents the paper "Automated Social Engineering Attacks using ChatBots on Professional Social Networks". It is the last submission before the final version of this dissertation submitted and works as a summarized version of the entire content of this work, including all the new methodology and tests executed, besides the final conclusions of the research.

- **Title:** "Automated Social Engineering Attacks using ChatBots on Professional Social Networks"
- **Conference:** "XXVIII Workshop de Gerência e Operação de Redes e Serviços (WGRS 2023)"
- **URL:** https://sol.sbc.org.br/index.php/wgrs/issue/view/1109
- **Date:** May 22-26, 2023
- **Venue:** Brasília, DF, Brazil
- **Digital Object Identifier (DOI):** https://doi.org/10.5753/wgrs.2023

Automated Social Engineering Attacks using ChatBots on Professional Social Networks

Maurício Ariza[1], **Antonio João Gonçalves de Azambuja**[1],
Jéferson Campos Nobre[1], **Lisandro Zambenedetti Granville**[1]

[1]Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)
Caixa Postal 15.064 – 91.501-970 – Porto Alegre – RS – Brazil

{mariza,antonio.azambuja,jcnobre,granville}@inf.ufrgs.br

Abstract. *The growth of the internet and social networks has intensified human interactions, raising the risk of cyberattacks. Social Engineering targets those human relationships in the cyber environment, using technology as a support to exploit natural human failures. Research has shown the capacity of Social Engineering attacks, however, there are few papers focusing on the evolution and trust of ChatBots and automation as a support for those attacks. This paper presents an analysis of the capacity of professional social networks to detect and block automated Social Engineering threats to their users. The approach developed allowed us to identify the characteristics of the trust relationship between the user, the social network, and the ChatBot resulting from the established interaction, and failures on the part of social networks to identify and block this kind of behavior. To this end, an automated Social Engineering bot was developed. The analysis and discussion of the results allow demonstration of the security vulnerabilities present in professional networks and in building the user's trust relationship with the ChatBot.*

1. Introduction

Social networks have been used as a vector for cyber attacks to obtain sensitive information from users using virtual profiles [Paradise et al. 2019]. Attackers make use of the connectivity of these social networks to expand their area of operation, a fact that exponentially increases the challenges of cybersecurity. The interconnectivity of social networks and the growth of the cognitive dimension of work are making human resources one of the pillars of security [Culot et al. 2019] [Greitzer et al. 2019].

Cyber attacks carried out on social networks have exploited human interaction in conjunction with technological gaps, weakening the cybersecurity chain. Organizations have used defense solutions to face cyber attacks, such as firewalls, Intrusion Detection Systems (IDS), Intrusion Prevention Systems (IPS), and antivirus. However, these defense mechanisms have not been sufficient to fully prevent Social Engineering (SE) actions in the cyber environment [Salahdine and Kaabouch 2019] [Klimburg-Witjes and Wentland 2021].

Cybersecurity experts characterize attacks that focus on human behavior as SE attacks that aim to manipulate users to reveal sensitive information. These attacks combine human interaction with the exploitation of [Klimburg-Witjes and Wentland 2021] technological vulnerabilities. The increasing use of social networks to establish personal and professional relationships opens a field for the actions of Automated Social Engineering (ASE) bots. [Huber et al. 2009]. Social Engineers have sought to develop bots with intelligence, making automated interaction unnoticed by users.

Bots can be used for positive actions, such as helping the user in their online experience. However, bots developed for ASE attacks have made it possible for a single attacker to contact a large number of potential victims simultaneously because of their scalability. The attacker aims to get the victim to reveal sensitive information, which can be used for data theft [Huber et al. 2009] [Dewangan and Kaushal 2016]. Bots are automated software, which sometimes uses features such as artificial intelligence. In its functionality it has the ability to execute operation and control commands to impersonate humans, simulating the activities of real users [Shafahi et al. 2016].

During the literature review, few papers were identified that present analyses on ASE with the use of bots. The studies focus on human behavior in the face of SE actions [Huber et al. 2009]. In this sense, the strategies of a cyber attack are being framed as a social issue and not just a technical vulnerability, according to the authors [Klimburg-Witjes and Wentland 2021].

SE attacks using fake accounts with identity theft on social networks with impacts on users' privacy and information security were already analyzed [Al-Charchafchi et al. 2019]. The discussion of the use of SocialBots for social media conviction campaigns is present in the papers that evaluate the impact of these tools on user behavior [Boshmaf et al. 2013]. The use of bots to influence users of Twitter, aiming to gain followers and compromise the network structure are presented in the work of some authors [Freitas et al. 2015] and [Messias et al. 2018].

An analysis of SE tasks automation through a bot on Facebook [Huber et al. 2009] concludes that persuasion is an essential resource in the ASE process. However, no papers were identified that present bots with intelligence to perform an automated human interaction to search for sensitive information, without the perception of the user being targeted by the ASE technique.

This paper proposes a proof-of-concept bot with intelligence to perform an ASE attack, having the offer of attractive jobs as a stimulus for user interaction with the bot. This attack focuses on a specific group of users of the professional social network LinkedIn. The main contributions of this work are: (i) evaluate LinkedIn capacity to detect and mitigate an ASE attack; (ii) implement a proof of concept to validate the technical viability of this attack; and (iii) propose control improvements that can be implemented by those social networks to decrease the risk of SE attacks to their users.

The paper is organized as follows. In Section 2 the concepts related to the theory to support the research. Next, Section 3, presents the methodological proposal of the study and the identified limitations. Section 4, performs an evaluation of the proposal and a discussion of the results. Section 5, analyzes the related works. Finally, it presents in Section 6 the conclusion and an approach for future work.

2. Background

2.1. Bots

In cyberspace, there are authentic bots that aim to perform useful activities for users. However, there are also malicious bots, which can perform attacks to obtain relevant information or maintain control of the accessed device. Bots can be used for spreading false information (fake news), spam, and phishing. [Freitas et al. 2015].

Cybercriminals use malicious bots to simulate human behavior, bypassing security mechanisms. As mentioned by some authors [Huber et al. 2009], ASE attacks using bots take SE to a new level of scalability of attacks. In the context of ASE, cybercriminals use malicious bots to simulate human behavior, avoiding security mechanisms.

With the growth of social networks and the large volume of data in cyberspace, social engineers have started to spread bots with human-like behavior to a large number of users. These bots simulate human conversations, known as ChatBots and, those that operate on social networks, SocialBots [Shafahi et al. 2016].

SocialBots are defined as effective tools to perform SE attacks, with the aim of gaining access to sensitive information. It is a tool that has the ability to compromise the structure of social networks, aiming to: (i) steal identity; (ii) influence users; (iii) increase the number of followers; and (iv) inflate the popularity ratings of a particular profile account [Boshmaf et al. 2011] [Camisani-Calzolari 2012] [Dewangan and Kaushal 2016].

As such, SocialBots need a technical infrastructure, with a combination of a social networking platform and technical requirements for automating the behavior of an account, using an Application Programming Interface (API) or proprietary mechanisms to interact with the platform [Assenmacher et al. 2020]. As a certain degree of intelligence is incorporated into SocialBots to simulate human behavior, it sparks interest in research on the topic [Ferrara et al. 2016].

It is a tool that simulates human behavior to perform automated interactions on social networks [Rouse 2013]. SocialBots for the most part are automated social media accounts that impersonate people. These interactions are artificial intelligence activities that have shown growth in the online environment with the use of this tool [Freitas et al. 2015] [Hepp 2020].

ChatBots are the integration of systems, tools, and scripts that promote instant messaging conversations with or without human participation [Stoeckli et al. 2018]. They are developed to help human users in specific service situations and are not exhaustive. Here are three (3) examples: customer service, communication service, and digital education service [Grimme et al. 2017].

This tool originated in the field of Computer Science, in this sense it is a tool developed with the help of artificial intelligence mechanisms that interact with users. The use of natural language in ChatBots, a language used for human communication, is a challenge to be overcome for the development of the tool [Khan and Das 2018].

The growing use of personal assistants demonstrates the popularity of ChatBots. However, as the use of this tool grows, it is important to keep in mind the increase of attacks on typical ChatBots architectures, for example: client module, communication module, response generation module, and database [Ye and Li 2020].

These tools, SocialBots and ChatBots, have been developed with the help of artificial intelligence mechanisms that interact with users [Freitas et al. 2015]. Artificial intelligence is similar to human intelligence, developed with automation as per the need of the application [Ferrara et al. 2016]. As a certain degree of intelligence is built into the tools to simulate human behavior, the capacity and scalability of attacks increase.

2.2. Automated Social Engineering

Social engineers make use of automated Bots, which are able to impersonate humans to carry out an ASE [Shafahi et al. 2016]. These attacks seek to establish a trust relationship to obtain sensitive information about the user and require little intervention to establish the relationship, enabling greater reach by their scalability [Mitnick and Simon 2003] [Huber et al. 2009]. Human communication has been based on the development of human-machine interfaces. The disruptive technologies are inspiring studies on this communication [Guzman and Lewis 2020].

Social networks are facilitating communication, social interaction, and sharing of personal and corporate information, increasing their popularity in the cyber environment. These networks represent an attractive virtual space for attackers to exploit technical vulnerabilities and users' lack of knowledge and awareness of SE actions [Al-Charchafchi et al. 2019].

The relationships formed in this environment allow greater exchange of information, ratifying the statement of [Castells 2009], that in social relationships, networks are communicative structures. Cyberspace constitutes a promising scenario for the practice of all sorts of illicit acts, without respecting geopolitical borders. The growth of social networks has enabled the creation of a large number of fake profiles, with the use of automated Bots for this activity [Tiwari 2017].

It is common the SE attacks to require time to establish a trusting relationship and resources. However, SE can be accomplished through automated mechanisms. ASE attacks require little human intervention to establish the relationship and have greater reach because of their scalability [Huber et al. 2009]. Automated attacks can be prepared using valuable information and/or influencing certain groups in social networks [Gallegos-Segovia et al. 2017]. ASE attacks using Bots and Phishing have become more frequent due to the increasing use of social networks for personal and professional activities [De Kimpe et al. 2020].

3. Methodology

It is already a common concept to classify humans/users as the weakest link of the cybersecurity chain, as attacks exploiting their failures have better success rates and sometimes require less technical skills and risk for the attacker [Darwish et al. 2012].

Social networks have their essence based on creating interaction between humans. But beyond allowing a space where distance boundaries can be bypassed to enhance connections, they also bring to the virtual world many of the threads from the real world. But there is a difference as it is a space where people don't have the same awareness and capacity to recognize risks, which together with the stronger capacity of anonymity and impersonation become a perfect environment for SE [Crossler and Bélanger 2014].

Comparing to other social networks like Facebook, Twitter, and Instagram, professional social networks create a more corporate environment, focused on business connections and career growth. This scenario creates a sense of trust and credibility, attracting headhunters looking for candidates as well as companies looking for potential new customers. These relations are already exploited by social engineers, especially impersonating recruiters using attractive job opportunities as bait to steal internal information or personal data of the victims[1].

Currently, LinkedIn is the most popular professional social network, with more than 850 million members in more than 200 countries. Their User Agreement defines in section 8.2[2] the actions that are not allowed to users, highlighting forbidden use of false information or impersonation in the profile and usage of bots and automation to realize actions in the platform.

Considering the SE attacks already mentioned, we can easily find references to fake profiles or false information used for several reasons. Talking specifically about automation, if we search on the internet or code repositories like GitHub we could find several bots and scripts specifically designed for LinkedIn. Those references indicate a potential lack or insufficient implementation of controls by the platform, that looks to focus the enforcement of policies on complaints or reports done by the users. Our goal is to evaluate the risk level for LinkedIn users to face an ASE attack, understand the platform's capacity to detect and block those attacks, and offer suggestions to improve their controls to decrease risk with no or minimal impact on usability.

3.1. Limitations

SE was born in the field of psychology, as they aim to exploit human failures, using technology as a support to achieve those goals. For a full understanding of the impact of a SE attack, we will need to only validate technical aspects, but also tricky subjects to observe their behavior and actions. The field of social psychology explored those matters for a while to could identify which are the boundaries for ethical research, understanding that the goals do not justify the methods to avoid extreme cases like the famous Milgram experiments in the 70s.

Exposing people to situations where they will be deceived, and having their vulnerabilities exploited without their consent (or full understanding) violates ethical dilemmas. As result, later they can create frustration and stress due to expectations created, broken promises, or the feeling of being fooled. Understanding and respecting those boundaries was one of the drivers of this paper, and even as discussing research ethics was not our goal it's not possible to run a work like that without raising questions in this matter.

Our first challenge was how to validate our proposal within keeping compliance with ethical policies. In order to achieve that, we break the entire attack life cycle in separated steps and try to do testing and validation of each one individually, based on the results we could have a fair understanding of the application response and the potential of the full attack.

[1] https://www.forbes.com/sites/reneemorad/2017/06/30/how-to-avoid-the-latest-linkedin-scam/?sh=13e1d13849c1

[2] https://www.linkedin.com/legal/user-agreement#dos

The main difficulties happened in the Approach and Interview steps, as they would require at least some level of contact with subjects. For the Approach step, we achieved a tiny line between keeping our premises and violating ethical barriers. We then decided to limit to a minimum number the quantity of LinkedIn users receiving the request and the message. In this way, we could evaluate if the platform will identify the automated behavior, and then cancel/exclude all actions immediately as a damage control mechanism. This format allows us to avoid any individual really having contact with our testing accounts.

For the Interview step, we focused on the main functionality of our Recruiter Chatbot: do a job interview. As the malicious action would happen by making the victim believe it is a real job interview happening, and them being expected to have a process involving signing a contract and sending documents for identification, for example. So our tests tried to validate the bot's capacity to run a convincing job interview as a way to consider their potential to have the same results in a full cycle attack.

Considering all those mechanisms we evaluate as being achieved enough results to validate the potential of the proposed attack without violating any associated ethical requirements. We believe this is a core topic for any kind of research and a deeper analysis of the impact of ethical research matters especially on the SE field it's an intriguing topic for further studies.

4. Proposal and Evaluation

To evaluate the attack we used a proof of concept scenario with 2 bots. The first one interacts with the social network to search and contact the victims with the bait - the Platform Bot. The second one it's a Chat Bot service that would act directly with the victims to execute the step of the job interview - the Recruiter Bot.

For the Platform Bot role, we developed a Python code to connect with LinkedIn. LinkedIn offers a very rich API for software interaction, but considering the characteristics of the attack and the kind of validation we are looking for it would not be the best option. Then we evaluate that a connection is done through a browser - like done by any regular user - would provide us a better understanding of the social network response than a channel for software connections. In order to achieve this we use the Selenium library, which allowed our bot to act in a request-response through the browser.

For the Recruiter Bot, there were several options available that could fit into the need to execute the necessary actions without the need to develop custom code. Using a pre-defined set of job interview questions, plus information scrapped from the victim's LinkedIn profile, the Recruiter Bot would basically conduct a false job interview with the victim with the goal to collect sensitive information from current and past jobs. As it can execute both RH-like interviews with more generic questions and a technical interview, the entire process can be executed by the same bot and then, in the end, as the victim is 'accepted' for the position, personal information can then be stolen for identity theft to sign the fake work contract.

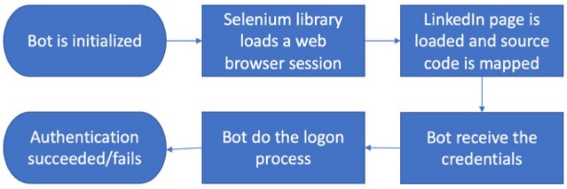
Our attack proposal follows the SE attack stages structure [Mitnick and Simon 2003], organizing it in 4 steps: i) Authentication, ii) Search, iii) Approach, and iv) Interview. Our goals in each are detailed below:

1. Authentication: As also illustrated in Figure 1, the goal of this step is to verify if the social network detects or has different behaviors when the user logon process it's done using automation. For this evaluation, our Platform Bot opens the LinkedIn website in the browser, maps the source code of the main page to identify the credential fields, fills them with the values received, and then submits to conclude the authentication process and access the main page of a logged user.

2. Search: This step aims to check the detection of automated searching of users. Similarly to the first step, the Platform Bot maps the page source code, identifies the search field, runs the search for the provided terms and then stores temporarily the returned profiles, creating a database of potential victims. Figure 2 also refers to this step.

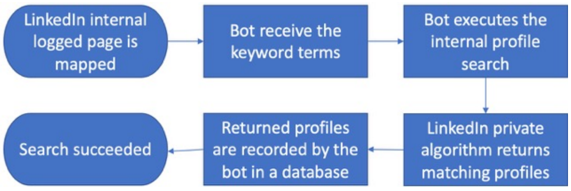
3. Approach: Goal here is to start the interaction with the profiles of potential victims collected in the previous step. As also seen in Figure 3, using the created temporary database, the Platform Bot adds them as contacts and sends a custom message, which serves as the bait for interaction. This action happens to all profiles captured.

4. Interview: Based on the results of the bait sent in step 3, a script scrapes data from the victim's LinkedIn profile to feed the Recruiter Bot database, which they have enough information to execute a job interview with the victim. Figure 4 also illustrates the full cycle of this step.



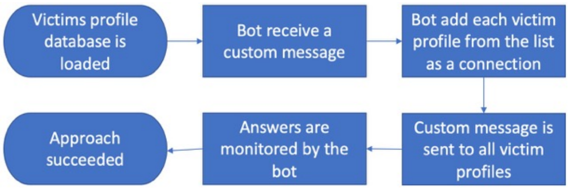
Bot is initialized
Selenium library loads a web browser session
LinkedIn page is loaded and source code is mapped
Bot receive the credentials
Bot do the logon process
Authentication succeeded/fails

Figure 1. Attack authentication phase.



LinkedIn internal logged page is mapped
Bot receive the keyword terms
Bot executes the internal profile search
LinkedIn private algorithm returns matching profiles
Returned profiles are recorded by the bot in a database
Search succeeded

Figure 2. Attack search phase.



Victims profile database is loaded
Bot receive a custom message
Bot add each victim profile from the list as a connection
Custom message is sent to all victim profiles
Answers are monitored by the bot
Approach succeeded

Figure 3. Attack approach phase.

4.1. Testing

The testing phase followed the steps of our proposed Attack Flow. The Platform bot was executed in a Windows machine running Python, the Selenium library, and Google Chrome as a browser. For the Interview bot we used the SAP Conversational AI platform.

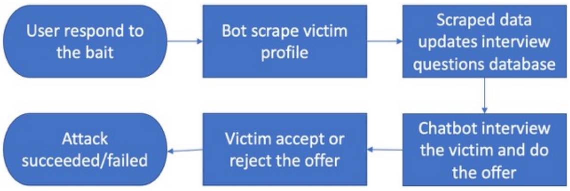

User respond to the bait
Bot scrape victim profile
Scraped data updates interview questions database
Chatbot interview the victim and do the offer
Victim accept or reject the offer
Attack succeeded/failed

Figure 4. Attack interview phase.

Testing Step 1 - Authentication: The success criteria of this step is to execute authentication in the platform following different behaviors to observe if we have any impact from the application side or that could demonstrate controls or blockers due to automation characteristics. For comparison criteria, we defined three basic behavior patterns to be tested: (1) Do the logon process 10 times simultaneously, (2) Do the logon process 10 times with 5 seconds waiting time between each attempt, and (3) Do the logon process 10 times with 10 seconds waiting time between each attempt. Those patterns try to replicate behaviors not expected from a real human user due to the quantity or speed of attempts, especially as the execution is happening through the web browser. Also, for each one of them, we tested using the following variations to observe if they impact the results in any way:

- Receive the credentials of the created fake profile in execution time through the script.
- Read the credentials of the created fake profile from a file.
- Use of wrong/invalid credentials.
- Use of a public proxy to execute the logon from a random country, different from the one defined as the location of the user in the created fake profile.

In the results of our tests, we did not observe any difference in the social network behavior when the tests were executed using the valid credentials of the fake account. We also executed the test of each pattern on different days to guarantee that the execution of one of them would impact the results of the others. As an alternative variation, we also executed 10 sequential login attempts using the valid fake profile credentials and invalid ones, but manually (not using the bot), through a web browser, which did not demonstrate as well any difference in the results. When testing using invalid credentials, both through a bot or manually, after the 6th attempt LinkedIn starts requiring a puzzle (similar to a Captcha verification) and/or additional validation like a code sent by email/SMS to proceed with login, indicating that brute force behaviors are identified and blocked, which not happens for all kind of automated access attempts.

Testing Step 2 - Search: In this step we evaluated the capacity of the Platform Bot to execute queries in the social network without being detected. The authors defined a series of keywords to be used for the queries, based on some standard IT skills associated with this project only for reference. It is important to highlight that the quantity, order, and all other information related to query results are all associated with the search algorithm used by LinkedIn, which analysis is out of the scope of this work. Keywords used were only a manner to evaluate the response for automated queries through the web browser. For the test, we enumerate ten keywords, ”test”, ”Social Engineering”, ”Bot”, ”Chatbots”, ”Social Networks”, ”Information Security”, ”Python”, ”Automation”, ”GitHub” and ”API”. Similar to what was done in step 1, we used the following variations:

- Querying the same keyword 10 times simultaneously.
- Querying the same keyword 10 times with 5 seconds waiting time between each.
- Querying 10 simultaneous sessions, each one using a different keyword.
- Querying the 10 different keywords in the same session, in sequence, with 5 seconds waiting time between each.

Was not a goal of this step to do stress/load testing or cause a denial of service in the application. The tested behaviors used a speed and/or quantity not expected to be executed by a human user, especially as they were executed through the web browser. Besides the expected differences in the results (considering the different terms used and the LinkedIn algorithm, out of the scope of this paper), we did not observe any differences in the variations, and all queries received correctly the results with a list of profiles associated with the keyword term.

Testing Step 3 - Approach: In the previous step, after running the queries, the Platform Bot keeps a reference of the returned profiles to be used for this step. Here we had our main challenge on the already discussed ethical implications. Looking to have a limit on the impact of our research without prejudice to the results, we implemented the following controls to our bot:

- For each query, instead of the several pages of results, we keep only the ones on the first page, which were around 15-21 profiles per query.
- We only executed the approach 10 times, one per keyword, disregarding variations on queries.
- After executing the approach, the bot keeps a record of the success and then deletes/cancels all their actions within the user/victim.

The approach happened with a connection request to the user and the sending of a custom message, using tags to use the real user name instead of generic terms like 'dear user'. Again, once validated the request and the message, the request was canceled and the message was deleted for both sides, avoiding any further interaction with the users. Again no impact or actions from the side of a social network were identified during any of the tests.

Testing Step 4 - Interview: For this step our validation followed a different direction. We used the SAP Conversational IA platform, with a proof of concept chatbot to work as the Recruiter Bot. Based on an initial database of common questions for a job interview, we used a script to scrape the data from the victim profile and use them as input for additional questions, creating questions like "How was your experience in Company X?", "Can you talk more about your skills on technology Y?". Based on the observation of the chatbot interview for different profiles randomly selected from the previous step, it was capable to conduct a job interview without the need for management or additional command/control. This result allows us to demonstrate their capacity to be used for the proposed attack without having a realistic approach with subjects/victims, violating the already discussed ethical limitations.

4.2. Evaluation of Results

The main contribution of this paper was to look to validate the hypothesis of lack or insufficient controls implemented by social networks, no matter whether this could be known or expected in the technology field, we could not found official research or academic references as foundations. It was not the goal to explore the human factor and the psychological matters associated with SE, but to observe if the technology channels allow or at least do not offer barriers to avoid those aspects being exploited in their users, especially in cases like our proposal were it's possible to achieve high scalability by the attacker.

Certainly the main impact of rigid controls on a social network it's on the user experience, which can directly affect the usage, base of users, and several important success indicators in this market. As a result, users can migrate to concurrent platforms, for example. But should be possible to find a balance and increase the security levels with no or minimal impact on the users.

Considering that the current LinkedIn User Agreement already forbidden the usage of automation, some automation-detection controls can be applied. This will not only avoid ASE but an entire behavior that is already forbidden. But even this control can be implemented with some flexibility. A regular user connecting through a web browser has some human limitations on their speed and quantity of requests - below a certain limit, not only you have a certain or potential automated behavior, but you also have high chances of SPAM and other unsolicited interactions. Based on our results, some examples of simple controls to detect automated behavior are:

- More than one simultaneous login of the same user (with some variations, like if you consider a user logged on the laptop and the smartphone at the same time), or several successful logins in a short time period.
- Several simultaneous and/or continuous requests (not only search queries but for any action) in a quantity or time frame higher than the average capacity of an human being.
- Do several contact requests and/or send several messages to different users simultaneously or in a short time frame (also potentially indicating SPAM).

The enforcement of controls on those behaviors doesn't need to be the block or cancellation of the request. Requiring additional fields like a Captcha, similar to what is already used to avoid brute force attacks, can be an excellent way to avoid automation, as they would be required only for certain scenarios that will not affect most of the regular users.

Going a bit beyond, imagining the need of certain users/scenarios where some automation can be useful or required - for real recruiters, for example, the enforcement of controls can be more rigid through a web browser (where regular users, not automation, is expected) and more flexible through API, for example. This will allow better monitoring and control by the platform, even being able from a business perspective to offer a certain quantity of free requests for minor customers (like independent small headhunters) or more robust professional services sold by the platform, like the already existing LinkedIn Recruiter.

Those examples of controls can solve the automation issue, enforcing already existing policies with minimum impact on users. But a more complex challenge is how easy is to create fake profiles, a problem not only for LinkedIn but for any social network in the current days. It is not difficult to build a base of interactions and contacts that can create a sense of legitimacy, organically through real users or even through a network of other false profiles.

There is no easy answer to doing it without complex validations. But the impact of fake profiles has been growing so fast that discussions over mandatory user validation are already happening on other social networks like Twitter. Of course, those networks have a different set of users and characteristics of usage, but if we analyze that LinkedIn's mission is to ”connect the world's professionals to make them more productive and successful”[3], would not be credibility and veracity of users a matter of interest for all their users? Verified profiles already exist usually for social influencers, and potentially even without the enforcement, many users will potentially look for this validation as a way to recognize their work and responsibility - or at least some groups like recruiters can be targeted. There are several opportunities, each with pros and cons, but certainly, some kind of control in this direction will be necessary to make it at least a bit harder to personify attacks.

5. Related Work

[Boshmaf et al. 2013], evaluate the vulnerabilities of social networks arising from a large-scale infiltration campaign using SocialBots. This study presents in their results an infiltration success rate of 80 % on Facebook, an index that demonstrates an unauthorized disclosure of private user data.

[Dewangan and Kaushal 2016], presents a model for detecting SocialBots used in political campaigns and marketing of products, having as input the behavior analysis. These actions bring with them security risks, considering the use of social networks for disseminating political positions and monitoring the consumption profile of users.

[Aroyo et al. 2018], discuss how SE exploits the trust relationship between users and bots. Based on the four (4) stages of an SE attack [Mitnick and Simon 2003], a bot was developed to simulate this task. First, the bot sought to obtain information with private questions. Then, it established a relationship of trust with the users, for a virtual and anonymous approach to the target.

With these actions, authors present in the research results that users have established a trust relationship with the tool. Among the requirements in the interaction with users, the ethical aspects were considered, by these authors.

[Al-Charchafchi et al. 2019], present a review of research on privacy and threats in social networks. For the authors, although the literature presents work on privacy, more effort is needed. The social networking environment is a rich source of personal data, making it an attraction for actions in social engineers, who exploit the users' lack of awareness and knowledge on security-related issues.

The complexity of SE attacks is related to the combination of social strategies and techniques used to carry out a cybercrime [Al-Charchafchi et al. 2019]. In this context to mitigate the impacts of attacks, [Piovesan et al. 2019] claims that security policies can provide a higher level of information security. However, they do not guarantee complete security.

[Freitas et al. 2014], present a discussion on the impact of the use of SocialBots on Twitter to characterize the behavior of the tool on a large database. In the results, the authors highlight that the method they developed to characterize and detect SocialBots, had a 92% successful detection indicator.

[3] https://about.linkedin.com

[Messias et al. 2018], claim that a simple Bot can achieve high levels of influence on Twitter. [Shafahi et al. 2016], on the other hand, points to the need to raise the level of awareness about phishing actions that use SocialBots. The authors state that these actions pose a threat to organizations.

[Paradise et al. 2019], analyze in the organizational context the strategies to monitor organizational social networks and detect SocialBots that aim to obtain data from the organization. The strategies were analyzed considering different levels of attacker knowledge using a simulation with real social network data.

[Huber et al. 2009], present the cycle of an ASE attack using a Bot. The attack demonstrated how social networks can be used by social engineers to obtain information. To this end, two (2) experiments were conducted in the study. The first analyzed the ability of Bots to obtain information from social networks. The second performed the Turing test, which seeks to evaluate the ability of a machine to imitate a human being.

Finally, for the authors, ASE with Bots is scalable and requires fewer human resources. The tool was used in a proof of concept on Facebook. The two (2) experiments allowed to ratify that it is possible to automate SE actions to obtain information and to demonstrate that the Bot used was not identified by the security measures of Facebook. The increasing number of users' social interactions on networks makes SE automation Bots an interesting tool for social engineers.

6. Conclusion

Cyber attacks have been exposing the vulnerabilities of computer networks and applications. Especially the context of social networks, becoming each day more important in people's lives, are being a promising scenario for several malicious actions, and the current defense mechanisms are not being efficient to mitigate or avoid them, highlighting the exploitation of trust using bots.

ASE bots offer great scalability with no need for more exposure from the attacker. This paper presented the development of a bot-based approach to simulate ASE attacks using job proposals as bait. Through a fake recruiter profile on LinkedIn, is it possible to identify and contact potential victims using automated mechanisms, looking for leakage of personal or corporate data. The complete absence of controls demonstrates the potential for similar SE actions in the real world, which should raise awareness and important actions to address those threats in an Information Security strategy.

The main contributions of this research are: i) Implement a Proof of Concept to validate the technical viability of this attack; ii) Evaluate the defense and response mechanisms of the social network to an ASE attack; and iii) Offer some potential ways to mitigate those attacks with minimum impact to user experience.

As a future work, the implementation of improvements to the proof of concept Platform Bot would allow to map and evaluate the limits for automated activities supported by the platform, and a more realistic simulation of an automated attack end-to-end. Also, more testing over the Recruiter Bot using real subjects outside an attack scenario will also help to better understand the capacity of similar chatbots to convince a real person and create a trust connection necessary to conclude a job interview, offering a deeper analysis for the last step of this attack life cycle validation.

References

Al-Charchafchi, A., Manickam, S., and Alqattan, Z. N. (2019). Threats against information privacy and security in social networks: A review. In *International Conference on Advances in Cyber Security*, pages 358–372. Springer.

Aroyo, A. M., Rea, F., Sandini, G., and Sciutti, A. (2018). Trust and social engineering in human robot interaction. *IEEE Robotics and Automation Letters*, 3(4):3701–3708.

Assenmacher, D., Clever, L., Frischlich, L., Quandt, T., Trautmann, H., and Grimme, C. (2020). Demystifying social bots: On the intelligence of automated social media actors. *Social Media+ Society*, 6(3):2056305120939264.

Boshmaf, Y., Muslukhov, I., Beznosov, K., and Ripeanu, M. (2011). The socialbot network: when bots socialize for fame and money. In *Proceedings of the 27th annual computer security applications conference*, pages 93–102.

Boshmaf, Y., Muslukhov, I., Beznosov, K., and Ripeanu, M. (2013). Design and analysis of a social botnet. *Computer Networks*, 57(2):556–578.

Camisani-Calzolari, M. (2012). Analysis of twitter followers of the us presidential election candidates: Barack obama and mitt romney. *http://digitalevaluations. com.*

Castells, M. (2009). Communication power. nueva york: oxford university press.

Crossler, R. and Bélanger, F. (2014). An extended perspective on individual security behaviors. *ACM SIGMIS Database*, 45(4):51–71.

Culot, G., Fattori, F., Podrecca, M., and Sartor, M. (2019). Addressing industry 4.0 cybersecurity challenges. *IEEE Engineering Management Review*, 47(3):79–86.

Darwish, A., Zarka, A. E., and Aloul, F. (2012). Towards understanding phishing victims' profile. In *2012 International Conference on Computer Systems and Industrial Informatics*, pages 1–5.

De Kimpe, L., Ponnet, K., Walrave, M., Snaphaan, T., Pauwels, L., and Hardyns, W. (2020). Help, i need somebody: Examining the antecedents of social support seeking among cybercrime victims. *Computers in Human Behavior*, 108:106310.

Dewangan, M. and Kaushal, R. (2016). Socialbot: Behavioral analysis and detection. In *International Symposium on Security in Computing and Communication*, pages 450–460. Springer.

Ferrara, E., Varol, O., Davis, C., Menczer, F., and Flammini, A. (2016). The rise of social bots. *Communications of the ACM*, 59(7):96–104.

Freitas, C., Benevenuto, F., Ghosh, S., and Veloso, A. (2015). Reverse engineering socialbot infiltration strategies in twitter. In *IEEE/ACM ASONAM 2015)*, pages 25–32. IEEE.

Freitas, C., Benevenuto, F., and Veloso, A. (2014). Socialbots: Implicações na segurança e na credibilidade de serviços baseados no twitter. *SBRC, Santa Catarina, Brasil*, pages 603–616.

Gallegos-Segovia, P. L., Bravo-Torres, J. F., Larios-Rosillo, V. M., Vintimilla-Tapia, P. E., Yuquilima-Albarado, I. F., and Jara-Saltos, J. D. (2017). Social engineering as an attack vector for ransomware. In *CHILECON 2017*, pages 1–6. IEEE.

Greitzer, F. L., Purl, J., Leong, Y. M., and Sticha, P. J. (2019). Positioning your organization to respond to insider threats. *IEEE Engineering Management Review*, 47(2):75–83.

Grimme, C., Preuss, M., Adam, L., and Trautmann, H. (2017). Social bots: Human-like by means of human control? *Big data*, 5(4):279–293.

Guzman, A. L. and Lewis, S. C. (2020). Artificial intelligence and communication. *New Media & Society*, 22(1):70–86.

Hepp, A. (2020). Artificial companions, social bots and work bots. *Media, Culture & Society*, 42(7-8):1410–1426.

Huber, M., Kowalski, S., Nohlberg, M., and Tjoa, S. (2009). Towards automating social engineering using social networking sites. In *2009 International Conference on Computational Science and Engineering*, volume 3, pages 117–124. IEEE.

Khan, R. and Das, A. (2018). Build better chatbots. *A complete guide to getting started with chatbots*.

Klimburg-Witjes, N. and Wentland, A. (2021). Hacking humans? social engineering and the construction of the "deficient user" in cybersecurity discourses. *Science, Technology, & Human Values*, 46(6):1316–1339.

Messias, J., Benevenuto, F., and Oliveira, R. (2018). Bots sociais: Como robôs podem se tornar pessoas influentes no twitter? *Revista Eletrônica de Iniciação Científica em Computação*, 16(1).

Mitnick, K. D. and Simon, W. L. (2003). *The art of deception*. John Wiley & Sons.

Paradise, A., Shabtai, A., and Puzis, R. (2019). Detecting organization-targeted socialbots by monitoring social network profiles. *Networks and Spatial Economics*, 19(3):731–761.

Piovesan, L. G., Silva, E. R. C., de Sousa, J. F., and Turibus, S. N. (2019). Engenharia social: Uma abordagem sobre phishing. *REVISTA CIENTÍFICA DA FACULDADE DE BALSAS*, 10(1):45–59.

Rouse, M. (2013). What is socialbot? *WhatIs.com*.

Salahdine, F. and Kaabouch, N. (2019). Social engineering attacks: a survey. *Future Internet*, 11(4):89.

Shafahi, M., Kempers, L., and Afsarmanesh, H. (2016). Phishing through social bots on twitter. In *2016 IEEE International Conference on Big Data*, pages 3703–3712. IEEE.

Stoeckli, E., Uebernickel, F., and Brenner, W. (2018). Exploring affordances of slack integrations and their actualization within enterprises-towards an understanding of how chatbots create value. In *Proceedings of the 51st Hawaii International Conference on System Sciences*.

Tiwari, V. (2017). Analysis and detection of fake profile over social network. In *ICCCA 2017*, pages 175–179. IEEE.

Ye, W. and Li, Q. (2020). Chatbot security and privacy in the age of personal assistants. In *IEEE/ACM SEC 2020*, pages 388–393. IEEE.