Computers, Materials & Continua CMC, vol.63, no.1, pp.131-150, 2020

A DDoS Attack Information Fusion Method Based on CNN for Multi-Element Data

Jieren Cheng[1, 2], Canting Cai[1, *], Xiangyan Tang[1], Victor S. Sheng[3], Wei Guo[1] and Mengyang Li[1]

Abstract: Traditional distributed denial of service (DDoS) detection methods need a lot of computing resource, and many of them which are based on single element have high missing rate and false alarm rate. In order to solve the problems, this paper proposes a DDoS attack information fusion method based on CNN for multi-element data. Firstly, according to the distribution, concentration and high traffic abruptness of DDoS attacks, this paper defines six features which are respectively obtained from the elements of source IP address, destination IP address, source port, destination port, packet size and the number of IP packets. Then, we propose feature weight calculation algorithm based on principal component analysis to measure the importance of different features in different network environment. The algorithm of weighted multi-element feature fusion proposed in this paper is used to fuse different features, and obtain multi-element fusion feature (MEFF) value. Finally, the DDoS attack information fusion classification model is established by using convolutional neural network and support vector machine respectively based on the MEFF time series. Experimental results show that the information fusion method proposed can effectively fuse multi-element data, reduce the missing rate and total error rate, memory resource consumption, running time, and improve the detection rate.

Keywords: DDoS attack, multi-element data, information fusion, principal component analysis, CNN.

1 Introduction

With the rapid development and popularization of information technology, the internet has penetrated into all aspects of society. According to the "Statistical Report on the Development of China's Internet" [China Internet Network Information Center (2018)], as of June 2018, the number of Chinese Internet users has reached 802 million. Distributed Denial of Service (DDoS) attacks are mainly targeted at network bandwidth and server host, and focus on sending a large number of seemingly legitimate but useless

[1] School of Information Science and Technology, Hainan University, Haikou, 570228, China.

[2] Key Laboratory of Internet Information Retrieval of Hainan Province, Hainan University, Haikou, 570228, China.

[3] Department of Computer Science, University of Central Arkansas, Conway, AR 72035, US.

[*] Corresponding Author: Canting Cai. Email: canting@hainanu.edu.cn.

Received: 23 January 2019; Accepted: 13 June 2019.

network packets to a victim host from a large number of zombie hosts, and then, denial of service occurs because of network congestion and depletion of network resources. The result is that network packets from the legitimate users are submerged, and the legitimate users cannot access the network resources in the server.

The Arbor Networks Report [Arbor Networks (2018)] shows that the frequency and complexity of DDoS attacks are rising in recent years, and there are 7.5 million DDoS attack cases in 2017, covering approximately one-third of global Internet traffic. It can be seen that DDoS attacks are still very rampant and are the main method of cyber-attack. Therefore, it is necessary to find a way to effectively integrate multiple features of DDoS attack to identify DDoS attacks more accurately.

The paper is organized as follows, Section 2 is the introduction of related work, Section 3 is the definition of multi-element features of DDoS attacks, Section 4 is the introduction of the information fusion method based on multi-element features, Section 5 is the introduction of the experiment, and Section 6 is the summary and outlook of the full text.

2 Related works

In recent years, DDoS attacks have spread more and more widely, and more and more fields are involved. A large number of researchers have also done a lot of research on DDoS attacks. Sahoo et al. [Sahoo, Puthal, Tiwary et al. (2018)] studied the stream-based nature of software defined networks and proposed a detection method for low-rate DDoS attack on the control layer based on the measurement of generalized entropy (GE). Real-time detection of abnormal network activities in the system log can be achieved by an online unsupervised deep learning method [Tuor, Kaplan, Hutchinson et al. (2017)]. According to the DDoS attack characteristics, Cheng proposed many different types of classifiers [Cheng, Zhou, Liu et al. (2018); Cheng, Tang and Yin (2017); Cheng, Zhang, Tang et al. (2018)]. Idhammad et al. [Idhammad, Afdel and Belouch (2018)] proposed a semi-supervised method for DDoS attack detection. Liu et al. [Liu, Cai, Xu et al. (2015)] proposed a solution which offers physical isolation during virtual network embedding to make sure the security of network. Cheng et al. [Cheng, Xu, Tang et al. (2018)] proposed a method to detect DDoS attacks based on changes of old IP address and new IP address, which could be identified in the early stages of the attack. With the development and popularity of IoT devices, DDoS attack in IoT devices become more and more serious. In 2016, hacker Anna-senpai launched large-scale DDoS attack by mirai virus in IoT devices, which broke the history record of DDoS attack traffic. As DDoS attacks gradually penetrate into IoT devices, the attack methods are more and more complex and diverse, and the damage is getting bigger and bigger. Antonakakis et al. [Antonakakis, April, Bailey et al. (2017)] analyzed the emergence and evolution of Mirai, got the characteristic of the vulnerable IoT devices, and proposed technical and non-technical interventions measure. In conclusion, the attack methods of DDoS attacks are updated rapidly, and the method with single elements can't identify DDoS attacks well. It is necessary to put forward an information fusion method based on multi-element features to detected DDoS attack.

Nowadays, in the era of big data, it is full of massive, diverse, high-speed and variable data everywhere. Information fusion is a multi-level, multi-faceted and multi-dimensional

deep processing process for multi-source and heterogeneous data, and it can get more complete, more accurate, and timelier process results. At present, many researchers are working and searching in the field of information fusion, and there have been many scientific research results. Wu et al. [Wu and Wang (2018)] developed a game theory analysis of detection strategies combining Nash equilibrium theory. MüUller et al. [MüUller, Kuwertz, Mühlenberg et al. (2017)] designed data fusion component to achieve situational awareness and helped people make correct judgments. Golestan et al. [Golestan, Khaleghi, Karray et al. (2016)] solved the problem of road safety by combining information fusion with bayesian network. Lin [Lin (2016)] proposed a multi-sensor information fusion algorithm, mainly through the neural network and Bayesian. Costa et al. [Costa, Yu, Atiahetchi et al. (2018)] proposed an architecture that used probabilistic ontology to accelerate the process of network asset planning. Yuan et al. [Yuan and Li (2018)] proposed an information fusion mechanism that can effectively resist attacks. Li et al. [Li, Lu, Liu et al. (2018)] proposed a network security situation assessment method, and solved the limitation of some assessment methods that only focus on attack behaviors. Liu et al. [Liu, Pan, Zhang et al. (2017)] proposed a multi-source information fusion method which calculate the credibility of emergency messages by analyzing the data from on-board sensors to detect forged emergency messages. Smart devices raise the issue of data fusion [Esposito, Castiglione, Palmieri et al. (2018)]. Lian et al. [Lian, Zhang, Xie et al. (2018)] proposed a deep fusion model to improve learning ability.

In view of the characteristic of DDoS attacks, diversity, abruptness, high traffic and unpredictability, this paper proposes a new DDoS attack information fusion method based on multi-element data, which can fuse multi-element data effectively, improve the detection rate for DDoS attacks, lower the missing rate and total error rate, and reduce running time and memory usage.

3 Multi-element feature definition of DDoS attack

This paper analyzes the characteristics of DDoS attack and the difference between normal flow and attack flow, and extracts features according to the characteristics of DDoS attack.

3.1 DDoS attack feature analysis

Through the research on a large number of classical DDoS attacks cases, it is found that the types of DDoS attacks are increasingly diverse, and the attack methods are more and more complex. DDoS attacks are of distribution, concentration and high traffic abruptness as follows:

(1) Distribution. When an attack is launched, the attacker can send a large number of useless data packets to the target host by controlling a large number of puppet machines, and the attacker can forge a great quantity of fake source IP addresses continuously or randomly. Puppet machines choose to use a random source port to attack. Since attackers want to exhaust the network resources of the target host as much as possible, puppet machines occupy all ports of the target host as much as possible. Therefore, when an attack occurs, the source IP address is distributed, the source port is distributed, and the destination port is distributed.

(2) Concentration. When an attack is launched, the attacker chooses a specific target host

to attack, and the size of attacking packets is consistent. Therefore, when an attack occurs, the destination IP address is concentrated, and the size of packets is concentrated.

(3) High traffic abruptness. When an attack is launched, many puppet machines send a large number of useless data packets to the target host. Therefore, when an attack occurs, the number of packets suddenly increases.

According to the above summary, multiple elements in the network stream will change when an attack is launched. If only a single element is considered, the current network condition cannot be fully represented. Therefore, this paper extracts multi-element features for fusion.

3.2 Extraction of multi-element features

Assume that the network flow F in unit time T is $\left\langle \left(t_1, sip_1, dip_1, sp_1, dp_1, p_1\right), \cdots, \left(t_n, sip_n, dip_n, sp_n, dp_n, p_n\right)\right\rangle$, where $i = 1, 2, \cdots, n$, $t_i, sip_i, dip_i, sp_i, dp_i, p_i$ represent time, source IP address, destination IP address, source port, destination port, and packet's size of the $i-th$ data packet respectively.

Definition 1. During sampling time, the Source IP Address Feature (SIPAF) of the network flow F is defined as follows:

$$SIPAF = \left| \bigcup_{i=1}^{n} \{sip_i\} \right| \tag{1}$$

In the definition of SIPAF, the type number of source IP address of the network flow F per unit time is counted, and this feature can better reflect the network flow situation. DDoS attack is an attack that the attacker sends a large number of useless packets to a victim host from a large number of fake IP addresses and the requests from normal legitimate network users will be covered up, which can achieve the purpose of attacking a victim host and consuming network resources. According to the analysis, the number of different source IP addresses in the network stream should be less and stable in a period of time under normal circumstances. When an attack is launched, the number of different source IP addresses will increase suddenly because a large number of fake IP addresses flood in the network stream. SIPAF is larger under attacking than that in normal circumstances; therefore, it can effectively distinguish normal network flow and abnormal network flow.

Definition 2. During sampling time, the Destination IP Address Feature (DIPAF) of the network flow is defined as follows:

$$DIPAF = \left| \bigcup_{i=1}^{n} \{dip_i\} \right| \tag{2}$$

In the definition of DIPAF, the type number of destination IP address of the network flow F per unit time is counted. According to the analysis, the number of different destination IP addresses in the network stream will be more and stable under normal circumstances. When an attack is launched, the attacker will find a target host, and the destination IP

address is relatively concentrated. Therefore, DIPAF is less under attacking than that in normal circumstances, which can effectively distinguish normal network flow from abnormal network flow.

Definition 3. During sampling time, the Source Port Feature (SPF) of the network flow is defined as follows:

$$SPF = \left| \bigcup_{i=1}^{n} \{sp_i\} \right| \tag{3}$$

In the definition of SPF, the type number of source port of the network flow F per unit time is counted. DDoS attack is an attack that the attacker sends a large number of useless packets to the victim target host by controlling a large number of puppet machines which will select the ports randomly. Under normal circumstances, the type number of source port in the network stream will be less and stable. When an attack occurs, the number of source port will increase.

Definition 4. During sampling time, the destination Port Feature (DPF) of the network flow is defined as follows:

$$DPF = \left| \bigcup_{i=1}^{n} \{dp_i\} \right| \tag{4}$$

In the definition of DPF, the type number of destination port of the network flow F per unit time is counted. In order to exhaust network resources, attackers will occupy network resources as much as possible, which makes it impossible for normal users to access to the network. The puppet machine occupies different ports of the victim target host. Under normal circumstances, the number of different destination ports in the network flow is in a low level. On the contrary, it will increase suddenly when attacks occur.

Definition 5. During sampling time, the Packet Number Feature (PNF) of the network flow is defined as follows:

$$PNF = n \tag{5}$$

In the definition of PNF, the number of packets of the network flow F per unit time is counted. According to the analysis, the number of packets is less under normal circumstances than that under attacking.

Definition 6. During sampling time, the Packet Size Feature (PSF) of the network flow is defined as follows:

$$PSF = \left| \bigcup_{i=1}^{n} \{p_i\} \right| \tag{6}$$

In the definition of PSF, the type of packets' size of the network flow F per unit time is counted. In normal network, the size of a video and a text are different obviously, even the same text, their size may be different in different environments. However, the size of DDoS attack packets is of the same size. According to the analysis, under normal circumstances, there are almost different sizes of packets in the network stream. However, the size of packets is all the same when an attack is launched. Therefore, PSF will be less in attacking than that in normal circumstances.

The six multi-element features defined above can reflect the current network situation, but they are not applicable to all situations. For example, when network congestion occurs, there may be misjudged as the occurrence of DDoS attack. Therefore, this paper proposes a DDoS attack information fusion method based on CNN for multi-element features, it can fuse multi-element features form multiple perspectives, which can reflect the real situation of the network more accurately.

4 Information fusion method

Information fusion is the process of correlating and synthesizing data and information obtained from single and multiple sources to get accurate location and identity estimates, as well as a comprehensive and timely assessment of the threats and their importance. Nowadays, information fusion technology is especially needed in the era of big data, and many researchers have done some related work in the field of information fusion. For example, some researchers put for-ward several unique context development dynamics and architectures [Snidaro, García and Llinas (2015)]. Information fusion is a discussion and evaluation of the quality of information and con-textual quality, the relationship between them, and their impact on the performance of fusion systems [Rogova and Snidaro (2018)]. Paggi et al. [Paggi, Soriano and Lara (2018)] proposed a multi-agent information fusion system model to improve the quality of processed information. Guo et al. [Guo, Yin, Li et al. (2018)] proposed a method to improve the recommendation system by using improved Dempster-Shafer theory to fuse multiple sources of information. In the big data environment, the data is diverse, but they can be more comprehensively viewed through the way of information fusion.

4.1 Feature weight calculation model

Due to the expression abilities of different features in different network environment are different, the ability to describe the network is not the same. The features acquired on the victim and the attacker is also different. Therefore, we propose to obtain different weights through principal component analysis method to measure the expression abilities of different features in current network environment.

Principal Component Analysis (PCA) is a multivariate statistical method that investigates the correlation among multiple variables. It studies how to reveal the internal structure among multiple variables through a few principal components. Principal component analysis can eliminate the interferences among the evaluation indicators, because principal component analysis transforms the original data indicator variables to form mutually independent principal components. Since principal component analysis is a multivariate analysis method, it is suitable for the processing of multivariate features in this paper. The feature weight calculation model based on principal component analysis proposed in this paper mainly considers the contribution of each feature in the multivariate features to determine the value of weight.

First of all, according to the extraction of the multi-element features in Section 3, the multi-element features are obtained as follows:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{16} \\ x_{21} & x_{22} & \cdots & x_{26} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{n6} \end{bmatrix} = \begin{pmatrix} X_1 & X_2 & \cdots & X_6 \end{pmatrix} \tag{7}$$

$$X_i = \begin{bmatrix} x_{1i} \\ x_{2i} \\ \vdots \\ x_{ni} \end{bmatrix} \tag{8}$$

where X_1, X_2, X_3, X_4, X_5 and X_6 represent *SIPAF*, *DIPAF*, *SPF*, *DPF*, *PNF* and *PSF* respectively, n represents the number of samples.

Through normalizing the matrix X in Formula 8, the data matrix Z is obtained:

$$Z_{ij} = \frac{x_{ij} - \overline{x}_j}{\sigma_{x_j}},\ i \neq j,\ i = 1, 2, \cdots, n;\ j = 1, 2, \cdots, 6 \tag{9}$$

$$Z = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{16} \\ z_{21} & z_{22} & \cdots & z_{26} \\ \vdots & \vdots & \vdots & \vdots \\ z_{n1} & z_{n2} & \cdots & z_{n6} \end{bmatrix} = \begin{pmatrix} Z_1 & Z_2 & \cdots & Z_6 \end{pmatrix} \tag{10}$$

where $\overline{x}_j$ represents the mean of the $j\text{-}th$ column, σ_{x_j} represents the standard deviation of the $j\text{-}th$ column.

By using Formula 10, covariance matrix R is obtained:

$$R = \frac{1}{n-1} ZZ^T \tag{11}$$

Calculate the characteristic root λ_i and characteristic vector γ_i of the matrix R, six linear combinations of principal components are obtained:

$$\begin{aligned} F_1 &= \gamma_{11} X_1 + \gamma_{21} X_2 + \cdots + \gamma_{61} X_6 \\ F_2 &= \gamma_{12} X_1 + \gamma_{22} X_2 + \cdots + \gamma_{62} X_6 \\ &\cdots\cdots \\ F_6 &= \gamma_{16} X_1 + \gamma_{26} X_2 + \cdots + \gamma_{66} X_6 \end{aligned} \tag{12}$$

Calculate the variance contribution rate of the $j\text{-}th$ principal component according to Formula 12:

$$\alpha_i = \frac{\lambda_i}{\sum_{k=1}^{6} \lambda_k} \tag{13}$$

When the cumulative variance contribution rate of the current $m\text{-}th$ principal components is greater than 85%, the m principal components are selected. Calculate the weight of each feature by using Formula 13, then get the final weight of each feature by normalization.

$$\omega_i = \frac{\sum_{k=1}^{m} \gamma_{ik} \alpha_k}{\sum_{k=1}^{m} \alpha_k} \tag{14}$$

where w_1, w_2, w_3, w_4, w_5 and w_6 represent the weight of *SIPAF*, *DIPAF*, *SPF*, *DPF*, *PNF* and *PSF* respectively.

4.2 Weighted multi-element feature fusion

The current network situation is more and more complex, and the single-element feature can only unilaterally express the network situation. For the characteristics of high flow and changeability of DDoS attack, single-element feature cannot identify DDoS attack accurately. This paper proposes a multi-element features information fusion method to consider information from multiple perspectives. A weighted feature-level fusion method is proposed to deal with the six multi-element features extracted in Section 3, and it considers the information of multi-element features comprehensively, which can reflect the current network environment more accurately.

This paper defines a Multi-element Fusion Feature (MEFF) which are calculated from the six multi-element features of information, including *SIPAF*, *DIPAF*, *SPF*, *DPF*, *PNF* and *PSF*.

$$\begin{aligned} MEFF = {} & \omega_1 \lg(SIPAF) + \omega_2 \lg(DIPAF) + \omega_3 \lg(SPF) + \\ & \omega_4 \lg(DPF) + \omega_5 \lg(PNF) + \omega_6 \lg(PSF) \end{aligned} \tag{15}$$

where w_1, w_2, w_3, w_4, w_5 and w_6 represent the weight of the six features respectively by calculating from principal component analysis in Section 4.1. This paper calculates the logarithm of *SIPAF*, *DIPAF*, *SPF*, *DPF*, *PNF* and *PSF* because the values of features are highly differentiated. If logarithm operation is not carried out, the direction of the gradient will deviate, the training time will be too long and the effect will be not ideal when taking gradient descent. After carrying out the logarithm operation, the value of features is relatively concentrated, and the precision and convergence speed are improved.

4.3 Classification model based on CNN

In order to verify the correctness of the information fusion method proposed in this paper, we construct a classification model based on convolutional neural network. Convolutional Neural Network (CNN) is a typical artificial feed-forward neural network,

which essentially extracts the characteristics of input data by establishing multiple filters. As the number of network layers increases, CNN continuously analyzes the extracted features to obtain the final features. CNN has two characteristics: local connection and weight sharing. The convolutional layer and the previous layer are connected by local connection and weight sharing, which greatly reduces the number of parameters, reduces the network complexity, makes the network more robust, and can effectively prevent over-fitting.

The basic structure of the convolutional neural network: input layer, convolutional layer, pooling layer, fully connected layer, and output layer. In general, the convolutional layer and the pooling layer alternately appear. Finally, the features of the pooling layer are connected to form a feature vector, and the feature vector obtains a classification vector through the fully connected layer.

Convolutional layer. The convolution layer is composed of multiple feature maps, and each feature map is composed of multiple neurons. Each neuron is connected to the upper feature map by the convolutional kernel. Convolutional layer extracts the features of different levels of input layer through convolution. The form of convolutional layer is as follows:

$$x_j^l = f\left(\sum_{i \in M_j} x_j^{l-1} k_{ij}^l + b_j^l\right) \tag{16}$$

where l represents the current layer, b represents the bias of the current layer, k represents the convolutional kernel, M_j represents the convolution window of the $j\text{-}th$ convolutional kernel. Activation functions are commonly used sigmoid, tanh, relu. In this paper, we choose relu activation function. Relu activation function is defined as follows:

$$f(x) = \max(0, x) \tag{17}$$

When $x > 0$, the gradient is always 1, and there is no gradient dissipation problem, and convergence is fast. When $x < 0$, the output of this layer is 0. The more neurons that are 0 after training, the more and more sparse they will be. The extracted features will be approximately representative and the stronger the generalization ability will be.

Pooling layer. The pooling layer is also composed of multiple feature maps behind the convolutional layer. Each feature map of the pooling layer only corresponds to one feature map of the previous layer, and the number of feature maps does not change. The convolutional layer is the input layer of the pooling layer. The form of pooling layer is as follows:

$$x_j^l = f\left(\beta_j^l down(x_j^{l-1}) + b_j^l\right) \tag{18}$$

where $down(x_j)$ represents down-sampling of the $j\text{-}th$ neuron. Each output feature map has weight β and bias b.

Fully connected layer. After multiple convolutional layers and pooling layers, one or more fully connected layers are connected. Each neuron in the fully connected layer is fully connected to all neurons in the previous layer. The activation function of each

neuron in the fully connected layer usually chooses relu function, and the output value of the last fully connected layer is delivered to an output layer, which can be classified by softmax.

In this paper, a one-dimensional convolutional neural network consisting of three convolutional layers, three pooling layers and two fully connected layers is constructed.

5 Experiment

5.1 Experimental data set and evaluation standard

This paper selects the data set of CAIDA DDoS Attack 2007. The size of data set is 21 GB. We introduce five related performance evaluation standards to evaluate the experimental results including Detection Rate (DR), Missing Rate (MR), False alarm rate (FR), Error Rate (ER), Accuracy, Running Time (RT) and Memory Usage (MU). The calculation formula of evaluation standards define as follows:

$$DR = \frac{TN}{TN + FN} \tag{19}$$

$$MR = \frac{FN}{TN + FN} \tag{20}$$

$$ER = \frac{FN + FP}{TP + FP + TN + FN} \tag{21}$$

$$Accuracy = 1 - ER \tag{22}$$

where TN represents the number of attack samples that are correctly identified, FN represents the number of attack samples that are misidentified, TP represents the number of normal samples that are correctly identified, FP represents the number of normal samples that are misidentified.

$$RT = \frac{\sum_{i=1}^{n} time}{n} \tag{23}$$

where *time* represents the running time of program each time, n represents the times of running program.

$$MU = \frac{\sum_{i=1}^{n} memory}{n} \tag{24}$$

where *memory* represents the amount of memory used of program each time.

5.2 Experimental results and analysis

In this paper, normal data samples and attack data samples are obtained from the data set of DDoS Attack 2007. First of all, the six features of *SIPAF*, *DIPAF*, *SPF*, *DPF*, *PNS* and *PSF* were extracted according to the rule of multi-element features extraction in Section 3 with $\triangle t = 0.1s$. Then, according to the weight determination method in Section 4.1, the weight of the features in Formula 14 are obtained that w_1, w_2, w_3, w_4, w_5 and w_6 are 0.186,

0.122, 0.185, 0.19, 0.186 and 0.131 respectively. Finally, *MEFF* is obtained from multi-element features fusion formula.

In order to verify the validity and universality of the multi-element feature information fusion method proposed in this paper, we performed comparison experiments, and the specific steps and the results of the comparison experiment are as follows.

5.2.1 Comparison of the performance of MEFF and other features based on CNN

In this experiment, the number of training set samples is unchanged, and five different test set samples are randomly selected from the test set that it contains normal flow and attack flow. The number of the five test set samples are 500, 1000, 2000, 5000 and 10000. In this experiment, we compare the performance of detection rate, missing rate and error rate between MEFF feature and other six features in different number of samples based on CNN model.

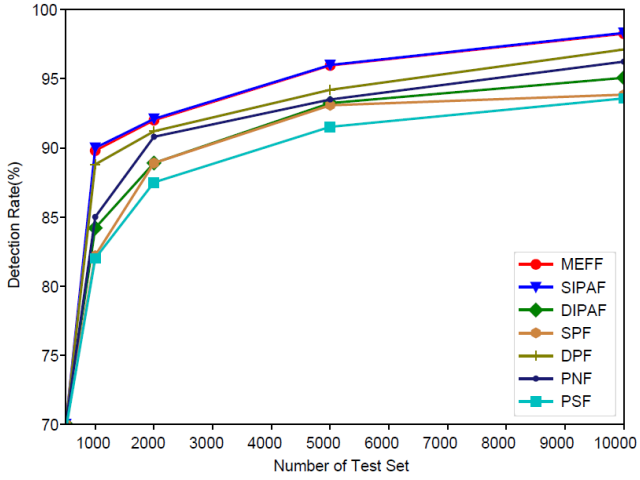

Figure 1: Comparison of detection rate in different test set samples based on CNN

It can be seen from Fig. 1 that MEFF, SIPAF, DIPAF, SPF, DPF, PNF and PSF can detect DDoS attacks better. When the number of test samples is 500, the detection rate of each feature is 70% except PSF whose detection rate is 69.6%. However, when the number of samples is 1000, it is obvious that the detection rate of each feature is greatly different. Among these features, the two features with high detection rate are MEFF and SIPAF, and they are 89.8% and 90% respectively. The gap between them is only 0.2%. It can be seen that the detection effect of these two features is not much different. However, the detection rates of DIPAF, SPF, DPF, PNF and PSF are 84.2%, 82.2%, 88.8%, 85% and 82% respectively. Compared with the detection rate of MEFF feature, the detection rates of other features are quite worse than MEFF feature. MEFF feature has better detection effect. When the number of samples in the test set is 2000, the detection rate of PSF feature is the lowest, only 87.5%, while the detection rate of MEFF feature and SIPAF feature are 92% and 92.1%, the difference gap between them is only 0.1%. When the number of samples in the test set is 5000, the detection rate of MEFF and SIPAF remains high, and the gap between them is smaller, only 0.04%. On the contrary, the

detection rate of DPF feature and PNF feature increases slowly compared with the case when the sample size is 2000. When the number of samples is 10000, the trend of features' detection rate is relatively stable. According to the experimental results, we can find that with the increase of sample size, the gap of detection rate between MEFF and SIPAF feature is getting smaller and smaller, which indicates that the fusion feature MEFF proposed in this paper can effectively identify DDoS attacks. When the number of samples is different, the detection rate of MEFF feature is generally higher than that of other features, such as DIPAF, SPF, DPF, PNF and PSF. Because MEFF takes into account the information of multiple elements, including source IP address, destination IP address, source port, destination port, packets size and number of packets, it has a higher detection rate than those features that only consider a single aspect. Fig. 1 shows that the detection rate of features increases with the increase of the number of samples in the test set. Based on the CNN model, the detection rate grows rapidly at the beginning and slowly at the later stage.

Table 1: Comparison of different test set samples based on CNN in MR and ER

Feature	Evaluation	Number of Test Set				
	Indexes	500	1000	2000	5000	10000
MEFF	MR (%)	30	10.2	8	4.04	1.74
	ER (%)	15	5.1	4	2.02	0.87
SIPAF	MR (%)	30	10	7.9	4	1.68
	ER (%)	15	5	3.95	2	0.84
DIPAF	MR (%)	30	15.8	11.1	6.76	4.93
	ER (%)	15	7.9	5.55	3.38	2.47
SPF	MR (%)	30	17.8	11.1	6.92	6.15
	ER (%)	15	8.9	5.55	3.46	3.08
DPF	MR (%)	30	11.2	8.8	5.8	2.88
	ER (%)	15	5.6	4.4	2.9	1.44
PNF	MR (%)	30	15	9.2	6.5	3.76
	ER (%)	15	7.5	4.6	3.25	1.88
PSF	MR (%)	30.4	18	12.5	8.48	6.42
	ER (%)	15.4	9.2	6.35	4.28	3.85

Tab. 1 shows the performance of missing rate and total error rate of MEFF, SIPAF, DIPAF, SPF, DPF, PNF and PSF with different samples of test set. When the number of samples is 500, the missing rate and total error rate of each feature are basically the same, indicating that in the case of small samples, the performance of missing rate and total error rate of each feature is similar. However, with the increase of the number of samples, the missing rate and total error rate of different features are obviously different. When the number of samples is 1000, MEFF feature and SIPAF feature maintain lower missing rate and lower total error rate, and even reduce 20% compared with the sample size of 500. On the contrary, other features have higher missing rate and total error rate, especially PSF feature whose missing rate is 18%. When the sample size is 2000, the missing rate of SIPAF is the lowest among these features, only 7.9%, while that of MEFF is 8%, which means it is not much different between them. When the number of samples is 2000, DIPAF feature and SPF feature have the same missing rate and total error rate which are 11.1% and 5.55% respectively. When the sample size is 5000, the missing rate of MEFF

feature is 4.04%, the total error rate of it is 2.02%, the missing rate of SIPAF feature is 4%, and the total error rate of it is 2%. However, other features have a higher missing rate and total error rate. When the sample size is 10000, the performance of missing rate and total error rate of MEFF feature and SIPAF feature is still not much different, but the performance of MEFF feature is much better than that of the other five features. In terms of missing rate, MEFF feature is 3.19% lower than DIPAF feature, 4.41% lower than SPF feature, 1.14% lower than DPF feature, 2.02% lower than PNF feature, and 4.68% lower than PSF feature. As the number of samples increases, the performance of missing rate and the total error rate of MEFF feature are getting better and better, while the performance of missing rate and the total error rate of PSF feature are getting worse than other features. According to the experimental results, it can be seen that under the circumstances of different sample sizes, the missing rate and total error rate of MEFF feature proposed in this paper perform better than most features. It is because that MEFF feature considers the information of many aspects rather than those features that only consider the information of a single element.

5.2.2 Comparison of the time and memory of MEFF and other features based on CNN

In this experiment, we choose the training set and test set with a fixed sample to comprehensively consider information, including source IP address, destination IP address, source port, destination port, the type of packets size, and number of packets. Also, this experiment compares the performance of MEFF feature and other six features from the aspects of running time (RT) and memory usage (MU) when detecting whether a sample is a DDoS attack or not based on CNN model.

Table 2: Comparison of MEFF and other six features based on CNN in RT and MU

Evaluation Indexes	MEFF	The sum of the six features
RT (S)	23.54	146.27
MU (MB)	33.84	225.74

It can be seen from Tab. 2, without changing the number of training set and test set, MEFF feature and the other six features have a large gap in the performance of running time and memory usage. In terms of running time, MEFF feature takes 23.54 seconds, while the sum of other six features take 146.27 seconds. As for memory usage, MEFF feature uses 33.84 MB and the sum of other six features uses 225.74 MB. From the experimental results, we can see that the total running time and total memory usage of MEFF feature is much lower than that of the sum of other six features.

5.2.3 Comparison of the performance of MEFF and other features based on SVM

In order to verify that the information fusion method proposed in this paper is not only applicable to the detection model based on CNN model proposed in this paper, but also applicable to other models. The comparison experiments are conducted based on Support Vector Machine (SVM) model.

SVM is a supervised learning model in the field of machine learning, which is usually used for pattern recognition, classification and regression analysis. When dealing with small sample, SVM can get a better performance than other models, and its

generalization performance is higher. This paper selects C-SVM model and the kernel function is radial basis function. In this experiment, we set the parameter c is 1 and g is 0.1 based on matlab platform.

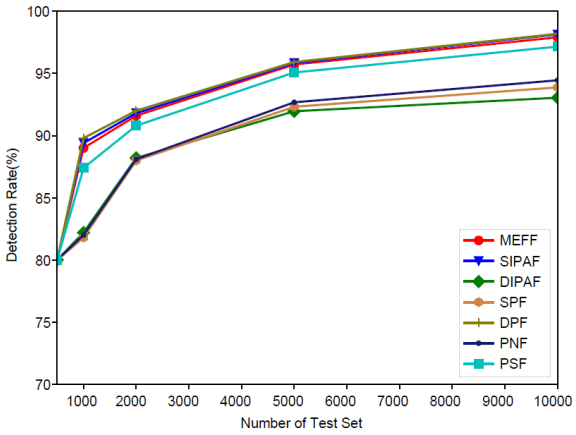

Figure 2: Comparison of detection rate in different test set samples based on SVM

From Fig. 2, based on SVM model, MEFF, SIPAF, DIPAF, SPF, DPF, PNF and PSF still have high detection rates. When the sample size of test set is 500, the detection rate of each feature is 80%. When the sample size is 1000, the detection rates of MEFF feature and SIPAF feature are around 90%, while that of PNF, SPF and DIPAF feature are lower than 85%, indicating that the detection effect of MEFF feature and SIPAF feature is significantly better than that of them. When the number of test set is 2000, the detection rates of PNF, SPF and DIPAF are significantly higher than that of other features, and their detection rates are increased by 6%. When the sample size is more than 2000, the detection rate of each feature grows slowly and becomes steadily. However, it can be clearly seen that MEFF and SIPAF have the same detection rate and always maintain a high detection rate. In particular, when the sample size is larger, for example, when the sample size is 10000, the detection rates of MEFF, SIPAF and PSF are more than 95%. By comparing PSF feature based on CNN model and SVM model, we can find that PSF feature are more applicable to SVM model, because based on CNN model, the detection rate of PSF feature is the lowest among all features, but based on SVM model, the detection rate is at a higher level. The MEFF feature proposed in this paper can maintain a high detection rate based on CNN model and SVM model. It can be seen that MEFF feature proposed in this paper can effectively fuse the information of multiple elements and detect DDoS attacks correctly.

Table 3: Comparison of different test set samples based on SVM in MR and ER

Feature	Evaluation Indexes	Number of Test Set				
		500	1000	2000	5000	10000
MEFF	MR (%)	20	11	8.4	4.28	2.09
	ER (%)	10	5.5	4.2	2.14	1.25
SIPAF	MR (%)	20	10.6	8.2	4.16	1.86
	ER (%)	10	5.3	4.1	2.08	1.11

DIPAF	MR (%)	20	17.8	11.8	8.04	6.95
	ER (%)	10	8.9	5.9	4.02	4.15
SPF	MR (%)	20	18.2	12	7.68	6.12
	ER (%)	10	9.1	6	3.84	3.06
DPF	MR (%)	20	10.2	8	4.08	1.81
	ER (%)	10	5.1	4	2.04	0.91
PNF	MR (%)	20	8	11.9	7.32	5.55
	ER (%)	10	4	5.95	3.66	2.78
PSF	MR (%)	20	12.6	9.2	4.92	2.84
	ER (%)	10	7.1	5.05	2.76	1.98

As can be seen from Tab. 3, with the increase of the number of test set, the missing rate and the total error rate of each feature show a declining trend. When the sample size is 500, the missing rate of each feature is 20%, and the total error rate is 10%. It can be seen that in the case of small samples, the missing rate and total error rate of every feature are relatively high. When sample size of test set is 1000, the missing rate and total error rate of SPF feature are the highest among all features, they are 18.2% and 9.1% respectively. When the sample size is 2000, MEFF, SIPAF and DPF feature maintain relatively low missing rate and total error rate, they are around 8%. When the sample size is 10000, the total error rate of MEFF feature is 1.25%, the total error rate of SIPAF feature is 1.11%, and the maximum total error rate among these features is DIPAF feature, it is 4.15%. It can be seen that with the increase of sample size, the missing rate and the total error rate of each feature are becoming lower and lower, which indicates that these features can detect DDoS attacks better. However, the missing rate and total error rate of MEFF feature are generally lower than that of most features, which means that MEFF feature can effectively fuse the information of multiple elements.

Table 4: Comparison of MEFF and other six features based on SVM in RT and MU

Evaluation Indexes	MEFF	The sum of the six features
RT (S)	9.6	101.44
MU (MB)	19.64	125.57

It can be seen from Tab. 4 that the running time and memory usage of MEFF are significantly lower than that of other six features when the samples are unchanged. From Tab. 4, the running time of MEFF feature is only 9.6 seconds while the running time of the sum of other six features needs 101.44 seconds based on SVM model. At the same time, the memory usage of MEFF feature is small and it is much less than that of the sum of six multiple elements features. The memory usage of MEFF feature is 19.64 MB, and that of the sum of other six features is 125.57 MB. From the experimental results, it can be seen that MEFF feature performs better in terms of running time and memory usage. MEFF feature can consider multi-element information and take the shortest running time and the least memory usage.

5.2.4 Comparison of the performance of MEFF and NWMEFF based on CNN

In order to verify the validity of weights in MEFF feature. In this paper, we conduct a comparative experiment to compare the accuracy of each batch of MEFF feature and no

weight MEFF feature (NWMEFF) in the training process based on the CNN model, and compare the detection rate, missing rate, error rate of MEFF and NWMEFF during testing.

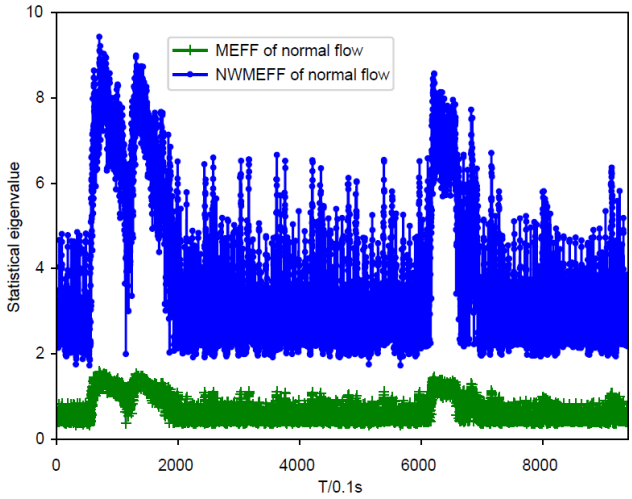

Figure 3: Statistical eigenvalue of MEFF and NWMEFF in normal flow

It can be seen from Fig. 3 that the eigenvalue of MEFF is relatively concentrated than that of NWMEFF. Since MEFF feature takes into account the importance of every element feature and measures each feature by weight, the eigenvalue of MEFF is relatively stable and does not fluctuate much. On the contrary, the eigenvalue of NWMEFF fluctuates greatly, the maximum value is more than 10 and the minimum value is less than 2. From Fig. 3, we can see that in the 500th sampling time, the 1500th sampling time and the 6000th sampling time, the value of NWMEFF is at the peak of network access, and it is likely to be misjudged as an attack. However, the value of MEFF is relatively stable, so misjudgment will not occur.

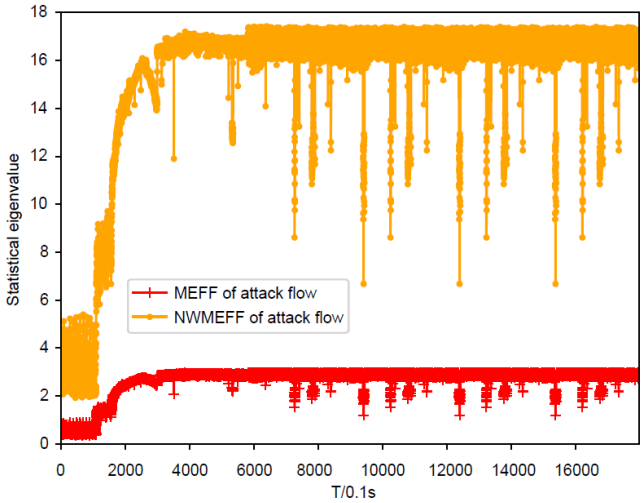

Figure 4: Statistical eigenvalue of MEFF and NWMEFF in attack flow

As can be seen from Fig. 4, at the beginning of attack, the eigenvalue of NWMEFF fluctuates greatly from 2 to 16. There are several fluctuations during the middle and late stages of the attack. Therefore, the attack flow may be misjudged as normal flow based on NWMEFF feature. However, MEFF does not fluctuate much in the early and late stages of an attack, so the possibility of misjudgment is much lower.

Fig. 5 shows the comparison of accuracy of MEFF and NWMEFF during training. At the beginning of training, the accuracy of NWMEFF is higher than that of MEFF feature. From the batch of 20th, the accuracy of MEFF is higher than that of NWMEFF feature, the accuracy of MEFF feature is around 80%. In the batch of 40th, the accuracy of MEFF is nearly up to 90%. It can be seen from Fig. 5 that the accuracy of MEFF is basically higher than that of NWMEFF during training. In the case of normal flow and attack flow, the eigenvalue of NWMEFF fluctuates greatly, which makes it impossible to accurately express the network situation at that time and prone to wrong judgment. However, the eigenvalue of MEFF is relatively stable and the possibility of wrong judgment is small. Therefore, in the whole training process, the accuracy of MEFF is relatively higher than that of NWMEFF.

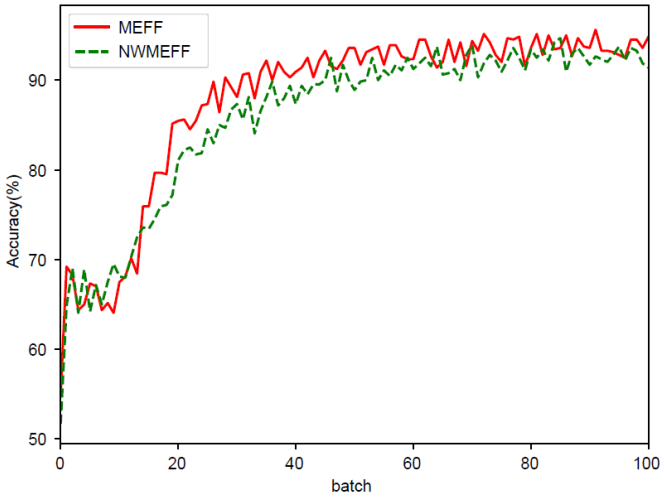

Figure 5: Comparison of accuracy of MEFF and NWMEFF during training

Table 5: Comparison of MEFF and NWMEFF in DR, FR, MR and ER

Feature	Evaluation Indexes		
	DR (%)	MR (%)	ER (%)
MEFF	93.76	6.24	3.70
NWMEFF	89.93	10.07	6.61

It can be seen from Tab. 5 that the detection rate of MEFF feature is more than that of NWMEFF feature. The missing rate of MEFF is 6.24%, and the missing rate of NWMEFF is 10.07%. The error rate of MEFF is much less than that of NWMEFF, the error rate of MEFF feature is 3.70%. The reason why the performances of MEFF feature are better than those of NWMEFF feature is that the weight of each feature is different. By increasing and decreasing the weight of every feature, MEFF can more accurately express the current network flow.

According to the experiments above, it can be seen that MEFF feature has a high detection rate, low missing rate and total error rate based on both CNN model and SVM model. Detecting DDoS attacks through MEFF feature runs faster and uses less memory. DPF feature has a high detection rate under the SVM model, but a low detection rate under the CNN model, indicating that DPF feature is only applicable to the SVM model. According to experiments, it can be found that the detection rate of MEFF feature is generally higher than that of other features, and the running time and memory usage are lower when other features' information is taken into account. Furthermore, the weight of MEFF can effectively measure the importance of each element feature, effectively fuse features, and the accuracy is high. To sum up, the information fusion method proposed in this paper can effectively fuse the information of multi-element features and consider the information of multiple element features at the same time, which cause the detection rate is high, missing rate is low and total error rate is low. Furthermore, this method is not only applicable to the CNN detection model proposed in this paper, but also applicable to other models.

6 Conclusion

In the context of Big Data, DDoS attack is of diversity, abruptness and high traffic. Many methods based on single-element leads to large resources consumption, high missing rate, and low detection rate. This paper proposes a DDoS attack information fusion method based on CNN for multi-element data, which fuses the information of multi-element data. By using the information fusion classification model proposed in this paper to detect DDoS attack, it is found that the information fusion method has a fast processing speed, low memory consumption, high detection rate, low missing rate and total error rate.

A possible goal for our future research would be to consider multiple source heterogeneous data in the field of information fusion.

Acknowledgement: This work was supported by the Hainan Provincial Natural Science Foundation of China [2018CXTD333, 617048]; National Natural Science Foundation of China [61762033, 61702539]; Hainan University Doctor Start Fund Project [kyqd1328]; Hainan University Youth Fund Project [qnjj1444].

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

Arbor Networks (2018): 13th worldwide infrastructure security report. https://pages.arbornetworks.com/rs/082-KNA-087/images/13th Worldwide Infrastructure Security Report.pdf.

Antonakakis, M.; April, T.; Bailey, M.; Bernhard, M.; Bursztein, E. et al. (2017): Understanding the Mirai botnet. *USENIX Security Symposium*, pp. 1093-1110.

Cheng, J.; Tang, X.; Yin, J. (2017): A change-point DDoS attack detection method based on half interaction anomaly degree. *International Journal of Autonomous and Adaptive Communications Systems*, vol. 10, no. 1, pp. 38-54.

Cheng, J.; Xu, R.; Tang, X.; Sheng, V. S.; Cai, C. (2018): An abnormal network flow feature sequence prediction approach for DDoS attacks detection in big data environment. *Computers, Materials and Continua*, vol. 55, no. 1, pp. 95-119.

Cheng, J.; Zhang, C.; T, X.; Sheng, V. S.; Dong, Z. et al. (2018): Adaptive DDoS attack detection method based on multiple-kernel learning. *Security and Communication Networks*, vol. 2018.

Cheng, J.; Zhou, J.; Liu, Q.; Tang, X.; Guo, Y. (2018): A DDoS detection method for socially aware networking based on forecasting fusion feature sequence. *Computer Journal*, vol. 61, no. 7, pp. 959-970.

China Internet Network Information Center (2018): Statistical report on the development of China's Internet. http://www.cac.gov.cn/2018-08/20/c1123296882.htm.

Costa, P. C.; Yu, B.; Atiahetchi, M.; Myers, D. (2018): High-level information fusion of cyber-security expert knowledge and experimental data. *International Conference on Information Fusion*, pp. 2322-2329.

Esposito, C.; Castiglione, A.; Palmieri, F.; Ficco, M.; Dobre, C. et al. (2018): Event-based sensor data exchange and fusion in the Internet of things environments. *Journal of Parallel and Distributed Computing*, vol. 118, no. 2, pp. 328-343.

Golestan, K.; Khaleghi, B.; Karray, F.; Kamel, M. S. (2016): Attention assist: a high-level information fusion framework for situation and threat assessment in vehicular ad hoc networks. *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 5, pp. 1271-1285.

Guo, Y.; Yin, C.; Li, M.; Ren, X.; Liu, P. (2018): Mobile e-commerce recommendation system based on multi-source information fusion for sustainable e-business. *Sustainability*, vol. 10, no. 1, pp. 147.

Idhammad, M.; Afdel, K.; Belouch, M. (2018): Semi-supervised machine learning approach for DDoS detection. *Applied Intelligence*, vol. 48, no. 10, pp. 3193-3208.

Li, X.; Lu, Y.; Liu, S.; Nie, W. (2018): Network security situation assessment method based on markov game model. *KSII Transactions on Internet and Information Systems*, vol. 12, no. 5, pp. 2414-2428.

Lian, J.; Zhang, F.; Xie, X.; Sun, G. (2018): Towards better representation learning for personalized news recommendation: a multi-channel deep fusion approach. *International Joint Conference on Artificial Intelligence*, pp. 3805-3811.

Lin, L. (2016): Multi-sensor information fusion method based on BP neural network. *International Journal of Online Engineering*, vol. 12, no. 5, pp. 53-57.

Liu, J.; Pan, H.; Zhang, J.; Zhang, Q.; Zheng, Q. (2017): Detecting bogus messages in vehicular ad-hoc networks: an information fusion approach. *China Conference on Wireless Sensor Networks*, vol. 812, pp. 191-200.

Liu, S.; Cai, Z.; Xu, H.; Xu, M. (2015): Towards security-aware virtual network embedding. *Computer Networks*, vol. 91, pp. 151-163.

MüUller, W.; Kuwertz, A.; Mühlenberg, D.; Sander, J. (2017): Semantic information fusion to enhance situational awareness in surveillance scenarios. *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, pp. 397-402.

Paggi, H.; Soriano, J.; Lara, J. A. (2018): A multi-agent system for minimizing information indeterminacy within information fusion scenarios in peer-to-peer networks with limited resources. *Information Sciences*, vol. 451-452, pp. 271-294.

Rogova, G. L.; Snidaro, L. (2018): Considerations of context and quality in information fusion. *21st International Conference on Information Fusion*, pp. 1925-1932.

Sahoo, K. S.; Puthal, D.; Tiwary, M.; Rodrigues, J. J. P. C.; Sahoo, B. et al. (2018): An early detection of low rate DDoS attack to SDN based data center networks using information distance metrics. *Future Generation Computer Systems*, vol. 89, pp. 685-697.

Snidaro, L.; García, J.; Llinas, J. (2015): Context-based information fusion: a survey and discussion. *Information Fusion*, vol. 25, pp. 16-31.

Tuor, A.; Kaplan, S.; Hutchinson, B.; Nichols, N.; Robinson, S. (2017): Deep learning for unsupervised insider threat detection in structured cybersecurity data streams. *Association for the Advance of Artificial Intelligence*, pp. 1-9.

Wu, H.; Wang, Z. (2018): Multi-source fusion-based security detection method for heterogeneous networks. *Computers & Security*, vol. 74, pp. 55-70.

Yuan, J.; Li, X. (2018): A reliable and lightweight trust computing mechanism for IoT edge devices based on multi-source feedback information fusion. *IEEE Access*, vol. 6, pp. 23626-23638.