IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Automating The Web Application Reconnaissance Process - All In One Recon

Ranjeet Kumar Singh (11802170)
Dr. Om Prakash Yadav, (UID-26121) Assistant Professor,
School of Computer science and engineering, Lovely Professional University,
Jalandhar - Delhi, Grand Trunk Rd, Phagwara, Punjab 144001

ABSTRACT - Generally in reconnaissance phase of hacking we have to use lots of different tools like for finding subdomains we have to use tools like Subfinder, Assetfinder, Sublister etc after than extract all unique subdomains from these tools then remove common subdomains and gather all subdomains one file. After this whole time consuming process we run some tools or manually check one by one to find which are active subdomains and which of them are dead so here lots of tools are present but when we first gather then extract then send collected list to this tools it will takes lots of time. So after this in third process we run nmap tool to gather open ports so we can identify which services are running on which port so we can do this either by running one by one command or directly send all collected subdomain list to nmap tool but again this will again consume lots of time if we do manually. After then in fourth step we go to wayback url tool and collect manually one by one or using some tool to collect all the past used endpoints or java script file. After this in fifth process we take screen shot of all web applications to see what is interface of website application and which services is running so this all process consumes lots of time and but using my tool I am going to solve these problems.

KEYWORD: Web App Pentesting, Reconnaissance, Vulnerability Assessment, Pentest

INTRODUCTION – Reconnaissance is key to any successful hacks. On average, approximately three-fourths of any hack should be spent performing accurate and precise reconnaissance. Reconnaissance is the act of gaining information about our targets. Such as all subdomain, active subdomain, open ports, operating systems, what services those ports are running, screenshot of UI and any vulnerable applications they have installed. All of this information will be absolutely important to choosing an attack. How are we supposed to hack if we don't know what we are hacking into?

There are two base types of recon, active and passive. Both have their pros and cons, so let's cover these types of recon briefly:

1. Active Recon: This type of reconnaissance requires that we interact with the target. This reconnaissance is faster and more accurate, but it also makes much more noise. Since we have to interact with the target to gain information, there's an increased chance that will get caught by a firewall or one of the network security devices. (Intrusion Detection Systems, any firewalls, etc.)

2. Passive Recon: This type of reconnaissance doesn't require any interaction with the target, so it is far less likely to be detected. The trade off is that the information gained is not as accurate and it's much slower than it active counterpart. Passive reconnaissance is the act of watching the targets. Instead of interacting with them, we can watch their traffics and gain information without so much as pinging targets.

Steps followed in reconnaissance –

- Collect subdomains
- Check which of them are active
- Recognize all ports
- Collect endpoints
- Take screenshot
- Reveal services used on ports
- Understand the network nmap

Now that we've covered the two base types of reconnaissance, let's go over some of the recon terms that we'll hear commonly:

Discovery : This is the act of discovering possible victim targets. Discovery is essential to recon as it tells us who our potential victims are. There are lots of tools available using which we can do discovery like assetfinder, subfinder, sublister.

Collect active subdomains : As the name implies, this is a process where we check which subdomains are active and which of them are working after than collect all active subdomains in one file.

Port Scanning: As the name implies, this is the act of scanning a range of ports on a victim targets. A port is used to make connections and manage communication for net-workable services or applications. Any open port is a possible avenue of attacks. There are multiple kinds of port scan, but those go beyond the scope of this introductory article. So for this we are going to use nmap tool.

Wayback url : So for finding all previous and deleted files we are going to use this tool and using this tool we can collect all javascript , php, txt, etc all files.

Screenshot : So after performing all initial recon we are now on final step and here we are going to take screenshot of all available active subdomains.After taking screenshot we can easily scroll and identify which service and which of them subdomains has same UI, server.

LITERATURE SURVEY

Nagendran Ketal in [1] explains the technical approach to perform a manual penetration tests in web applications for testing the security of the applications and it serves as a great guide to look for security vulnerabilities. It provides us with various technique to secure web applications from hackers. Ahana Royetal in [2] says that they proposed a tool which gather the footprint of a corporation, helpful for information gathering phase during a penetration test and it is found that there is a lack of an easy tool which can help in the first stage of such penetration tests, reconnaissance. The Java-based tool greatly help in gathering organization specific data. These data storages help greatly invulnerability evaluation of a firm. Kristian Beckers in [3] describes the details of a survey done on tools in 2017 which are there for social engineering and intelligence gathering. It presents an outline of their specifications and capabilities. It describes that attackers have a wide range of Opensource intelligence gathering tools which greatly increases the likelihood of the attacks in the future. Usman Ali Dar in [4] explores different kinds of recon techniques that are used by an attacker or hacker to collect information regarding the target. It determines which techniques gathers the most info about the target while keeping itself hidden to the internet.

Dr Arun Kumar in [5] explains that as Web Applications are 2022 International Conference for Advancement in Technology (ICONAT) Goa, India. Jan 21-22, 2022978-1-6654-2577-3/22/$31.00 ©2022 IEEE1increasingly used for complex services, they become a popular and great targets for security attacks. Plenty of techniques have been developed to secure website applications and stop the attacks towards web apps, there is a very little effort devoted to drawing conclusions among these techniques and developing a broader view of the web application security researches. This papers gives an outlined examination of assaults against picked critical parameters.

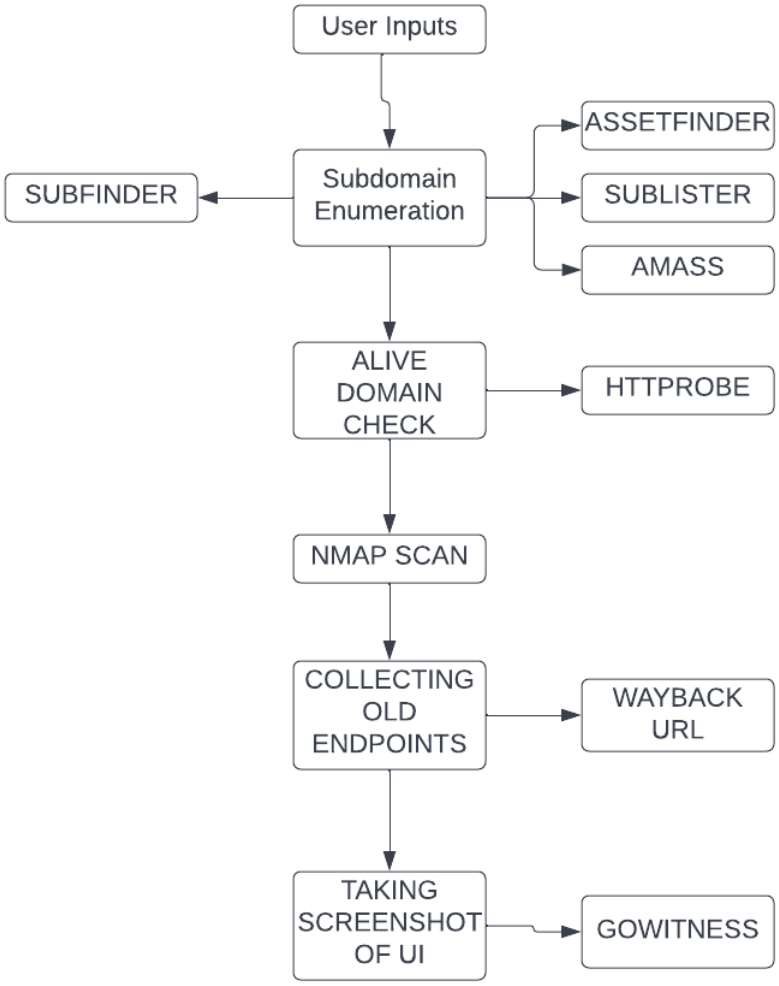
S.M. Zia Ur Rashid in [6] describes that the Domain name system has been an essential part of cyber security and an essential part of the web service used. The nameserver are completely responsible for the safety and functionality of their domain names. But as there is lag of sample security and DNS misconfiguration, there can be a chance to take over the subdomain from the external services. These papers mainly focuses on detailed analysis on subdomain takeover, map out the bug's impact on the firm. Tae Hyun Kim et.al in [7] describes that DNS is used to provide scalable name resolution services to the users in an easy and efficient manners. However, DNS was developed without security initially, and the data is not secure. We describe the overview of DNS bug, DNS attack, and even protection systems. In detail, attacks are divided by purpose and technique for defending against the attacks that are introduced and analysed. The important finding of this work is to introduce basic vulnerabilities of the DNS.

The paper [8] describe that is easy to find log and bugs in server-side applications but when we use client-side web applications it is more complex. The front end of client-side app uses Angular, React etc which flags the way for vulnerabilities. The static analysis is performed to find vulnerabilities like secret key to API, finding domains, Potential wild card entry etc. Script Hunter by Robre is used for finding JavaScript file. But before using this, we need to install Golang properly. The paper [9] uses JavaScript enumerations, DOM XSS vulnerability can be exploited. JavaScript enumeration can be time consuming. The steps included in JavaScript enumeration are extracting JavaScript file, beautify the JavaScript code, JavaScript enumeration using grep.

METHODOLOGY - Workflow of the proposed model :

- Take the input (top level domain) from the user as a command line argument to the reconnaissance script.
- Perform subdomain enumerations on the target (top level domain names)
- Extract all the live subdomains which have a web server running on them from the enumerated subdomain list.
- Also gather the status code and titles of the live subdomains.
- Check dead subdomain.
- Perform google dork on the subdomains using tools.
- Get all the URLs once present on the target using wayback machine or any tools.
- Perform nmap scan on the targets.
- Perform screenshot of web app.

Flowchart :



This flow diagram describes the working of the proposed model.

So in (1) image user is giving input.

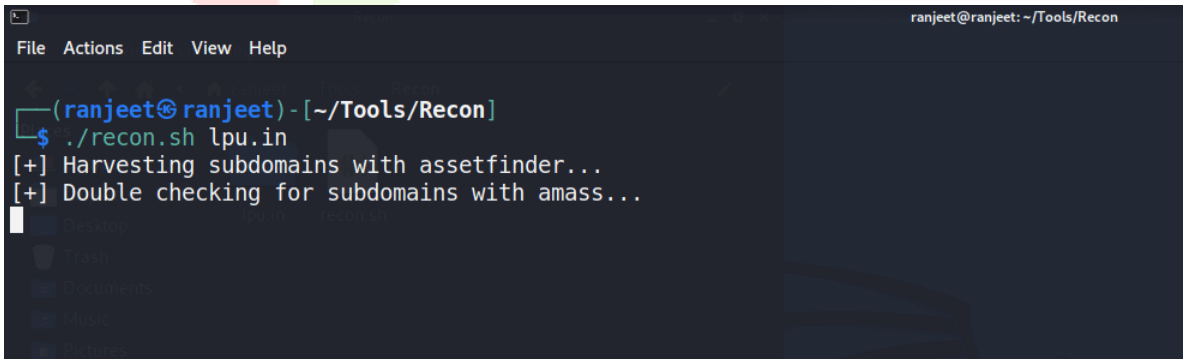


Figure 1: user input

In (2) image it is creating directories.

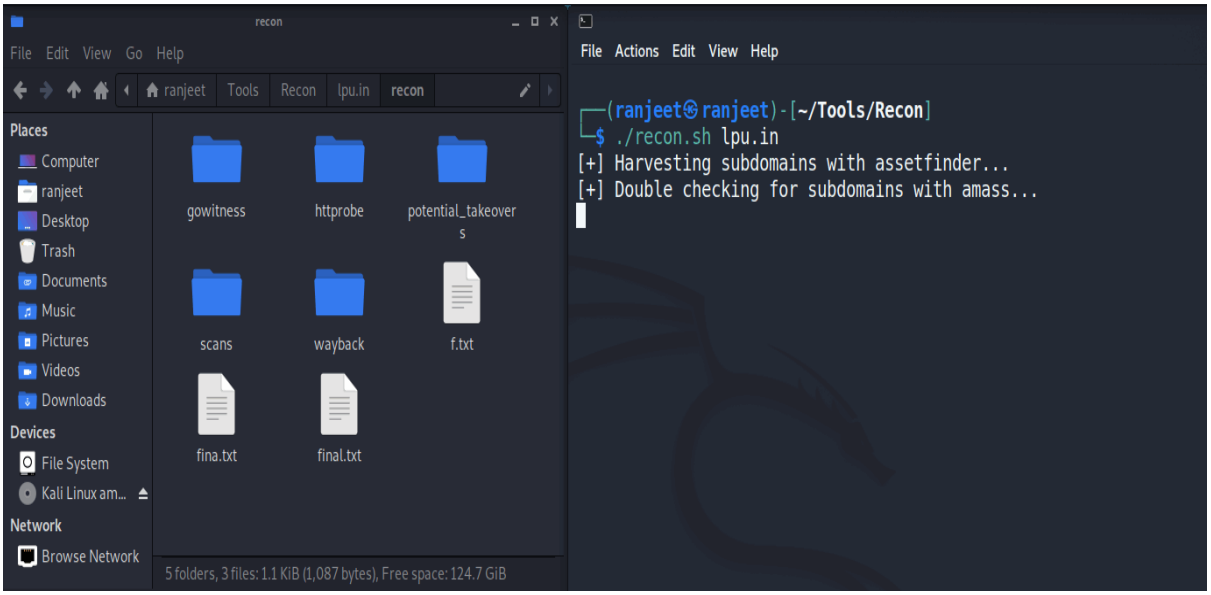


Figure 2 creating directories for saving data

In (3) image it is gathering subdomains using assetfinder, amass, subfinder and sublister then checked for alive subdomains and then for subdomain takeovers after than it has done nmap scan and at last gathered past endpoints and taken screenshot.

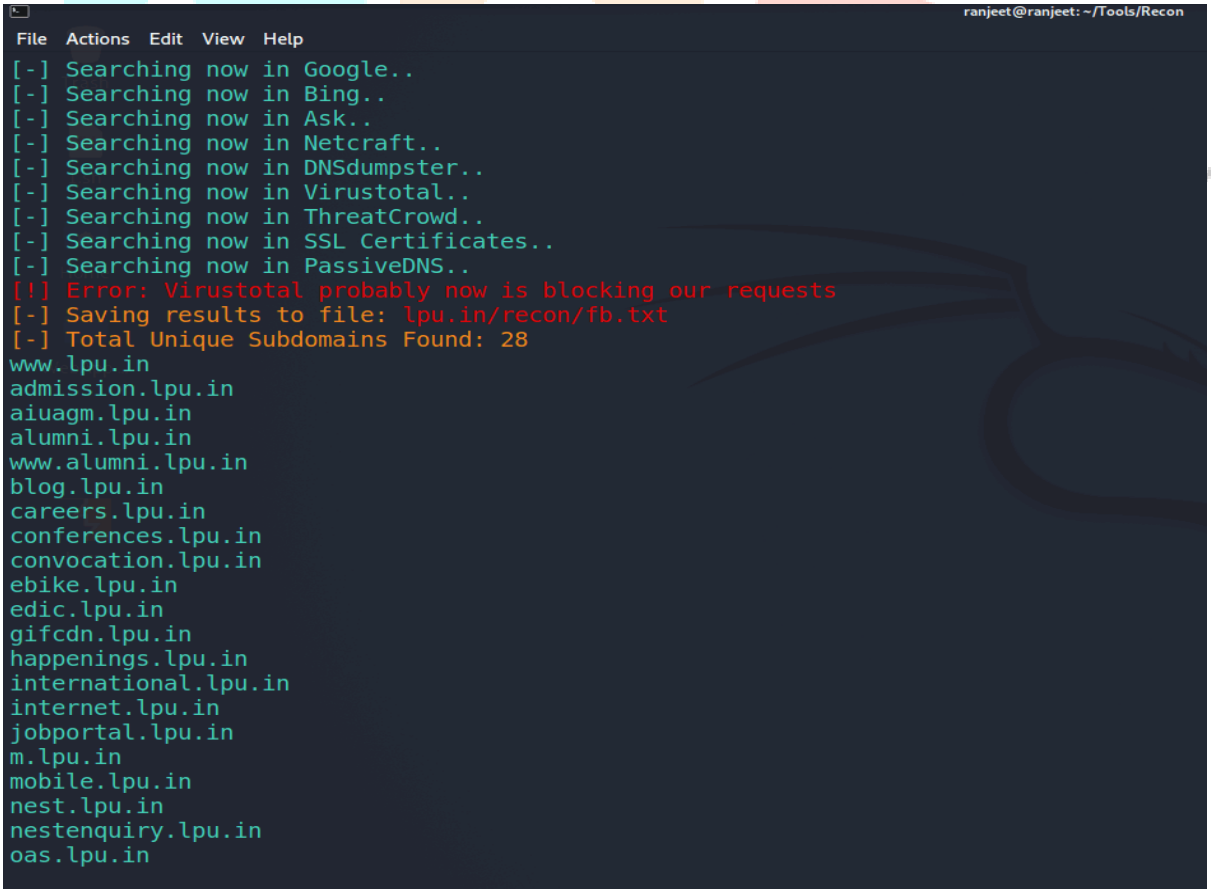


Figure 3 Checking alive subdomains

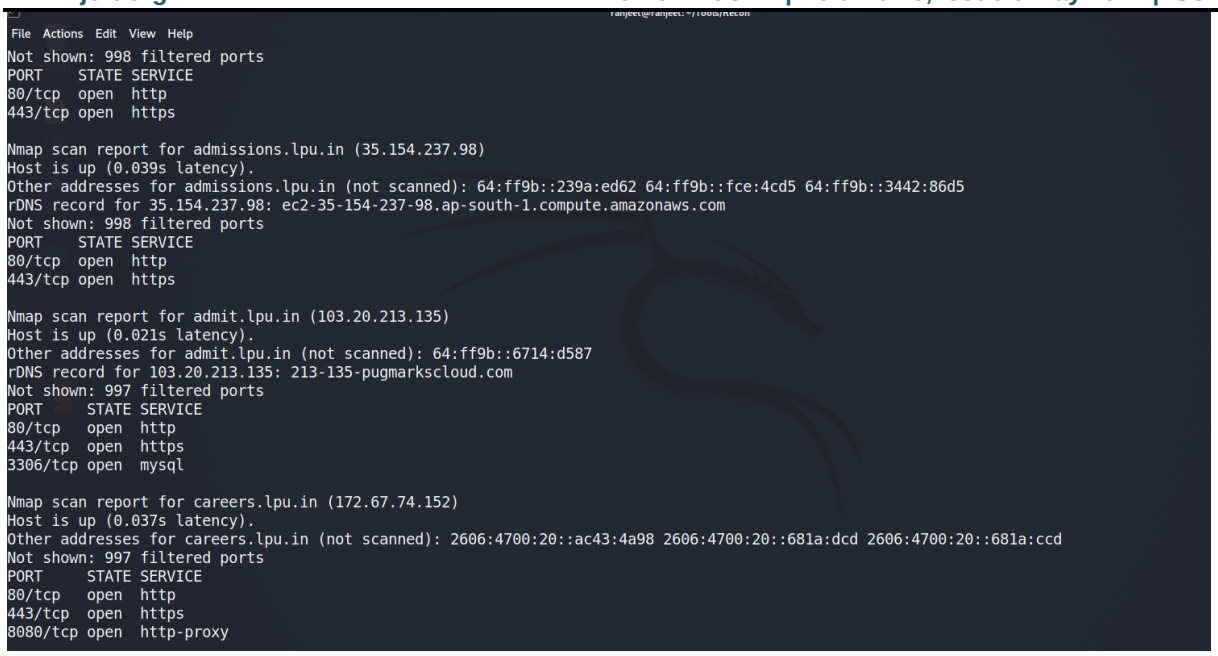


Figure 4 Running Nmap

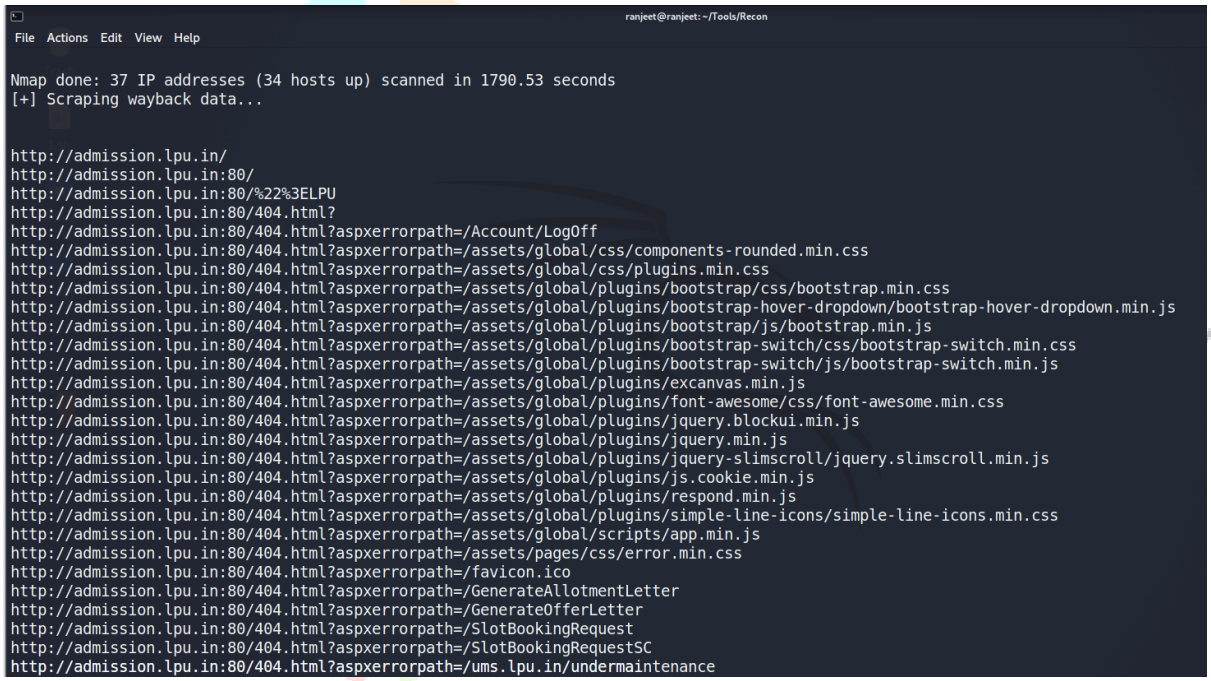


Figure 5 Gathering old java script files and taking screenshot

RESULTS AND DISCUSSIONS

The script first creates a few empty directories where the result of the script are stored. Then the subdomain enumeration starts showing and all the subdomains are stored in a text file in subdomains folder. Website (http or https) with their status code and title are extracted from the subdomains list. And then credential from breached data are collected. All the past URLs of the website are collected using wayback url. Then, JavaScript

enumeration start and collects all the available .js files of the targets Then, all in one recon starts its scan and finds all active subdomains then do ports scanning and then using way back url to gather all past endpoints and after that it takes screen shot of web applications. Port scanning does its job by collecting all the open ports, services running on them, and their version. Finally, all the results are stored in their respective directories. All a user needs to do is navigate to the directory of his choice and view the text files using any text editor like vim ,nano or he can do cat. The penetest job is greatly reduced. He/she just needs to run the model which is basically a shell script by giving the target's domain as an argument to

the script, sit back and then relax!! The script does its job and show the result after the execution is completed. Each module's output is stored in its own directory.

CONCLUSION

The various task of information gathering phase of a pentesting are very tedious and require a great human time and effort. This project simply aims at the automation of the project and makes the life of a penetration tester easier than never by shifting his/her task to the computers. This project is divided into various modules where each module does a specific job and at the end, the result of each module combinedly, is the output of this reconnaissance project and this can be helpful for a penetration tester to proceed with the other phases of penetration testing like exploitation. All a user needs to do is navigate to the directory of his choice and view the text files using any text editor like vim , nano or simply he can use cat. The pentesting job is greatly reduced.He/she just needs to run the models which is basically a shell script by giving the target's domain as an argument to the script, sit back and relax!!Nmap Scan probes for all the open ports present on the target's web app, the services running on those ports and their versions too. Information leakages, lack of security headers, subdomain takeovers etc.

Future Work

I present a quick analysis tool to locate sensitive information of organizations (should be only used by penetration testers and ethical hackers). I will be working on creating a master list of data which would contain sensitive pattern or signatures. This master list will be used by the tool to compare with the file in search. Only those file will be downloaded which may contain critical information, which should have been kept private and as a secured entity. In future, we also intend to build a tool in bash that will automatically find bugs in website application. We will also try to implement a vulnerability assessment tool like Nessus, Acunetix as part of our future work.

REFERENCES

[1] Ahana Roy, Louis Mejia, Paul Helling, Aspen Olmsted "Automation of Cyber Reconnaissance: A Java based open-source tool for information gathering",published at 2017 12th International Conference for InternetTechnology and Secured Transactions (ICITST)

[2] Kristian Beckers, Sebastian Pape, Peter Schaab, Daniel Schosser, "Conference: International Conference on Trust and Privacy in Digital Business", August 2017

[3] Tae Hyun Kim, Douglas Reeves, "A survey of domain name system vulnerabilities and attacks", January 2017

[4] Usman Dar, Arsalan Iqbal, "The Silent Art of Reconnaissance: The Other Side of the Hill", January 2018.

[5] Arun Kumar, Sandeep Arora, "A Review on Web Application Security", March 2018

[6] S M Zia Ur Rashid, MD Imtiaz Kamrul, Asraful Islam, "Understanding the Security Threats of Esoteric Subdomain Takeover and Prevention Scheme", published at 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), Feb 2019.

[7] Nagendran K, Adithyan A, Chethana R, Camillus P, Bala Sri Varshini K B "Web Application Penetration Testing," at International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-10, August 2019

[8] Andres Ojamaa, Karl Duuna, "Assessing the security of Node.js platform", published at 2012 International Conference for Internet Technology and Secured Transactions, Dec 2012

[9] Nataliia Bielova, "Survey on JavaScript Security Policies and their Enforcement Mechanisms in a Web Browser",published at Journal of Logic and Algebraic Programming, November 2019

[10] Ankur Taly, Ulfar Erlingsson, John C. Mitchell, Mark S. Miller, Jasvir Nagra,"Automated Analysis of Security-Critical JavaScript APIs"

[11] Arjun Guha, Claudiu Saftoui, Shriram Krishnamurthy, "The Essence of JavaScript "

[12] Jaspher Kathrine, Ronnie T Baby, V. Ebenzer, "COMPARATIVE ANALYSIS OF SUBDOMAIN ENUMERATION TOOLS AND STATIC CODE ANALYSIS", ISSN (Online) : 2454 -7190 Vol.-15, No.-6, June (2020) pp 158-173 ISSN (Print) 0973 8975

[13] Rizdqi Akbar Ramada, Redho Maland, Dedi Hariyadi, "Sudomy: Information Gathering Tools for Subdomain Enumeration and Analysis", The 2nd International Conference on Engineering and Applied Sciences 2019 (2nd InCEAS 2019)At: Yogyakarta, Indonesia, Volume: 771, March 2020

[14] Mayur Parmar, "Google Dorks -Advance Searching Technique", August 2019

[15] Marco Squarcina, Mauro Tempesta, and Lorenzo Veronese, TU Wien; Stefano Calzavara, "Can I Take Your Subdomain? Exploring Same-Site Attacks in the Modern Web", Università Ca' Foscari Venezia & OWASP; Matteo Maffei, TU Wien

[16] Suraj S.Mundalik, "Penetration Testing: An Art of Securing the System (Using Kali Linux)", published at International Journal of Advanced Research in Computer Science and Software Engineering, October 2019

[17] Sushmita Reddy Mamilla, "A Study of Penetration Testing Processes

and Tools", May 2019.

[18] Monawar H. Bhuyan, Dhruba K. Bhattacharya, Jugal Kalita, "Surveying Port Scans and Their Detection Methodologies", The Computer Journal 54(10):1565-1581, October 2011

[19] Marco de Vivo, Le Ke, Germinal Isern, Gabriela O. de Vivo, "A review of port scanning techniques", ACM SIGCOMM Computer Communication Review 29(2):41-48, April 2019

[20] Vinitha K P, "Ethical Hacking", published at INTERNATIONAL

JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY, May 2019

[21] Bowman H.Miller, "Open Source Intelligence (OSINT): An Oxymoron?", December 2019

[22] Javier Paster-Galindo, Pantaleone Nespoli, Felix Gomez Marmol, Gregorio Martinez Perez, "The Not Yet Exploited Goldmine of OSINT: Opportunities, Open Challenges and Future Trends", January 201

[23] Himanshu Singh, "Distributed Port Scanning Detection", 2021

[24] Muharman Lubis, Nurul Ibtisaam Yacoob, Hafizah Binti Reh, Montadzah Ambag Abdulgani, "Study on Implementation and Impact of Google Hacking in Internet Security", Regional Conference on Knowledge Integration in ICT 2010At: Selangor, June 2020

[25] Mamta Bhavsar, Dr Priyanka Sharma, Manik Gokani, "Port Scanning using Nmap", published at International Journal of Engineering Development and Research, December 2020.

[26] R. Vijaya Saraswathi, L. Padma Sree , K. Anuradha, "Dynamic group key management scheme for clustered wireless sensor networks", International Journal of Grid and Utility Computing,September 2021.

[27] Vijaya Saraswathi R., Padma Sree L., Anuradha K. (2020) Secured Cluster-Based Distributed Dynamic Group Key Management for Wireless Sensor Networks. In: Pant M., Sharma T., Basterrech S., Banerjee C. (eds) Computational Network Application Tools for Performance Management. Asset Analytics (Performance and Safety Management). Springer, Singapore. https://doi.org/10.1007/978-981-32-9585-8_18.

[28] RV Saraswathi, LP Sree, K Anuradha," Support Vector Based Regression Model to Detect Sybil Attacks in WSN", International Journal of Advanced Trends in Computer Science and Engineering, June 2021.

[29] V. S. Manvith, R. V. Saraswathi and R. Vasavi, "A Performance Comparison of Machine Learning Approaches on Intrusion Detection

[30] Dataset," 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), 2021, pp. 782-788, doi: 10.1109/ICICV50876.2021.9388502. 5

[31] R. V. Saraswathi, L. P. Sree and K. Anuradha, "Dynamic and probabilistic key management for distributed wireless sensor networks," 2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), 2016, pp. 1-6, doi: 10.1109/ICCIC.2016.7919666.

[32] RV Saraswathi, LP Sree, K Anuradha, "Key management schemes in wireless sensor networks: a survey", CiiT International Journal of Wireless Communication, 2016.

[33] RV Saraswathi, LP Sree, K Anuradha, "Multi-stage key management scheme for cluster based WSN" International Journal of Communication Networks and Information Security, December 2021.

[34] Mandala Mounica, R Vijayasaraswathi, R Vasavi, "Detecting Sybil Attack In Wireless Sensor Networks Using Machine Learning Algorithms", IOP Conference Series: Materials Science and Engineering,2021.