International Journal on Perceptive and Cognitive Computing (IJPCC) Vol 10, Issue 2 (2024)
https://doi.org/10.31436/ijpcc.v10i2.484

Analyzing Threat Level of the Backdoor Attack Method for an Organization's Operation

Muhammad Zafran Syahmi Mohd Nasharuddin, Adamu Abubakar*
Dept. of Computer Science, KICT, International Islamic University Malaysia, 53100 Kuala Lumpur, Malaysia.

*Corresponding author: adamu@iium.edu.my
(Received: 4th June 2024; Accepted: 25th June 2024; Published on-line: 30th July 2024)

Abstract— Backdoor attacks played a critical part in the catastrophe, as well as the overall impact of cyberattacks. Backdoor assaults are additionally influencing the landscape of malware and threats, forcing companies to concentrate more on detecting and establishing vulnerability tactics in order to avoid hostile backdoor threats. Despite advances in cybersecurity systems, backdoor assaults remain a source of concern because of their propensity to remain undetected long after the attack vector has been started. This research is aimed to examine the threats of backdoor attack methods in an organization's operational network, provide a full-scale review, and serve as direction for training and defensive measures. The fundamental inspiration was drawn from the alarming and involving threat in cybersecurity, which necessitates a better awareness of the level of risk and the concurrent requirement for increased security measures. Most traditional security solutions usually fail to detect harmful backdoors due to the stealthy nature of backdoor code within the system, necessitating a unique approach to full-scale threat analysis. A multi-phase approach that begins with considerable reading and examination of existing literature to get insight into typical backdoor attack methodologies and application methods. Following analysis, testing was carried out in a virtual lab in a controlled environment because thorough malware analysis testing must adhere to ethical and legal cyber testing laws to avoid any penalties or foolish breaches. This methodology also included testing on numerous attack channels combined with backdoor attacks, such as detecting software vulnerabilities, phishing emails, and direct payload injection, to determine the complexity of the different attack vectors. Each of the collected data is utilized to create a threat model that predicts the amount of risk associated with the backdoor attack approach. The finding contributes to the development of more resilient defence mechanisms, while also strengthening the overall organization's security architecture and protocols.

Keywords— Cybersecurity, Backdoor attack, Malware, Jitter, Direct payload injection

I. Introduction

In this modernization century, current technologies are heavily relied on efficiency, innovation, and competitive benefits for either the enterprise or the individual. However, when it comes to security issues in computing, modernization generates an entry point and significant vulnerabilities in computing system. Specifically, a security system that can exploit any weakness in system. One such issue is a "backdoor attack". This is a scenario where in system code that was implemented or existed during the implementation process can be potentially exploited [1]. The backdoor attack provides one of the most serious forms of infiltration because it could supply multiple attack paths from a single entry point. That is why, a backdoor attack is an approach of introducing a vulnerability in the system that allows other types of malicious attacks to enter or acquire unauthorized access to the system silently [2]. The backdoor's invisibility allows the attacker to move undetected by the detection system while targeting the victims and jeopardizing the confidentiality, integrity, and availability of the organization's operational systems and services, in particular.

The primary goal of this research is to offer a thorough analysis of the cyber threat level posed by the backdoor payload. The attack's developments and innovations complicate backdoor malicious attacks, making it more difficult for security measures in the business to keep up with the latest and sophisticated attacks in order to avoid disruptions in network system services. The fact that a backdoor is a means of breaking into a system by exploiting an existing or future implementation of malware to gain unauthorized access. This means that backdoors can easily bypass traditional security protections and authentication procedures due to their discrete character and ability to remain hidden in the absence of appropriate security monitoring or assessment [3]. The problem comes when a user's conducts a security evaluation but unable to finds

https://doi.org/10.31436/ijpcc.v10i2.484

backdoor codes since they are hiding in the system. Even if current security measures have improved, backdoor attacks remain a significant issue. The fundamental reason advanced security technology cannot deal with current cyber threats is a lack of knowledge in security defensive measures, which makes them difficult to implement a proper degree of defence comply with the attacks [4].

In general, majority of backdoor research focuses on evaluating malware analysis to identify between different levels of a backdoor attack in order to correctly strategy mitigation and security assessments [5]. Most existing security mechanisms frequently fail to detect the backdoor due to its ability to move silently throughout network networks, leaving the organization's system open to unauthorized access and possible exploit. Extensive testing and analysis of the threat level of a backdoor attack were carried out to provide a comprehensive risk assessment, a mitigation strategy, and insightful knowledge about the particular attack to improve the organization's framework and defensive measures [6]. Other issues encountered were the integration of new technologies such as multi-source transfer learning, machine learning, and artificial intelligence into the current system, which made it more vulnerable and created a lot of new loopholes in the system, requiring more advanced and up-to-date information about current cyber-attacks [7].

The final motivation of this study lies within "Attackers framework". It was determined that there were frequently among the internal workforce, as it is quite easy to compromise critical data and spy on the organization's network [1]. Considering that backdoor attack allows the attacker to conduct a threatening attack by exploiting software vulnerabilities or transmitting a malicious payload and easily manipulating system information [8]. This significantly increases the frequency of cyberattacks, particularly within the corporation.

In light of the present circumstances, it is imperative for companies to carry out security evaluations and testing in order to identify any abnormal or detrimental activity. In order to carry out a comprehensive examination, it is crucial to have access to rules that ensure adherence to corporate standards and the complete utilisation of defence and prevention systems. This research is valuable because it offers extensive guidance and understanding of the seriousness of backdoor attack approaches. This enables the adoption of appropriate preventive and security measures. Hence, this ongoing investigation is of utmost importance.

The remaining sections of this work is organized as follows: Section 2 discuss the related work, Section 3 provide the research methodology, and Section 4 presented the analysis and results. Section 5 discuss the conclusions of the study.

II. Literature Review

There are many previous research studies on backdoor attacks. Crucial to these is the work of Hashemi and Zarei [9] which purpose to investigate backdoor attacks in Internet of Things (IoT) environments, with a particular focus on issues of resource management and security. A number of different detection methods and recommendations for improving the security of Internet of Things devices against backdoor attacks has been presented in the paper. Finally, the paper concluded that vulnerabilities that are associated with Internet of Things systems that need security solutions.

Qiu et al. [10] presents "Deepsweep," a framework specifically developed to counteract backdoor attack on deep neural networks (DNNs) by employing data augmentation techniques. The paper illustrates that through the diversification of the training data, Deepsweep can significantly diminish the success rate of backdoor attacks, hence bolstering the resilience of DNNs against these types of threats.

Liu et al. [11] examine backdoor attack that utilise the process of machine unlearning, a technique designed to exclude specific input from models. The paper demonstrates the ability of intentionally designed unlearning requests to introduce backdoors into the model. Finally, the study advocates for the implementation of stronger unlearning techniques in order to mitigate these weaknesses.

Chen et al. [12] research presents a technique for identifying backdoor attacks on DNN by utilising activation clustering. The approach detects clusters that exhibit abnormal behaviour by analysing the activations of neurons in response to inputs, hence identifying potential backdoor attacks. The efficacy of the technique is demonstrated in several attack scenarios, offering a means to detect corrupted models.

Al Kader et al. [13] research investigates the phenomenon of backdoor attacks on video action recognition systems. The method presents a new approach to launching attacks by utilising both visual and audio signals to embed and activate backdoors concurrently. The results emphasise the necessity for implementing more extensive safeguards in multimodal video action recognition systems.

Dong et al. [14] work introduces a technique for identifying backdoor attack using a black-box approach, even when there is a scarcity of information and data available. The paper utilises an innovative testing approach to detect anomalous model behaviours that suggest the presence of backdoors. The approach exhibits resilience and

efficacy, even in the absence of knowledge regarding the attack specifics and model structure.

Wan et al. [15] examines the occurrence of data and model poisoning backdoor attacks in wireless federated learning. The findings identify crucial obstacles and suggests future research paths to enhance the security of wireless federated learning systems.

Goldblum et al. [16] provides a thorough examination of dataset security in the context of machine learning, with a specific emphasis on the topics of data poisoning and backdoor attacks. The paper classifies different attack techniques and defence mechanisms, highlighting the significance of safe data management practices. The paper put out recommendations for improving the security of datasets and reducing possible risks.

Nguyen et al. [17] introduces a method for executing irreversible backdoor attacks in federated learning settings. Unlike conventional backdoors that may be perhaps reduced or eliminated, the proposed system guarantees the persistence of the backdoor even after substantial model changes and retraining, thereby presenting a considerable risk to the security of federated learning.

The evaluated publications collectively examine several aspects of backdoor attacks in machine learning models, with a particular emphasis on deep neural networks (DNNs) and other advanced architectures like vision transformers and LSTMs. The approaches presented aim to identify and reduce the impact of backdoor attacks through techniques such as data augmentation, activation clustering, and black-box testing. Thorough surveys and evaluations emphasise the difficulties in ensuring the security of datasets, the risks associated with backdoor learning, and the vulnerabilities present in IoT contexts. These findings underscore the importance of implementing strong defence mechanisms. The papers also explore innovative attack techniques such as switchable backdoors and irreversible backdoor attacks in federated learning, demonstrating the dynamic nature of risks and the significance of adaptable security mechanisms.

Although there have been significant improvements, there are still some areas of research that have not been fully explored. First and foremost, there is a requirement for stronger and more scalable defence systems that can adjust to different methods of assault and model structures. Existing techniques frequently depend on particular assumptions or restricted data, hence diminishing their capacity for generalisation. Moreover, the interaction between various forms of attacks, such as data poisoning and backdoor attacks, has not been well investigated, hence neglecting possible vulnerabilities. Further work is needed to understand the impact of backdoor assaults on new technologies, such multimodal learning and IoT systems. Furthermore, there is a deficiency in the establishment of standardised evaluation frameworks and standards to measure the efficiency of defence methods.

Future research should prioritise the development of comprehensive defence frameworks that incorporate various detection and mitigation strategies, hence improving their resilience and scalability. Investigating the interconnections between various forms of attacks and defences could yield a more holistic comprehension of security weaknesses. Researchers should give priority to developing standardised benchmarks and evaluation criteria to simplify the comparison and enhancement of defence mechanisms. Examining the consequences of backdoor assaults on emerging technologies like edge computing and quantum machine learning can reveal innovative risks and countermeasures. Furthermore, the development of more sophisticated methods for securely managing data and protecting privacy in federated and distributed learning settings will be essential in addressing the changing nature of backdoor attacks.

III. Research Methodology

In the methodology section, a proper assessment and penetration testing were performed on the level of effect of a backdoor attack on organization network system administration. This test is necessary to analyse the behaviour of the backdoor payload to estimate the degree of threat posed by the backdoor. This methodology outlines a step-by-step approach to properly conducting penetration testing on advanced backdoor attacks, particularly on organization operations utilizing the Kali Linux penetration operating system. The algorithms employed in this methodology are replications of actual cyber-attacks. The testing will help an organization to understand the backdoor attack mechanism, and a comprehensive insight to prepare for cyber-attacks and system exploitation.

First, due to the risk of handling the backdoor during malware analysis, a proper virtual lab is required during the environment setting-up phases. The virtual environment allows the tester to be more flexible and independent during the penetration testing because the virtual machine is a separate network environment, so it will not affect the local machine or network system from any unintended breach or unexpected attack from the malicious software. then create the representation of organization network architecture to be able to assess the backdoor attack and analyse the level of the attack vector and also to look out if there is the possibility of system vulnerabilities that might occur. This allows for a close comparison of the findings to real-world attack scenarios.

A. Architectural Framework

The architectural framework of this current study lies with the general computer network security issues. For the

https://doi.org/10.31436/ijpcc.v10i2.484

purpose of this specific research, the architectural framework is based on a traditional network environment. Consequently, the architecture that has been suggested is shown in Figure 1. What follows is a condensed version of it: An attack that occurs outside of the network, which is the presence of the attacker, is characterised by the attacker's ability to rely on an open port, where it searches for and eventually locates one. By the "listening port" is where you will find acquire.

Following the acquisition of knowledge regarding both the opening and listening port, the attacker proceeded to create a connection of its own. It is possible that the connection is with a "Internet of Things device," a "control server," or a "Network that remain undetectable" in the system, but this cannot be determined without doing an inspection. In both of these open instances, the attacker has the potential to get root privileges of the control server as well as remote access to the "Internet of Things device." When taking into consideration the scenarios that have been provided thus far, the most important and core aspect of the backdoor attack is the capability to enter the network zones without going via the door that is expected, despite the fact that it is not permitted to go through the door that everyone is familiar with. The term "Backdoor" refers to the fact that attackers have the potential to establish another door, which is why it is an attack mechanism. In light of this, the purpose of this research is to investigate the potential backdoor that the attackers utilised, as well as to assess the possibilities of accessing these doors and the ways in which they might be presented.

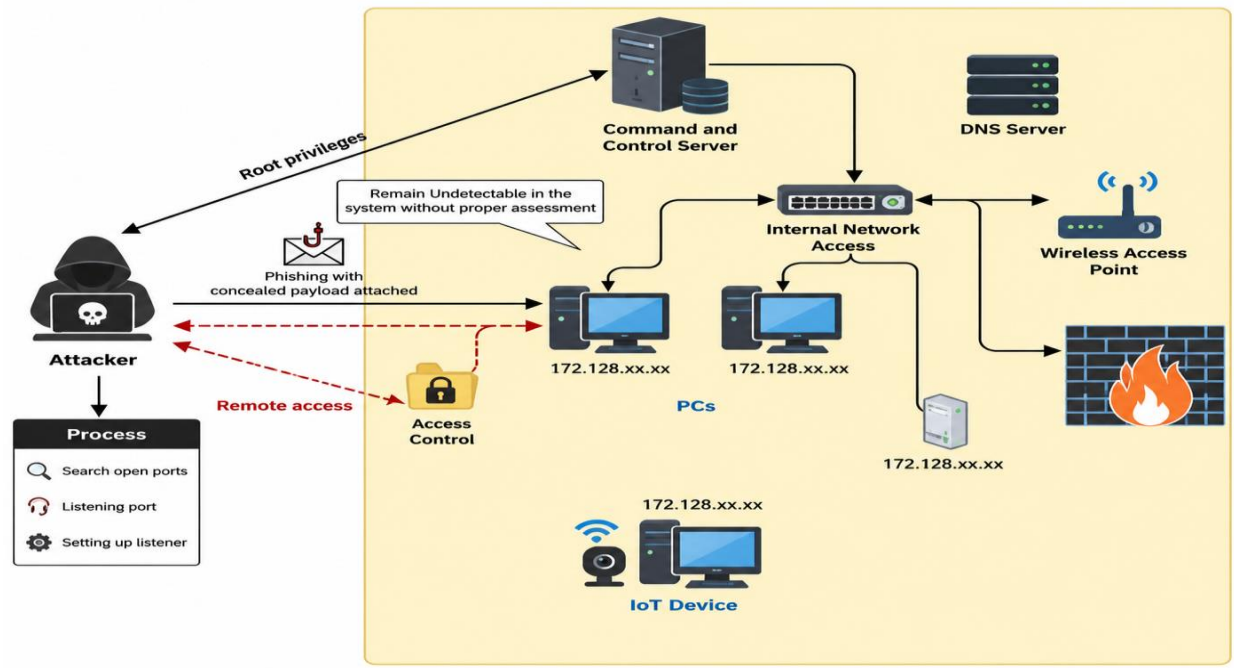

Fig. 1 The Research Architectural Framework

B. Experimental Simulation

The experimental simulation focuses on accurately created simulation approaches to estimate the threat level of backdoor attacks. This technique was implemented in a controlled environment, where all possible scenarios were thoroughly analysed and experimented with. In order to guarantee the effectiveness of the methodological approach, it is crucial to include the requirement for testing. The simulation requirement includes the necessary system specifications for evaluating the extent of the backdoor attack approach. The "Attack Simulation on the Kali Linux" was conducted with the "Metasploit Framework" serving as the listener and for deploying the backdoor. The simulation framework includes Test Servers that utilise a "Kali Linux" operating system, which is installed on a virtual machine. The "msfvenom" tool is employed to generate payloads, while specialised Python scripts were created to target certain attack situations and targets.

An experimental setup was created using a network environment that replicated the organization's network architecture and important systems. A test environment was employed within the Virtual Machine to guarantee the isolation of the network from the production network, thereby avoiding any unwanted repercussions. A strategic plan for launching an assault and a technique for delivering

https://doi.org/10.31436/ijpcc.v10i2.484

hidden malicious software using msfvenom were devised in response to the given circumstances. The covert malicious software was successfully executed. Execution and monitoring are responsible for initiating the connection session between the system and the target in order to get access. After obtaining access, an assault was initiated utilising the Metasploit framework. A thorough examination and assessment of the functionalities of the backdoor, together with the system's.

After clearly outlining the simulation objective, the testing then began with building an isolated environment by configuring the virtual system to minimize any unwanted breaches and effects while also ensuring a safe testing environment.

Then, mimic the organization's network system architecture for exploitation testing. In this situation, the research will use Kali Linux technologies such as msfvenom, metasploit framework, custom Python script, veil framework, and MulVal. The purpose of msfvenom is to construct a backdoor payload containing malicious code that will be sent to the target machine.

While the use of the Metasploit framework is to listen to the victim's machine via a connected backdoor payload that was delivered, the Python script used was to construct a specific malware payload to exploit the specific target within the system. Other than analysing the capability of the monitoring tools to detect backdoor payloads, detecting any intrusion within the system is critical when utilizing the Windows Defender system.

Next, create possible scenarios for the simulation that closely reflect the actual cyberattack.

- Scenario 1: Phishing attack using the delivery emails method
- Scenario 2: Search for any vulnerabilities in outdated software
- Scenario 3: Manually installing backdoor payload in the victim’s machine

Further the simulation process by defining the entry point of each scenario above to completely compromise the system and gain full control to enable the remote access control. after the acknowledgment of the entry point, then by using the created payload to deliver to the target system. After delivery is successful, establish the session to connect the victim's machine via the delivered backdoor. The attack can only be delivered if the backdoor is triggered even after the session is established. Select the attack vector method offered by the Metasploit framework, for example accessing command prompt, network monitoring, activity logs and control remotely, then observe the system activity and the interaction between backdoor and detection system. Is the backdoor visible by the intrusion detection system even after launching an attack?

Following the penetration, to avoid leaving a visible footprint or traceable artifact, clean the activity by ending the session and erase all the data artifacts and residual malware to remain hidden in the system. Collect all the information gained and start analyzing the capability of the backdoor payload and detection system towards the integration of the current security technologies. Document all the findings, including timestamp, simulation process, backdoor, and detection system capabilities. All the insight collected can be used for training, assessing software vulnerability, and estimating the degree of threat related to backdoor attacks. To ensure the effectiveness of the testing, schedule the continuous testing within the organization to effectively respond to the attack, and strategize the mitigation plans.

C. Experimental Setup

The experiment setup for this research is presented in Figure 2. The flowchart provides an overview of all of the procedures that are involved in the experimental scenarios. It also illustrates how a traditional backdoor functions by utilising particular tools such as Msfvenom and the Metasploit framework. In order to gain a deeper comprehension of the manner in which backdoor attack operate in a variety of contexts, the experimental scenarios were designed for that. Furthermore, the purpose of this experiment was to repeat the various attack scenarios that were discussed before, as well as to investigate how the backdoor evolved within the context of the incorporation of new technologies.

The first approach involves updating to the most recent version of Kali Linux. This version is required to obtain cutting-edge tools and security patches if the operating system was installed.

Next, launch the terminal and begin the process. Make use of msfvenom in order to create the payload. An application that can construct pre-created payloads is called Msfvenom. As an illustration, in order to develop a payload that employs reverse TCP as a type of attack that is tailored exclusively for Windows operating systems, Msfvenom was used.

As we move further, we begin the exploitation session by opening a listener, which is a component of the Metasploit framework. This allow us to link the payload to the host system and the listener port. In order to configure the Listener, start the msfconsole application and use the (multi/handler) command. After that, enable the remote exploit module. Take advantage of the exploit known as (multi/handler) and set the payload to the one that was generated by msfvenom in order to handle the reverse connection.

https://doi.org/10.31436/ijpcc.v10i2.484

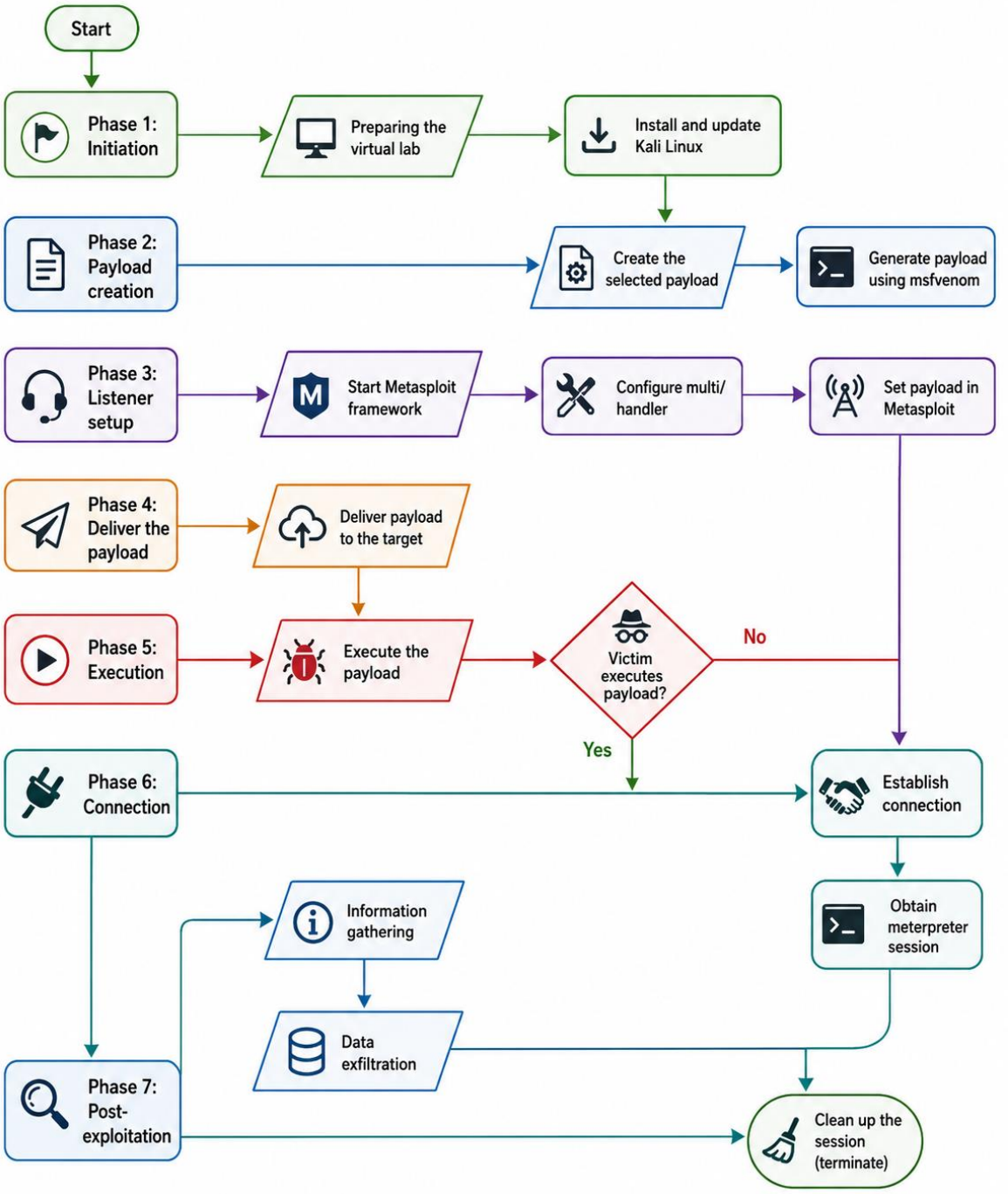

Fig. 2. The Research Methodological Flow

Next, in order to configure the listener, you will need to ensure that the values of LHOST (the attacker host) and LPORT (the listening port) are identical to those of the payload. You can start the listener by using the exploit command, and you can view the several attack routes by using the help/options command afterwards. When everything is ready, the next step is to hand over the payload to the victim. While there are a variety of approaches to delivering the payload in this experiment, we make use of two approaches: "Social engineering" and "Direct Transfer". The social engineering involves manipulating the victim in order to deliver the payload to the target. Sending an email that contains a malicious file attached to it or hosting it on a website that has been infiltrated is what this entails.

https://doi.org/10.31436/ijpcc.v10i2.484

Whereas the "Direct Transfer" deals with providing the payload in a direct manner. In the event that physical access was available or a pre-existing network connection was obtained, the payload may be sent immediately through the use of any storage devices.

After the payload has been properly delivered, the victim or the machine must activate the backdoor in order to create a successful connection in Msfconsole. Only then will the payload be executed. Following the establishment of a session in Metasploit, the listener should be able to detect the reverse connection, which will provide you with a Meterpreter session. This allow obtaining access to the victim system or machine. Several post-exploitation actions can be carried out once access has been gained. These operations include "Privilege Manipulation", "Data Exfiltration", "Installing Malware/Trojans," and "eavesdropping network".

Last but not least, the session is terminated by removing any leftover files that have been created. Once the testing phase is finished, it is imperative to verify that the session is terminated correctly in order to eliminate all traces of the footprint and conceal the backdoor. Erase any artefacts that may be present on the target system, such as any files that are still open or system logs, in order to prevent detection and guarantee the test's dependability

IV. Results

The experimental result of this study is presented in this section. Following extensive testing conducted under conditions mirroring real-world settings, we can now share the initial findings of the penetration testing. The research primarily aimed to assess the level of harm posed by backdoor attack methods to the organization's operations. The testing involved replicating different backdoor attack paths with a single payload development, interpreting software vulnerabilities, and analysing rogue websites in a controlled setting to prevent any unwanted consequences.

The effects of the breach were primarily focused on identifying vulnerabilities and assessing the organization's ability to recognise and respond to them. During the testing, the attack scenario was carried out. The first scenario involves a phishing attempt that aims to deliver a backdoor payload. The objective is to assess the effectiveness of email filters and user knowledge in preventing this type of attack. The method employed entailed dispatching phishing emails to targeted people, which included a malevolent attachment. A total of twenty employees were specifically selected as targets, leading to a 10% chance of successfully opening the payload and a 5% probability of successfully executing it (see Figure 3). The SIEM system identified 2 out of 5 instances where the payload was successfully executed, while the remaining 3 instances were spotted by EDR tools. The mean duration from detection to initial response was 15 minutes. The attack's impact was mitigated by achieving isolation before any important data was accessed.

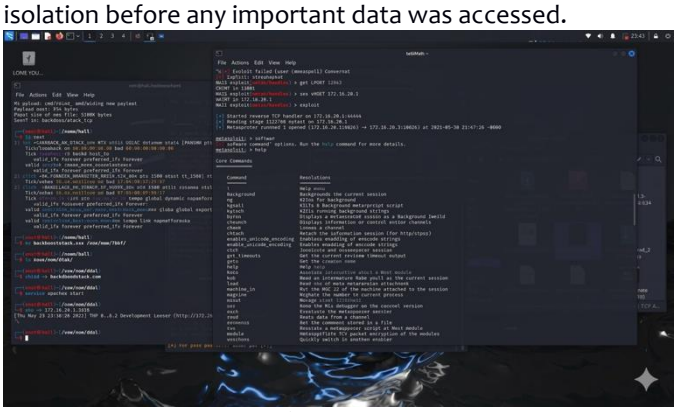

Fig 3. The first Scenario Attack Vector

Scenario 2: Exploiting Software Vulnerability: This scenario examines the impact of a backdoor on old software. The software is susceptible to exploitation due to identified flaws, such as inadequate security patches. The backdoor payload can readily infiltrate the system via the susceptible software. Thankfully, the Intrusion Detection System (IDS) promptly detected the abnormal behavior within the software at an early stage. Nevertheless, the system's average reaction time for isolating the affected software was 10 minutes, a somewhat sluggish performance for response isolation (see Figure 4). During the time it took for a response to be initiated, the attackers were able to effectively steal and remove half of the essential data and monitor network logs. This has had a substantial impact on the organization's operations, which could be further disrupted if an appropriate strategy to mitigate the situation is not implemented.

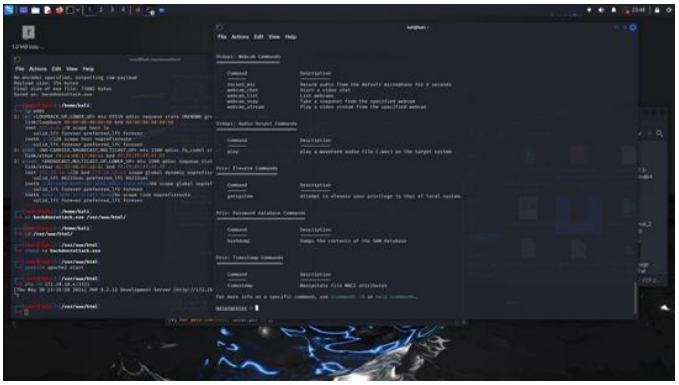

Fig 4. The Second Scenario Attack Vector

This research emphasizes the suggestions and recommendations for enhancing overall security management, such as the integration of sophisticated email filtering. Technologies and methods used to detect unauthorized access or intrusion. Enhancing phishing

awareness is crucial for mitigating the majority of phishing-related attacks. Furthermore, in order to minimize software vulnerabilities and maintain high levels of security, it is imperative to regularly perform data backups, manage patches, and update software. Regular testing is necessary to prevent potential "zero-day" malware. Enhancing security information and event management, intrusion detection systems (IDS), and intrusion prevention systems (IPS), along with adopting strong authentication and access control, can bolster security measures by enabling staff training, continuous monitoring, and authentication.

Finally, . A phishing assault has a 5% success rate in delivering a backdoor payload. This attack method involves using phishing emails to transfer the backdoor payloads, which has led to successful executions. Although SIEM and EDR systems detected most occurrences, a few initially went unnoticed. During the test, once the machine system was compromised, it required an average of 7 minutes to completely gain control. The testing yielded restricted horizontal movement and no essential data retrieval before confinement. Exploiting software flaws, particularly a known weakness in an outdated software program, is the most effective method of breaching security in this test, when compared to other types of attacks. The Intrusion detected the anomalous behavior within a span of 3 minutes by utilizing Intrusion Detection Systems (IDS). The system response involves the isolation of affected data or software within a time frame of ten minutes. Exploiting software vulnerabilities can result in the partial exfiltration of data before the machine is isolated. Insiders executing insider threats may employ a USB device to install harmful or backdoor malware, either acquiring unauthorized access or detecting vulnerabilities in the code. The endpoint protection system recognized it within a span of two minutes. The system will be quarantined within an average duration of 5 minutes. The intrusion detection system, also known as SIEM, has a little impact due to its rapid and efficient detection and response capabilities.

V. Conclusions

The primary objective of this study's penetration testing was to analyses and evaluate the level of risk posed by backdoor attack techniques, specifically within organizational operations, in order to quantify the extent of the backdoor threat. The test involved many attacks, which yielded valuable data on the detection capability, response time to backdoor attacks, stealthiest of the backdoor attack, and effectiveness of mitigation measures. The vulnerabilities revealed encompass a deficiency in email security, leading to a significant rate of success for phishing emails in circumventing the organization's current safeguards. Moreover, the high occurrence of crucial vulnerabilities in obsolete software programs presents a substantial danger as security threats and attacks persistently develop. Insider threats are a worry because of the insufficient monitoring and access controls for personnel within an organization, which can make them potentially dangerous. Ultimately, this study emphasizes the significance of regularly conducting testing, monitoring systems, and upgrading Intrusion Detection Systems (IDS) or Intrusion Prevention Systems (IPS) inside an organization. This ensures that vulnerabilities can be quickly identified and backdoor assaults may be prevented in their early phases. As advanced technology becomes more integrated, security measures must enhance defensive strategies to keep up with the quickly growing cyber threats. This abstract provides a concise overview of the research results, presenting thorough and unambiguous recommendations for organizations to enhance their defensive strategies and offering guidelines for penetration testers conducting assessments on backdoor Attack.

Acknowledgment

This research is made possible and supported by UMP-IIUM Sustainable Research Collaboration 2022 Research Grant (IUMP-SRCG22-014-0014).

Conflict of Interest

The authors declare that there is no conflict of interest.

References

[1] H. Yang, K. Xiang, M. Ge, H. Li, Lu R, S. Yu. A comprehensive overview of backdoor attacks in large language models within communication networks. IEEE Network. 2024 Feb 20.

[2] J. Dai, C. Chen, Y. Li. A backdoor attack against lstm-based text classification systems. IEEE Access. 2019 Sep 13; 7:138872-8.

[3] S. Yang, J. Bai, K. Gao, Y. Yang, Y. Li, S.T. Xia. Not all prompts are secure: A switchable backdoor attack against pre-trained vision transfomers. InProceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2024 (pp. 24431-24441).

[4] Y. Li, Y. Jiang, Z. Li, S.T. Xia. Backdoor learning: A survey. IEEE Transactions on Neural Networks and Learning Systems. 2022 Jun 22;35(1):5-22.

[5] S. Liang, M. Zhu, A. Liu, B. Wu, X. Cao, E.C. Chang. Badclip: Dual-embedding guided backdoor attack on multimodal contrastive learning. InProceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2024 (pp. 24645-24654).

[6] S. Seo, Kim. D. Study on inside threats based on analytic hierarchy process. Symmetry. 2020 Jul 29;12(8):1255.

[7] Y. Gao, B.G. Doan, Z. Zhang, S. Ma, J. Zhang, A. Fu, S. Nepal, H. Kim. Backdoor attacks and countermeasures on deep learning: A comprehensive review. arXiv preprint arXiv:2007.10760. 2020 Jul 21.

[8] B. Li, Y. Cai, H. Li, F. Xue, Z. Li, Y. Li. Nearest is not dearest: Towards practical defense against quantization-conditioned backdoor attacks. InProceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2024 (pp. 24523-24533).

[9] S. Hashemi, M. Zarei. Internet of Things backdoors: Resource management issues, security challenges, and detection methods.

Transactions on Emerging Telecommunications Technologies. 2021 Feb;32(2):e4142.

[10] H. Qiu, Y. Zeng, S. Guo, T. Zhang, M. Qiu, Thuraisingham B. Deepsweep: An evaluation framework for mitigating DNN backdoor attacks using data augmentation. InProceedings of the 2021 ACM Asia Conference on Computer and Communications Security 2021 May 24 (pp. 363-377).

[11] Z. Liu, T. Wang, Huai M. Miao C. Backdoor attacks via machine unlearning. InProceedings of the AAAI Conference on Artificial Intelligence 2024 Mar 24 (Vol. 38, No. 13, pp. 14115-14123).

[12] B. Chen B, Carvalho W, Baracaldo N, Ludwig H, Edwards B, Lee T, Molloy I, Srivastava B. Detecting backdoor attacks on deep neural networks by activation clustering. arXiv preprint arXiv:1811.03728. 2018 Nov 9.

[13] A.l. Kader H.A. Hammoud, S. Liu, M. Alkhrashi, F. Albalawi, Ghanem B. Look Listen and Attack: Backdoor Attacks Against Video Action Recognition. InProceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2024 (pp. 3439-3450).

[14] Y. Dong, X. Yang, Z. Deng, T. Pang, Z. Xiao, H. Su, J. Zhu. Black-box detection of backdoor attacks with limited information and data. InProceedings of the IEEE/CVF International Conference on Computer Vision 2021 (pp. 16482-16491).

[15] Y. Wan, Y. Qu, W. Ni, Y. Xiang, L. Gao, Hossain E. Data and model poisoning backdoor attacks on wireless federated learning, and the defense mechanisms: A comprehensive survey. IEEE Communications Surveys & Tutorials. 2024 Feb 7.

[16] M. Goldblum, D. Tsipras, Xie C, Chen X, Schwarzschild A, Song D, Mądry A, Li B, Goldstein T. Dataset security for machine learning: Data poisoning, backdoor attacks, and defenses. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2022 Mar 25;45(2):1563-80.

[17] T.D. Nguyen, Nguyen T.A, Tran A, Doan K.D, Wong K.S. Iba: Towards irreversible backdoor attacks in federated learning. Advances in Neural Information Processing Systems. 2024 Feb 13;36.