The Detection Method of Website Administrator Account Password Brute-Force Attack by Flow Characteristics

Yue Liu

College of Automation, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

Abstract

As one of the main means of threatening the security of the site, brute-force attacked by the administrator all the possible account and password combination to remotely log site backstage, tampering or theft of information, the site caused significant losses. This paper presents a method of detecting the brute-force attack of the website administrator password based on the traffic characteristics. By obtaining the statistical characteristics of the traffic flow, the attacking behavior is obvious based on the number of packets requested. The data is analyzed and reexamined by the packet characteristics The Experimental results show that the method has high accuracy and can identify more than 95% of violent attacks.

Keywords

Web Site Security; Characteristics of Network Flow; Brute-Force Attacks.

1. Introduction

With the development of the Internet, the popularity of a large number of intelligent devices, people's lifestyles and habits has undergone great changes, a variety of APP inadvertently penetrate into the user's lives in all aspects of a large number of shared information to bring people all aspects of convenience, but because of the openness and vulnerability of the network, to bring convenience to users at the same time, but also for users to bring a security risk can not be overlooked. In the case of a web application, the user stores his personal information in the background of the website, which is managed by the administrator. However, the attacker can crack the administrator's password through violence, obtain the website user information, cause the user information to leak, Property security has had a great impact.

2. Summary of Domestic and Foreign Research

In recent years, domestic and foreign experts have put forward a series of brute-force attack site background attack detection methods, including the method. In literature, an anomaly detection algorithm based on K-Nearest Neighbor algorithm is proposed, which identifies violence by comparing historical data with recent data sampling. Document [2] through the penetration test to detect whether the site was cracked, the method used to simulate the hacker attacks on the site to assess the security. [3] investigated the user in a number of sites to reuse the password caused by hackers on the user account password cracking effect. In [4], the concept of a false password is proposed. When an attacker tries to log in with a false password, the system issues an alarm. [5] from the analysis of HTTP log, the establishment of a normal user behavior model, used to detect network attacks. These studies have analyzed or proposed a method of detecting brute-force attacks from different perspectives. However, with the development of modern technology, users pay attention to privacy, the usually detection methods can not meet the existing security needs.

In this paper, according to the existing shortcomings of the background detection method of background violence, the paper analyzes the attack behavior from the viewpoint of website traffic, that is, proposes a website background violence detection method based on traffic characteristics, which can effectively improve its detection rate.

3. Violence crack flow characteristics selection

3.1 Packet size characteristics

In the process of website communication, the general use of http or https protocol for communication, and HTTP or https protocol is two different hosts in the form of exchange of messages to complete the page request and response. And the data packet is the basic unit of all Internet transmission data. Packets store their basic information in the header, including source IP, destination IP, packet size, protocol, and so on. Through a large number of experimental data analysis, in the usual network communication, the size of the packet is uncertain, but in the brute force attack, the attacker in order to speed up the attack process, will continue to try the account password, and the speed was faster than ordinary people. In the case of Figure 1, this article simulates a site brute-force attacks.

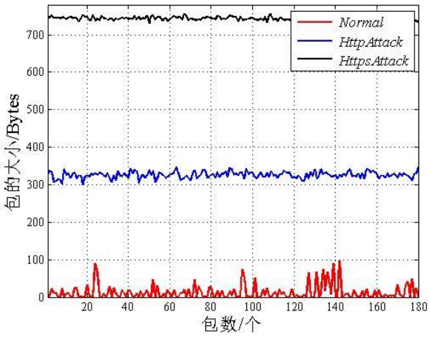


Fig.1 The number of fluctuations in the process

As can be seen from Figure 1, normal web access does not result in a large number of packets of the same size, because the normal access to the page, because the object is different, so the size of the package is not the same. In the process of violent attacks, due to the attacker constantly try to account password, continue to send the same request, and the attacker in order to speed up the attack speed, the use of scripting software, attack frequency is much faster than normal access, it will appear in a short time multiple packet sizes are similar to the case of packets. Analysis of packet size between hosts needs to communicate between the host data packets, the steps are as follows:

Step 1 build the simulation site background login page;

Step 2 Use the violent attack software to violate the background password of the website;

Step 3 use the capturing tool to grab the packets flowing through the network card in seconds as pcap files N_i (i = 1, 2, 3 ...);

Step 4 resolves the five-tuple (source IP address, source port, destination IP address, destination port, and application layer protocol) inN_i;

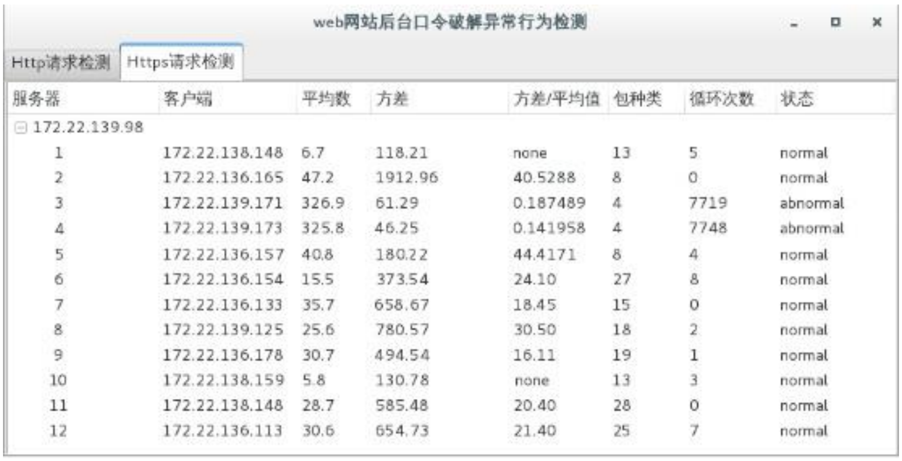
Step 5 to keep N_i contains useful information packets, the packet grouping, the same source IP to the destination IP packet is divided into a group;

Step 6 calculates the total sizeT_i, the packet number B_i of the packet of each packet according to the packet header information, and calculates the average size of the packets of each packet $\overline{t_i}$=T_i/B_i.

3.2 Number of packets

As a result of violent attacks using script software, by analyzing a large number of normal packets and abnormal packets, abnormal packets found in the number of packets on a very strong statistical law, and normal user site access does not exist such a rule, through statistical data the number of packages to analyze the existence of statistical characteristics as a basis for brute-force attack. In this paper, the use of Hydra violence crack software to test, and with the normal user data comparison, Table 1 is the case of cases of the number of changes in the number of packets.

Tab.1 the changes of package



服务器	客户端	平均数	方差	方差/平均值	包种类	循环次数	状态
⊟ 172.22.139.98							
1	172.22.138.148	6.7	118.21	none	13	5	normal
2	172.22.136.165	47.2	1912.96	40.5288	8	0	normal
3	172.22.139.171	326.9	61.29	0.187489	4	7719	abnormal
4	172.22.139.173	325.8	46.25	0.141958	4	7748	abnormal
5	172.22.136.157	40.8	180.22	44.4171	8	4	normal
6	172.22.136.154	15.5	373.54	24.10	27	8	normal
7	172.22.136.133	35.7	658.67	18.45	15	0	normal
8	172.22.139.125	25.6	780.57	30.50	18	2	normal
9	172.22.136.178	30.7	494.54	16.11	19	1	normal
10	172.22.138.159	5.8	130.78	none	13	3	normal
11	172.22.138.148	28.7	585.48	20.40	28	0	normal
12	172.22.136.113	30.6	654.73	21.40	25	7	normal

As can be seen from Table 1, in the case of attack, the size of the packet per second because the attack protocol varies, but the use of the same attack software packet size will not appear too much volatility. And the normal communication, the size of the packet is random changes, and fluctuations are large, there is no statistical law. Determine the number of changes in the number of packets first of all, the statistical processing of data packets, statistical steps are as follows:

Step 1 Use the capturing tool to grab the packets flowing through the network card in seconds as pcap files N_i (i = 1, 2, 3 ...);

Step 2 Resolve the five-tuple (source IP address, source port, destination IP address, destination port, and application layer protocol) in N_i;

Step 3 Keep the packets containing the protocol in N_i, group the packets into packets with the same source IP to destination IP packets.

Step 4 calculates the size of each packet for each packet according to the packet header information, divides the packets of the same size into a group, and counts the numberM_i.

4. Detection method construction

4.1 Packet size detection

From the above analysis, the number of normal packets per second fluctuations, and the number of changes per second packet is not regular, but the number of abnormal data packets fluctuates smaller, and every second changes in the situation there is a clear statistical law. By calculating the standard deviation can determine the size of the data fluctuations, the standard deviation σ is

$$\sigma=\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(x_i-\overline{X}\right)^2} \tag{1}$$

Where the number of samples is the mean of the sample, the smaller the standard deviation is, the smaller the fluctuation range of the data is. The smaller the σ_t of the set of samples is less than the threshold σ_T. It can be seen that the smaller the σ, the higher the detection rate, but with the decrease of σ, the number of samples increases, which in turn increases the risk of the site being cracked. Therefore, you need to find a suitable value to achieve the best results. In this paper, the judgment value $P=\sigma^2/\overline{t_i}$ is chosen as the criterion. After the experimental analysis, when $P < 0.5$, it is determined that the brute-force attacks; when $100 < P < 0.5$, uncertainty is abnormal, needs to do further testing. Through this classification can filter a large number of obvious brute-force attacks, reduce the amount of calculation, and speed up the detection process. By detecting the size of the packet, it can quickly detect the single-plane violence attack, but because the distributed attack is not continuous in time, this method is not high for the distributed attack detection rate, need to do further testing.

4.2 Number of packets

From the above analysis, we can see that there is no strong statistical law of the number of communication packets between hosts in the normal website communication every second. However, in the process of brute-force attack, due to the existence of software script, there are statistical laws, Attack type of discontinuity, only by counting the number of packets per second, is not enough to determine whether the violence attack. Based on this situation, this paper is counted by the number of packets, and the number of packets of the same size of each packet is counted by intercepting consecutive packets and grouping them by twice the number. Comparison found that the same size of the number of packets is a multiple increase. In other words, you need to count the number of packets at least two seconds, and the same size of the number of packets for statistics, you can determine whether the background of the site was brute-force attack attack.

4.3 Detection method to achieve

According to the above method, we use Python language to achieve detection system of brute-force. Change the system can be used to find the page account password single machine brute-force attack and distributed brute-force crack behavior. Into the pcap format for the packet, the output is included in the violence to crack the IP group. Detection process as shown:

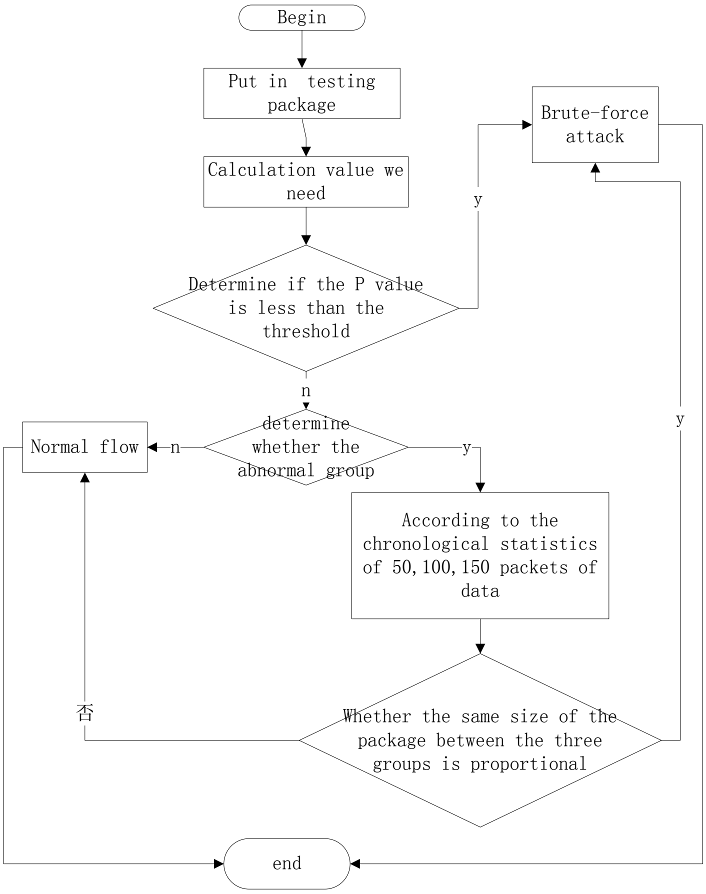

Begin
Put in testing package
Calculation value we need
Determine if the P value is less than the threshold
y
Brute-force attack
n
determine whether the abnormal group
n
Normal flow
y
According to the chronological statistics of 50,100,150 packets of data
Whether the same size of the package between the three groups is proportional
否
y
end

Fig.2 Brute-Force Attacks testing flow chart

5. Experimental results and analysis

In order to verify the effectiveness of the brute-force attack recognition system, this paper builds a simulation page for testing. The experimental environment consists of sixteen PC and a switch, one host as the attack server, two hosts as an attacker, the rest for the normal user. Test continued communication for 500 minutes, during a total of 30 attacks, for http, https remote login account password violent 10 times, distributed attack 5 times, each attack for 10 minutes, a total attack of 150 minutes. The test results are shown in the table.

	the time determined as attack M′	The time determined as normal L'
The length of attack M=150min	146min	4min
The length of attack L=350min	3min	347min

Define the performance evaluation criteria of the brute-force attack identification algorithm as follows:

(1) Accuracy

$$ACC = \frac{T_{L\to L'} + T_{M\to M'}}{T_{L\to L'} + T_{M\to M'} + T_{M\to L'} + T_{L\to M'}} \tag{2}$$

(2) False Positive Rate

$$FPR = \frac{T_{L\to M'}}{T_{L\to L'} + T_{L\to M'}} \tag{3}$$

(3) false negative rate

$$FNR = \frac{T_{M\to L'}}{T_{M\to L'} + T_{M\to M'}} \tag{4}$$

In the case, $T_{M\to M'}$ the length of the attack is detected correctly; $T_{L\to M'}$ the length of the attack is incorrectly determined to include the attack; $T_{M\to L'}$ the length of time that the attack is judged to be not included in the attack; $T_{L\to L'}$ the length of the attack is not detected.

False Positive Rate	false negative rate	Accuracy
0.8%	2.7%	98.6%

After analysis, the test results in three false positives are due to three groups of fluctuations caused by a small misjudgment.

6. Conclusion

Based on the analysis of a large number of brute-force attacked traffic, this paper summarizes the traffic characteristics and description methods that characterize the abnormal data, and put forward the method of attack detection based on the flow of the website administrator password and the system development and implementation of the detection method. The detection system uses the simulation data to extract the traffic data, extracts the traffic characteristics of the data, and realizes the detection and judgment of the brute-force attack of the website administrator password. The experimental results show that the effectiveness of the method based on the traffic feature detection method can be used to identify the background of the Http and https sites, and can obtain high detection accuracy and lower error.

References

[1] Zhang Lijuan.Massnetworkmonitoringdataautomaticfusion and correlation analysis[J]. School of Electronic and Information Engineering, Shenzhen Polytechnic, Shenzhen 518055, Guangdong, China, 2011, 8: 21-23.

[2]TIAN Lijun. Techniques and methods for penetration testing[J]. Department of Information Technology, Xi’an Railway Administration, Xi’an 710054, China

[3]Das A, Bonneau J, Caesar M, et al. The Tangled Web of Password Reuse[C]//NDSS. 2014, 14: 23-26.

[4]Kaur J, Singh R, Kaur P. Prevention of DDoS and brute force attacks on web log files using combination of genetic algorithm and feed forward back propagation neural network[J]. International Journal of Computer Applications, 2015, 120(23).

[5]Zolotukhin M, Hämäläinen T, Kokkonen T, et al. Analysis of HTTP requests for anomaly detection of web attacks[C]//Dependable, Autonomic and Secure Computing (DASC), 2014 IEEE 12th International Conference on. IEEE, 2014: 406-411.

[6]Bonneau J, Herley C, van Oorschot P C, et al. Passwords and the evolution of imperfect authentication[J]. Communications of the ACM, 2015, 58(7): 78-87.

[7]Zhao Z, Li S, Kang Y, et al. A weak password cracker of UHF RFID tags[C].Ubiquitous Intelligence and Computing and 2015 IEEE 12th Intl Conf on Autonomic and Trusted Computing and 2015 IEEE 15th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom), 2015: 1563-1566.

[8]Surian D, Chawla S. Detection of spatiotemporal outlier events in social networks[J]. Encyclopedia of Social Network Analysis and Mining, 2014: 364-369.

[9]Howell C J M. The restrictive deterrent effect of warning banners in a compromised computer System[J]. 2016.

[10]Abdou A R, Barrera D, Van Oorschot P C. What lies beneath? analyzing automated SSH bruteforce attacks [C].International Conference on Passwords. Springer International Publishing, 2015: 72-91.

[11]Kuo C, Romanosky S, Cranor L F. Human selection of mnemonic phrase-based passwords [C]//Proceedings of the second symposium on Usable privacy and security. ACM, 2006: 67-78.

[12]Florencio D, Herley C. A large-scale study of web password habits[C]//Proceedings of the 16th international conference on World Wide Web. ACM, 2007: 657-666.

[13]Halderman J A, Waters B, Felten E W. A convenient method for securely managing passwords[C]//Proceedings of the 14th international conference on World Wide Web. ACM, 2005: 471-479.

[14]Florêncio D, Herley C, Coskun B. Do strong web passwords accomplish anything?[J]. HotSec, 2007, 7(6).

[15]Farmer D, Venema W. Improving the security of your site by breaking into it[J]. 1993.