Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/43/20230841

Analysis of a new firewall constructed on Pfsense with Snort to defend against common internet intrusions

Wang Buqing

College of computer and information engineering, Nanjing Tech University, Nanjing, 211816, China

15953454928@163.com

Abstract. This paper aims to design a new firewall based on the detection of common forms of network attacks in the current Internet environment using Snort in combination with the Pfsense network security platform. Firstly, the functions of Snort and Pfsense are introduced, and the shortcomings of traditional firewalls in the current network environment are analyzed. Then analyze the common port scanning attacks, DOS attacks and algorithm complexity attacks, and by means of reviewing the literature, demonstrate that it adopts the network proxy, CNN-BiLSTM intrusion detection model, IBM algorithm, VLDC algorithm and other means to specifically detect common network attacks, so as to achieve the function of specifically defending against network attacks, and to improve the operation efficiency and detection accuracy of the firewall. accuracy of the firewall, etc.

Keywords: Firewall, Snort, Pfsense, Network Security, Network Intrusion.

1. Introduction

The rapid development of the Internet has brought great convenience to human beings, even changed the way of life of human beings and promoted the development of various emerging technologies. However, at the same time, the connectivity of the Internet has also brought about corresponding security problems, and the rapid development of the Internet has also brought new problems and challenges to network security.

In the traditional sense, the use of firewalls to defend against intrusion is the preferred operation of network security, such as the Windows system that comes with the defender, which can reduce the occurrence of network attacks. However, with the development of information technology, the forms and principles of attacks continue to innovate, the traditional firewall against network attacks is not advanced enough, and there are still a large number of network security issues that continue to be addressed. In 2020, China's National Center for Internet Emergency Response (CNCERT) released the latest China Internet Network Security Report 2020[1], which shows that in 2020 alone, the number of computer malware samples captured throughout the year exceeded 42.98 million, and the average daily propagation of these malware programs reached more than 4.82 million, involving more than 348,000 malware code families, and new These malicious programs were disseminated more than 4.82 million times per day, involving more than 348,000 malicious code families, with 235 new families added. It can be seen that the rapid expansion of network scale provides invaders with invasion opportunities, and

DOI: 10.54254/2755-2721/43/20230841

they invade other people's systems for various purposes, steal important information, and damage the network. Therefore, the research and promotion of network security technology is very important.

This paper discusses the feasibility of a novel firewall based on the commonly used open-source router firewall software Pfsense which adds the open-source intrusion detection system Snort on top of it to cope with various network attacks. Meanwhile, the new firewall uses a CNN-BiLSTM intrusion detection model developed based on machine learning to categorize the invasive behavior, which facilitates the setting of different intrusion detection rules for different types of attacks in the Snort system, and enhances the ability of the firewall to resist intrusions, in order to make up for the deficiencies of the traditional firewall in today's network environment.

2. Software Environment Introduction

2.1. Snort Intrusion Detection System Introduction

Snort Intrusion Detection System, is an intrusion detection system based on pattern matching of anomalous behavior, designed and implemented by Marty Rosech in 1998, and is currently one of the most active open-source network intrusion detection projects. In use, the user configures the Snort file to set custom detection ranges according to the needs of the user to match the network data and rules to determine whether it contains intrusion behavior. The five modules are the packet capture module, packet decoding module, preprocessing module, detection engine module, and alarm output module.

The Detection Engine module is the core part of Snort. The Detection Engine module matches the processed packets with each packet through the pre-defined rule files. If the match is successful, the alarm will be notified.

2.2. Rule Construction of Snort

Snort's rule construction consists of two sub-modules, the rule header and the rule options. The rule header can be partially filtered to effectively reduce the time consumed by the rule body to process the data, and its content and structure are shown in Table 1.

Table 1. Content and Structure of Rule Construction.

Operation	Protocol	Destination address	Destination port	Direction	Source address	Destination address

The second module is the rule options, the content of which exists in pairs, and each pair of options is edited with each other, only if the conditions in the rule options are satisfied, the corresponding action will be executed.

2.3. Pfsense Introduction

PfSense is a FreeBSD-based, open source open-source routing + firewall software customized for firewall and router functionality. It is installed on the computer as a firewall and router in the network exist, it can be configured through the WEB page, upgraded and managed without the need for users to have the underlying knowledge of FreeBSD, with easy-to-start, easy-to-configure user-oriented features. However, with its comprehensive intrusion detection and protection rules on flexibility, compared with professional protection software, there are still some shortcomings [2].

3. Firewall Construction Direction

3.1. Shortcomings of Traditional Firewalls

Traditional firewalls adopt static defense policies, most of which are set based on historical intrusion data [3], and cannot quickly respond to real-time network attacks or new types of attacks, and increase the maintenance and configuration workload of network engineers.

Increase the detection speed of intrusion detection systems to handle large numbers of packets. It is necessary to match the speed of detection with the speed of data transmission. If it cannot keep up with

DOI: 10.54254/2755-2721/43/20230841

the data transmission, it will produce many missed reports, and unstable intrusion detection equipment is meaningless. At the same time, after the interception of the packet, the analysis of the packet will take a lot of resources and time, the current Ethernet, industrial control networks are very sensitive to the processing speed, and the speed of intrusion detection has lagged behind the 100-megabit network.

Traditional firewalls do not have the ability to monitor the security of the internal network and operate using standard network protocols, so they cannot effectively defend against network attacks that take advantage of protocol flaws, such as DOS or DDOS.

3.2. Advantages of the New Firewall Based on Pfsense with Snort

Snort can be used as a functional plug-in to run directly on Pfsense and provide the corresponding services normally. Therefore, the Pfsense firewall equipped with Snort plug-in can not only retain the advantages of Pfsense such as easy to operate and maintain, but also have the multi-platform operation of Snort, real-time traffic analysis and other network intrusion detection functions, and play a Snort update timely and the advantages of the flexibility of the rule set, which makes the new firewall (Snort and Pfsense) This makes the new firewall (Snort and Pfsense) able to withstand many types of network attacks as described below.

4. Analysis of Common Network Intrusion Principle

4.1. Port Scanning Attacks

The port scanning attack itself does not do much harm to the network [4], but before a cyber attacker conducts an attack, he first utilizes various port scanning tools to collect detailed information about the target host or network and then discovers the vulnerabilities or fragile points of the target system. Therefore, port scanning attacks are often regarded as a prelude to intrusion attacks. Accordingly, if the port scanning detection technology can be used, in the scanning phase of the attacker can distinguish the malicious scanning and normal access in order to prevent the attack so as to avoid or mitigate the damage caused by the attack to a certain extent.

4.2. DOS

DOS attack is called a denial-of-service attack, which refers to the attacker through special means to make the target machine cannot normally provide services. For example, forcibly filling the server's buffer or even exhausting the target machine's network bandwidth, so that the server cannot receive new requests, resulting in the stop of normal services or even causing the host to die [5].

DOS attacks converge in intrusion behavior, and machine learning modules can be used to learn a large number of DOS intrusion behaviors as a way to detect DOS intrusion.

4.3. Algorithm Complexity Attack

An algorithm complexity attack is aimed at the core detection algorithm of network intrusion detection and completes the attack on the system, so that the pattern-matching algorithm of the intrusion detection system reaches the worst time complexity of the algorithm, which in turn drastically affects the matching time, resulting in the algorithm cannot be successfully detected packets containing intrusion behavior.

5. Algorithm Analysis for Common Network Intrusions

5.1. Network Agent

In order to enhance the active prevention of intrusion detection technology, agent technology is usually utilized as a plug-in for active defence, the agent is equivalent to a daemon running in the background, once Snort detects an intrusion, the band is triggered to notify the defense equipment to carry out the relevant interception as well as the corresponding processing. Usually use Guarding and Snortsam agents. The following describes the active defense steps of the intrusion prevention system developed by the combination of Guarding and Snortsam, as well as its problems: When an intruder conducts a

DOI: 10.54254/2755-2721/43/20230841

malicious scan of a server for information collection purposes, Wireshark is driven to capture packets for analysis of the scanning process, and Nmap scanning sends UDP probe reports and generates matching Snort rules accordingly, and adds them to the Snort rule base. Snort opens the intrusion detection according to the rules, performs defense functions, and prevents intruders from accessing the server by cutting off network connections. By cutting off the network connection, Snort can prevent further malicious behavior by organizing the access of intruders.

5.2. CNN-BiLSTM Intrusion Detection Modeling

In traditional firewalls, the misuse-based anomaly detection system relies too much on known rule bases, and when the number of rule bases is more complete, the false alarm rate of detection will be lower. However, not all attacks are known and contain features, and the improvement of computer network technology also promotes the updating of attacks. When the attack behavior is updated but the rule base is not updated for this type of attack, a serious under-reporting event occurs.

In the face of unknown attacks, the Snort intrusion detection system based on misuse technology is slightly insufficient. To address the shortcomings of unknown attacks, the CNN-BiLSTM intrusion detection model is used to increase the ability of the Snort intrusion detection system to predict unknown attacks by investigating machine learning methods. Increased the ability of the module Snort Intrusion Detection System. Two common anomaly detection methods are described below:

Anomaly detection method based on behavioral whitelist [6] includes a data collection module that matches the whitelisted users and their corresponding permissions and assigns the permissions to an authentication module that detects the legitimacy of the users through identification and access management; a data detection module that compares the collected data with the actual accessed users, the application processes, and the network behaviors of the application processes, and when illegitimate behaviors are detected, it will issue a warning and notify the administrator, then generates logs.

A method for detecting abnormal behavior based on time-correlated baselines [7] includes the following modules: A data acquisition module for obtaining data on network traffic, user behavior, and process/control behavior; a data processing module that samples a series of similar time periods of collected data to obtain a baseline behavior over an observable period and calculates an expected baseline based on the actual situation; and a data comparison module that compares the actual data with the expected baseline, and in case of anomalous data, issues a warning and generates logs.

Using the CNN-BiLSTM intrusion detection model can get rid of the dependence on the known intrusion behavior library and rule base to classify the invasive behavior, which facilitates the setting of different intrusion detection rules for different types of attacks in the Snort system, and cope with a variety of new types of cyber-attacks in the Internet at the present time, e.g., when a large number of high-frequency access requests are recognized, the behavior is the same as the precursor of invasion of DOS attack For example, when a large number of high-frequency access requests are recognized, the behavior is the same as the precursor of DOS attack invasion, which can be used as DOS attack detection to enhance the ability of the firewall to resist invasion.

5.3. IBM Algorithm

In past Snort-based systems developed using the BM algorithm, the traditional string comparison algorithm performs a one-time search for and matching process. The algorithm runs very fast if the number of attack patterns in the network is relatively small and the attacks are homogeneous. However, the content of the packet captured by the network intrusion detection system can lead to an increase in some of the matching rules, that is to say, the number of matching patterns increases, and every time the comparison algorithm is re-run will perform matching operations, reducing the efficiency of the system.

The IBM (Improved BM) algorithm [8] was proposed by Zhenxiong Zhang et al. from the China University of Weights and Measures, this algorithm takes into account the bad character jumping rules, but also the text string characters that are postpositioned when the match fails to be made, and finally obtains a distance to the right that is as large as possible, and at the same time, it can also not miss all the possible matching fields.

DOI: 10.54254/2755-2721/43/20230841

The IBM algorithm has two main considerations for calculating the maximum distance to the right. First, when a match failure occurs, view the next text string character immediately to the right of the match string, determine whether it exists in the pattern string, if not, then continue to scan to the right; secondly, if it exists, the character is sure to participate in the next step of the text string to match, and at the same time, locate the position of the character in the pattern string, and the text string is placed in the same position for comparison. It is concluded that the algorithm improves the matching efficiency in two ways, by ignoring the characters that do not exist in the pattern string, reducing the number of comparisons with the pattern string, reducing the number of matches, and improving the effectiveness of each match, and at the same time, the characters that can be effectively located in the matched characters is also a means of improving efficiency [9].

The new algorithm has two phases, a preprocessing phase and a matching phase. In the preprocessing phase, an array is created “Found Times”, the size of the array is 256, representing a collection of 256 characters, using this collection to mark the characters in the text string, when the value in the array is 0, it represents that the character in the text string has never appeared in the pattern string. Create an array “next” which is used to store the previous character of the pattern string that is aligned to the rightmost position of the character in the text string when the character appears in the pattern string [10].

The main judgment in the algorithm is divided into two parts, one is to determine whether a mismatch character appears in the pattern string, and the other is, whether a combination character appears in the pattern string in a combined form. In addition, there is a repetition process in the two matching processes, so the time consumed on the judgment can be well reduced. In the optimal case, the time complexity of the algorithm depends on the similarity between the text string and the pattern string. That is, the algorithm can get maximum computational efficiency when the similarity between the text string and pattern string is low. When the degree of similarity between the text string and the pattern string is high, good matching efficiency can also be achieved due to the low probability of the simultaneous existence of double characters. The main function is as Figure 1:

```
1.  if(nocase)
2.  {
3.  success = newSearch(base_ptr, depth, pmd->pattern_buf, pmd->pattern_size);
4.  }
```

Figure 1. The code of the Main Function.

The IBM algorithm reduces the number of matches as well as the number of moves by considering all aspects of the text as well as by maximizing the shift to the right, the algorithm is much faster than the BM algorithm's brute-force matching, and the statistics of the time required are shown in Figure 2, which improves the operational efficiency of the firewall, enhances the speed of intrusion detection, and positively affects the defense against cyber-attacks.

DOI: 10.54254/2755-2721/43/20230841

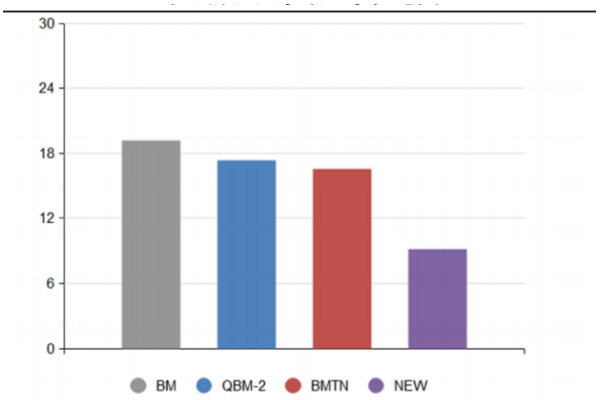

Figure 2. Snort System with Different Algorithms and Time Required to Measure Data [10].

5.4. VLDC Algorithm

In response to the algorithmic complexity attack presented above, the VLDC algorithm, proposed by the literature [11] in 2011, VLDC uses a simple and fast method to realize the problem of multi-pattern matching, and then the patterns can be separated from each other by an arbitrary distance in multi-pattern matching, given the alphabet ∑, the set of patterns P is p1@p2... @pm, pi is a combination of any characters in the alphabet and is referred to as a keyword, @ is not in the alphabet ∑, @ is called a length-independent pattern matching symbol, only if the input t=u0p1u1p2.... .um-1pmum, and ui is also composed of any character in the alphabet, then the pattern set P is considered to be matched successfully. [5] The VLDC multi-pattern matching algorithm can be regarded as a special regular expression matching algorithm, and the VLDC multi-pattern set can be called a set of regular expressions. Efficient multi-pattern matching is a core function in deep packet inspection, intrusion detection and virus scanning.

Therefore, for the Snort network intrusion detection system, the new VLDC algorithm has better applicability. Because Snort's rules are divided into multiple rules, each of which is independent of the other, and there may be multiple repeated pattern substrings. Due to the characteristics of the VLDC algorithm, it can cope with the situation of duplicate rules and optimize the construction of matching automata. The VLDC algorithm can cope with the duplicated rules optimize the construction of matching automata, and avoid redundant comparisons in the process of matching by pooling, further optimizing the matching time. The matching time is further optimized.

After applying the VLDC algorithm in the system, the matching and detection of all the attack packets can be well accomplished, and a 100% packet detection rate is achieved. Compared with the original system with Snort-VLDC algorithm as the core algorithm, the VLDC algorithm-centered Snort system has better performance, and it can detect and process all the packets in the face of the hybrid attack, ensuring that all the packets that flow through the network are detected.

6. Conclusion

By analyzing the detection of common network attacks based on Snort's multiple intrusion detection algorithms in the application of Pfsense firewall, it is demonstrated that the new firewall (Pfsense with Snort) can provide effective and specific intrusion detection and intrusion defence against new complex network attacks under the complex situation of the computer network nowadays. Limited to space and personal ability reasons, this paper only selected part of the common types of attacks and corresponding algorithms to analyze and demonstrate.

In the future development of new firewall construction, it should tend to use new technologies such as big data, neural networks, data mining, etc. to develop smarter matching algorithms, and jointly establish a larger linkage intrusion detection system in order to face a larger network environment and a more complex system structure, and to realize that after the emergence of a certain new type of network

DOI: 10.54254/2755-2721/43/20230841

attack, all the intrusion detection mechanisms under the linkage system can update the detection library at the same time to deal with the A new type of intrusion, work together to reduce false alarms.

References

[1] Li Xiangning, Lu Wei, Yan Hanbing, et al. China Internet Network Security Report 2021 [R]. Beijing:National Computer Network Emergency Response Technology Processing Coordination Center,2021.7.

[2] Dai Shan Guo. (2022). Network intrusion detection and protection system based on Pfsense+Snort. Network Security and Informatization (09), 123-126.

[3] Wang Daxian, Zhang Jishan, and Yu jiujiu. 2020. Research on intelligent Firewall for network security. In Proceedings of the 2020 2nd International Conference on Robotics, Intelligent Control and Artificial Intelligence (RICAI '20). Association for Computing Machinery, New York, NY, USA, 255–258.

[4] Wang, Longye & Roger. (2016). A security detection method for internet port scanning attacks. Information Security and Technology (02), 44-45+64.

[5] Huang, Hsiao Nan.Implementation and Detection of Denial of Service Attacks against Snort (Master's thesis, Jilin University).

[6] Ye Peng,Zhang Zhenxiong. Anomaly detection method, apparatus, and electronic device based on behavioral whitelist [P]. China Patent: 2018111809412, 2018-10-10

[7] Zhenxiong Zhang, Hao Zhang. A method, device and apparatus for detecting abnormal behavior based on a time-dependent baseline [P]. China Patent: 2018109739816, 2018-08-24.

[8] Zhang Zhenxiong. (2020). Design and Implementation of Snort-based Intrusion Detection System (Master's Thesis, China University of Weights and Measures).

[9] K. Dinakaran,D. Rajalakshmi,P. Valarmathie. Efficient pattern matching for uncertain time series data with optimal sampling and dimensionality reduction[J]. Microprocessors and Microsystems,2020,

[10] Jiahui Li. (2021). Optimization and Implementation of Snort Intrusion Detection Method (Master's thesis, Northeast Normal University).

[11] Zhang M, Zhang Y, Hu L. A faster algorithm for matching a set of patterns with variable length don't cares[J]. Information Processing Letters, 2010, 110(6):216-220.