On the use of information theory metrics for detecting DDoS attacks and flash events: an empirical analysis, comparison, and future directions

Jagdeep Singh*, Navjot Jyoti, Sunny Behal
Dept. of CSE, Shaheed Bhagat Singh State University,
Ferozepur, Punjab, India
**Corresponding author: jagdeepsinghdahien@gmail.com*

Abstract

A Distributed Denial of Service (DDoS) attack is one of the lethal threats that can cripple down the computing and communication resources of a web server hosting Internet-based services and applications. It has motivated the researchers over the years to find diversified and robust solutions to combat against DDoS attacks and characterization of flash events (a sudden surge in the legitimate traffic) from HR-DDoS (High-Rate DDoS) attacks. In recent times, the volume of legitimate traffic has also magnified manifolds. It results in behavioral similarities of attack traffic and legitimate traffic that make it very difficult and crucial to differentiate between the two. Predominantly, Netflow-based techniques are in use for detecting and differentiating legitimate and attack traffic flows. Over the last decade, fellow researchers have extensively used distinct information theory metrics for Netflow-based DDoS defense solutions. However, a comprehensive analysis and comparison of these diversified information theory metrics used for particularly DDoS attack detection are needed for a better understanding of the defense systems based on information theory. This paper elucidates the efficacy and effectiveness of information theory-based various entropy and divergence measures in the field of DDoS attack detection. As part of the work, a generalized NetFlow-based methodology has been proposed. The proposed detection methodology has been validated using the traffic traces of various real benchmarked datasets on a set of detection system evaluation metrics such as Detection rate (Recall), Precision, F-Measure, FPR, Classification rate, and Receiver-Operating Characteristics (ROC) curves. It has concluded that generalized divergence-based information theory metrics produce more accuracy in detecting different types of attack flows in contrast to entropy-based information theory metrics.

Keywords: DDoS attacks; divergence; entropy; empirical investigation; information theory

1. Introduction

Over the years, the Internet has witnessed the exponential growth of online services and applications. The current statistics show that around 57% of World's population is the user of the Internet Clicks (2019). The increasing usage of these interactive web-based services and applications have also en-couraged an exponential increase in risks and pos-sibilities of misuse of the Internet. The researchers are developing sophisticated and robust solutions to design better networks of the future, such as HTTP as the narrow waist Popa *et al.* (2010), Named Data Networking (NDN) Zhang *et al.* (2010), programmable networks Campbell *et al.* (1999) and Software-Defined Networking Fundation (2012). However, the Internet is still vulnerable and open to various security threats arise due to worms, port scans, Trojans, and DDoS attacks, etc. Among several security threats, DoS (Denial of service) and DDoS (Distributed-Denial of service) attacks are the most ruthless security threats.

A DoS attack characterizes as an explicit attempt by an attacker to prevent legitimate users from accessing a website, web service, or a computer system. Such an attack is launched using millions of computer systems (if online) in a coordinated manner, called a distributed denial of service (DDoS) attack. DDoS attacks deny the target service bysending the redundant stream of packets to a victim rendering it unavailable to legitimate users. Fur-ther, attackers also send these requests at a low rate to elude the defense system Zhijun et al. (2020). The DDoS attacks can be mainly divided into two categories as

- Network DDoS attacks (Layer 3/4): These attacks target the network and transport layers. Such attacks occur when the network traffic overwhelms a network and consumes all the available resources of the network.

- Application DDoS attacks (Layer 7): These DDoS attacks target the application layer of the OSI model. In such types of DDoS attacks, the attacker exploits seemingly legitimate HTTP GET or POST requests to attack a web server or application.

As per the latest global DDoS threat landscape Q4 report of 2019 *Global Threat landscape Report* (2019), the trend has shifted from network layer DDoS attacks, which were based on spoofing, to application-layer DDoS attacks which are based on legitimate TCP connections. These attacks send multiple HTTP GET requests to over-whelm either victim resources such as CPU cycles, memory, buffers, and file descriptors or network resources such as bandwidth. Under such situa-tions, the web server spends most of its CPU time in processing useless attack packets instead of le-gitimate packets leading to a denial of service to normal users. Even the 2019 annual worldwide infrastructure security report (WISR) from Netscout has reported the volume of such attacks touched to 1.7 Tbps *Netscout WISR Report* (2019).

While DDoS attack traffic targets to bring the service down intentionally, there is a type of legitimate traffic that can also bring the target service down, such type of traffic is known as a Flash Event (FE) Jung *et al.* (2002). An FE cause surge in traf-fic volume likely to a DDoS attack when hundreds of legitimate users try to access the same website simultaneously Bhandari *et al.* (2016), Sachdeva *et al.* (2016). Sometimes because of the occurrence of some sudden events such as the launching of a sale, death of some celebrity, budgetary sessions, new product launch, etc. lead to an immense the volume of network traffic toward web servers host-ing that news. It causes delays in the web server's responses and thus, do not intend to bring the service down. Instead, these events are the result of overuse of service for which the service provider may or may not be ready before the occurrence of these events. However, there is a high risk of misinterpreting a FE as a DDoS attack or vice versa, if the proper mechanism of discrimination is not in place and hence, leading to blacklisting of legitimate IP addresses. So, it is very crucial to detect DDoS attacks and discriminate them from behaviorally similar flash events in-time to the making timely available services and applications based on the Internet.

Through the extensive survey of existing literature, we observed that the majority of the DDoS defence solutions use the concept of flow-similarity and have used information theory-based metrics as the underlying detection logic. There are several salient features of using information theory metrics such as (1) they usually have a small complexity in term of space, time, and computation as only header fields are used for calculation, (2) they have lesser need of storage, so there is no need to accumulate a large number of network traces, (3) they require lesser packet header features to differentiate various network traffic types, (4) high scalability,(5) low false-positive rates, and (6) high sensitivity towards capturing meek deviations Xiang *et al.*(2011).

In recent times, researchers have used diversified information theory-based entropy and divergence measures. Information entropy metrics such as Shannon entropy Sachdeva *et al.* (2016), Be-hal & Kumar (2017*a*), Singh *et al.* (2020), Renyi's generalized entropy Xiang *et al.* (2011), Behal & Kumar (2017*a*,*b*), Bereziński *et al.* (2015) - Sa-hoo *et al.* (2018), Tsallis entropy Bereziński *et al.* (2015), Ma & Chen (2013), Basicevic *et al.* (2015), ϕ-entropy Behal & Kumar (2017*b*), Basicevic & Ocovaj (2019) and divergence based metrics such as KL divergence Behal & Kumar (2017*a*,*b*), Sa-hoo *et al.* (2018), Li *et al.* (2009), Hellinger dis-tance Jeyanthi & Iyengar (2012), Yu *et al.* (2009) - Saravanan *et al.* (2016), generalized information divergence Xiang *et al.* (2011), Behal & Kumar (2017*a*,*b*), Bhuyan *et al.* (2016), Li *et al.* (2009), Total variation distance Li *et al.* (2009), Rahmani *et al.* (2012*a*,*b*) and Jeffrey distance Behal & Kumar (2017*a*), Yu *et al.* (2009), Yu *et al.* (2011) have been used extensively in the literature to detect the network anomalies. In this paper, we have presented a comprehensive and exhaustive empirical analysis of predominantly used five distinct entropy measures and ten divergence measures.

The major contributions of this paper can be summarized as:

- A Netflow-based generalized detection methodology is proposed. It uses various mathematical models to classify network traffic into different types of Netflows such as normal, HR-DDoS attack, LR-DDoS attack and FEs.
- A scalable emulation-based DDoSTB testbed has been used to validate the proposed generalized detection algorithm by replaying the traffic traces of benchmarked datasets of MIT Lincoln (for Normal traffic), FIFA (for FE traffic), CAIDA (for LR-DDoS and HR-DDoS traffic), and DDoSTB (Normal, HR-DDoS and FE traffic) to generate various types of normal and attack Netflows. A scalable emulation-based DDoSTB represents near to real high-speed network traffic for the validation purpose.
- The efficiency and effectiveness of various en-tropy and divergence measures are measured using detection systems evaluation parameters such as detection rate (recall), precision, false positive rate (FPR), F-measure, and classification rate.
- A comprehensive evaluation of the detection efficiency of each entropy and divergence metric is performed by visualizing the tradeoff between detection rate and FPR in terms of ROC (receiver operating characteristic) curves.

The Rest of the paper is organized in the following way. The background of information theory metrics as well as related work, is given in Section 2. Section 3 explains generalized proposed detection methodology, and experimental setup details are given in Section 4. In Section 5, the results are compared and discussed, and finally, Section 6 concluded the work by highlighting future directions.

2. Background of information theory and related work

DDoS attacks cause significant deviations in the packet header features of network traffic. Recently, information theory-based metrics such as entropy and divergence measures, have been used progressively in detecting DDoS attacks.

2.1 Background of information theory

Information theory is based on probability theory and statistics. Information theory often concerns itself with the measure of information of probability distributions associated with random variables. This section briefly explains the present prominently used information theory-based entropy and divergence measures.

2.1.1 Information entropy measures

Information entropy is a measure of the uncertainty associated with a random variable—the more random the information variable, the bigger the en-tropy. In contrast, the greater certainty of the infor-mation variable, the smaller the entropy. Initially, in 1948, Claude Shannon defined information en-tropy metric to calculate the uncertainty, random-ness or disorder in the physical system. Mathematically, Shannon entropy Shannon (2001) is given by:

$$H(x) = -\sum_{i=1}^{n} p_i log_2 p_i \tag{1}$$

where p_i denotes the probability of the occurrence of an event x. Further, Alfred Renyi, in 1961, defined a more general form of Shannon entropy to quantify the arbitrariness in a system. Renyi entropy has many applications in statistical and coding theory statistics and is known as an index of diversity. The generalized information entropy (GE) of order α is defined as:

$$H_\alpha(x) = \frac{1}{1-\alpha} log_2 \left(\sum_{i=1}^{n} p_i^\alpha \right) \tag{2}$$

where α is an entropic index. Using different order of α values, GE can compute different contributions of the numerous proportions of probability distributions. When $\alpha < 0$, generalized entropy is more sensitive to the events which are having lesser frequency whereas when $\alpha \geq 0$, GE is more sensitive to occur events frequently. It means when $\alpha \geq 0$, events with high probability contribute more to GE and when $\alpha < 0$ events with low probability contribute more in GE Rényi (1965). For exam-ple, when $\alpha = 0$, Hartley entropy can be obtained which gives the maximum value of information and is defined as

$$H_0(x) = log_2 n \tag{3}$$

When $\alpha \to 1$, GE tends to Shannon entropy.

Plastino in 1993 defined Tsallis entropy as another generalized form of entropy Plastino & Plastino (1993). It is also called non-extensive entropy Tsallis (1988) and is a one-parameter generalization of traditional Shannon entropy. When $\alpha \to 1$, Tsallis entropy tends to Shannon entropy. Like GE, on different values of α Tsallis entropy gives diverse contributions of a probability distribution.

$$H'(x) = \frac{1}{1-\alpha}\Big(1 - \sum_{i=1}^{n} p_i^{\alpha}\Big) \qquad (4)$$

In 2009, Ubriaco (2009) proposed a new fractional entropy having similar properties as of Shannon entropy but it does not follow the additivity property like Shannon entropy Machado (2010). Ubriaco entropy is defined as:

$$H(x) = \sum_{i=1}^{n} \Big(- log\Big(p(z_i)\Big)\Big)^{q} p(z_i) \qquad (5)$$

Further, Bhatia *et al.* (2012) in 2012 proposed a new set of metrics based on information theory. Authors claimed that their given metrics have better differentiation power than existing Renyi's GE and GID. They defined a new generalized entropy as ϕ-entropy:

$$\phi - entropy = -\frac{1}{sinh(\alpha)}\left(\sum_{i=1}^{n}\Big(p_i sinh(\alpha log_2 p_i)\Big)\right)(6)$$

ϕ-entropy is an improved version of Shannon entropy and Renyi's GE. It's adjustable behavior is more acceptable to discover attack behavior and pattern in the early stage.

2.1.2 Divergence measures

A plethora of divergence measures is available in the literature which can be used to measure the quantitative difference among two different probability distributions. Initially, Pearson in 1900 de-fined a Pearson divergence, also known as Chi-square, χ^2-divergence, Quadratic divergence, Kagen divergence, least-squares Pearson (1900) as:

$$Pearson(P||Q) = \frac{1}{2}\sum_{i=1}^{n}\frac{\big(p_i - q_i\big)^2}{q_i} \qquad (7)$$

where P and Q are probability distributions. χ^2 divergence can range from 0 to ∞. χ^2 is 0 iff P and Q are equal. It increases as distributions become dis-similar and reach to ∞ when $P \neq Q$. For normal traffic, χ^2 function must be close to 0 and shows large deviation when distributions change. Then, Ernst Hellinger introduced the concept of Hellinger distance in 1909 in terms of the Hellinger integral Hellinger (1909). It is defined as:

$$HD(P||Q) = \left(\sum_{i=1}^{n}\Big(\sqrt{p_i} - \sqrt{q_i}\Big)^2\right)^{1/2} \qquad (8)$$

where Hellinger Distance (HD) satisfies the inequality 0≤ $HD(P||Q) \leq 1$, and $HD(P||Q)$ = 0 iff P = Q. HD is a symmetric distance i.e. $HD(P||Q) = HD(Q||P)$. Further, in 1943, Bhattacharyya measures the similarity of two proba-bility distributions called a Bhattacharyya distance (BD). For probability distributions P and Q over the same domain X, BD is defined as:

$$BD = -log_2\Big[\sum_{i=1}^{n}\sqrt{P(x)Q(x)}\Big] \qquad (9)$$

It is used to determine the relative closeness of the two probability density functions being considered. In 1946, H. Jeffreys gave the concept of divergence as a measure of the divergence between two probability distributions. Jeffreys distance (JD) is defined as:

$$JD = \frac{1}{2}\Big[D(P,Q) + D(Q,P)\Big] \qquad (10)$$

Further, Solomon Kullback and Richard Leibler in 1951 extended the concept of Bhattacharyya distance as the directed divergence between two probability distribution, known as Kullback Leibler (KL) divergence. Though it is not a true detection metric since it is not symmetric, nor does it obey the triangle inequality. For computing KL divergence both the probability distributions must have the same sample space. For any two discrete probability distributions $P = (p_1, p_2, ..., p_n)$ and $Q = (q_1, q_2, ..., q_n)$ with $\sum_{i=1}^{n} p_i = \sum_{i=1}^{n} q_i = 1$, i = 1,2,....,n, KL divergence is defined as:

$$D(P||Q) = \sum_{i=1}^{n} p_i log_2\Big(\frac{p_i}{q_i}\Big) \qquad (11)$$

KL metric is always non-negative. i.e. $D(P||Q) \geq 0$ and is also known as Relative entropy.

Further, Alford Renyi in 1961 gave more general definition of information divergence, know as generalized information divergence (GID) defined as:

$$D_{\alpha}(P \parallel Q) = \frac{1}{1-\alpha} log_2\left(\sum_{i=1}^{n} p_i^{\alpha} q_i^{1-\alpha}\right), \alpha \geq 0. \quad (12)$$

Here, the range of α values can be used to get different useful formulas. When $\alpha \to 1$ then, GID is reduced to KL divergence. Total Variational Distance is also a kind of divergence measure between two probability distributions defined in 1966 and is given as:

$$V(P, Q) = \frac{1}{2}\sum_{i=1}^{n} |p_i - q_i| \tag{13}$$

where $P = (p_1, p_2, p_3 ... p_n)$ and $Q = (q_1, q_2, q_3 ... q_n)$ and $1 \geq p_i \geq 0$, $1 \geq q_i \geq 0$, $i = (1, 2, 3, ..., n)$. It is a symmetric measure because $V(P, Q) = V(Q, P)$. It tends to give the maximum distance between two distributions as claimed by the authors. In 1991, Lin *et al.* gave a new measure derived from Jensen's inequality and the Shannon entropy called Jensen-Shannon divergence (JSD) Lin (1991). Its importance is that different weights can be assigned to each probability distribution. With this property, it became best suited for the study of decision problems. JSD is defined as:

$$JSD(P||Q) = \frac{1}{2}\left[\sum_{i=1}^{n} p_i log\left(\frac{p_i}{m_i}\right) + \sum_{i=1}^{n} q_i log\left(\frac{q_i}{m_i}\right)\right] \tag{14}$$

where $m_i = \frac{p_i + q_i}{2}$. Sometimes, the probability distribution is treated as a vector in Euclidean vector space and distance Crooks (2017) between these distributions, known as Euclidean Distance (ED), can be defined as

$$E(P||Q) = \sqrt{\sum_{i=1}^{n} |(q_i) - (p_i)|^2} \tag{15}$$

.

In 2013, *Bhatia et. al.* Bhatia & Singh (2013) proposed a new divergence metric based on Csiszar's f-divergence measure called a ϕ-divergence, defined as

$$\phi - div(P||Q) = \sum_{i=1}^{n} \frac{p_i sinh(\alpha log_2(\frac{p_i}{q_i}))}{sinh(\alpha)}, \alpha \to 1 \tag{16}$$

2.2 Related work

In recent times, the information theory-based metrics have been used increasingly for detection of network anomalies. The idea behind information theory is to identify the deviations in the probability distribution of packet header fields of network flows. This section briefly explains the existing work done for detection of DDoS attacks using information theory-based entropy and divergence metrics.

2.2.1 Entropy based methods

In 2003, Feinstein *et al.* (2003) proposed a method to detect DDoS attacks using Shannon entropy and Chi-square test. They calculated the randomness of source IP address using these statistical metrics and compared them with the baseline val-ues to determine the abnormal behavior. They used NZIX, Bell Labs, University, and Small Company, four real traffic traces to validate their approach. Further, Chen & Yonezawa (2005) improved the idea of Feinstein and improved the accuracy with the use of destination IP and source IP as a key to filter out the legitimate traffic. They used the protocol packet header feature to devise a better strategy to create better filter rule for attack traffic. Kumar *et al.* (2007) proposed a distributed approach using Shannon entropy to detect HR-DDoS attacks in the ISP domain. Li *et al.* (2007) made improvements in the entropy-based detection systems by taking the cumulative sum of entropy to enhance accuracy. They used the source IP feature for entropy calculation and the Winpcap system for capturing and analyzing packets.

Further, Lee *et al.* (2008) used Shannon entropy at the first phase to observe the distributions and then, Euclidean distance is used to analyze the clusters to detect DDoS attacks. Tritilanunt *et al.* (2010) used Shannon entropy of input traffic as well as output traffic for the same. Wen *et al.*(2010) detect various application-layer DDoS attacks using entropy value of source IP and page access order. Xiang *et al.* (2011) used a novel generalized entropy metric called Renyi entropy to detect low rate DDoS attacks. They calculated entropy value at a different order of α to get optimal values of information distance for detection of DDoS attacks which look similar to legitimate traffic flows. Further, Tao & Yu (2013) used Shannon entropy of flow similarity to detect DDoS attacks in the local area network. Moreover, Ma & Chen (2013) used Tsallis entropy of the volume of network traffic to determine anomaly in the network. Basicevic *et al.* (2015) also used Tsallis entropy to detect SYN flood DDoS attacks. In another work Basicevic & Ocovaj (2019), authors also applied Shannon, Renyi, Tsallis, Ubriaco, and ϕ-entropy formulas to detect DDoS attacks. Saleh & Abdul Manaf (2015) detect HTTP based DDoS attacks using Shannon entropy formula. They calculate the entropy of incoming requests and clicks' average of the web page. Patil *et al.* (2019) used Shannon entropy to detect anomalies in the Hadoop-based clusters. Wang & Liu (2020) used Shannon entropy and deep learning for the detection of DDoS attacks in SDN-based networks. They use information theory for inspection of malicious traffic and then used the convolutional neural network (CNN) model to differentiate attack traffic from the legitimate one. Bhuyan Bhuyan *et al.* (2016) developed a lightweight detection system for DDoS flood at-tacks using extended Shannon entropy. They calculate different entropies (Hartley, Shannon, Renyi) within a time interval and add all to develop the extended entropy. Then this entropy value is compared with the threshold value to find the anomaly in the network. They simulate the MIT Lincoln and CAIDA data sets for validation of their approach.

2.2.2 Divergence based methods

Various diversified divergence and distance metrics are used by fellow researchers to detect DDoS attacks. In 2008, Yu *et al.* (2008) developed a dis-tance algorithm to detect suspicious flow. They used KL-distance (Relative entropy) to calculate the distance of probability distributions at the ag-gregate router and then compared this distance with a small number called threshold to determine anomalous behavior. Further, Hellinger distance is used by Sengar *et al.* (2009) to detect anomaly in the network. They used source IP, source port, destination IP, destination port as parameters and validate their approach using traffic traces of Abilene OC48c backbone link. MANETs are also vul-nerable to many cyber-attacks because of dynamic topology and lack of centralized control Nadeem & Howarth (2013). Nadeem & Howarth (2009) use chi-square test and control chart to detect the intrusion in MANETs. Authors use GloMoSim to validate their approach. Li *et al.* (2009) calcu-late information distance using KL-divergence and GID (Generalized Information Divergence) to de-tect DDoS attacks. They used MIT Lincoln DDoS 2.0.2 dataset for validation of their approach.

Xiang *et al.* (2011) used GID to detect low rate DDoS attacks and observed GID outperforms KL divergence. They take MIT Lincoln data set for normal traffic and CAIDA for low rate attack traffic traces for validation of their approach. Salem *et al.* (2012) used dynamic threshold with chi-square to increase the accuracy and compare the results with Hellinger distance and JSD (Jensen-Shannon Divergence). They used traffic trace from MAWI repository to validate their claims. Rahmani *et al.* (2012*a*) used the Total Variation Distance (TVD) metric to calculate the distance of the number of packets per connection to find the abnormal behavior. They measured the degree of similarity of traffic traces from the CAIDA dataset and MAWI repository to validate their approach. Bhuyan & Elmroth (2018) also used generalized Total Varia-tion Distance to detect multi-scale low rate DDoS attacks.

2.2.3 Discrimination of DDoS attacks from FE traffic

Researchers have made many efforts to detect low rate and high rate DDoS attacks. Still, similar efforts are needed to characterize flash events from similar-looking DDoS attacks to enhance the reliability of the system. Information theory metrics are also used by many researchers to distinguish flash events from DDoS attacks. Yu *et al.*(2009) used JSD (Jensen-Shannon Distance) to dif-ferentiate the flash crowd from a DDoS attack. In their scheme, two routers calculate probability distribution and then applied JSD to these distributions. Further, it was compared with the thresh-old to classify network flow as FE or DDoS at-tack. They used NLANR PMA Aukland VIII data set for the flash crowd and MIT Lincoln DDoS 1.0 data set for DDoS attacks to validate their approach. Li *et al.* (2009) take two probability metrics such as TVD and Bhattacharyya distance to differentiate flash event and DDoS attacks. Jeyanthi & Iyengar (2012) use Hellinger distance to dis-tinguish the DDoS attack and FE in the VoIP network. Prasad *et al.* (2013) used variation in entropy value for the discrimination of DDoS attack and flash events. Sachdeva *et al.* (2016) calculate the entropy of small clusters to discriminate both. They take FIFA98 data set for the flash crowd and CAIDA dataset for attack traffic to validate their approach. Saravanan *et al.* (2016) used Hellinger distance to capture the flow similarity, client legitimacy, and page reference for discrimination of two. Behal & Kumar (2017*b*) proposed novel information theory metrics called ϕ-entropy and ϕ-divergence to detect and discriminate DDoS attacks and flash events. They compared their results with GE and GID and claimed better accuracy. They validate their claims using FIFACup98 data set for the flash crowd, MIT Lincoln for normal traffic, and CAIDA data set for DDoS attacks.

2.3 Discussion

DDoS attacks are an austere menace to the network security that can cripple down a business in no time. Though enormous DDoS defense solutions have been proposed as mentioned above but still the problem of combatting against DDoS attacks is obstinate. After the extensive review of existing prominent research in the previous section, the following observations are made:

- The majority of the recent work has used a distinct set of entropy such as Shannon entropy, Renyi entropy, ϕ-entropy, Tsallis entropy; and divergence based measures such as Bhattacharyya distance, Sibson distance, Hellinger distance, Total variation distance, Renyi divergence, ϕ-divergence, etc. to detect various DDoS attacks.
- Both FEs and DDoS attacks share many behavioral characteristics which make the distinction very difficult between the two traffic types. During the severe DDoS attacks, the legitimate traffic has to be dropped by the defense solutions. So, it is very crucial to discriminate such traffic from attack traffic so that only attack traffic could be dropped out.
- The majority of the proposed methods have used the real dataset of MIT Lincoln to represent normal traffic, CAIDA dataset to represent HR-DDoS and LR-DDoS traffic and FIFA dataset to represent FE traffic.
- Most of the existing research has proposed isolated methods for detection of low rate DDoS attacks, high rate DDoS attacks and FE traffic. There is a need to introduce a collective generalized approach that is capable of detecting these types of traffic.
- Although information theory-based metrics are used mostly in the detection of DDoS, however, the choice of the appropriate value of the generalized parameter α for a particular network is still very challenging.

Table 1. Notations and Symbols used

Notation	Definition
T	Sampling period
T_w	Time Window
Δ	size of T_w
j	initialized to 1 and increment after each T_w
$flowid_i$	Unique flow id of i^{th} network flow
n_C	number of packets per T_w in current traffic
n_N	number of packets per T_w in baseline traffic
σ_1	threshold based on n_N computed from baseline traffic
σ_2	threshold based on ID_N computed from baseline traffic
ID_C	Information distance between current and normal traffic flows
ID_N	Information distance between normal traffic flows
$P(x), Q(x)$	Probability distributions of network flows in different T_w base on x packet header feature
E'	Information Theory Metric such as Shannon entropy, Ubriaco entropy, Renyi entropy, Tsallis entropy, ϕ-entropy, Divergence Metric such as KL divergence, Jeffrey Distance, Bhattacharyya Distance, JSD, Hellinger Distance, Pearson Distance, Total Variation Distance, Euclidean Distance, GID, ϕ-divergence
E'_C	Metric value of current netflow
E'_N	Metric value of baseline netflow

3. Proposed methodology

The proposed detection approach works on the knowledge of flow similarity. Attack traffic flows are generated through the common shared logic. So, there are more chances that there is a similarity in traffic flows when there will be attack traffic. On the other hand, normal traffic is highly dynamic, which causes a significant deviation in the packet header features of normal traffic and attack traffic. The various notations used in the subsequent sections are explained in Table 1.

We use the generalized detection methodology as given in Algorithm 1. We use this common methodology for all information theory-based metrics and different datasets for evaluation of the results based on standard criteria for better comparisons, so called generalized detection methodology. In this, the detection process starts by sampling the network traffic in each time window T_w = 1 seconds, sampling period = 120 seconds. We extract the appropriate packet header features and classify the traffic into particular traffic type. A Netflow is defined as a 5-tuple destination IP, source IP, source port, protocol, and destination port. We use the packet header features of the protocol, destination IP, source IP, and incoming packet rate to represent a Netflow. The destination IP field of the packet

Table 2. Effect of network anomalies on traffic header features

S.No.	Netflow type	Nature of Netflow	Traffic Feature	Distribution
1.	DoS/DDoS	Anomaly	Source IP	Dispersed
			Destination IP	Skewed
2.	Port Scan	Anomaly	Destination IP	Skewed
			Destination Ports	Dispersed
3.	Network Scan	Anomaly	Destination IP	Dispersed
			Destination Ports	Skewed
4.	FE	Legitimate	Source IP	Dispersed
			Destination IP, Ports	Skewed
5.	Normal	Legitimate	Source IP	Dispersed
			Destination IP, Ports	Skewed

Algorithm 1 A Proposed Generalized Detection Algorithm

```
Set T=120 seconds, $T_w$ =1 second $\sigma_1, \sigma_2 \leftarrow$ standard thresholds
while $T_w <=$ T do
    Extract packet header features $\{srcIP, dstIP, proto, n_c\} \in T_w$.
    Calculate $P(srcIP) \in T_w$.
    Compute $E'$ using $P(srcIP)$
    Calculate the average information distance $ID = |E'_C - E'_N|$.
    if $n_c$ >$\sigma_1$ then
        $ID$>$\sigma_2$ ? HR-DDoS Attack : Flash Event
    else
        $ID$>$\sigma_2$ ? LR-DDoS Attack : Normal traffic.
    end if
    $T_w + +$
end while
```

header classifies the network traffic flow, which is destining towards a specific target IP and protocol field is used to find the type of network traffic. There are other packet header features which can be used to detect the anomaly of the network. How-ever, we used only those features which are enough to detect flash events and DDoS attacks in current experiments. Even, the pattern of attack traffic remains hidden in the collected data, and it can be anticipated using packet header fields Bhandari *et al.*(2016)-Xiang *et al.* (2011), Bhuyan *et al.* (2016), Yu *et al.* (2009), Wang *et al.* (2012).

The prominent existing research has extensively used flow similarity of Netflow in the network anomaly detection domain. There are wide varieties of network anomalies present in a network, as shown in Table 2. Despite the type, every traffic anomaly exhibit a common behavioral characteristic, i.e. they all cause significant deviations in the packet header feature distributions such as changes in source/destination IP addresses, source/destination port numbers, protocol, etc. For example, when there is DDoS attack, the distribution of destination IPs will be concentrated towards a specific victim address. In contrast, the distribution of source IPs will be more dispersed. Similarly, during a network scan for a particularly vulnerable port, there would be a skewed distribution. So, by examining the deviations in the probability distribution of packet header features, we can detect and classify a broad set of network anomalies. Both types of traffic, such as DDoS attack traffic and flash event, cause noticeable differences in packet header features of these traffic types. How-ever, high-rate DDoS attacks are easy to detect as their traffic profiles significantly deviate from nor-mal traffic profiles.

We use the concept of information distance (ID) to find the quantitative difference between attack traffic and normal traffic. ID is defined as the diference between metric values of attack and normal traffic. For the set E' i.e. set of different entropy metrics, it is computed as ID= $|E'_C - E'_N|$. Here, E'_C, E'_N represent entropy values of current Netflows and normal Netflows, respectively. How-ever, there is no need to compute the ID, i.e. difference of divergence metrics, because divergence is already computed concerning two probability distributions of Netflows. More is the value of ID between two Netflows; higher is the detection efficiency. As we defined the ID, the mathematical models of different types of Netflow are defined as follows.

- **Definition 1:** Normal Netflow: For a given sampled Netflow (s_i) if the incoming rate of packets in a time window is less than the incoming rate of packets in a baseline behavior of the network, and the value of the ID, i.e. information distance is also less than equal to ID values between baseline behavior Net-flows, then, it is termed as Normal Netflow. Mathematically, it can be represented as:

$$n_C <= n_N \pm a * sd_n \wedge ID_C <= ID_N \pm b * sd_{ID_N} \quad (17)$$

- **Definition 2:** LR-DDoS Attack: For a given sampled Netflow (s_i) if the incoming rate of packets in a time window is less than the incoming rate of packets in a baseline behavior of the network, but the value of ID, i.e. information distance is more than the ID values between baseline behavior Netflows, then, it is termed as LR-DDoS attack Netflow. Math-ematically, it can be represented as:

$$n_C < n_N \pm a * sd_n \wedge ID_C > ID_N \pm b * sd_{ID_N} \quad (18)$$

- **Definition 3:** HR-DDoS Attack: For a given sampled Netflow (s_i) if the incoming rate of packets in a time window is more than the incoming rate of packets in a baseline behavior of the network, and the value of ID, i.e. information distance is more than the ID values between baseline behavior Netflows, then, it is termed as HR-DDoS attack Netflow. Math-ematically, it can be represented as:

$$n_C > n_N \pm a * sd_n \wedge ID_C > ID_N \pm b * sd_{ID_N} \quad (19)$$

- **Definition 4:** FE Traffic: For a given sampled Netflow (s_i) if the incoming rate of packets in a time window is more than the incoming rate of packets in a baseline behavior of the network, but the value of ID, i.e. information distance is less than the ID values between base-line behavior Netflows, then, it is termed as FE Netflow. Mathematically, it can be repre-sented as:

$$nC > nN \pm a * sdn \wedge IDC <= IDN \pm b * sdIDN \quad (20)$$

Here a, b are known as Tolerance factors. sd_n is the standard deviation of the incoming number of packets. sd_{ID_N} is the standard deviation in ID values between normal Netflow. Accordingly,

$$\sigma_1 = n_N \pm a * sd_n \quad (21)$$

$$\sigma_2 = ID_N \pm b * sd_{ID_N} \quad (22)$$

These two threshold values are calculated by analyzing the Netflows during the baseline behavior of the network. We compute FPR and FNR values from the normal Netflows to compute the threshold values (details in section 5.3.1). We choose the values of tolerance factors a and b where FNR and FPR curves intersect each other. The detection process analyzes incoming Netflows, and it separates low-rate, and high-rate Netflows by comparing the current number of incoming packets in each T_W, i.e. n_C is compared with threshold σ_1. Af-ter that, ID values from the entropy metrics mentioned in the set E' are calculated. Whenever there is a significant deviation of ID_C from σ_2, the LR-DDoS or HR-DDoS attack is said to be detected. The high-rate Netflows may also be declared as legitimate traffic (FE). Flash events cause a gradual increase and decrease in traffic rate over a period. This change in traffic happens when thousands of legitimate users start accessing the same web re-sources at the same time. Since high rate DDoS attacks and flash events, in a short time, lead to a sudden change in network traffic volume, so both impact almost in the same way on their entropy values which results in minimal information distance between these two types of traffic, it makes the char-acterization of these traffics types very challenging. In the case of generalized entropy, the larger value of the entropic index parameter magnifies the information distance between these two types of traffics, resultantly there are better chances of their characterizing.

4. Experimental setup and datasets used

Real network testing, simulation, and emulation are three experimental schemes to validate new and existing ideas. Every technique has its benefits over the other. Testing in the real network is challenging in the case of DDoS attacks Sachdeva & Kumar (2014). On the other hand, simulation gives the controlled environment and opportunities to perform repeated experiments. However, all the devices, network operating systems, links between devices are virtual in simulation, which affect the results drastically as compared to the real environment. An emulation is a hybrid approach that combines the real elements with synthetic or abstracted elements to design the test environment White *et al.*(2002), Fall (1999).

We used an emulation-based DDoSTB testbed developed by Behal & Kumar (2017*a*) to validate our approach. The scaled architecture of DDoSTB is shown in Figure 1. DDoSTB testbed is composed of 75 physical nodes organized into three different clusters of 25 nodes each, 3 D-Link

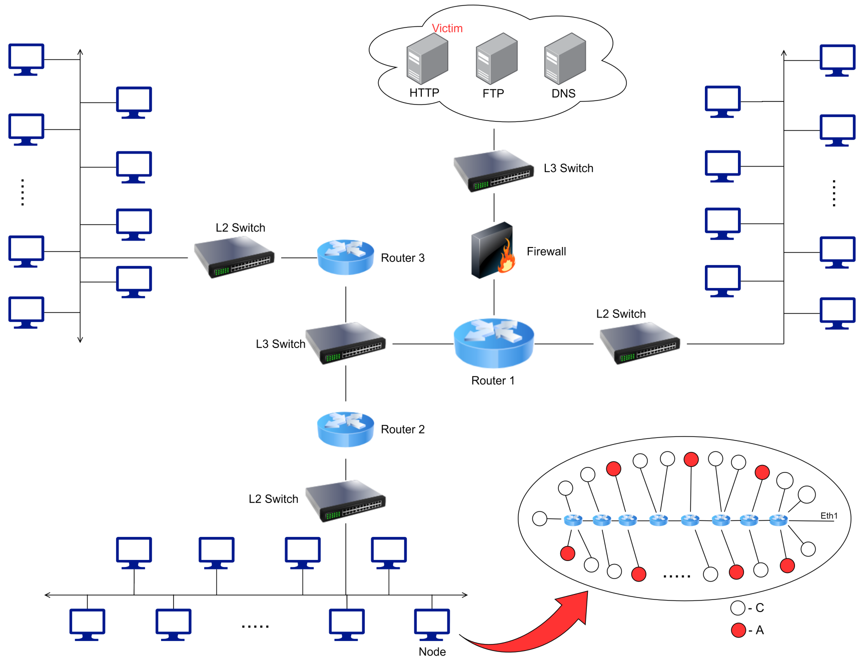


Fig. 1. Distributed denial of service testbed (DDoSTB)

physical routers, 3 Layer2 switches (1Gbps bandwidth), 3 Layer3 switches (1 Gbps bandwidth), two dual-processor 8-core Linux servers that act as the victims.

The nodes run ubuntu and windows OS instances. In this testbed, CORE emulator *The vCORE Emulator, http://www.nrl.navy.mil/itd/ncs/products/core*(2016) is used to escalation the number of virtual nodes. One CORE node was made up of sixteen virtual clients and four soft routers to generate real network traffic. We extended the configuration of CORE node from 20 nodes to 48 nodes. It increases the capacity of DDoSTB testbed from 2250 nodes to around 5000 nodes with randomly distributed normal clients (C) and attackers (A). The BONESi Alcorn & Chow (2014) botnet simulator is used to generate high-rate DDoS Netflow.

We used MIT Lincoln Laboratory lincoln laboratory LLSDDos0.2.2 dataset (n.d.) dataset to represent normal traffic, CAIDA *The CAIDA DDoS Attack Dataset, "Coop-erative Analysis for Internet Data Analy-sis", https://www.caida.org/data/passive/ddos-20070804-dataset.xml* (2010) dataset to represent HR-DDoS and LR-DDoS traffic, and the FIFA World Cup ITA (1998) dataset represents the flash event. It is important to note that the FE scenario is taken from the 66^{th} day of the FIFA dataset because it contains the maximum number of GET-requests to the webserver. Further, the novel DDoSTB is used to represent HR-DDoS and FE traffic scenarios.

5. Results and discussion

We compute different entropy and divergence metrics on entropic index parameter α from 0.1 to 15. The different ID (Information Distance) values computed using these metrics are shown in Table 3 and Table 4. In these Tables, ID_1 represents the ID values between normal traffic profile taken from MIT Lincoln dataset and HR-DDoS attack traffic profile taken from CAIDA dataset. ID_2 represents the ID values between normal traffic profile which is taken from MIT Lincoln dataset and LR-DDoS attack traffic profile taken from CAIDA dataset. ID_3 represents the ID values between normal traffic profile taken from MIT Lincoln dataset and flash event traffic profile taken from FIFA dataset. ID_4 represents the ID values between normal traffic profile taken from DDoSTB dataset and HR-DDoS attack traffic profile taken from DDoSTB dataset. ID_5 represents the ID values between normal traffic profile taken from DDoSTB dataset and FE traffic profile taken from DDoSTB dataset. ID_6 represents the ID values between HR-DDoS attack traffic profile taken from CAIDA dataset and

Table 3. Temporal variation of entropy metrics and ID values

Entropy Type	α-order	MIT Lincoln Normal	CAIDA HR-DDoS	CAIDA LR-DDoS	FIFA FE	DDoSTB FE	DDoSTB HR-DDoS	ID_1	ID_2	ID_3	ID_4	ID_5	ID_6
Shannon	-	1.791	0.369	1.795	2.590	0.532	2.490	1.422	0.550	0.825	1.259	0.784	2.220
Ubriaco	-	0.384	0.188	0.387	0.758	0.286	0.712	0.218	0.126	0.377	0.1593	0.335	0.570
Renyi	2	1.599	11.138	3.226	3.615	10.572	3.419	9.539	1.771	2.015	8.973	1.820	7.524
	4	1.435	8.107	2.695	3.092	7.914	2.657	6.671	1.405	1.662	6.479	1.230	5.014
	6	1.362	7.445	2.560	2.916	7.310	2.438	6.083	1.346	1.561	5.948	1.088	4.529
	8	1.321	7.150	2.496	2.824	7.031	2.340	5.829	1.322	1.510	5.710	1.034	4.326
	10	1.295	6.982	2.459	2.767	6.869	2.284	5.687	1.310	1.479	5.573	1.005	4.215
	12	1.278	6.874	2.435	2.728	6.763	2.248	5.596	1.302	1.458	5.485	0.988	4.146
	14	1.265	6.799	2.418	2.700	6.689	2.223	5.534	1.297	1.442	5.423	0.976	4.099
ϕ-entropy	1	6.523	26.451	6.139	15.648	38.938	25.908	19.928	0.384	9.125	32.415	18.779	10.820
	1.1	9.022	48.139	7.914	21.601	71.040	37.579	39.117	1.108	12.579	62.017	27.513	26.538
	1.2	13.033	88.383	10.524	30.371	130.830	55.311	75.350	2.509	17.339	117.797	40.706	58.012
	1.3	19.630	163.467	14.502	43.447	242.918	82.498	143.838	5.128	23.818	223.289	61.369	120.020
	1.4	30.741	304.226	20.801	63.180	454.322	124.556	273.485	9.940	32.439	423.581	93.898	241.046
	1.5	49.872	569.219	31.162	93.312	855.319	190.192	519.347	18.710	43.440	805.448	146.907	475.907
	1.6	83.471	1069.953	48.825	139.839	1620.098	293.506	986.482	34.646	56.367	1536.626	233.782	930.114
T-Sallis	0.1	3.439	3.523	3.434	7.293	5.662	10.271	1.520	0.946	3.881	2.481	6.832	3.770
	0.5	2.014	0.817	1.643	3.208	0.299	4.214	2.831	0.989	1.244	1.714	2.210	4.025
	0.6	1.804	1.587	1.334	2.546	1.179	3.358	3.391	1.063	0.834	0.730	1.569	4.133
	0.7	1.628	2.580	1.157	1.903	2.246	2.598	4.208	1.126	0.572	0.689	1.043	4.483
	0.8	1.479	4.299	1.429	1.140	3.991	1.807	5.778	1.053	0.640	2.513	0.685	5.439
	2	0.654	0.999	0.857	0.915	0.999	0.899	0.346	0.244	0.261	0.346	0.245	0.085
	3	0.425	0.500	0.476	0.493	0.500	0.489	0.075	0.069	0.069	0.075	0.064	0.007

Table 4. Temporal variation of ID values of Divergence metrics

Divergence type	α-order	ID_1	ID_2	ID_3	ID_4	ID_5	ID_6
KL Divergence	-	4.915	1.019	1.166	4.390	0.423	2.426
Jeffrey Distance	-	2.372	0.490	0.437	2.106	0.176	1.121
Bhattacharyya Distance	-	2.446	0.490	0.540	2.354	3.623	2.504
JSD	-	0.411	0.153	0.214	0.383	0.037	0.274
Hellinger Distance	-	0.822	0.464	0.589	0.819	0.437	0.668
Pearson Divergence	-	0.460	0.463	0.266	0.071	0.139	0.328
T. Variational Distance	-	0.486	0.339	0.402	0.491	0.153	0.353
Euclidean Distance	-	0.563	0.369	0.370	0.561	0.102	0.268
GID	2	5.063	1.224	1.279	4.900	0.925	3.191
	4	5.197	1.443	1.397	5.016	2.436	3.747
	6	5.264	1.533	1.461	5.074	2.812	3.913
	8	5.304	1.583	1.500	5.110	2.995	4.002
	10	5.330	1.615	1.527	5.134	3.104	4.059
	12	5.349	1.636	1.546	5.151	3.176	4.099
	14	5.362	1.652	1.561	5.163	3.228	4.129
ϕ-Divergence	0.5	1.524	0.300	0.341	1.467	0.210	0.738
	0.9	1.735	0.284	0.319	1.647	0.208	0.766
	1.5	2.317	0.251	0.272	2.134	0.205	0.839
	1.8	2.755	0.233	0.246	2.494	0.204	0.888
	2	3.114	0.221	0.229	2.786	0.203	0.926
	2.5	4.313	0.196	0.188	3.742	0.202	1.040
	2.6	4.616	0.192	0.180	3.979	0.202	1.066
	4	12.787	0.169	0.098	10.077	0.211	1.600
	6	64.586	0.317	0.040	45.219	0.261	3.374
	8	374.081	1.002	0.020	238.006	0.393	8.520

FE traffic profile taken from the FIFA dataset. Further, the entropic index parameter α in the case of General-ized entropy and Generalized divergence measures can be adjusted by calculating the coefficients of correlation as done by Berezi ński *et al.* (2015). Xi-ang *et al.* (2011) also proposed a method based on reduced FPR to select appropriate entropic index parameter.

5.1 Empirical analysis of entropy metrics

This section empirically investigates the performance of each entropy metrics defined in Section 2. For the validation purpose, these entropy metrics are computed on the datasets mentioned in Section 4. The temporal variation in the entropy values and corresponding ID values are shown in Table 3.

The proposed generalized detection methodology works based on Netflow similarity. More the value of ID more will be detection accuracy. It has observed from Table 3, that ϕ-entropy gives the maximum value of ID as compared to Shannon, Ubriaco, Renyi, and Tsallis entropy metrics. So, the detection accuracy of ϕ-entropy shall be more to detect these different types of Netflow. Further, as per the concept of reduced FPR given by Xiang *et al.* (2011) to select the optimal value of α (entropic index parameter), we chose the value of α=2 for Renyi entropy, α=1.5 for ϕ-entropy and α=0.7 for Tsallis entropy. It has observed from Table 3 that ID values for Renyi generalized entropy keep on decreasing with the increase in entropic index parameter α whereas ID values of generalized Tsallis entropy keep on increasing up to entropic index parameter α =0.8. For ϕ-entropy, ID values keep on increasing with the increase in entropic index parameter α. It means ϕ-entropy comes out to be more appropriate as compared to other entropy metrics for detecting different types of attack Netflow.

Further, it has observed that the incoming rate of both HR-DDoS Netflows and FE Netflows are almost equal. So, the ID values for both types of traffic are almost the same as evident from ID_1, ID_3 and ID_6 values of Table 3, which makes the distinction between HR-DDoS and FE Netflows very difficult.

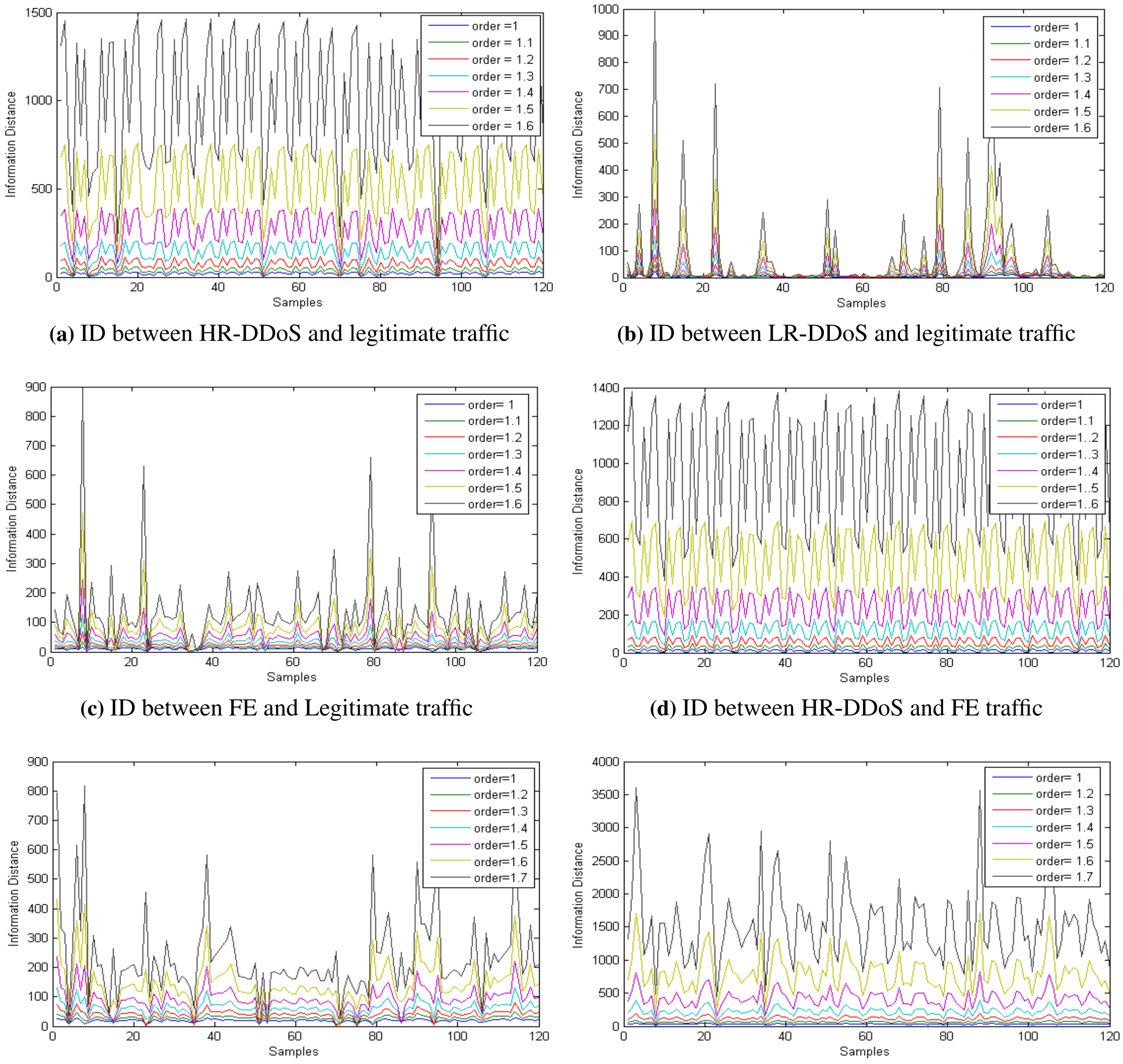
Further, we found (Figure 2 and Figure 3) that ϕ-entropy is capable of correctly predicting the type of pattern of ongoing attack traffic. Figure 2(a) to 2(f) and 3(a) to 3(f) represent the ID_1 to ID_6 values using ϕ-entropy and Renyi entropy respectively. It has been observed that as the value of the entropic index parameter increases, the value of ID also increases in the case of ϕ-entropy leading to more detection accuracy. Further, the type of attack pattern is more predictable in the case of ϕ-entropy as compared to Renyi entropy.

5.2 Empirical analysis of divergence metrics

This section empirically investigates the performance of various divergence metrics defined in Section 2. For the validation purpose, all of these divergence metrics are computed on the datasets mentioned in Section 4. The temporal variation in ID values of all these metrics are shown in Table 4.

Since all attack nodes operate in a distributed manner and in a coordinated way to direct attack traffic toward the victim using a common program logic, their probability distributions are of a similar kind. It results in ID value near to 0 between attack Netflows as compared to normal Netflow. Further, the number of packets in the probability distributions need to be normalized within a particular T_w (time window) for computing ID values using a divergence metric. As ID is the difference of divergence value of current Netflows and normal baseline Netflows, more the value of ID, more will be detection accuracy. We found (Table 4), that ϕ-divergence metric gives the maximum value of ID as compared to other metrics. We choose the value of α=14 for GID, α=8 for ϕ-divergence for all types of Netflow.

We observed that GID metric is more capable of differentiating between normal and FE traffic but this type of distinction is not much relevant to a network administrator as both represent legitimate traffic. The only difference is the frequency of Netflow. Divergence metrics are more susceptible to find such minor differences as compared to entropy metrics. It leads to more detection accuracy of divergence metrics in contrast to entropy metrics. Further, apart from the generalized divergence measure, KL metric is best suited to detect different types of Netflows in a network. Also, as the incoming rate of both HR-DDoS Netflows and FE Netflows are almost equal, the divergence based ID values for both types of traffic shows the considerable difference as clear from ID_1, ID_3, and ID_6 values of Table 4. It means divergence based metrics are more appropriate and useful for discriminating HR-DDoS and FE Netflows as compared to entropy metrics.



(a) ID between HR-DDoS and legitimate traffic

(b) ID between LR-DDoS and legitimate traffic

(c) ID between FE and Legitimate traffic

(d) ID between HR-DDoS and FE traffic

(e) ID between FE DDoSTB and legitimate traffic

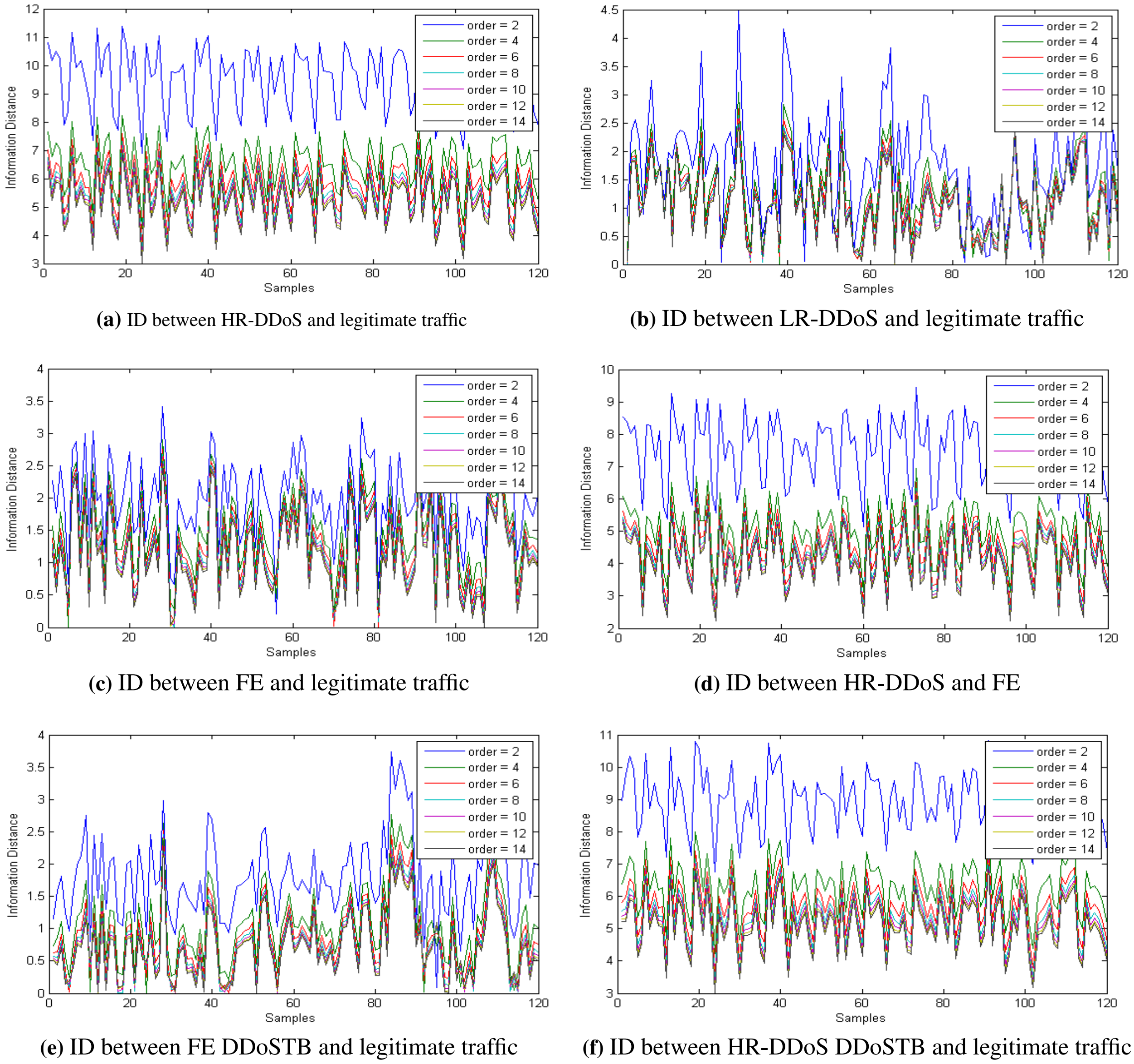
(f) ID between HR-DDoS DDoSTB and legitimate traffic

Fig. 2. Information distance (ID) between various types of traffics using ϕ-entropy

In contrast to entropy ID values which keeps on decreasing with increase in entropic index parameter α, the divergence based ID values keep on increasing with the entropic index parameter α value. It results in more differentiation power of divergence metrics, hence, lead to more detection accuracy.

It is worth mentioning that in the case generalized information theory metric, the selection of optimal value entropic index parameter α is very crucial. More the value of this parameter more will be the computational complexity of the detection metric, hence would lead to enforce more delay in the detection process. Most of the traditional divergence measures (except JSD) compares at most two probability distributions at a time. In the case of a large-scale ISP where many edge routers are working in parallel, it would be computationally expensive to use these divergence measures. In such cases, JSD divergence is best suited, which is capable of comparing N parallel probability distributions at the same time. We empirically investigate the same by mathematically and experimentally.

Suppose, in an ISP domain, there are n_1, n_2, n_3......N edge routers that forward the traffic towards the particular target IP (victim webserver). There will be N different traffic



(a) ID between HR-DDoS and legitimate traffic

(b) ID between LR-DDoS and legitimate traffic

(c) ID between FE and legitimate traffic

(d) ID between HR-DDoS and FE

(e) ID between FE DDoSTB and legitimate traffic

(f) ID between HR-DDoS DDoSTB and legitimate traffic

Fig. 3. Information distance (ID) between various network traffics using Renyi entropy

distributions that are to be monitored at the victim end. If the server uses the old-style of calculations for comparison of two probability distributions, for example, KL divergence method. Then for comparison of total N traffic distributions, there will exist ${}^{N}C_2$ combinations to compare all the traffic distributions with one another. Suppose it takes t seconds to compare two distributions then the overall time for comparison will be ${}^{N}C_2 * t$ seconds. However, JSD divergence metric compares all different probability distributions in one phase, resultantly reducing time complexity.

So, if there are N edge routers and k distributions for different edge routers, there shall be a total of ${}^{n}C_k = \frac{n!}{k!(n-k)!}$ combinations that are too complex to calculate using the old style of computation. On the other hand, JSD measure can compute the value of metric between all these possible combinations in only N x O(k), where k is the number of different probability distributions and N is the number of edge routers. We empirically investigated that for N=14, JSD took 0.003921 seconds for computation, whereas it took 0.08421 seconds to compute KL divergence.

5.3 Performance evaluation

This section focuses on the performance evaluation of various information theory-based entropy and divergence metrics. For evaluating the efficacy of any network traffic-based defense system, initially, there is a need to understand some basic terminologies.

Table 5. Confusion Matrix

		Predicted class	
		Normal	Attack
Actual Class	Normal	TN	FP
	Attack	FN	TP

A DDoS defence system classifies events into the attack and normal events. With the perspective of a defence system, a positive event is considered to be an attack event, while a negative event is considered to be a normal event. There are four combinations of these two decision variables, as shown in Table 5. The value of TP (True Positive) increases when the defence system correctly classifies it as an attack event. In contrast, FP (false positive) value increases when a legitimate event incorrectly classified as an attack event. In the same way, TN (true negative) increases when a normal event is correctly categories as a legitimate event, and FN (false negative) increases when the defence system does not detect abnormal behavior. To evaluate the performance of information theory metrics, we use the detection system evaluation parameters as defined by Ghorbani *et al.* (2010) such as detection rate (D_r) also known as True Positive Rate (TPR) or Recall, Precision, F-Measure (F_m), False Negative Rate (FNR), False Positive Rate (FPR), and Classification Rate (C_r). The mathematical formulae of these detection metrics are shown in Table 8.

A detection rate parameter is used to measure the fraction of attack events which are detected correctly, and on the other hand, the classification rate represents the ratio of genuinely classified events to the total occurred events. A FPR represents the effectiveness of the defense system, and FNR measures the reliability of the detection system.

5.3.1 Design parameters

There are several detection system design parameters that also need to tune optimally for the efficient working of a DDoS defense system. The performance of a DDoS defense system depends on how well these design parameters are selected. However, the optimal tuning of these parameters depends on the network conditions, which can be different for different networks, situations, or applications. As the network traffic behavior is highly dynamic nowadays, we need to choose the value of these parameters very carefully.

- **Analysis of Time window size**: The time series analysis of Netflows can be performed in two ways (1) time window analysis, or (2) packet window analysis. In this work, we opted for time window analysis because for detecting DDoS attacks, it is very crucial to identify the periodicity of the incoming traffic patterns, which is not possible with the packet counting approach. The size of the time window plays a vital role in the efficient working of the detection systems. If it is not adjusted appropriately, then the system may detect happening of the event but may not detect the type of the event effectively. A larger window size would lead to a high false-negative rate as well as a low false-positive rate and vice-versa.

  We monitor the baseline behavior of the network using diverse window sizes of t = 0.1, 0.3, 0.5, 1, 1.2, 1.5, and 2 seconds for computation of standard deviation of entropy (or divergence) metric values. When standard deviation having the least value, it indicates the stable behavior of the network. So the size of the time window should be chosen at that value where standard deviation has the lowest value. In our case, we choose T_w=1 second.

- **Setting up of generalized parameter α**: The value of α plays a vital role in the accuracy of the detection system based on generalized information theory metrics because the network has a very dynamic nature. According to Xiang *et al.* (2011), each α value discloses different facets of probability distributions which are used for categorization of the network. The value of α depends on the particular type of anomalies present in the network traffic Bereziński *et al.* (2015); authors used the concept of correlation to set the optimal range of generalized α parameter. However, we used the concept (reduced FPR) of Xiang *et al.* (2011) to identify the appropriate value of α parameter for both types of the metrics. We empirically investigated and analyzed the normal baseline behavior of the net-

Table 6. Comparison of entropy metrics on various detection system evaluation parameters

Metric	High-Rate Attack						Flash Event						Low-Rate Attack					
	D_r	P	F_m	C_r	FNR	FPR	D_r	P	F_m	C_r	FNR	FPR	D_r	P	F_m	C_r	FNR	FPR
Shannon Entropy	0.513	0.976	0.672	0.667	0.488	0.042	0.163	0.929	0.277	0.433	0.838	0.042	0.050	0.667	0.093	0.513	0.950	0.042
Tsallis Entropy	0.100	0.615	0.172	0.358	0.900	0.125	0.100	0.615	0.172	0.358	0.900	0.125	0.125	0.500	0.200	0.500	0.875	0.125
Ubriaco Entropy	0.067	0.762	0.123	0.364	0.933	0.025	0.225	0.915	0.361	0.469	0.775	0.025	0.033	0.444	0.062	0.496	0.967	0.025
Renyi Entropy	1.000	0.984	0.992	0.989	0.000	0.033	0.483	0.967	0.644	0.644	0.517	0.033	0.558	0.944	0.702	0.763	0.442	0.033
ϕ-Entropy	1.000	0.885	0.937	0.911	0.004	0.258	0.863	0.870	0.866	0.822	0.138	0.258	0.158	0.380	0.224	0.450	0.842	0.258

Table 7. Comparison of divergence metrics on various detection system evaluation parameters

Metric	High-Rate Attack						Flash Event						Low-Rate Attack					
	D_r	P	F_m	C_r	FNR	FPR	D_r	P	F_m	C_r	FNR	FPR	D_r	P	F_m	C_r	FNR	FPR
KL Divergence	1.000	0.984	0.992	0.989	0.000	0.033	0.379	0.958	0.543	0.575	0.621	0.033	0.592	0.947	0.728	0.779	0.408	0.033
Jeffrey Distance	1.000	0.960	0.980	0.972	0.000	0.083	0.097	0.697	0.170	0.373	0.903	0.083	0.258	0.756	0.385	0.588	0.742	0.083
Bhattacharayya	1.000	0.980	0.990	0.986	0.000	0.042	0.892	0.977	0.932	0.914	0.108	0.042	0.750	0.947	0.837	0.854	0.250	0.042
Hellinger	0.975	1.000	0.987	0.983	0.025	0.000	0.096	1.000	0.175	0.397	0.904	0.000	0.075	1.000	0.140	0.538	0.925	0.000
JSD	0.996	0.984	0.990	0.986	0.004	0.033	0.181	0.915	0.303	0.445	0.819	0.033	0.150	0.818	0.254	0.558	0.850	0.033
Pearson Divergence	0.625	0.955	0.756	0.731	0.375	0.058	0.200	0.873	0.325	0.447	0.800	0.058	0.167	0.741	0.272	0.554	0.833	0.058
Total Variation	1.000	0.992	0.996	0.994	0.000	0.017	0.346	0.976	0.511	0.558	0.654	0.017	0.475	0.966	0.637	0.729	0.525	0.017
Euclidian Distance	0.875	0.972	0.921	0.900	0.125	0.050	0.142	0.850	0.243	0.411	0.858	0.050	0.250	0.833	0.385	0.600	0.750	0.050
ϕ-Divergence	1.000	0.984	0.992	0.989	0.000	0.033	0.871	0.981	0.923	0.903	0.129	0.033	0.775	0.959	0.857	0.871	0.225	0.033
GID	1.000	0.976	0.988	0.983	0.000	0.050	0.563	0.957	0.709	0.692	0.438	0.050	0.142	0.739	0.238	0.546	0.858	0.050

Table 8. Detection Metrics

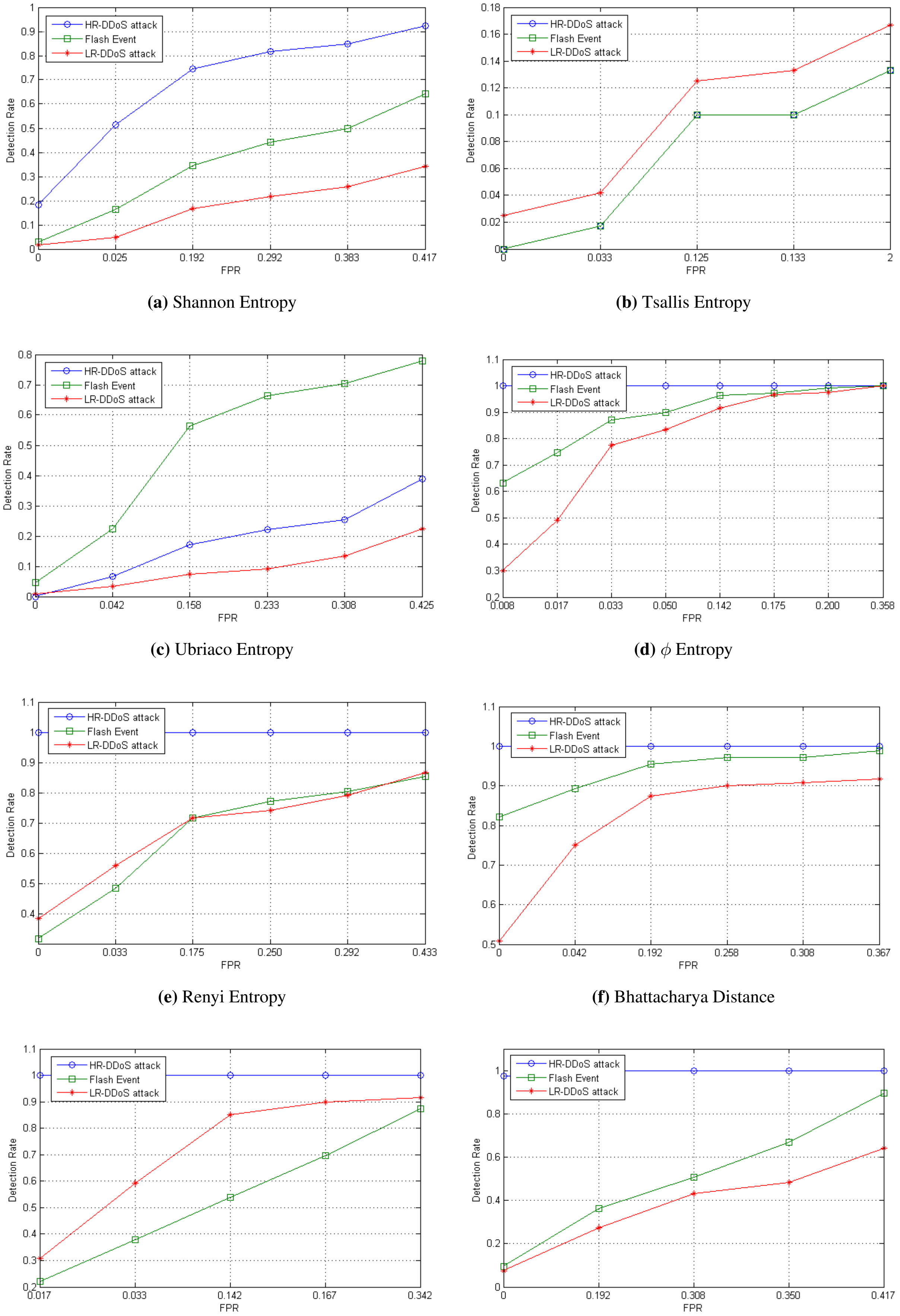
Sr. No.	Detection Metric	Formulae
1	Precision	$\frac{TP}{TP+FP}$
2	Detection Rate	$\frac{TP}{TP+FN}$
3	False Positive Rate	$\frac{FP}{TN+FP}$
4	F-Measure	$\frac{2*P*R}{P+R}$
5	Classification Rate	$\frac{TP+TN}{TP+TN+FP+FN}$
6	False Negative Rate	$\frac{FN}{TP+FN}$

work (without attack) as published in Behal & Kumar (2017*b*) and choose the value of α=0.5 for the current network under analysis.

- **Threshold calibration**: The dynamics of the network under consideration should be considered whenever we are choosing the range of threshold limits. A complete analysis of the existing network has been done done to select the appropriate limits of thresholds. A low tolerance factor value always results in a high detection rate and FPR, and minimum FNR. So, if our goal is to detect all types of attacks, then the detection system may signal some of the normal states as attack states, which leads to an increase in FPR. On the contrary, if the threshold value is set to high, then it leads to low FPR as well as low detection rate. Consequently, the DDoS defence system may miss some attack events to detect. So, to get the system with balanced FPR and FNR, the value of the tolerance factor is must be in between low and high bounds. FPR signifies the effi-cacy of a detection system, whereas FNR (1-Recall) signifies the reliability of a detection system. We have used the temporal variation of tolerance factor (k) to select a threshold value. The point where both FNR and FPR curves intersect can be selected as the optimal threshold value. Alternatively, where the point where Precision-Recall (PR) curves intersect can also be used to select the optimal thresh-old value.

We compute the values of various detection system evaluation parameters, as shown in Table 8. For generalized Renyi and Tsallis entropy metrics, we compute the results on the entropic index parameter α=15, and for generalized ϕ-entropy, we use α=1.2 as per the reduced FPR value computed using the method adopted by Xiang *et al.* (2011). For generalized information divergence metric, we use α=15 and for generalized ϕ-divergence, we use α=0.5. The results of various divergence and entropy metrics are shown in Table 6 and Table 7 respectively. We compute separate detection system evaluation parameters for each type of Netflow, i.e. HR-DDoS attack, FE traffic, and LR-DDoS attack.

As stated in the previous sections, more is the information distance between different types of Netflow, and more will be the detection accuracy. It has observed that for HR-DDoS attack, Tsallis entropy, Renyi entropy and ϕ-entropy produce the best de-



(a) Shannon Entropy

(b) Tsallis Entropy

(c) Ubriaco Entropy

(d) ϕ Entropy

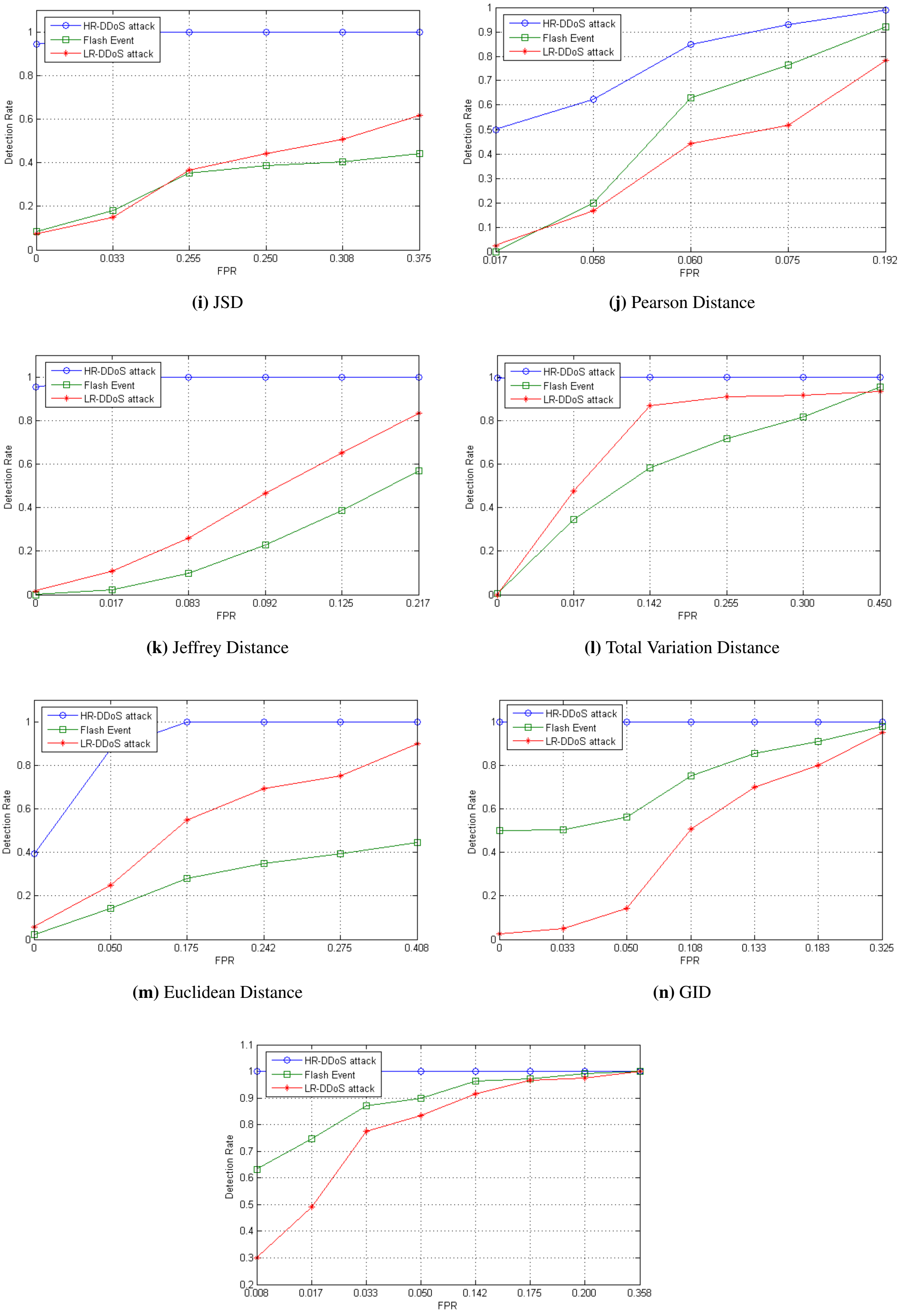
(e) Renyi Entropy

(f) Bhattacharya Distance

(g) KL Divergence

(h) Hellinger Distance

Fig. 4. ROC curves of various Entropy and Divergence metrics (conti.)



(i) JSD

(j) Pearson Distance

(k) Jeffrey Distance

(l) Total Variation Distance

(m) Euclidean Distance

(n) GID

(o) ϕ-Divergence

Fig. 4. ROC curves of various Entropy and Divergence metrics

Table 9. Comparison of Proposed work with other similar works

Authors	Experimental	Detection	Datasets	Type of Netflows Detected		
Year	Technique	Metrics	Used	LR-DDoS	HR-DDoS	FE
Yu *et al.* (2009)	Real Datasets	Sibson Distance Hellinger Distance Jeffrey Distance	MIT LLSDDOS NLANR Auckland VIII	-	√	√
Xiang *et al.* (2011)	Real Datasets	Shannon Entropy Renyi Generalized Entropy Renyi Generalized Information Divergence KL Divergence	MIT Lincoln CAIDA	√	-	-
Bhuyan *et al.* (2015)	Real Experiment	Shannon Entropy Renyi Generalized Entropy Renyi Generalized Information Divergence KL Divergence	MIT Lincoln CAIDA TUIDS	-	√	-
Berezinski *et al.* (2015)	Simulation	Shannon Entropy, Tsallis Entropy Renyi Generalized Entropy	Synthetic	√	√	-
Behal *et al.* (2017)	Emulation Testbed	Shannon Entropy,Renyi Generalized Entropy Variation Distance,Sibson Distance KL Divergence,Renyi Generalized Information Divergence	MIT Lincoln CAIDA, FIFA DDoSTB	√	√	√
Basicevic *et al.* (2019)	Simulation	Shannon Entropy,Renyi Generalized Entropy Tsallis Entropy, Bhatia-Singh Entropy Ubraico Entropy, Tsallis Divergence KL Divergence,Renyi Divergence	Simulated	-	√	-
Proposed Work (2020)	Emulation Testbed	Shannon Entropy, Ubriaco Entropy Renyi Generalized Entropy, ϕ-Entropy Tsallis Entropy KL Divergence,Renyi Generalized Information Divergence Bhattacharyya Distance, Jensen Shannon Distance Hellinger Distance, Pearson Distance Total Variation Distance, Euclidean Distance Jeffrey Distance, ϕ-Divergence	MIT Lincoln CAIDA, FIFA DDoSTB	√	√	√

tection rate of 100% in comparison to Shannon entropy (51.3%) and Ubriaco entropy (6%). For all other detection system evaluation parameters, Renyi entropy produces the best results as compared to other entropy metrics.

For FE traffic, ϕ-entropy produces the best results on all detection system evaluation parameters. For LR-DDoS attack traffic, Renyi entropy produces the best detection accuracy and precision.

For divergence metrics, it has observed that for HR-DDoS attack, KL divergence, Jeffrey distance, Bhattacharyya distance, total variation distance, GID and ϕ-divergence produces 100% detection accuracy as compared to Hellinger distance (97.5%), JSD (999.6%) and Euclidean distance (87.5%). For FE traffic and LR-DDoS attack, both ϕ-divergence and Bhattacharyya distance gave the best results.

Further, Figure-4 shows the trade-off between detection rate and FPR in terms of receiver operating characteristic (ROC) curves for all the entropy and divergence metrics. ROC curves depict that with an increase in FPR, detection rate increases, i.e. if we compromise on FPR, a better detection rate can be achieved and vice-versa.

5.4 Comparison with similar works

Many authors have compared the performance of information theory-based metrics in the past. However, the proposed work presented in this paper differ from existing works in many ways and is summarized in Table 9:

- Xiang *et al.* (2011) compared the performance of Shannon entropy, Generalized entropy, KL divergence, and Generalized information divergence in detecting low-rate DDoS (LR-DDoS) attacks from legitimate traffic. The authors used the existing datasets of MIT Lincoln to represent legitimate traffic and CAIDA to represent HR-DDoS traffic. We extended the idea of this paper to detect LR-DDoS attack, HR-DDoS attack and discriminating the attack traffic from similar-looking FE traffic. We performed experiments in emulation based DDoSTB testbed for validating the proposed approach, whereas the authors of Xiang *et al.* (2011) used real datasets. Moreover, in this paper, we empirically investigated the performance of five entropy metrics along with ten divergence measures.

- Behal & Kumar (2017*a*) also compared the performance of Shannon entropy, Generalized entropy (GE), KL divergence and Generalized information divergence (GID) in detecting HR-DDoS attack, LR-DDoS attack, and FE traffic from legitimate traffic. The authors claimed that GE and GID metrics are more suitable to identify different types of DDoS attacks and FEs. In contrast, our contribution in this paper is that in most of the cases, ϕ-generalized entropy and ϕ-generalized divergence measures are best suited to detect different types of Netflow. Moreover, in this paper, a generalized detection algorithm along with mathematical models of various Netflows are proposed.

- Bereziński *et al.* (2015) proposed a generalized entropy-based anomaly detection system. The authors compared the performance of generalized Tsallis entropy, generalized Renyi's entropy and Shannon entropy. They used an AI-based WEKA tool to validate their proposed work. They used simulation tools to create a dataset and synthetically inserted anomalies in the traffic. The authors claimed that Tsallis entropy is best suited to detect HR-DDoS attack traffic, whereas the scope of the current paper is more as it can detect HR-DDoS and LR-DDoS attacks along with FEs.

- Bhuyan *et al.* (2015) in their paper, empirically investigated the performance of Shannon entropy and Renyi's generalized entropy in detecting HR-DDoS attacks from legitimate traffic. Authors used existing CAIDA dataset and their own created TUIDS dataset for validating the proposed approach. Whereas, authors in this paper have detected LR-DDoS attacks, HR-DDoS and FE traffic using real datasets as well as synthetically generated datasets using DDoSTB.

- Yu *et al.* (2009) compared the performance of information theory-based divergence measures like Sibson distance, Hellinger distance and Jeffrey distance for detecting HR-DDoS attacks and FE traffic based on the idea of flow similarity. They used real datasets of NLANR Auckland VIII for representing FE traffic, MIT Lincoln DDOS dataset for attack traffic. Authors observed that out of the three divergence measures, Sibson distance is more suitable to discriminate DDoS attack traffic from FE traffic. However, the authors did not consider differentiating LR-DDoS attack traffic and normal traffic from FE traffic as we did.

- Basicevic & Ocovaj (2019) compared the performance of Shannon entropy, Tsallis entropy, Renyi entropy, Bhatia-Singh entropy, Ubriaco entropy, KL divergence, Tsallis divergence and Renyi divergence similar to our work. The authors perform simulation-based experiments to create a synthetic dataset for validation purpose, whereas we have used real datasets for validating the proposed approach. Further, the scope of their work is minimal. Their proposed method is capable of detecting only network layer SYN flooding, whereas our proposed approach is more generalized and has a broader scope. Our proposed approach can identify different types of LR-DDoS attack and HR-DDoS attack along with discriminating them from similar-looking FE traffic.

6. Conclusion and future directions

The DDoS attack poses a severe threat to online services and network resources. The impact of DDoS attacks can be very devastating. So, in-time detection of such attacks is an essential characteristic of any network anomaly detection system. In recent times, information theory-based distinct entropy and divergence metrics have been used increasingly in the domain of network anomaly detection. This paper attempts to empirically investigate the performance of these predominantly used information theory-based entropy metrics such as Shannon entropy, Ubriaco entropy, Renyi's generalized entropy, Tsallis entropy, ϕ-entropy, and various divergence metrics such as KL divergence, Jeffrey distance, Bhattacharyya distance, Jensen Shannon divergence, Hellinger distance, Pearson distance, Total variation distance, Euclidean distance, generalized information divergence and ϕ-divergence measures.

As part of the work, a generalized flow-based detection algorithm is used that works based on information distance (ID) between the different types of Netflow. We observed that the divergence metrics exploit more information distance between different Netflows as compared to entropy metrics, hence lead to more detection accuracy as compared to entropy metrics. Out of the various entropy metrics, for HR-DDoS attack, Tsallis entropy, Renyi entropy and ϕ-entropy produce the best detection accuracy. For all other detection system evaluation parameters, Renyi entropy produces the best results. For detecting FE traffic, ϕ-entropy produce the best results on all detection system evaluation parameters. However, for detecting meek variations in the network traffic such as LR-DDoS attack traffic, Renyi entropy produce the best results.

Similarly, we observed in the case of divergence metrics that for HR-DDoS attack, KL divergence, Jeffrey distance, Bhattacharyya distance, total variation distance, GID, and ϕ-divergence produce 100% detection accuracy as compared to Hellinger distance (97.5%), JSD (99.6%) and Euclidean distance (87.5%). Whereas for detecting FE traffic and LR-DDoS attack traffic, both ϕ-divergence and Bhattacharyya distance gave the best results.

For future work, the researchers shall 1) propose an ISP level distributed approach to mitigate the impact of FEs and DDoS attacks on the network resources using information theory-based generalized divergence metric, and (2) implement and validate the distributed algorithm in Software Defined Networking (SDN) domain.

ACKNOWLEDGEMENT

This research work has been supported by the All India Council for Technical Education(AICTE), New Delhi, India under Grant of Research Promotion Scheme(RPS) under Grant No. 8023/RID/RPS-93/2011-12.

References

Alcorn, J. A. & Chow, C. E. (2014), A framework for large-scale modeling and simulation of attacks on an openflow network, *in* '2014 23rd International Conference on Computer Communication and Networks (ICCCN)', IEEE: pp. 1–6.

Basicevic, I. & Ocovaj, S. (2019) , 'Application of entropy formulas in detection of denial-of-service attacks', *International Journal of Communication Systems*: p. e4067.

Basicevic, I., Ocovaj, S. & Popovic, M. (2015) , 'Use of tsallis entropy in detection of syn flood dos attacks', *Security and Communication Net-works* **8**(18): 3634–3640.

Behal, S. & Kumar, K. (2017*a*) , 'Detection of ddos attacks and flash events using informa-tion theory metrics–an empirical investigation', *Computer Communications* **103**: 18–28.

Behal, S. & Kumar, K. (2017*b*), 'Detection of ddos attacks and flash events using novel information theory metrics', *Computer Networks* **116**: 96–110.

Bereziński, P., Jasiul, B. & Szpyrka, M. (2015), 'An entropy-based network anomaly detection method', *Entropy* **17**(4): 2367–2408.

Bhandari, A., Sangal, A. L. & Kumar, K. (2016) , 'Characterizing flash events and distributed denial-of-service attacks: an empirical investi-gation', *Security and Communication Networks* **9**(13): 2222–2239.

Bhatia, P. & Singh, S. (2013), 'On a new csiszar's f-divergence measure', *Cybernetics and information technologies* **13**(2): 43–57.

Bhatia, S., Schmidt, D. & Mohay, G. (2012), Ensemble-based ddos detection and mitigation model, *in* 'Proceedings of the Fifth International Conference on Security of Information and Networks', ACM: pp. 79–86.

Bhuyan, M. H., Bhattacharyya, D. & Kalita, J. K. (2015), 'An empirical evaluation of information metrics for low-rate and high-rate ddos attack detection', *Pattern Recognition Letters* **51**: 1–7.

Bhuyan, M. H., Bhattacharyya, D. & Kalita, J. K. (2016), 'E-ldat: a lightweight system for ddos flooding attack detection and ip traceback using extended entropy metric', *Security and Communication Networks* **9**(16): 3251–3270.

Bhuyan, M. H. & Elmroth, E. (2018) , Multi-scale low-rate ddos attack detection using the generalized total variation metric, *in* '2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)', IEEE: pp. 1040–1047.

Campbell, A. T., De Meer, H. G., Kounavis, M. E., Miki, K., Vicente, J. B. & Villela, D. (1999), 'A survey of programmable networks', *ACM SIGCOMM Computer Communication Review* **29**(2): 7–23.

Chen, E. Y. & Yonezawa, A. (2005) , Practical techniques for defending against ddos attacks, *in* 'The 3rd ACS/IEEE International Conference onComputer Systems and Applications, 2005.', IEEE: p. 72.

Clicks (2019), ‘https://www.clickz.com/ internetgrowth-usage-stats-2019-time-online-devicesusers/ 235102/’.

Crooks, G. E. (2017), ‘On measures of entropy and information’, Tech. Note 9: v4. Fall, K. (1999), Network emulation in the vint/ns simulator, in ‘Proceedings IEEE International Symposium on Computers and Communications (Cat. No. PR00250)’, IEEE: pp. 244–250.

Feinstein, L., Schnackenberg, D., Balupari, R. & Kindred, D. (2003), Statistical approaches to ddos attack detection and response, in ‘Proceedings DARPA informationsurvivability conference and exposition’, Vol. 1, IEEE: pp. 303–314.

Fundation, O. N. (2012), ‘Software-defined networking: The new norm for networks’, ONF White Paper **2**: 2–6.

Ghorbani, A. A., Lu,W. & Tavallaee, M. (2010), Network attacks, in ‘Network Intrusion Detection and Prevention’, Springer: pp. 1–25.

Global Threat landscape Report (2019). URL: https://www.imperva.com/blog/2019-global-ddos-threat-landscape-report/

Hellinger, E. (1909), ‘Neue begr¨undung der theorie quadratischer formen von unendlichvielen ver¨anderlichen.’, Journal f¨ur die reine und angewandte Mathematik (Crelles Journal 1909(136): 210–271.

ITA(1998),‘http://ita.ee.lbl.gov/html/contrib/ worldcup.html’. 25/3/2019:Accessed on 16/03/2020.

Jeyanthi, N. & Iyengar, N. C. S. N. (2012), ‘An entropy based approach to detect and distinguish ddos attacks from flash crowds in voip networks.’, IJ Network Security 14(**5**): 257–269.

Jung, J., Krishnamurthy, B. & Rabinovich, M. (2002), 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’: pp. 293–304.

Kumar, K., Joshi, R. & Singh, K. (2007), A distributed approach using entropy to detect ddos attacks in isp domain, in ‘2007 International Conference on Signal Processing, Communications and Networking’, IEEE:pp. 331–337.

Lee, K., Kim, J., Kwon, K. H., Han, Y. & Kim, S. (2008), ‘Ddos attack detection method using cluster analysis’, Expert systems with applications 34(3): 1659–1665.

Li, K., Zhou, W., Li, P., Hai, J. & Liu, J. (2009), Distinguishing ddos attacks from flash crowds using probability metrics, in ‘2009 Third International Conference on Network and System Security’, IEEE: pp. 9–17.

Li, K., Zhou, W. & Yu, S. (2009), ‘Effective metric for detecting distributed denial-of-service attacks based on information divergence’, IET communications 3(**12**): 1851–1860.

Li, L., Zhou, J. & Xiao, N. (2007), Ddos attack detection algorithms based on entropy computing, in ‘International Conference on Information and Communications Security’, Springer: pp. 452–466.

Lin, J. (1991), ‘Divergence measures based on the shannon entropy’, IEEE Transactions on Information theory 37(1): 145–151.

Lincoln laboratory LLSDDos0.2.2 dataset, M. (n.d.) , ‘https://www.ll.mit.edu/rd/datasets/2000-darpa-intrusion-detectionscenario-specific-datasets’.:25/03/2019.

Ma, X. & Chen, Y. (2013), 'Ddos detection method based on chaos analysis of network traffic entropy', IEEE Communications Letters 18(**1**): 114–117.

Machado, J. T. (2010), 'Entropy analysis of integer and fractional dynamical systems', Nonlinear Dynamics 62(**1-2**): 371–378**.**

Nadeem, A. & Howarth, M. (2009), Adaptive intrusion detection & prevention of denial of service attacks in manets, in 'Proceedings of the 2009 international conference on wireless communications and mobile computing: Connecting the world wirelessly': pp. 926–930.

Nadeem, A. & Howarth, M. P. (2013), 'A survey of manet intrusion detection & prevention approaches for network layer attacks', IEEE communications surveys & tutorials 15(**4**): 2027– 2045.

Netscout WISR Report (2019). URL:https:// www.netscout.com/pressreleases/netscout-releases-14th-annualworldwide-infrastructure

Patil, N. V., Krishna, C. R., Kumar, K. & Behal, S. (2019), 'E-had: A distributed and collaborative detection framework for early detection of ddos attacks', Journal of King Saud University- Computer and Information Sciences .

Pearson, K. (1900), 'X. on the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling', The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 50(**302**): 157–175.

Plastino, A. & Plastino, A. (1993), 'Tsallis' entropy, ehrenfest theorem and information theory', Physics Letters A 177(**3**): 177–179.

Popa, L., Ghodsi, A. & Stoica, I. (2010) , Http as the narrow waist of the future internet, in 'Proceedings of the 9th ACM SIGCOMMWorkshop on Hot Topics in Networks': pp. 1–6.

Prasad, K. M., Reddy, A. R. M. & Rao, K. V. (2013), 'Discriminating ddos attack traffic from flash crowds on internet threat monitors (itm) using entropy variations', African J. Comput. ICT 6(**2**): 53–62.

Rahmani, H., Sahli, N. & Kamoun, F. (2012a), 'Ddos flooding attack detection scheme based on f-divergence', Computer Communications 35(**11**): 1380–1391.

Rahmani, H., Sahli, N. & Kamoun, F. (2012b), 'Distributed denial-of-service attack detection scheme-based joint-entropy', Security and Communication Networks 5(**9**): 1049–1061.

R´enyi, A. (1965), 'On the foundations of information theory', Revue de l'Institut International de Statistique :pp. 1–14.

Sachdeva, M. & Kumar, K. (2014), 'A trafficcluster entropy based approach to distinguish ddos attacks from flash event using deter testbed', ISRN Communications and Networking: 2014.

Sachdeva, M., Kumar, K.&Singh, G. (2016), 'A comprehensive approach to discriminate ddos attacks from flash events', Journal of information security and applications **26**: 8–22.

Sahoo, K. S., Puthal, D., Tiwary, M., Rodrigues, J. J., Sahoo, B. & Dash, R. (2018), 'An early detection of low rate ddos attack to sdn based data center networks using information distance metrics', Future Generation Computer Systems **89**: 685–697.

Sahoo, K. S., Tiwary, M. & Sahoo, B. (2018), Detection of high rate ddos attack from flash events using information metrics in software defined networks, in '2018 10th International Conference on Communication Systems&Networks (COMSNETS)', IEEE: pp. 421–424.

Saleh, M. A. & Abdul Manaf, A. (2015), 'A novel protective framework for defeating httpbased denial of service and distributed denial of service attacks', The Scientific World Journal :2015.

Salem, O., Na¨ıt-Abdesselam, F. & Mehaoua, A. (2012), Anomaly detection in network traffic using jensen-shannon divergence, in '2012 IEEE International Conference on Communications (ICC)', IEEE: pp. 5200–5204.

Saravanan, R., Shanmuganathan, S. & Palanichamy, Y. (2016), 'Behavior-based detection of application layer distributed denial of service attacks during flash events', Turkish Journal of Electrical Engineering & Computer Sciences 24(**2**): 510–523.

Sengar, H.,Wang, X.,Wang, H.,Wijesekera, D. & Jajodia, S. (2009), Online detection of network traffic anomalies using behavioraldistance, in '2009 17th InternationalWorkshop on Quality of Service', IEEE: pp. 1–9.

Shannon, C. E. (2001), 'A mathematical theory of communication', ACM SIGMOBILE Mobile Computing and Communications Review 5(**1**): 3–55.

Singh, K., Dhindsa, K. S. & Nehra, D. (2020), 'T-cad: A threshold based collaborative ddos attack detection in multiple autonomous systems', Journal of Information Security and Applications 51: 102457.

Tao, Y. & Yu, S. (2013), Ddos attack detection at local area networks using information theoretical metrics, in '2013 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications', IEEE: pp. 233–240.

The CAIDA DDoS Attack Dataset, "Cooperative Analysis for Internet Data Analysis",https://www.caida.org/data/passive/ddos-20070804-dataset.xml (2010).

The vCORE Emulator, http://www.nrl.navy.mil/itd/ncs/products/core(2016).

.

Tritilanunt, S., Sivakorn, S., Juengjincharoen, C. & Siripornpisan, A. (2010), Entropy-based input-output traffic mode detection scheme for dos/ddos attacks, in '2010 10th International Symposium on Communications and Information Technologies', IEEE: pp. 804–809**.**

Tsallis, C. (1988), 'Possible generalization of boltzmann-gibbs statistics', Journal of statistical physics 52(1-2): 479–487.

Ubriaco, M. R. (2009), 'Entropies based on fractional calculus', Physics Letters A 373(**30**): 2516–2519.

Wang, F., Wang, H., Wang, X. & Su, J. (2012), A new multistage approach to detect subtle ddos attacks', Mathematical and Computer Modelling 55(1-2): 198–213

Wang, L. & Liu, Y. (2020), A ddos attack detection method based on information entropy and deep learning in sdn, in '2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC', Vol. 1, IEEE: pp. 1084–1088.

Wen, S., Jia, W., Zhou, W., Zhou, W. & Xu, C. (2010), Cald: Surviving various applicationlayer ddos attacks that mimic flash crowd, in '2010 Fourth International Conference on Network and System Security', IEEE: pp. 247–254.

White, B., Lepreau, J., Stoller, L., Ricci, R., Guruprasad, S., Newbold, M., Hibler, M., Barb, C. & Joglekar, A. (2002), 'An integrated experimental environment for distributed systems and networks', ACM SIGOPS Operating Systems Review 36(**SI**): 255–270.

Xiang, Y., Li, K. & Zhou, W. (2011), 'Lowrate ddos attacks detection and traceback by using new information metrics', IEEE transactions on information forensics and security 6(**2**): 426– 437.

Yu, S., Thapngam, T., Liu, J., Wei, S. & Zhou, W. (2009), Discriminating ddos flows from flash crowds using information distance, in '2009 Third International Conference on Network and System Security', IEEE: pp. 351–356.

Yu, S., Zhou,W. & Doss, R. (2008), 'Information theory based detection against network behavior mimicking ddos attacks', IEEE Communications Letters 12(**4**): 318–321.

Yu, S., Zhou, W., Jia, W., Guo, S., Xiang, Y. & Tang, F. (2011), 'Discriminating ddos attacks from flash crowds using flow correlation coefficient', IEEE Transactions on Parallel and Distributed Systems 23(**6**): 1073–1080.

Zhang, L., Estrin, D., Burke, J., Jacobson, V., Thornton, J. D., Smetters, D. K., Zhang, B., Tsudik, G., Massey, D., Papadopoulos, C. (2010), 'Named data networking (ndn) project', Relat ´orio T ´ecnico NDN-0001, Xerox Palo Alto Research Center-PARC 157: 158.

Zhijun, W., Qing, X., Jingjie, W., Meng, Y. & Liang, L. (2020), 'Low-rate ddos attack detection based on factorization machine in software defined network', IEEE Access **8**: 17404–17418.

Submitted: 21/09/2020
Revised: 18/11/2020
Accepted: 01/12/2020
DOI: 10.48129/kjs.v48i4.10612