Vol 1, Ed. 2. e-ISSN 2716-666X

The Asian Journal of Professional and Business Studies

Please cite this article as: Yusof, M.A.A., Samad, N.H.A., & Adnan, R. (2020). Packet threshold algorithm coupled with machine learning for DDoS classification attacks. *The Asian Journal of Professional and Business Studies, Volume 1* (2).

PACKET THRESHOLD ALGORITHM COUPLED WITH MACHINE LEARNING FOR DDoS CLASSIFICATION ATTACKS

[1]MOHD AZAHARI MOHD YUSOF*
azhari@kuptm.edu.my
[2]NOR HAFIZA ABD SAMAD
hafiza@kuptm.edu.my
[3]RUKHIYAH ADNAN
rukhiyah @kuptm.edu.my

Corresponding Author*

[1,2,3]*Kolej Universiti Poly-Tech MARA*

ABSTRACT

Today, DDoS attacks are the most common Internet threats. DDoS attacks are generated by attackers from anywhere to disable a company's servers from being accessed by users worldwide. An attacker can easily launch one or more types of DDoS attacks at a time. DDoS attacks that can be generated by attackers include Slowloris, UDP flood, Smurf, HTTP flood, TCP SYN flood and more. Therefore, we have proposed a technique called the Packet Threshold Algorithm (PTA) in this paper, where it is combined with several machine learning to classify normal packet and DDoS attacks, namely UDP flood, Smurf, TCP SYN flood and Ping of Death. There are four machine learning, which are K-Nearest Neighbor (KNN), Naïve Bayes, Logistic Regression and Support Vector Machine (SVM) combined with the Packet Threshold Algorithm (PTA) to reduce false positive rate to obtain high detection accuracy. Among the four combinations of techniques, PTA-KNN has been considered as the best technique in the context of reduction of false positive rate. The determination of this best technique is based on the PTA-KNN has achieved the highest detection accuracy (99.83%) compared to the other three techniques with only 0.02% false positive rate. The determination of this best technique is based on the PTA-KNN has achieved the highest detection accuracy (99.83%) compared to the other three techniques with only 0.02% false positive rate.

ARTICLE INFO

Keywords:

DDoS attack,
False positive rate,
Detection accuracy,
Machine learning

1.0 INTRODUCTION

Nowadays, computer networks are essential and widely used by users around the world. It provides many functions to network users, including information exchange, even when they are in different places. Computer networks are formed by combining several devices to form several types of networks, where it can be local area network, metropolitan area network, wide area network or wireless network (Caulfield and Fielder 2015). However, computer networks are often interrupted by several network attacks, including Distributed Denial of Service (DDoS) attacks.

DDoS attacks will avoid users from accessing the network system even if they are legitimate users (Azahari Mohd Yusof, Hani Mohd Ali, & Yusof Darus, 2018). Usually, attackers use botnets to launch DDoS attacks, where the botnet helps the DDoS attack more smoothly. There are three categories of DDoS attacks based on (Gadelrab, Elsheikh et al. 2018), the first category is protocol attack, followed by application layer attack, and the third category is volume-based attack. TCP SYN flood, Smurf and Ping of Death are three types of DDoS attacks in the category of protocol attack. The attacker launches the TCP SYN flood by sending a large number of SYN requests to the target server at a time to keep the target server inoperable (Kolahi, Alghalbi et al. 2014). Attackers build Smurf attacks by sending large amounts of ICMP packets to all broadcast addresses of a target server, where the target server fails to respond to all requests (Sandeep and Rajneet 2014). Meanwhile, Ping of Death happens when an attacker attempts to override a target server by sending more than 65,535 bytes of ICMP packets (Gunasekhar, Rao, Saikiran, & Lakshmi, 2014). Two examples of DDoS attacks in the application layer attack category are Zero-day attack and Slowloris, where Zero-day attack was created by the attacker to avoid software developers fixing the defects that have been detected in the software that has been developed (Singh, Joshi et al. 2017). Meanwhile, Slowloris is a tool used by attackers to disable a target server through a single computer by sending multiple HTTP partial requests consistently (Calvert & Khoshgoftaar, 2019). Another two DDoS attacks called ICMP flood and UDP flood have been considered as the volume-based attack category. Attackers can turn off a target server by generating ICMP flood, where they send large amounts of ICMP packets to that target server (Alqahtani, Balushi et al. 2014). Apart from that, attackers can also send multiple UDP packets to generate UDP flood to disrupt target server stability (Badis, Doyen, & Khatoun, 2014).

We have prepared this paper to present a technique for detecting normal packets and DDoS packets whether they are Smurf, UDP flood, Ping of Death or TCP SYN flood using our proposed algorithm called Packet Threshold Algorithm (PTA). There are several types of DDoS attacks that can be generated by attackers from wherever they are, but the four types of DDoS attacks mentioned are focused in our study with a few justifications. According to (Arora and Dalal 2019), these four types of DDoS attacks are the most popular types of attacks launched by attackers at any given time because DDoS has a very simple concept by generating a large amount of traffic to the target server. When a DDoS attack is successful, it is extremely difficult to stop unless the attacker is tired enough to continue the attack on the target server. Most importantly, PTA will be combined with four machine learning called KNN, Naïve Bayes, Logistic Regression and SVM to compare them in terms of detection accuracy and false positive rate.

This paper contains five main sections, where section 2 presents some of the strengths and weaknesses found in previous studies. They relate to DDoS attack detection techniques. The methodology and evaluation in section 3 are designed to describe several phases of our study, including the measurement parameters used to obtain the results. Results and discussion are continued in section 4 to present the results of our study and conclusion presented in section 5 to summarize this paper's contents.

2.0 LITERATURE REVIEW

As we know, DDoS attacks are intended to disable the target server so that it is not accessible to anyone at that time. Therefore, most organizations need to have specific techniques to address DDoS issues against their computer network environment. There are several solutions have been developed by previous researchers. It includes the least squares support vector machine designed by (Sahi, Lai et al. 2017), where the technique has loaded a function named as CS_DDoS. It is a classifier to decide whether incoming packets are normal packets or DDoS packets. The classifier contains two important stages, the first stage is detection stage and the other stage is prevention stage, where the detection stage is able to determine

the packet present in the cloud is a normal packet or TCP SYN flood. Meanwhile, the prevention stage will act to record and blacklist the source IP if the incoming packet is detected as a DDoS attack.

Density-based spatial clustering and application with noise technique have formed by (Al-mamory & Algelal, 2017), where it is based on the concept of entropy. They used the DARPA dataset to perform experimental activities to see the accuracy of their techniques. They compared the technique with other techniques such as k-means, FCM and GKFCM to see the strength of their technique in terms of detection accuracy and false positive rate. Apart from that, a technique named artificial neural network was designed by (Peraković, Periša et al. 2016) to determine incoming packets are DDoS attacks or non-network attacks. They conducted a simulation to see the technique in terms of detection accuracy. All packets that have been classified as DDoS attacks will be migrated into a new dataset. They begin the simulation using the MatLab tool by applying several network packets from four different datasets. Another technique is the hop-count filter proposed by (Li, Yang et al. 2015), where it is useful for detecting clean packets or DDoS attacks. The technique is loaded with the packet threshold algorithm, where if the incoming packet exceeds the specified packet threshold, the packet is known as a DDoS attack.

Unfortunately, all of the techniques mentioned above have two problems that are still unresolved, false positive rate and detection accuracy. The main reason for the high false positive rate on the technique that has been built is that the technique does not differentiate between clean packet and DDoS traffic well. There are also techniques that achieve a high detection rate (for specific network traffic) but it has a 0.9% false positive rate, which is considered a high false positive rate. Additionally, some techniques mistakenly detect two different types of DDoS attacks, for example, UDP packets are detected as TCP packets due to flow inequality. This will make the technique unsuccessful to detect different types of DDoS attacks based on attack behaviour. As a major conclusion, the problem of a high false positive rate will have a severe impact on the detection accuracy of a technique developed. Apart from that, some of these techniques can only detect DDoS attacks in general, where it does not focus on the specific type of DDoS attacks such as TCP SYN flood, Smurf and Ping of Death. Detection of a specific type of DDoS attack is important because it can be launched easily to disrupt server activity using different packet types, whether UDP, ICMP or TCP.

3.0 METHODOLOGY

To support this study, we need to consider four important phases as shown in Figure 1. Each of these phases must be completed to produce a successful study based on the objectives set out in Section 1.

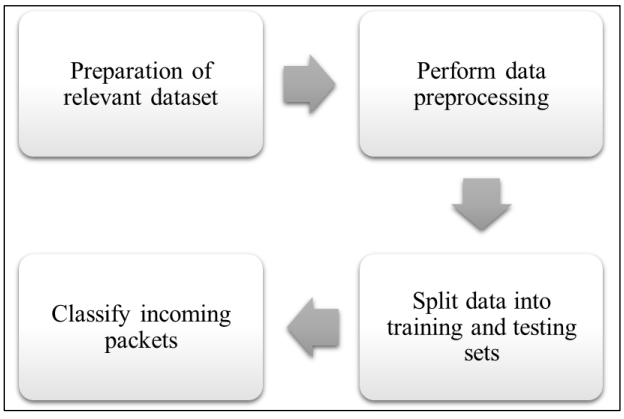

Figure 1. Research Methodology

The first phase of our study is the preparation of relevant datasets which contains several incoming packets that have been captured, where it is an open source dataset as presented in Table 1. There are several open datasets available, but the

datasets we use are particularly suitable for our study because the recorded incoming packets consist of normal and DDoS packets. Moreover, we focus on five types of packets, they are normal, Smurf, UDP flood, Ping of Death and TCP SYN flood packets.

Table 1. Sample of DDoS Dataset

SOURCE ADDRESS	DESTINATION ADDRESS	PACKET TYPE	PACKET SIZE	…	PACKET CLASS
192.168.10.87	192.168.10.78	ICMP	65,535		Ping of Death
192.168.10.97	192.168.10.124	UDP	1,192		UDP flood
192.168.10.124	192.168.10.97	ICMP	1,540		Smurf
192.168.10.54	192.168.10.87	TCP	55		Normal
192.168.10.78	192.168.10.65	TCP	77		TCP SYN flood

Our second phase is performing data preprocessing, where it involves two important steps, data cleaning and data reduction. We use data cleaning method to ensure that the data stored in the dataset is accurate and consistent without any error. Then, we implemented a data reduction method, where we selected important data in our study to make the simulation run smoothly. Both of these methods are very important to our study because originally the data we obtained were inconsistent, incomplete and noisy.

After complete the data preprocessing activities, we split the data into training set and testing set. We provide 80% for the training set and the remaining 20% for the testing set. There are 240,000 samples in the dataset that we used in our study. This means that there are 192,000 samples for the training set and 48,000 samples for the testing set.

The final phase to complete this study is to classify incoming packets, whether they are normal packets or DDoS packets. We propose a technique called Packet Threshold Algorithm (PTA) and combined with four machine learning techniques, they are Naive Bayes (NB), K-Nearest Neighbor (KNN), Logistic Regression (LR) and Support Vector Machine (SVM) as presented in Figure 2. The PTA works to check the type of packet class that will be sent to the server, either normal packet, Ping of Death, UDP flood, Smurf or TCP SYN flood. Normal packets are detected if the packet type is TCP, ICMP or ICMP sent to the server not exceeding 60 packets per second. The PTA will detect Ping of Death if the ICMP packet received by the server exceeds 65,535 bytes per second. Meanwhile, Smurf can be detected by the PTA if the ICMP packet received by the server exceeds 60 and less than 65,535 bytes per second. Next, the UDP flood is detected by the PTA if the UDP packet size received by the server exceeds 60 packets per second.

We apply two performance metrics to our proposed technique, namely detection accuracy and false positive rate. Detection accuracy is used to determine the correct number of packets detected. Meanwhile, false positive rate is used to determine the number of normal packets inaccurately detected as DDoS packet or DDoS packet inaccurately detected as DDoS packet itself.

```
if[(ps < 60) & (pt == 'tcp') | (pt == 'udp') | (pt == 'icmp')]:
    print(ddos_df.iloc[[0], [-1]])
    print('Action: pass packets')

if[(ps >= 60) & (pt == 'tcp')]:
    print(ddos_df.iloc[[159], [-1]])
    print('Action: drop packets')

if[(ps >= 60) & (pt == 'udp')]:
    print(ddos_df.iloc[[2], [-1]])
    print('Action: drop packets')

if[(ps >= 65535) & (pt == 'icmp')]:
    print(ddos_df.iloc[[25], [-1]])
    print('Action: drop packets')

if[(ps >= 60) & (ps < 65535) & (pt == 'icmp')]:
    print(ddos_df.iloc[[17], [-1]])
    print('Action: drop packets')
```

Figure 2. Packet Threshold Algorithm (PTA)

4.0 FINDINGS AND DISCUSSION

Referring to Table 2, it shows the results obtained in our study. It shows that PTA-KNN achieved 99.83% detection accuracy, which is the highest percentage compared to the other three techniques with only 0.02% false positive rate. There are 47,920 packets were successfully detected as 42,914 normal packets, 133 Ping of Death, 393 Smurf, 88 TCP SYN flood and 4,392 UDP flood. The second best technique is PTA-SVM, where it has achieved 99.63% detection accuracy with 0.02% false positive rate. The technique has successfully detected 47,821 packets, where it came from 42,916 normal packets, 133 Ping of Death, 298 Smurf, 85 TCP SYN flood and 4,389 UDP flood. Next, the third useful technique is PTA-LR, which it has successfully detected 42,916 normal packets, 133 Ping of Death, 220 Smurf, 84 TCP SYN flood and 4,249 UDP flood as the packet itself. Hereby, the technique has achieved 99.17% detection accuracy with 0.26% false positive rate. Meanwhile, PTA-NB has achieved 98.68% detection accuracy with 1.08% false positive rate. It has successfully detected 47,367 packets as 42,643 normal packets, 133 Ping of Death, 299 Smurf, 83 TCP SYN flood and 4,209 UDP flood.

Table 2: Performance Comparison of PTA Coupled With Four Machine Learning

TECHNIQUE	DETECTION ACCURACY	FALSE POSITIVE RATE	PACKET CLASS				
			NORMAL	PING OF DEATH	SMURF	TCP SYN FLOOD	UDP FLOOD
PTA-SVM	99.63%	0.02%	42,916	133	298	85	4,389
PTA-NB	98.68%	1.08%	42,643	133	299	83	4,209
PTA-LR	99.17%	0.26%	42,916	133	220	84	4,249
PTA-KNN	99.83%	0.02%	42,914	133	393	88	4,392

5.0 CONCLUSION

We have designed the PTA technique to detect five types of incoming packets based on the specified packet threshold and packet type. Most importantly, the technique has been combined into four machine learning techniques, they are KNN, Naïve Bayes, Logistic Regression and SVM. The technique has been tested to determine the percentage of detection accuracy and false positive rate, and it is found that the PTA-KNN technique is the best compared to the other three techniques.

REFERENCES

1. Azahari Mohd Yusof, M., Hani Mohd Ali, F., & Yusof Darus, M. (2018). Detection and Defense Algorithms of Different Types of DDoS Attacks. International Journal of Engineering and Technology, 9(5), 410–444. https://doi.org/10.7763/ijet.2017.v9.1008.
2. Caulfield, T. and A. Fielder (2015). Optimizing time allocation for network defence. Journal of Cybersecurity 1(1): 37-51.
3. Gadelrab, M., M. Elsheikh, M. Ghoneim and M. Rashwan (2018). BotCap: Machine Learning Approach for Botnet Detection Based on Statistical Features. International Journal of Communication Networks and Information Security 10: 563-579.
4. Kolahi, S. S., A. A. Alghalbi, A. F. Alotaibi, S. S. Ahmed and D. Lad (2014). Performance comparison of defense mechanisms against TCP SYN flood DDoS attack. International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT).
5. Sandeep and Rajneet (2014). A Study of DoS & DDoS – Smurf Attack and Preventive Measures, International Journal of Computer Science and Information Technology Research.
6. Gunasekhar, T., Rao, K.T., Saikiran, P., & Lakshmi, P. (2014). A Survey on Denial of Service Attacks. International Journal of Computer Science and Information Technologies.
7. Singh, U., C. Joshi and S. Singh (2017). ZDAR System: Defending Against the Unknown.International Journal of Computer Science and Mobile Computing 512: 143-149.
8. Calvert, C. and T. Khoshgoftaar (2019). Impact of class distribution on the detection of slow HTTP DoS attacks using Big Data. Journal of Big Data 6: 1-18.
9. Alqahtani, S. M., M. A. Balushi and R. John (2014). An Intelligent Intrusion Detection System for Cloud Computing (SIDSCC). 2014 International Conference on Computational Science and Computational Intelligence.
10. Badis, H., G. Doyen and R. Khatoun (2014). Understanding botclouds from a system perspective: A principal component analysis. 2014 IEEE Network Operations and Management Symposium (NOMS).
11. Arora, S and S. Dalal (2019). DDoS Attacks Simulation in Cloud Computing Environment. International Journal of Innovative Technology and Exploring Engineering (IJITEE), 9(1),414–417.
12. Sahi, A., D. Lai, Y. Li and M. Diykh (2017).An Efficient DDoS TCP Flood Attack Detection and Prevention System in a Cloud Environment. IEEE Access 5: 6036-6048.
13. Al-mamory, S. O. and Z. M. Algelal (2017). A modified DBSCAN clustering algorithm for proactive detection of DDoS attacks. 2017 Annual Conference on New Trends in Information & Communications Technology Applications (NTICT).
14. Peraković, D., M. Periša, I. Cvitić and S. Husnjak (2016). Artificial neuron network implementation in detection and classification of DDoS traffic. 2016 24th Telecommunications Forum (TELFOR).
15. Li, C., J. Yang, Z. Wang, F. Li and Y. Yang (2015). A Lightweight DDoS Flooding Attack Detection Algorithm Based on Synchronous Long Flows. 2015 IEEE Global Communications Conference (GLOBECOM).