2022 International Conference for Advancement in Technology (ICONAT)
Goa, India. Jan 21-22, 2022

Automation of Recon Process for Ethical Hackers

Vijaya Saraswathi R[1]
Assistant Professor,CSE
VNRVJIET
Vijayasaraswathi_r@vnrvjiet.in

Iftequar Ahmed Sk[2]
Department of CSE
VNRVJIET
iftequar.154@gmail.com

Sriveda Reddy M[3]
Department of CSE
VNRVJIET
sreeveda.reddy18@gmail.com

Akshay S[4]
Department of CSE
VNRVJIET
shadipuramakshay27634@gmail.com

Vrushik Reddy M[5]
Department of CSE
VNRVJIET
vrushik.2k@gmail.com

Sanjana Reddy M[6]
Department of CSE
VNRVJIET
morasanjanareddy@gmail.com

***Abstract*— The need for Recon automation is rapidly increasing as ethical hackers are being lazy in performing every little check manually. So as to make the Recon process (Info gathering phase) of penetration testing easy, fast and accurate, a Recon framework with highly sophisticated tools written in languages like bash, go and python needs to be developed and made open source to everyone. Manually doing this task can be very intimidating since a lot of time and efforts are needed in accomplishing this task. So, automation of this task can be very handy to the penetration testers and saves a lot of time as they can focus on other tasks of the further tasks of a penetration test. So, our project is an automation to the tedious task of information gathering. This Recon Framework just takes the main top-level domain of the organization as the input, does the recon and stores the result in an organized manner in the corresponding directories. The output of this framework is ready to be used to perform further security tests as the results are generated in a neat greppable format and can be passed to other tools to further filter the data according to the ethical hacker's wish and need. In addition to that, the results are displayed in a graphical interface in the form of a web application. So, all a user needs to do is enter the top-level domain name of the organization on which he/she wants to perform penetration testing.**

Keywords*— Recon, penetration testing, subdomain enumeration, port scanning.

I. INTRODUCTION

The explosive growth of the web has brought many goodies like E-commerce, E-mail, Cloud computing, but there's also a black side of Hacking, malwares etc. Hacking is one of the biggest problems faced by tech companies, governments, and citizens across the world. An Ethical Hacker or a Penetration tester can assist or help the people that are suffering from these cyber-attacks. Ethical hacking is usually termed as online geeks or groups that legally access the company's online assets after obtaining official approval. Reconnaissance means the preparatory stage where an Ethical Hacker seeks to collect the maximum amount of information as possible a few targets before launching a security test. It involves 3 phases namely, scanning, foot printing, and enumeration of the organization's network. In this project, we'll be handling automation of foot printing of an organization. Foot printing is nothing but the blueprint of the safety of a corporation, undergone during a procedural manner. It finds all information available on the internet about the target. It is a time-taking process to flick through the sites and collect info; hence in this paper, we investigate the solution for tedious web search and propose a proficient way to organize, extract and store data from the search engines employing a new tool, Search Simplified. Info gathering techniques are often broadly divided into the following:

Active: This includes intrusive recon that sends (specially built) data to the target, for example, port-scanning. Advanced network foot printing techniques dodge direct connections with the target host.

Passive: This is the kind of reconnaissance that either does not contact or communicate directly to the target system or that uses publicly available information, and not normally found from standard logs. This paper also focuses on this technique.

Both active and passive reconnaissance can cause the invention of useful data to use in a malicious activity. This information may enable an attacker to seek out vulnerabilities in the OS's version and exploit the loophole to gain more access. Shell-script based Recon Framework is a fully-featured recon framework which is written in shell script. It provides a great environment where open-source reconnaissance is often carried out in a timely and thorough manner.

II. RELATED WORK

Nagendran K et.al in [1] explains the technical approach to perform a manual penetration test 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 techniques to secure web apps from hackers. Ahana Roy et.al in [2] says that they proposed a tool which gathers 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 helps in gathering organization-specific data. These data storages help greatly in vulnerability evaluation of a firm.

Kristian Beckers et.al 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 attacker have a wide range of Opensource intelligence gathering tools which greatly increases the likelihood of the attacks in the future. Usman Ali Dar et.al in [4] explores different kinds of reconnaissance techniques that are used by an attacker or hacker to collect information regarding the target. It determines which technique gathers the most info about the target while keeping itself hidden to the internet. Dr Arun Kumar et.al in [5] explains that as Web Apps are

2022 International Conference for Advancement in Technology (ICONAT) | 978-1-6654-2577-3/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICONAT53423.2022.9726077

increasingly used for complex services, they become a popular and great target for security attacks. Plenty of techniques have been developed to secure web 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 paper gives an outlined examination of assaults against picked critical parameters. S.M. Zia Ur Rashid et.al in [6] describes that the Domain name system has been an essential part of cyber security and an essential part of the web services used. The nameservers are completely responsible for the safety and functionality of their domain names. But as there is lag of ample security and DNS misconfiguration, there can be a chance to take over the subdomain from the external services. This paper 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 manner. However, DNS was developed without security initially, and the data is not secured. We describe the overview of DNS bugs, DNS attacks, and even protection systems. In detail, attacks are divided by purpose and techniques for defending against the attacks that are introduced and analysed. The important findings of this work is to introduce basic vulnerabilities of DNS. The paper [8] describes that is easy to find logs and bugs in server-side applications but when we use client-side applications it is more complex. The front end of client-side applications uses Angular, React etc which flags the way for vulnerabilities. The static analysis is performed to find vulnerabilities like secret keys to API, finding domains, Potential wild card entries etc. Script Hunter by Robre is used for finding JavaScript files. But before using this, we need to install Go properly. The paper [9] uses JavaScript enumeration, DOM XSS vulnerabilities can be exploited. JavaScript enumeration can be time consuming. The steps included in JavaScript enumeration are extracting JavaScript files, beautify the JavaScript code, JavaScript enumeration with grep.

The paper [10] tells us that most of the companies are using frameworks like React, Vue, Ember etc instead of JavaScript. However, there are various tools to convert them to JavaScript. Dev Tools tab is used to inspect JavaScript code in Chrome. Map files give us a way to go through the original source code. Arjun Guha et.al in [11] describes that in static analysis, we need to identify and gather JavaScript files, make the JavaScript file readable and finding security issues. Jaspher Kathrine et.al in [12] tells us about the patterns which play a vital role in Subdomain Enumeration. Since, the patterns like qa, dev, staging, api, uat are repeatedly used by the developers in naming the subdomains, enumeration of subdomains can be done easily by brute forcing these patterns. Rizdqi Akbar Ramadan et.al in [13] explains the importance of performing ample reconnaissance that will increase the chance of finding vulnerabilities. Subdomain Enumeration techniques are also explained here. Mayur Parmar in [14] uses search string which uses advanced options to find the hidden information. It is useful for bug bounties for performing network mapping, port scanning, information gathering etc. Marco Squarcina et.al in [15] defines the threats posed by related-domain attackers to web application security. The paper describes on some vulnerabilities like CORS(Cross-Origin Resource Sharing), CSP(Content-Security Policy) bypass, postMessage and domain relaxation. Suraj S. Mundalik et.al in [16] explores information gathering techniques and the attack simulation implementation that is done particularly with Kali Linux OS by using various pentesting tools. Kali Linux OS makes it easier to perform pen testing on the target host with the help of its huge tool set which is free of cost and present in open source. He also emphasises on a basic overview of the various tools present in the Kali Linux Operating System. Sushmitha Reddy Mamilla in [17] explains her project that compares some basic scanning tools in terms of the no. of ports found open and the time taken by the tool to find those open ports. A cmprehensive analysis of the results generated by the tools will be used to find the efficient tools. Monowar et.al in [18] tells us about port scans performed by attackers to discover weak systems to compromise. This paper also discusses about the common port scanning attacks. Marco de Vivo et.al in [19] describes the techniques that the TCP port scanners use. Administrators also use port scanning to prevent the unwanted exploitation by the port scanners. It also describes the various services' vulnerabilities that can be found on those pen ports. Vinitha K in [20] explains the basic security issues in the modern web applications and also describes the types of hackers. This paper also focuses on the various ethical hacking phases which are being used by both hackers and penetration testers too.

Bowman Miller in [21] focuses on the importance of OSINT by telling the fact of having special wings for OSINT in US Air Force, US Dept of State and many other government organizations of the nation. This paper also describes the various OSINT techniques used in World War II. Javier Pasto-Galindo et.al in [22] describes the Opportunities, open challenges and Future Trends of the OSINT techniques and methodologies. This paper emphasises on the fact that a ton of information available to gather information about everything. Himanshu Singh in [23] describes the various ways of detecting active port scans running against an organization. This paper also describes the patterns found in TCP packets while scanning for ports and details the working of the packet sniffer. Muharman Lubis et.al in [24] describes the importance of Google dorking before proceeding into further steps of a penetration test. This paper also focuses on the Google Hacking Database which is a repository of various google dorks written by people across the globe. Mamta Bhavsar et.al in [25] dives deep into the various scanning techniques of the famous port scanning tool called Nmap. This paper also describes the various TCP packets sent in this process.

III. EXISTING MODELS

The existing models available today have very limited features and are not compatible with the modern web development frameworks like MEAN and MERN stacks, Django, Flask or Spring Framework. As these new technologies and tech stacks came into limelight, there are many things which are overlooked and often many vulnerabilities are missed when tried with the existing recon frameworks. Some of the important things missed by the existing recon frameworks are:

- JavaScript file enumeration and analysis.
- Automation of Google Dorking

- Automation of some known OWASP vulnerabilities like XSS, SSRF etc.
- Absence of project discovery's nuclei at the time of writing old recon frameworks.
- Automation of fuzzing for endpoints on the target.

Since the existing frameworks did not incorporate multi-threading in their tools, the recon process takes a lot of time. The output management hasn't been up to the mark in any of the frameworks. And in addition to that, each recon framework lacks one or the other features like speed, accuracy, etc. These are the limitations of the existing models and thus there is a need for a fast and accurate framework which automates every single module of the information gathering phase.

IV. Proposed Model

Keeping in view the existing models, this proposed model is an attempt to overcome the limitations of the existing models and having updated tools and techniques which are mostly based on fingerprints of various endpoints of the target web application.

TABLE I. Key features of proposed model

S.No	Features of the project
1.	The speed of the recon process is great as the tools are written in Go and Python and support multi-threading.
2.	Nuclei is another great tool in finding low hanging vulnerabilities.
3.	JavaScript enumeration is made simple than ever before.
4.	Dorking is now just one click away as all the main dorks for a target are incorporated in the project.
5.	With evolved wordlists, content enumeration is very effective with our project.

A. Workflow of the proposed model

- Take the input(top-level domain) from the user as a command line argument to the recon script.
- Perform subdomain enumeration on the target(top level domain name)
- Extract all the live subdomains which have a web server running on them from the enumerated subdomains list.
- Also gather the status codes and titles of the live subdomains
- Perform google dorking on the subdomains.
- Get all the URLs once present on the target from waybackmachine.
- Perform credential stuffing on the target.
- Perform JavaScript enumeration on all the live subdomains.
- Perform fuzzing to find the hidden functionality and content on the subdomains.
- Perform a simple port scan to have an idea of what ports are open and what services are running on them.
- Perform nuclei scan on the target.
- Look for some simple vulnerabilities like Open Redirects, XSS, SSRF on some parameters obtained from the waybackurls.

B. Flow Diagram of the Proposed Model

The step-by-step process of the working of the proposed model is explained in the form of a flowchart in the figure 1.The flowchart clearly depicts the working flow of the process of how each module of the model helps in gathering information which can be a greatly useful for further phases of a penetration test. Each major task like subdomain enumeration, dorking, javascript analysis, content enumeration etc. are termed as a module in this model and thus each module contributes to the final output of the proposed model.

V. Results and Discussion

The script first creates a few empty directories where the results of the script are stored as shown in the figure 2. Then, the subdomain enumeration starts as shown in the figure 3 and all the subdomains are stored in a text file in subdomains folder. Websites (http or https) with their status codes and titles are extracted from the subdomains list as shown in the figure 4. And then credentials from breached data are collected as shown in the figure 5. All the past URLs of the website are collected using waybackurls. Then, JavaScript enumeration starts and collects all the available .js files of the target as shown in the figure 6.

Then, project discovery's nuclei starts its scans and finds some low hanging vulnerabilities as shown in the figure 7. Port scanning does its job by collecting all the open ports, services running on them, and their versions too as shown in the figure 8. Then, hidden content is found using file and directory enumeration as shown in the figure 9. 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 or nano. The penetration tester's 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 relax!! The script does its job and shows the results after the execution is completed. Each module's output is stored in its own directory.

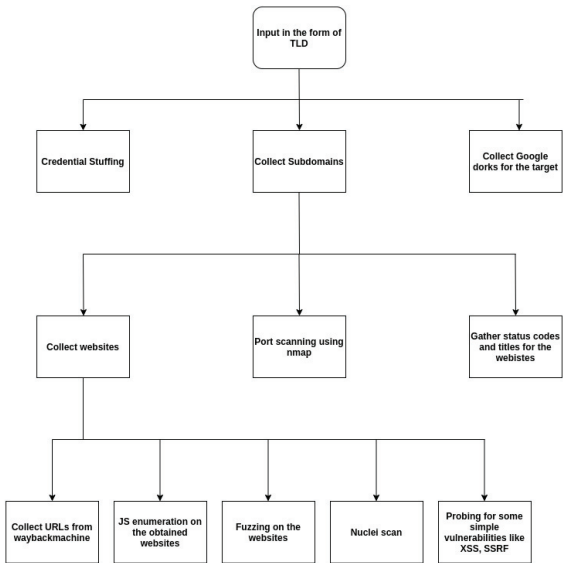

Fig 1: Flow Diagram of the proposed model

This flow diagram describes the working of the proposed model on a broader view. The output of one module is sometimes taken as input to another module as shown in the figure 1.

Fig 2: Output Directories for each module

Each module has its own directory where the output of that module is stored. Subdomains are stored in subdomains directory; nuclei results are stored in the nuclei directory and similarly for the other modules.

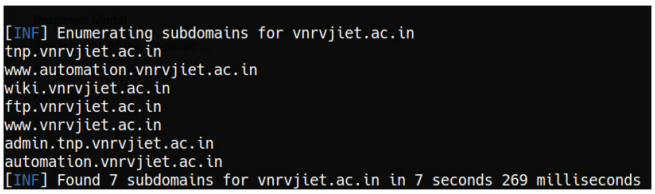

Fig 3: Subdomain enumeration in progress

Subdomain enumeration is the process of finding all the subdomains that are present on the given subdomain. The script collects all the subdomains and stores them in the subdomains directory.

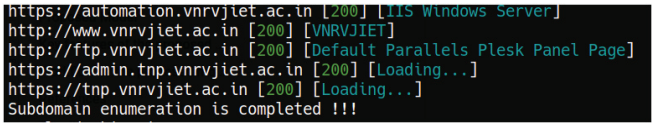

Fig 4: Collecting website titles and Status codes

Collecting websites' titles and status codes can help penetration testers to have a rough idea on the content present on that website. It can sometimes disclose the server being used on backend of the web infrastructure as shown in the figure 4.

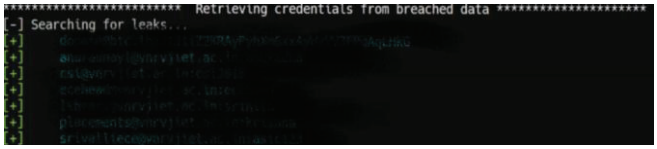

Fig 5: Credential Stuffing in progress

Credential Stuffing is a process where all the credentials of the users of the target are gathered from breached data available on the internet. If those people haven't changed their password yet, it can be a major security threat to the organization too.

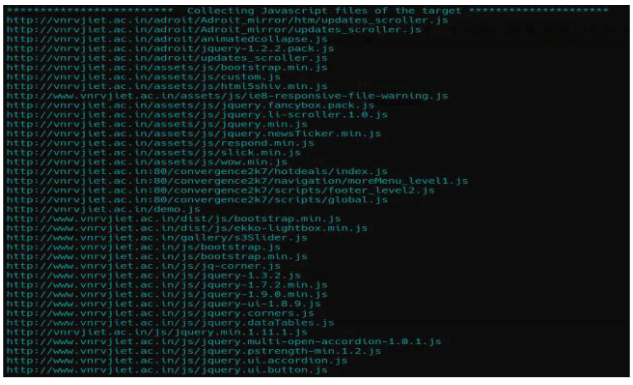

Fig 6: Collecting JavaScript files of the target

JavaScript enumeration is very important because sometimes developers hardcoded credentials, passwords and hidden admin endpoints in JavaScript files.

Fig 7: Nuclei scan in progress

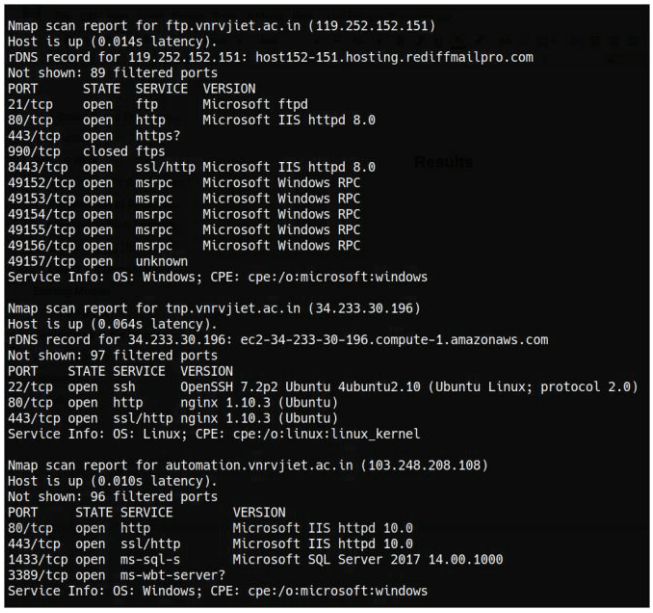

Fig 8: Port Scanning by Nmap

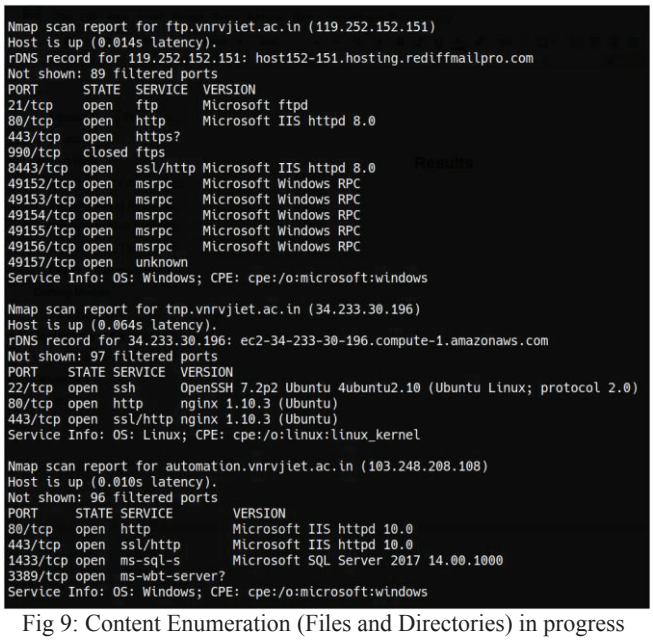

Fig 9: Content Enumeration (Files and Directories) in progress

VI. CONCLUSION

The various tasks of information gathering phase of a penetration test 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 computer. This

project is divided into various modules where each module does a specific job and at the end, the results of each module combinedly, is the output of this recon 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 or nano. The penetration tester's 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 relax!!Nmap Scan probes for all the open ports present on the target's websites, the services running on those ports and their versions too.

Information leakages, lack of security headers etc. Nuclei Scan finds low hanging vulnerabilities like using advanced wordlists, sometimes hidden content (files and directories) on the websites can be found.

VII. Future Scope

Ethical hacking and penetration testing is slowly increasing its demand in India, as a result of heists, the cyber-hacks are rapidly increasing in India due to vulnerabilities present in the websites. Indian websites are extremely prone to Cross - Site Script attacks, which can further lead to web defacement. While making a statement about Digital India, the honourable Prime Minister, Mr. Narendra Modi Criticized that the Indian Websites are easily prone to cyber-attacks from the enemy countries due to lack of cyber security experts in India. Ethical hacking is the field which is growing exponentially in India. Every IT organization, including the tech giants, which is dealing with users and their information with privacy needs an ethical hacker to protect their network and domain. Ethical hacking is not just the use to automated tools against web apps or servers, its deeply about how you can pre-test a project manually using your out box thinking, tools are just a way to perform your action.

References

[1] 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

[2] 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 Internet Technology and Secured Transactions (ICITST)

[3] Kristian Beckers, Sebastian Pape, Peter Schaab, Daniel Schosser, "Conference: International Conference on Trust and Privacy in Digital Business", August 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] Tae Hyun Kim, Douglas Reeves, "A survey of domain name system vulnerabilities and attacks", January 2020

[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 2013

[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 2015

[17] Sushmita Reddy Mamilla, "A Study of Penetration Testing Processes and Tools", May 2021.

[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 1999

[20] Vinitha K P, "Ethical Hacking", published at INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY, May 2018

[21] Bowman H.Miller, "Open Source Intelligence (OSINT): An Oxymoron?", December 2018

[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 2020

[23] Himanshu Singh, "Distributed Port Scanning Detection", 2009

[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 2011

[25] Mamta Bhavsar, Dr Priyanka Sharma, Manik Gokani, "Port Scanning using Nmap", published at International Journal of Engineering Development and Research, December 2017.

[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 2020.

[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 2020.

[29] V. S. Manvith, R. V. Saraswathi and R. Vasavi, "A Performance Comparison of Machine Learning Approaches on Intrusion Detection Dataset," *2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*, 2021, pp. 782-788, doi: 10.1109/ICICV50876.2021.9388502.

[30] 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.

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

[32] RV Saraswathi, LP Sree, K Anuradha, "Multi-stage key management scheme for cluster based WSN" International Journal of Communication Networks and Information Security, December 2018.

[33] 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.