International Journal of Engineering and Information Systems (IJEAIS)
ISSN: 2643-640X
Vol. 5 Issue 8, August - 2021, Pages: 35-39

A Review on Internet intrusion attack Dataset

Anwar Mira

University of Babylon
anwar.jaafar@uobabylon.edu.iq

Abstract: *One of the most important goals of SDN (Software-defined networking)technology is the central control and managements of Internet networks, which gives the ability to programmable control of network devices at the basic data level that restores network traffic, and this control has expanded to include network data security management and detection of intrusions that may occur on the Internet, and in this research We will address the most important data of the Internet networks proposed to work on them for the purpose of detecting the types of spoofing that the networks may be exposed to. We have dealt with this data according to the type of intrusion that the network may be exposed to.*

1. Introduction

The attacks faced by internet networks have become a major challenge to network security, where intrusion on network must be detected accurately ,in addition to the presence of many intruders who are waiting for the opportunity to launch various attacks on the network and as a result of the rapid developments in the fields of internet and communications, which led to an increase in the size of communications networks ,Therefore, increasing the data .Intrusion detection system(IDS) has become one of the very important tools that scan the network traffic in order to prevent the network from any kind of potential intrusion.

Recently, machine learning[1] and deep[2] learning methods have been adopted as proposed solutions to detect potential intrusion on the network in an effective manner and with the aim of improving the accuracy of intruder detecting, where training is done on data representing penetration cases in order to be able to know similar cases of intrusion[3].

1. 1 What is the Network Intrusion?

Network intrusion has been defined as any unauthorized activity on the internet with the aim of stealing important network resources, therefore the security of the network and its data will be at risk [4].

To reduce the risk of network intrusion, the proactive detection of network intrusion will help in this, and as an essential step there must be a comprehensive understanding of the mechanism of network intrusion and how to implement intrusion and network attack systems and methods of cover up followed.

2. Types of internet intrusion:

There are many network intrusion attack techniques [5], we will discuss the most common of them, given the difficulty of identifying all the anomalies that may indicate the occurrence of a network intrusion due to the large expansion of the activities that occur constantly on the network [6].

2.1 Buffer Overflow

The sequential section in memory called Buffer which is considered to allocate any kind of variables such as character, integer, string or array of them.

Buffer overflow occurs when the specified buffer is exceeded, where the entire buffer is exploited to infringe on additional spaces, which may lead to system failure or overwriting to other areas of memory and damage the original data, or it may give an opportunity for the attacker to add malicious code of procedures and commands [9].

2.1.1 Why does it occur?

Most of buffer overflow problems are caused by coding error, a common application that may support this kind of data overflow is C/C^{++} , which does not have built-in protection against buffer overflows. Consequently, to be often targets of buffer overflow attacks [10].

2.1.2 Dataset

CDX 2009 dataset [11]which focused on buffer overflow attacks found on SNORT intrusion preventing log were performed only on two services, postfix email and Apache web server, the data contains 2 types of labels,the first is **label_2** ,recognize whether an actual record represent network buffer overflow attack or not,while the next is **label_poly** which is composed of 2 parts one to detect legitimate and malicious communications and the second to represent the type of communication on particular network service. Software Assurance Reference Dataset (SARD) had provided a set of known security flaw includes Buffer Overflow problems [12].

2.1.3 How to avoid?

Programmers of C/C++ applications are typically advised to avoid of using standard library functions that are not bounds-checked, such as gets, scanf and strcpy.

Using automatic protection at the language level is one way to avoid buffer overflow, in addition to bounds-checking enforced at run-time, by making automatically checking for the data written to a buffer is within acceptable boundaries[13].

2.2 Denial of service (DOS)

It represents the interruption of access to network resources by legitimate users of the network under the influence of attacker's control of the network service, where attackers flood the resources of victim's web servers, systems or networks with so many requests that it becomes impossible for a service requester to access them.

System restarting will usually recover the crash, but flooding attacks are more difficult to recover from, where the denial of service attack traffic from a large number of sources so that the weaknesses in the network protocols and how they deal with the network traffic are taken advantage of by the attacker confusing the network service by sending many packets with different IP addresses, which leads to weakening or stopping the network jobs[14].

2.2.1 How does it occur?

The most common OSI targets attacker include Layer 3 (network), Layer 4 (transport), Layer 6 (presentation) and Layer 7 (application), where attackers target one or more of these layers[15].

Using User Datagram Protocol (UDP) packets is one common way of attacking the OSI layers. UDP speeds transmission transferring data before the receiving party sends its agreement. Another common attack method is SYN (synchronization) packet attacks. In these attacks, packets are sent to all open ports on a server, using spoofed, or fake, IP addresses. UDP and SYN attacks typically target OSI Layers 3 and 4.

Attacks on Layers 6 and 7 now commonly used, where handshakes launched from internet of things (IOT) devices. These attacks can be difficult to identify and pre-empt because IOT devices are everywhere and each is a discrete intelligent client [16].

You can check if DOS is occurring by notice the following:

1. Degrading in network performance that is particularly noticeable when trying to access a website or open files on the network

2. Access a websites is enabling.

3. Unusual spam email.

2.2.2 Dataset

A lot of datasets had presented to deal with denial of services attack the recent one were generated using mininet emulator [17] which recognised between benign and malicious traffic on TCP,UDP and ICMP.it has 23 features some were extracted from switch and others were calculated ,this data consists of 104,345 rows of data. CIC-DDoS2019 Dataset are resembles the true real-world data, for this dataset, the abstract behaviour of 25 users was constructed based on the HTTP, HTTPS, FTP, SSH, and email protocols with labelled flows based on the time stamp, source, and destination IPs, source and destination ports, protocols and attack[18]

2.2.3 How to avoid?

Contact internet service provider (ISP) to determine whether the slow performance or other indications are from an attack or some other factor. The ISP can reroute the malicious traffic to counter the attack. ISP can also use load balancers to mitigate the severity of the attack[19].

Major steps have to taken to reduce the likelihood that an attacker will use your computer:

● Install and maintain anti-virus software.

● Install a firewall, and filter.

● updating the operating system.

2.3 IP spoofing

It is a technique often used by bad actors to invoke DDOS attacks against a target device or the surrounding infrastructure, by creating Internet Protocol (IP) packets in order to modify source address either to hide the identity of the sender, to impersonate another computer system, or both[20].

2.3.1 How does it occur?

The primary way in which networked computers and other devices communicate, and can constitute the basis of the modern internet is by Sending and receiving IP packets [21].

In general, all IP packets consist of a head and a body, where the header contains important routing information, including the source address or the address of the sender of the packet. A basic security that may DDOS attacks take advantage of it becomes even more difficult for cyber security to track the perpetrator of the attack when the IP identity is hidden [22].

2.3.2 Dataset

dataset of BGP announcements had presented by O.Fonseca[23] which includes four features (announcement configuration, logs with the timestamps , route measurements , tcpdump information)the data had collected poisoning and not poisoning examples, on the other hand data model were provided by CAIDA spoofer data API [24]collected by Spoofer service which is a public data interface, the data model contains a unique integer identifier for each session with a timestamp and parameters for read access, the IPv4 client address, the IPv6 client address, the country, NAT4 address, NAT6 address, and information about private and routed addresses. The API allows the user to query based on date, session id, or by Autonomous System Number (ASN). The API returns a paginated list of measurement sessions.

2.3.3 How to avoid?

A very common defence against spoofing is ingress filtering, which is a form of packet filtering usually applied on a network edge device which examines incoming IP packets and looks at their source headers. If the source headers on those packets don't match their origin or they otherwise look fishy, the packets are rejected, while some networks will also apply egress filtering, which looks at IP packets exiting the network, ensuring that those packets have legitimate source headers to prevent someone within the network from launching an outbound malicious attack using IP spoofing [25].

3. Conclusion

There are many intrusion attacks that network may faced, in order to be able to choose the appropriate IDS (Intrusion Detection System), we must build a sufficient background on the circumstances surrounding the attack and how it occurs. Given the large expansion and spread of the internet, the methods of attack and network intrusion are subject to change and development constantly, which may lead to the difficulty of detecting the intrusion and thus protection from it.

References

[1] L. Dhanabal, S.P. Shantharajah Dr.,**"A Study on NSL-KDD Dataset for Intrusion Detection System Based on Classification Algorithms"**
International Journal of Advanced Research in Computer and Communication Engineering, 4 (6) (2015)June.

[2] Mohammad Samar Ansari, Vaclav Bartos, Brian Lee,'Shallow and Deep Learning Approaches for Network Intrusion Alert Prediction',Procedia Computer Science, Volume 171,2020,Pages 644-653,ISSN 1877-0509,

[3] C. So-In, N. Mongkonchai, P. Aimtongkham, K. Wijitsopon, and K. Rujirakul, "An evaluation of data mining classification models for network intrusion detection," in 2014 fourth international conference on digital information and communication technology and its applications (DICTAP), 2014, pp. 90–94

[4]Ghansela Siddharth**,Network Security: Attacks, Tools and Techniques",**ijarcsse, Volume 3 (Issue 6) (2013 June)

[5]Mohamed A.B., Idris N.B., Shanmugum B. (2012) A Brief Introduction to Intrusion Detection System. In: Ponnambalam S.G., Parkkinen J., Ramanathan K.C. (eds) Trends in Intelligent Robotics, Automation, and Manufacturing. IRAM 2012. Communications in Computer and Information Science, vol 330. Springer, Berlin, Heidelberg.

[6] Liao H-J, Lin C-HR, Lin Y-C, Tung K-Y (2013b) Intrusion detection system: a comprehensive review. J Netw Comput Appl 36(1):16–24

[8] Jiadong Ren, Zhangqi Zheng, Qian Liu, Zhiyao Wei, Huaizhi Yan, "A Buffer Overflow Prediction Approach Based on Software Metrics and Machine Learning", *Security and Communication Networks*, vol. 2019, Article ID 8391425, 13 pages, 2019.
[9] Keromytis A.D. (2011) Buffer Overflow Attacks. In: van Tilborg H.C.A., Jajodia S. (eds) Encyclopedia of Cryptography and Security. Springer, Boston, MA. https://doi.org/10.1007/978-1-4419-5906-5_502

[10] C Cowan, C Pu, D Maier**,** J Walpole, P Bakke, S Beattie,' Stackguard: automatic adaptive detection and prevention of buffer-overflow attacks',
USENIX security symposium, 1998
[11] Brno University Security LABoratory research group (BUSLAB) ,available on : ASNM Datasets (vutbr.cz)
[12] National Institutes of standards and Technology , Available on:Software Assurance Reference Dataset (nist.gov)

[13] Sahel Alouneh, Mazen Kharbutli, Rana AlQurem,'Stack Memory Buffer Overflow Protection based on Duplication and Randomization',Procedia Computer Science,Volume 21,2013,Pages 250-256,ISSN 1877-0509,

[14] Mahjabin T, Xiao Y, Sun G, Jiang W. A survey of distributed denial-of-service attack, prevention, and mitigation techniques. *International Journal of Distributed Sensor Networks*. December 2017.

[15] Raiyn J (2014) A survey of cyber attack detection strategies. International Journal of Security and Its Applications 8(1):247–256
Sadotra P, Sharma C (2016) A survey: intelligent intrusion detection system in computer security. Int J Comput Appl 151(3):18–22

[16] Mahjabin T, Xiao Y, Sun G, Jiang W. A survey of distributed denial-of-service attack, prevention, and mitigation techniques. *International Journal of Distributed Sensor Networks*. December 2017.

[17] Ahuja, Nisha; Singal, Gaurav; Mukhopadhyay, Debajyoti (2020), "DDOS attack SDN Dataset", Mendeley Data, V1, doi: 10.17632/jxpfjc64kr.1

[18] University of New Brunswick, "DDoS Evaluation Dataset (CICDDoS2019)," unb.ca, 2019. [Online]. Available: https://www.unb.ca/cic/datasets/ddos-2019.html.

[19] Al-Ani AK, Anbar M, Manickam S, Al-Ani A (2019) DAD-match; Security technique to prevent denial of service attack on duplicate address detection process in IPv6 link-local network. PLoS ONE 14(4): e0214518.

[20] Kováčik M., Kajan M., Žádník M. (2013) Detecting IP Spoofing by Modelling History of IP Address Entry Points. In: Doyen G., Waldburger M., Čeleda P., Sperotto A., Stiller B. (eds) Emerging Management Mechanisms for the Future Internet. AIMS 2013. Lecture Notes in Computer Science, vol 7943. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-38998-6_9

[21] P. Ramesh Babu , D.Lalitha Bhaskari , CH.Satyanarayana ,'A Comprehensive Analysis of Spoofing', (IJACSA) International Journal of Advanced Computer Science and Applications,
Vol. 1, No.6, December 2010

[22] Seo JW, Lee SJ. A study on efficient detection of network-based IP spoofing DDoS and malware-infected Systems. *Springerplus*. 2016;5(1):1878. Published 2016 Oct 26. doi:10.1186/s40064-016-3569-3

[23] O. Fonseca *et al*., "Identifying Networks Vulnerable to IP Spoofing," in *IEEE Transactions on Network and Service Management*, doi: 10.1109/TNSM.2021.3061486.

[24] CAIDA project,Available on : Spoofer - CAIDA

[25] Gunjan Agrawal, 2019, Detection and Prevention of ARP-Spoofing Attacks, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 08, Issue 10 (October 2019),