International Journal of New Developments in Engineering and Society
ISSN 2522-3488 Vol. 4, Issue 1: 33-40, DOI: 10.25236/IJNDES.040106

A Fast-Single Pattern Matching Algorithm of Next Generation Firewall

Wei Anlei [1], Wang Zhaoshun[2], Lv Shuwang[3]

1 China Center for Information Industry Development, Beijing, 100044 China
2 School of Computer & Communication Engineering, University of Science and Technology Beijing, Beijing 100083 China
3 State Key Laboratory of Information Security, Chinese Academy of Science, Beijing, 100864 China

ABSTRACT. *Based on the existing single pattern matching algorithm, we propose a fast-single pattern matching algorithm. A next generation firewall system based on DPI (Deep Packet Inspection) technology is designed and implemented where the proposed pattern matching algorithm is applied. Tests with respect to different performance indicators have demonstrated that the performance of the next generation firewall has superior performance to the unimproved one.*

KEYWORDS: *DPI technology, next generation firewall, pattern matching algorithm*

1. Introduction

As the computer network has made its applications in every industry, the security of network becomes increasingly prominent. Firewall locating on the border of trusted network and untrusted network, is the gateway of data exchange between the two networks. Since its performance, availability, reliability, security and other indicators largely determine the transmission efficiency and security of the network, building a well-behaved firewall system is an effective method to protect network security.

At present, most firewalls are built on the state detection technology. They function on the network layer detecting and filtering data flow on the basis of packet header information such as source IP address, destination IP address, source port, destination port and communication protocol, and also making protection decisions according to session status. However, when people were studying and designing state detection technology, they did not take into account the attacks based on application layer protocol. Indeed, the attack programs can hide themselves in application layer payload. Along with the rapid increase of application layer protocol types, innumerable application layer attacks have emerged. Statistics have shown that about 70% of network attacks and threats come from the application layer. In

ISSN 2522-3488 Vol. 4, Issue 1: 33-40, DOI: 10.25236/IJNDES.040106

this background, it was urgent to adopt new firewall mechanisms to prevent attacks and threats on application layer. One of the inventions that has been gradually applied to the network security devices is called DPI (Deep Packet Inspection) technology.

Traditional security measures and products, such as UTM devices, are simply aggregates of functions such as firewalls, virus protection, and intrusion protection. When these functions are turned on simultaneously, the performance of the products decreases significantly to fail the requirements. Therefore, it is necessary to develop a new product that is all-round and multi-functional. Unlike the traditional ones, the next generation firewall inherits and integrates all the functions (traffic control, intrusion detection, virus killing and so on) from traditional firewall, and in addition highlights the comprehensive detection of data traffic. In particular, the identification and control of information from application layer which is the focus of the design and implementation of the new products, relies on the depth and ability of traffic detection that can be promoted by the improvement of DPI technology, especially the optimization of pattern matching algorithm. In another word, more sophisticated DPI technology will nourish firewall to achieve better performance, higher reliability and stronger intelligence.

2. Foundation knowledge

KMP algorithm is an advanced string single pattern matching algorithm proposed by Knuth, Morris and Ratt, which has higher efficiency than the original pattern matching algorithm. KMP introduced “next[] array” to determine the new position of j’s matching if j fails to match in current state. The value of next [j] indicates that the length of the longest suffix in P [0... j-1] is equal to the prefix of the same character sequence.

The next [] array is defined as follows:

$$\text{next[j]} = \begin{cases} -1 & j = 0 \\ \max(k) & 0 < k < j \;\; P[0 \ldots k-1] = P[j-k, j-1] \\ 0 & \text{else} \end{cases}$$

When next [j] = k > 0, P [0... k-1] = P [j-k, j-1].

Another algorithm frequently used is the BMH2C algorithm, the basic idea of which is to use both text[i] and text[i+1] characters at the same time, as the substring determine the right shift of the mode. In this algorithm, the text character corresponding to the last character of the pattern string and its next text character are regarded as a substring. When the substring appears in the pattern, the pattern moves to the right, aligning the substring with its appearance on the rightmost side of the pattern; otherwise, when the mode moves right, it skips the substring directly, that is, the right shift is m+1. Since two characters are used to determine the right offset, BMH2C algorithm uses a two-dimensional subscript to index the offset array, denoted as skip[char1][char2]. In the first step of array initialization, the values of the two-dimensional array are set to m+1. The algorithm also considers a special case, that is, when the last character text [i+1] of the substring text [i] text [i+1] is

ISSN 2522-3488 Vol. 4, Issue 1: 33-40, DOI: 10.25236/IJNDES.040106

the same as the first character pat [0] in the pattern, right shift m+1 may lead to the miss of a matching, even though the substring text [i] text [i+1] does not appear in the pattern. So the pattern should have m as the right shift such that the first character pat [0] of the pattern is aligned with the last character text [i+1]. As a consequence, in the second step of initialization, the value of skip [char] [pat [0]] is modified to m. The following step of initialization is to set the right shift for all substrings that appear in the pattern. The schematic diagram of initialization is shown in Figure 1.

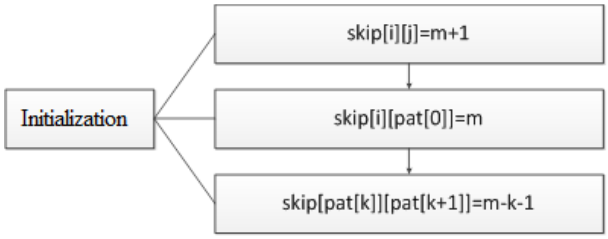

Figure.1 Schematic diagram of initialization

3. The Fast-Single Pattern Matching Algorithm

By the learning from KMP and BMH2C, a new single pattern matching algorithm BMH2CKMP is proposed. The design of BMH2CKMP is as follows.

First, modify BMH2C algorithm to forward matching.

Second, preprocessing, i.e., calculate the next array for Patten strings.

Third, determine whether the skip value is positive or negative when the character mismatches in pattern matching. If positive, skip to get the maximum shift, otherwise, chooses to look up the next array and move the position of j to force i not to backtrack.

In the initialization, BMH2CKMP algorithm presets skip value and next arrays. When the matching starts, text string S aligns left with pattern string P, and then matches characters in turn. As the mismatch occurs, look up the skip array to obtain the shift. At this point we judge whether the shift is positive or negative. If positive, it will skip according to the shift. If negative, the next array is queried in order to obtain the new shift. The procedure loop will not terminate until a successful match or the end of the text string. The positiveness of the shift of next array guarantees that the J value of the algorithm will never be traced back, and the skips have maximum shifts, and complexity O (n).

The algorithm flow is shown in figure 2.

Taking advantages of both KMP and BMH2C algorithm, BMH2CKMP algorithm realizes a single pattern matching procedure that speeds up the pattern matching. It can be applied to DPI technology for the further improvement of the efficiency of the next generation firewall.

ISSN 2522-3488 Vol. 4, Issue 1: 33-40, DOI: 10.25236/IJNDES.040106

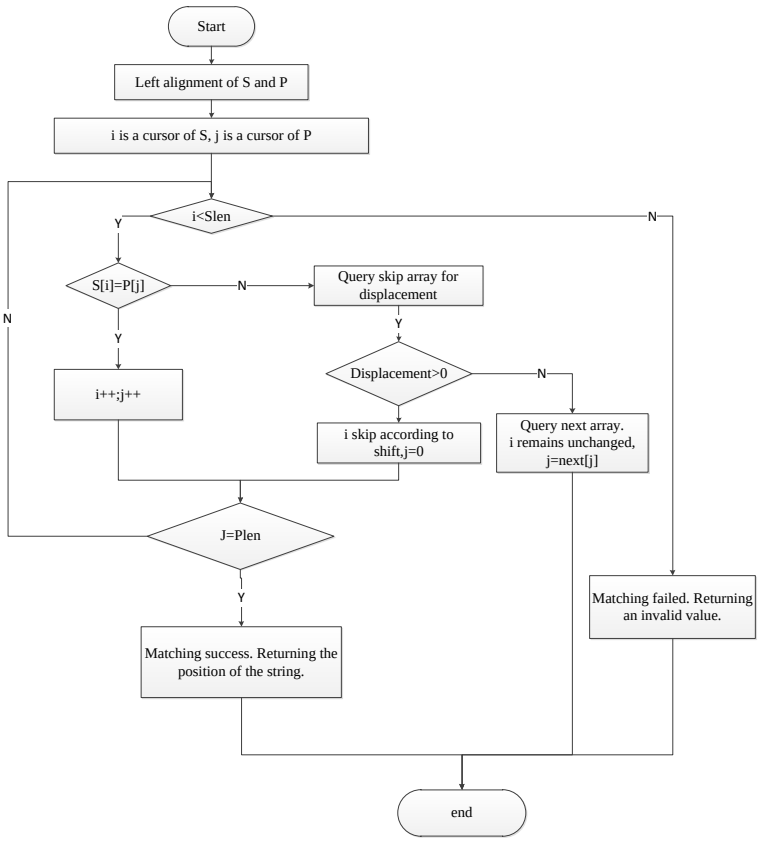

Figure.2 Algorithm Flow

4. Configuration for the case study

A relatively complete next generation firewall system based on DPI technology using BMH2CKMP algorithm is designed and implemented, which can meet the security needs of users of various types of network architecture.

The next generation firewall system is composed of account management, system management, security policy management, log management, security detection, network billing, VPN virtual private network. It includes function modules such as network configuration, routing management, IP-MAC binding, rule configuration.

The physical architecture of the system is shown in Figure 3. Using the browser, the user interacts with HTTP server in two directions. HTTP server connects the user management interface of firewall. This mechanism allows the user to access or manage the firewall directly through the browser without having to deal with its underlying structure.

ISSN 2522-3488 Vol. 4, Issue 1: 33-40, DOI: 10.25236/IJNDES.040106

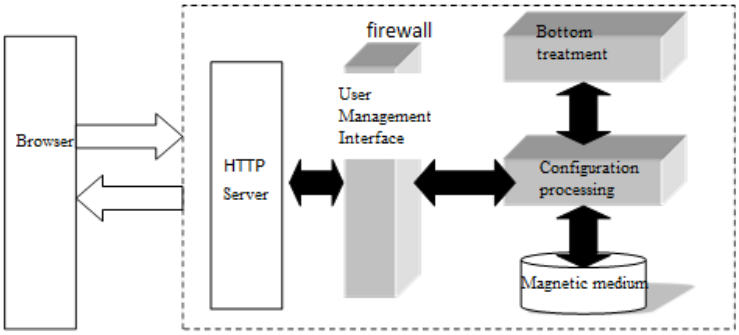

Figure.3 Physics Architecture Diagram

The system program structure is shown in Figure 4. The function modules cooperate with each other through user interface to accomplish the firewall system facility.

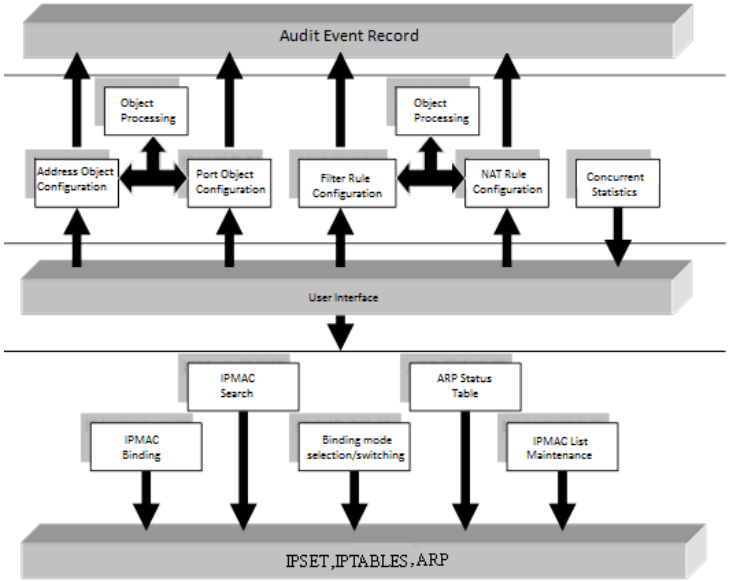

Figure. 4 System Program Architecture Diagram

We test the performance of the next generation firewall with respect to the indicators including the maximum number of concurrent connections, the throughput, and the number of TCP connections opened per second.

4.1 Maximum number of concurrent connections

ISSN 2522-3488 Vol. 4, Issue 1: 33-40, DOI: 10.25236/IJNDES.040106

The maximum number of concurrent connections in a firewall is the maximum number of simultaneous connections that a firewall can support, The number reflects firewall's support for the network. The bigger the number, the larger the scale of the network supported. This indicator is tested by Smart Bits network performance tester and WebSuite test software. The test network topology is shown in Figure 5.

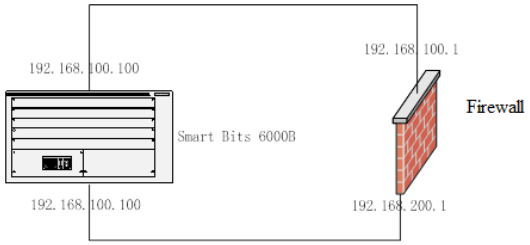

Figure.5 Concurrent Test Topology

4.2 Throughput

Firewall throughput is the maximum rate of the tested device. The larger the throughput, the better the performance of the firewall. In order to comprehensively measure the firewall throughput, according to RFC recommendations, a two-way test is adopted. The entire test time is 120 seconds, and the throughput test is carried out within the multi-stream settings. The multi-streams are set to be bidirectional, 100 convection, UDP (that is, a total of distinct 200 addresses in internal and external networks). Furthermore, source addresses and target addresses change simultaneously, to be specific, there will be 200 state connections in the state table of the firewall. Smart Bits network performance tester and Smart Flow test software are used to test the two-way throughput index of firewall. Expected throughput: 64-byte frame of length 500 Mbps, 512-byte frame of length 4 Gbps, 1518-byte frame of length 10 Gbps, and the impact of the number of rules on throughput should be less than 2%. The actual test results are shown in Table 1.

In fact, the more TCP connections a firewall can open per second, the more requests the firewall can handle at the same time, and hence faster the firewall will be. Smart Bits Network Performance Tester is used to test this indicator. The test network topology is shown in Figure 6.

Table 1 Results of throughput test

Frame length (bytes)	Throughput (%)
64	86.35%
128	100%
256	100%
512	100%

ISSN 2522-3488 Vol. 4, Issue 1: 33-40, DOI: 10.25236/IJNDES.040106

1024	100%
1280	100%
1518	100%

C.Number of TCP connections opened per second

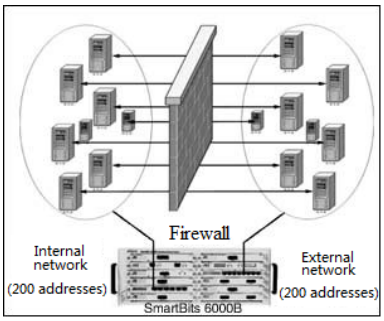

Figure.6 New test topology diagram per second

The number of TCP connections that can be opened per second by firewall before improvement is 200,000 while the actual test result of the improved firewall is 220,000 connections, implying that the our algorithm yields a faster firewall.

5. Discussion and conclusions

Given the facts that KMP algorithm has a small skip distance and BMH2C algorithm cannot guarantee that the pattern is not retrospective after character matching, this paper designs a fast single pattern matching algorithm BMH2CKMP, and applies it to the design of a new firewall. Through testing with respect to three performance indicators including throughput, maximum number of concurrent connections and number of new connections per second, it can be found that our designed firewall with the improved pattern matching algorithm in it has better performance than does the one without it.

References

[1] Katic T, Pale P (2007). Optimization of firewall rules. Proceedings of the ITI 29th International Conf on Information Technology Interfaces, no. 29, pp.685-690.

[2] Nazief H M, Sabastian T A, Presekal A, et al (2014). Development of University of Indonesia next generation firewall prototype and access control with deep packet inspection. International Conference on Advanced Computer Science and Information Systems. IEEE, pp.47-52.

[3] CORWIN E H (2011). Deep Packet Inspection: Shaping the Internet and the Implications on Privacy and Securit. Information Security Journal: A Global Perspective, vol.20, no.6, pp.311-316.

[4] Min Lianting, Zhao Tingting (2006). Research and Improvement of BM Algorithms. Journal of Wuhan University of Technology (Transportation

ISSN 2522-3488 Vol. 4, Issue 1: 33-40, DOI: 10.25236/IJNDES.040106

Science& Engineering), vol.30, no.3, pp.528-30.

[5] Nazief H M, Sabastian T A, Presekal A, et al. Development of University of Indonesia next generation firewall prototype and access control with deep packet inspection. International Conference on Advanced Computer Science and Information Systems. IEEE, 2014:47-52.

[6] L Maccari, RL Cigno (2013). Waterwall: a cooperative, distributed firewall for wireless mesh networks. Eurasip Journal on Wireless Communications and Networking, no.1, pp.1-12.

[7] Olesenbagneux O (2017). The library before print and after the computer: The similarities between string search algorithms and mnemonic retrieval in pre-print libraries. Information Society, vol.33, no.4, pp.205-214.

[8] Horspool R N (2010). Practical fast searching in strings. Software Practice & Experience, 2010, vol.10, no.6, pp.501-506.