Journal of Computer Science

Review

Evaluating Common Reconnaissance Tools and Techniques for Information Gathering

[1]Isaac Odun-Ayo, [1]Emmanuel Owoka, [1]Otavie Okuoyo, [1]Opeyemi Ogunsola, [1]Obaro Ikoh, [1]Olumide Adeosun, [1]Deborah Etukudo, [1]Victoria Robert and [2]Gabriel Oyeyemi

[1]Department of Computer and Information Sciences, Covenant University, Ota, Nigeria
[2]Covenant Applied Informatics and Communications - Africa Centre of Excellence, Covenant University, Ota, Nigeria

Article history
Received: 10-08-2021
Revised: 23-10-2021
Accepted: 04-11-2021

Corresponding Author:
Isaac Odun-Ayo
Department of Computer and Information Sciences, Covenant University, Ota, Nigeria
Email: isaac.odun-ayo@covenantuniversity.edu.ng

Abstract: A reconnaissance attack is a commonly overlooked step in penetration testing but a critical step that could help increase the effectiveness of an attack on a target. However, it is a common attack faced by companies and institutions, among others. It enables the attacker or penetration tester to gain valuable information on the target and select the best tools and methods that would make the attack successful. This study aims to identify and review existing state-of-the-art methodology for reconnaissance attacks based on certain techniques and evaluation metrics which will be beneficial to professional, ethical hackers in selecting the best-fit tool for a successful reconnaissance attack and enlighten organizations and the general public of the potential harm of a successful reconnaissance attack. This study explored seven online databases, which include Springer, Elsevier, Wiley, IEEE, ACM, ArXiv and Google Scholar. A total of 1306 publications were retrieved. Several screening criteria were executed to select relevant articles. Finally, 19 articles were identified for in-depth analysis. A quantitative evaluation was conducted on the selected articles using two search strategies: Techniques and source. A Quantitative Analysis (QA) was conducted on the selected articles and the outcome based on existing reconnaissance tools shows that 95.2% of the tools allowed experts to gather information by running their necessary command from the command line. While 4.8% of the tools do not provide a command-line interface allowing users to launch it from the command line interface while specifying some parameters to customize how it runs. 61.9% of the tools are network-based as they can be used to gather about the target's network infrastructure such as protocols, ports, DNS, IP address, hosts and the general network architecture. At the same time, 28.5% could be classified as hybrid as they allow the attacker to gather system-based and network-based information. This study concludes with findings that show that some of the tools operate in a different capacity; the best-fit tool is massively dependent on the attacker or penetration tester and the scenario's situations. Therefore, a tool should be selected based on the user's preference and the attack style.

Keywords: Reconnaissance, Information Gathering, Cybersecurity, Social Engineering, Techniques

Introduction

These days, almost all we do is on the internet available to the public. This opens us all to a cyber-attack (being hacked). The phases of cyber-attack generally follow the same pattern as a traditional crime (Mazurczyk and Caviglione, 2021). Benjamin Franklin said, "By failing to prepare, you are preparing to fail." Abraham Lincoln said, "I will prepare and someday my chance will come." When starting something, as a matter of priority, preparation is key. Also, when thieves want to rob a bank, they observe vital details of the bank to coordinate their operation. Observing essential information of the bank is a big part of their preparation and attack. This is known as Reconnaissance.

A reconnaissance attack is also known as information gathering, is the first phase of a cyber-attack. It is an unauthorized retrieval of information about a target to identify vulnerabilities that an attacker can exploit. It helps the attacker or penetration-tester select the best tools and methods to carry out a successful attack on a target. It is a very dangerous attack that could directly impact the magnitude of the effect of a completed attack. Reconnaissance can be passive or active. Passive reconnaissance is more like a light, non-in-depth approach to gathering information on a target. This goes on without alerting the target. An active reconnaissance is a more in-depth approach to gathering information on a target and this alerts the target.

The primary objective of the reconnaissance phase is, therefore, to map a "real-world" target (a company, corporation, government, or other organization) to a cyber world target, where "cyber world target" is defined as a set of reachable and relevant IP addresses (Cuppens-Boulahia, 2012). You can say that when an attacker gathers information about a target, the attacker seeks to understand what the target does when the target does what it does, how the target carries out operations and why the target carries out operations, all in an attempt to spot vulnerabilities and intelligently craft or execute an exploit on found vulnerabilities in the cyber-space.

This potentially dangerous attack is very much common and it more often than not happens without the targets' knowledge. With the world being more digital as time goes by, we believe there would be a rise in cyber-attacks. Individuals and companies, mainly medium to small scale enterprises, do not understand the gravity of safeguarding their cyber-space. Most do not actively manage their information in the cyber-space as well as in the real world.

This study aims to help individuals and companies comprehend the impact of reconnaissance attacks and how to mitigate them and also to help cybersecurity professionals, ethical hackers make an educated decision in selecting the best-fit tool for performing a reconnaissance attack. The objective of this study is to evaluate common reconnaissance attack tools, what they do and their features.

This study also contributes to the United Nation's Sustainable Development Goals. It positively promotes SDG 4 - Quality Education, SDG 8 - Decent Work and Economic Growth and SDG 9-Industry, Innovation and Infrastructure (United Nations, 2015).

Out of the 17 goals adopted by all United Nations Member States in 2015 for the 2030 Agenda for Sustainable Development, this study was able to identify three goals which revolves around promoting quality education and enhancing lifelong learning opportunities, adding to economic growth and decent work and contributing to innovation and industry (United Nations, 2015).

Related Works

The five pillars of information assurance are confidentiality, integrity, availability, non-repudiation and authentication. An attack that violates one or more of these pillars is considered a cyber-attack. Cyber-attacks are in phases, but the first phase of any cyber-attack is reconnaissance. In this phase, the weak points of the target are identified. Critical information like the victim's IP addresses, home address, telephone number, frequent hangout, dark secrets, security policies, etc., are collected and ways of bypassing the victim's defense systems are also noted (Sanghvi and Dahiya, 2013).

Cyber-attacks are growing in terms of complexity and volume. According to Industry Week, in 2018, spear-phishing and spoofing attempts of business emails increased by 70 and 250%, respectively and ransomware campaigns targeting enterprises had an impressive 350% growth. In general, economic damages are relevant due to the need to detect and investigate the attack and restore the compromised hardware and software. To give an idea of the impact of the problem, the average cost of a data breach has risen from $4.9 million in 2017 to $7.5 million in 2018. To make things worse, attackers can now use a wide range of tools for compromising hosts, network appliances and Internet of Things (IoT) devices simply and effectively, for example, via a Crime-as-a-Service business model (Mazurczyk and Caviglione, 2021).

Social engineering is probably the oldest family of techniques used for reconnaissance and it is extraordinarily effective as it exploits the weakest link in security: Humans. In essence, social engineering tries to manipulate and deceive victims by misusing their trust and convincing them to share confidential information or to perform activities that can be useful to the attacker, for example, downloading and installing a key logger. It can also significantly decrease the time needed to gather information and often requires minimal or non-technical skills (Mazurczyk and Caviglione, 2021)

In the discussion of Kumar *et al*. (2015), it was seen that social engineering is a non-technical method hackers use to trick people into breaking security procedures. It was also identified as a manipulative way hackers use to get confidential information from people. it was discussed in their paper that social engineering is essential because it deals highly with human interactions.

Social engineering attacks have grown beyond hardware and software and are targeted at the humans in those organizations. In the study of (Chantler and Broadhurst, 2008) it was discussed that the attacks done by hackers to gain access into an organization are in levels. The attack first starts at getting the user's information and progresses to having access to the organization's computer and then attacking the organization's control program, which is the target, where the whole database of the organization that involves the financial information is exposed (Aldawood and Skinner, 2019).

In their paper, (Aldawood and Skinner, 2019) discussed knowledge-based measures and Geoffrey Skinner noted that the more education given to employees

and users about social engineering, the less vulnerability to employee activity performance online. Also, a technique that was considered is email security; organizations should try to use Domain-based Message Authentication Reporting and Conformance (DMARC) and real-time blocking, which maintains email security gateways through tools to identify origins of email (SPF) sent and included cryptographic signatures to know the validity of the email and also identifies malicious emails (Aldawood and Skinner, 2019).

A survey was conducted in Aldawood and Skinner, (2019) and the results showed that most people did not know about the intrusion they had experienced in past years. The authors stated that employees are responsible for securing the organization's data. If they are careless, they expose vital information to hackers that will gain access to them-some tools such as proxy configurations, malware detection tools, neuro-fuzzy inference systems.

Organizations can adapt the neuro-fuzzy inference system, which uses neural networks to create a self-predicting phishing detection that places hackers on blacklists making it difficult to generalize (Aldawood and Skinner, 2019).

Unfortunately, the evolution of the Internet, the diffusion of online social networks, as well as the rise of services for scanning smart appliances and IoT nodes lead to an explosion of sources that can make the reconnaissance phase quicker, easier and more effective. This could also prevent contact with the victim or limit its duration, thus making it more difficult to detect early and block reconnaissance attempts. Therefore, investigating the evolution of techniques used for cyber reconnaissance is of paramount importance to deploy or engineer effective countermeasures (Mazurczyk and Caviglione, 2021).

Research Methods

The research protocol for this study was based on several search strategies such as techniques, source, attack types and status. This systematic review started with preliminary searches to identify existing studies and assess the volume of potentially relevant articles in this study domain, which was included as sources. The search techniques in were adopted and modified for this study and the specific objective of this study is to identify, review, analyze and establish an easy understanding of the state-of-the-art methodology utilized in evaluating common techniques and tools for reconnaissance attacks. To achieve the aforementioned objectives, research questions were prepared as shown below:

RQ1: What are the relevant values of the existing studies for reconnaissance tools and techniques
RQ2: What are the existing reconnaissance tools and techniques
RQ3: What are the methods of evaluating common tools for reconnaissance attacks

The above-started research questions which form the foundation for embarking on this study are entwined and concurrently explored

Search Strings

The search strategy utilized for obtaining the relevant studies in this study was done based on the following criteria:

1. Obtain relevant keywords from the research questions
2. Recognize distinct synonyms and spellings for the keywords
3. Identify keywords in relevant articles
4. Utilize "AND" and "OR" to relate relevant keywords

The outcome of the search query used for searching relevant articles is as follows: (Reconnaissance AND tools OR techniques) OR ("reconnaissance attack" OR "reconnaissance attack techniques") OR (reconnaissance AND evaluation OR metrics) OR ("description of reconnaissance techniques" OR "concept of reconnaissance tools").

Search Selection Strategy

The primary search for this study was targeted at online research databases such as Springer, Elsevier, Wiley, IEEE, ACM, ArXiv and Google Scholar. The aforementioned search string was utilized for advanced search in the highlighted databases. The systematic review process and selection of the relevant studies at different phases are illustrated in the PRISMA flowchart as shown in Fig. 1.

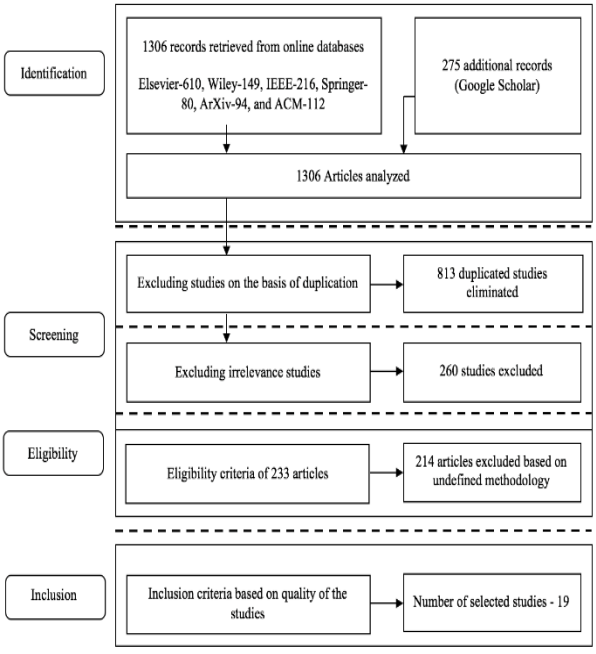

Fig. 1: PRISMA flowchart of the review process and selection of the articles

After defining the sources, criteria and process for selecting the studies, a quantitative evaluation was further demonstrated to identify new contributions, techniques, measures and applications presented by researchers in the study domain.

Search stage 1 (Extracting information) is based on extracting relevant information from the aforementioned databases. A thorough search was executed in the seven databases with their respective output summing up to 1306 articles in total and this acts as a set of possible articles for further selection, as shown in Table 1.

Search stage 2 (Screening criteria): Based on the output obtained from Table 1, a total of 1306 possible articles were retrieved. The first screening was executed based on duplication criteria and 813 articles were found to be duplicated among the seven databases. The screening process was further conducted based on the irrelevant title of the articles and 260 articles were considered irrelevant for these studies.

Search stage 3 (Eligibility): A full text-based selection criterion was executed to extract relevant studies and 214 articles were removed based on undefined methodology.

Search stage 4 (Inclusion): Based on the above-stated research questions, a quality assessment was initiated for the remaining articles. Issues concerning the selection of the studies were resolved by the authors and finally, 19 primary studies were selected for this study.

Search Strategy

This section presents and discusses the various search strategy adopted for this study using 19 selected relevant studies. The 19 relevant studies were published as follows: 7 Articles from workshops, 6 appears in journals,

4 from conference proceedings and 2 from books, as shown in Fig. 2. In this study, the search strategy was divided into four classes to explore all contributions made by previous researchers in the study domain.

The aforementioned databases were used for information extraction in this study. The keywords, titles and abstracts were used to execute a search query for published books, conference proceedings, workshops and journals.

Techniques

This subsection classified the techniques used in the selected studies as follows:

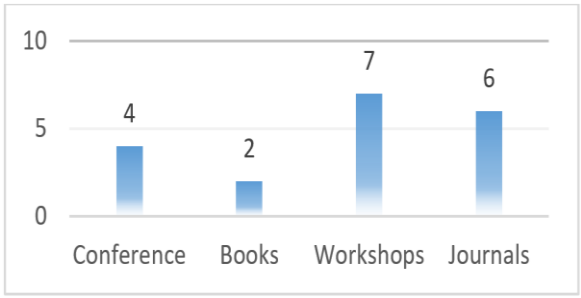

Fig. 2: The number of collated studies

1. Existing reconnaissance tools: Twenty-one (21) existing reconnaissance tools were identified from the selected articles. The Twenty-one (21) existing techniques are Msfconsole, Sqlmap, Nmap, Dnseunm, Dmitry, Shodan, Dnmap, Hping3, Burp Suite, Dns Recon, Nbstat, Metasploit, Dnswalk, Nbtscan, Wireshark, Dns Tracer, Nikto, Massscan, Faraday, Ghost Phisher, Theharvester. These tools are discussed in Section 3.1, which answers RQ2
2. Existing reconnaissance techniques: Seven (7) existing reconnaissance techniques were identified from the selected studies. The seven (7) selected techniques are phishing, pretexting, baiting, tailgating, dumpster diving, watering hole attack and spear phishing. These techniques are discussed in 3.2
3. Methods of evaluating common techniques for reconnaissance attacks: Fourteen (14) existing methods for evaluating common techniques for reconnaissance attacks were identified. These techniques are discussed in 4, which answers RQ3

Analysis of Papers

This section discussed the findings of this study and the result of the two search strategies were analyzed in detail. A detailed discussion of our findings with respect to the outline research questions was stated in different subsections with a concise interpretation of our findings. A word cloud analysis was done using the titles of the selected studies on Orange machine learning development environment and the result obtained is depicted in Fig. 3, with 'Penetration' having the most frequent occurrence followed by 'Testing' and 'Security'.

Search Strategy 1: Source

The first search exercise was executed using a hierarchical search strategy to identify related articles using relevant keywords and the paper titles before a final search strategy was developed. The search for related works was conducted for articles between 2013 and 2021 from the aforementioned online databases.

Fig. 3: Generated keywords from Titles of selected studies

The answer to RQ1 can be expressed in Fig. 4, which shows the returned results for relevant studies sources, while Fig. 5 shows the number of relevant articles based on the publication year.

RQ2: Reconnaissance Tools

In this section, we present and discuss the existing reconnaissance tools explored from the 19 selected studies. Some of the existing reconnaissance tools include.

Msfconsole

It is an interface to the Metasploit Framework. It provides an all-in-one centralized console that allows one to efficiently have access to virtually all the options available in the MSF. It makes hacking easier and an indispensable tool for the red hat hackers and blue hat hackers. Metasploit integrates with other tools like Nmap during the information-gathering phase, thus making it a cherished tool for most pen testers. Once a weakness has been identified, all you need do is search the database of Metasploit for the exploit (a script or program software that helps hackers to have control over the system) that will crack open and give you access to the system. It provides a command-line interface only. An upgraded version of Metasploit is the Metasploit pro that not only uses a command-line interface but comes with a web interface. It is used for both active and passive information gathering and helps pen testers simulate real-world attacks, collect data and provide remedies for found exploits. It runs on Mac, Linux and Windows operating systems Metasploit framework is a powerful tool for exploiting a remote target machine. With more than 900 attacks obtained by multiple combinations of payloads and exploit types, the ever-increasing need for patching the vulnerabilities in the system can be dealt with a great deal of information about them and the risk of an attack happening by exploiting a particular vulnerability (Timalsina and Gurung, 2017).

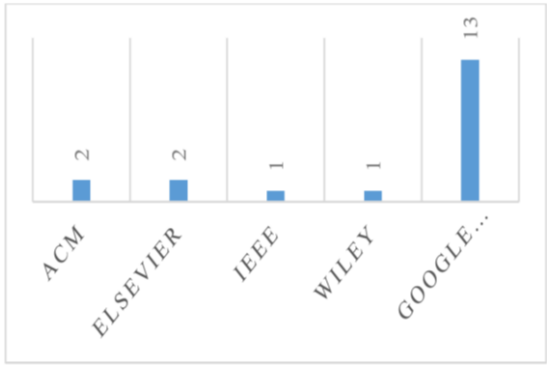

Fig. 4: Sources and number of relevant articles

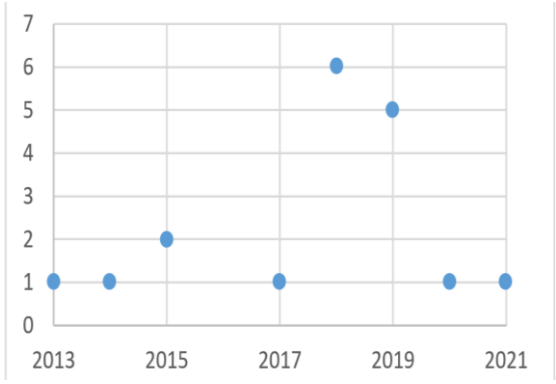

Fig. 5: Published year and number of relevant articles

Table 1: Number of articles extracted from databases

S/N	Database	Number of articles
1	Springer	80
2	Elsevier	610
3	Wiley	149
4	IEEE	216
5	ACM	112
6	ArXiv	94
7	Google Scholar	275

Table 2: Features of reconnaissance tools

Reconnaissance tools	Open-source	Supporting Platform (mobile)	Web Based	Supporting Platform (windows)	Supporting (Linux)	Supporting Platform (macOS)	Platform license	Commercial passive	Active	OS fingerprinting	Network fingerprinting	CLI	GUI	Multiuser
MSFCONSOLE	✓			✓	✓	✓	✓		✓	✓		✓		✓
DNSEUNM	✓				✓	✓		✓			✓	✓		
DNMAP	✓				✓	✓		✓			✓	✓		
DNS RECON	✓				✓			✓			✓	✓		
DNSWALK	✓				✓			✓			✓	✓		
DNS TRACER	✓				✓	✓		✓	✓		✓	✓		✓
NMAP	✓		✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	
SHODAN	✓	✓	✓	✓	✓	✓	✓		✓	✓		✓	✓	✓
BURP SUITE	✓	✓	✓	✓		✓	✓	✓	✓	✓			✓	✓
METASPLOIT	✓	✓		✓	✓	✓	✓	✓	✓	✓		✓	✓	✓
WIRESHARK	✓		✓	✓	✓	✓	✓	✓		✓		✓	✓	
SQLMAP	✓			✓	✓		✓		✓	✓		✓		✓
DMITRY	✓		✓	✓	✓	✓	✓	✓		✓		✓		✓
HPING 3	✓			✓	✓	✓	✓		✓	✓		✓		
NBSTAT				✓					✓	✓	✓	✓		
NBTSCAN			✓	✓	✓	✓			✓	✓	✓	✓		
NIKTO	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
GHOST PHISHER	✓				✓			✓	✓		✓	✓	✓	
THEHARVESTER	✓				✓			✓	✓		✓	✓		
FARADAY	✓			✓	✓	✓		✓	✓	✓	✓	✓	✓	
MASSSCAN	✓			✓	✓	✓		✓	✓	✓	✓	✓		

Dnseunm

It is a command-line tool that is used in the information gathering stage to help pen testers gather DNS information about the target. It helps to locate all DNS servers and DNS entries for an organization. A pen tester or hacker can also use it to perform Google scraping. Google scraping is the process of sending queries to Goggle to discover all the domain names linked to the target domain. The Supporting base is Kali Linux. It is used for passive information gathering.

Dnmap

It is a framework that uses a client/server architecture to distribute Nmap scans among several clients. The output from the Nmap scan is stored on both the server and the client. Nmap or Network Mapper is an open-source Linux command-line tool that helps a pen tester in the information gathering stage to discover hosts and services and detect vulnerabilities on a network. It does this by sending packets and analyzing the response gotten. OS fingerprinting is done with it.

DNS Recon

It is written in Python Language and is used when conducting DNS enumeration. It provides the pen tester the ability to perform: Google scanning for subdomains and host, reverse lookup against an IP range, general DNS query, amongst others. It runs on the Linux Operating system.

Dnswalk

It is a tool that helps to check the target database for internal consistency and accuracy. It is a DNS debugger. It can be used to initiate zone transfer, that is, copying contents of the zone file on a primary DNS server (a copy of part of its database) to a secondary DNS server (zone transfer). It runs on Linux operating system.

DNS Tracer

It is a tool used by pen testers in information gathering to extract unique DNS information about a domain. DNS records extracted are NS (Name server records), MX (Mail exchanger records), etc., amongst others. It determines where a given DNS gets its information from a given hostname following the chain of DNS servers back to the DNS server hosting the primary copy of the DNS record that responded to your lookup.

Nmap

Network Mapper (Nmap) is a free and open-source platform that is used to perform initial device or network scanning. Because of its benefits, this tool is often used in the initial step of penetration testing. The most valuable Nmap tools are for gaining insight into a targeted network, including discovering accessible hosts, operating systems and port discovery. It can also be used to search both (Božić *et al*., 2019). It scans individual IP addresses and ranges and returns valuable information such as the operating system, utilities found and available ports (Ankele *et al*., 2019). It is by far the best port scanner in the arena and an excellent component of our host security equipment. We discover IP scope with the aid of the Nmap tool, which we can use in our running system. Nmap aids in the discovery of accessible ports and facilities. It's used to find something (Shah *et al*., 2019).

Shodan

Shodan is a search engine that can be used to locate individual devices and types of devices. Webcams and cisco are the most used searches. The Shodan search engine scans the entire Internet before parsing the banners returned by the scanned machines. When the search is over, the information returned by the Shodan search will most likely be about web servers and their models, as well as anonymous FTP servers if they operate in a specific area and system model information. The use of the Shodan search engine for security analysis around the Internet of Things is expected to grow ever further (Božić *et al*., 2019).

Burp Suite

Burp Suite is a tool for conducting web application security monitoring. It is a tool that helps with the whole testing process, from plotting and analyzing an application's attack surface to discovering and leveraging security flaws (Božić *et al*., 2019). It can be used to search for popular site vulnerabilities automatically, but it also includes specialized manual scanning procedures to help with each step of penetration testing (Ankele *et al*., 2019).

Metasploit

Metasploit is a vulnerability discovery, leveraging and validation tool that includes Metasploit Framework and some commercial equivalents. The Metasploit Framework is an open-source initiative that offers infrastructure, content and resources for penetration testing. Anti-forensics and specialized avoidance tools are also available, with some of them being integrated into the Metasploit Framework (Božić *et al*., 2019).

Wireshark

Wireshark is a network packet inspection tool (also known as a network sniffer) that captures and displays packets in a human-readable format in real-time. Wireshark is a passive network traffic analyzer that does not transmit data. This ensures that if Wireshark is used on a network, other parties would not be able to spot it. This tool is open source and works for UNIX, Windows

and various other operating systems. It has a graphical user interface, which distinguishes it from other packet analyzers like tcpdump, which shares several features with Wireshark (Božić *et al*., 2019).

SQLmap

SQLmap is a free and open-source penetration testing platform that automates the task of manipulating SQL databases. It can also be used to determine the database and version being used. As a result, the tool can be used both during reconnaissance and during the gaining access process (Ankele *et al*., 2019).

Dmitry

Dmitry is a data collection tool that can be used to find out who is who, uptime records, email addresses and subdomains. Additionally, the instrument can be used to conduct port scans (Ankele *et al*., 2019).

Hping3

Hping is a command-line-oriented TCP/IP packet assembler/analyzer. The interface is inspired by the ping (8) UNIX command, but hyping isn't only able to send ICMP echo requests; it supports TCP, UDP, ICMP and RAW-IP protocols, has a traceroute mode, the ability to send files between a covered channel and many other features. While hping was mainly used as a security tool in the past, it can be used in many ways by people that don't care about security to test networks and hosts (Sanfilippo, 2006).

Nbtstat

Nbtstat is a network utility that is used to verify the status of ongoing TCP/IP connections, according to ethical hacking experts. Nbtstat displays all of the network connections that are active in the Windows operating system. Because this utility is preinstalled in Windows, you won't need any additional program to utilize it. It's a valuable tool for determining all of the Windows workstations' TCP/IP connections (Gill, 2018).

Nbtscan

The NBTScan utility may be used to look for NetBIOS name information in IP addresses. It will generate a report with the connected computers' IP addresses, NetBIOS computer names, services accessible, logged-in usernames and MAC addresses. This data will come in handy throughout the penetration testing process. The difference between nbtstat and Windows' NBTScan is that NBTScan can scan a wide range of IP addresses. You should be warned that utilizing this program generates a lot of traffic, which the target computers may log (Allen *et al*., 2014).

Nikto

Sullo, CIRT, Inc. was the first to write and maintain Nikto. David Lodge is the current maintainer, though other individuals have also contributed to the project. It was included in the Kali Linux Penetration Testing distribution and is designed to work on any platform having a Perl environment. It is an open-source program that supports SSL, proxies, host authentication, IDS evasion and other features. It provides the optional sending of new version data back to the maintainers and may be updated automatically from the command line (Obbayi, 2018).

Ghost Phisher

Ghost Phisher is a wireless and Ethernet security auditing and attacks software application built-in Python with the Python Qt GUI framework. It can impersonate access points and distribute malicious code (Savio-Code, 2017).

The Harvester

This program's goal is to collect emails, subdomains, hosts, employee names, open ports and banners from a variety of public sources, such as search engines, PGP key servers and the SHODAN computer database. This application is designed to assist penetration testers in understanding the client's footprint on the Internet during the early phases of a penetration test. It's also valuable for anyone curious about what an attacker sees about their company (Martorella, 2021).

Faraday

IPE (Integrated Penetration-Test Environment), a multiuser Penetration-Testing IDE, is a novel idea introduced by Faraday. Faraday is a program that distributes, indexes and analyzes the information gathered during a security audit. It was created to allow you to use the community's various tools in a genuine multiuser manner (John, 2021).

Masscan

This is the quickest scanner for Internet ports. It can scan the whole Internet in less than six minutes and transmit ten million packets per second. It delivers findings that are comparable to those of Nmap, the most well-known port scanner. Internally, it uses asynchronous communication and works similarly to Scanand, Unicornscan and ZMap. The main distinction is that it is far faster than these other scanners. Furthermore, it is more adaptable, allowing for any address and port ranges (Graham, 2021).

RQ2: Reconnaissance Techniques

In this section, we present and discuss the existing reconnaissance techniques explored from the 19 selected studies. Some of the existing reconnaissance techniques include.

Phishing

This is the most common way to get information through the use of email, phone calls and SMS. In phishing, the hackers observe the target by learning what kind of mails the target is interested in, SMS of interest, etc. The hacker then sends malicious emails to the target with familiar features such as passwords, bank, etc., on what the target is used to and when it is clicked on, it gives the hacker access to the target's personal information (Lohani, 2019).

Baiting

This is a process of physically attacking an individual or an organization through the use of various mediums such as an infected USB drive so that when that drive is inserted into the organization or individual's devices, it tends to affect the device (Yasin *et al*., 2019).

Dumpster Diving

This is another technique used for getting information from a target. Dumpster diving involves the monitoring of the waste bin or dust bin of an organization by an attacker and collecting information such as an important document that may have been disposed of carelessly by staff or workers in the organization (Yasin *et al*., 2019).

Spear Phishing

This is also a technique used to obtain private information, just like phishing; the difference is that Spear phishing focuses on specific persons and sends emails based on the information gathered of that person. Once the emails are clicked on, it gains access to private information. The success rate in spear-phishing attacks is higher than the phishing attack (Abass, 2018).

Pretexting

This is a process of acting or pretending to be someone you are not or using a persuasive approach to get information from an organization or individual, for example, acting as an investigator to obtain confidential company records (Abass, 2018).

Tailgating

This process involves an attacker gaining access into an organization simply by following an employee or an authorized person working in the organization (Yasin *et al*., 2019).

Watering Hole Attack

The watering holes attack involves tracking all websites the target has visited and how frequently the target visits those websites. The hacker then gets the website that is most visited by the target. The hacker then finds a loophole in those websites, such as reflected XSS, host header, etc. and obtains private and confidential information from their target (Lohani, 2019).

Discussion

Upon exploring the 19 selected relevant studies, the authors present a summary of the similarities and differences with respect to the features associated with them as shown in Fig. 6. The reviewed tools were associated with one or more features that could help cyber security experts determine their suitability in a given IT environment. Such instance includes the commercial license, we can see that even though 19 of the 21 tools are open sources, at least 10 of them require a commercial license.

These tools were investigated and benchmarked against the features which could help cybersecurity experts and practitioners briefly identify their usage scenarios, environment and requirements associated with these tools. This investigation will be beneficial to cybersecurity research practitioners to enable them to test and determine which information-gathering tools apply to a given scenario and if the tool meets their deployment requirement. Additionally, these features were also identified as a means of identifying these tools based on their capabilities.

Table 2 illustrates the summary of our findings which answers RQ3. The features of these tools can be used to evaluate them. Before considering and deploying tools to be used for information gathering, experts often review and gather information on them. Features like open-source promote trust and dependability as it offers readability into the coding structure and development history of these tools and a community that supports them. The Supported Platform is also an essential feature that experts need to consider; this will help them determine if their existing infrastructure is well equipped for the usage of the tools. Most of these tools are available for free to developers and professionals for free, with basic features available. However, experts need to be fully informed if they might need to pay for the tools in an enterprise environment.

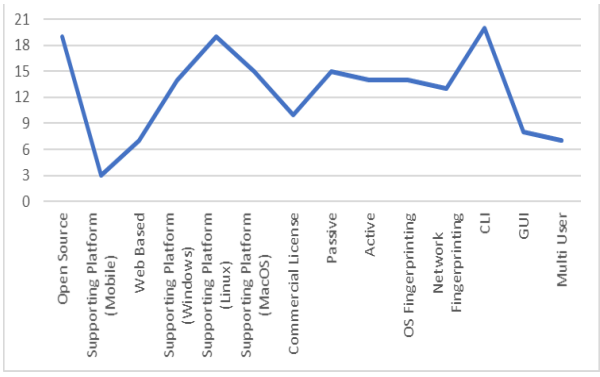

Fig. 6: Summary of the tools reviewed

Other important metrics used to evaluate these tools include the mode of accessing the tools in terms of the user interface, which can be graphically or in a command-line fashion; the methodology of these tools, i.e., if it actively tries to gather information about the target or it passive generated this information; Most of these tools often collect information about a target host or a target network, which can be categorized as OS fingerprinting or Network fingerprinting; Experts may also want to know if their InfoSec team can collaborate within this tools, which defines the multiuser feature.

Platforms Supported: Operating Systems:

From the result of the table summarized above, we can see that the Linux operating system is richly favored when it comes to the availability of information-gathering tools. It recorded 95.2% support for all the tools evaluated, which means they are all available to be installed on a Linux machine. Some of these tools even come preinstalled on Kali Linux (formerly known as Backtrack), Parrot Security OS, Backbox and other Linux distributions that were built for hacking and penetration testing.

As shown in Fig. 7, macOS recorded more support for these tools than Windows, probably because it is built on a UNIX kernel at its core. But on the PC platform, Windows recorded the lowest number of supports from these tools. This implies that Windows might not be the choice OS for ethical hackers and cybersecurity professionals.

IOS/Android Support

Even with the advancement in the hardware and software of mobile operating system in recent times, mobile operating systems like Android and iOS has not been given much consideration as a penetration testing tool. This is evident from this research as shown in Fig. 8, as only about 19% of these tools are supported on Android/iOS. This low-rate support for mobile can be attributed to the fact that a lot of these tools need a shell or command-line interface to run, which is not readily available on the OSes.

This means they are not optimized and convenient for most of the information-gathering tools. Until the developers of these OS consider creating a built-in command-line interface as part of the user's space, the mobile platform may not be considered as an essential device for information gathering. Alternatively, although it may not be convenient, experts may decide to use 3rd party terminal emulators like Termux, Terminal Emulator and Terms, if they need to perform penetration testing using these platforms.

Web-Based Support

Although within the last few decades, the IT world has evolved from the deployment of stand-alone applications into deploying web-based/cloud-based applications, unfortunately, this has not been evident in the cybersecurity sphere, based on the result of this analysis as shown in Fig. 9 and 10. The result of this research shows that out of the 21 tools reviewed, only 33% are web-based.

This could be attributed to the fact that most of the information-gathering tools often need to be deployed within the target infrastructure for active testing. However, because of the emerging trend of web applications, we can expect more information-gathering tools to become web-based in the coming years.

Open-Source

IT Practitioners often associate open-source software with flexibility, community support, security and reliability. When it comes to choosing information-gathering tools for either testing or live deployment, one of the significant determinants for experts is whether a tool is Open-Source or not.

From the analysis of these tools, we discovered that about 90.4% of these tools are Open-Source, while 9.6% were not proprietary tools. This implies that the majority of these tools have their source code hosted on public communities like GitHub, where collaboration and contributions can be made. Organizations can take advantage of the licenses that come with these Open-Source tools (e.g., BSD, GNU, CDDL, etc.) and decide if they want to modify these tools to suit their needs and environment.

Commercial/Enterprise Licensing

Another consideration that is important to cybersecurity experts and enterprises looking to deploy these tools in their organization is the financial cost implication of selecting any of the tools. They need to know if a license fee is required from enterprises, whether the tool needs some skillset and if they will need to hire a professional. They also need to know how much it will cost to reduce the surface area for the information gathered by these tools because the amount of information a threat actor can gather determines the success rate of the attack.

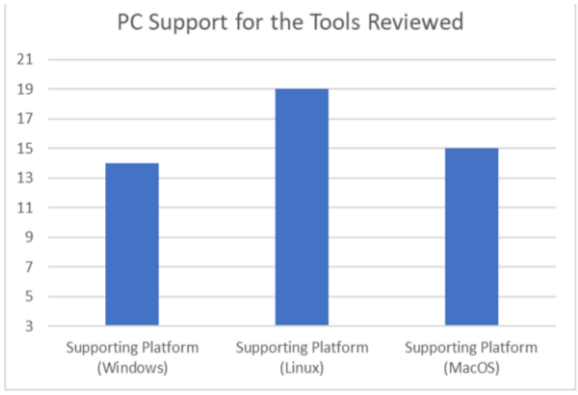

Fig. 7: PC supported for the tools reviewed

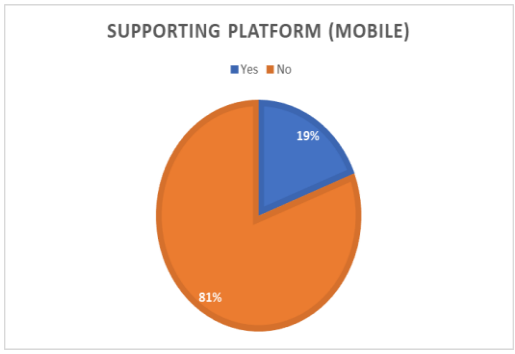

Fig. 8: Mobile supported for the tools reviewed

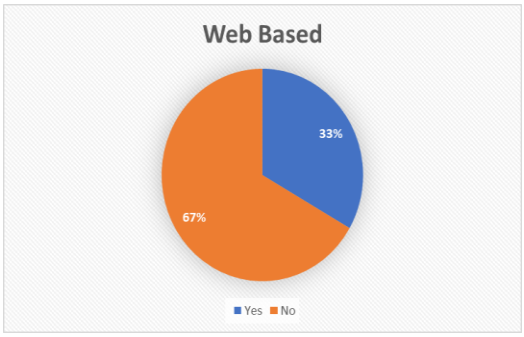

Fig. 9: Tools reviewed that are web-based

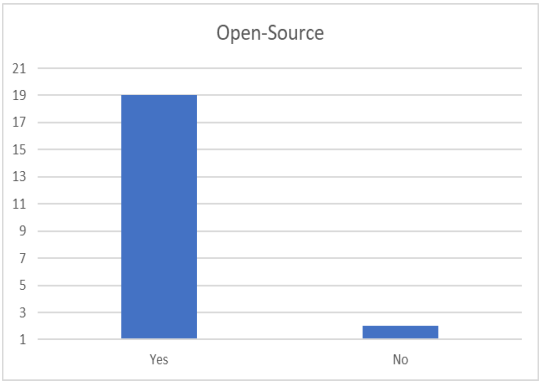

Fig. 10. Tools reviewed that are open-source

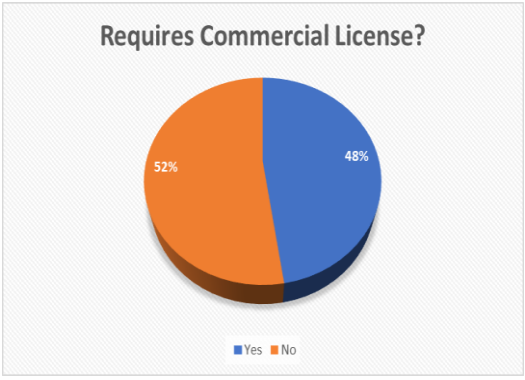

Fig.11: Tools reviewed that requires a commercial licensing

Figure 11 shows the result of the analysis on the tools reviewed indicates that about 48%, almost half of the tools reviewed, have, or require licensing for enterprise deployment. This would mean that they may need to adjust their IT budget or investigate alternative tools that may be "license-free."

Command-Line and Graphical User Interface

A software that has a Graphical User Interface (GUI) is typically more appealing to users who are not programmers or scripters, especially when they come from a Windows background and seldom access its built-in command-line tools (PowerShell and Command Prompt) because unlike the command line interface, it allows them to operate the application using a mouse or keyboard to navigate through windows, menus and clickable icons. However, this is not usually a preference for ethical hackers with solid scripting skills. This can be attributed to the fact that navigating through the GUI is typically slower than the CLI, however, CLI gives more options for customization and advanced interfacing. This may be one of the reasons why the majority of the information-gathering tools reviewed had a command-line interface.

The expectation that ethical hackers should be able to write codes or scripting has now become an industry standard. This is visible in the result of this review as shown in Fig. 12 and 13, where only 8 of the 21 tools, accounting for 38% of the tools evaluated, provide a graphical interface for users to interact with, while 20 out of 21 (95.2%) of the tools provide a command-line interaction for these applications. Even 4.8% of the tools that do not offer a command-line interface allowed users to launch it from the command line interface while specifying some parameters to customize how it runs.

Also, 95.2% of the tools reviewed allowed experts to gather information by running their necessary commands from the command line. Although this option requires a level of coding skills and knowledge of the command for the application, it offers greater flexibility, faster management, greater control and an automation option. This means that ethical hackers can do faster within the designed scope of the application.

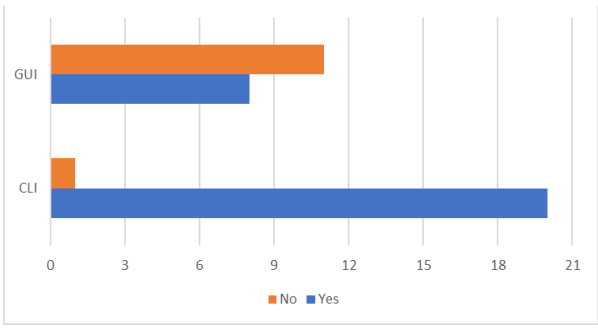

Fig. 12: User interface tools reviewed

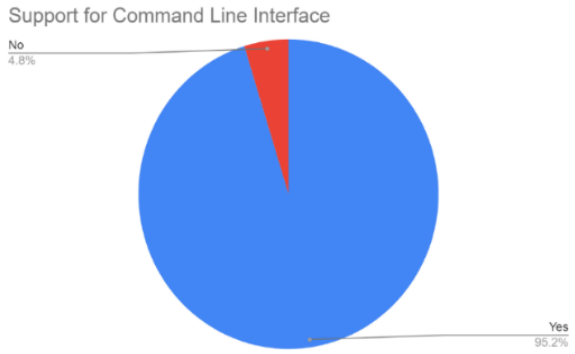

Fig. 13: Tools reviewed that support graphical user interface

Network-Based and System-Based Fingerprinting

Information gathering tools can be classified into various categories based on their usage scope. In this study, we evaluated and classified these tools into two information-gathering categories, which are OS fingerprinting and Network fingerprinting. This means the tools can either be used to get information about the target system's hardware and operating system or gather data about the target network.

Out of the 21 tools that were reviewed, 66.7% of them (14 tools), could be used for gathering and extracting information actively relating to the target system, as shown in Fig. 14 and 15. Examples of such information include machine names, operating systems, usernames, network resources and services. 61.9% (13 tools) of these tools are network-based as they can be used to gather about the target's network infrastructure such as protocols, ports, DNS, IP address, hosts and the general network architecture. At the same time, 28.5% could be classified as hybrid as they allow the attacker to gather system-based and network-based information.

Active and Passive Information Gathering

Passive information gathering refers to gathering as much information as possible without establishing contact between the pen tester (yourself) and the target about which you are collecting information. Active information gathering involves contact between the pen tester and the actual target. Some tools used in information gathering are dynamic as they can be used in an Active or Passive way, depending on the goal of the attacker and the resources available. From the study, we discovered that while we could classify the tools into either active or passive information-gathering tools, some of the tools could function as both depending on how the attacker uses them. Figure 16 and 17 shows that 15 tools passively gather information, 14 tools actively gather information, while 6 tools could be used as both an active and passive approach to information gathering.

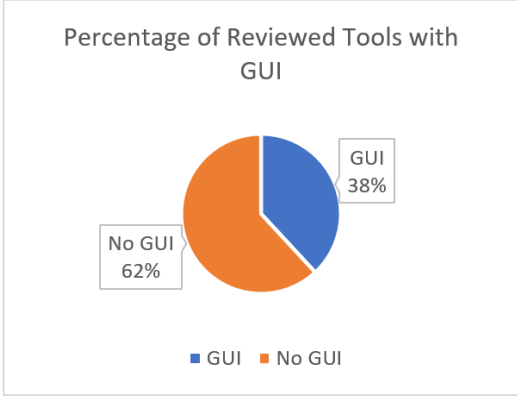

Fig. 14: Usage scope tools reviewed

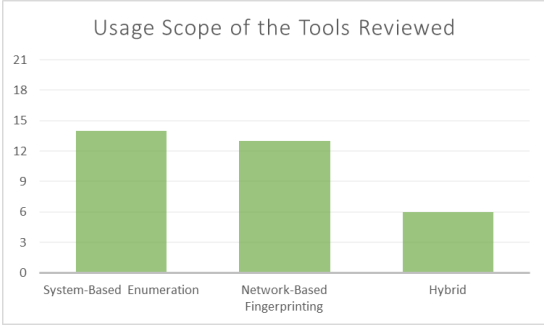

Fig. 15: Usage scope tools reviewed

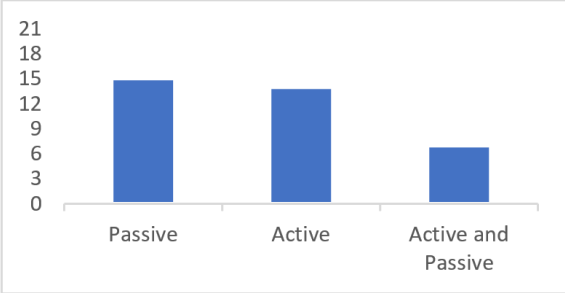

Fig. 16: Passive and Active tool count

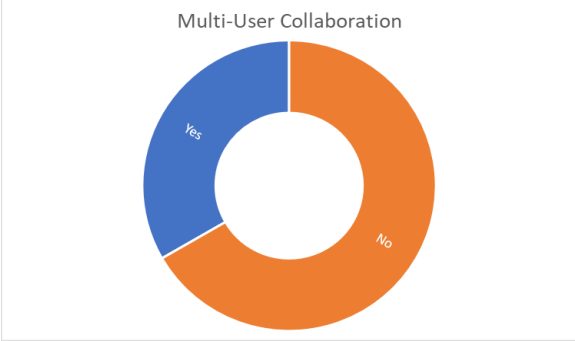

Fig. 17: Tools reviewed that support multi-user collaboration

Multiuser Environment

An information gathering tool with multiuser-support provides cybersecurity experts with the ability to collaborate on an engagement or reconnaissance with other team members. The team can log into the management console or interface to perform tasks, review data and share projects. It also provides enterprises with a dedicated team of cybersecurity experts with the option to implement role-based access control. Each team member will only be able to use the tools or files necessary for their job description.

From the analysis done, most of the tools do not have a comprehensive option when it comes to multiuser and collaboration. Although, out of the 21 tools reviewed, 33.3% had features like basic communication features like chatting that classified that as a multiuser information gathering tool, other features like task management, scheduling and other multiuser features that could better enhance collaboration was missing. One of the recent trends in IT is the implantation of Role-Based Access Control (RBAC) to ensure that in organizations, IT staffs only get just the privilege or permission they need to do their work and nothing more. This means that information technology has evolved from traditional roles like system administrator and enterprise administrator into job-specific roles like Security Administrator, Compliance Administrator, User Access Administrator, Authentication Administrator, Security Operator and Password Administrator. Unfortunately, this trend is not yet visible in the information gathering tools that were reviewed, as they still function as though only one person should perform reconnaissance and not a team.

The five pillars of information assurance are confidentiality, integrity, availability, non-repudiation and authentication. An attack that violates one or more of these pillars is considered a cyber-attack. Cyber-attacks are in phases, but the first phase of any cyber-attack is reconnaissance. In this phase, the weak points of the target are identified. Critical information like the victim's IP addresses, home address, telephone number, frequent hangout, dark secrets, security policies, etc., are collected and ways of bypassing the victim's defense systems are also noted (Sanghvi and Dahiya, 2013).

Cyber-attacks are growing in terms of complexity and volume. According to Industry Week, in 2018, spear-phishing and spoofing attempts of business emails increased by 70 and 250%, respectively and ransomware campaigns targeting enterprises had an impressive 350% growth. In general, economic damages are relevant, as there is the need to detect and investigate the attack as well as restore the compromised hardware and software. To give an idea of the impact of the problem, the average cost of a data breach has risen from $4.9 million in 2017 to $7.5 million in 2018. To make things worse, attackers can now use a wide range of tools for compromising hosts, network appliances and Internet of Things (IoT) devices simply and effectively, for example, via a Crime-as-a-Service business model (Mazurczyk and Caviglione, 2021).

Conclusion

Reconnaissance attacks may not seem disastrous; however, a successful one could lead to a successful devastating attack. This review identified existing reconnaissance tools and techniques, showing what they are about and what they involve. Also, the methods used to evaluate common reconnaissance tools were highlighted.

A Quantitative Analysis (QA) was conducted on the selected articles. The results, among others, show that 95.2% of the tools allowed experts to gather information by running their necessary command from the command line. While 4.8% of the tools do not provide a command-line interface allowing users to launch it from the command line interface while specifying some parameters to customize how it runs. 61.9% of the tools are network-based as they can be used to gather about the target's network infrastructure such as protocols, ports, DNS, IP address, hosts and the general network architecture. In comparison, 28.5% could be classified as hybrid as they allow the attacker to gather system-based and network-based information.

Having evaluated twenty-one common reconnaissance attack tools, it was observed that although some of the tools operate in a different capacity, the best-fit tool is massively dependent on the attacker or penetration tester and the situations surrounding the scenario. Therefore, a tool should be selected based on the user's preference and the attack style. However, enumerating these tools and how they operate alongside the social engineering techniques in gathering information of victims, enlightens common users, cybersecurity professionals and organizations to the risk of a successful reconnaissance attack and possible ways of avoiding or mitigating the effects of the attack.

Acknowledgement

We appreciate the support and funding of the Covenant University Center for Research Innovation and Development (CUCRID).

Author's Contributions

Isaac Odun-Ayo, Gabriel Oyeyemi and Emmanuel Owoka: Conceived and formulated the research concept, contributed to the writing of the paper and made necessary corrections.

Otavie Okuoyo, Opeyemi Ogunsola, Obaro Ikoh, Olumide Adeosun, Deborah Etukudo and Victoria Robert: Performed analysis, gathered data and

contributed to the writing of the paper.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all other authors have read and approved the manuscript. There is no conflict of interest.

References

Abass, I. A. M. (2018). Social Engineering Threat and Defense: A Literature Survey. Journal of Information Security, 09(04), 257-264. doi.org/10.4236/JIS.2018.94018

Aldawood, H., & Skinner, G. (2019). Contemporary Cyber Security Social Engineering Solutions, Measures, Policies, Tools and Applications: A Critical Appraisal. International Journal of Security, 10(1).

Allen, L., Heriyanto, T., & Ali, S. (2014). Kali Linux - Assuring Security by Penetration Testing. In Network Security (Vol. 2014). Packt Publishing. doi.org/10.1016/s1353-4858(14)70077-7

Ankele, R., Marksteiner, S., Nahrgang, K., & Vallant, H. (2019). Requirements and recommendations for IoT/IIoT models to automate security assurance through threat modelling, security analysis and penetration testing. WISI 2019 - Workshop on Industrial Security and IoT. Canterbury: Association for Computing Machinery. doi.org/10.1145/3339252.3341482

Božić, K., Penevski, N., & Saša, A. (2019). Penetration Testing and Vulnerability Assessment: Introduction, Phases, Tools and Methods. Sinteza 2019 - International Scientific Conference on Information Technology and Data Related Research, 229–234. Singidunum University. doi.org/10.15308/sinteza-2019-229-234

Chantler, N., & Broadhurst, R. (2008). Social Engineering and Crime Prevention In Cyberspace. Proceedings of the Korean Institute of Criminology. Retrieved from http://eprints.qut.edu.au/7526/

Cuppens-Boulahia, N. (2012). Penetration Tester's Open Source Toolkit. Computers & Security, 31(4), 630–632. doi.org/10.1016/j.cose.2012.03.008

Gill, J. (2018, November 28). NETBIOS OVER TCP/IP – NBTSTAT USAGE IN DETAIL. https://www.securitynewspaper.com/2018/11/28/netbios-over-tcp-ip-nbtstat-usage-in-detail/

Graham, R. D. (2021, December 23). MASSCAN: Mass IP port scanner. https://github.com/robertdavidgraham/masscan

John, K. (2021, May 23). Faraday - Penetration Testing IDE & Vulnerability Management Platform. https://computingforgeeks.com/faraday-penetration-test-vulnerability-management-ide/

Kumar, A., Chaudhary, M., & Kumar, N. (2015). Social Engineering Threats and Awareness: A Survey. European Journal of Advances in Engineering and Technology, 2(11), 15-19. www.ejaet.com

Lohani, S. (2019). Social Engineering: Hacking into Humans. International Journal of Advanced Studies of Scientific Research, 4(1), 10.

Martorella, C. (2021, November 25). The Harvester. https://github.com/laramies/theHarvester

Mazurczyk, W., & Caviglione, L. (2021). Cyber reconnaissance techniques. Communications of the ACM, 64(3), 86–95. doi.org/10.1145/3418293

Obbayi, L. (2018, March 30). Introduction to the Nikto web application vulnerability scanner - Infosec Resources. https://resources.infosecinstitute.com/topic/introduction-nikto-web-application-vulnerability-scanner/

Sanfilippo, S. (2006). Hping - Active Network Security Tool. ttp://www.hping.org/

Sanghvi, P. H., & Dahiya, S. M. (2013). Cyber Reconnaissance: An Alarm before Cyber Attack. International Journal of Computer Applications, 63(6), 36–38. https://doi.org/10.5120/10472-5202

Savio-Code. (2017). Ghost-Phisher. https://github.com/savio-code/ghost-phisher

Shah, M., Ahmed, S., Saeed, K., Junaid, M., Khan, H., & Ata-ur-rehman. (2019). Penetration Testing Active Reconnaissance Phase – Optimized Port Scanning With Nmap Tool. 2019 2nd International Conference on Computing, Mathematics and Engineering Technologies (ICoMET), 1–6. IEEE.

Timalsina, U., & Gurung, K. (2017). Use of Metasploit Framework in Kali Linux. doi.org/10.13140/RG.2.2.12377.93284

United Nations. (2015). United Nations Sustainable Development. https://www.un.org/sustainabledevelopment/

Yasin, A., Fatima, R., Liu, L., Yasin, A., & Wang, J. (2019). Contemplating social engineering studies and attack scenarios: A review study. Security and Privacy, 2(4). doi.org/10.1002/spy2.73