Entropy-score: A Method to Detect DDoS Attack and Flash Crowd

Akshat Gaurav
Department of Computer Engineering
National Institute of Technology, Kurushatra
Haryana, INDIA
akshat_gaurav@yahoo.co.in

Awadhesh Kumar Singh
Department of Computer Engineering
National Institute of Technology, Kurushatra
Haryana, INDIA
aksingh@nitkkr.ac.in

***Abstract*—Nowadays the Internet plays a vital role in the growth of the economy for any nation. DDoS attacks are one of the major threat that hurting this growth as it affects the systems and network which uses the Internet for their business work. In DDoS attacks, victims bandwidth is flooded with the excessive amount of malicious or fake traffic due to which, the victim is unable to serve the legitimate users. There have been many different techniques proposed by the researchers which can detect DDoS attack efficiently. But they have many limitations and one of the important limitation of these techniques is their inability to differentiate flash crowd from DDoS attacks. Flash crowd is a scenario in which plenty of legitimate users tries to access a common server or system, so filtering of this kind of traffic may lead to business loss or credibility loss of the victim. In this context, we proposed a new detection method, Entropy-score. Which uses a hierarchical structure to analysis the incoming packets. In the proposed approach first, the entropy-based method is used for characterizing the incoming packets and then packet score based method is used for filtering the malicious packets. We implement this proposed method by using OMNET++ simulation tool and the experimental results show that Entropy-score method not only differentiates DDoS attacks traffic from Flash crowd but can also differentiate the attack traffic from the normal traffic.**

Keywords- DDoS attack, Packet score, Entropy, Flash crowd

I. Introduction

DDoS attacks [1] are one of the attacks which have been affecting the IT industry for past many years. Attackers are using different attack tools like trinoo [2], Mstream [3], Shaft [4], Kinght [4] or botnets [5], and worms [6] to generate malicious traffic which can affect the victim's system or its network. Fig.1 represents basic DDoS attack architecture. In general, there are many defense techniques for DDoS attacks like attack prevention i.e. Honeypots [7], attack filtering i.e. ScoreForCore [8], attack trace back i.e. Probabilistic packet marking technique [9]. All the perversely defined methods are using specific characteristics of incoming packets to identify the DDoS attacks. But, nowadays attackers easily imitate these characteristic and hence, bypass the filtering methods. For example, attackers spoof the source IP address of the malicious packets so that the different Ingress/Egress filters [10] not able to differentiate between malicious packets and legitimates packets; sometimes attackers manipulate the TTL value of the malicious packets to fool the hop count filters [11]; sometimes attackers also tries to impersonate the characteristics of flash crowd [12] to bypass the detection filters. Flash crowds is a particular situation in which many users

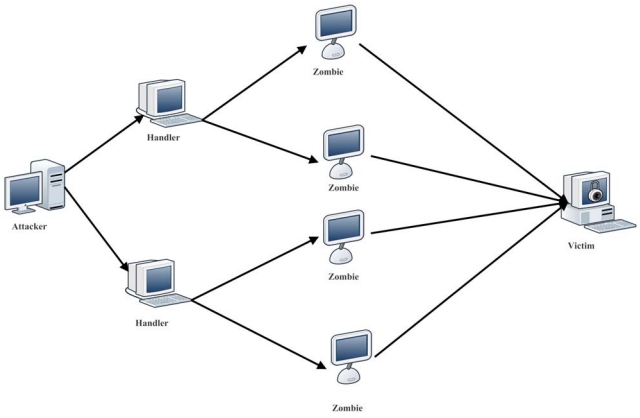


Figure 1: DDoS attack Architecture

simultaneously access one specific service like web server which in turn reduces the efficiency of the service.

The basic characteristics of DDoS attack and flash crowds are same due to which it is difficult to differentiate both of them [12] . In flash crowd and DDoS attacks, a lot of request packets are generated which leads to slow response or packet drop at the destination point. Jung et al.,2002 gives different methods to differentiate flash crowds from DDoS attacks but these methods are not applicable to all attack scenarios. Different detection methods based on entropy calculation [13] were not able to differentiate flash crowds from DDoS attacks.

In this paper, two different techniques i.e. packet scoring [14] and entropy based method [15] are used to filter the incoming packets. Packet scoring method is used to analysis each packet or group of packets according to their score values. And entropy based method is used to analysis the entire group of packets according to their entropy values. So by using these two techniques malicious packets from the group of incoming packets can be identified and deleted more efficiently. More details are given in sectionIII. The major contributions of this paper are as follows:

- This paper uses packet score method, and entropy based method to detect DDoS attacks, and for differentiating flash crowds from DDoS attacks.
- The proposed approach is applicable to any filter in the network.
- The proposed approach is attack independent i.e applicable at any type DDoS flooding attack.

The remaining paper is organized as follows. Section II

978-1-5090-3704-9/17/$31.00 ©2017 IEEE

presents the related work in recent years to detect DDoS attacks and flash crowds. In Section III, proposed approach is explained in detail. Result and discussion is present in Section IV. Finally, Section V concludes the paper and discuss the future work.

II. Related Work

The previous section briefly explains different detection methods for DDoS attack and flash crowd. This section gives some details about some of the recent and important detection techniques.

Xi Qin et. al. [16] gives the approach to detect DDoS attacks in high speed networks. For fast analysis the incoming packets are arranged in different groups according to their packet size due to this processing time of filter is reduced, after this entropy of each packet is calculated. It uses K-means clustering algorithm to model the incoming patterns, in this period different clusters with center C and radius R are formed. At attack time incoming packets are again grouped in these clusters but if any packet does not belong to any of the clusters then it is treated as a malicious packet. The main limitation of this approach is that proposed model required frequent training.

Monika Sachdev et. al. [17] It compares the entropy of an incoming packet with the average entropy and also from the average entropy of the group which generates it. So by analyzing the entropy of individual packet and generating cluster, it is decided that the packet is malicious of not. The main limitation of this scheme is that it uses real-time data to train the model and due to two layer of comparison the processing speed of the model is reduced.

Xiao et. al. [18] tries to find the correlation between the flow of generated data and its source, as if data is generated from software controlled bots then it shows a high degree of correlation. So if this correlation is analyzed then malicious packets can be detected. But if the attacker uses different software to generate the attack traffic then this approach did not work.

Shui et. al. [19] uses the Sibson distance between two incoming traffic and if this distance is more than a predefined threshold then the incoming traffic is marked as attack traffic or flash crowd. But it cannot detect DDoS attack and flash crowd at the same time.

Ke li et. al. [20] uses probability matrix to differentiate the DDoS attacks from the flash crowd. It calculates the variation coefficient and similarity coefficient of the incoming traffic and if these coefficients do not fulfill the desired criteria then respective flow is discarded.

Theerasak Thapngam et. al. [21] uses the correlation between the different arrival rates to differentiate DDoS attacks and flash crowd, for this it calculates Pearsons correlation coefficient. But for this method, the probability distribution of incoming traffic should be known and this also not able to detect DDoS attack and flash crowd at the same time.

III. Proposed Approach

This proposed method is based on proactive and individual filtering method. So each filter can filter out the malicious traffic at the early stage of the attack. Each filter uses three different steps to analysis the incoming packets which are given in Fig. 2

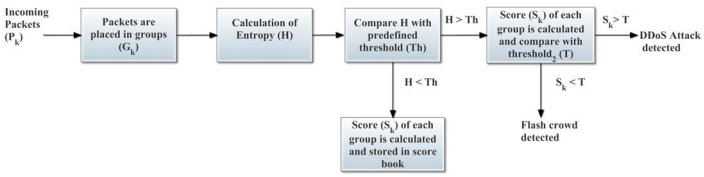


Figure 2: Framework of Proposed Approach

- Parameter extraction - The main aim of our proposed approach is to differentiate different incoming packets. So we use IP address of the host, which generates the packet as our analysis parameter. So firstly, the source IP address is extracted from each incoming packet and then these packets are arranged in different groups according to that IP address. Each packet is placed in a specific group if its distance from the centroid of that respective group is less than some predefined value. If no such group is present then a new group is formed to store that packet.
- Entropy calculation- Entropy [15] of each group is calculated. Entropy is the measure of the randomness of the samples, and it calculated by using equation 1.

$$Entropy(H) = -\sum_{X}\{p(x_i)\log_2(p(x_i))\} \quad (1)$$

where $p(x_i)$ is the probability of occurrence of i^{th} sample and X is the total sample space. During normal scenario source IP addresses are not very scattered, so for no attack period entropy is less as compare entropy during to DDoS attacks, in which attacker uses IP spoofing so source IP addresses are random in nature. But in flash crowd source IP addresses are completely random so its entropy is high as compare to the entropy of other two scenarios as given in equation 2.

$$H_{no\ attack} < H_{DDoS\ attack} < H_{Flash\ crowd} \quad (2)$$

- Packet scoring- During normal operation packet score of each group is calculated according to the packets scoring method, as given in equation 3

$$Score\ (S) = \frac{p(x_i)}{p(y_i)} \quad (3)$$

Where $P(x_i)$ is the current probability of occurrence of the group and $P(y_i)$ is the stored probability of the occurrence of the group.
If the entropy less than the threshold value then these scores are used to update the score book. But if entropy is more than the threshold value then score of the group is compare with the stored score in the score book and packets in that group are marked normal or suspicious according to set threshold.
- Threshold calculation- As explained in above step to mark the incoming traffic as; DDoS attack, flash crowd or normal flow, proper threshold selection is important. In our proposed approach threshold value in entropy based method for making the selection weather score book is updated or not is fix and it is given by 10%

of Th. Where Th is a predefined threshold value for normal operation. For packet score method threshold is calculated by load shedding algorithm. According to load shedding algorithm, the threshold value is given as follows:

$$T = 1 - \frac{\psi}{\phi} \tag{4}$$

Where ψ is the probability of current traffic and ϕ is the probability of excitable traffic.

A. Description Of Algorithm

This section explains the algorithm that is going to be executed in the filters for analyzing the incoming packets and then filtering them. Before the explanation of the algorithm, Table I defines the different attributes that are used in the algorithm.

TABLE I: Attributes used in Algorithm

Terms	Explanation
G_i	i^{th} group
p_i	Probability of i^{th} packet
H	Entropy of the system
Th	Threshold for Entropy
S_i	Score of G_i^{th} group
T	Threshold for Packet Score

Algorithm 1: Algorithm to Analysis Incoming Packets

Input : Incoming Packets
Output : Weather Packet is Malicious or Legitimate
Start
for Every incoming packet (P_k) **do**
  Place the packet in the appropriate group according to the distance of its source IP address from the group center;
  Calculate the Entropy (H);
  if $H < Th$ **then**
    Score (S_i) of each group is calculated;
    Update the score values in score book;
  end
  else
    Score (S_i) of the group (G_i) is calculated;
    Compare the score with $threshold_2(T)$;
    if $S_i < T$ **then**
      Flash crowd detected;
      Packets in the group (G_i) are Legitimate ;
    end
    else if $S_i > T$ **then**
      DDoS attack detected;
      Packets in group (G_i) are Malicious;
    end
  end
end
END

According to equation 1, every incoming packet (P_k) is placed into its respective group (G_i), according to its distance from the group center. After grouping, the entropy of the whole system (all groups together) is calculated by equation 1. If this entropy is less than the predefined threshold then the packet score of each group is calculated by using equation 3 and this score value is stored in the score book for next evaluation process. But if the entropy value of the system is more than the predefined threshold then the calculated scores of each group are again compared with threshold$_2$ (T), which is calculated by equation 4, and if their score values are more than the threshold$_2$ (T), then the packets of that group are discarded (DDoS Attack detected). But if score value of a group is less than the threshold$_2$ (T) then packets of that group are forwarded uninterruptedly (normal traffic of flash crowd).

Algorithm 1 has a time complexity of O(n) and space complexity of O(g), where n and g are the number of packets received and the number of groups formed during measuring time interval.

IV. Result And Discussion

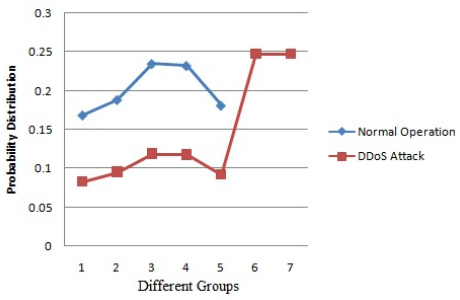
The proposed approach is implemented with the help of OMNET++ simulator [22]. Which is a discrete event simulator. Due to the discrete nature of OMNET++, it is easy to analysis each and every process in the simulation. Moreover, it supports common network models like IPv4 and, IPv6, also in this it is easy to implement new protocols as compare to other available simulators.

In OMNET++ a small network is created with 50 different hosts, but 15 new hosts join the network during DDoS attack scenario, and during Flash crowd scenario 40 new hosts joins the network and each host have a unique IP addresses. Table II represents the summary of these scenarios.

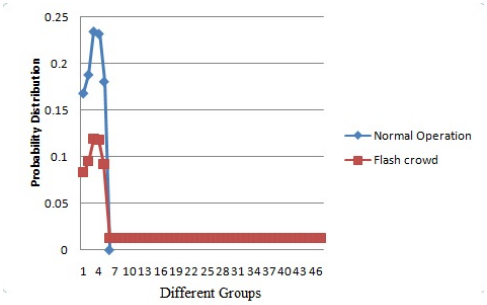
TABLE II: Simulated Output of Proposed Approach

Different Scenarios	Simulated Environment		
	Number of hosts	Number of Groups	Packet Generated
Normal Operation	50	5	409
DDoS Attack	15	7	809
Flash Crowd	40	47	809

- Step 1: Put the incoming packets in different groups according to their source IP address. In this simulation, five different groups are formed during a normal scenario. But during DDoS attack scenario, fifteen new hosts are added to the system due to which the number of groups increased to seven. Moreover, during flash crowd, forty random users added to the network which increased the number of groups to forty-seven. The probability distribution for each group during normal operation, DDoS attack and Flash crowd is represented in Fig. 3. So from Fig. 3(a) and Fig. 3(b) it is clear that during normal operation most of the packets are get grouped in first five groups but during DDoS attack scenario, more number of packets is grouped in newly formed groups and during the flash crowd scenario all the packets are randomly distributed in all the groups.
- Step 2: For each situation i.e. normal operation, DDoS attacks scenario and, flash crowd, we calculate the entropy value of the samples in a time window of 30 seconds. Fig. 4 represents the entropy values for different scenarios. From the Fig. 4, it is clear that if the entropy value of the current situation is less than the threshold (Th) then the respective traffic flow is considered as normal flow but if the entropy is more than the threshold value then the flow is considered as suspicious and it need further analysis.



(a)



(b)

Figure 3: Probability distribution Between (a) Normal and DDoS attack Scenario, (b) Normal and Flash Crowd Scenario

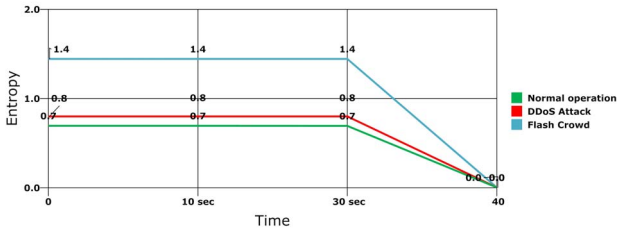


Figure 4: Entropy values during Normal operation, DDoS attack and Flash crowd scenarios

- Step 3: Each packet gets the packet score according to the frequency of the group to which it belongs. Packet score is calculated by equation 3. If some group is newly created then its stored probability is considered as 0.01. Score value during DDoS attack and Flash crowd scenarios are represented in Fig. 5. So it is clear from Fig.5 that there is a large difference between the score value of each group during DDoS attack and flash crowd scenario, hence these two events are easily differentiated by the threshold value.

A. *Result comparison*

Algorithm 1 has a time complexity of O(n) and space complexity of O(g), where n and g are the numbers of packets received and the number of groups formed during measuring time interval. But all the previously explained detecting methods in related work has space complexity of 2O(n), because they have to store data for each incoming packet. Moreover, our proposed approach uses packet score method for filtering, so its false positive (FP) rate is nearly equal to zero. This makes make our proposed approach

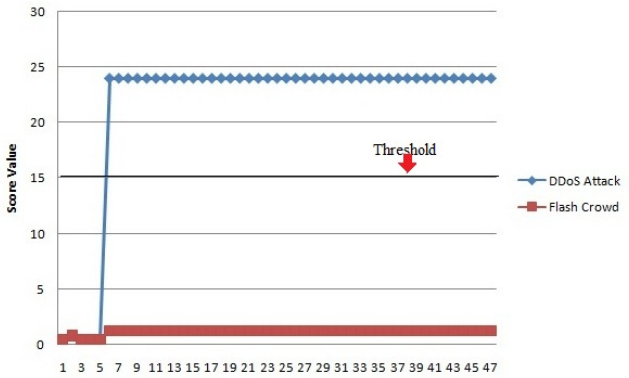


Figure 5: Score values during DDoS attack and Flash crowd scenarios

more accurate and sensitive. This comparison is represented in table III.

TABLE III: Result Comparison

Different Detection Methods	Qualities Considered			
	Detects DDoS attack	Detect Flash Crowd	Space Complexity	False Positive Rate
Packet Score [14]	YES	NO	O(g)	LOW
Entropy Based Methods [16-21]	YES	YES	2O(n)	High
Entropy-score	YES	YES	O(g)	LOW

Where ‘g’ is the number of groups formed for calculation of packet score and ‘n’ is the number of packets received during analysis period and $g \ll n$.

So from above discussion, we can say that our proposed approach is more accurate than previously define techniques [16-21] and also requires less storing space in the filter.

V. Conclusion and Future Work

In this paper, a new approach ‘Entropy-score’ was proposed, which uses an entropy-based method and packet score based method to detect the DDoS attack and also differentiate flash crowd from DDoS attack. In Entropy-score approach at first, the entropy of entire time window is calculated and compared with the set threshold to differentiate DDoS attack from the flash crowd or normal operation and then, the score of each group is calculated and compared with the other threshold values to filter the malicious traffic. Through simulation results, we showed that our proposed approach clearly differentiate DDoS attack from the flash crowd and normal flow. In addition to that, it is also capable of filtering the required malicious packets. Moreover, our proposed approach decrease false positive rate, so it increased detection sensitivity.

Our future work will be focus on testing the proposed approach in the real-time environment, as well with more attack scenarios.

Reference

[1] J. Mirkovic and P. Reiher, “A taxonomy of ddos attack and ddos defense mechanisms,” *ACM SIGCOMM Computer Communication Review*, vol. 34, no. 2, pp. 39–53, 2004.

[2] D. Dittrich, “The dos projects trinoodistributed denial of service attack tool,” 1999.

[3] D. Dittrich, G. Weaver, S. Dietrich, and N. Long, “The mstream distributed denial of service attack tool,” *URL http://staff. washington. edu/dittrich/misc/mstream. analysis. txt*, vol. 3, 2000.

[4] B. B. Gupta, R. C. Joshi, and M. Misra, "Distributed denial of service prevention techniques," *arXiv preprint arXiv:1208.3557*, 2012.
[5] M. Abu Rajab, J. Zarfoss, F. Monrose, and A. Terzis, "A multifaceted approach to understanding the botnet phenomenon," in *Proceedings of the 6th ACM SIGCOMM conference on Internet measurement*. ACM, 2006, pp. 41–52.
[6] U. S. CERT, "w32/mydoom. b virus, united states computer emergency readiness team," 2004.
[7] N. Weiler, "Honeypots for distributed denial-of-service attacks," in *Enabling Technologies: Infrastructure for Collaborative Enterprises, 2002. WET ICE 2002. Proceedings. Eleventh IEEE International Workshops on*. IEEE, 2002, pp. 109–114.
[8] K. Kalkan and F. Alagöz, "A distributed filtering mechanism against ddos attacks: Scoreforcore," *Computer Networks*, vol. 108, pp. 199–209, 2016.
[9] Y. Bhavani and P. N. Reddy, "An efficient ip traceback through packet marking algorithm," *International Journal of Network Security and Its Applications, IJNSA*, pp. 132–142, 2010.
[10] D. Senie and P. Ferguson, "Network ingress filtering: Defeating denial of service attacks which employ ip source address spoofing," *Network*, 1998.
[11] C. Jin, H. Wang, and K. G. Shin, "Hop-count filtering: an effective defense against spoofed ddos traffic," in *Proceedings of the 10th ACM conference on Computer and communications security*. ACM, 2003, pp. 30–41.
[12] J. Jung, B. Krishnamurthy, and M. Rabinovich, "Flash crowds and denial of service attacks: Characterization and implications for cdns and web sites," in *Proceedings of the 11th international conference on World Wide Web*. ACM, 2002, pp. 293–304.
[13] G. Carl, G. Kesidis, R. R. Brooks, and S. Rai, "Denial-of-service attack-detection techniques," *IEEE Internet computing*, vol. 10, no. 1, pp. 82–89, 2006.
[14] Y. Kim, W. C. Lau, M. C. Chuah, and H. J. Chao, "Packetscore: a statistics-based packet filtering scheme against distributed denial-of-service attacks," *IEEE transactions on dependable and secure computing*, vol. 3, no. 2, pp. 141–155, 2006.
[15] C. E. Shannon, "A mathematical theory of communication," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 5, no. 1, pp. 3–55, 2001.
[16] X. Qin, T. Xu, and C. Wang, "Ddos attack detection using flow entropy and clustering technique," in *Computational Intelligence and Security (CIS), 2015 11th International Conference on*. IEEE, 2015, pp. 412–415.
[17] M. Sachdeva, K. Kumar, and G. Singh, "A comprehensive approach to discriminate ddos attacks from flash events," *Journal of Information Security and Applications*, vol. 26, pp. 8–22, 2016.
[18] P. Xiao, W. Qu, H. Qi, and Z. Li, "Detecting ddos attacks against data center with correlation analysis," *Computer Communications*, vol. 67, pp. 66–74, 2015.
[19] S. Yu, T. Thapngam, J. Liu, S. Wei, and W. Zhou, "Discriminating ddos flows from flash crowds using information distance," in *Network and System Security, 2009. NSS'09. Third International Conference on*. IEEE, 2009, pp. 351–356.
[20] K. Li, W. Zhou, P. Li, J. Hai, and J. Liu, "Distinguishing ddos attacks from flash crowds using probability metrics," in *Network and System Security, 2009. NSS'09. Third International Conference on*. IEEE, 2009, pp. 9–17.
[21] T. Thapngam, S. Yu, W. Zhou, and G. Beliakov, "Discriminating ddos attack traffic from flash crowd through packet arrival patterns," in *Computer Communications Workshops (INFOCOM WKSHPS), 2011 IEEE Conference on*. IEEE, 2011, pp. 952–957.
[22] A. Varga and R. Hornig, "An overview of the omnet++ simulation environment," in *Proceedings of the 1st international conference on Simulation tools and techniques for communications, networks and systems & workshops*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2008, p. 60.