Journal of Parallel and Distributed Computing 152 (2021) 79–87

Contents lists available at ScienceDirect

J. Parallel Distrib. Comput.

journal homepage: www.elsevier.com/locate/jpdc

Discriminating flash crowds from DDoS attacks using efficient thresholding algorithm

Jisa David [a,*], Ciza Thomas [b]

[a] Rajagiri School of Engineering and Technology, Kerala, India[1]
[b] Directorate of Technical Education, Government of Kerala, India

ARTICLE INFO

Article history:
Received 4 February 2020
Received in revised form 27 December 2020
Accepted 20 February 2021
Available online 28 February 2021

Keywords:
Network security
Tsallis entropy
DDoS attack

ABSTRACT

Distributed Denial-of-Service attacks have been a challenge to cyberspace, as the attackers send a large number of attack packets similar to the normal traffic, to throttle legitimate flows. These attacks intentionally disrupt the services offered by the systems resulting in heavy cost. A flash crowd or flash event is an unexpected surge in the number of visitors to a particular website resulting in a sudden increase in server load. Flash crowds, which are legitimate flows, are difficult to be discriminated from Distributed Denial-of-Service attacks that are illicit flows. Effective and accurate detection of Distributed Denial of Service attacks still remains a challenge due to the difficulty in its detection and the false alerts generated in the case of flash crowds. There is a trade off between detection rate and false positive rate. This work deals with an efficient and early detection of distributed denial of service attacks and discriminates flash crowd by considering two network traffic parameters such as packet size and destination IP address. Using these traffic features two attributes are computed and its generalized entropies are calculated. The threshold is computed using the mean value of network attributes to detect the attacks. Threshold updater can automatically adjust the threshold values according to the changes in the channel conditions. The data sets used to evaluate the performance of the proposed approach are the MIT Lincoln Laboratory DARPA data set and a data set generated in a University network. Experimental results show this research approach achieves higher detection rate and lower false positives in a much reduced processing time as compared to the existing methods.

1. Introduction

We are living in a world where we cannot imagine a single day without the Internet. Over the last few decades, the Internet has become a robust platform that significantly transformed and made progress in the Information and Communication Technology (ICT). Malicious usage, attacks and sabotage have been on the rise as more and more computers are put into use. The attacks on the Internet have become easier to implement because of the ubiquity of the Internet and the pervasiveness of easy-to-use operating systems and development environments. There are different types of network attacks, but Denial of Service (DoS) attacks and Distributed Denial of Service (DDoS) attacks have become a common problem over the last few decades.

In DoS attack, an attacker attempts to prevent legitimate users from accessing information on computer systems or network resources such as a website or web service. Thus an attacker may be able to prevent a user from accessing email, website, or online accounts. The most common and obvious type of DoS attack occurs when an attacker floods a network with a lot of bogus information thereby depleting the bandwidth or through CPU utilization. In Distributed Denial of Service attack, an attacker may use other systems to attack a target system. By targeting the security vulnerabilities or weaknesses of the target system, an attacker could take control of any system on the network. The attacker could force these systems to send a massive amount of data request to a server or send spam to a particular email address. Here the attack is distributed because the attacker is using multiple systems to launch the DoS attack.

There are lot of different DDoS attack tools on the Internet. These tools do not require any specific expertise to use. Some of them have special properties like disable or uninstall itself, when certain conditions exist. Even stealthier attacks can easily be launched by using different tools like Trin00, Stacheldraht, Metasploit, Hping, Hyenae, TFN and Tribe Flood Network 2000 (TFN2K).

Efficient DDoS attack detection has high relevance in an industrial application with IoT (Internet of Things) based embedded devices, which envisions the autonomous interaction of sensors and actuators offering all sort of smart services. However, these

* Corresponding author.
E-mail address: jisa_d@rajagiritech.edu.in (J. David).
[1] Since 1880.

https://doi.org/10.1016/j.jpdc.2021.02.019
0743-7315/

IoT devices are limited in computation, storage, and network capacity, which makes them easy to hack and compromise. Hence, the detection of DDoS attacks has high significance. In 2009, a DDoS attack was launched that disrupted the network services of most popular websites like Live Journal, Facebook, Amazon, and Twitter (Acohido & Swartz, 2009). In 2010 and 2011, more than 75,000 computer systems in 2500 organizations and 4 million computers in 100 countries respectively were affected by DDoS attacks (Li et al. 2015). According to the records in Woolf (2016), the largest DDoS attacks in the history was orchestrated in October, 2016 using a new Mirai botnet against the servers of an American company named as Dyn, that steer much of the Internet's Domain Name System (DNS) infrastructure. The most significant DDoS attack to date took place in February of 2018. This attack targeted GitHub, a popular online code management service used by millions of developers. At its peak, this attack saw incoming traffic at a rate of 1.3 terabytes per second (Tbps) disrupting the target server offering the service. Along with the extreme growth of the Internet, the high prevalence of the DoS attacks over the Internet has been the reason for the security personnel to think of efficient DDoS detection systems. DDoS attacks are increasing exponentially, and every networking or computing device connected to the Internet is potentially susceptible to such attacks. Thus, the study of DDoS attacks, and the development of techniques to accurately and reliably detect the DDoS attacks is a main area of research and development.

The aim of this research work is the efficient and early detection of distributed denial of service attacks and discriminating it from flash crowd by considering two network traffic parameters such as packet size and destination IP address. Using these traffic features two attributes are computed. Then generalized entropy of the attribute is calculated. The threshold is computed using the mean value of network attributes to detect the attacks. Threshold updater can automatically adjust the threshold value according to the changes in the channel conditions. This contributes to the improvement in the detection accuracy and the effective reduction in the false positives. A data set developed in the University laboratory and DARPA data set are used to validate the analytical results.

Our contributions are listed as follows:

1. We propose flow-based approach, where instead of considering all packets passing through a network, only aggregated details of packets of network traffic in the form of flow is considered there by reducing the amount of data to be analysed. Also, and looking only two traffic features are used to detect DDoS attack. So it reduces processing time [1,7].
2. To measure the distribution of the packet size and destination IP, Tsallis entropy is introduced. It can discriminate DDoS attacks from flash crowds.
3. Proposed approach introduces dynamic thresholding algorithm, which provides a high detection rate with less false positives.
4. We analyse the simulations using different available data sets and our own generated data sets. By comparing with existing approaches, this method achieves high detection rate with less false positives and less processing time.
5. The proposed approach does not depend on network topology, packet arrival pattern, etc. and hence avoids signature-based metric.

The rest of this paper is organized as follows: Section 2 discusses the related work. Section 3 describes the proposed approach. Section 4 provides details on simulation topology. Section 5 describes simulation results and performance analysis and Section 6 concludes the paper.

2. Related work

Many researchers have proposed different DDoS attack detection techniques. Based on the distribution difference of the packet size, Zhou et al. [30] propose an approach to distinguish two typical low-rate DDoS attacks, the constant attack and the pulsing attack, from the legitimate traffic. An Expectation of Packet Size (EPS) based approach is used to measure the distribution difference of the packet size. The probability distributions of the packet size of the constant attack and the pulsing attack are quite different compared with that of legitimate traffic. The simulations are done using real datasets to demonstrate that the false-negative rate is small. The method is independent of network topology, arrival patterns and pulse patterns of attack packets. However, the network delay caused by network congestion is not considered. Xiang et al. [28] propose an approach to detect DDoS attacks using generalized entropy metric and the information distance metric. Hoque et al. [13] propose a DDoS attack detection method using Feature Feature Score (FFSc). In this method for each sample of network traffic, similarity value using FFSc is computed, and if the score is greater than a threshold, attack alarm is generated. But this approach shows less detection efficiency because of a static threshold.

Zhang et al. [29] propose DDoS detection based on Congestion Participation Rate (CPR). This approach of CPR-based approach can achieve per-flow-level detection of Low rate DDoS attacks. When CPR threshold value is 0.63, method can achieve a 100% detection rate with a 1.625% false positive rate.

Bhuyan et al. [6] suggest several major information metrics approach such as Hartley entropy, Shannon entropy, Renyi entropy, generalized entropy, Kullback–Leibler divergence and generalized information distance measure in their ability to detect both low-rate and high-rate DDoS attack.

The researchers have suggested many effective approaches to discriminate DDoS attacks from the flash crowd. Thapngam et al. [26] propose a behaviour based detection using Pearson's correlation coefficient that can discriminate DDoS attack traffic from traffic generated by real users. A number of observing data may cause delay in detection. This approach requires more processing time.

Prasad et al. [22] propose an entropy-based approach to detect security attack. Entropy variation of network traffic is calculated to identify the flash crowd traffic, but the threshold value for detection depends on the size of the network traffic. There should be a mechanism to control the threshold according to the volume of the network traffic, otherwise it may lead to high false positives. Gupta et al. [12] discuss details of cloud security issues and associated challenges and importance of DoS and DDoS attacks in a cloud environment. Performance parameters that are used to measure the accuracy and performance of the defense systems are included.

Matta et al. [16] suggest a formal model for the aforementioned class of attacks, and devise an inference algorithm that estimates the botnet hidden in the network, converging to the true solution as time progresses. Here, the average fraction of correctly identified bots is relatively high (greater than 80%). Razumov et al. [24] create an emulation software tool that implements the developed algorithm for detecting and blocking HTTP flood attack. This method uses an unlimited number of devices in a single filtering system. It is also possible to build infrastructure by increasing the number of proxies. Praseed et al. [23] discuss the taxonomy of application layer distributed denial of service attacks. A review of the existing research directions and defense mechanisms has also been presented to bring out the different features used for detecting these attacks, and the different methods of detection.

Table 1
Summary of related work.

References	Methodology	Performance		
Bhushan et al.[5]	Flow-table sharing approach to protect the SDN-based cloud from flow table overloading DDoS attacks.	The method has low communication overhead.		
Gupta et al. [11]	Detect DDoS attack traffic in cloud environment using chaos theory. To predict the network traffic state, nonlinear time series model GARCH model is used.	**True Detection**	**False positive**	
		99.4%	0.3%	
Kamarudin et. al [14]	Correlation feature selection (CFS) together with three different search techniques known as best-first, greedy stepwise and genetic algorithm is used for intrusion detection.	**Features**	**False positive**	**Detection**
		12	0.03%	99.99%
Bhuyan et al. [6]	Suggests several major information metrics approach such as Hartley entropy, generalized entropy, Kullback–Leibler divergence and generalized information distance measure in their ability to detect both low-rate and high-rate DDoS attack for.	**Generalized metric**	**Normal traffic**	**Attack traffic**
		Entropy metric (high-rate)	0.03249	9.39344
		Entropy metric (low-rate)	0.00649	1.87868

Table 2
Summary of related work.

References	Methodology	Performance		
Zhou et al.[30]	Based on the distribution difference of the packet size, distinguish two typical low-rate DDoS attacks, the constant attack and the pulsing attack, from the legitimate traffic. An Expectation of Packet Size (EPS) based approach is used to measure the distribution difference of the packet size.		**Pulsating attack**	**Legitimate traffic**
		EPS	87.2663	756.1894
		σ^2	3.1434	690.4353
Hoque et al. [13]	DDoS attack detection method using Feature Feature Score (FFSc). In this method for each sample of network traffic, similarity value using FFSc is computed, and if the score is greater than a threshold, attack alarm is generated.	**Threshold**	**Detection accuracy**	
		0.05	99.55%	
		1	96.6 %	
Zhang et al. [29]	DDoS detection based on Congestion Participation Rate (CPR). This approach of CPR-based approach can achieve per-flow-level detection of low rate DDoS attacks.	**Threshold**	**Detection**	**False positive**
		0.63	100 %	1.625%

Few researchers mentioned reviews of paper related to discriminate DDoS attack from the flash crowd [2,3,25]. Most of the techniques use a static threshold value for detection. But network activities and user's behaviour could vary over time so that it reduces detection accuracy. These approaches are not suitable for detecting low rate DDoS attacks. Related work is summarized in Tables 1–3 which highlight the differences of existing approaches in terms of methodology and performance (see Bhushan et al. [5], Gupta et al. [11], Kamarudin et. al [14], Bhushan et al. [4]).

3. Proposed approach

In this section, we describe the proposed approach that includes computation of attributes and dynamic thresholding algorithm.

3.1. Computation of various attributes

Discrimination of DDoS attack from the flash crowd is identified by using a statistic approach. Network traffic features like packet size and destination IP address are considered to detect DDoS attack. This approach utilizes the features that are generally used for flooding based attack detection. The packet size will be the same in most flooding attacks. On the other hand, depending on requests and responses or data and acknowledgements the legitimate network traffics have different packet sizes [9]. Wireshark tool is used to extract the traffic features.

Attackers usually generate small packets or even empty packets to reduce the resources required. In contrast, the legitimate traffic are packets filled with user data, which are typically large. Hence, during DDoS attack, it is observed that the packet size will be very small. Therefore, this distribution difference between

Table 3
Summary of related work.

References	Methodology	Performance
Xiang et. al [28]	Detect DDoS attacks using generalized entropy metric and the information distance metric.	Generalized Entropy: α=2, α=5, α=10 — Reduced False positive Rate: 161.71%, 172.11%, 199.19%
Prasad et al. [22]	Entropy-based approach to detect security attack. Entropy variation of network traffic is calculated to identify the flash crowd traffic.	Generalized Entropy: α=3, α=5 — Normal traffic: 0.451, 0.376 — Attack traffic: 0.187, 0.156
Bhushan et al. [4]	Flow confidence based discrimination algorithm to distinguish between flash crowd event and DDoS attack.	Few parameters are consider for discrimination algorithm. It reduces detection accuracy. In cloud environment more resources has to consider to ensure users QoE during flash crowd events.

attack traffic and normal traffic can be measured by considering packet size as one of the traffic features to discriminate DDoS from flash crowd. One of the attributes for DDoS detection is hence derived from packet size. Attacker generally targets a machine making the destination address the same, whereas considers random source IP addresses for simultaneous distributed attack. Hence, the distribution of destination IP is concentrated, and its entropy value will be minimal.

To improve the detection accuracy, generalized entropy such as Tsallis entropy is used to compute the second attribute using the distribution of destination IP.

Entropy defines a mathematical measure of the degree of randomness in a set of data. Higher the randomness, higher will be the entropy. Hence, predictability is expected to give lower entropy. In DDoS attack, feature characteristics will become more distributed for spoofed source addresses and more concentrated for destination addresses. Entropy could capture this distributed and concentrated phenomenon. Generalized entropy metric is proposed to detect DDoS attacks.

Shannon Entropy:A distribution of probabilities $P = p_1, p_2, \ldots, p_k$, with k elements, where $0 \leq p_i \leq 1$ and $\sum_{i=1}^{k} = 1$, Shannon entropy is given by Eq. (1).

$$\bar{H}_S = \sum_{i=1}^{k} P_i log P_i \tag{1}$$

Tsallis Entropy: Given a distribution of probabilities $P = p_1, p_2, \ldots, p_k$, with k elements, where $0 \leq p_i \leq 1$ and $\sum_{i=1}^{k} = 1$, Tsallis entropy is given by Eq. (2).

$$\bar{H}_T = \frac{1 - \sum_{i=1}^{k} P_i^{\alpha}}{\alpha - 1} \tag{2}$$

where, α defines one of the parameters of entropy. Different values of α modify the relative contribution of a given event to the whole. When the entropic parameter α is equal to one, Tsallis entropy becomes same as that of Shannon entropy. Hence, Tsallis entropy is the generalized Shannon entropy.

Renyi Entropy: In information theory Alfred Renyi introduced the spectrum of Renyi entropies of order α giving an important generalization of extensive entropy (Shannon entropy). The Renyi entropy is defined by Eq. (3).

$$\bar{H}_T = log \frac{\sum_{i=1}^{k} P_i^{\alpha}}{1 - \alpha} \tag{3}$$

The generalization of Shannon entropy determines either the diversity uncertainty or randomness of a system, which is called the generalized information entropy. It is a critical metric in statistics as an index of diversity. Renyi and Tsallis are the generalized entropies. Comparing the formulas of Shannon and generalized information entropy, the high probability event can contribute more to the final entropy in generalized information entropy than in Shannon entropy while $\alpha > 1$. The low probability event can contribute more to the final entropy in generalized information entropy than in Shannon entropy while $0 < \alpha < 1$. Therefore, different final entropy values are obtained by adjusting the value of α according to the requirements. Better detection for high dispersion and concentration is possible with the Tsallis and Renyi entropies. Comparing Tsallis and Renyi entropies, it is observed that more false positives occur with higher values of threshold (more significant peaks of entropy values) with Tsallis entropy, whereas the Renyi entropy is not so sensitive to anomalies (the excess of the threshold was smaller than Tsallis) and hence less affected with false positives.

Due to the expected variations in the network activities and user behaviour, the threshold value need to be updated continuously. The dynamic threshold algorithm applied on different attributes helps in detecting DDoS attack with enhanced detection performance.

Traffic features are extracted for identifying the DDoS attacks, and the packets are grouped. Packets are classified into a cluster, which share the same destination address in every time interval Δt. Let n be the number of clusters in Δt time interval. Let $DIP(\Delta t)_1, DIP(\Delta t)_2, \ldots, DIP(\Delta t)_n$ be the frequency of occurrence of destination IP addresses in different clusters in a given time interval Δt. During the DDoS attack, the number of packets to a particular destination IP will be very high. Hence the frequency of that destination IP will be very high. $PS(\Delta t)_1, PS(\Delta t)_2, \ldots, PS(\Delta t)_n$ be the array of the packet size of each cluster in Δt as shown in Eq. (4).

$$\begin{bmatrix} PS(\Delta t)_1 \\ PS(\Delta t)_2 \\ \ldots \\ PS(\Delta t)_n \end{bmatrix} = \begin{bmatrix} len_{11} & len_{12} & \ldots & \ldots & len_{1max} \\ len_{21} & len_{12} & \ldots & \ldots & len_{2max} \\ \ldots & \ldots & \ldots & \ldots & \\ len_{i1} & len_{i2} & \ldots & \ldots & len_{imax} \\ \ldots & \ldots & \ldots & \ldots & \\ len_{n1} & len_{n2} & \ldots & \ldots & len_{nmax} \end{bmatrix} \tag{4}$$

where $len_{i1}, len_{i2}, \ldots, len_{imax}$ are the different packet size of ith cluster in any given time interval Δt. $UPS(\Delta t)_1, UPS(\Delta t)_2, \ldots, UPS(\Delta t)_n$ be the number of unique packet size of each cluster in that time interval. During the DDoS attack the number of unique packet size of each cluster will be very small, as most of the attack packets will have same size (60 bytes headers only). However, this value may not be small in the case of flash crowd. Let $I(\Delta t)_1, I(\Delta t)_2 \ldots, I(\Delta t)_n$ be the ratio of unique packet size count

to destination IP count in each cluster in a given time interval Δt. It is computed using Eq. (5). The ratio of these components will be very small during DDoS attack and this computation improves the performance of the proposed approach.

$$I(\Delta t)_1 = \frac{UPS(\Delta t)_1}{DIP(\Delta t)_1} \tag{5}$$

So let $I(\Delta t)_{min}$ be the minimum value of $I(\Delta t)_1, I(\Delta t)_2 \ldots, I(\Delta t)_n$ in a given time interval Δt as shown in Eq. (6). If this attribute value is less than the dynamic threshold, it is indicated as an attack.

$$I(\Delta t)_{min} = min(I(\Delta t)_1, I(\Delta t)_2 \ldots, I(\Delta t)_n) \tag{6}$$

To compute second attribute, we need to compute probability of occurrence of destination IP. Let $P(\Delta t)_1, P(\Delta t)_2, \ldots, P(\Delta t)_n$ be the probability of occurrence of destination IP in a given time interval, and it is given by Eq. (7).

$$p(\Delta t)_1 = \frac{DIP(\Delta t)_1}{\sum_{i=1}^{n} DIP(\Delta t)_i} \tag{7}$$

Let $H(\Delta t)_T$ be the Tsallis entropy of destination IP for given time interval Δt. It is given by Eq. (8).

$$H(\Delta t)_T = \frac{1 - \sum_{i=1}^{n} p(\Delta t)_i^{\alpha}}{\alpha - 1} \tag{8}$$

Here, when α value is less than 10, the false positives are high. On decreasing the value of α from 10, the false positives also increases. Similarly, if the α value is above 64, the false negatives are seen to increase. Hence the optimum value of α lies between 10 and 64. In our experiment, we have limited the α value between 10 and 20. Attacker generally considers random source IP addresses and destination address (target) will be the same. So, the distribution of destination IP is concentrated, and its entropy value will be minimal.

Hence $I(\Delta t)_{min}$ and $H(\Delta t)_T$ are computed for all the time intervals. NI_{min} (A1) and NH_T (A2) are the normalized attributes used for DDoS detection. If the two attributes are small compared to the threshold, it is an indication of DDoS attack. Based on the mean and variance of the each attribute, the threshold value is updated.

3.2. Algorithm to discriminate ddos attack from flash crowd

Different traffic features are used to identify and discriminate DDoS attacks from the flash crowd. Using Wireshark tool different packet header features are extracted. Based on traffic features two attributes A_1 and A_2 are calculated. The mean (μ) is calculated by using a sliding window concept. DDoS attack is detected when there is a violation of threshold.

Algorithm 1 explains the steps involved in detection process. It is a dynamic threshold algorithm applied to the two attributes of the network traffic. Since the thresholding parameter β is updated depending on traffic condition and user's behaviour. Here, the mean and variance of entropy of each attribute are computed for window size K and with overlapping interval. The threshold for each attribute is determined by mean, variance and multiplication factor β. The multiplication factor β changes in accordance with the packet condition of the network traffic used in the simulation. The attack detector is unable to detect the attacks with high value of β when the attacker sends malicious traffic with small variation in traffic at the time when the channel is stable. Due to the steady channel condition and stealthy attack pattern, the detection becomes difficult with highly set value of β. On the other hand, with a burst channel and the detector having small β, the detector will be extremely sensitive. This will lead to several false positives, resulting in an inefficient detection. The trade-off between the detection rate and the false positive rate provides experimental guidance for choosing β value in practice [20,21]. Here β is initialized with 0.5 (value chosen between 0 and 1) and threshold $Th = \mu * \beta$. If the mean value of current interval is 80% (value chosen between 50% to 90%) less than the previous interval, then β value is decreased by 0.1 otherwise it is increased by 0.1 and threshold is computed as $Th = \mu/\beta$. If the beta value reaches less than 0.3, then beta value is assigned as 0.3. If the beta value reaches greater than 07, then beta value is assigned as 0.7. This process is continued for all the intervals. DDoS attack is detected if and only if the threshold gets violated for the two attributes. This algorithm not only detects DDoS attack but also discriminates from the flash crowd. The proposed DDoS attack detection approach is shown in Fig. 1.

Algorithm 1 Proposed Detection algorithm

1: Set
   $T \leftarrow$ Sampling interval-Aggregate each traffic feature
   $N \leftarrow$ Sliding window with window size to compute mean of each attribute
   $Th \leftarrow$ Threshold
   $\beta \leftarrow$ Thresholding factor
   $\mu \leftarrow$ mean
2: Initialize $\beta = 0.5$
3: Analyse the Traffic and extract packet header features and cluster every time interval
4: Compute two Attributes A1 and A2
5: Consider first attribute A1
6: Compute μ_j of A1; with window size = N
7: Compute $Th = \mu_j * \beta$
8: **if** $A1 < Th$ **then**
9: DDoS attack
10: **else**
11: Not DDoS attack
12: **end if**
13: Compute μ_{j+1} of A1;
14: **if** $\mu_{(j+1)} < 0.8 * \mu_j$ **then**,
15: $\beta = \beta - 0.1$ (where j+1 is the next overlapping sampling interval)
16: **if** $\beta < 0.3$ **then**
17: $\beta = 0.3$
18: **end if**
19: $Th = \mu_{j+1}/\beta$
20: **else**
21: $\beta = \beta + 0.1$
22: **if** $\beta > 0.7$ **then**
23: $\beta = 0.7$
24: **end if**
25: $Th = \mu_{j+1} * \beta$
26: **end if**
27: Go to step 9
28: Execute this algorithm for the other attributes.
29: DDoS attack is confirmed if and only if the two attributes satisfy the condition for the threshold.

4. Simulation topology

Many researchers have great difficulty in obtaining proper dataset to evaluate their detection algorithms. Due to privacy issues, many datasets for the evaluation of IDSs are not shared. At the same time, the available datasets are not updated; it does not include any latest trends in attacks. Nehinbe 2011 [18] mentioned the review of challenges related to the dataset for

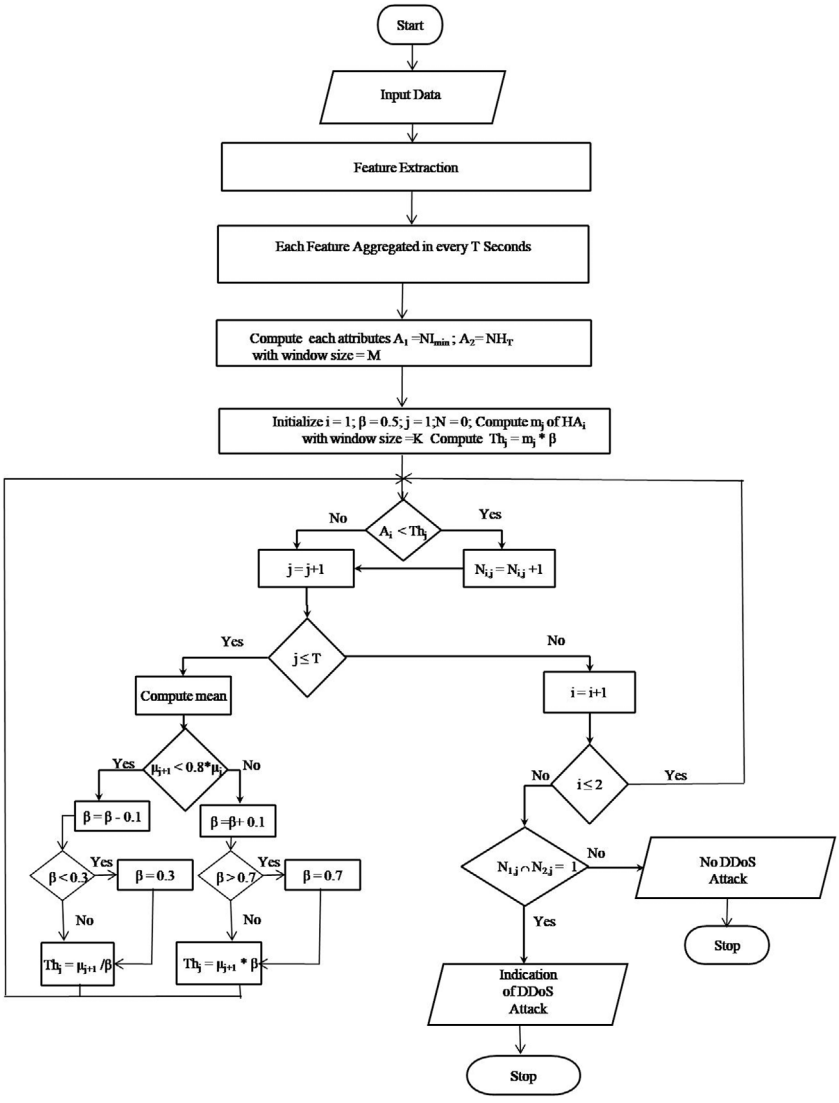

Fig. 1. DDoS detection Flow Diagram.

evaluation of intrusion detection and prevention algorithms. Koch et al. [15] encourage ISPs to cooperate on the development of realistic dataset to lay the foundation for the assessment of intrusion detection algorithms and generated real datasets.

It is very difficult to create a real world scenario of DDoS except through simulation. Omnet, Omnet++, Opnet, NS2, NS3 and Netsim are some of the tools used to create network setup mimicking real world scenario. However, these tools have the disadvantage of setting numerous parameters regarding LAN set up.

In this work a virtual network is set up using virtualbox tool. The LAN set up is shown in Fig. 2, with four virtual machines configured using virtualbox. The legitimate clients are identified as host1 and host2. Host1 is installed with Windows 8.1 and host2 is installed with Ubuntu 14.04. Attacker runs Kali OS, which is equipped with several DDoS attacking tools such as Hping3, Hyenae, and Metasploit.

Hping3 is normally used by system administrators and ethical hackers basically for pinging or for advanced tasks as it can bypass the firewall filter. It uses TCP, UDP, ICMP and RAW-IP protocols. It also has the traceroute mode and the ability to send files between covered channels. Hping3 is a network tool that is able to send custom TCP/IP packets and display the target replies like how the ping program does with ICMP replies. Hping3 handles fragmentation, with arbitrary packets to transfer files encapsulated with supported protocols. Hping3 enables one to perform at least the following stuff:

- Test firewall rules
- Advanced port scanning
- Test net performance using different protocols, packet size, TOS (type of service) and fragmentation
- Path MTU discovery
- Transferring files between even really fascist firewall rules
- Traceroute-like under different protocols
- Firewall-like usage
- Remote OS fingerprinting
- TCP/IP stack auditing

It is also a good didactic tool to learn TCP/IP. Hping3 is developed and maintained by antirez@invece.org and is licensed under GPL version 2.

Hyenae is a packet generator tool that is used to create forged packets for dumping the server. This tool is platform independent and is used as a network packet producer that can initiate DDoS, DoS and MITM attacks. This tool is easily configurable. Address patterns are used in Hyenae, reducing the number of

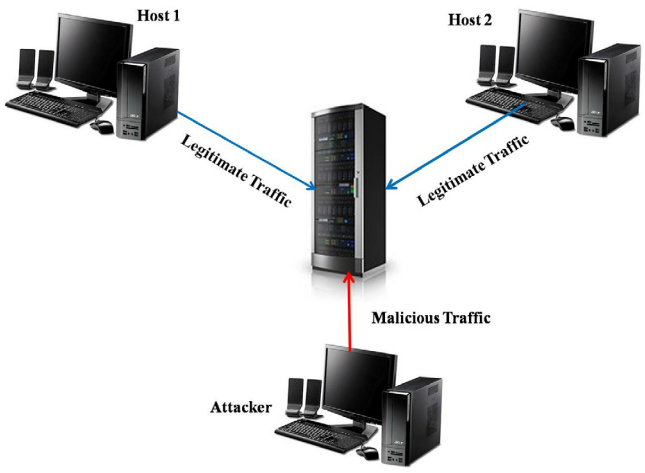

Fig. 2. System modelling.

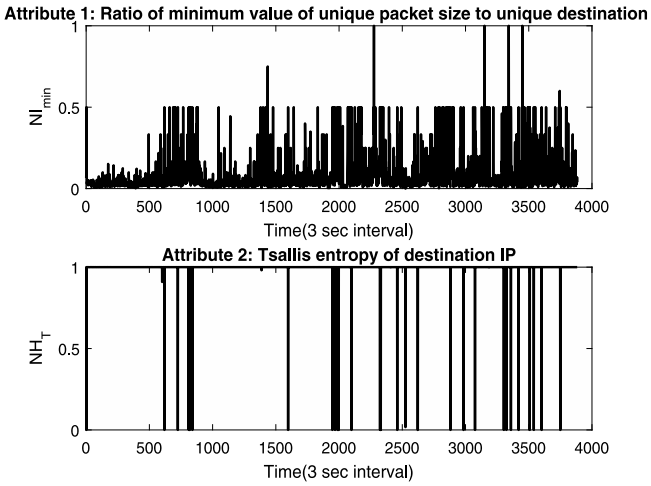

Fig. 3. Different attributes in every three seconds.

arguments because all the necessary parameters, such as the way we want to randomize addresses or the IP address version to use, can be derived from the pattern format that is provided. Besides smart wildcard-based address randomization, a highly customizable packet generation control and an interactive attack assistant, Hyenae comes with a clusterable remote daemon for setting up distributed attack networks.

Metasploit tool can trigger remote attacks by choosing and configuring an exploit code that enters a target system by taking advantage of one of its vulnerabilities. Metasploit runs on Unix and on Windows. This tool can also group zombies and is capable of triggering remote attacks.

The victim machine is having Windows12 R2 server OS and is installed with Apache web server.

4.1. Traffic generation procedure

The steps involved in generation of attack traffic are as follows:

1. Turn on virtual box.
2. Turn on all virtual machines (host1, host2, attacker and victim).
3. Run Wireshark on victim machine to capture packets.
4. Attacker uses hping3 or hyenae to launch different types of attacks (syn flood, IP spoof etc.).
5. Host1 and host2 access Apache web server (victim) by typing 192.168.101.28 (web browser of Ubuntu system) or airliss.vnet (web browser of Ubuntu system).
6. Wireshark captures network traffic (Attack and normal Traffic) from victim machine.

4.2. Wireshark tool

Wireshark is one of the most common network traffic analyser frequently used by the system administrator or security professional nowadays. Wireshark is open software. It is possible to analyse the network in real time; at the same time, this tool can troubleshoot some of the network issues. Wireshark can troubleshoot, dropped packets, malicious activity on network and latency issues. Network administrator usually uses Wireshark tool to detect damaged network appliances; it includes dropping packets, latency issues caused by machines routing traffic halfway around the world, and hacking attempts against an organization.

The generated attack traffic, DARPA 98, DARPA 2000(LLDoS1.0) as the relevance of DARPA data set is illustrated in the work of Thomas et al. [27] and CAIDA data sets are used for evaluation of the proposed method.

5. Simulation results and performance analysis

In the proposed method MIT Lincoln Laboratory DARPA 98 (Friday of week five is used in standard), DARPA 2000 (LLDoS1.0) and CAIDA data set is also used for analysis. According to Moore et al. [17], it is a high-rate attack if there are more than 10,000 packets per second over the network, with 1000 attack packets per second covering 60% of the attack traffic. By analysing packet features, DDoS attack can be detected. During DDoS attack, packet features like the destination IP address are concentrated to a single victim and variance of packet size will be small. This is because of considering packet size as one feature to distinguish flash crowd from DDoS attack.

We choose to aggregate the packets every three seconds (DARPA 2000 data set) and its execution is done on MATLAB. In the analysis, we divide the packets into clusters which share the same destination IP then compute the frequency of occurrence of destination IPs and its entropy. Packet size is another traffic feature, during DDoS attack packet size will be small.

In Fig. 3 different attributes are plotted. If the values of two attributes are smaller than the threshold value, then it indicates the DDoS attack. Dynamic threshold algorithm is used to discriminate DDoS attack from flash crowd. This adaptive algorithm is applied for each attribute. Fig. 4 shows the attack detection using different attributes with some false positives. If two attributes violate the threshold, then DDoS detection is confirmed.

Fig. 5 shows an efficient way of detection of DDoS attack ('0' indicates normal traffic and '1' indicates attack traffic). The processing time of the proposed algorithm is less as only two attributes are considered and the method is not based on any time series model.

The accuracy, precision, sensitivity and specificity are used for performance evaluation of the proposed algorithm. Detection Rate or sensitivity (TPR) measures the percentage of correctly identified attacks over all the actual attacks and is computed using Eq. (9).

$$Sensitivity(TPR) = \frac{TP}{TP + FN} \tag{9}$$

Accuracy measures the percentage of true detection over the entire traffic trace and is computed using Eq. (10).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

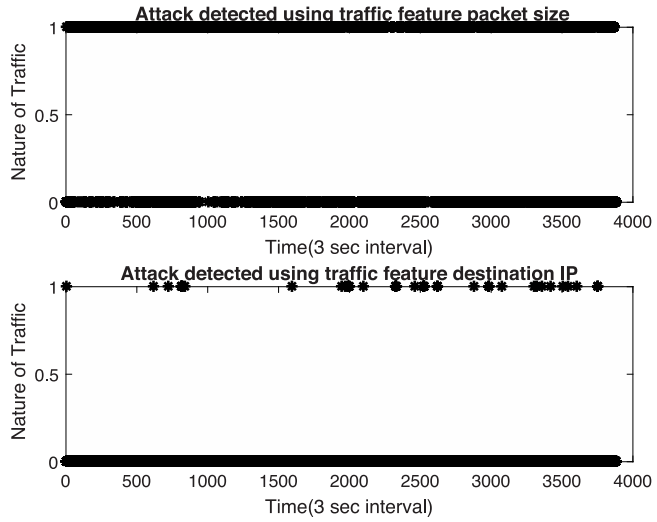

Fig. 4. Attack detection using two attributes.

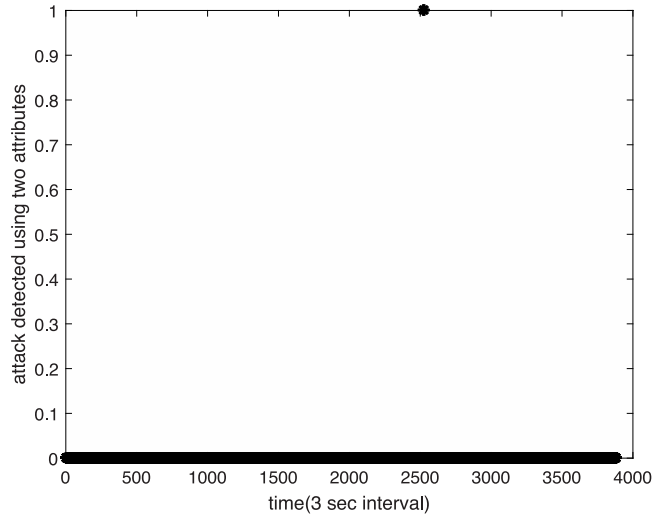

Fig. 5. DDoS attack detection.

Specificity(TNR) relates to the ability to correctly detect traffic without attack. The specificity is computed using Eq. (11).

$$Specificity(TNR) = \frac{TN}{TN + FP} \tag{11}$$

Precision is the positive predictive value or the fraction of the positive predictions that are actually positive and it is computed using Eq. (12).

$$Precision = \frac{TP}{TP + FP} \tag{12}$$

where True Positive (TP) indicates the correct predictions of the DDoS attack and False Positive (FP) refers to the normal traffic incorrectly classified as DDoS attack. True Negative (TN) refers to the normal data correctly classified as normal data. False Negative (FN) refers to the DDoS attack traffic incorrectly classified as normal traffic [10].

Table 4 shows the performance comparison of Shannon and Tsallis entropy. It shows the generalized entropy such as Tsallis entropy perform much better than Shannon entropy.

Tables 5–7 show a comparative study of existing methods and performance of the proposed method. Three data sets are considered for the evaluation. Low rate DDoS attack data are considered in CAIDA data set. DARPA 2000 contains high rate DDoS attacks. Here time interval(T) allotted DARPA 2000 and generated data set is 3 s. Generated data set is used to identify the flash crowd from DDoS attack. Two of the existing methods [8,19] with which we have compared this work fail to discriminate DDoS attack from the flash crowd and cannot detect low rate DDoS attack.

Table 4
Comparison with different types of entropy.

Entropy method	Data set	Detection	FP	Accuracy
Shannon entropy	DARPA 2000	100%	8.3%	91.66%
Tsallis entropy	DARPA 2000	100%	0%	100%

Table 5
Existing method [19].

Data set	Detection	FP	Accuracy	Processing time
DARPA 98	94.4%	0.12%	99.5%	150 s
DARPA 2000	100%	13%	86.9%	75 s
CAIDA	66.6%	0%	66.6%	35 s

Table 6
Existing method [8].

Data set	Detection	FP	Accuracy	Processing time
DARPA 98	98%	0.43%	99.5%	100 s
DARPA 2000	100%	0%	100%	65 s
CAIDA	83.3%	0%	83.3%	26 s

Table 7
Proposed method.

Data set	Detection	FP	Accuracy	Processing time
DARPA 98	96%	0.0008%	99.7%	100 s
DARPA 2000	100%	0%	100%	63 s
CAIDA	100%	0.005%	100%	23 s

6. Conclusion

Efficient DDoS attack detection has high relevance in an industrial application with IoT based embedded devices, which envisions the autonomous interaction of sensors and actuators offering all sorts of smart services. However, these IoT devices are limited in computation, storage, and network capacity, which makes them easy to hack and compromise. Hence, the detection of DDoS attacks has high significance in the present time. Here, the method is based on the distribution of packet size and the destination IP. The DDoS attack is identified using dynamic thresholding algorithm because network activities and user behaviour vary with time, and this improves the detection rate and reduces the false positive rate. This method can effectively discriminate DDoS attacks from the flash crowd with high detection rate, low false positive rate and low processing time as compared to the existing approaches. This work has focused only on the detection of flooding type of DDoS attack without considering the detection of application-layer DDoS. It also has the limitation that neither it locates nor mitigates the attack flows. As a total defense solution for the DDoS attacks, it is necessary to include the mitigation of attack flows as a necessary part of the system and such preventive systems can also be incorporated as a future work.

CRediT authorship contribution statement

Jisa David: Conception and design of study, Acquisition of data, Analysis and/or interpretation of data, Writing - original

draft, Writing - review & editing. **Ciza Thomas:** Conception and design of study, Acquisition of data, Analysis and/or interpretation of data, Writing - original draft, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

All authors approved the version of the manuscript to be published.

References

[1] H. Alaidaros, M. Mahmuddin, A. Al-Mazari, et al., An Overview of Flow-Based and Packet-Based Intrusion Detection Performance in High Speed Networks, Naif Arab University for Security Sciences, 2011.
[2] S. Behal, K. Kumar, M. Sachdeva, Characterizing DDoS attacks and flash events: Review, research gaps and future directions, Comp. Sci. Rev. 25 (2017) 101–114.
[3] S. Behal, K. Kumar, M. Sachdeva, Discriminating flash events from DDoS attacks: A comprehensive review, Int. J. Netw. Secur. 19 (5) (2017) 734–741.
[4] K. Bhushan, B.B. Gupta, A novel approach to defend multimedia flash crowd in cloud environment, Multimedia Tools Appl. 77 (4) (2018) 4609–4639.
[5] K. Bhushan, B.B. Gupta, Distributed denial of service (DDoS) attack mitigation in software defined network (SDN)-based cloud computing environment, J. Ambient Intell. Humaniz. Comput. 10 (5) (2019) 1985–1997.
[6] M.H. Bhuyan, D.K. Bhattacharyya, J.K. Kalita, An empirical evaluation of information metrics for low-rate and high-rate DDoS attack detection, Pattern Recognit. Lett. 51 (2015) 1–7.
[7] J. David, C. Thomas, Intrusion detection using flow-based analysis of network traffic, in: International Conference on Computer Science and Information Technology, Springer, 2011, pp. 391–399.
[8] J. David, C. Thomas, Efficient DDoS flood attack detection using dynamic thresholding on flow-based network traffic, Comput. Secur. (2019).
[9] P. Du, S. Abe, IP packet size entropy-based scheme for detection of DoS/DDoS attacks, IEICE Trans. Inf. Syst. 91 (5) (2008) 1274–1281.
[10] M.E. Elhamahmy, H.N. Elmahdy, I.A. Saroit, A new approach for evaluating intrusion detection system, CiiT Int. J. Artif. Intell. Syst. Mach. Learn. 2 (11) (2010).
[11] B.B. Gupta, O.P. Badve, GARCH and ANN-based DDoS detection and filtering in cloud computing environment, Int. J. Embed. Syst. 9 (5) (2017) 391–400.
[12] B.B. Gupta, O.P. Badve, Taxonomy of DoS and DDoS attacks and desirable defense mechanism in a cloud computing environment, Neural Comput. Appl. 28 (12) (2017) 3655–3682.
[13] N. Hoque, D.K. Bhattacharyya, J.K. Kalita, A novel measure for low-rate and high-rate DDoS attack detection using multivariate data analysis, in: COMSNETS, 2016, pp. 1–2.
[14] M.H. Kamarudin, C. Maple, T. Watson, Hybrid feature selection technique for intrusion detection system, Int. J. High Perform. Comput. Netw. 13 (2) (2019) 232–240.
[15] R. Koch, M. Golling, G.D. Rodosek, Towards comparability of intrusion detection systems: New data sets, in: TERENA Networking Conference, Vol. 7, 2014.
[16] V. Matta, M. Di Mauro, M. Longo, Botnet identification in randomized DDoS attacks, in: 2016 24th European Signal Processing Conference, EUSIPCO, IEEE, 2016, pp. 2260–2264.
[17] D. Moore, C. Shannon, D.J. Brown, G.M. Voelker, S. Savage, Inferring internet denial-of-service activity, ACM Trans. Comput. Syst. (TOCS) 24 (2) (2006) 115–139.
[18] J.O. Nehinbe, A critical evaluation of datasets for investigating IDSs and IPSs researches, in: 2011 IEEE 10th International Conference on Cybernetic Intelligent Systems, CIS, IEEE, 2011, pp. 92–97.
[19] S.M.T. Nezhad, M. Nazari, E.A. Gharavol, A novel DoS and DDoS attacks detection algorithm using ARIMA time series model and chaotic system in computer networks, IEEE Commun. Lett. 20 (4) (2016) 700–703.
[20] G. No, I. Ra, An efficient and reliable DDoS attack detection using a fast entropy computation method, in: Communications and Information Technology, 2009. ISCIT 2009. 9th International Symposium on, IEEE, 2009, pp. 1223–1228.
[21] G. No, I. Ra, Adaptive DDoS detector design using fast entropy computation method, in: Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS), 2011 Fifth International Conference on, IEEE, 2011, pp. 86–93.
[22] K.M. Prasad, A.R.M. Reddy, K.V. Rao, Discriminating ddos attack traffic from flash crowds on internet threat monitors (itm) using entropy variations, Afr. J. Comput. ICT 6 (2) (2013) 53.
[23] A. Praseed, P.S. Thilagam, DDoS attacks at the application layer: Challenges and research perspectives for safeguarding Web applications, IEEE Commun. Surv. Tutor. 21 (1) (2018) 661–685.
[24] P.V. Razumov, O.A. Safaryan, I.A. Smirnov, V.M. Porksheyan, N.V. Boldyrikhin, D.A. Korochentsev, L.V. Cherckesova, S.A. Osikov, Developing of algorithm of HTTP FLOOD DDoS protection, in: 2020 3rd International Conference on Computer Applications & Information Security, ICCAIS, IEEE, 2020, pp. 1–6.
[25] P.R. Reddy, R. Siva, C. Malathi, Techniques to differentiate ddos attacks from flash crowds, Int. J. Adv. Res. Comput. Sci. Softw. Eng. 3 (6) (2013).
[26] T. Thapngam, S. Yu, W. Zhou, 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.
[27] C. Thomas, V. Sharma, N. Balakrishnan, Usefulness of DARPA dataset for intrusion detection system evaluation, in: Data Mining, Intrusion Detection, Information Assurance, and Data Networks Security 2008, Vol. 6973, International Society for Optics and Photonics, 2008, 69730G.
[28] Y. Xiang, K. Li, W. Zhou, Low-rate DDoS attacks detection and traceback by using new information metrics, IEEE Trans. Inf. Forensics Secur. 6 (2) (2011) 426–437.
[29] C. Zhang, Z. Cai, W. Chen, X. Luo, J. Yin, Flow level detection and filtering of low-rate DDoS, Comput. Netw. 56 (15) (2012) 3417–3431.
[30] L. Zhou, M. Liao, C. Yuan, H. Zhang, Low-rate DDoS attack detection using expectation of packet size, Secur. Commun. Netw. 2017 (2017).

Dr. Jisa David is an assistant professor at Rajagiri School of Engineering and Technology, Kochi, Kerala, India since 2008. She received her M. Tech degree and Ph.D from Kerala University in 2008 and 2020 respectively. Her area of interests includes network security and image processing.

Dr. Ciza Thomas is currently working as Senior Joint Director, Directorate of Technical Education, Government of Kerala. Her area of expertise is Cyber Security with publications in more than 50 International Journals and International Conference Proceedings, and more than 50 national conference publications. She has edited six books and published eight book chapters. She is a reviewer of more than ten reputed International journals including IEEE transactions on Signal Processing, IEEE transactions on Neural Networks, International Journal of Network Security, International Journal of Network Management, and IEEE-John Wiley International Journal on Security and Communications Network. She is a guest editor of the IEEE Security and Privacy Magazine and International Journal of Cyber Situational Awareness. She is a recipient of achievement award in 2010 and the e-learning IT award in 2014 from Government of Kerala.