Port Scanning: Techniques, Tools and Detection

By Ali Mirza

Abstract—In cyber security, it has long been understood that some cyber security procedures can be utilised for malicious purposes as well as cyber defence. The focus of this paper lies on one such procedure called port scanning. In port scanning, vulnerabilities are discovered through searching hosts and networks for open ports and closed ports. In this paper, a thorough analysis of modern port scanning tools, techniques and detection methods is presented. Finally, this paper will discuss and propose an argument for the implementation of one specific port scanning detection technique aside from any other techniques.

Keywords—Port Scanning, UDP, TCP, Bounce, Stealth, Idle, ports, packet, connection and firewalls.

I. Introduction

As the symbiosis of people and technology increases, the vulnerability of confidential data has also increased exponentially. Due to the increase in vulnerability, it is paramount that potential cyber attacks are detected early and that unauthorised users are refused access to sensitive data instantly. Port scanning allows both attackers and cyber security experts to gather information such as the presence of firewalls, open ports and the overall structure of a network. Port scanning attacks can cause organisations to breach the confidentiality, integrity and availability CIA triad and therefore, severely compromise the sensitive data of individuals. Existing work which is relevant to port scanning is examined in order to establish a strong link between existing research and my own arguments put forward in this paper. This paper contains a critical analysis of UDP scans, TCP scans and SYN scans etc as well as an in-depth discussion of the legal, social, ethical and professional issues (LSEPI) surrounding the whole port scanning procedure.

II. Related Work

Most recently, a lot of research has been carried out in order to solve problems related to port scanning. More specifically, the accuracy of port scanning techniques has been sacrificed in order to increase the speed of port scans. For instance, [1] argue that the Scanrand tool is fast mostly due to it compromising reliability for a faster scanning speed through changing the timeout period of connections. The detection of port scanning attacks is a popular trend in the research relating to port scanning especially where machine learning is implemented in order to aid the detection of port scanning attacks. A machine learning approach is proposed by [2] to detect port scanning attacks through applying traditional supervised learning algorithms such as Random Forest, AdaBoost, K-nearest neighbors and Linear SVM. On the other hand, the authors in [3] propose a more traditional approach for detection which is anomaly detection that is based on a mathematical model which then analyses the packets that have been captured. This mathematical model is then implemented through software which detected ACK scans within 19 seconds on average. Port scanning detection involves challenges such as false positives and false negatives, in order to combat this particular issue, [4] put forward an application of a custom fuzzy logic controller with Snort in order to combat false positives during the detection process. However, the major risk of this approach is that it has not been tested thoroughly enough to accurately conclude that it can in fact outperform Snort without the additional fuzzy logic controller.

Many different techniques to carry out port scanning rather than detecting it have also been proposed. A UDP scanning technique presented by [5] is claimed to be around 190 times faster than other traditional port scanning techniques. However, the technique proposed by them has a prerequisite that the connection must exist with a network that has no port address translation (PAT) or network address translation (NAT).

III. Techniques

A. UDP Scans

The user datagram protocol (UDP) is used to transmit time-sensitive data over the internet as it does not require a connection to be established prior to the transmission of any data. Due to the connectionless property of the UDP protocol, there are very few indications from which we can collect any information about the current status of a UDP port. A UDP port scan is initiated by either an attacker or a cyber security professional through dispatching a UDP packet to a port. Upon

receiving the UDP packet, the port will not respond with a reply if it is open and if the port is closed, it will reply with an error message alongside an ICMP packet. From an attacker's perspective, a UDP scan would be avoided due to it's unreliability from being easily blocked and as it is connectionless. The authors [6] state that the UDP scan is not detectable at the firewall level. However, a UDP scan would immediately fail as the packet would be dropped as even though in theory it would not be detected at a firewall level, it would be filtered out as most organisations block UDP by default because it has no mechanism to prevent congestion [20]. Therefore, UDP scans are not used very often by attackers to carry out reconnaissance prior to a major cyber attack and they do not pose a big threat to a network. Nmap is the most widely used tool for udp port scan due to it's ability to adjust the rate of a scan in order to avoid flooding a target network. A major issue surrounding UDP scans which is not noted by [7] is that if no ICMP response is received, it should not be assumed that the port is open if there is a lack of ICMP response as some systems do not send ICMP responses by default even if UDP ports are open. A consequence of this could be that network security administrators could waste valuable time and resources on trying to solve issues with open UDP ports even though they are not open in reality and the scanner is detecting false positives due to not receiving ICMP responses.

B. TCP Scans

The transmission control protocol (TCP), allows the exchange of messages between devices and programs over a network. Unlike UDP, TCP creates a connection between a source and a destination prior to any transmission of data [21]. SYN and ACK scans are two forms of TCP scans. The most widely implemented form of a TCP scan is SYN due to it's speed and ability to withstand firewalls. Another reason for the popularity of an SYN scan is that it can easily distinguish closed, filtered and open ports from one another. The most valuable trait of the SYN scan is that it does not require it's user to have much prior knowledge of TCP. Many port scanning tools such as Nmap, carry out SYN scans by delivering a TCP packet with an SYN flag to a port. The port then responds with a SYN flag and for an ACK scan, it will also return an ACK flag alongside the SYN flag [11].

In the domain of SYN detection, many different algorithms have been developed and discussed. The rank-based SYN detection algorithm implemented by [8] is a union of three different techniques used which are Entropy-based failed connections detection, Half open connections detection and vertical-horizontal port scans detection. The largest and the most obvious advantage of the rank-based SYN detection algorithm is that if attackers try to counter one of the three techniques included in this algorithm, their port scanning process would more than likely be detected by one of the two other techniques [22]. On the other hand, a machine learning approach for detection could also be applied here for SYN port scans. For instance, a logistic regression machine learning model was applied to a port scan attack dataset by [9], this model could perhaps be applied specifically to SYN port scan detection. The large distinction between the two approaches and accuracy results for detection between [9] and [8] suggests to me that the modern approach of using machine learning for scan detection is more feasible due to higher accuracy of "99.4%", which is the case when linear regression in particular is applied.

If SYN scan methods i.e. SYN packets are blocked by hosts, the form of TCP scans called ACK scans can be applied. ACK scans are carried out by port scanning tools by sending a TCP packet along with an ACK flag set to the port 80. Compared to the other port scanning techniques such as SYN, ACK does not explain to the user which ports are open or closed but it can only compute if certain ports are filtered or unfiltered with the use of "nmap -sP -PA" [10]. The ACK scan allows users to discover if the firewalls being used by the host are stateful or stateless. Firewalls that apply rules to filter packets are described as stateless and firewalls which can monitor the states of all of the traffic on a network are known as stateful. Due to not being able to connect to a remote host, the major drawback to using the ACK scan procedure is that no open ports can be identified by the user.

Similarly to SYN scans, ACK scans can be detected with the use of both supervised and unsupervised machine learning models. Sequential neural networks are implemented to aid port scan detection by [12] whereas, [14] opted for the supervised deep learning approach by implementing convolutional neural networks. The "MAWI 2017" dataset used by [14] to train their model is much larger compared to the dataset used by [12] however, the proportions of anomalies in the "MAWI 2017" dataset vary a lot and could lead to poor generalization or detection of port scanning. However, the model developed by [12] struggles to detect TCP scans but performs very well when other forms of port scans take place. The lower accuracy of the model developed by [12] at detecting TCP scans when compared to Wireshark could be due to the TCP used by them not having labels for each of the different forms of TCP scans such as SYN scans.

When carrying out port scans, attackers or network security administrator need to make the important decision on which ports to scan. This decision can be made easier with the implementation of the Naive Bayes classifier. The Naive Bayes classifier can aid in the prediction of potential ports that are likely to change state from closed to open or from open to closed [13]. The event probabilities of the ports and their corresponding states can be vital in saving time and resources as time is not wasted scanning ports which will always remain

closed and such ports can be explicitly excluded from scans when tools like Nmap are used.

C. Bounce Scans

FTP bounce scanning is an old technique that capitalises on the vulnerabilities in the FTP protocol. By design, the FTP protocol is extremely vulnerable to misuse as it allows connection to an FTP server so that files can then be delivered to a third-party server of an attacker's choice. An FTP bounce scan is carried out by delivering a file to the ports of a target host and the status of the port is indicated through an error message received [23]. However, this form of FTP scan is no longer feasible as vulnerabilities in FTP servers and most FTP servers will block any port commands by default [15]. The command below was most commonly used for FTP bounce scans.

$$nmap{-}b < targetFTPServer > \quad (1)$$

D. Idle Scans

Idle port scans are a subset of TCP scans where the user sends deceptive packets to a passive host. This form of scan allows the attacker to mask their digital location with the use of a zombie computer. Idle port scans are often labelled as stealth scans due to the fact that they can bypass intrusion detection systems.The advantage to this property of being able to use a zombie computer is not just anonymity but it also prevents the attacker from being blocked from a network directly as the zombie computer would be blocked instead. The zombie computer allows an attacker to scan each and every TCP port.

The results obtained by the attacker from this scan can indicate to the attacker where they can initiate an attack on a network. Reference [16] highlight an obstacle in idle scans which is the fact that idle scans make the assumption that the zombie computer being used by an attacker is idle. Despite the anonymity Idle scans can provide an attacker, the speed of the scan is dependent on the latency between the zombie computer and the attacker [24]. Therefore, this means that the attacker may have to compromise on speed to gain anonymity when using this form of port scanning. However, zombie computers are very readily available to attackers and if the latency between the attacker and the zombie computer is very low, the idle scan will be completed relatively quickly. One major issue surrounding idle scanning is that false positives for open ports can easily occur if the zombie delivers packets to any irrelevant hosts and it's IP id increases leading to a false positive open port being found by the tool used.

IV. ARGUMENT FOR A DETECTION TECHNIQUE

The most suitable approach for the detection of port scans is the application of a linear support vector machine (SVM) machine learning model. Compared to the other approaches for detection such as carrying out anomaly detection through implementing clustering algorithms (Fuzzy C-means and KNN etc), SVM outperforms every other classifier when classifying scans into more classes as demonstrated by [17] and [18]. The difference in accuracy alone is not the only justification for the SVM approach [25]. I believe that due to the limited datasets available for port scans, a SVM approach for detection is also favourable over deep learning as a deep learning model will struggle to generalize when trained on a small dataset and produce many false negatives/positives leading to the breach

Algorithm	Application	Training	Classification
Naïve Bayes	Classification	$O(nf)$	$O(f)$
Decision Tree	Classification/Regression	$O(n^2 f)$	$O(f)$
SVM (kernel based)	Classification/Regression	$O(n^2 f + n^3)$	$O(n_{SV} f)$
KNN	Classification/Regression	-	$O(nf)$
Linear Regression	Regression	$O(f^2 n + f^3)$	$O(f)$
Random Forest	Classification/Regression	$O(n^2 f n_{trees})$	$O(f n_{trees})$

Fig. 1. Complexity of Machine Learning algorithms [19]

of sensitive data. One downside to my argument for SVM application could be the higher amount of time it takes to train an SVM classifier compared to for instance, a Random Forest (RF) classifier [26]. I believe that an SVM classifier would take longer and require more resources due to its higher computational complexity.

V. LSEPI

Several legal, social, ethical and professional issues (LSEPI) arise when attackers carry out port scans and when port scanning is implemented by network security administrators. The legality of port scans is widely debated as some classify port scanning as illegal no matter what the circumstances are while others believe that under certain circumstances port scanning is legal. In the UK, the Computer Misuse Act 1990 regards unauthorised access to a computer as a criminal offence. As a result of the Computer Misuse Act, social, ethical and professional issues also arise when port scanning is carried out. The trust between an organisation and individuals would deteriorate if sensitive data of individuals was leaked after port scanning from attackers and the organisation was not aware of vulnerabilities as it could not carry out port scans on its own

network because of the Computer Misuse Act. An ethical issue that would also arise is that if an employer's network security officials for instance are carrying out port scanning on employee's computers, employees may feel that their privacy is being invaded despite the intention of the security officials being positive. Here, there are two sides to the ethics of and views towards port scanning as there is no distinct line to determine if a scan is malicious or administrative. I believe that this split in legal and ethical views is due to vague laws and because of the fact that a person's real intentions can never be known. The vague laws around port scanning present attackers with a grey area and an opportunity to steal confidential information and therefore, destroy the CIA triad organisations try to follow meticulously.

In order to combat these issues, I believe that organisations should have clear policies regarding port scanning permissions and a disciplinary policy that clearly states consequences for those who carry out port scanning within an organisation despite not being authorised to. On the other hand, outside of organisations, it is unfortunately very difficult to outline clear policies to combat these legal, social, ethical and professional issues. Governments or internet service providers (ISP) could possibly collaborate and produce less vague policies and laws to draw a distinguished line between circumstances where port scanning is legal and other circumstances where it is illegal.

VI. CONCLUSION

The rate at which cyber attacks occur is rising exponentially and port scanning is a preliminary step which allows such attacks to take place. In this report, I presented some different techniques and tools for carrying out port scans as well as detecting port scans. My personal view of utilising a support vector machine model for port scan detection is also presented. I also discussed and referred to different detection methods proposed by other authors. Properties of different forms of port scans were compared like their speed, reliability, usability and their popularity. I believe that much more progress is required in both the detection of port scans and the different laws and policies surrounding port scans to counteract the rapid rise in cyber crime. Finally, I believe that proactive efforts need to take place to educate members of society on port scans and how they could prevent themselves from becoming targets of scans which can lead to cyber attacks and sensitive data being breached.

REFERENCES

[1] Yuan, C., Du, J., Yue, M. and Ma, T. (2020) The design of large scale IP address and port scanning tool. Sensors (Basel, Switzerland) 20 (16), 1-12.

[2] Algaolahi, A. Q. M., Hasan, A. A., Sallam, A., Sharaf, A. M., Abdu, A. A. and Alqadi, A. A. (2021) Port-Scanning Attack Detection Using Supervised Machine Learning Classifiers. 2021 1st International Conference on Emerging Smart Technologies and Applications (eSmarTA). 10-12 Aug. 2021.

[3] M. Lefoane, I. Ghafir, S. Kabir, and I. Awan, "Machine Learning for Botnet Detection: An Optimized Feature Selection Approach". International Conference on Future Networks & Distributed Systems. Association for Computing Machinery, New York, NY, USA, 2021.

[4] Ananin, E. V., Nikishova, A. V. and Kozhevnikova, I. S. (2017) Port scanning detection based on anomalies. 2017 Dynamics of Systems, Mechanisms and Machines (Dynamics). 14-16 Nov. 2017.

[5] I. Ghafir, V. Prenosil, A. Alhejailan and M. Hammoudeh, "Social Engineering Attack Strategies and Defence Approaches." International Conference on Future Internet of Things and Cloud. Vienna, Austria, pp. 145-149, 2016.

[6] El-Hajj, W., Hajj, H., Trabelsi, Z. and Aloul, F. (2011) Updating snort with a customized controller to thwart port scanning DETECTING PORT SCANNING ATTACKS. Security and communication networks 4 (8), 807-814.

[7] Kumar, S. and Sudarsan, S. D. (2014) An Innovative UDP Port Scanning Technique. International Journal of Future Computer and Communication 3 (6), 381-384.

[8] Bhuyan, M. H., Bhattacharyya, D. K. and Kalita, J. K. (2011) Surveying Port Scans and Their Detection Methodologies. The Computer Journal 54 (10), 1565-1581.

[9] Sagatov, E. S., Mayhoub, S., Sukhov, A. M., Esposito, F. and Calyam, P. (2021) Proactive Detection for Countermeasures on Port Scanning based Attacks. 2021 17th International Conference on Network and Service Management (CNSM). 25-29 Oct. 2021.

[10] M. Lefoane, I. Ghafir, S. Kabir and I. U. Awan, "Multi-stage Attack Detection: Emerging Challenges for Wireless Networks," International Conference on Smart Applications, Communications and Networking, Palapye, Botswana, 2022.

[11] Aniello, L., Lodi, G. and Baldoni, R. Inter-domain stealthy port scan detection through complex event processing. 2011. ACM. https://go.exlibris.link/W7LqK5yH.

[12] Al-Haija, Q. A., Saleh, E. and Alnabhan, M. Detecting Port Scan Attacks Using Logistic Regression. 2021. IEEE. https://go.exlibris.link/pNWcBjf3.

[13] Pale, P. C. (2012) Nmap 6: Network Exploration and Security Auditing Cookbook : Network exploration and security auditing Cookbook. Birmingham, UNITED KINGDOM: Packt Publishing, Limited.

[14] I. Ghafir, M. Hammoudeh, V. Prenosil, L. Han and R. Hegarty, K. Rabie and F. J. Aparicio-Navarro, "Detection of Advanced Persistent Threat Using Machine-Learning Correlation Analysis," Future Generation Computer Systems, vol. 89, pp. 349-359, 2018.

[15] Goodrich, M. T. and Tamassia, R. (2014) Introduction to computer security. First Pearson new international edition. Harlow, Essex: Pearson.

[16] Hartpence, B. and Kwasinski, A. Combating TCP Port Scan Attacks Using Sequential Neural Networks. 2020. IEEE. https://go.exlibris.link/S6q3FSLH.

[17] I. Ghafir, V. Prenosil, M. Hammoudeh, T. Baker, S. Jabbar, S. Khalid and S. Jaf, "BotDet: A System for Real Time Botnet Command and Control Traffic Detection," IEEE Access, vol. 6, pp. 1-12, 2018.

[18] Malani, R., Putra, A. B. W. and Rifani, M. (2020) Implementation of the Naive Bayes Classifier Method for Potential Network Port Selection. International journal of computer network and information security 12 (2), 32-40.

[19] A. Abdulhamid, S. Kabir, I. Ghafir and C. Lei, "Dependability of the Internet of Things: Current Status and Challenges," International Conference on Electrical, Computer, Communications and Mechatronics Engineering, Maldives, Maldives, 2022.

[20] Chockwanich, N. and Visoottiviseth, V. Intrusion Detection by Deep Learning with TensorFlow. 2019. Vol. 2019-. Global IT Research Institute (GIRI). https://go.exlibris.link/m3yXFRQS.

[21] Baloch, R. (2015) Ethical Hacking and Penetration Testing Guide. 1 edition. Philadelphia, PA: CRC Press.

[22] Zhang, X., Knockel, J. and Crandall, J. R. ONIS: Inferring TCP/IPbased Trust Relationships Completely Off-Path. 2018. Vol. 2018-. IEEE. https://go.exlibris.link/hmfmw9j2.

[23] Aamir, M., Hussain Rizvi, S. S., Hashmani, M. A., Zubair, M. and Ahmed, J. (2021) Machine Learning Classification of Port Scanning and DDoS Attacks: A Comparative Analysis. Mehran University Research Journal of Engineering and Technology 40 (1), 215-229.

[24] Rustam, Z. and Ariantari, N. P. A. (2018) Comparison between support vector machine and fuzzy Kernel C-Means as classifiers for intrusion detection system using chi-square feature selection.

[25] Younes, H., Alameh, M., Ibrahim, A., Rizk, M. and Valle, M. (2020) Efficient Algorithms for Embedded Tactile Data Processing. 113-138.

[26] S. Eltanani and I. Ghafir. "Coverage Optimisation for Aerial Wireless Networks." 2020 14th International Conference on Innovations in Information Technology (IIT). IEEE, 2020.