Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Volume 13, Issue 03 March 2024

Footprinting Toolkit: A Toolset Streamlined to Perform Multiple Information-gathering Techniques

Reddyvari Venkateswara Reddy, Emerald Sharon, Kolli Sanjana, Vennam Srinivas Reddy
Associate Professor & Head of the Department, Department of CSE (Cyber Security), CMR College of Engineering & Technology, Hyderabad, Telangana, India.
Student, Department of CSE (Cyber Security), CMR College of Engineering & Technology, Hyderabad, Telangana, India.

***Abstract*— In the field of cybersecurity, the reconnaissance stage is essential for finding weaknesses and evaluating a network's overall security posture. The objective is to develop a dependable, user-friendly toolkit to gather crucial data about target systems, services, and potential attack points while adhering to ethical standards. The proposed solution provides functionalities of multiple footprinting and information-gathering tools streamlined into a single toolset, which can be used over a Linux operating system setup. Our footprinting toolset utilizes DNS enumeration, active scanning, and passive information gathering to provide a comprehensive digital footprint picture for users. The toolkit enables security professionals to assess and evaluate potential risks, identify vulnerabilities, and enhance the security posture of target organizations effectively, in a user-friendly setup as required.**

***Keywords:* Cybersecurity, reconnaissance, network, footprinting, digital footprint, DNS enumeration, active scanning, passive information gathering, security posture, Linux.**

INTRODUCTION

In today's world, with the growing technology, the number of cyber-attacks has also been increasing and advancing. Hence, an organization or a business must safeguard its data and strategies. To ensure this the business or organization must invest in creating its security architecture and posture, which includes thorough testing and analysis of the business architecture, and products. A penetration test is one of many cybersecurity procedures where authorized personnel conduct evaluations by simulating a real attack and examining the architecture and the security measures that have been implemented. The phase of reconnaissance, also known as information collecting, is the first and most important step in the method of implementing this approach, which is extensively used for detecting vulnerabilities.

The phase of reconnaissance aims to collect comprehensive intelligence on the system, network, or organization being targeted. System configurations, IP addresses, domain names, and network infrastructure are among the items identified in this process. Through gathering information about a potential target, reconnaissance can help in revealing possible attack vectors, and vulnerabilities within the system or network. These may include misconfigured systems, and outdated software among others that can be exploited. Reconnaissance is central to defining the attack surface – a term used to describe all ways possible that an attacker could use to gain unauthorized access to an organization's system or network, with additional details of physical locations and technology stack, information about the target organization's infrastructure including its network architecture may also be provided through reconnaissance. This kind of knowledge gained from understanding can necessarily be used for planning subsequent phases of penetration tests or cyberattacks effectively.

Reconnaissance is an essential part of cybersecurity as it facilitates the detection of vulnerabilities in networks and systems, and footprinting is a part of reconnaissance. As a component of reconnaissance, footprinting is vital for gathering intel. It enables cybersecurity professionals to gather information in two ways, particularly: active scanning, and passive scanning. Passive scanning is the method in which target intel is collected without directly interacting with it, whereas, in active scanning the information is gathered directly by interacting with the target organization or network.

The gathered information reveals the online presence and susceptible areas of an organization to intrusions. Organizations can strengthen their security and mitigate potential threats by acquiring knowledge of their attack surface. Cybersecurity experts can get useful data from the tool, like website addresses, names, email addresses, employee details, previous organization structures, domain names, IP addresses, and metadata. This data reveals the digital footprint of the target domain and shows the vulnerabilities that may lead to potential data breach attacks. Organizations can fortify their security measures and mitigate potential threats by developing an in-depth awareness of their attack surface.

The objective of implementing the project is a streamlined toolkit that comprises functionalities of multiple footprinting tools in one. With this toolkit being capable of performing multiple types of information-gathering techniques, enumeration, and harvesting methods, we aim to establish a robust, user-friendly, and effective toolset that enables users to adhere to a comprehensive digital footprint of the target.

I. LITERATURE REVIEW

Reconnaissance is an essential part of penetration testing, ethical hacking, and attack processes. This review of literature explores works, shedding light on progress and contributions to reconnaissance, footprinting, and various tools used for the information-gathering process.

In their research paper, "Penetration Testing – Reconnaissance with NMAP Tool" published in IJRAC, Volume 8, No. 3, March-April 2017, Gurline Kaur, and Navjot Kaur provide insights into the operation of the NMAP tool for the reconnaissance process. They discuss how the NMAP tool works to find IP addresses and information about available ports and services of the target system [1].

Sheetal Temara suggests the integration of ChatGPT (an AI Language model) into the reconnaissance phase of penetration testing in her 2024 research paper, "Maximizing Penetration Testing Success with Effective Reconnaissance Techniques using ChatGPT" in the Asian Journal of Research in Computer Science. She emphasizes the potential improvements like enhancing the efficiency and depth of the information-gathering process during critical security assessments, that can be made to traditional approaches, by integrating ChatGPT into the reconnaissance phase [2].

Bandar Abdulrhman Bin Arfaj, Shailendra Mishra, and Mohammed Alshehri, in their May 2021 Article publication, " Efficacy of Unconventional Testing Practices", in Tech Science Press, explain conventional penetration testing method in detail and then explain the slight variance between the conventional and unconventional penetration testing methods, further stating that usage of multiple tools gives better results, regardless of the tools being autonomous [3].

Manikanta Singirikonda, in his June 2023 Article, "Penetration Testing Tool Guide", in the Journal of Cybersecurity, briefly describes the stepwise process of penetration testing and further provides insights into essential cybersecurity tools with their example usages and exploitation commands [4].

Fouz Barman, Nora Alkaabi, Hamda Almenhali, Mahra Alshedi, and Richard Ikuesan, in their ECCWS 2023 publication, "A Methodical Framework for Conducting Reconnaissance and Enumeration in Ethical Hacking Life Cycle", have proposed a framework, a process model managing the reconnaissance and enumeration phases in an agnostic way. They have also provided a technical guide to NMAP and NETCAT [5].

II. OBJECTIVE

The primary objective of the Footprinting Toolkit project is to integrate the functionalities and features of multiple reconnaissance and enumeration tools into one, to offer a better user experience, reduce time consumption, and provide a better digital footprint.

The core objective of the proposed solution is to offer security experts a powerful solution for collecting information, finding possible vulnerabilities, profiling targets, measuring the attack surface, and gathering intelligence in a simplified and effective manner.

The project will involve the implementation of features of various known and autonomous tools used for active and passive footprinting methods. By doing so, we intend to improve the effectuality of the security assessments and simplify operations related to reconnaissance and enumeration. Additionally, the functionality of the toolkit is achieved by using the capabilities of several footprinting techniques, such as email harvesting and domain profiling, metadata extraction from documents, and subdomain enumeration. The purpose of this system is to aid in the detection of possible attack vectors and security threats by collecting, analyzing, and displaying the information that has been acquired from a variety of sources.

Furthermore, the project aims to enhance the organization's vulnerability identification and analysis, and attack surface management capabilities by integrating features of useful tools into a single toolset that is compatible with the Linux operating system. This precautionary approach will aid us in swiftly detecting potential security loops, vulnerabilities, and attack vectors of the target.

Ultimately, the Footprinting Toolkit Project is geared towards establishing a resilient, user-friendly, and efficient information-gathering package that leads to better target profiling, data harvesting and enumeration, attack surface discovery, and vulnerability assessment. The effective execution of this initiative will make a substantial contribution to the organization's overall cybersecurity strategy by enhancing phases of penetration testing.

III. SYSTEM REQUIREMENTS

1. Hardware: A virtual machine (VM) with sufficient resources such as CPU, RAM, and storage to perform enumeration, initiate connections, and handle the network traffic and operations.
2. Software and Operating System: The toolkit is compatible with the Linux operating system – Kali Linux, and Python.
3. Internet Connection: A stable and reliable internet connection is necessary for the toolkit to perform active and passive scanning or information-gathering techniques, and to search across sources effectively.

IV. PROBLEM DEFINITION

Given the circumstances today, to perform reconnaissance and enumeration, multiple tools must be used, to gather different types of information. The usage of multiple tools usually disrupts the workflow of an individual user or a security professional. It is also a time-consuming process, as using various tools, one at a time is a real deal, and consumes a lot of time. Another problem is that, to use multiple tools one must learn the usage and functionalities of each tool, and gain hands-on experience in using each one of them, which indicates that it has a steep learning curve. Having a single tool, or even a single toolkit that can perform all the activities performed by the various tools, can solve most of the problem. The availability of this kind of solution is non-existent, hence leading to the idea of our proposed solution.

V. EXISTING SOLUTIONS

1. theHarvester**:** It is a reconnaissance tool that was developed to acquire information. It can extract data from public sources such as search engines, PGP key servers, and SHODAN, which subsequently provide information such as email addresses, subdomains, and related information.

Fig-1: theHarvester

2. Sublist3r**:** It is a Python-based application that utilizes multiple resources & search engines to enumerate subdomains of a target domain. It helps find possible starting points where additional investigation can be done. Subdomain enumeration and support for multiple search engines are its essential features.

Fig-2: Sublist3r

3. Nmap (Network Mapper): Nmap is a popular open-source program for security audits and network discovery. It facilitates the identification of hosts, open ports, and operating services on a network, thereby furnishing crucial data for footprinting purposes.

Fig-3: NMAP

4. Metagoofil**:** Metagoofil is a tool that extracts metadata from publicly available documents. It is used to get details from online documents regarding a target's infrastructure, including usernames, software versions, and server information.

Fig-4: Metagoofil

5. Whois**:** The Whois command-line tool gives you information about registered domains, such as the owner of the domain, the administrator, and the name servers. It is a fundamental tool for footprinting and domain reconnaissance.

Fig-5: Whois lookup

VI. LIMITATIONS OF THE EXISTING SYSTEM

1. Fragmented Workflow: Working with several separate tools can cause a broken process, which means users have to switch between different command-line settings or user interfaces. This can make the entire procedure less united and effective.

2. Learning Curve: Users need to learn how to use each tool properly by understanding its command-line choices, output forms, and working methods. This may take plenty amount of time, creating a steep learning curve for new users.

3. Inconsistent Output Forms: It might be difficult to reliably compile and evaluate data since various tools may provide outputs in different forms. Multiple output formats may require users to adjust, which could make data handling more difficult.

4. Redundant Work: Attempts may be made in redundant ways when certain activities are completed using different instruments, and repeating the same task is a waste of time. A combined toolkit could cut down on unnecessary work and speed up the information gathering process.

5. Maintenance and Updation: It can be hard to keep track of upkeep and changes for many tools. Users have to keep track of and update each tool separately, which makes it more likely that they will use old versions.

VII. WORK FLOW

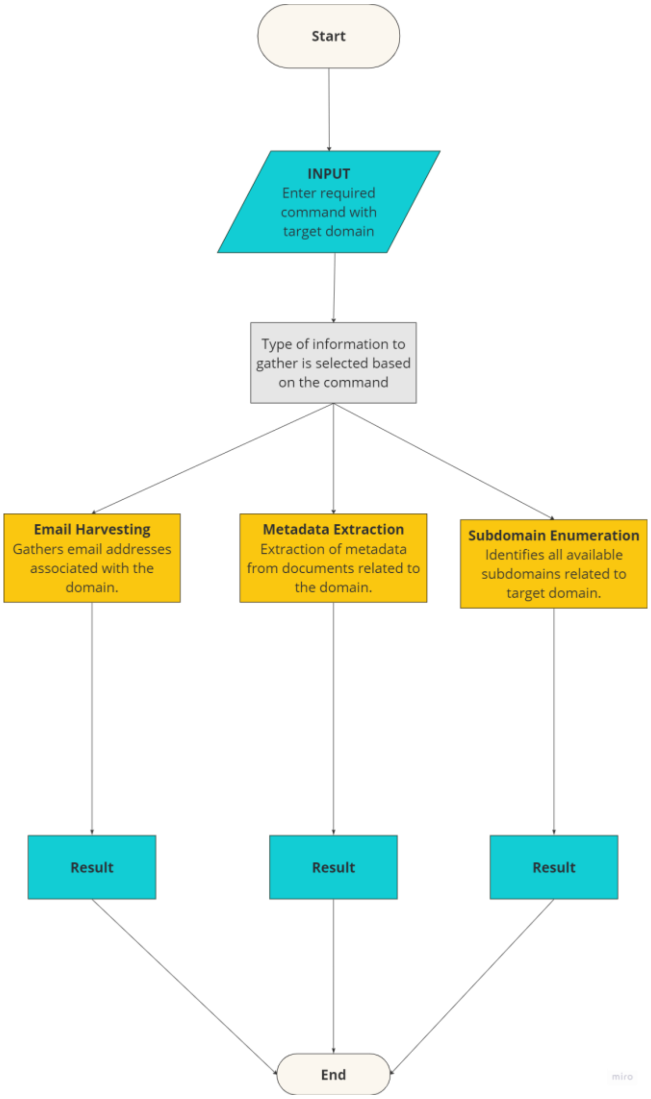

Start
INPUT
Enter required command with target domain
Type of information to gather is selected based on the command
Email Harvesting
Gathers email addresses associated with the domain.
Metadata Extraction
Extraction of metadata from documents related to the domain.
Subdomain Enumeration
Identifies all available subdomains related to target domain.
Result
Result
Result
End
miro

Fig-6: Work Flow

The workflow of the toolkit is represented in the *fig-6,* the explanation is as follows:

1. Start: This is the starting point of the workflow.
2. Input: The user must enter the required command.
3. Method Selection: As per the command specification, the required type of reconnaissance or information-gathering method is performed.
   a. Email Harvesting: Gathers email addresses associated with the domain.
   b. Metadata Extraction: Extraction of metadata from documents associated with the domain.
   c. Subdomain Enumeration: Identifies all available subdomains of the target domain.

4. RESULTS:

a) Displays the collected email addresses related to the target domain.
b) The extracted metadata is stored in the specified file as a result.
c) Provides subdomains, requests, and open ports as a result.

5. End of Process: This is the endpoint of the workflow.

VIII. ARCHITECTURE

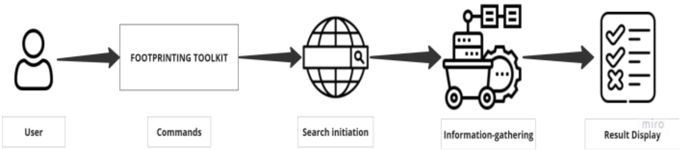


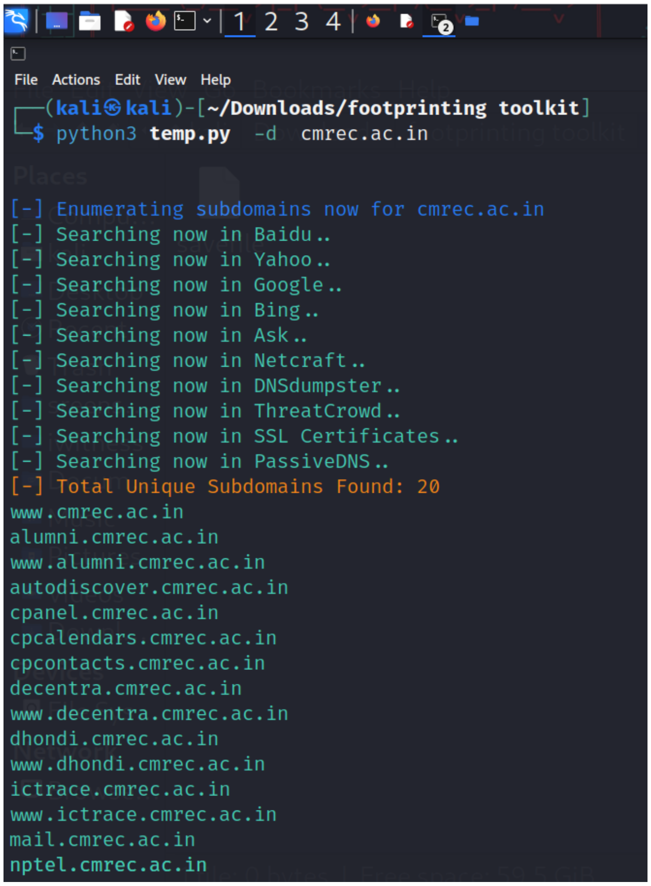
Fig-7: Architecture

1. The user has to install and set up the footprinting toolkit.
2. After set up, the user can perform various footprinting techniques like active and passive scanning. The type is selected per the command entered by the user.
3. Once the procedure is selected, the search is initiated. Passive scanning includes searching across various search engines, available databases, PGP key logs, and many other resources.
4. Associated information of the target domain is collected.
5. The gathered information is then displayed to the user.

IX. CONCLUSION

In conclusion, our project introduces a footprinting toolkit which is a streamlined toolset, with features and functionalities of multiple footprinting tools, integrated into one. It solves the problem of disrupted workflow, by providing multiple features in a single package, further enhancing the workflow of the user. It is user-friendly, time-saving, efficient, robust, and has a small learning curve, making it the best solution to all the stated problems. Our toolkit enables security professionals, or any of its users, to adhere to a comprehensive digital footprint of the target.

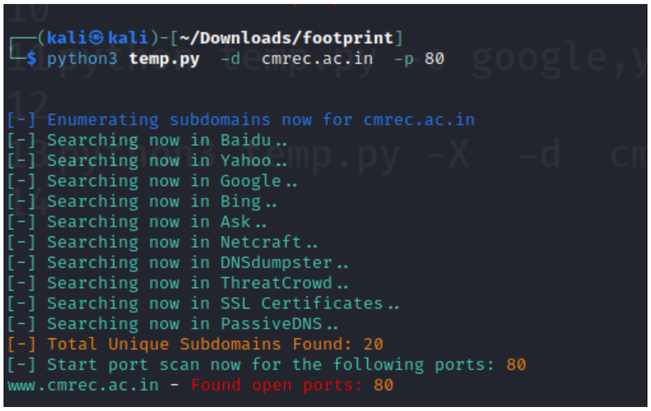
X. RESULTS



```
┌──(kali㉿kali)-[~/Downloads/footprinting toolkit]
└─$ python3 temp.py  -d  cmrec.ac.in

[-] Enumerating subdomains now for cmrec.ac.in
[-] Searching now in Baidu..
[-] Searching now in Yahoo..
[-] Searching now in Google..
[-] Searching now in Bing..
[-] Searching now in Ask..
[-] Searching now in Netcraft..
[-] Searching now in DNSdumpster..
[-] Searching now in ThreatCrowd..
[-] Searching now in SSL Certificates..
[-] Searching now in PassiveDNS..
[-] Total Unique Subdomains Found: 20
www.cmrec.ac.in
alumni.cmrec.ac.in
www.alumni.cmrec.ac.in
autodiscover.cmrec.ac.in
cpanel.cmrec.ac.in
cpcalendars.cmrec.ac.in
cpcontacts.cmrec.ac.in
decentra.cmrec.ac.in
www.decentra.cmrec.ac.in
dhondi.cmrec.ac.in
www.dhondi.cmrec.ac.in
ictrace.cmrec.ac.in
www.ictrace.cmrec.ac.in
mail.cmrec.ac.in
nptel.cmrec.ac.in
```

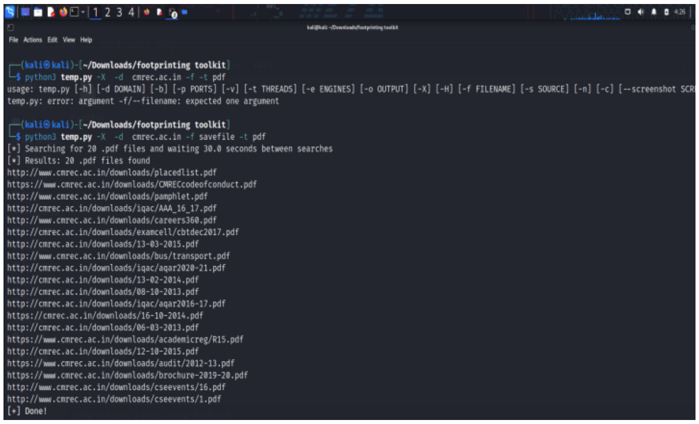
Fig-8: Subdomain Enumeration



```
┌──(kali㉿kali)-[~/Downloads/footprint]
└─$ python3 temp.py  -d  cmrec.ac.in  -p 80

[-] Enumerating subdomains now for cmrec.ac.in
[-] Searching now in Baidu..
[-] Searching now in Yahoo..
[-] Searching now in Google..
[-] Searching now in Bing..
[-] Searching now in Ask..
[-] Searching now in Netcraft..
[-] Searching now in DNSdumpster..
[-] Searching now in ThreatCrowd..
[-] Searching now in SSL Certificates..
[-] Searching now in PassiveDNS..
[-] Total Unique Subdomains Found: 20
[-] Start port scan now for the following ports: 80
www.cmrec.ac.in - Found open ports: 80
```

Fig-9: Subdomain Enumeration

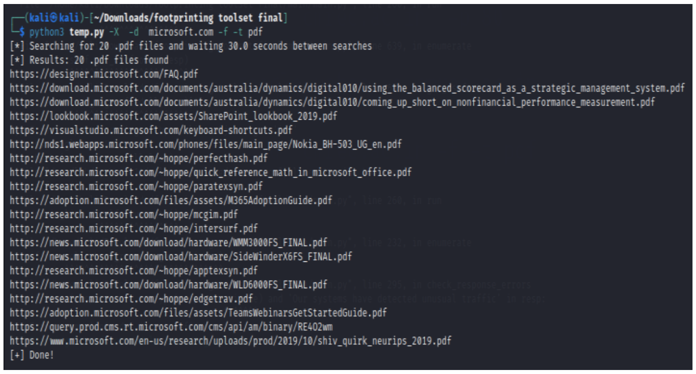


```
┌──(kali㉿kali)-[~/Downloads/footprinting toolkit]
└─$ python3 temp.py -X  -d  cmrec.ac.in -f -t pdf
usage: temp.py [-h] [-d DOMAIN] [-b] [-p PORTS] [-v] [-t THREADS] [-e ENGINES] [-o OUTPUT] [-X] [-H] [-f FILENAME] [-s SOURCE] [-n] [-c] [--screenshot SCR
temp.py: error: argument -f/--filename: expected one argument

┌──(kali㉿kali)-[~/Downloads/footprinting toolkit]
└─$ python3 temp.py -X  -d  cmrec.ac.in -f savefile -t pdf
[*] Searching for 20 .pdf files and waiting 30.0 seconds between searches
[*] Results: 20 .pdf files found
http://www.cmrec.ac.in/downloads/placedlist.pdf
https://www.cmrec.ac.in/downloads/CMRECcodeofconduct.pdf
http://www.cmrec.ac.in/downloads/pamphlet.pdf
http://cmrec.ac.in/downloads/iqac/AAA_16_17.pdf
http://www.cmrec.ac.in/downloads/careers360.pdf
http://cmrec.ac.in/downloads/examcell/cbtdec2017.pdf
http://cmrec.ac.in/downloads/13-03-2015.pdf
http://www.cmrec.ac.in/downloads/bus/transport.pdf
http://cmrec.ac.in/downloads/iqac/aqar2020-21.pdf
http://cmrec.ac.in/downloads/13-02-2014.pdf
http://cmrec.ac.in/downloads/08-10-2013.pdf
http://cmrec.ac.in/downloads/iqac/aqar2016-17.pdf
https://cmrec.ac.in/downloads/16-10-2014.pdf
http://cmrec.ac.in/downloads/06-03-2013.pdf
https://www.cmrec.ac.in/downloads/academicreg/R15.pdf
http://cmrec.ac.in/downloads/12-10-2015.pdf
https://www.cmrec.ac.in/downloads/audit/2012-13.pdf
https://www.cmrec.ac.in/downloads/brochure-2019-20.pdf
http://www.cmrec.ac.in/downloads/cseevents/16.pdf
http://www.cmrec.ac.in/downloads/cseevents/1.pdf
[+] Done!
```

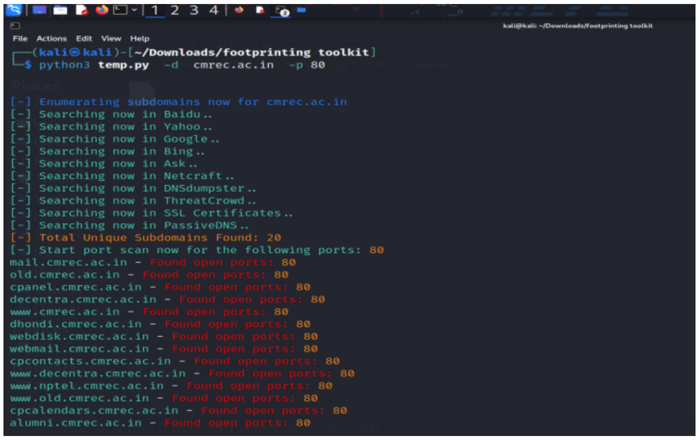
1 2 3 4
kali@kali: ~/Downloads/footprinting toolkit
File Actions Edit View Help

Fig-10: File extraction & Metadata extraction



```
┌──(kali㉿kali)-[~/Downloads/footprinting toolset final]
└─$ python3 temp.py -X  -d  microsoft.com -f -t pdf
[*] Searching for 20 .pdf files and waiting 30.0 seconds between searches
[*] Results: 20 .pdf files found
https://designer.microsoft.com/FAQ.pdf
https://download.microsoft.com/documents/australia/dynamics/digital010/using_the_balanced_scorecard_as_a_strategic_management_system.pdf
https://download.microsoft.com/documents/australia/dynamics/digital010/coming_up_short_on_nonfinancial_performance_measurement.pdf
https://lookbook.microsoft.com/assets/SharePoint_lookbook_2019.pdf
https://visualstudio.microsoft.com/keyboard-shortcuts.pdf
http://nds1.webapps.microsoft.com/phones/files/main_page/Nokia_BH-503_UG_en.pdf
http://research.microsoft.com/~hoppe/perfecthash.pdf
http://research.microsoft.com/~hoppe/quick_reference_math_in_microsoft_office.pdf
http://research.microsoft.com/~hoppe/paratexsyn.pdf
https://adoption.microsoft.com/files/assets/M365AdoptionGuide.pdf
http://research.microsoft.com/~hoppe/mcgim.pdf
http://research.microsoft.com/~hoppe/intersurf.pdf
https://news.microsoft.com/download/hardware/WMM3000FS_FINAL.pdf
https://news.microsoft.com/download/hardware/SideWinderX6FS_FINAL.pdf
http://research.microsoft.com/~hoppe/apptexsyn.pdf
https://news.microsoft.com/download/hardware/WLD6000FS_FINAL.pdf
http://research.microsoft.com/~hoppe/edgetrav.pdf
https://adoption.microsoft.com/files/assets/TeamsWebinarsGetStartedGuide.pdf
https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE4O2wm
https://www.microsoft.com/en-us/research/uploads/prod/2019/10/shiv_quirk_neurips_2019.pdf
[+] Done!
```

Fig-11: Metadata extraction



```
┌──(kali㉿kali)-[~/Downloads/footprinting toolkit]
└─$ python3 temp.py  -d  cmrec.ac.in  -p 80

[-] Enumerating subdomains now for cmrec.ac.in
[-] Searching now in Baidu..
[-] Searching now in Yahoo..
[-] Searching now in Google..
[-] Searching now in Bing..
[-] Searching now in Ask..
[-] Searching now in Netcraft..
[-] Searching now in DNSdumpster..
[-] Searching now in ThreatCrowd..
[-] Searching now in SSL Certificates..
[-] Searching now in PassiveDNS..
[-] Total Unique Subdomains Found: 20
[-] Start port scan now for the following ports: 80
mail.cmrec.ac.in - Found open ports: 80
old.cmrec.ac.in - Found open ports: 80
cpanel.cmrec.ac.in - Found open ports: 80
decentra.cmrec.ac.in - Found open ports: 80
www.cmrec.ac.in - Found open ports: 80
dhondi.cmrec.ac.in - Found open ports: 80
webdisk.cmrec.ac.in - Found open ports: 80
webmail.cmrec.ac.in - Found open ports: 80
cpcontacts.cmrec.ac.in - Found open ports: 80
www.decentra.cmrec.ac.in - Found open ports: 80
www.nptel.cmrec.ac.in - Found open ports: 80
www.old.cmrec.ac.in - Found open ports: 80
cpcalendars.cmrec.ac.in - Found open ports: 80
alumni.cmrec.ac.in - Found open ports: 80
```

1 2 3 4
kali@kali: ~/Downloads/footprinting toolkit
File Actions Edit View Help

Fig-12: Port Scanning

XI. REFERENCES

[1] Gurline Kaur, & Navjot Kaur (2017). Penetration Testing – Reconnaissance with NMAP Tool. International Journal of Advanced Research in Computer Science.

[2] Sheetal Temara, (2024). Maximizing Penetration Testing Success with Effective Reconnaissance Techniques using ChatGPT. Asian Journal of Research in Computer Science.

[3] Bandar Abdulrhman Bin Arfaj, Shailendra Mishra, & Mohammed Alshehri, (2021). Efficacy of Unconventional Testing Practices. Tech Science Press. ResearchGate.

[4] Manikanta Singirikonda, (2023). Penetration Testing Tool Guide. Journal of Cybersecurity. ResearchGate.

[5] Fouz Barman, Nora Alkaabi, Hamda Almenhali, Mahra Alshedi, & Richard Ikuesan, (2023). A Methodical Framework for Conducting Reconnaissance and Enumeration in Ethical Hacking Life Cycle. ECCWS.

[6] Arunima Santhosh, Rinimol Kurian, Amal Jyothi College of Engineering, (2021). Identifying Subdomains of the Website Using Sublist3r and Comparing Sublist3rAmass to Knockpy. Proceedings of the National Conference on Emerging Computer Applications (NCECA).

[7] M. Alby, I. Ruslan, Muharmah, I., (2023). Information Security Test on Websites and social media using Footprinting Technique.

[8] Yuvraj Singh, (2022). Footprinting using Nmap. Journal of Informatics, Electrical and Electronics Engineering.

[9] Shreya, S., (2020), Footprinting: Techniques, Tools, and Countermeasures for Footprinting, Journal of Critical.

[10] Matthew Denis; Carlos Zena; Thaier Hayajneh, (2016). Penetration testing: Concepts, attack methods, and defense strategies, IEEE.