Modeling User Browsing Activity for Application Layer DDoS Attack Detection

TungNgai Miu[1], Chenxu Wang[2,3(✉)], Daniel Xiapu Luo[3], and Jinhe Wang[2]

[1] NexusGuard Limited, Hong Kong, China
[2] Xi'an Jiaotong University, Xi'an, Shaanxi, China
cxwang@mail.xjtu.edu.cn
[3] The Hong Kong Polytechnic University, Hong Kong, China

Abstract. Application layer distributed denial of service (App-layer DDoS) attacks are becoming a severe threat to the security of web servers. In this paper, we model user browsing activity in order to detect abnormal requests. User access patterns are analyzed to detect anomaly at the session level. The likelihood of a browsing session is then calculated to distinguish abnormal behaviors from normal ones. We evaluate our methods based on a real dataset collected from a commercial website that suffered from actual DDoS attacks. The experimental results validate the effectiveness of the proposed methods.

Keywords: DDoS attack · Browsing activity · User access pattern

1 Introduction

Application layer DDoS attacks attempt to disrupt legitimate users' services by exhausting the resources of the victims [1]. Since such attacks masquerade as flash crowds (a large number of normal users access to a web server simultaneously) by generating legitimate traffic [2], conventional signature-based intrusion detection systems (IDS) become ineffective to them. Moreover, compared with the botnet-induced volumetric attacks that generate a significant amount of traffic, low-volume DDoS attacks are even more pernicious and problematic from a defensive standpoint since attacks generally consume less bandwidth and are stealthier in nature [3].

The eventual criterion to distinguish illegitimate users from legitimate ones is the intentions of visiting users [4], which could be well inferred from the browsing activity. In this paper, we propose a Markov model to profile users' browsing activity. The model well characterizes user's access patterns in HTTP sessions. Then the likelihood of a user's browsing activity is calculated based on the access patterns to detect abnormal sessions. The proposed method affiliates the detection of stealthy DDoS attacks, thus reducing the false positive rate. Finally, we evaluate the performance of the detection method based on a real DDoS attack dataset collected from a busy e-commercial web server.

R. Deng et al. (Eds.): SecureComm 2016, LNICST 198, pp. 747–750, 2017.
DOI: 10.1007/978-3-319-59608-2_42

2 Detection Schemes

In an HTTP session, the user browses the website by jumping from one web page to another. We assume that in a single session the next page a user will browse only depends on the current browsing page, and employ the Markov Chain Model to describe user access patterns. The Markov property of user access patterns has been validated in [5]. We further use a directed weighted graph to represent the Markov Chain, where each node representing a main page and the weights of the edge representing the transition probabilities from one page to another. Formally, the transition probability from page i to page j is defined as

$$p(i|j) = \frac{n_{ij}}{\sum_{j=1}^{N} n_{ij}}, \tag{1}$$

where n_{ij} is the number of observations that page i is followed by page j in a single session; N is the total number of pages. Denote a session as $\{MP_1, MP_2, \ldots, MP_n\}$, where n is the length of the session representing the number of main pages. Then, the log likelihood of the session is defined as

$$\ln L = \ln p(MP_1) + \sum_{i=1}^{n-1} \ln p(MP_i|MP_{i+1}) \tag{2}$$

where $p(MP_1)$ is the probability of page MP_1, and $p(MP_i|MP_{i+1})$ is the transition probability from the i_{th} to $(i+1)_{\text{th}}$ page.

3 Experiments

We conduct the experiments based on real data collected from a commercial web server. Table 1 lists a brief summary of the dataset.

Table 1. Summary of the dataset

Date	Requests	Users	Max. RR[1]	Min. RR[1]	Suspected IPs
2015/12/29	30,933,159	30,242	283	20	845
2015/12/30	32,202,986	32,886	290	18	1023
2015/12/31	30,850,731	31,063	341	19	1139
Total	93,986,876	74,773	-	-	1270

[1] RR is the abbreviate of request rate with a time unit of second.

The website has a total of 8464 pages and 14036 objects. The access patterns are closely related to the web structure, which exhibits hierarchical clusters. The transition matrix of the top 80 most accessed pages are shown in Fig. 1. These pages dominates 90% of the total requests. Then, we use the transition matrix to calculate the likelihood of all sessions and the results versus the session

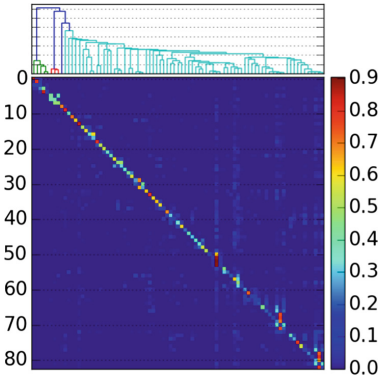


Fig. 1. Transition matrix of web pages

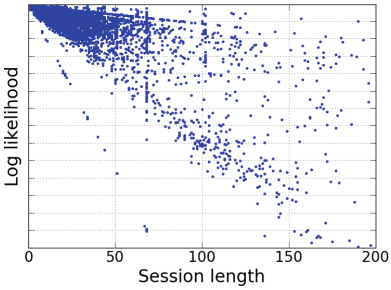
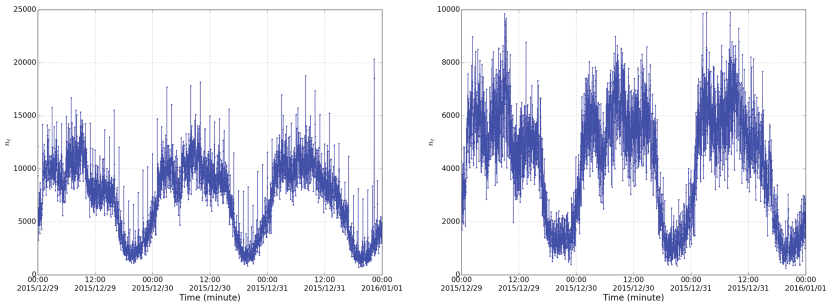


Fig. 2. Session likelihoods v.s. length

length are shown in Fig. 2. It is clearly shown that there are some outliers for different session lengths. This demonstrates that our methods can distinguish the abnormal sessions from the normal ones.

Testing and validation of a detection method is complicated due to the lack of adequate datasets that clearly identify attack behaviors and legitimate human users particularly flash crowd. Following we conducted statistical experiments to evaluate the effectiveness of our method [6].

Denote n_t as the number of requests received by the server in a time unit, and the request rate is plotted versus the time in Fig. 3(a). It is observed that the server suffered from periodic DDoS attacks which result in the comb-shape. Figure 3(b) shows the request rate after filtering the attacking traffic based on the detection results of the combined method. The comparison indicates that the detection method is effective to reduce the burden of the server. In addition, it is noticeable that the request rate varies periodically, suggesting that detection methods should avoid the impacts of fluctuations raised by such periodicity.



(a) Original request rate

(b) Filtered request rate

Fig. 3. The request rate versus the time

Figure 4 compare the access frequency of pages. It is shown that the filtered activity follows the Zipf distribution. Figure 5 presents the distribution of the inter-request times between two consecutive accessed pages. It is shown that the filtered data follows Pareto distribution.

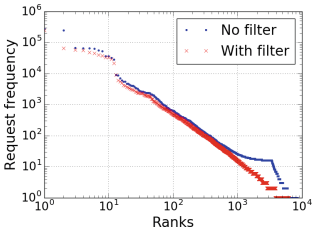


Fig. 4. Page frequency distributions

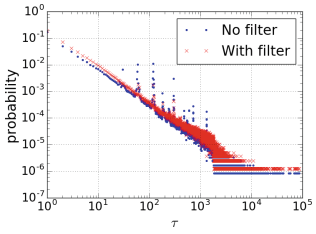


Fig. 5. Time interval distributions

4 Conclusion

We propose a new mechanism to detect application layer DDoS attacks. Based on the access log at the sever end, we propose a Markov model to describe the browsing activity of a user in an HTTP session. Then, the likelihood of a session is calculated and the results are used to distinguish the attack behaviors from the normal ones. We use a real dataset to evaluate the effectiveness of our method.

Acknowledgments. We thank the anonymous reviewers for their quality reviews and suggestions. This work is supported in part by the Hong Kong ITF (No. UIM/285) and Shenzhen City Science and Technology R&D Fund (No. JCYJ20150630115257892).

References

1. Xie, Y., Shun-Zheng, Y.: A large-scale hidden semi-markov model for anomaly detection on user browsing behaviors. IEEE/ACM Trans. Networking **17**(1), 54–65 (2009)
2. Zargar, S.T., Joshi, J., Tipper, D.: A survey of defense mechanisms against distributed denial of service (DDoS) flooding attacks. IEEE Commun. Surv. Tutorials **15**(4), 2046–2069 (2013)
3. DARPA. Extreme ddos defense (2015). http://www.darpa.mil/program/extreme-ddos-defense. Accessed 17 Apr 2016
4. Xie, Y., Shun-Zheng, Y.: A novel model for detecting application layer DDoS attacks. In: Proceedings IMSCCS (2006)
5. Li, Z., Tian, J.: Testing the suitability of markov chains as web usage models. In: Proceedings COMPSAC (2003)
6. Sivabalan, S., Radcliffe, P.J.: A novel framework to detect and block DDoS attack at the application layer. In: Proceedings IEEE TENCON (2013)