Low-Rate DDoS Attacks Detection and Traceback by Using New Information Metrics

Yang Xiang, *Member, IEEE*, Ke Li, and Wanlei Zhou, *Senior Member, IEEE*

***Abstract*—A low-rate distributed denial of service (DDoS) attack has significant ability of concealing its traffic because it is very much like normal traffic. It has the capacity to elude the current anomaly-based detection schemes. An information metric can quantify the differences of network traffic with various probability distributions. In this paper, we innovatively propose using two new information metrics such as the generalized entropy metric and the information distance metric to detect low-rate DDoS attacks by measuring the difference between legitimate traffic and attack traffic. The proposed generalized entropy metric can detect attacks several hops earlier (three hops earlier while the order $\alpha = 10$) than the traditional Shannon metric. The proposed information distance metric outperforms (six hops earlier while the order $\alpha = 10$) the popular Kullback–Leibler divergence approach as it can clearly enlarge the adjudication distance and then obtain the optimal detection sensitivity. The experimental results show that the proposed information metrics can effectively detect low-rate DDoS attacks and clearly reduce the false positive rate. Furthermore, the proposed IP traceback algorithm can find all attacks as well as attackers from their own local area networks (LANs) and discard attack traffic.**

***Index Terms*—Attack detection, information metrics, IP traceback, low-rate distributed denial of service (DDoS) attack.**

I. Introduction

THE distributed denial of service (DDoS) attack is a serious threat to the security of cyberspace. It typically exhausts bandwidth, processing capacity, or memory of a targeted machine or network. A DDoS attack is a distributed, cooperative and large-scale attack. It has been widely spread on wired [1] or wireless networks [2]. A low-rate DDoS attack is an intelligent attack as the attacker can send attack packets to the victim at a sufficiently low rate to elude detection. Today, a large-scale DDoS attack is usually combined with multiple low-rate attacks, which are distributed on the Internet to avoid being detected by current detection schemes. An attacker can use botnets to launch a low-rate DDoS attack, producing network behavior that appears normal. Therefore, it is difficult to detect and mitigate such attacks [3].

Manuscript received May 30, 2010; revised November 29, 2010; accepted January 11, 2011. Date of publication January 20, 2011; date of current version May 18, 2011. This work was supported in part by ARC Discovery Project DP1095498 and in part by ARC Linkage Project LP100100208. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Yong Guan.

The authors are with the School of Information Technology, Deakin University, Burwood, VIC 3125, Australia (e-mail: yang@deakin.edu.au; like309@gmail.com; wanlei@deakin.edu.au).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIFS.2011.2107320

A. Motivation

Currently, DDoS attack detection metrics are mainly separated into two categories: the signature-based metric and anomaly-based metric. The signature-based metric depends on technology that deploys a predefined set of attack signatures such as patterns or strings as signatures to match incoming packets. The anomaly-based detection metric typically models the normal network (traffic) behavior and deploys it to compare differences with incoming network behavior. Anomaly-based detection has many limitations. First, in anomaly-based detection systems, attackers can train detection systems to gradually accept anomaly network behavior as normal. Second, the false positive rate using the anomaly-based detection metric is usually higher than the one using the signature-based detection metric. It is difficult to set the proper thresholds which help to balance the false positive rate and the false negative rate. Third, it is very difficult to extract the features of normal and anomalous network behaviors precisely. An anomaly-based detection metric uses a predefined specific threshold, such as an abnormal deviation of some statistical characteristics from normal network traffic, to identify abnormal traffic amongst all normal traffic. Therefore, the utilization and choice of statistical methods and tools is vitally important [4]. It is generally accepted that the fractional Gaussian noise function can be used to simulate real network traffic in aggregation and the Poisson distribution function can be used to simulate the DDoS attack traffic in aggregation [5]–[9].

Therefore, many information-theory-based metrics have been proposed to overcome the above limitations. In information theory, information entropy is a measure of the uncertainty associated with a random variable. Information distance (or divergence) is a measure of the difference between different probability distributions. Shannon's entropy and Kullback–Leibler's divergence methods have both been regarded as effective methods for detecting abnormal traffic based on IP address-distribution statistics or packet size-distribution statistics [10]–[12]. Early detection and detection accuracy (such as a low false positive rate) of DDoS attacks are the two most important criteria for the success of a defense system. In this paper, we innovatively propose two new and effective anomaly-based detection metrics which not only identify attacks earlier, but also produce lower false positive rates when compared with the traditional Shannon's entropy method and the Kullback–Leibler divergence method.

B. Contributions

The main contributions of this paper are as follows.

1) It analyzes and highlights the advantages of generalized entropy and information distance compared with Shannon entropy and Kullback–Leibler distance, respectively.

2) It proposes the generalized entropy and information distance metrics outperform the traditional Shannon entropy and Kullback–Leibler distance metrics for the low-rate DDoS attack detection in terms of early detection, lower false positive rates, and stabilities.
3) It proposes an effective IP traceback scheme based on an information distance metric that can trace all attacks back to their own local area networks (LANs) in a short time.

II. Detection Algorithms and IP Traceback Analysis

In this section, we propose and analyze two effective detection algorithms and an IP traceback scheme. In this paper, we make the following reasonable assumptions:

1) we have full control of all the routers;
2) we have extracted an effective feature of network traffic (e.g., the unforged source IP addresses) to sample its probability distribution;
3) we have obtained and stored the average traffic of the normal, as well as the local thresholds σ_{fi} and $\sigma_{f(Ri)}$ on their own routers in advance;
4) on all routers, the attack traffic obeys Poisson distribution and the normal traffic obeys Gaussian noise distribution.

A. Generalized Entropy Metric

In information theory, the information entropy is a measure of the uncertainty associated with a random variable, forming the basis for distance and divergence measurements between probability densities. The more random the information variable, the bigger the entropy. In contrast, the greater certainty of the information variable, the smaller the entropy [13]. The generalized information entropy as a generalization of Shannon entropy is one of a family of functions for quantifying either the diversity uncertainty or randomness of a system. It is a very important metric in statistics as an index of diversity.

The generalized information entropy of order α is defined as follows:

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

where P_i are the probabilities of $\{x_1, x_2 \dots x_n\}$, $P_i \geq 0$,

$$\sum_{i=1}^{n} p_i = 1, \alpha \geq 0, \alpha \neq 1.$$

When $\alpha = 0$ or the probabilities of $\{x_1, x_2 \dots x_n\}$ are all the same, we have the maximum information entropy as follows:

$$H_0(x) = \log_2 n$$

which indicates the probability density of information is maximum decentralization.

When $\alpha \to 1$, $H_\alpha(x)$ converges to Shannon entropy, the equation is as follows:

$$H_1(x) = -\sum_{i=1}^{n} p_i \log_2 p_i. \tag{2}$$

When $\alpha \to \infty$, we can obtain the minimum information entropy $H_\infty(x)$. When $H_\infty(x) = 0$, this indicates the probability density of information is at the maximum concentration.

$H_\infty(x) = -\log_2 p_i^{\max}$, where $p_i^{\max}$ is the largest probability among P_i.

In the case $\alpha > 0$, we have $(\partial/\partial\alpha)H_\alpha(x) \leq 0$; therefore, the generalized information entropy is a nonincreasing function of α.

$$\text{Namely}: H_{\alpha_1}(x) \geq H_{\alpha_2}(x) \quad \text{for} \quad \alpha_1 < \alpha_2. \tag{3}$$

Karol discussed the relations between Shannon entropy and generalized entropies of integer order α [14]. The value of generalized entropy depends on the parameter α. In particular, the more important performance for generalized entropy $(\alpha > 1)$ is that it can increase the deviation between the different probability distributions compared to when Shannon entropy is used [15], [16].

To observe and analyze the formulas of Shannon and generalized information entropy, we know that 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 $\alpha < 1$. Therefore, we can obtain different final entropy values by adjusting the α value according to different requirements.

In particular, when $\alpha = 2$, we have

$$H_2(x) = -\log_2 \sum_{i=1}^{n} p_i^2. \tag{4}$$

Based on the above analysis, we consider the different characteristics of probability distribution between the human-participating legitimate network traffic and the automatic machine-generated DDoS attack traffic and include the property of generalized information entropy of order α. We design our anomaly-based DDoS detection system based on the above analysis.

In theory, the Shannon entropy value of fractional Gaussian noise distribution is higher than that of the Poisson distribution. The generalized information entropy value is lower than the Shannon entropy value and the higher probability event can have a greater influence on the final entropy in generalized information entropy compared to Shannon entropy while $\alpha > 1$. In contrast, the generalized information entropy value is higher than the Shannon entropy value and the lower probability event can have greater influence on the final entropy in generalized information entropy than in Shannon entropy while $0 \leq \alpha < 1$.

Therefore, we can obtain much better detection results by using generalized information entropy by adjusting the value of order α of generalized entropy in DDoS detection.

B. Information Distance Metric

We consider two discrete complete probability distributions $P = (p_1, p_2, \dots, p_n)$ and $Q = (q_1, q_2, \dots, q_n)$ with $\sum_{i=1}^{n} p_i = \sum_{i=1}^{n} q_i = 1$, $1 \geq p_i \geq 0$, $1 \geq q_i \geq 0$, $i = \{1, 2, \dots n\}$.

The information divergence is a measure of the divergence between P and Q and is shown below

$$D_\alpha(P\|Q) = \frac{1}{\alpha - 1}\log_2\left(\sum_{i=1}^{n} p_i^\alpha q_i^{1-\alpha}\right), \quad \alpha \geq 0. \tag{5}$$

In fact, this is information divergence of order α and it is always nonnegative if $\alpha \geq 0$. $D_\alpha(P\|Q) = 0$ must be the minimum of the distance if, and only if $P = Q$. The exceptional case is that if P and Q are incomplete probability distributions or $\alpha < 0$, then $D_\alpha(P\|Q)$ may be negative.

As α is an arbitrary positive parameter, we can assume the following special and useful formulas according to the different α value:

$$D_0(P\|Q) = -\log_2\left(\sum_{i=1}^{n} q_i\right), \quad \alpha = 0 \tag{6}$$

$$D_1(P\|Q) = \sum_{i=1}^{n}\left(p_i \log_2(p_i/q_i)\right), \quad \alpha \to 1$$

which is the Kullback–Leibler divergence [17].

Similarly, we can test and validate the inequality as follows:

$$D_1(P\|Q) < D_1(Q\|P)$$

while P is the Poisson probability distribution and Q is the fractional Gaussian noise probability distribution.

We discuss three important properties of the information divergence: additive, asymmetric, and increasing function of α. To begin with, we prove the additive property.

Assertion 1: Let P_1 and Q_1 be two different probability distributions on the same set and let P_2 and Q_2 be two different probability distributions on another set. This means P_1 and P_2 are two statistically independent distributions of each other. The same is true for Q_1 and Q_2.

Proof:

$$\begin{aligned} &D_\alpha(P_1 \times P_2\|Q_1 \times Q_2) \\ &\quad= \frac{1}{\alpha-1}\log_2\sum_{i=1}^{n}\left[(p_1 \times p_2)_i^\alpha \times (q_1 \times q_2)_i^{1-\alpha}\right] \\ &\quad= \frac{1}{\alpha-1}\log_2\sum_{i=1}^{n} p_{1i}^\alpha q_{1i}^{1-\alpha} + \frac{1}{\alpha-1}\log_2\sum_{i=1}^{n} p_{2i}^\alpha q_{2i}^{1-\alpha} \\ &\quad= D_\alpha(P_1\|Q_1) + D_\alpha(P_2\|Q_2). \end{aligned} \tag{7}$$

In general, if $P = P_1 \times P_2 \times \cdots \times P_n$, $Q = Q_1 \times Q_2 \times \cdots \times Q_n$, we have

$$D_\alpha(P\|Q) = D_\alpha(P_1\|Q_1) + D_\alpha(P_2\|Q_2) + \cdots D_\alpha(P_n\|Q_n).$$

Particularly, when $P = P_1 \times P_2$, $Q = Q_1 \times Q_2$, $P_1 = P_{11} \times P_{12}$, $P_2 = P_{21} \times P_{22}$, $Q_1 = Q_{11} \times Q_{12}$, $Q_2 = Q_{21} \times Q_{22}$, we have

$$\begin{aligned} D_\alpha(P\|Q) &= D_\alpha(P_1\|Q_1) + D_\alpha(P_2\|Q_2) \\ &= D_\alpha(P_{11}\|Q_{11}) + D_\alpha(P_{12}\|Q_{12}) \\ &\quad + D_\alpha(P_{21}\|Q_{21}) + D_\alpha(P_{22}\|Q_{22}). \end{aligned} \tag{8}$$

This additive property is very useful because it implies that aggregated traffic can be seen as the sum of individual traffic and, therefore, it is the theoretical basis of the collaborative detection or multipoint detection [15], [18]. We will design a collaborative DDoS detection algorithm later (shown as Listing 1) based on this property.

Second, we discuss the asymmetric property of divergence.

Assertion 2: Let P and Q be two different probability distributions on the same set, then $D_\alpha(P\|Q)$ is a directed divergence, and in general, $D_\alpha(P\|Q) \neq D_\alpha(Q\|P)$ while $P \neq Q$. This in fact means $D_\alpha(P\|Q)$ is not a metric.

Proof: We assume $P \neq Q$, so we have $\sum_{i=1}^{n} q_i^{1-2\alpha} \neq \sum_{i=1}^{n} p_i^{1-2\alpha}$, we further have $\sum_{i=1}^{n} p_i^\alpha q_i^{1-\alpha} \neq \sum_{i=1}^{n} q_i^\alpha p_i^{1-\alpha}$, and finally we have

$$\frac{1}{\alpha-1}\log_2\left(\sum_{i=1}^{n} p_i^\alpha q_i^{1-\alpha}\right) \neq \frac{1}{\alpha-1}\log_2\left(\sum_{i=1}^{n} q_i^\alpha p_i^{1-\alpha}\right).$$

Namely, $D_\alpha(P\|Q) \neq D_\alpha(Q\|P)$.

The asymmetric property is an important property of information divergence as the direction of divergence used in detecting DDoS attacks can influence the effectiveness of the method. Namely, in general,

$$D_\alpha(P\|Q) \neq D_\alpha(Q\|P), \quad \text{while} \quad P \neq Q. \tag{9}$$

When P is the Poisson probability distribution, and Q is the fractional Gaussian noise probability distribution, we can test and validate the inequality

$$D_\alpha(P\|Q) < D_\alpha(Q\|P). \tag{10}$$

To use information divergence as a metric, we need to overcome the asymmetric property. Here we propose the information distance as defined as follows.

Definition: We name $D_\alpha(P, Q)$ defined as follows as the information distance:

$$\begin{aligned} &D_\alpha(P, Q) \\ &\quad= D_\alpha(P\|Q) + D_\alpha(Q\|P) \\ &\quad= \frac{1}{\alpha-1}\log_2\left(\sum_{i=1}^{n} p_i^\alpha q_i^{1-\alpha} \times \sum_{i=1}^{n} q_i^\alpha p_i^{1-\alpha}\right). \end{aligned} \tag{11}$$

Obviously, $D_\alpha(P, Q)$ is a symmetric measure and always is not less than $D_\alpha(P\|Q)$ and $D_\alpha(Q\|P)$. It should be noted that $D_\alpha(P, Q)$ is undefined if $p_i = 0$ or $q_i = 0$ while $\alpha > 1$. This means that the distribution of P and Q must be absolutely continuous with respect to each other.

Likewise, we can also define the Kullback–Leibler distance as

$$\begin{aligned} D_1(P, Q) &= D_1(P\|Q) + D_1(Q\|P) \\ &= \sum_{i=1}^{n}\left((p_i - q_i)\log_2\frac{p_i}{q_i}\right). \end{aligned} \tag{12}$$

Since $p_i \neq 0$ and $q_i \neq 0$, this means that P and Q have to be absolutely continuous probability distributions.

From the above definitions and formulas, we can see that both $D_\alpha(P, Q)$ and $D_1(P, Q)$ are symmetric measures and are more than their own asymmetric divergences.

We can now verify that both $D_\alpha(P, Q)$ and $D_1(P, Q)$ are metrics. In order to be considered metrics, they must wholly satisfy the properties of identity, symmetry, and triangle inequality. In fact we can show that

$$\forall\, P, Q, L \in R^+.$$

We have the following:

1) Identity property:

$$D_\alpha(P,Q)=0, D_1(P,Q)=0;\ \text{while } P=Q. \tag{13}$$

2) Symmetry property:

$$D_\alpha(P,Q)=D_\alpha(Q,P), D_1(P,Q)=D_1(Q,P). \tag{14}$$

3) Triangle inequality:

$$\begin{aligned} D_\alpha(P,Q) &\le D_\alpha(P,L)+D_\alpha(L,Q) \\ D_1(P,Q) &\le D_1(P,L)+D_1(L,Q). \end{aligned} \tag{15}$$

Therefore, both $D_\alpha(P,Q)$ and $D_1(P,Q)$ are metrics and can be used as distance measures in DDoS attack detection.

Finally, we discuss the third property of information divergence.

Assertion 3: Both $D_\alpha(P\|Q)$ and $D_\alpha(Q\|P)$ are the increasing functions in α while $\alpha>1$.

This is because they are both the convex functions in α while $\alpha>1$ [19]. Obviously the information distance also has additive and increasing properties.

According to the above discussion, we design the collaborative detection algorithm as shown in Listing 1 to detect a DDoS attack and discard its packets.

Listing 1. A collaborative DDoS attack detection algorithm

1. Set the sampling frequency as f, the sampling period as T, and the collaborative detection threshold as σ.
2. In routers R_1 and R_2 of Fig. 1, sampling the network traffic comes from the upstream routers R_3, R_4, R_5, R_6 and LAN_1, LAN_2 in parallel.
3. Calculate in parallel the numbers of packet which have various recognizable characteristics (e.g., the source IP address or the packet's size, etc.) in each sampling time interval $\tau(\tau=1/f)$ within T.
4. Calculate the probability distributions of the network traffic come from R_3, R_4, LAN_1 and R_5, R_6, LAN_2 in parallel.
5. Calculate their distances on router R_1 and R_2, respectively, using the formula
$$D_\alpha(P,Q)=D_\alpha(P\|Q)+D_\alpha(Q\|P).$$
6. Sum the distances.
7. If the summed distance is more than the collaborative detection threshold σ, then the system detects the DDoS attack, and begins to raise an alarm and discards the attack packets; otherwise the routers forward the packets to the downstream routers.
8. Return to step 2.

To illustrate this algorithm, we use the network topology of Fig. 1 as an example. Our algorithm can not only detect DDoS attacks at router R_0 via a single-point detection, but also can detect attacks using a collaborative detection at routers R_1, R_2 or at R_3, R_4, R_5, and R_6. The processing flowchart of the collaborative detection algorithm is shown as Fig. 2. Compared with single-point detection, we can detect attacks earlier by using the collaborative detection approach because traffic can be analyzed in upper stream routers instead of just in the victim's router.

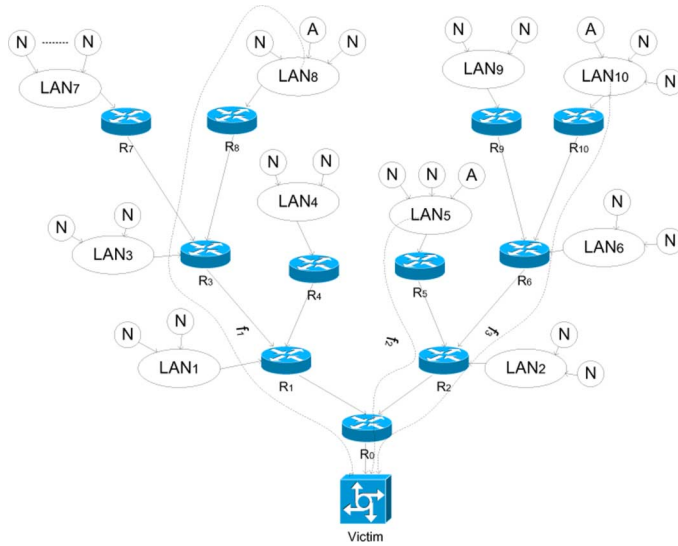


Fig. 1. Simple and partial scenario of low-rate DDoS attacks on a victim; N indicates normal client and A indicates attacker.

In information theory, we know that both information divergence and information distance are nonnegative values and the sum of the divergences or distances is always greater than themselves. In the meantime, both the divergence and distance are increasing with order α. While $\alpha>1$, we can increase the divergence or distance between legitimate traffic and attack traffic to distinguish DDoS attacks easily and earlier by increasing the value of order α and summing the divergences or distances in collaborative detection. Therefore, in DDoS attack detection, we can take full advantage of the additive and increasing properties in α of the information divergence and the information distance to enlarge the distance or gap between legitimate traffic and attack traffic. This means we can find and raise alarms for DDoS attacks early and accurately with a lower false positive rate.

C. IP Traceback Analysis

IP traceback [20] is the ability to find the source of an IP packet without relying on the source IP field in the packet, which is often spoofed. We combine our DDoS attacks detection metric with IP traceback algorithm and filtering technology together to form an effective collaborative defense mechanism against network security threats in Internet.

In hop-by-hop IP tracing, the more hops the more tracing processes, thus the longer time will be taken. In order to convenience for IP traceback algorithm analysis, we classify two types of traffic in Figs. 1 and 3 as local traffic and forward traffic, respectively. The local traffic of R_i is the traffic generated from its LAN_i, the forward traffic of R_i is the sum of its local traffic and the traffic forwarded from its immediate upstream routers. In this paper, we propose an IP tracback algorithm that can trace the source (zombies) of the attack up to its local administrative network; Listing 2 illustrates this algorithm.

Listing 2. An IP traceback algorithm in DDoS attacks detection

```
IP_Traceback_Algorithm ()
  {
  while(true)
    call Check_ForwardTraffic(0)//check attacks on
  router R_0 (or victim)
  }

  Check_ForwardTraffic (i)
  {
  calculate information distance D_f(Ri)
  if D_f(Ri) > σ_f(Ri)
    call Check_LocalTraffic (i)
    for j = 1 to n
      k = the ID of the jth immediate upstream router
  of router Ri
      call Check_ForwardTraffic (k)
    end for
  end if
  }

  Check_LocalTraffic (i)
  {
  calculate information distance D_li
  if D_li > σ_li
    stop forwarding the attack traffic to downstream
  routers (or destination), label the zombie
  end if
  }
```

We discuss the proposed IP traceback algorithm based on a sample scenario of low-rate DDoS attacks on a victim as shown in Figs. 1 and 3. When the proposed attacks detection system detects an attack on a victim, the proposed IP traceback algorithm will be launched immediately.

On router R_0, the proposed traceback algorithm calculates information distances based on variations of its local traffic and the forward traffic from its immediate upstream routers; in this paper, we set LAN_0 of router R_0 include the victim. If the information distance based on its local traffic is more than the specific detection threshold σ_{10}, the proposed detection system detects an attack in its LAN_0; this means that the detected attack is an internal attack. If the information distances based on the forward traffic from its immediate upstream routers R_1 and R_2 are both more than the specific detection threshold $\sigma_{f(R1)}$ and $\sigma_{f(R2)}$, respectively, the proposed detection system has detected attacks in routers R_1 and R_2, then on R_1 and R_2 the proposed traceback algorithm calculates information distances based on variations of their local traffic and the forward traffic from their immediate upstream routers, and will find that there are no attacks in LAN_1 and LAN_2 and R_4; therefore, on routers R_3, R_5, and R_6, the proposed algorithm calculates continually information distances based on variations of their local traffic and the forward traffic from their immediate upstream routers, then can find there is an attack (zombie) in LAN_5 so the router R_5 will stop forwarding the traffic from the zombie immediately.

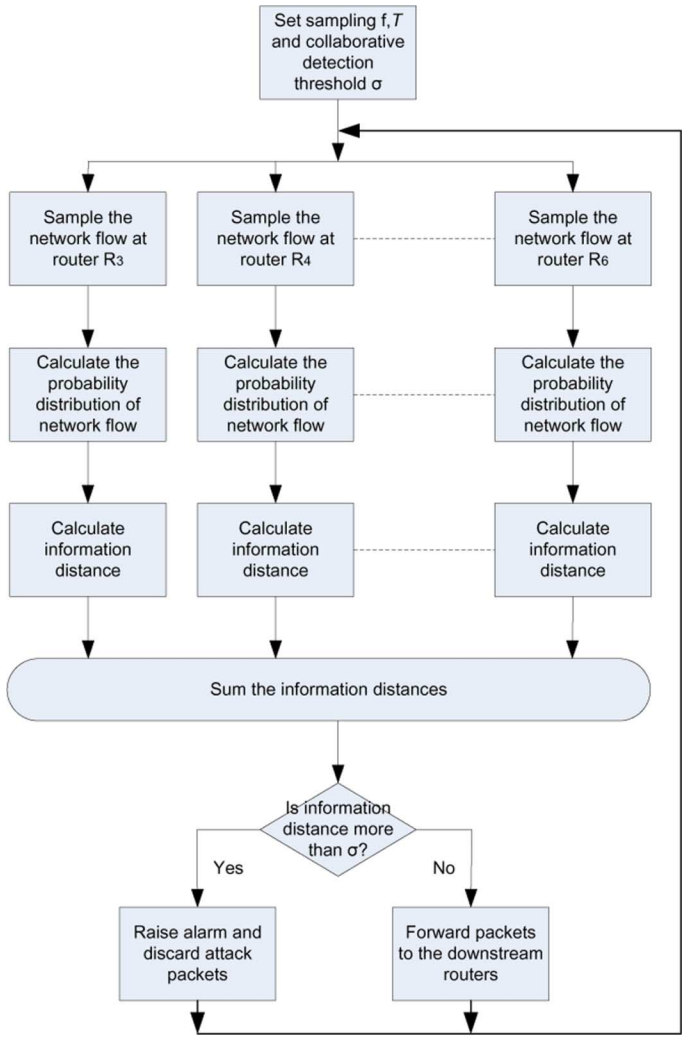


Fig. 2. Processing flowchart of the collaborative detection algorithm in DDoS attack detection system.

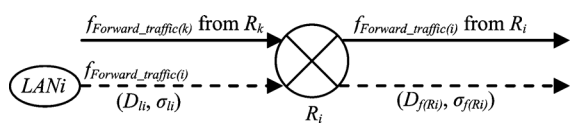


Fig. 3. Local traffic, forward traffic, information distance D, and threshold σ at a router.

Finally, the proposed algorithm can find attacks (zombies) in LAN_8 and LAN_{10}, respectively.

Therefore, based on the IP traceback algorithm, it is easy to trace back and figure out all attack routes f_1, f_2, and f_3 as shown in Fig. 1. From Listings 1 and 2, we know that the proposed traceback algorithm has lower computational cost (or time complexity) than the binary tree traversal algorithm in a binary attack tree, and has higher accuracy of traceback process as the proposed information distance metric has a lower false positive rate in attacks detection.

For the evaluation of the total traceback time, we consider the worst situation that the binary attack tree is a full branches tree and all zombies are distributed at the far ends evenly; the evaluation result is shown in Fig. 4. From Fig. 4, we know that there will be a short traceback time within 5 hops from the victim to the far end zombies, but with more than 6 hops the

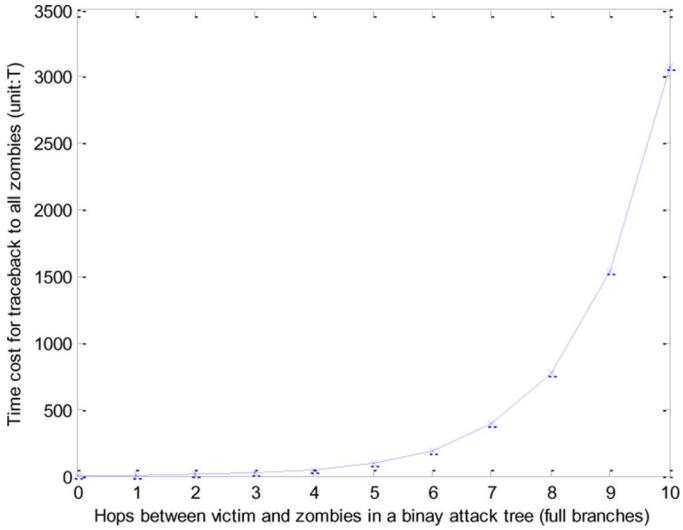


Fig. 4. Total traceback time with the variation of hops in a binary attack tree (full branches); T indicates one sampling period.

total traceback time will be increasing quickly. This is just the worst case; actually the distribution of zombies is uneven and not all of them are located in the far ends of the attack tree, thus the total traceback time will be decreasing sharply. Furthermore, further measures can be taken to reduce the total traceback time in the proposed traceback algorithm; for example, we can improve the traceback algorithm using the parallel processing method to trace back all zombies, also we can obtain and store the attack traffics of one sampling period on their own routers in advance while the proposed detection metric detects an attack on the victim.

III. Experiment Results

The proposed detection systems can use either the source IP address-based method or the IP packet size-based method to calculate the probability distribution of the traffic in the given time interval. The IP packet size-based method is to utilize the feature that attacks usually produce packets in defiance of a victim's response and when a flooding-based attack occurs, the same sized packets are generally used. On the other hand, the legitimate network traffics have typical packet sizes with respect to requests and responses or data and acknowledgments [5]. Therefore, the more concentrated the size distribution of observed IP packets, the smaller the entropy value. Similarly, the more dispersed the size distribution of IP packets size, the bigger its entropy value. The source IP address-based method is utilized when attacks from zombies occur because they usually have a more concentrated source IP address than legitimate access. Therefore, we can obtain the different information entropy value through calculating the probability distribution of the packets' source IP address. A bigger entropy value represents more randomness of the source IP addresses. Through detecting the change of the information entropy value, we can obtain the change of the source IP address distribution, and then decide whether the attack traffic is and then discard it.

In the experiment, we use the MIT Lincoln Laboratory Scenario (attack-free) inside tcpdump dataset [21] as the normal network traffic, and use the Low-rate DDoS attack scenario

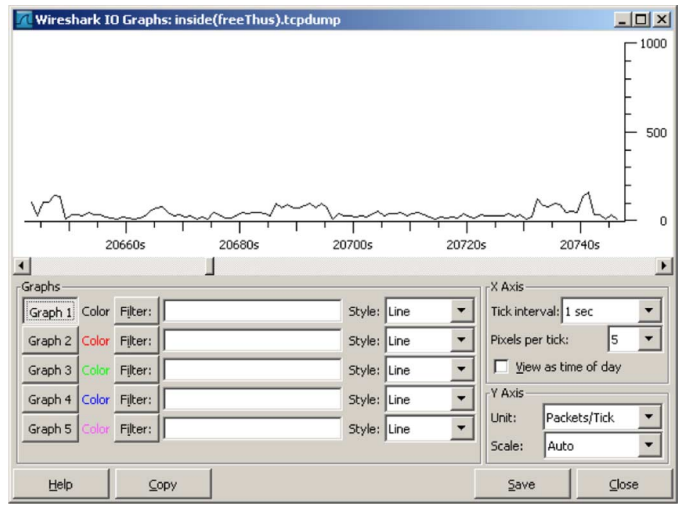


Fig. 5. Normal network traffic (attack-free) scenario from MIT/LL; X-axis denotes tick interval (second), and Y-axis denotes packets/tick (unit).

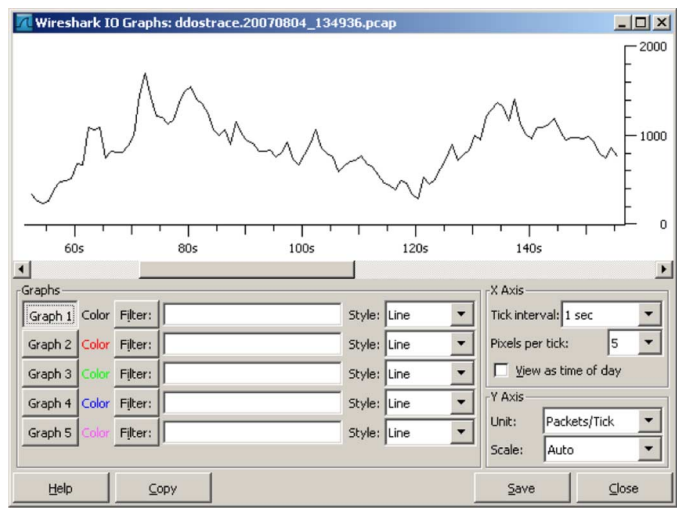


Fig. 6. Low-rate DDoS attack scenario from CAIDA; X-axis denotes tick interval (second), and Y-axis denotes packets/tick (unit).

from CAIDA [22] as the DDoS attack traffic to test the proposed algorithms. The normal network traffic scenario is the whole day data collected on Thursday in the third training week; the data do not contain any attacks. In this experiment, we let the sampling period be 300 s, so in this attack-free scenario we collect at random the normal traffic from the 20 650th to 20 950th as a sampling period. The partial traffic scenario is shown in Fig. 5. The attack scenario includes a DDoS attack run by an attacker and is performed over multiple networks. The attack dataset contains approximately 5 min (300 s) of anonymized traffic form a DDoS attack on August 4, 2007. The traces include only attack traffic to the victim and responses from the victim; nonattack traffic has been removed as much as possible. The partial attack scenario is shown in Fig. 6. Based on [23], more than 10 000 attack packets per second can achieve a high-rate attack; 1000 attack packets per second around can only achieve 60% of full attack. Therefore, this is a low-rate DDoS attack. The details of traffic feature are shown in Fig. 7.

We classify statistic IP packets and compute the probability distributions of the source IP addresses in attack and attack-free scenarios, respectively, as shown in Figs. 8 and 9. We consider

```
Maximum capture length for interface 0: 65000
First timestamp:          1186260576.487629000
Last timestamp:           1186260876.482457000
Unknown encapsulation:                       0
IPv4 bytes:                           37068253
IPv4 pkts:                              166448
IPv4 traffic:                             8079
Unique IPv4 addresses:                     136
Unique IPv4 source addresses:              132
Unique IPv4 destination addresses:         136
Unique IPv4 TCP source ports:             4270
Unique IPv4 TCP destination ports:        3348
Unique IPv4 UDP source  ports:               1
Unique IPv4 UDP destination ports:           1
Unique IPv4 ICMP type/codes:                 2
```

Fig. 7. Details of traffic feature of the low-rate DDoS attack scenario from CAIDA.

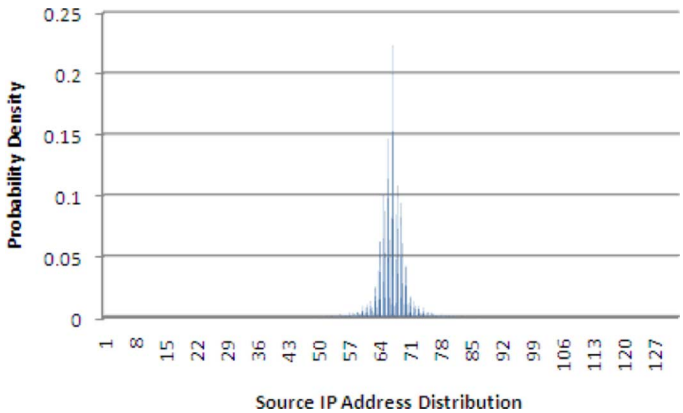


Fig. 8. Probability distribution of source IP address in low-rate DDoS attack (only attack traffic) scenario.

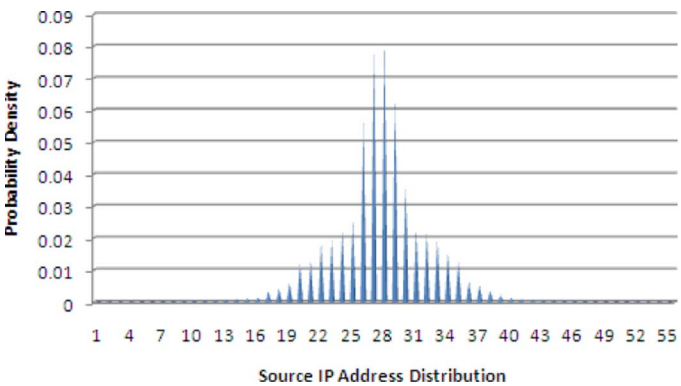


Fig. 9. Probability distribution of source IP address in normal network traffic (attack-free) scenario.

the real low-rate DDoS attack scenario in a real network environment, because the low-rate attack has not yet consumed the whole computing resources on the server or all of the bandwidth of the network connecting the server to the Internet. Therefore, a real low-rate DDoS attack scenario not only contains attack traffic but also contains attack-free traffic. In this experiment, we mix the low-rate DDoS attack traffic and the normal network traffic into a real low-rate DDoS attack scenario. Its probability distribution of source IP address is shown in Fig. 10.

A. Generalized Entropy Metric

As a comparison, we not only test the generalized entropies in varied α value but also test the Shannon entropies using the real dataset for normal (attack-free) traffic and attack traffic.

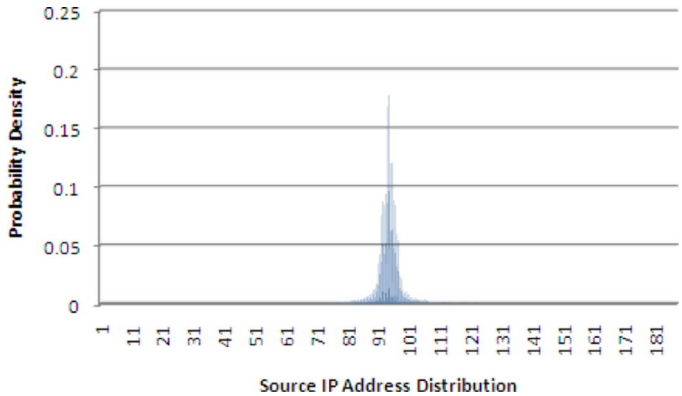


Fig. 10. Probability distribution of source IP address in a real low-rate DDoS attack (mixed traffic of attack and attack-free) scenario.

TABLE I
COMPARISON OF SHANNON ENTROPY AND GENERALIZED ENTROPY IN THE LOW-RATE DDOS ATTACK DETECTIONC

Entropy \ Metric		Normal Traffic	Attack Traffic	Spacing
Shannon		4.0892	3.9171	0.1721
Generalized	α=2	3.6204	3.1700	0.4504
	α=3	3.3858	2.9288	0.4570
	α=4	3.2489	2.7900	0.4589
	α=5	3.1618	2.6935	0.4683
	α=10	2.9765	2.4616	0.5149

Table I shows the Shannon and generalized entropies of normal traffic and attacks traffic along with their spacing, the spacing represents the distance of entropy value between normal traffic and attacks traffic. It demonstrates that the generalized entropy method outperforms the Shannon entropy method in low-rate DDoS attack detection as the spacing is more significant. It also shows that the spacing in generalized entropy method increases along with the order α gradually. This increase is almost linear. Therefore, we can adjust the order α value according to different requirements.

For the aim of evaluating the performance of generalized entropy metric globally, we test the proposed metric in the following situations, respectively: to increase DDoS attack intensity gradually and quickly, as well as to reduce DDoS attack intensity gradually then quickly to observe variations of the spacing. In this experiment, in the victim, we keep the normal traffic same, then increase the number of the pure low-rate DDoS attack traffic (as shown in Fig. 6) from 1 to 10 gradually and from 100 to 1000 quickly, as well as reduce the number of attack traffic to one half gradually then from one half reduced to one tenth quickly. The experimental results are shown in Figs. 11, 12, and 13, respectively.

Fig. 11 indicates that the spacing of Shannon and generalized metrics are increasing along with the increasing of number of DDoS attack traffic. There are rapid increases of spacing at the

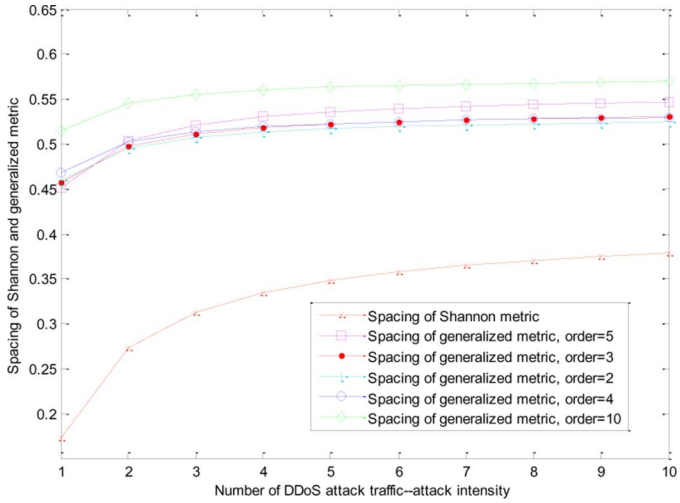


Fig. 11. Variations of spacing of Shannon and generalized metrics in increasing DDoS attack intensity gradually.

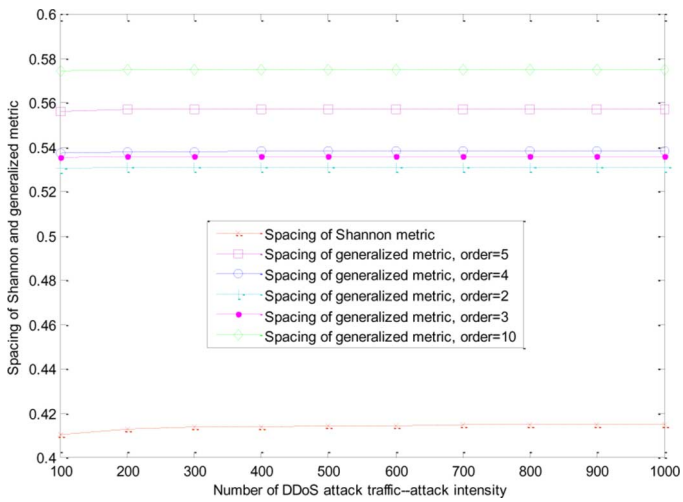


Fig. 12. Variations of spacing of Shannon and generalized metrics in increasing DDoS attack intensity quickly.

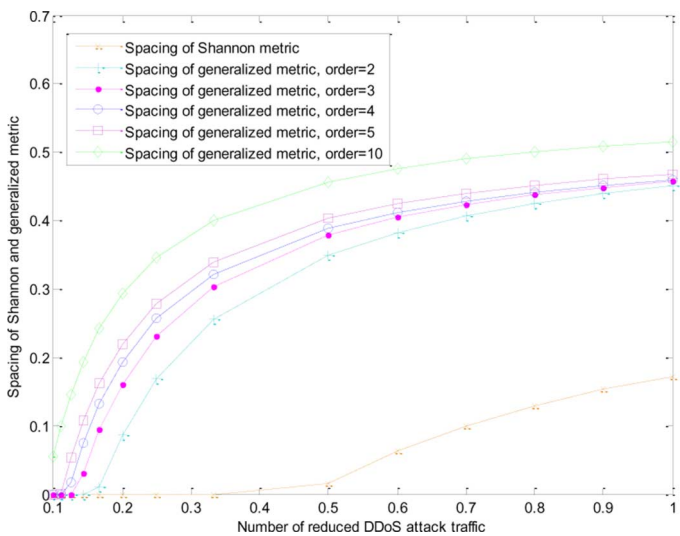


Fig. 13. Variations of spacing of Shannon and generalized metrics in reducing DDoS attack intensity gradually and then quickly.

beginning period whatever the Shannon or generalized metric, because the attacks still are low-rate attacks during this period. However, the spacing of the generalized metric can achieve stable values after three times the number increased of attack traffic in the order of $\alpha = 10$ or after four times the number increased of attack traffic in the order of $\alpha = 2$; the spacing of Shannon metric cannot achieve a stable value and is still increasing along with the increase of number of attack traffic. In order to evaluate the performance of the proposed metric in detecting high intensive (high-rate) DDoS attack, in the test we increase the number of attack traffic dramatically from 100 times up to 1000 times to observe variations of spacing. Fig. 12 shows that the spacing of the Shannon metric can achieve a stable value after 300 times the number increased of attack traffic. Therefore, the proposed generalized metric is a stable and better (larger spacing) metric for detecting low-rate DDoS attack, especially excellent for high-rate attacks in comparison with Shannon metric.

A very low-rate attack traffic will be drown by normal network traffic totally and become extremely difficult to detect using anomaly-based traffic detection approaches. It is important to know how low-rate DDoS attack traffic can be detect by the proposed metric. In this experiment, we first reduce the number of low-rate attack traffic gradually and then quickly reduce the number. Fig. 13 shows the experimental result that the spacing of the Shannon and generalized metrics are reducing along with reducing the number of low-rate attack traffic; there is stable reduction when the number of attack traffic reduces gradually for the generalized metric, but reduces quickly for the Shannon metric. When the number of attack traffic reduces sharply the spacing of Shannon and generalized metrics will reduce dramatically too, but for the Shannon metric, the spacing will reduce and up to zero (here, the number of attack traffic is just reduced to one third) extremely quickly, because at this situation the attack traffic becomes a very low-rate attack. Therefore, the proposed metric can detect a very low-rate DDoS attack well in comparison with the Shannon metric. For example, in this experiment, the proposed metric can still detect a very low-rate attack which is reduced to one tenth the number of the original low-rate attack traffic while the order $\alpha = 10$.

Now we discuss how early the proposed metric can detect a low-rate DDoS attack in comparison with the Shannon metric. To simplify the discussion, we assume a scenario of an attack based on a binary tree network topology, and the number of attack traffic in a local router is formed by the numbers of traffic from its two upstream routers (they are called brother), and let the numbers of traffic from every brother router be the same. Therefore, we have: the number of attack traffic in Hop0 (1) = two times number of attack traffic in Hop1 (1/2) = four times number of attack traffic in Hop2 (1/4) = eight times number of attack traffic in Hop3 (1/8) = $\cdots$, then based on this rule we test the proposed metric and the experimental result is shown in Table II. From this table we can see that the proposed metric can detect a low-rate DDoS attack two hops earlier than the Shannon metric while the order $\alpha = 2$, and three hops earlier approximately while $\alpha = 10$. According to [24], generally on the Internet, the normal route hops between two network ends is 15; therefore, the proposed metric should be a better metric in detecting attacks for several hops earlier, such as three hops earlier while the order $\alpha = 10$.

TABLE II
COMPARISON (HOP EARLY) OF GENERALIZED ENTROPY METRIC WITH SHANNON ENTROPY METRIC IN THE LOW-RATE DDOS ATTACK DETECTION

Entropy \ Metric		Spacing			
		Hop0(1)	Hop1(1/2)	Hop2(1/4)	Hop3(1/8)
Shannon		0.1721	0.0154	0.0000	0.0000
Generalized	α=2	0.4504	0.3499	0.1683	0.0000
	α=3	0.4570	0.3785	0.2306	0.0000
	α=4	0.4589	0.3890	0.2566	0.0175
	α=5	0.4683	0.4028	0.2786	0.0534
	α=10	0.5149	0.4567	0.3463	0.1459

TABLE III
REDUCED FALSE POSITIVE RATE OF THE PROPOSED METRIC IN COMPARISON WITH THE SHANNON ENTROPY METRIC

Generalized Entropy Metric	Reduced False Positive Rate
α=2	161.71%
α=3	165.54%
α=4	166.65%
α=5	172.11%
α=10	199.19%

We further compute and compare the false positive rate of the proposed approach with the Shannon metric. There will be different false positive rate values in different situations due to the real network being extremely dynamic and complex. Therefore, it is very difficult to obtain a definite value of the false positive rate by using a certain metric. For the purpose of the discussion in this paper, we assume that the false positive rate is known by using the Shannon metric in the real network situation. The false positive rate β is defined as the proportion of negative events (not attacks) that were mistakenly reported as being positive events (attacks) in the total of tested events. We choose the mid-value of the spacing as the threshold, and obtain a reduced false positive rate when using the proposed metric compared to the Shannon metric. The reduced false positive rate β' is defined as

$$\beta' = \frac{(\beta_{\text{Shannon}} - \beta_{\text{Generalized}})}{\beta_{\text{Shannon}}}. \tag{16}$$

This measurement represents how better the generalized metric outperforms the traditional Shannon metric. Table III shows that the proposed metric clearly reduces the false positive rate, from 161.71% to 199.19%, which is more than 1.6 times of the baseline false positive rate.

In summary, compared with the Shannon metric, the proposed generalized entropy metric is a stable and low false positive rate metric in low-rate DDoS attacks detection, it can not only effectively detect low-rate attacks but also detect attacks several hops early.

B. Information Distance Metric

In this experiment, we use the real normal network traffic and low-rate attack datasets shown above as the incoming traffic to test the effectiveness of the proposed metric in detecting a low-rate DDoS attack, and further to study the following properties.

1) When detecting low-rate DDoS attacks, our approach is much better than the Kullback–Leibler divergence approach because we are able to enlarge the distance rate and reduce the false positive rate.
2) Our metric definition is necessary because if the divergence is used inappropriately, the outcome will be unsatisfactory as the distance (gap) will be very small, and the false positive rate will be increased.
3) Our approach is able to achieve early detection of low-rate DDoS attacks.
4) By adjusting the value of α, we can adjust the resulting distance in our approach.

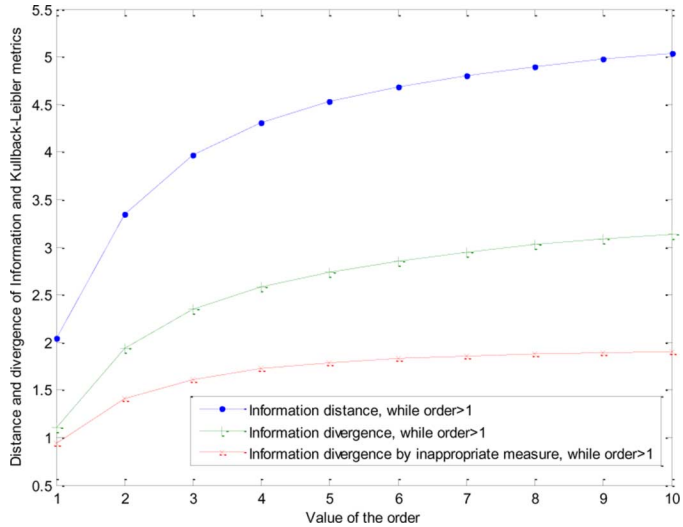


Fig. 14. Variations of information distance and divergence as well as divergence by inappropriate measure along with the value of the order $\alpha(\alpha > 1)$, the Kullback–Leibler distance and divergence as well as the divergence by inappropriate measure while $\alpha = 1$.

In order to test variations of distance and divergence of the Kullback–Leibler metric and information metric along with the order α, the normal network traffic and the low-rate attack traffic must have the same number of source IP addresses in a sampling period. Therefore, we sample the above low-rate DDoS attack traffic again to form a new low-rate attack which will have the same number of source IP addresses with the normal traffic, and have the same probability distribution of source IP addresses with the original attack traffic. The experimental result is shown in Fig. 14, which indicates that the information distance and divergence as well as the divergence by inappropriate measure all are increasing along with the increase of order α, but the information distance increases quickly, the divergence by inappropriate measure increases a little and keeps a stable value after the order $\alpha = 3$. The information distance has a bigger gap than the Kullback–Leibler distance and divergence. Therefore, the proposed metric outperforms the Kullback–Leibler metric in a low-rate DDoS attack detection. It is important that we can adjust the resulting (detecting) distance as a requirement by adjusting the value of α to achieve better detection.

We first discuss the detection effectiveness of the proposed information distance metric under a very low-rate DDoS attack condition, then discuss how many hops early with the proposed metric in comparison with the Kullback–Leibler metric

TABLE IV
COMPARISON (HOP EARLY) OF INFORMATION DISTANCE METRIC WITH KULLBACK–LEIBLER DISTANCE METRIC IN THE LOW-RATE DDOS ATTACK DETECTION

Metric Distance		Value of Distance			
		Hop0(1)	Hop1(1/2)	Hop2(1/4)	Hop3(1/8)
Kullback-Leibler		2.0479	1.7441	1.4982	1.1502
Information	α=2	3.3438	2.8811	2.5268	2.0232
	α=3	3.9646	3.4628	3.0863	2.5589
	α=4	4.3070	3.8056	3.4216	2.8959
	α=5	4.5279	4.0399	3.6493	3.1233
	α=10	5.0417	4.6011	4.1931	3.6365

TABLE V
COMPARISON (HOP EARLY) OF INFORMATION DISTANCE METRIC WITH KULLBACK–LEIBLER DISTANCE METRIC IN THE LOW-RATE DDOS ATTACK DETECTION (CONTINUED)

Metric Distance		Value of Distance		
		Hop4(1/16)	Hop5(1/32)	Hop6(1/64)
Kullback-Leibler		0.7963	0.5524	0.3832
Information	α=2	1.4856	1.0614	0.7857
	α=3	1.9844	1.4631	1.1516
	α=4	2.3239	1.7602	1.4440
	α=5	2.5547	1.9736	1.6597
	α=10	3.0690	2.4689	2.1832

in a low-rate attack. We assume the low-rate attack scenario and the network topology are the same as above (used in generalized metric test). The experimental results are shown in Tables IV and V. From Tables IV and V, we know that the value of distance is decreasing gradually along with an increase of hop count, namely the lower the rate of attack traffic, the smaller the distance. The proposed metric can detect a very low-rate attack better than using the Kullback–Leibler metric; for example, the distance still has a big gap (2.1832, while the order $\alpha = 10$; the larger distance will give a better accuracy in an attack detection) when the attack traffic is reduced to 1/64 of itself, but for the Kullback–Leibler distance it becomes a little gap (0.3832). Furthermore, the experimental results also show that the proposed metric can detect a low-rate DDoS attack early in comparison with the Kullback–Leibler distance metric, such as there should be three hops early while the order $\alpha = 2$, four hops early while the order $\alpha = 4$, and while the order $\alpha = 10$ it can have six hops early. Therefore, the information distance metric is a good metric for detecting low-rate DDoS attacks; it can not only detect very low-rate attacks but also have successful detection several hops earlier than the Kullback–Leibler distance metric.

In order to evaluate the performance of the proposed information distance metric in detecting high intensive (high-rate) DDoS attack, we increase the number of attack traffic dramatically from 100 times up to 1000 times to observe the variations of distance. Fig. 15 indicates that the distances of the proposed metric are increasing gradually along with the increase of

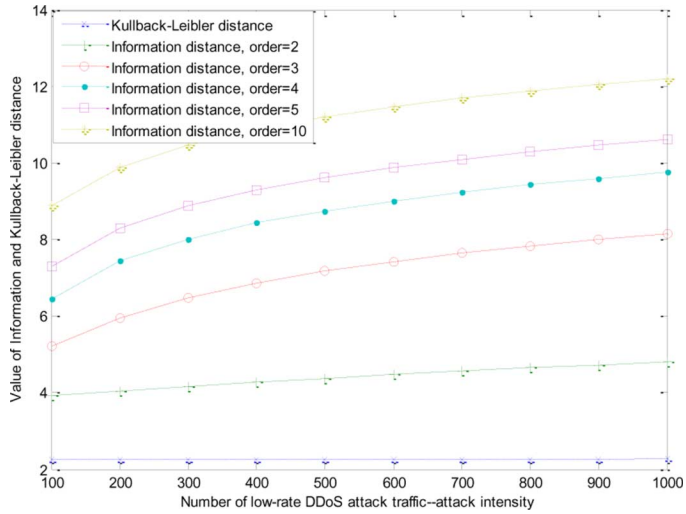


Fig. 15. Variations of distance of the information and Kullback–Leibler metrics in increasing DDoS attack intensity quickly.

TABLE VI
REDUCED FALSE POSITIVE RATE OF THE PROPOSED INFORMATION DISTANCE METRIC IN COMPARISON WITH THE KULLBACK–LEIBLER DISTANCE METRIC

Information Distance Metric	Reduced False Positive Rate
α=2	63.28%
α=3	93.60%
α=4	110.32%
α=5	121.10%
α=6	128.78%
α=7	134.61%
α=8	139.24%
α=9	143.03%
α=10	146.19%

the number of attacks traffic. There are rapid increases of distance at the beginning period, because the attack after aggregation is still a low-rate attack during this period. Then there should be the stable increase by the rapid increase of attack intensity. Therefore, the proposed metric is a stable and better (larger gap) metric for detecting low-rate DDoS attack, and is perfect for high-rate attacks detection in comparison with the Kullback–Leibler metric.

We compute the false positive rate of the proposed information distance metric under the same conditions as the above (used for generalized entropy metric). Similarly, the reduced false positive rate β'' is defined as

$$\beta'' = \frac{(\beta_{\text{Kullback-Leibler}} - \beta_{\text{Information distance}})}{\beta_{\text{Kullback-Leibler}}}. \qquad (17)$$

The result is shown in Table VI, which indicates the reduced false positive rate by the proposed metric in different values of order α in comparison to the Kullback–Leibler distance metric. It can clearly reduce the false positive rate up to 146.19% of the Kullback–Leibler metric while the order $\alpha = 10$ for the proposed information distance metric.

For all situations, it has been shown that the proposed information distance metric is a better metric because first, it can be used in real measurements in comparison with the information

divergence approach which is not a real metric (it is asymmetric and the distance gap between the normal network traffic and the attack traffic will be smaller if we used the inappropriate divergence measurement, which may result in decreasing the detection sensitivity during a low-rate DDoS attack); second, it is a stable metric which holds a low false positive rate (it can not only effectively detect low-rate attacks but also detect the attacks several hops earlier in comparison with the Kullback–Leibler distance metric). Confidence intervals will be beneficial for estimating the possible attacks based on the outputs of the system parameters, especially for predicting future attacks. Confidence intervals are a good indication for the reliability the prediction system. Due to the paper length limit, obtaining the confidence intervals of the system will be our future work.

IV. Related Work

The metrics of anomaly-based detection have been the focus of intense study for years in an attempt to detect intrusions and attacks on the Internet. Recently, information theory as one of the statistical metrics is being increasingly used for anomaly detection. Feinstein *et al.* [25] present methods to identify DDoS attacks by computing entropy and frequency-sorted distributions of selected packet attributes. The DDoS attacks show anomalies in the characteristics of the selected packet attributes, and the detection accuracy and performance are analyzed using live traffic traces from a variety of network environments. However, because the proposed detector and responder lack coordination with each other, the possible impact of responses on legitimate traffic and expenses for computational analysis are increased. Yu and Zhou [26] applied an information theory parameter (entropy rate) to discriminate the DDoS attack from the surge legitimate accessing. This is based on shared regularities with different DDoS attack traffic which are different from real surging accessing in a short period of time. However, attackers can adopt a multiple attack package generation function in one attack to easily fool the proposed detection algorithm. Lee and Xiang [27] used several information-theoretic measures, such as entropy, conditional entropy, relative conditional entropy, information gain, and information cost for anomaly detection. To some extent these measures can be used to evaluate the quality of anomaly detection methods and build the appropriate anomaly detection models even though it is very difficult to build an adaptive model that can dynamically adjust to different sequence lengths (or time windows) based on run-time information.

A low-rate DDoS attack is substantially different from the traditional (high-rate) DDoS attack. A few researchers have proposed several detection schemes against this type of attack. Sun *et al.* [28] proposed a distributed detection mechanism that used a dynamic time warping method to identify the existence of the low-rate attacks, and then a fair resource allocation mechanism will be used to minimize the number of affected flows. However, this method can lose the legitimate traffic to some extent. Shevtekar *et al.* [3] presented a light-weight data structure to store the necessary flow history at edge routers to detect the low-rate TCP DoS attacks. Although this method can detect any periodic pattern in the flows, it may not be scalable and can be deceived by the IP address spoofing. Chen *et al.* [18] present a collaborative detection of DDoS attacks. While focusing on detection rate, it is difficult for this scheme to differentiate the normal flash crowds and real attacks. As it heavily relies on the normal operation of participating routers, the false positives will increase if the routers are compromised. Zhang *et al.* [29] propose to use self-similarity to detect low-rate DDoS attacks. While the approach is claimed to be effective, the paper does not use real scenario data to evaluate it.

Kullback–Leibler divergence, as a well-known information divergence, has been used by researchers to detect abnormal traffic such as DDoS attacks [10], [11], [30]. The difference between previous work and our research is that we are the first to propose using information divergence for DDoS attack detection. Information divergence, as the generalized divergence, can deduce many concrete divergence forms according to different values of order α. For example, when $\alpha \to 1$, it can decipher the Kullback–Leibler divergence. It is very important and significant that we can obtain the optimal value of divergence between the attack traffic and the legitimate traffic in a DDoS detection system by adjusting the value of order α of information divergence. In addition to this, we also study the properties of Kullback–Leibler divergence and information divergence in theory and overcome their asymmetric property when used in real measurement. We successfully convert the information divergence into an effective metric in DDoS attack (including both low-rate and high-rate) detection.

V. Conclusion

In this paper, we propose two new and effective information metrics for low-rate DDoS attacks detection: generalized entropy and information distance metric. The experimental results show that these metrics work effectively and stably. They outperform the traditional Shannon entropy and Kullback–Leibler distance approaches, respectively, in detecting anomaly traffic. In particular, these metrics can improve (or match the various requirements of) the systems' detection sensitivity by effectively adjusting the value of order α of the generalized entropy and information distance metrics. As the proposed metrics can increase the information distance (gap) between attack traffic and legitimate traffic, they can effectively detect low-rate DDoS attacks early and reduce the false positive rate clearly. The proposed information distance metric overcomes the properties of asymmetric of both Kullback–Leibler and information divergences. Furthermore, the proposed IP traceback scheme based on information metrics can effectively trace all attacks until their own LANs (zombies). In conclusion, our proposed information metrics can substantially improve the performance of low-rate DDoS attacks detection and IP traceback over the traditional approaches.

References

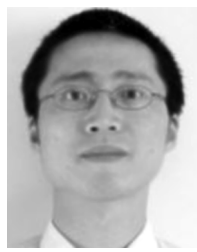
[1] A. Chonka *et al.*, "Cloud security defence to protect cloud computing against HTTP-DoS and XML-DoS attacks," *J. Netw. Comput. Applicat.* Jun. 23, 2010 [Online]. Available: http://dx.doi.org/10.1016/j.jnca.2010.06.004

[2] X. Jin *et al.*, "ZSBT: A novel algorithm for tracing DoS attackers in MANETs," *EURASIP J. Wireless Commun. Netw.*, vol. 2006, no. 2, pp. 1–9, 2006.

[3] A. Shevtekar, K. Anantharam, and N. Ansari, "Low rate TCP Denial-of-Service attack detection at edge routers," *IEEE Commun. Lett.*, vol. 9, no. 4, pp. 363–365, Apr. 2005.

[4] G. Carl *et al.*, "Denial-of-service attack-detection techniques," *IEEE Internet Comput.*, vol. 10, no. 1, pp. 82–89, Jan./Feb. 2006.

[5] P. Du and S. Abe, "IP packet size entropy-based scheme for detection of DoS/DDoS attacks," *IEICE Trans. Inf. Syst.*, vol. E91-D, no. 5, pp. 1274–1281, 2008.
[6] S. Ledesma and D. Liu, "Synthesis of fractional Gaussian noise using linear approximation for generating self-similar network traffic," *Comput. Commun. Rev.*, vol. 30, no. 2, pp. 4–17, 2000.
[7] E. Perrin *et al.*, "Nth-order fractional Brownian motion and fractional Gaussian noises," *IEEE Trans. Signal Process.*, vol. 49, no. 5, pp. 1049–1059, May 2001.
[8] E. Perrin *et al.*, "Fast and exact synthesis for 1-D fractional Brownian motion and fractional Gaussian noises," *IEEE Signal Process. Lett.*, vol. 9, no. 11, pp. 382–384, Nov. 2002.
[9] Y. Bao and H. Krim, "Renyi entropy based divergence measures for ICA," in *Proc. IEEE Workshop on Statistical Signal Processing*, 2003, pp. 565–568.
[10] Y. Gu, A. McCallum, and D. Towsley, "Detecting anomalies in network traffic using maximum entropy estimation," in *Proc. ACM SIGCOMM Conf. Internet Measurement (IMC 2005)*, 2005, pp. 32–32.
[11] R. Sekar *et al.*, "Specification based anomaly detection: A new approach for detecting network intrusions," in *Proc. ACM Conf. Computer and Communications Security (CCS 2002)*, 2002, pp. 265–274.
[12] A. Patcha and J.-M. Park, "An overview of anomaly detection techniques: Existing solutions and latest technological trends," *Comput. Netw.*, vol. 51, no. 12, pp. 3448–3470, 2007.
[13] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423 and 623–656, 1948.
[14] K. Zyczkowski, "Rényi extrapolation of Shannon entropy," *Open Syst. Inf. Dynamics*, vol. 10, no. 3, pp. 297–310, 2003.
[15] K. J. Kumar, R. C. Joshi, and K. Singh, "A distributed approach using entropy to detect DDoS attacks in ISP domain," in *Proc. Int. Conf. Signal Processing, Communications and Networking (ICSCN 2007)*, 2007, pp. 331–337.
[16] A. R. Barron, L. Gyorfi, and E. C. van der Meulen, "Distribution estimation consistent in total variation and in two types of information divergence," *IEEE Trans. Inf. Theory*, vol. 38, no. 5, pp. 1437–1454, Sep. 1992.
[17] M. Broniatowski, "Estimation of the Kullback–Leibler divergence," in *Mathematical Methods of Statistics*. Princeton, NJ: Princeton Univ. Press, 2003.
[18] Y. Chen, K. Hwang, and W.-S. Ku, "Collaborative detection of DDoS attacks over multiple network domains," *IEEE Trans. Parallel Distrib. Syst.*, vol. 18, no. 12, pp. 1649–1662, Dec. 2007.
[19] J.-F. Bercher, "On some entropy functionals derived from Rényi information divergence," *Inf. Sci.*, vol. 178, no. 12, pp. 2489–2506, 2008.
[20] Y. Xiang, W. Zhou, and M. Guo, "Flexible deterministic packet marking: An IP traceback system to find the real source of attacks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 20, no. 4, pp. 567–580, Apr. 2009.
[21] MIT Lincoln Laboratory Data Sets [Online]. Available: http://www.ll.mit.edu/mission/communications/ist/corpora/ideval/data/2000/LLS_DDOS_0.2.2.html
[22] CAIDA, 2010 [Online]. Available: http://data.caida.org/datasets/security/ddos-20070804/
[23] D. Moore *et al.*, "Inferring Internet denial-of-service activity," *ACM Trans. Comput. Syst.*, vol. 24, no. 2, pp. 115–139, 2006.
[24] T. K. T. Law, J. C. S. Lui, and D. K. Y. Yau, "You can run, but you can't hide: An effective statistical methodology to trace back DDoS attackers," *IEEE Trans. Parallel Distrib. Syst.*, vol. 16, no. 9, pp. 799–813, Sep. 2005.
[25] L. Feinstein *et al.*, "Statistical approaches to DDoS attack detection and response," in *Proc. DARPA Information Survivability Conf. Exposition*, 2003, pp. 303–314.
[26] S. Yu and W. Zhou, "Entropy-Based collaborative detection of DDoS attacks on community networks," in *Proc. 6th IEEE Int. Conf. Pervasive Computing and Communications (PerCom 2008)*, 2008, pp. 566–571.
[27] W. Lee and D. Xiang, "Information-Theoretic measures for anomaly detection," in *Proc. IEEE Symp. Security and Privacy*, 2001, pp. 130–143.
[28] H. Sun, J. C. S. Lui, and D. K. Y. Yau, "Defending against low-rate TCP attacks: Dynamic detection and protection," in *Proc. IEEE Int. Conf. Network Protocols (ICNP 2004)*, 2004, pp. 196–205.
[29] S. Zhang *et al.*, "Detection of low-rate DDoS attack based on self-similarity," in *Proc. Int. Workshop on Education Technology and Computer Science*, 2010, pp. 333–336.
[30] S. Yu, W. Zhou, and R. Doss, "Information theory based detection against network behavior mimicking DDoS attacks," *IEEE Commun. Lett.*, vol. 12, no. 4, pp. 319–321, Apr. 2008.



Yang Xiang (A'08–M'09) received the Ph.D. degree in computer science from Deakin University, Victoria, Australia, in 2007.

He is currently with the School of Information Technology, Deakin University, Australia. His research interests include network and system security, and wireless systems. He has published more than 100 research papers in international journals and conferences. He has served as Program/General Chair for many international conferences such as ICA3PP 11, TrustCom 11, IEEE HPCC 10/09, IEEE ICPADS 08, and NSS 10/09/08/07. He has been a PC member for many international conferences such as IEEE ICC, IEEE GLOBECOM, SECRYPT, Malware, and IEEE ICPADS. He has served as reviewer for many international journals such as IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, IEEE COMMUNICATIONS LETTERS, and IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS. He is on the editorial board of the *Journal of Network and Computer Applications*.

Ke Li received the Ph.D. degree from Deakin University, Victoria, Australia, in 2009.

He is currently a postdoctoral research fellow in the School of Information Technology, Deakin University, Australia. Previous to this, he was a network engineer at Telstra in Australia. His research interests include network security, distributed and parallel systems, and digital signal processing.

Wanlei Zhou (M'92–SM'09) received the Ph.D. degree in 1991 from the Australian National University, Canberra, Australia, and the D.Sc. degree from Deakin University, Victoria, Australia, in 2002.

He is currently the chair professor of Information Technology and the Head of School of Information Technology, Deakin University, Melbourne. His research interests include distributed and parallel systems, network security, mobile computing, bioinformatics, and e-learning. He has published more than 200 papers in refereed international journals and refereed international conference proceedings. Since 1997, he has been involved in more than 50 international conferences as the general chair, a steering chair, a PC chair, a session chair, a publication chair, and a PC member.