Research Article

Characterizing Network Anomaly Traffic with Euclidean Distance-Based Multiscale Fuzzy Entropy

Renjie Zhou ^{1,2,3}, Xiao Wang ², Jingjing Yang ⁴, Wei Zhang ^{2,3} and Sanyuan Zhang ¹

¹College of Computer Science and Technology, Zhejiang University, Hangzhou 310058, China

²College of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou 310018, China

³Key Laboratory of Complex Systems Modeling and Simulation of the Ministry of Education, Hangzhou Dianzi University, Hangzhou 310018, China

⁴Zhuoyue Honors College, Hangzhou Dianzi University, Hangzhou 310018, China

Correspondence should be addressed to Renjie Zhou; renjie_zhou@163.com, Wei Zhang; magherozhw@hdu.edu.cn, and Sanyuan Zhang; syzhang@cs.zju.edu.cn

Received 19 February 2021; Accepted 21 May 2021; Published 16 June 2021

Academic Editor: Zhe-Li Liu

Copyright © 2021 Renjie Zhou et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The prosperity of mobile networks and social networks brings revolutionary conveniences to our daily lives. However, due to the complexity and fragility of the network environment, network attacks are becoming more and more serious. Characterization of network traffic is commonly used to model and detect network anomalies and finally to raise the cybersecurity awareness capability of network administrators. As a tool to characterize system running status, entropy-based time-series complexity measurement methods such as Multiscale Entropy (MSE), Composite Multiscale Entropy (CMSE), and Fuzzy Approximate Entropy (FuzzyEn) have been widely used in anomaly detection. However, the existing methods calculate the distance between vectors solely using the two most different elements of the two vectors. Furthermore, the similarity of vectors is calculated using the Heaviside function, which has a problem of bouncing between 0 and 1. The Euclidean Distance-Based Multiscale Fuzzy Entropy (EDM-Fuzzy) algorithm was proposed to avoid the two disadvantages and to measure entropy values of system signals more precisely, accurately, and stably. In this paper, the EDM-Fuzzy is applied to analyze the characteristics of abnormal network traffic such as botnet network traffic and Distributed Denial of Service (DDoS) attack traffic. The experimental analysis shows that the EDM-Fuzzy entropy technology is able to characterize the differences between normal traffic and abnormal traffic. The EDM-Fuzzy entropy characteristics of ARP traffic discovered in this paper can be used to detect various types of network traffic anomalies including botnet and DDoS attacks.

1. Introduction

The prosperity of network technologies, such as mobile networks and social networks, brings revolutionary changes to our daily lives. However, due to the complexity and fragility of the network infrastructures, network anomalies and attacks frequently cause serious problems and significant loss to people. Researchers are studying various cybersecurity awareness technologies to help people understand the security status and trend of networks. Characterization of network anomaly traffic is one of the key technologies commonly used to model and detect network anomalies and then to raise the cybersecurity awareness

capability of network administrators. The existing approaches of network anomaly detection can be mainly classified into six categories [1]: classification-based methods [2–4], clustering-based methods [5–9], statistical methods [10, 11], stochastic methods [12, 13], deep-learning-based methods [14–17], and others [18–21].

Network anomaly detection via traffic feature distributions is becoming more and more popular these days. As the measure of uncertainty, entropy can be used to summarize feature distributions in a compact form [22]. There are many forms of entropy, but only a few have been applied to network anomaly detection [23–27]. On this basis, we apply a Euclidean Distance-Based Multiscale Fuzzy Entropy

(EDM-Fuzzy) algorithm which we proposed to detect abnormal network traffic as a useful supplement of other approaches.

Investigation irregularity of signals generated by complex systems is valuable to predict the future states as well as detect abnormal behaviors [28]. In order to quantitatively analyze signal irregularity and diagnose system anomalies, researchers have proposed various signal complexity and uncertainty indicators, such as algorithmic complexity [29], Shannon Entropy [30], Approximate Entropy [31], Sample Entropy [32], Fuzzy Entropy [33], Multiscale Entropy (MSE) [34], and Composite Multiscale Entropy (CMSE) [35]. Entropy-based technologies have been widely applied in diagnosing the anomalies of various systems. For example, Shannon Entropy was applied in detecting faults of mechanical systems [36], MSE was applied in fault diagnosis of power systems [37], and so on.

However, the existing methods calculate the distance between vectors solely using the two most different elements of the two vectors. Furthermore, the similarity of vectors is calculated with Heaviside function, which has a problem of bouncing between 0 and 1. To this end, we proposed a novel entropy technology named EDM-Fuzzy in the paper [38]. The EDM-Fuzzy technology uses the sum of the Euclidean distances of the elements corresponding to two vectors instead of the largest element difference between the two vectors and uses the hyperbolic function to calculate the similarity between the two vectors. Thus, the EDM-Fuzzy technology avoids the two disadvantages inherent in the other entropy technologies and measures entropy values of system signals more precisely, accurately, and stably. In this paper, we apply the EDM-Fuzzy algorithm to characterize network anomaly traffic. We first briefly introduce the EDM-Fuzzy algorithm and then introduce the botnet CTU-13 dataset and the Distributed Denial of Service (DDoS) attack CICDDoS2019 dataset used in this paper. Then, the basic characteristics of these two datasets are introduced. Then, the EDM-Fuzzy entropy value analysis is performed on two datasets. Finally, we analyze the characteristics of the normal traffic and investigate the characteristics of the malicious traffic by comparing the differences between the normal and malicious traffic.

The rest of this paper is organized as follows. The related works are introduced in Section 2, and the EDM-Fuzzy entropy technology and network traffic traces are introduced in Sections 3 and 4, respectively. Section 5 is the analysis of network anomaly traffic with EDM-Fuzzy entropy. Section 6 concludes the paper and introduces the future work.

2. Related Works

2.1. Network Anomaly Traffic Detection Approaches. Network anomaly traffic detection approaches have been extensively explored. The existing approaches can be mainly classified into six categories [1]: classification-based methods [2–4], clustering-based methods [5–9], statistical methods [10, 11], stochastic methods [12, 13], deep-learning-based methods [14–17], and others [18–21]. A classification-based approach is a supervised learning algorithm. Classification

algorithms such as logistic regression, k -nearest neighbor algorithm, decision tree, and support vector machine are commonly used. More recently, several hybrid classification models were proposed [3–5]. However, in most cases, labeling data manually is highly time-consuming and inefficient. Clustering techniques are used to identify clusters and outliers in multiple low-dimensional spaces. The evidence of traffic structure provided by these multiple clusters is then combined to produce an abnormality ranking of network traffic [39]. Several distance-based metrics are commonly used in anomaly detection, such as the Euclidean distance, Manhattan distance, and dynamic time warping (DTW) distance. However, the number of clusters is difficult to decide and different numbers of clusters would produce extremely different results. In a statistical method, an abnormality is often determined by checking whether the traffic complies with the assumed distribution model and whether the value is larger than a preset threshold. The most frequent assumptions are Gaussian distributions, Poisson distributions, multivariate Gaussian distributions, and so on. The model systematically analyzes abnormal behaviors of the network, but detection of such abnormalities is difficult since there will be cases that do not obey the presumed distributions. Stochastic processes like Hidden Markov Model and Conditional Random Field were also frequently applied in detection of traffic anomaly [12, 13]. Due to the success of deep-learning technologies in image processing and natural language processing, they have been intensively studied in network intrusion detection [14, 15], network traffic tracking [16], and network traffic abnormal behavior detection [17]. Besides, time-series density analysis [18], wavelet [19], principal components analysis [20], and ensemble learning technologies [21] have been extensively investigated in network anomaly detection.

2.2. Entropy-Based Technologies. Entropy-based technologies are highly valued in detecting the degree of disorder or irregularity of a complex system. Thus, there have been a number of entropy-based technologies being proposed and being widely applied in detecting anomalies of complex systems. Khan et al. [37] presented an entropy-based approach for detecting faults in power systems. An entropy-based methodology was proposed in paper [40] to extract characteristics from signals of smart meters to effectively classify power quality problems. The Kullback-Shannon Entropy was applied as a standalone feature to predict failure in lubricated surfaces [41].

Pincus [31], Richman, and Moorman [32] and Costa et al. [34, 42] proposed Approximate Entropy, Sample Entropy, and MSE to measure signal complexity, respectively. Although MSE has been widely applied, the variance of the entropy values increases significantly as the time series is coarse-grained for larger time scales [43]. In order to solve the problem, Wu et al. proposed CMSE [35] and introduced a composite averaging method to reduce the variance. Niu and Wang [44] applied CMSE to study the characteristics of stock market indices and found that CMSE is more stable and reliable than MSE. Chen et al. [33] proposed Fuzzy

Approximate Entropy (FuzzyEn) and applied it in the study of surface muscle signal. Wang et al. [45] proposed fractional fuzzy entropy to study physics financial dynamics. Li et al. [46] integrated fractional fuzzy entropy with a binary tree support vector machine to perform early diagnosis of rolling bearing faults. Composite multiscale fuzzy entropy is proposed in paper [47] and is applied to extract the hidden features of vibration signals.

Entropy-based network anomaly detection via traffic feature characterization is becoming more and more popular these days. Ranjan et al. [23] proposed a worm detection algorithm that measures Shannon Entropy values for traffic and alarms on sudden bursts. Gu et al. [24] applied Shannon maximum entropy estimation to draw the network baseline distribution and to build a multiperspective view of network traffic. Paper [25] presented a novel network intrusion detection system using Shannon Entropy and traffic distributions of the source port. Paper [26] proposed a hybrid DDoS detection method, which integrates Kernel Online Anomaly Detection (KOAD), Shannon Entropy, and Mahalanobis Distance. In this study, Shannon Entropy is utilized with an online machine learning method to detect malicious traffic including DDoS attacks and Flash Event traffic. Paper [27] presented anomaly detection in activities of daily living based on entropy measures.

However, there are still two disadvantages in the existing state-of-the-art entropy algorithms, such as MSE, CMSE, RCMSE, MMSE, and FuzzyEn. That is, the existing methods calculate the distance between vectors solely based on the two most different elements of the two vectors. Furthermore, the similarity of vectors is calculated using Heaviside function, which has a problem of bouncing between 0 and 1. In order to address the shortcomings of existing state-of-the-art entropy algorithms, we proposed novel entropy technology [38], named EDM-Fuzzy.

3. EDM-Fuzzy Technology

EDM-Fuzzy measures the distance of the two vectors with Euclidean distance taking all the corresponding elements in the two vectors into the computation. Furthermore, in order to solve the problem of instability, we choose the hyperbolic function as the fuzzy function instead of the Heaviside function to define the similarity between vectors with full-range continuous values from zero to one based on the Euclidean distance of the two vectors. The computation process of EDM-Fuzzy is formally described in Algorithm 1.

The goal of the algorithm is to measure the complexity and irregularity of time series more accurately and stably. The input of the algorithm is a time series $X = \{x_1, x_2, \dots, x_N\}$, time scale τ , vector dimension m , tolerance coefficient r , and standard deviation SD of time series X . The output of the algorithm is the EDM-Fuzzy entropy value of time series X at time scale τ . The general process of the algorithm is first to coarse-grain the time series with time scale τ , then split the time series into m -dimensional vectors, move the vectors to its centroid, and finally, calculate the Euclidean distance of the two vectors and compute the Euclidean distance based on fuzzy sample entropy value of time series. For parameters m

and r , m is usually set to 2 and r generally ranges from 0.1 to 0.2. In our experiments, r is set to 0.15; that is, the similarity tolerance is set to $0.15 \cdot SD$. Here, SD represents the standard deviation of the original time series.

4. Network Traffic Trace

A suitable network traffic trace is essential to the research of the characterization of network anomaly traffic. The traces used in this paper are publicly accessible, within which anomaly activities including botnet and DDoS attack were recorded. Through analysis of these public traces with EDM-Fuzzy algorithm, we can further discover the characteristics of such anomaly activities.

4.1. Botnet Traffic Trace. The botnet traffic trace used in this section is the CTU-13 trace that was collected and provided by the Stratosphere Laboratory of CTU University in the Czech Republic [48, 49]. This trace contains botnet traffic as well as normal background traffic. The CTU-13 trace contains 13 botnet samples in different scenarios. In each sample, a specific malware is executed and different operations were performed accordingly. The brief information of the trace is shown in Tables 1 and 2.

Table 1 shows the characteristics of 13 types of botnet scenarios. Each type of botnet has different characteristics of malicious behavior. In Table 1, IRC represents the network relay chat protocol, SPAM represents spam, CF represents malicious clicks, PS represents port scan, FF represents fast flux, P2P refers to end-to-end, DDoS refers to Distributed Denial of Service, and US refers to a protocol that is controlled and completed by humans. The basic characteristics of each botnet can be seen in Table 1.

Table 2 shows the duration, the number of data packets, the type of malware, and the number of infected computers of these 13 types of botnet scenarios. The duration of botnet scenarios varies from 15 minutes to 66 hours. The number of infected hosts for most scenarios is 1 host. Neris-3, Rbot-3, Rbot-4, and NSIS.ay scenarios have 10 and 3 infected hosts, respectively.

4.2. DDoS Traffic Trace. DDoS attack is an abnormal network behavior designed to exhaust server resources. It will cause server congestion and thus will be unable to provide services to users. The traffic trace used in this paper is the CICDDoS2019 which was published by the Canadian Cyber Security Institute (CIC) [50]. The CICDDoS2019 trace contains common and latest DDoS attacks. There are mainly two categories of DDoS attack methods involved in this trace, DDoS reflection attack and DDoS direct attack. DDoS reflection attack method utilizes routers, servers, and other facilities to respond to requests, thus reflecting the attack traffic to hide the source of the attack. The direct DDoS attack method is to directly attack the target using the controlled hosts. Compared with the reflection type attack, the direct attack method has a lower degree of anonymity. The specific attack types and attack duration time in the CICDDoS2019 dataset are shown in Tables 3 and 4.

Inputs:

Time series: $X = \{x_1, x_2, \dots, x_N\}$.
 Time scale: $\tau \geq 1$.
 Vector dimension: $m \geq 2$.
 Tolerance coefficient: $0.1 \leq r \leq 0.2$.
 Standard deviation of time series X : SD .

Output:

EDM-Fuzzy entropy value of time series X at time scale τ .

- (1) **for** $k = 1$ to τ
- (2) $p = \lfloor (N - k + 1)/\tau \rfloor$;
- (3) **for** $j = 1$ to p
- (4) Coarse-graining the time series $y_{k,j}^{(\tau)} = (1/\tau) \sum_{i=(j-1)\tau+k}^{j\tau+k-1} x_i$;
- (5) **end for**
- (6) **end for**
- (7) **for** $k = 1$ to τ
- (8) **for** $i = 1$ to $p - m + 1$
- (9) Calculate the mean of each vector
 $u_{k,m}^{(\tau)}(i) = (1/m) \sum_{q=0}^{m-1} y_{k,i+q}^{(\tau)}$;
- (10) Move the vectors
 $Y_{k,m}^{(\tau)}(i) = \{y_{k,i}^{(\tau)}, y_{k,i+1}^{(\tau)}, \dots, y_{k,i+m-1}^{(\tau)}\} - u_{k,m}^{(\tau)}(i)$;
- (11) **end for**
- (12) **for** $i = 1$ to $p - m$
- (13) **for** $j = i + 1$ to $p - m + 1$
- (14) Calculate the Euclidean distance of the two
 vectors $Y_{k,m}^{(\tau)}(i)$ and $Y_{k,m}^{(\tau)}(j)$:
 $d_{k,m}^{(\tau)}(i, j) = d[Y_{k,m}^{(\tau)}(i), Y_{k,m}^{(\tau)}(j)]$
 $= \sqrt{\sum_{w=0}^{m-1} ((y_{k,i+w}^{(\tau)} - u_{k,m}^{(\tau)}(i)) - (y_{k,j+w}^{(\tau)} - u_{k,m}^{(\tau)}(j)))^2}$;
- (15) Calculate the similarity between $Y_{k,m}^{(\tau)}(i)$ and $Y_{k,m}^{(\tau)}(j)$ vectors
 $S_{k,m}^{(\tau)}(i, j) = \mu(d_{k,m}^{(\tau)}(i, j), r) = (1/(1 + d_{k,m}^{(\tau)}(i, j)/r))$;
- (16) **end for**
- (17) Calculate the average similarity between vector
 $Y_{k,m}^{(\tau)}(i)$ and the other vectors
 $B_{k,m}^{(\tau,r)}(i) = (1/(p - m - 1)) \sum_{j=i+1}^{p-m} S_{k,m}^{(\tau)}(i, j)$;
- (18) **end for**
- (19) Compute the average of $B_{k,m}^{(\tau,r)}(i)$, that is,
 $B_{k,m}^{(\tau,r)} = (1/(p - m)) \sum_{i=1}^{p-m} B_{k,m}^{(\tau,r)}(i)$;
- (20) **Set** dimensional length of vectors to $m + 1$ and **repeat** step 8~19 to calculate average similarity between each pair of $m + 1$
 points vectors in coarse-grained time series; you can get $A_{k,m+1}^{(\tau,r)}(i)$ and $A_{k,m+1}^{(\tau,r)}$
- (21) $A_{k,m+1}^{(\tau,r)}(i) = (1/(p - m - 2)) \sum_{j \neq i}^{p-m-1} S_{k,m+1}^{(\tau)}(i, j)$;
- (22) $A_{k,m+1}^{(\tau,r)} = (1/(p - m - 1)) \sum_{i=1}^{p-m-1} A_{k,m+1}^{(\tau,r)}(i)$;
- (23) Compute the Euclidean distance based on fuzzy
 sample entropy value for every $y_k^{(\tau)}$,
 $EuclideanFuzzyEn(y_k^{(\tau)}, m, r) = \lim_{p \rightarrow \infty} \{-\ln(A_{k,m+1}^{(\tau,r)}/B_{k,m}^{(\tau,r)})\}$;
- (24) **end for**
- (25) Compute the fuzzy sample entropy value for the original time series at time scale τ
 $EDM\ Fuzzy(m, r) = (1/\tau) \sum_{k=1}^{\tau} EuclideanFuzzyEn(y_k^{(\tau)}, m, r)$.

ALGORITHM 1: EDM-fuzzy algorithm.

Two days of traffic were collected in this trace, which were November 3 and December 1, as shown in Tables 3 and 4, respectively. There were 10 types of DDoS attacks on December 1, that is, NTP, DNS, LDAP, MSSQL, NetBIOS, SNMP, SSDP, UDP, UDPLag, and TFTP attack that lasted from 10:30 to 17:15. On November 3, there were 7 types of DDoS attacks including PortMap, NetBIOS, LDAP, MSSQL, UDP, UDPLag, and SYN attacks; the duration is from 9:40 to 17:35. The attack method of each type of DDoS attack in the CICDDoS2019 dataset is shown in Figure 1.

As shown in Figure 1, there are two types of DDoS attacks in the CICDDoS2019 trace, namely, reflection DDoS attacks and direct DDoS attacks. Both DDoS attacks are based on TCP/UDP protocol execution. As shown in the figure above, 9 types of DDoS attacks such as MSSQL, SSDP, DNS, LDAP, NetBIOS, SNMP, PortMap, NTP, and TFTP, are distributed reflective denial attacks, while SYN, UDP, and UDPLag Flood are direct DDoS attacks. In Figure 1, TCP-based attacks include MSSQL, SSDP, and SYN, and UDP-based attacks include NTP, TFTP, UDP, and UDPLag. The remaining types of attacks such as DNS, LDAP,

TABLE 1: Scenarios of Botnet traffic.

ID	IRC	SPAM	CF	PS	DDoS	FF	P2P	US	HTTP
1	✓	✓	✓	—	—	—	—	—	—
2	✓	✓	✓	—	—	—	—	—	—
3	✓	—	—	✓	—	—	—	✓	—
4	✓	—	—	—	✓	—	—	✓	—
5	—	✓	—	✓	—	—	—	—	✓
6	—	—	—	✓	—	—	—	—	—
7	—	—	—	—	—	—	—	—	✓
8	—	—	—	✓	—	—	—	—	—
9	✓	✓	✓	✓	—	—	—	—	—
10	✓	—	—	—	✓	—	—	—	✓
11	✓	—	—	—	✓	—	—	—	✓
12	—	—	—	—	—	—	✓	—	—
13	—	✓	—	✓	—	—	—	—	✓

TABLE 2: Traffic volume of 13 types of Botnet scenarios.

ID	Duration (hours)	Packets	Malware type	Infected hosts
1	6.15	71971482	Neris-1	1
2	4.21	71851300	Neris-2	1
3	66.85	167730395	Rbot-1	1
4	4.21	62089135	Rbot-2	1
5	11.63	4481167	Virut-1	1
6	2.18	38764357	Menti	1
7	0.38	7467139	Sogou	1
8	19.5	155207799	Murlo	1
9	5.18	115415321	Neris-3	10
10	4.75	90389782	Rbot-3	10
11	0.26	6337202	Rbot-4	3
12	1.21	13212268	NSIS.ay	3
13	16.36	50888256	Virut-2	1

TABLE 3: DDoS attack time on November 3.

Type ID	Attack type	Attack time
1	PortMap	9:43–9:51
2	NetBIOS	10:00–10:09
3	LDAP	10:21–10:30
4	MSSQL	10:33–10:42
5	UDP	10:53–11:03
6	UDPLag	11:14–11:24
7	SYN	11:28–17:35

TABLE 4: DDoS attack time on December 1.

Type ID	Attack type	Attack time
1	NTP	10:35–10:45
2	DNS	10:52–11:05
3	LDAP	11:22–11:32
4	MSSQL	11:36–11:45
5	NetBIOS	11:50–12:00
6	SNMP	12:12–12:23
7	SSDP	12:27–12:37
8	UDP	12:45–13:09
9	UDPLag	13:11–13:15
10	TFTP	13:35–17:15

NetBIOS, SNMP, PortMap, and other types of attacks are executed by using TCP or UDP.

5. Analysis of Network Anomaly Traffic with EDM-Fuzzy Entropy

Entropy-based time-series complexity measurement methods are widely used in fault diagnosis and anomaly detection of various complex systems. In this section, we apply EDM-Fuzzy in network traffic anomaly characterization and detection. The analysis of anomaly traffic characteristics based on MSE of Euclidean distance is an important part of the study of abnormal traffic. In this section, two anomalies of botnet and DDoS attack will be analyzed by Euclidean distance multiscale entropy. This section will calculate the entropy value of these two abnormal network protocol time series to obtain the entropy curves of the two and study the characteristics of the abnormal traffic by comparing the difference in the entropy curves.

5.1. Botnet Traffic in ARP. In this section, we will study the EDM-Fuzzy entropy characteristics of 13 types of botnets abnormal ARP traffic in the CTU-13 dataset. According to the TCP/IP architecture, the ARP protocol is located in the IP layer of the network layer, and its main function is to provide address translation services and find the network physical address of the host corresponding to the IP address. We first calculate the entropy values for each type of botnet using ARP protocol traffic data in the CTU-13 dataset at time scales from 1 to 40. The entropy curves of 13 types of botnets in the CTU-13 dataset with scale factors from 1 to 40 are shown in Figure 2.

As can be seen from the figure, there are common trends shared by entropy curves of most types of botnet traffic. More specifically, there is a reflection point for 11 entropy curves (Neris-1, Neris-2, Rbot-1, Rbot-2, Virut-1, Menti, Sogou, Murlo, Neris-3, NSIS.ay, and Virut-2) when the time scale is 20, and the second reflection point appears at the time scale of 30 for all entropy curves. For the above 11 types of botnet ARP traffic, the entropy values between the inflection points increase first and then decrease. The trend of the entropy curves of Rbot-3 and Rbot-4 is different from other types of abnormal behavior. Entropy curves of Rbot-3 and Rbot-4 are in a steady growth state when the time scale is around 20, but when the time scale is 30, there is also an inflection point. Moreover, entropy values of Rbot-4 are significantly larger compared to those of other types of anomalies. The above results illustrate that the attack methods of Rbot-3 and Rbot-4 are different from the other types of botnets. This difference is caused by the way they infect hosts, and the complexity of the botnet is consistent with the complexity of the ARP protocol.

5.2. DDoS Traffic in ARP. In this section, we will study the EDM-Fuzzy entropy characteristics of the malicious traffic of the distributed denial attacks on November 3 and December 1 in the CICDDoS2019 dataset. Through analysis of the trend of entropy values, it is possible to understand more characteristics of DDoS attack traffic. As introduced in the dataset, there were seven and ten types of distributed denial

FIGURE 1: Types of DDoS attacks.

FIGURE 2: Entropy curves of 13 types of botnets in the CTU-13 dataset.

FIGURE 3: Entropy curves of DDoS attacks on November 3.

attacks launched on November 3 and December 1, respectively. In this section, we first calculate the entropy value of the ARP traffic of each type of DDoS attack in the CICDDoS2019 dataset at time scales from 1 to 40, and the entropy value curves are shown in Figures 3 and 4.

Figures 3 and 4 show the entropy curves of the ARP traffic for DDoS attacks on November 3 and December 1, respectively. As can be seen from Figures 3 and 4, there are three characteristics of the entropy values of DDoS attacks on November 3 and December 1. Firstly, for both November 3 and December 1, all of the entropy values of DDoS attacks are larger than 0.18 when the time scale is larger than 4. Secondly, the entropy values of most types of DDoS attacks gradually stabilized to a value between 0.3 and 0.5 when the time scale is larger than 10. There is only the NetBIOS attack that has a relatively big fluctuation that may exceed the upper bound. Thirdly, the entropy values of the same type of DDoS attack for the two different days are quite similar. The possible underlying principle of the characteristics is that, in DDoS attacks, attackers continuously use distributed attacks to attack hosts, and the attacked hosts continue to

FIGURE 4: Entropy curves of DDoS attacks on December 1.

communicate during the attack. In the communication process, the ARP protocol continuously performs address resolution, while the number of resolved source IP addresses and destination IP addresses remains stable, so the entropy values of ARP traffic are gradually stabilized.

5.3. Normal Traffic in ARP. In this section, we will analyze the characteristics of network traffic under normal status. The normal traffic trace used in this paper is captured and published by the Stratosphere laboratory.

In order to study the entropy characteristics of normal traffic, the EDM-Fuzzy entropy values are calculated on the CTU-Normal-20 and CTU-Normal-23 traces with time scales from 1 to 40 and the results are shown in Figure 5.

As can be seen from Figure 5, the entropy values of normal ARP traffic exhibit different characteristics. The entropy values grow steadily when the time scale grows from 1 to 30, and then the entropy values grow slowly as the time scale increases. Furthermore, for all time scales, the entropy values of normal ARP traffic are smaller than 0.18.

Compared with the time series of the CTU-13 dataset, Figure 5 shows that the entropy value of the ARP protocol in the CTU-13 dataset exhibits its own unique laws. Compared with the CICDDoS2019 dataset, the basic law of the ARP protocol is that the entropy curve increases first and then gradually stabilizes.

5.4. Malicious versus Normal. In this section, we will compare the ARP traffic entropy curves between botnet, DDoS attack, and normal status and then characterize the differences between normal and abnormal traffic.

By comparing the entropy curves of ARP traffic of botnet, DDoS attack, and normal status, we find out the following main differences between normal traffic and malicious traffic. In the entropy curves of 13 types of botnets, 11 entropy curves (Neris-1, Neris-2, Rbot-1, Rbot-2, Virut-1, Menti, Sogou, Murlo, Neris-3, NSIS.ay, and Virut-2) have a reflection point at the time scale of 20, and all entropy curves have a reflection point at the time scale of 30. In the entropy curves of DDoS traffic, all of the entropy values of DDoS attacks are larger than 0.18 when the time scale is larger than 4, and most types of DDoS attacks gradually stabilized to entropy values between 0.3 and 0.5 when the time scale is larger than 10. In contrast, the entropy values of normal ARP traffic grow slowly as the time scale increases and the entropy values are smaller than 0.18 for all time scales.

In order to be presented more intuitively, the main characteristics of entropy curves of ARP traffic of botnet, DDoS attack, and normal status are listed in Table 5.

On the basis of the above analysis, it is reasonable to summarize that the characteristics of entropy curves of ARP traffic of botnet, DDoS, and normal status are quite distinguishable. Thus, the characteristics are easy to be used to detect these types of network traffic anomalies. In the future, we will study characteristics of EDM-Fuzzy entropy curves of more types of network traffic anomalies and utilize the learned characteristics of network traffic anomalies in

FIGURE 5: Entropy values of normal ARP traffic.

TABLE 5: Characteristics of EDM-Fuzzy entropy curves of ARP traffic of botnet, DDoS, and normal status.

	Botnet	DDoS	Normal
Value	Mostly between 0.1 and 0.4.	Larger than 0.18 when the time scale is larger than 4.	Smaller than 0.18.
Trend	Entropy curves of all types of botnet traffic have an inflection point at a time scale of 30; 11 types have an inflection point at a time scale of 20.	Gradually stabilized to a value between 0.3 and 0.5 when the time scale is larger than 10.	Increase steadily from 0 to 0.18.

combination with intelligent algorithms to automatically detect network anomalies.

6. Conclusions

In order to raise the cybersecurity awareness capability of network administrators, it is necessary to develop new technologies for detecting network anomalies more accurately and efficiently. The basis of such network anomaly detection technologies is to understand the characteristics of abnormal network traffic. In this paper, we apply the EDM-Fuzzy technology as a tool to analyze the characteristics of abnormal network traffic such as botnet network traffic and DDoS attack traffic. The EDM-Fuzzy is a technology that we proposed for analyzing and diagnosing faults/anomalies of complex systems by measuring the complexity and regularity of their time-series signals. The experimental analysis shows that the EDM-Fuzzy entropy curve is capable of characterizing the difference between normal traffic and abnormal traffic and the characteristics are easy to be used to detect various types of network traffic anomalies. In the current work, we have not investigated other types of

network anomalies and have not finished the automatic detection of network traffic anomalies. In the future, we will investigate EDM-Fuzzy entropy characteristics for more types of network anomalies and then integrate the EDM-Fuzzy entropy and deep-learning technologies to propose the novel network anomaly detection method.

Data Availability

The botnet traffic trace used in this section is the CTU-13 trace that was collected and provided by the Stratosphere Laboratory of CTU University in the Czech Republic [48, 49]. The traffic trace used in this paper is the CIC-DoS2019 which was published by the Canadian Cyber Security Institute (CIC) [48].

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was in part supported by the National Key Research and Development Program of China under Grant 2019YFB2102100, the China Postdoctoral Science Foundation under Grant 2016M600465, the Key Research and Development Program of Zhejiang Province under Grant 2019C03134, and the National Natural Science Foundation of China under Grant 61772165.

References

- [1] H. Zhang, Y. Luo, Q. Yu, L. Sun, X. Li, and Z. Sun, "A framework of abnormal behavior detection and classification based on big trajectory data for mobile networks," *Security and Communication Networks*, vol. 2020, Article ID 8858444, 15 pages, 2020.
- [2] J. Gou, W. Qiu, Z. Yi, Y. Xu, Q. Mao, and Y. Zhan, "A local mean representation-based K -nearest neighbor classifier," *ACM Transactions on Intelligent Systems and Technology*, vol. 10, no. 3, pp. 1–25, 2019.
- [3] V. Dutta, M. Choraś, R. Kozik, and M. Pawlicki, "Hybrid model for improving the classification effectiveness of network intrusion detection," in *Proceedings of the 13th International Conference on Computational Intelligence in Security for Information Systems (CISIS 2020)*, Burgos, Spain, September 2020.
- [4] F. Al-Obeidat and E.-S. M. El-Alfy, "Hybrid multicriteria fuzzy classification of network traffic patterns, anomalies, and protocols," *Personal and Ubiquitous Computing*, vol. 23, no. 5–6, pp. 777–791, 2019.
- [5] J. Dromard, G. Roudiere, and P. Owezarski, "Online and scalable unsupervised network anomaly detection method," *IEEE Transactions on Network and Service Management*, vol. 14, no. 1, pp. 34–47, 2017.
- [6] E. Bigdeli, M. Mohammadi, B. Raahemi, and S. Matwin, "Incremental anomaly detection using two-layer cluster-based structure," *Information Sciences*, vol. 429, pp. 315–331, 2018.
- [7] S. Baek, D. Kwon, S. C. Suh, H. Kim, I. Kim, and J. Kim, "Clustering-based label estimation for network anomaly detection," *Digital Communications and Networks*, vol. 7, no. 1, pp. 37–44, 2020.
- [8] G. Pu, L. Wang, J. Shen, and F. Dong, "A hybrid unsupervised clustering-based anomaly detection method," *Tsinghua Science and Technology*, vol. 26, no. 2, pp. 146–153, 2020.
- [9] J. Mao, Y. Hu, D. Jiang, T. Wei, and F. Shen, "CBFS: a clustering-based feature selection mechanism for network anomaly detection," *IEEE Access*, vol. 8, pp. 116216–116225, 2020.
- [10] H. H. Jazi, H. Gonzalez, N. Stakhanova, and A. A. Ghorbani, "Detecting HTTP-based application layer DoS attacks on web servers in the presence of sampling," *Computer Networks*, vol. 121, no. 5, pp. 25–36, 2017.
- [11] M. Thottan, G. Liu, and C. Ji, "Anomaly detection approaches for communication networks," in *Algorithms for Next Generation Networks*, G. Cormode and M. Thottan, Eds., Springer, London, UK, pp. 239–261, 2010.
- [12] J.-H. Bang, Y.-J. Cho, and K. Kang, "Anomaly detection of network-initiated LTE signaling traffic in wireless sensor and actuator networks based on a Hidden semi-Markov Model," *Computers & Security*, vol. 65, pp. 108–120, 2017.
- [13] G. Zheng, X. Xu, and J. Yan, "SD-CRF: a DoS attack detection method for SDN," in *Proceedings of the 2020 IEEE 20th International Conference on Communication Technology (ICCT)*, Nanning, China, October 2020.
- [14] Z. Wang, Y. Zeng, Y. Liu, and D. Li, "Deep belief network integrating improved kernel-based extreme learning machine for network intrusion detection," *IEEE Access*, vol. 9, no. 1, pp. 16062–16091, 2021.
- [15] Z. Wang, Y. Liu, D. He, and S. Chan, "Intrusion detection methods based on integrated deep learning model," *Computers & Security*, vol. 103, p. 102177, 2021.
- [16] D. K. K. Reddy, H. S. Behera, J. Nayak, P. Vijayakumar, B. Naik, and P. K. Singh, "Deep neural network based anomaly detection in Internet of Things network traffic tracking for the applications of future smart cities," *Transactions on Emerging Telecommunications Technologies*, p. e4121, 2020.
- [17] N. Marir, H. Wang, G. Feng, B. Li, and M. Jia, "Distributed abnormal behavior detection approach based on deep belief network and ensemble SVM using spark," *IEEE Access*, vol. 6, pp. 59657–59671, 2018.
- [18] K. Flanagan, E. Fallon, P. Connolly, and A. Awad, "Network anomaly detection in time series using distance based outlier detection with cluster density analysis," in *Proceedings of the 2017 Internet Technologies and Applications (ITA)*, Wrexham, UK, September 2017.
- [19] C. B. Zerbini, L. F. Carvalho, T. Abrão, and M. L. Proença, "Wavelet against random forest for anomaly mitigation in software-defined networking," *Applied Soft Computing*, vol. 80, pp. 138–153, 2019.
- [20] Sharipuddin, B. Purnama, Kurniabudi et al., "Features extraction on IoT intrusion detection system using principal components analysis (PCA)," in *Proceedings of the 2020 7th International Conference on Electrical Engineering, Computer Sciences and Informatics (EECSI)*, Yogyakarta, Indonesia, October 2020.
- [21] Y. Zhong, W. Chen, Z. Wang et al., "HELAD: a novel network anomaly detection model based on heterogeneous ensemble learning," *Computer Networks*, vol. 169, p. 107049, 2020.
- [22] P. Bereziński, B. Jasiul, and M. Szpyrka, "An entropy-based network anomaly detection method," *Entropy*, vol. 17, no. 4, pp. 2367–2408, 2015.

- [23] S. Ranjan, S. Shah, A. Nucci, M. Munafo, R. Cruz, and S. Muthukrishnan, "DoWitcher: effective worm detection and containment in the internet core," in *Proceedings of 26th IEEE International Conference on Computer Communications (INFOCOM 2007)*, pp. 2541–2545, Anchorage, AL, USA, May 2007.
- [24] Y. Gu, A. McCallum, and D. Towsley, "Detecting anomalies in network traffic using maximum entropy estimation," in *Proceedings of the 5th ACM SIGCOMM Conference on Internet Measurement (IMC '05)*, p. 32, Berkeley, CA, USA, October 2005.
- [25] S. Ransewa, N. Elz, N. Thanon, and S. Intajag, "Anomaly detection using source port data with shannon entropy and EWMA control chart," in *Proceedings of the 2018 18th International Conference on Control, Automation and Systems (ICCAS)*, pp. 596–601, PyeongChang, Korea, October 2018.
- [26] S. Daneshgadeh, T. Kemmerich, T. Ahmed, and N. Baykal, "An empirical investigation of DDoS and Flash event detection using Shannon entropy, KOAD and SVM combined," in *Proceedings of the 2019 International Conference on Computing, Networking and Communications (ICNC)*, pp. 658–662, Honolulu, HI, USA, February 2019.
- [27] A. Howedi, A. Lotfi, and A. Pourabdollah, "An entropy-based approach for anomaly detection in activities of daily living in the presence of a visitor," *Entropy*, vol. 22, no. 8, p. 845, 2020.
- [28] J. S. Cánovas, G. García-Clemente, and M. Muñoz-Guillermo, "Comparing permutation entropy functions to detect structural changes in time series," *Physica A: Statistical Mechanics and its Applications*, vol. 507, no. 1, pp. 153–174, 2018.
- [29] A. N. Kolmogorov, "Three approaches to the quantitative definition of information," *International Journal of Computer Mathematics*, vol. 2, no. 1–4, pp. 156–168, 1968.
- [30] S. Claude, "A mathematical theory of communications," *Bell Labs Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.
- [31] S. M. Pincus, "Approximate entropy as a measure of system complexity," *Proceedings of the National Academy of Sciences*, vol. 88, no. 6, pp. 2297–2301, 1991.
- [32] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *American Journal of Physiology: Heart and Circulatory Physiology*, vol. 278, no. 6, pp. 2039–2049, 2000.
- [33] W. Chen, J. Zhuang, W. Yu, and Z. Wang, "Measuring complexity using FuzzyEn, ApEn, and SampEn," *Medical Engineering & Physics*, vol. 31, no. 1, pp. 61–68, 2009.
- [34] M. Costa, A. L. Goldberger, and C. K. Peng, "Multiscale entropy analysis of biological signals," *Physical Review E*, vol. 71, no. 2, pp. 1–18, 2005.
- [35] S.-D. Wu, C.-W. Wu, S.-G. Lin, C.-C. Wang, and K.-Y. Lee, "Time series analysis using composite multiscale entropy," *Entropy*, vol. 15, no. 3, pp. 1069–1084, 2013.
- [36] L. Dou, S. Wan, and C. Zhan, "Application of multiscale entropy in mechanical fault diagnosis of high voltage circuit breaker," *Entropy*, vol. 20, no. 5, pp. 325–329, 2018.
- [37] I. Khan, Y. L. Xu, S. Kar, M. Chow, and V. Bhattacharjee, "Compressive sensing and morphology singular entropy-based real-time secondary voltage control of multi-area power systems," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 7, pp. 3796–3807, 2019.
- [38] R. Zhou, X. Wang, J. Wan, and N. Xiong, "EDM-fuzzy: an euclidean distance based multiscale fuzzy entropy technology for diagnosing faults of industrial systems," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 6, pp. 4046–4054, 2020.
- [39] P. Casas, J. Mazel, and P. Owezarski, "UNADA: unsupervised network anomaly detection using sub-space outliers ranking," in *Networking 2011*, J. Domingo-Pascual, P. Manzoni, S. Palazzo, A. Pont, and C. Scoglio, Eds., Springer, Berlin-Germany, 2011.
- [40] F. A. S. Borges, R. A. S. Fernandes, I. N. Silva, and C. B. S. Silva, "Feature extraction and power quality disturbances classification using smart meters signals," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 2, pp. 824–833, 2016.
- [41] S. A. Shevchik, F. Saeidi, B. Meylan, and K. Wasmer, "Prediction of failure in lubricated surfaces using acoustic time-frequency features and random forest algorithm," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 4, pp. 1541–1553, 2017.
- [42] M. Costa, A. L. Goldberger, and C. K. Peng, "Multiscale entropy analysis of complex physiologic time series," *Physical Review Letters*, vol. 89, no. 6, pp. 1–4, 2002.
- [43] C. M. Galanakis, "Modeling in food and bioproducts processing using Boltzmann entropy equation: a viewpoint of future perspectives," *Food and Bioproducts Processing*, vol. 106, no. 1, pp. 102–107, 2017.
- [44] H.-L. Niu and J. Wang, "Entropy and recurrence measures of a financial dynamic system by an interacting voter system," *Entropy*, vol. 17, no. 5, pp. 2590–2605, 2015.
- [45] Y. Wang, S. Zheng, W. Zhang, G. Wang, and J. Wang, "Fuzzy entropy complexity and multifractal behavior of statistical physics financial dynamics," *Physica A: Statistical Mechanics and its Applications*, vol. 506, no. 15, pp. 486–498, 2018.
- [46] Y. Li, Y. Yang, X. Wang, B. Liu, and X. Liang, "Early fault diagnosis of rolling bearings based on hierarchical symbol dynamic entropy and binary tree support vector machine," *Journal of Sound and Vibration*, vol. 428, no. 18, pp. 72–86, 2018.
- [47] J. Zheng, H. Pan, and J. Cheng, "Rolling bearing fault detection and diagnosis based on composite multiscale fuzzy entropy and ensemble support vector machines," *Mechanical Systems and Signal Processing*, vol. 85, no. 15, pp. 746–759, 2017.
- [48] "The CTU-13 Dataset. a labeled dataset with botnet, normal and background traffic on stratosphere research laboratory," <https://www.stratosphereips.org/datasets-ctu13>.
- [49] S. García, M. Grill, J. Stiborek, and A. Zunino, "An empirical comparison of botnet detection methods," *Computers & Security*, vol. 45, pp. 100–123, 2014.
- [50] Canadian Institute for Cybersecurity, *DDoS Evaluation Dataset (CIC-DDoS2019)* Canadian Institute for Cybersecurity, Fredericton, Canada, 2019, <https://www.unb.ca/cic/datasets/ddos-2019.html>.