Computers, Materials & Continua CMC, vol.60, no.3, pp.1263-1281, 2019

A DDoS Attack Situation Assessment Method via Optimized Cloud Model Based on Influence Function

Xiangyan Tang[1], Qidong Zheng[1, *], Jieren Cheng[1, 2] , Victor S. Sheng[3], Rui Cao[1] and Meizhu Chen[1]

Abstract: The existing network security situation assessment methods cannot effectively assess the Distributed denial-of-service (DDoS) attack situation. In order to solve these problems, we propose a DDoS attack situation assessment method via optimized cloud model based on influence function. Firstly, according to the state change characteristics of the IP addresses which are accessed by new and old user respectively, this paper defines a fusion feature value. Then, based on this value, we establish a V-Support Vector Machines (V-SVM) classification model to analyze network flow for identifying DDoS attacks. Secondly, according to the change of new and old IP addresses, we propose three evaluation indexes. Furthermore, we propose index weight calculation algorithm to measure the importance of different indexes. According to the fusion index, which is optimized by the weighted algorithm, we define the Risk Degree (RD) and calculate the RD value of each network node. Then we obtain the situation information of the whole network according to the RD values, which are from each network nodes with different weights. Finally, the whole situation information is classified via cloud model to quantitatively assess the DDoS attack situation. The experimental results show that our method can not only improve the detection rate and reduce the missing rate of DDoS attacks, but also access the DDoS attack situation effectively. This method is more accurate and flexible than the existing methods.

Keywords: DDoS attack, V-SVM, influence function, cloud model.

1 Introduction

With the rapid rise of industries such as the Internet of Things, cloud computing, and big data, the global informationization process has been greatly promoted. These new technologies have brought innovation vitality to economic and social development, and at the same time, traditional network security protection faces serious threats and new challenges. Among the various types of security threats, the danger of DDoS attacks is especially serious. By controlling multiple puppet machines in different locations on the network to attack the victim simultaneously, the attacker uses asymmetry of resources to

[1] Key Laboratory of Internet Information Retrieval of Hainan Province, Hainan Universty, Haikou, 570228, China.

[2] College of Information Science & Technology, Hainan University, Haikou, 570228, China.

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

[*] Corresponding Author: Qidong Zheng. Email: qidong@hainu.edu.cn.

produce plenty of attack flow, causing abnormal service, thereby preventing legitimate users from accessing the network normally [Idhammad, Afdel and Belouch (2018)]. According to the data provided by Tencent Cloud in 2018, the large-flow DDoS attacks in the first half of 2018 continued to increase [Teng, Liu, Liu et al. (2019)]. Among them, Memcached DDoS has sprung up everywhere, with a reflection magnification of up to 50,000 and a peak of up to 1.7 Tbps of attack traffic, which has become a new focus of the network security community. As one of the most important security threats the Internet faces, frequent DDoS attacks which leaded to major economic losses in the whole society. However, traditional single defense devices or detection devices are unable to meet security requirements. Different security factors can be integrated by the network security situation evaluation technology, dynamically reflect the network security status as a whole, and provide a reliable reference for enhancing network security. Therefore, the security situation assessment model and key technologies for the network have become the research hotspots at field of network security.

The rest of the paper is organized as follows: the second part is the related work of network security situation assessment at home and abroad in recent years; the third part introduces the extraction of network flow attack features and the identification of attacks; the fourth part is the establishment of the assessment indexes and the optimization by the influence function; the fifth part is the assessment method via cloud model; the sixth part is the specific experiment and the analysis of experimental results. The seventh part is the conclusion of this article, and the last part is the reference.

2 Related works

Endsley proposed the concept of Situation Assessment (SA) firstly [Endsley (1988)]. Then, Tim Bass applied the concept of SA at field of network security for the first time, and proposed the concept of network security situation assessment [Bass (2000)].

In recent years, researchers from different countries have made some progress in this field and have obtained some research results. Zhao et al. [Zhao and Liu (2018)] used wavelet neural network algorithm which is based on particle swarm method, and obtained situational awareness context value in big data environment. The algorithm had fast convergence and good adaptability. Pu [Pu (2018)] used dynamic Bayesian networks to evaluate the attack effects of network nodes, whose method was more accurate and efficient than traditional node importance assessment methods. Xu et al. [Xu, Cao and Ren (2017)] proposed an IoT network security situation assessment model which is based on semantic ontology. The perception of this method was more comprehensive, and the ability was more powerful than the traditional network security situational awareness method. Liu et al. [Liu, Dong and Lv (2017)] proposed the confidence-based situation awareness method, which uses integrated learning algorithm and D-S evidence theory to calculate confidence of attack and combines various situational elements to obtain situation information.

Lounis [Lounis (2018)] proposed a stochastic security evaluation method which used a tree model to establish related security scenarios, then used a stochastic model to evaluate security situation. Alali et al. [Alali, Almogren, Hassan et al. (2018)] used the fuzzy reasoning method to generate security assessment results that were based on four assessment indexes and conducted various analyses on these factors, the model could better

assess system risk. Jana et al. [Jana and Ghosh (2018)] proposed a novel interval controller of fuzzy logic which improved the evaluation model about network risk, and improved the possibility of predicting network security risk assessment. Fan et al. [Fan, Xiao and Nayak (2017)] established a risk situation evaluation model through Software Defined Network (SDN), analyzed the features of the common network attacks and Address Resolution Protocol (ARP) attack, and calculated the risk values of SDN via Hidden Markov Model (HMM). The network status can also be predicted by the proposed approach based on HMM. Dai et al. [Dai, Hu and Zheng (2017)] proposed a risk flow attack graph to construct a network assessment model to assess network risk whose method is capable of identifying network security situations effectively. Liu et al. [Liu, Zhang, Zhu et al. (2017)] put forward a hierarchical network menace situation evaluation method that was via D-S evidence method to assess the influence of DDoS attacks on the network, this method divided four layers, collection, extraction, fusion and assessment, DDoS attacks effect can be better reflected by the proposed method on the network security menace situation.

Liu et al. [Liu, Wang and Lu (2016)] proposed a fusion-based control model of evaluation in network security situation, which provided new methods and means for monitoring and managing networks. Wu et al. [Wu, Liu and Xu (2016)] used the complex network theory to model the Security Situation Awareness (SSA) data, which could mine current data's features more effectively. Li et al. [Li and Zhao (2016)] proposed a method to obtain observations value through sliding time window, then through Multiple Population Genetic Algorithm (MPGA) for obtaining HMM parameters, thus improved the reliability of parameters. Zhang et al. [Zhang, Cheng, Tang et al. (2018)] proposed one algorithm to evaluate situation of DDoS, which improved Fuzzy C-means clustering algorithm. Sara et al. [Sara, Jose and Skarmeta (2018)] to solve problems of large-scale Internet of Things deployment, proposed a security certification method which was designed for IoT,this method could evaluate the security grade of a device automatically. Jordan et al. [Jordan and Niall (2018)] used Real-Time Dynamic Network Anomaly (ReTiNA) to combine anomalies on each edge of the network graph and then performed situational awareness across the whole network. Zhong et al. [Zhong, Lin, Liu et al. (2018)] proposed the data triage working retrieval system which can achieved cyber situational evaluation in Security Operations Center (SOC). This system can provide on-the-job advice for novice analysts. Yusuf et al. [Yusuf, Ge, Hong et al. (2018)] used a Temporal Hierarchical Attack Representation method to assess security index, when the status of systems changes, the indexes also changes. The results provide some insights about security indicators under network changes.

Zahid-Hasan et al. [Zahid-Hasan, Zubair-Hasan and Sattar (2018)] used the Deep Convolution Neural Network for detecting DoS attack, this model is more effective than K-Nearest Neighbor (KNN). Cheng et al. [Cheng, Xu, Tang et al. (2018)] aim to deal with DDoS detection problems in big data environment, put an abnormal network flow characteristic prediction method forward, this method can identify DDoS attack better and solve the problem of compute resources consumption. Kshira et al. [Kshira, Deepak, Mayank et al. (2018)] applied information distance as a measurement in DDoS detecting which based on logically centralized controller,this method improved the detection accuracy. Khundrakpam et al. [Khundrakpam and Tanmay (2017)] applied for multi-layer perceptron and inherited algorithm in detecting DDoS attacks in application layer analysis.

This method provided the minimum error compared to others traditional methods. Cheng et al. [Cheng, Zhou, Liu et al. (2018)] used the method of multi-protocol-fusion to define usual network flows and used Autoregressive Integrated Moving Average (ARIMA) to detect DDoS attack. Sabrina et al. [Sabrina, Alessandra, Daniele et al. (2018)] proposed a risk evaluation technology to assess dynamic and static components of an IoT system, this method was based on end-to-end systems. Alireza et al. [Alireza, Saygin, HungDa et al. (2018)] presented a method to use the game theory method to solve network security problems, this method could be applied for other area. Jin et al. [Jin, Simon, Dong et al. (2018)] put forward a different series of indexs to assess the Moving Target Defense technology, these indexs could offer direction in choosing the best effective Moving Target Defense (MTD) technology.

3 A DDoS attack detection based on V-SVM

3.1 Features extracting

Given a normal network flow U with n samples and a network flow V with m samples which need to be detected, this paper define each sample as (T_i, S_i, D_i), where T_i represents the arrival time of packet i, S_i denote its source IP and D_i represents the destination IP. This paper uses the normal network flow U to train a model which identifies the network flow V, in the process of training and detection, the parameter Δt is the same. After the above definition, during the training process, at the end of the k-th Δt, we will take out the IP packet of the network flow sample corresponding to the whole period from the normal network flow U, and define it as G_k. And for each IP packet in G_k , if the IP address contained in the IP packet is invalid, the IP packet will be thrown away from G_k. We will define the sample group after filtering as F_k.

As getting each F_k, we gradually establish a set of IP addresses O, which represents the old users of the current network. In the first Δt time period, we merge all source IP addresses S in F_1 into the IP addresses O, and the maximum number of old users is $O_{max} = ||\{S|S \in F_1\}||$. Subsequently, the number of old users of F_k calculated for the k-th (k>1) time is $F_k \cap O$, then we update the maximum number of old users $O_{max} = max(O_{max}, ||F_k \cap O||)$. Finally, after calculating the O_{max}, we put each source IP address in F_k into IP addresses set O. By repeating this operation, we can get a maximum value that represents the old user in a specific time period. While updating of O_{max} , we also need to calculate the number of new users in this time period $N_k = ||F_k| - ||F_k \cap O||$. Since we already know that the set $F_k \cap O$ represents old user, and the set $F_k \setminus O$ denotes new user at this time interval, N_k denotes the number of new users. In the same way, we can also calculate the average number of N in per Δt period as $\overline{N} = \sum_{i=1}^{k} N_i / k$.

After above calculation, we have got four model parameters, the network flow detection operation could be started. We read V with Δt as above, in every sample interval, we define the dictionary $W_k[S_{k,i}]$ denotes the number of visitor of source IP $S_{k,i}$ during each k-th time period. When every Δt be finished, we could get four parameters in k-th time period. And then they are defined as below.

$$\begin{cases} B_k = \dfrac{\left||G_k \cap O|\right| - O_{max}}{O_{max}} \\ M_k = \left|\left|G_k \backslash O\right|\right| - \bar{N} \\ L_k = \dfrac{||G_k \backslash O||}{O_{max}} \\ D_k = \dfrac{\sum \{W_k[S_{k,j}] | \forall_{S_{k,j}} \in (G_k \backslash O)\}}{||G_k \backslash O|| \Delta t} \end{cases} \tag{1}$$

In Eq. (1), B_k denotes a ratio of old IPs that appeared at present time interval k over the maximum number of old IPs in certain k intervals. M_k represents the changes in the number of new users in the current time interval k relative with number of average new IPs. L_k denotes a percentage of the new user to the old user in current time interval k. D_K denotes accessing ratio of new user, more specifically, the feature value refers to the number of visitors of per new user in per second during the k time period.

From actual experience, our old users will be accessed in a relatively fixed mode when the network is normal. This access mode can be used as a priori condition for DDoS attack identification. Assume that O_{max} = 400 during training process, then if $\left||G_k \cap O|\right|$=200 old users in the detection flow, and the result of the calculation is $B=\frac{200-400}{400}$=-0.5=-50%. It could represent the maximum value of O_{max} compared to the old users, the percentage of current old users is increased or decreased.

Based on the definitions of DDoS, it is assumed that there should be many new users in the network flow under the DDoS attack. So $||G_k \backslash O||$ is assumed more bigger than the number of new IPs under general conditions, the feature value M is a change amount indicating the average value of new IPs learned in the current time period relative to the new user learned in the training set.

Supposing that attack is taking place, there is a reason to speculate current number of new IPs $||G_k \backslash O||$ is bigger than current number of old IPs. But there may be only a few old IPs and little new IPs appear. At this condition, supposing new IPs more than old IPs, the $\frac{||G_k \backslash O||}{||O||}$ will cause fluctuation. Therefore, the ratio of new IPs currently found in maximum number of the old IPs in the training is used as L. Based on M and L, if a DDoS attack occurs, the accessing ratio of the new IPs will be a great value because the network will receive many flooded packets from the attacker. In the normal flow or network congestion, because each normal user will abide by the TCP/IP protocol, the access rate D will be a relatively small constant value.

Due to the way the eigenvalues are calculated, it is important to avoid any eigenvalues be 0. Therefore, the calculation methods of B_k and L_k are improved.

$$B_k = \begin{cases} \dfrac{||G_k \cap O|| - O_{max}}{O_{max}}, & if\ ||G_k \cap O|| - O_{max} \neq 0 \\ \dfrac{||G_k \cap O|| - O_{max} - 1}{O_{max}}, & others \end{cases} \tag{2}$$

$$L_k=\begin{cases}\frac{\|G_k\backslash O\|}{O_{max}}, & if\ \|G_k\backslash O\|\neq 0\\ \frac{\|G_k\backslash O\|-1}{O_{max}}, & others\end{cases} \tag{3}$$

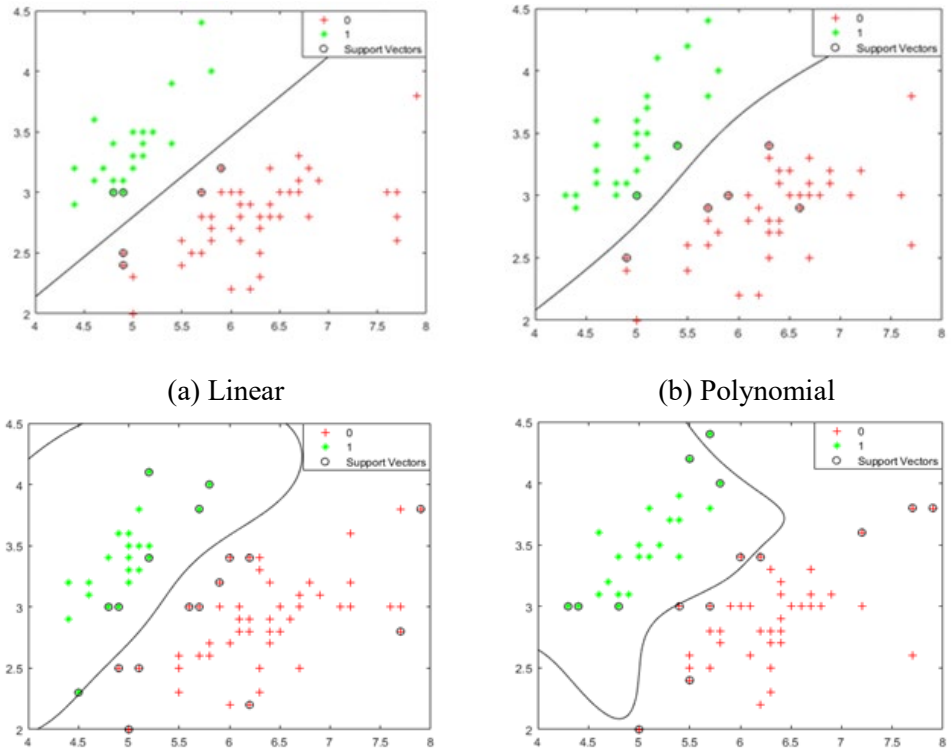
Within B_k,M_k,L_k and D_k have defined above, the product of them is calculated by us firstly, we then take the product's negative value which is defined as $BMLD_k$.

3.2 Attack recognition based on V-SVM

The support vector machine approach is based on statistical learning's VC dimension theory and the rule of morphological risk minimization. According to message of finite samples, it looks for the most excellent compromise between the model's complexity and the learning capability to acquire the best promotion capability. Compared with other machine learning algorithms such as neural network, decision tree and Adaboost, SVM classifier has simpler structure design, moderate computational complexity and better generalization performance. It shows a lot of great pluses in dealing with small sample, high and nonlinear dimensional pattern recognition. Mu et al. [Mu and Nandi (2005)] proposed a V-SVM that added a parameter V that could manage support vectors and error vectors in numbers. So as to show value of the parameter V and enhance the detection ratio of DDoS attacks and decrease false negative ratio, this paper used V-SVM method to detect DDoS attack.

Most of the classification problems existing in real life are nonlinear. The support vector machine method is to choose a kernel function for nonlinear expansion. According to the actual problem and the research object, choosing the appropriate kernel function often plays a decisive role in the classification effect of the model. Under the conditions of Linear, Polynomial, Radial Basis Function, and Sigmoid four different types of kernel functions, the first two dimensions of the partial training set are selected for visualization. The red and green marks represent the normal sample points and the abnormal sample points respectively, and the black marks indicate the support vectors. As shown in Fig. 1, the impact of the kernel function on the decision hyperplane position is visually reflected. Compared to other kernel functions, this paper chose RBF kernel function, because the RBF kernel function has the highest classification accuracy and the lowest false negative rate.

Although the parameter V controls the number of support vectors and error vectors very well, however, due to the difference in the specific objects of training and the different degrees of influence of each feature on the classification results, the expected detection results may not be achieved. At present, there is no uniform set of best methods for the selection of V values at home and abroad. Based on the specific data, this paper analyzed the impact of artificial experience value V on DDoS attack detection, thus selected the appropriate V value.



(a) Linear (b) Polynomial

(c) RBF (d) Sigmoid

Figure 1: Schematic diagram of V-SVM classification under different types of kernel functions

The detection method is as shown in Fig. 2 below. Firstly, the data set is analyzed, the appropriate features are selected, and the training set is normalized and reduced. The second step is to construct various SVM classification models, and study the process of classification model parameters and kernel function parameters optimization. The third step is to initialize the training model, import the training set to start the training of the model, the original issue is changed into a convex optimization issue, then the lagrangian function is constructed, the original problem is dualized and the kernel function is selected, the optimal solution α is solved by solving the dual problem through the Sequential Mini Optimization (SMO) algorithm, and then the optimal solutions w and b of the problem are solved. Finally, we obtain hyperplane and classification decision function.

The last step, the constructed classification model is tested by a data set containing unknown tags, we analyzed the experimental results. The SMO algorithm is different from the SVM algorithm, which is a heuristic type algorithm. When the solution of each variable in the algorithm reaches the Karush-Kuhn-Tucker (KKT) requirement for the optimization problem, the solution can be found.

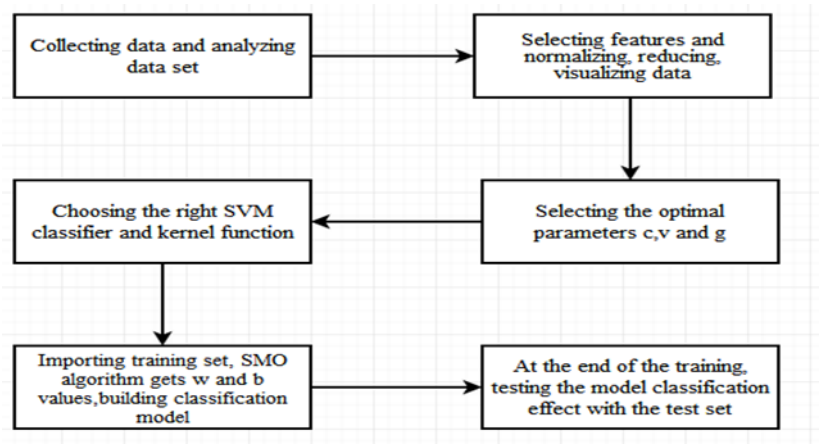


Figure 2: DDoS attack detection process

4 Assessment index optimized by influence function

4.1 Assessment index

Based on the above detection results, this paper began to evaluate the impact of the DDoS attack. Because DDoS attack situations are multiple factors, ambiguity and subjective judgement, this paper quantifies the original qualitative evaluation, and then proposed three indexs as follow.

On the one hand, under normal conditions, the number of new IP that accesses the website is small, because the users who access the website are basically old users. When a DDoS attack occurs, there will be accompanied by the emergence of a large number of new IP, so the rate of the number of new IP at time of attack to the maximum number of new IP at the normal time is a very meaningful value. So this paper propose New IP Ratio (*NIR*), the formula is as follows:

$$NIR = \frac{NI / N_{max}}{\Delta t} \tag{4}$$

where NI indicates the number of addresses for different new IP that change over time when a DDoS attack occurs, and N_{max} indicates the maximum number of new IP under normal conditions. NIR indicates the relative strength of the attack, that is, the degree of change of the attack new IP relative to the normal new IP maximum value relative to the normal conditions. For example, when a large number of new IP appear, it cannot be said that this is an attack, or it may be a network congestion or a hot event. So using relative values can better describe the changes in the new IP when an attack occurs.

In addition, when an attack occurs that will cause denial of service, that is, attacker convey lots of IP packets to the victim, but it does not accept the third handshake, so the network flow with no return can be defined as a One-Way-Flow (OWF). When an attack occurs, the attack strength will increase with time going by, and the IP packets sent by the attacker will gradually increase. So we can define the absolute strength of the attack as follow:

$$SAD=\frac{DFI}{\Delta t} \tag{5}$$

where *DFI* indicates different packets in One-Way-Flow of new IP, *SAD* indicates the service abnormality degree, that is, the number of packets of new IP increases with time, which can be used to measure the absolute strength of DDoS.

Under normal conditions, the number of old IP is generally larger than the new IP. When a DDOS attack occurs, the number of new IP gradually increases, while the number of old IP gradually decreases. To express the dynamics between the new and old IP, we defined the following formula:

$$ANF=\frac{NI}{OD} \tag{6}$$

where *NI* indicates the number of addresses for different new IP that change over time when a DDoS attack occurs, *OD* indicates the dynamic change in the number of old users' IP addresses, *ANF* represents the ratio of addresses of different new IP to addresses of different old IP.

4.2 Influence function

The influence function is an important concept in robust statistics. It describes the degree of influence of a single outlier on the statistic, and reflects the variation of the statistic of the model distribution when it is polluted by the unit. The statistical approach provides a new perspective and provides a simple and flexible tool for robust statistics. For the first time, Koh et al. [Koh and Liang (2017)] applied influence function in machine learning. The learning algorithm tracks the prediction of the model and returns its training data, so that it can find the part of the training data that has the greatest impact on a given prediction. Because the weights of *NIR*, *SAD* and *ANF* are different, we can use influence function to assign weights to them.

The formula that influence function is defined as follows:

$$\begin{aligned} I_{up,loss}(z,z_{test}) &\stackrel{\text{def}}{=} \frac{dL\left(z_{test},\hat{\theta}_{\epsilon,z}\right)}{d\epsilon}|_{\epsilon=0} = \nabla_{\theta}L(z_{test},\hat{\theta})^{\intercal}\frac{d\hat{\theta}_{\epsilon,z}}{d\epsilon}|_{\epsilon=0} \\ &= -\nabla_{\theta}L(z_{test},\hat{\theta})^{\intercal}H_{\hat{\theta}}^{I}\nabla_{\theta}L\left(z,\hat{\theta}\right) \end{aligned} \tag{7}$$

where $H_{\hat{\theta}}\stackrel{\text{def}}{=}\frac{1}{n}\sum_{i=1}^{n}\nabla_{\theta}^{2}L(z_i,\hat{\theta})$ is the Hessian matrix and the hypothesis is positive , $\nabla_{\theta}L(z,\hat{\theta})$ is the gradient, the effect of a single training sample z on the model parameters θ. We then substitute the three indicators into the influence function, and then calculate their influence function values separately, and assign weights to the three indicators by the size of the value. The calculated three indicators are as *NIR'*=*F*(*NIR*), *SAD'*=*F*(*SAD*), *ANF'*=*F*(*ANF*).

5 Assessment method based on cloud model

5.1 Risk Degree

We define the Risk Degree (RD) which means the degree of risk of the system under ddos attack, the formula is as follows:

$$RD=NIR'\times SAD'\times ANF' \tag{8}$$

This value is used to assess the risk level of each network node. Classify the *RD* of each

network node fusion, which is classified into normal, low-risk, medium-risk and high-risk. Then, the four levels of each network node are merged, and then the level is divided into four levels, which indicate that the whole network node status is normal, slightly affected, serious damaged and devastated.

Because the importance of each network node is different, this paper set the weight of each node to $\alpha_1, \alpha_2, \cdots, \alpha_n$. For backbone nodes, the weight will be larger. If the asset of the node is important, then the weight of the node should also be large. Fig. 3 is a topology diagram of network nodes and their corresponding weights.

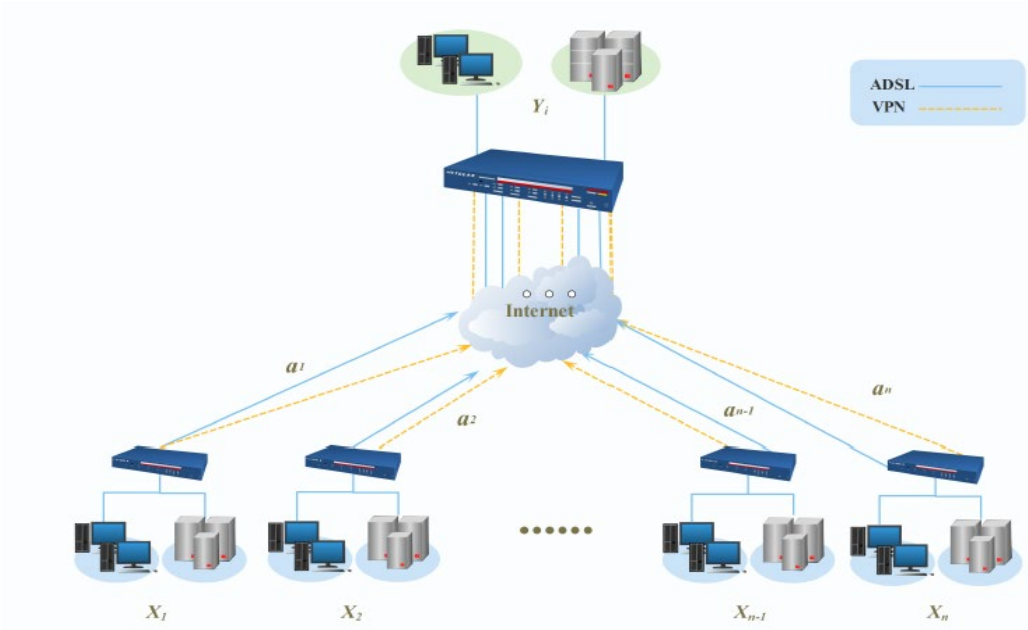


Figure 3: Network node graph

5.2 Assessment system based on cloud model

In our research, the cloud model was used as technical grade classification model. The cloud model was put foward by Li Deyi, he applied this model to the field of intelligent control for the first time [Wang, Peng and Zhang (2016)]. The model is used to handle an unsure conversion in qualitative concepts and quantitative characterization. This model is used in many fields such as data mining, decision analysis, and intelligent control.

Definition 1. Cloud and cloud drop. Suppose that U denote quantitative range of risk in accurate number representation, and C denote the qualitative concept in U, That is, the evaluation of the degree of attack, normal, slightly, serious and destructive, In case $x \in U$ denotes a random implementation of C, and degree of certain $\mu(x) \in [0,1]$ of x for C:μ:$U \rightarrow [0,1]$, $\forall x \in U, x \rightarrow \mu(x)$, and the distribution of x in the domain U is definded as cloud, which was indicated by the cloud $C(X)$. Every x was defined as cloud drop. Assuming that the domain of x is multi-dimensional, it can be extended to n-dimensional clouds accordingly. The calculation method of n-dimensional cloud is the same as that of one-dimensional.

The overall characteristics of cloud model related concepts could be reflected in the numeral features of cloud, which used a three-dimensional feature such as Expected value (Ex), Entropy (En), and Hyper entropy (He) to denote a concept. The overall characteristics of the multi-dimensional cloud model could be represented by multiple sets of digital features.

Ex denotes an expectation value, which is based on the distribution of cloud droplets in the universe, this point that best denotes the qualitative concept, it is also the most prominent sample value for this value quantification; En is a unit of measure, which was used primarily to denotes the uncertainty of qualitative concepts. In addition, this concept changes with entropy, when this entropy becomes larger, and concept is also macroscopic. Which is determined by randomness and ambiguity of this concept, both are related. In addition, En could be seen as the measure of randomness, which reflect a level of probability of cloud drop that could denotes this value; from another perspective, it could be a measure of qualitative concepts, reflecting the range of values that could be accepted with concept space; Last concept is He, which indicates a uncertainty measure of En, that means, the entropy of entropy, this value is determined by two properties of entropy, they are randomness and ambiguity. These overall characteristics of the qualitative concept denoted through the three digital features are indicated as C (Ex, En, He). This paper uses the reverse cloud generator to realize the transformation of the degree of risk to the influence level of a DDoS attack, the formula and algorithm are as follow:

$$E_n=\sqrt{\frac{\pi}{2}}\times\frac{1}{n}\sum_{i=1}^{n}|x_i-\overline{X}| \tag{9}$$

$$H_e=\sqrt{s^2-En^2} \tag{10}$$

Algorithm 1 Assessment

Input: risk degree value

Output: Ex, En, He

1: setup cloud drop N

2: **while** get value x_i **do**

3: get the sample mean of RD $\overline{X}=\frac{1}{n}\sum_{i=1}^{n}x_i$

4: get $\frac{1}{n}\sum_{i=1}^{n}|x_i-\overline{X}|$

5: calculate sample variance $s^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{X})^2$

6: calculate En according to Eq. (9)

7: calculate He according to Eq. (10)

8: **end while**

9: **return** Ex, En, He

6 Experiment

6.1 Experiment data set and evaluation standard

The experiments below were based the dataset CAIDA DDoS attack 2007, the size of dataset is 21 GB. In this data set, normal network flow started at 13:49:36, and abnormal flow attacks occurred at 14:15:56. We introduce three related performance evaluation standards to evaluate the experimental results including Accurate Rate (AR), Detection Rate (DR), Missing Rate (MR). The calculation formula of evaluation standards define as follows:

$$AR=\frac{TP+TN}{TP+TN+FP+FN} \tag{11}$$

$$DR=\frac{TN}{TN+FN} \tag{12}$$

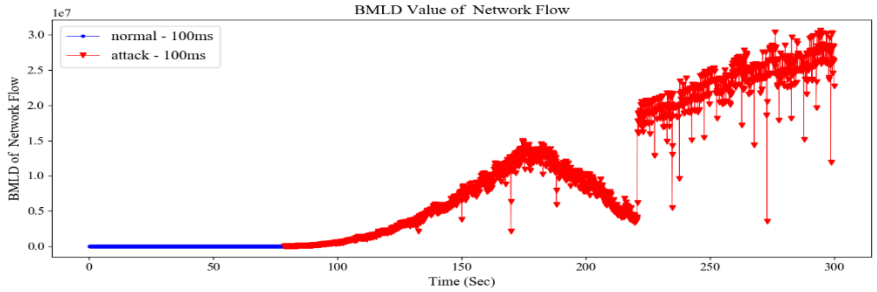
$$MR=\frac{FN}{TN+FN} \tag{13}$$

where *TN* denotes the number of attack samples correctly identified, the number of attack samples misidentified is denoted by *FN*, the number of normal samples correctly identified is denoted by *TP*, the number of normal samples misidentified is denoted by *FP*. With the formula above had definded, the related experiment could be started.

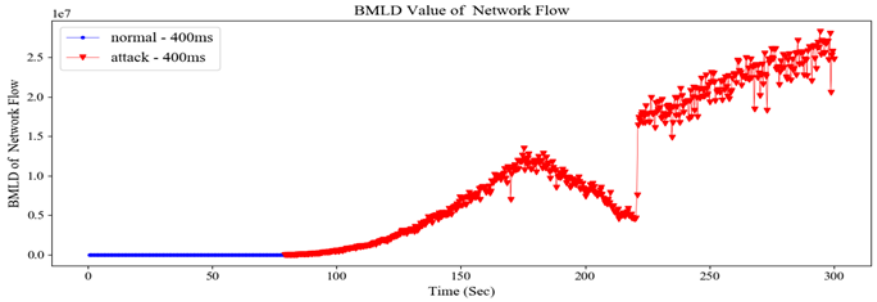
6.2 Experiment results and analysis

6.2.1 The experiment of detection

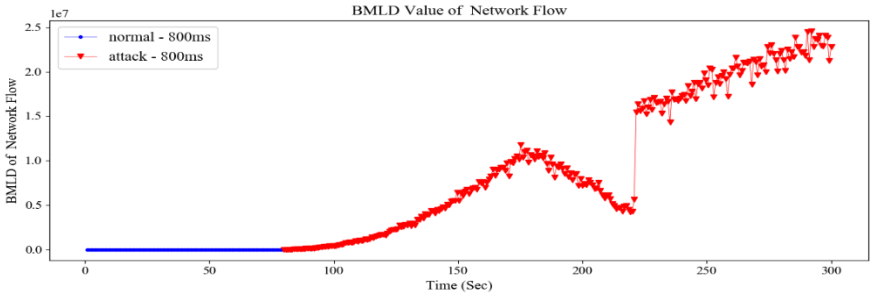
We used different Δt of 100 ms, 400 ms, 800 ms, 1000 ms, 1600 ms and 2000 ms as the parameters of training model, and the comparison of BMLD value of network flow with different Δt as shown in Fig. 4.



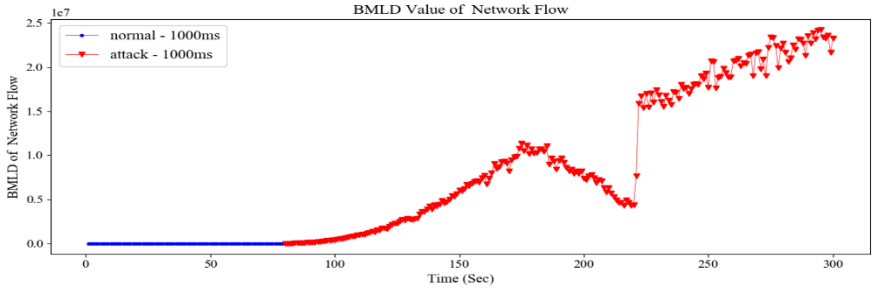
(a) 100 ms



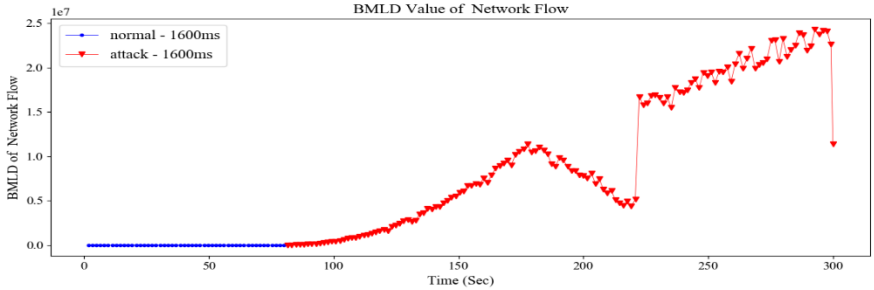
(b) 400 ms



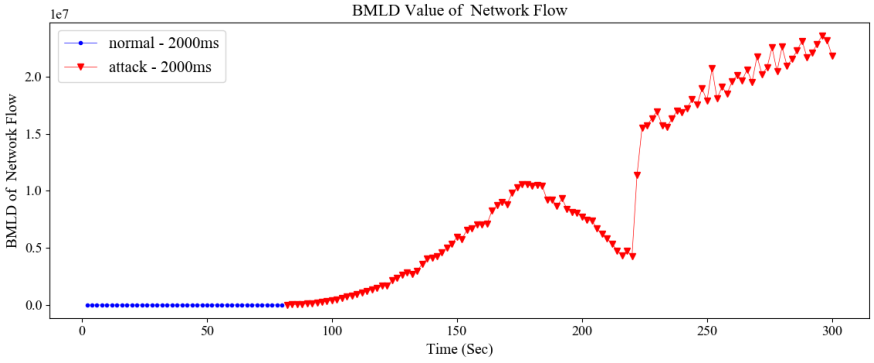
(c) 800 ms



(d) 1000 ms



(e) 1600 ms



(f) 2000 ms

Figure 4: BMLD value of network flow with different Δt

It can be seen from the comparison figures that the BMLD value could successfully distinguished normal flow and DDoS attack flow under six different Δt. The DDoS attack started at around the 79th second, it was worth noting that the BMLD value have dropped dramatically around 220th second.

The reason is factor M, because the defined factor M is the amount of change between the the current new IP value and new IP average value, M value is unstable, which could change at a time rapidly. When calculating the BMLD value, the high sensitivity of M value will cause the BMLD value to drop.

In the V-SVM experiment, we randomly selected 5 training samples from the training set containing normal flow and attack flow, the size was 2000, 4000, 6000, 8000 and 10000 records in sequence. As shown in Tab. 1, we chose the kernel function RBF and V=0.2, in different sample size, support vector is different, but their accurate rate are up to 99%. As the number of samples increase, detection rate also increases and missing rate decreases. As we could see in table is that when the sample size is 10000, the detection rate is the highest which is 99.05% and its missing rate is lowest which is 1.5%. When the sample size is 2000, the support vector is 403 and its accurate rate is 99.9%, but its detection is the lowest when compared to other sample sizes. When the sample size is 6000, its accurate rate is 99.7% which is 0.2% accuracy less than 2000 sample size.

Table 1: V-SVM classification result based on RBF kernel function

Sample size	Optimal parameter	TN(SV)	AR	DR	MR
2000	V=0.2	403	99.9%	95.0%	7.8%
4000	V=0.2	802	99.6%	97.4%	4.0%
6000	V=0.2	1201	99.7%	98.4%	2.0%
8000	V=0.2	1603	99.73%	98.85%	1.8%
10000	V=0.2	2002	99.81%	99.05%	1.5%

With the V-SVM experiment has been done, we also did C-SVM experiment. The purpose of these experiment is comparing the detection effects between two different types of support vector machines. The results of these experiments were shown in Fig. 5. To ensure the reliability of the comparison, training set samples is unchanged, and five different test set samples are randomly chosen from test set which contains normal and attack samples. We compared the different parameters, such AR, DR and MR between C-SVM and V-SVM in different number of samples.

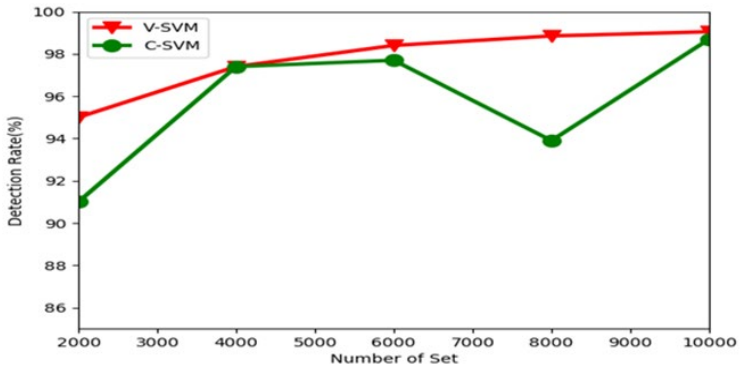


Figure 5: Detection rate of two different methods

From Fig. 5 we can see that V-SVM can detect DDoS attacks better. When the number of sample is 2000, the detection rate of C-SVM is lowest, only 91.7%. When the number of samples is 4000, the detection rate is the same as 97.4%. However, when the number of sample is 8000, the detection rate of C-SVM is 93.5%, and the gap between V-SVM and C-SVM is bigger. When samples increase to 10000, the detection rate of C-SVM and V-SVM remains high, and the gap between them is smaller, only 0.44%. According to the above results of comparison, when sample size increase, the detection rate of V-SVM is increasing then remains stable, indicating that the V-SVM method could accurately identify DDoS attacks. The detection rate of C-SVM is unstable, it also can identify DDoS attacks but its performance is not better than V-SVM.

6.2.2 The experiment of assessment based on cloud model

According to the weight assignment method in Section 4.2, the weight of the indexs are 0.254, 0.349,0.397 respectively. According to formula (8) $RD=NIR \times SAD \times ANF$, we calculate the fusion value RD and classify it according to the cloud model. Before importing the RD value into cloud model, we need to normalize data. The normalization of data is to scale the data so as to compare data in an equal dimension. This method changes the dimensions and magnitude of different data, and convert it to a uniform dimension, which makes it easy to compare indexes with different units or magnitudes. The method we used in this paper is Min-max normalization method, that is, the data is uniformly mapped to the interval [0,1]. After normalization of data, we divide the RD value into four levels, which are normal, low risk, medium risk and high risk respectively, and some test data are shown in Tab. 2.

Table 2: Test data of the cloud model

Normal	Low Risk	Medium Risk	High Risk
0.00	0.163649396	0.458527913	1.00
0.000723367	0.232171437	0.611188904	0.857719481
0.028869119	0.424151244	0.712969597	0.944596597
0.089123065	0.342824791	0.568682142	0.767582306
0.086380673	0.260025169	0.601610558	0.979374618

The corresponding network nodes are affected by the degree of normal, slightly affected, serious damaged and devastated respectively. In this paper, the qualitative transformation experiment based on inverse cloud generator is executed by using the calculated RD data. The cloud drop image generated according to the parameters of the RD cloud model is shown in Fig. 6.

We can see in Fig. 6 is that the distribution of RD corresponding to different risks is different, when network node is high risk, the value of RD are more concentrated in the subsequent interval, mostly between 0.75 and 1.0, this phase is the most violent phase of the DDoS attack, and the victim system is devastating. When network node is normal, RD is smaller than other risk degree and its maximum is 0.146580573. The four different risk degree of the system in the figure are represented by different colors, green means the system is normal, blue means the system suffers slightly, brown means the system suffers serious damage, and red means the system is devastating. From the figure, we can clearly see the risk status of the network node, and can know the extent to which the network node is affected by the DDoS attack. Experiments show that cloud models can effectively evaluate DDoS attacks which can provide network administrators with certain attack situation information.

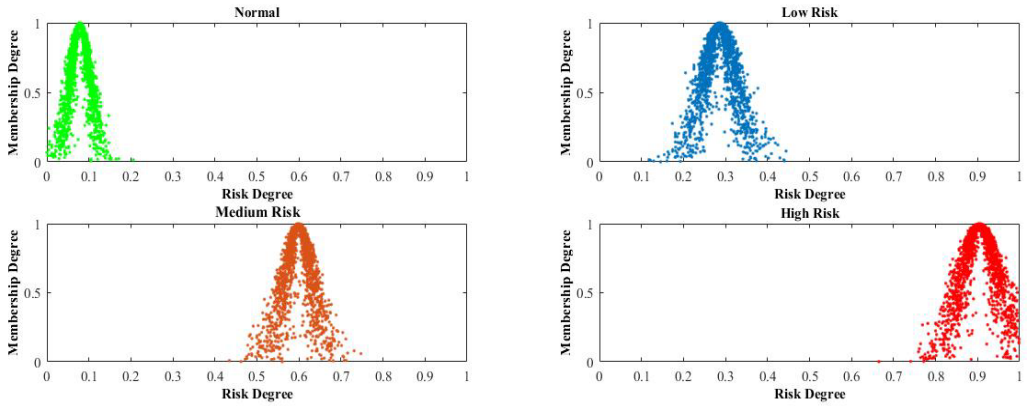


Figure 6: Different risk degree cloud model reduction map

Compared with the assessment model based on Analytic Hierarchy Process (AHP) [Mano, Guillen-Gosalbez, Jimenez et al. (2019)], the evaluation model based in cloud model we have proposed which is taking uncertainty as the starting point, the subjective and objective information is better merged together, and the ambiguity and randomness are inherited, which constitutes the mapping between qualitative and quantitative, and effectively process the uncertainty of DDoS attack situation. On the other hand, the evaluation results finally obtained by this model are presented in the form of cloud images, which are highly intuitive. The comparison of cloud model and assessment model based on AHP is shown in Tab. 3.

Table 3: Comparison of assessment model based on AHP

Function	Model	
	Assessment model based on cloud model	Assessment model based on analytic hierarchy process
Indicator system editing	Support	Not support
Dimensional situation display	Support	Not support
Comprehensive calculation	Support	Support
Comprehensive analysis of qualitative and quantitative indicators	Support	Not support

7 Conclusion

In this paper, we analyzed the characteristics of network flows during normal network flows and DDoS attacks. Then according to the state change characteristics of the new and old user network flow's IP address, we defined the fusion feature BMLD, and we used the V-SVM to identify DDoS attacks, V-SVM can better identify attacks with high accuracy and detection rate. The three proposed indicators better assess the impact of attack flow on network nodes. The cloud model used in this paper can effectively classify attack flow, four levels can better reflect the impact of network nodes on DDoS attacks. The experimental results show that the method can reasonably evaluate the security situation of DDoS attacks, which is more accurate and flexible than the existing evaluation methods.

Funding: 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]; Hainan philosophy and social science 2016 planning project achievements [HNSK(YB)16-86].

References

Alali, M.; Almogren, A.; Hassan, M. M.; Rassan, I. A.; Bhuiyan, M. Z. A. (2018): Improving risk assessment model of cyber security using fuzzy logic inference system. *Computers & Security*, vol. 74, pp. 323-339.

Alireza, Z.; Saygin, C.; HungDa, W.; Yooneun, L.; Alejandro, B. (2018): A game theory based cybersecurity assessment model for advanced manufacturing systems. *Procedia Manufacturing*, vol. 26, pp. 1255-1264.

Bass, T. (2000): Intrusion detection systems and multisensor data fusion. *Communications of the ACM*, vol. 43, no. 4, pp. 99-105.

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

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. *The Computer Journal*, vol. 61, no. 7, pp. 959-970.

Dai, F.; Hu, Y.; Zheng, K. (2017): Exploring risk flow attack graph for security risk assessment. *IET Information Security*, vol. 9, no. 6, pp. 344-353.

Endsley, M. R. (1988): Situation awareness global assessment technique (SAGAT). *National Aerospace and Electronics Conference*, vol. 3, pp. 789-795.

Fan, Z.; Xiao, Y.; Nayak, A. (2017): An improved network security situation assessment approach in software defined networks. *Peer-to-Peer Networking and Applications*, pp. 1-15.

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

Jana, D. K.; Ghosh, R. J. (2018): Novel interval type-2 fuzzy logic controller for improving risk assessment model of cyber security. *Information Security and Applications*, vol. 40, pp. 173-182.

Jin, B.; Simon, Y.; Dong, S.; Armstrong, N.; Noora, F. et al. (2018): Dynamic security metrics for measuring the effectiveness of moving target defense techniques. *Computers & Security*, vol. 79, pp. 33-52.

Jordan, N.; Niall, A. (2018): Real-time dynamic network anomaly detection. *IEEE Intelligent Systems*, vol. 33, no. 2, pp. 5-18.

Khundrakpam, J. S.; Tanmay, D. (2017): MLP-GA based algorithm to detect application layer DDoS attack. *Journal of Information Security and Applications*, vol. 36, pp. 145-153.

Koh, P. W.; Liang, P. (2017): Understanding black-box predictions via influence functions. *ArXiv Preprint ArXiv*, vol. 1703, pp. 730-756.

Kshira, S.; Deepak, P.; Mayank, T.; Joel, J.; Rodrigues, B. (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.

Li, X.; Zhao, H. (2016): Network security situation assessment based on HMM-MPGA. *International Conference on Information Management*, vol. 24, no. 6, pp. 57-69.

Liu, D.; Dong, L.; Lv, S. (2017): A novel approach to network security situation assessment based on attack confidence. *Network and System Security*, vol. 10394, pp. 450-463.

Liu, X.; Wang, H.; Lu, H. (2016): Fusion-based cognitive awareness-control model for network security situation. *Journal of Software*, vol. 27, no. 8, pp. 2099-2114.

Liu, Z.; Zhang, B.; Zhu, N.; Li, L. (2017): Hierarchical network threat situation assessment method for DDoS based on D-S evidence theory. *IEEE International Conference on Intelligence and Security Informatics: Security and Big Data (ISI)*, pp. 49-63.

Lounis, K. (2018): Stochastic-based semantics of attack-defense trees for security assessment. *Electronic Notes in Theoretical Computer Science*, vol. 337, pp. 135-154.

Mano, T. B.; Guillen-Gosalbez, G.; Jimenez, L.; Ravagnani, M. (2019): Synthesis of heat exchanger networks with economic and environmental assessment using fuzzy-Analytic Hierarchy Process. *Chemical Engineering Science*, vol. 195, pp. 185-200.

Matheu-García, S. N.; Hernández-Ramos, J. L.; Skarmeta, A. F. (2018): Risk-based automated assessment and testing for the cybersecurity certification and labelling of IOT devices. *Computer Standards & Interfaces*, vol. 62, pp. 64-83.

Mu, T.; Nandi, A. (2005): Detection of breast cancer using V-SVM and RBF networks with self-organized selection of centers. *IEEE International Seminar on Medical Applications of Signal Processing*, vol. 51, pp. 47-59.

Pu, Z. (2018): Network security situation analysis based on a dynamic Bayesian network and phase space reconstruction. *The Journal of Supercomputing*, pp. 1-16.

Sabrina, S.; Alessandra, R.; Daniele, M.; Alberto, C. P. (2018): A risk assessment methodology for the Internet of Things. *Computer Communications*, vol. 129, pp. 67-79.

Teng, H.; Liu, Y.; Liu, A.; Xiong, N. N.; Cai, Z. et al. (2019): A novel code data dissemination scheme for Internet of Things through mobile vehicle of smart cities. *Future Generation Computer Systems*, vol. 94, pp. 351-367.

Wang, J. Q.; Peng, J. J.; Zhang, H. Y. (2016): An uncertain linguistic multi-criteria group decision-making method based on a cloud model. *Group Decision and Negotiation*, vol. 24, no. 1, pp. 171-192.

Wu, Z.; Liu, J.; Xu, S. (2016): A cyberspace security situation awareness model based on complex network. *11th International Conference on Reliability, Maintainability and Safety*, pp. 1-12.

Xu, G.; Cao, Y.; Ren, Y. (2017): Network security situation awareness based on semantic ontology and user-defined rules for Internet of Things. *IEEE Access*, vol. 99, pp.1-15.

Yusuf, E. S.; Ge, M.; Hong, J. B.; Alzaid, H.; Kim, D. S. (2018): A systematic evaluation of cybersecurity metrics for dynamic networks. *Computer Networks*, vol. 144, pp. 216-229.

Zahid-Hasan, M. D.; Zubair-Hasan, K. M.; Sattar, A. (2018): Burst header packet flood detection in optical burst switching network using deep learning model. *Procedia Computer Science*, vol. 143, pp. 970-977.

Zhang, R.; Cheng, J.; Tang, X.; Liu, Q. (2018): DDoS attack security situation assessment model using fusion feature based on Fuzzy C-Means clustering algorithm. *Cloud Computing and Security*, vol. 11064, pp. 654-669.

Zhao, D.; Liu, J. (2018): Study on network security situation awareness based on particle swarm optimization algorithm. *Computers & Industrial Engineering*, vol. 125, pp. 764-775.

Zhong, C.; Lin, T.; Liu, P.; Yen, J.; Chen, K. (2018): A cyber security data triage operation retrieval system. *Computers & Security*, vol. 76, pp. 12-31.