DDoS Attack Detection Method for Space-Based Network Based on SDN Architecture

JIA Min[1], SHU Yuejie[1], GUO Qing[1], GAO Zihe[2], XIE Suofei[2]

(1. Communication Research Center, School of Electronics and Information Engineering,
Harbin Institute of Technology, Harbin 150006, China;
2. Institute of Telecommunication Satellite, China Academy of Space Technology, Beijing 100094, China)

Abstract: With the development of satellite communications, the number of satellite nodes is constantly increasing, which undoubtedly increases the difficulty of maintaining network security. Combining software defined network (SDN) with traditional space-based networks provides a new class of ideas for solving this problem. However, because of the highly centralized network management of the SDN controller, once the SDN controller is destroyed by network attacks, the network it manages will be paralyzed due to loss of control. One of the main security threats to SDN controllers is Distributed Denial of Service (DDoS) attacks, so how to detect DDoS attacks scientifically has become a hot topic among SDN security management. This paper proposes a DDoS attack detection method for space-based networks based on SDN architecture. This attack detection method combines the optimized Long Short-Term Memory (LSTM) deep learning model and Support Vector Machine (SVM), which can not only make classification judgments on the time series, but also achieve the purpose of detecting and judging through the flow characteristics of a period of time. In addition, it can reduce the detection time as well as the system burden.

Keywords: space-based network; SDN; DDoS attack; LSTM; SVM

1 Introduction

With the continuous development of satellite communications, the number of satellite network nodes is increasing and people are trying to pursue the overall grasp of the network. Because the control plane and forwarding plane of the software defined network (SDN) network are separated and the centralized network control mode is adopted, the rapid deployment of business and rapid update of information in the satellite network, such as link switching, traffic distribution and topology changes, can be realized. Therefore, people have turned their attention to SDN. In the space-based network architecture of the SDN network, the ground station is regarded as the SDN controller and the satellite is regarded as the OpenFlow switch, so as to realize a flexible and extensible network architecture. The SDN controller implements a highly centralized network management, making the SDN controller the focus of network attacks. Once the SDN controller has been attacked, the network it manages will be paralyzed due to loss of control. Therefore, the security of the controller is the key to the security of the entire SDN network. The Distributed Denial of Service (DDoS) attack is one of the main threats to the security of

the controller. How to quickly and accurately detect the DDoS attack has become a research hotspot in the field of SDN security. At present, the detection methods for DDoS attacks in SDN networks mainly use statistical analysis and machine learning methods to detect DDoS attacks in the network by deploying anomaly detection technology in the SDN controller. In Ref. [1], a detection method based on entropy anomaly is proposed, which can determine whether the current state is in an abnormally attacked state according to the entropy value and detect the DDoS attack through the change of the entropy value of the network characteristics. Another detection algorithm based on self-organizing maps (SOM) is proposed in Ref. [2], in which the self-organizing maps build an unsupervised artificial neural network, trained with traffic flow features. This is a lightweight DDoS attack detection method. The self-organizing mapping traffic analysis can obtain a higher detection rate and a lower false alarm rate. A hybrid machine learning model, Support Vector Machine (SVM)-SOM, is proposed in Ref. [3] to detect DDoS attacks. Compared with simple machine learning models, hybrid machine learning models provide higher accuracy, detection rate and fewer false alarm rates. In Ref. [4], a method based on K-means++ and fast K-nearest neighbors is proposed, as well as a modular detection system in the controller[5]. However, the above-mentioned DDoS attack detection methods still have some limitations[6]. The limitation of traditional machine learning for DDoS attack detection is that the historical characteristics of traffic cannot be used. Its main purpose is to improve the classification detection accuracy of a single sample without processing the time series[7]. However, when a DDoS attack occurs, the extracted traffic information is more suitable for the detection of time series samples, and the classification prediction of samples with time series relationships is more suitable for the use of some deep learning methods that can handle time series data, such as Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) model[8]. In this paper, a DDoS attack detection method for space-based network based on SDN architecture is proposed, which combines the optimized LSTM deep learning model and SVM. The optimized LSTM model is used to classify and judge the time series to reduce the false alarm problem of the traditional machine learning classifier for unstable abnormal traffic[9].

2 Detection Mechanism

Fig. 1 shows the proposed DDoS attack detection mechanism of space-based network based on SDN architecture.

After collecting the flow table information, the SDN control center of ground station extracts the feature vector according to the feature extraction algorithm, and caches the data extracted in real time into a file for storage. The extracted flow table feature vectors are sent to the SVM model for detection. The SVM uses the provided feature vector information to determine whether it is attack traffic or normal traffic. If it is normal traffic, the result is directly saved; if it is abnormal traffic, it is combined with the traffic information of the previous time after standardization to form a time series and sent to the optimized LSTM model for checking. If the optimized LSTM model judges that it is abnormal traffic, it outputs the information of abnormal traffic detected, indicating that it is under DDoS attack; if it is detected as normal traffic, it is judged as normal traffic. Moreover, the time series method is mainly used to solve the problem of false alarms of single abnormal traffic feature vectors in the previous machine learning of DDoS attack detection. Therefore, when the SVM classifier makes an abnormal judgment, the system will send the flow table feature information based on previous times to the optimized LSTM deep learning model to make a judgment on the next flow table information. The final result is given by the comprehensive judgment of the detection mechanism.

We will focus on the data preprocessing part, using the improved genetic algorithm to optimize the LSTM model and introducing SVM to solve the misjudgment problem caused by the LSTM data sensitivity at the initial stage of the network. Finally, we will build an experimental simulation platform to verify the feasibility of the proposed DDoS attack detection method for space-based network based on SDN architecture.

3 Data Preprocessing

3.1 Feature Extraction Algorithm of OpenFlow Flow Tables

Although the DDoS attack methods adopted by attackers are diverse, most of the attack traffic still has a certain pattern. Therefore, by collecting the flow table information and converting it into a feature vector, the change of network traffic distribution characteristics per unit time can be analyzed to distinguish normal traffic from attack traffic.

The original OpenFlow flow table information mainly includes the source port, destination port, source IP address, destination IP address, frame length, frame protocol, packet length, etc. Direct training on these data has a poor effect and is greatly affected by the network environment itself. Therefore, a feature extraction algorithm is first used for the flow table information to extract the following five features. We assume that the SDN controller samples the flow entries of the OpenFlow switch at period τ and the total number of flow entries obtained is $Total_F$:

(1) Average number of flow packets (ANFP).

The number of flow packets under normal network conditions and that under attacks are different. Attacks usually generate fake IP addresses continuously and randomly, and reduce the number of packets contained in a single flow entry. Eq. (1) defines the average number of flow packets, where $DataPackagesNum_k$ represents the number of data packets of the k-th flow entry.
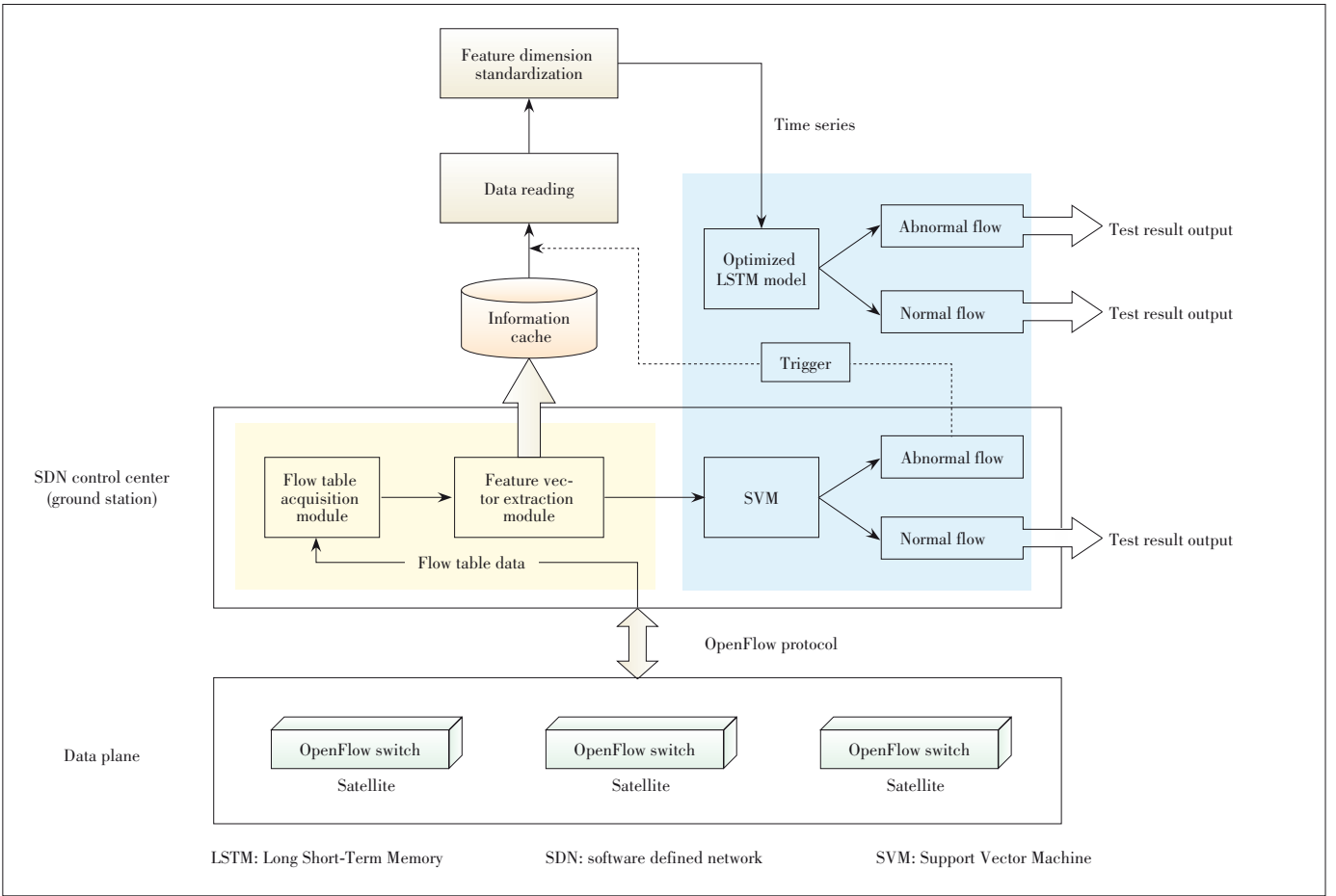
▲Figure 1. Distributed Denial of Service (DDoS) attack detection mechanism of space-based network based on SDN architecture.

$$ANFP = \frac{\sum\limits_{k}^{Total_F} DataPackagesNum_k}{Total_F}. \tag{1}$$

(2) Average number of bits in flows (ANBF).

Similar to the definition of ANFP, when a DDoS attack occurs, the attacker will send a large number of packets with a small number of bits, which also provides a basis for DDoS attack detection. Eq. (2) defines the average number of bits in flows, where $BytesNum_k$ is the number of data packet bits in the k-th flow entry.

$$ANBF = \frac{\sum\limits_{k}^{Total_F} BytesNum_k}{Total_F}. \tag{2}$$

(3) Flow generation speed (FV).

When a DDoS attack occurs, a large number of pseudo IP addresses will send data packets, resulting in an increase in the number of flow tables during the collection time. Eq. (3) defines the flow generation speed, where $FlowNum$ is the number of flow tables collected in period τ.

$$FV = \frac{FlowNum}{\tau}. \tag{3}$$

(4) Source IP address generation speed (SIPV).

The main attack feature of DDoS is to send a large number of data packets by forging the source IP address, which makes the growth rate of the number of source IP addresses at the time of the attack greatly increase at a fixed time. Eq. (4) defines the source IP address number generation speed, where $SIPNum$ is the number of different source IP addresses in the sampling period.

$$SIPV = \frac{SIPNum}{\tau}. \tag{4}$$

(5) Port growth (PV).

Under normal circumstances, the increase in the number of ports is relatively stable. In a DDoS attack, since the port number is randomly generated, when the attack occurs, the growth rate of the port will be greatly increased. Eq. (5) defines the port growth rate, where $PortNum$ represents the sum of different port numbers corresponding to different IP addresses in the flow table within the sampling period.

$$PV = \frac{PortNum}{\tau}.$$ (5)

3.2 Collecting of Flow Table Information

Normal network environment and DDoS attack environment are simulated in SDN. A total of more than 20 000 pieces of data are collected as a data set, which contains a ratio of normal traffic to attack traffic of about 1: 1. Besides, the normal traffic is labelled as 0 and the abnormal traffic is labelled as 1.

According to the proposed feature extraction algorithm, the data set is used to extract feature vectors to obtain a data set {*data*} that contains five columns of feature vectors and one column of labels.

3.3 Standardization and Processing of Time Series

The data is read in {*data*} and standard deviation standardization processing is conducted, that is, the mean value for each feature dimension is removed and the variance normalization operation is conducted so that the processed data conforms to the standard normal distribution (the mean is 0 and the standard deviation is 1). The conversion function is:

$$X = \frac{x - \mu}{\sigma}.$$ (6)

The standardized data set is a two-dimensional data set. Each data set has five feature dimensions, which can be marked as $R_i = [r_{i1}, r_{i2}, r_{i3}, r_{i4}, r_{i5}]$. As shown in **Fig. 2**, because the data type of the LSTM model is a set of time series, it is necessary to use a sliding time window to process the data, with *time_steps* = k, every time the time window slides to get a set of time series, from Seq_1 to $Seq_{(Num - k)}$. The resulting data set is a three-dimensional data set with a size of $(Num - k) \times k \times 5$. The processed time series data set {*Seq*} is sent to the LSTM deep learning model for classification.
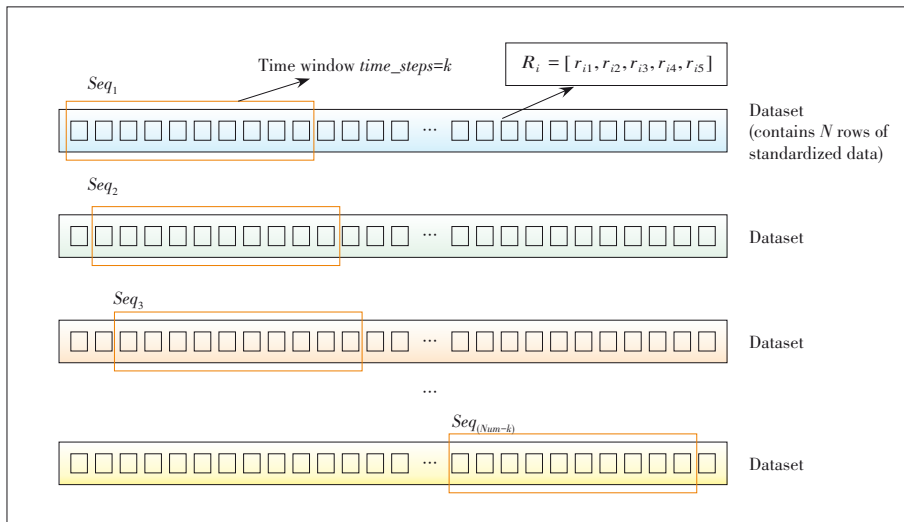
▲Figure 2. Schematic diagram of time window processing.

4 Optimization of LSTM Deep Learning Model

LSTM is a kind of time recurrent neural network, which is specially designed to solve the long-term dependence problem of general Recurrent Neural Network (RNN). Unlike a single neural network layer, the repeating module in LSTM contains four interactive layers. The main mathematical process is as follows：

$$\left. \begin{aligned} f_t &= \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \\ i_t &= \sigma(W_f \cdot [h_{t-1}, x_t] + b_i) \\ \tilde{C}_t &= \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \\ C_t &= f_t \times C_{t-1} + i_t \times \tilde{C}_t \\ o_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \\ h_t &= o_t \times \tanh(C_t) \end{aligned} \right\}.$$ (7)

The genetic algorithm (GA) is a heuristic search and an optimization method inspired by natural selection process. It is widely used to find the best solution to optimization problems with large parameter spaces. In addition, because it does not consider auxiliary information (such as derivatives), it can be used for discrete optimization and continuous optimization. In Ref. [10], GA is used to optimize the LSTM model to find the optimal time window size and the number of neurons and to reduce the problem of overfitting.

However, the root mean square error (RMSE) of the fitness function used in Ref. [10] is less effective for the classification of the binary classification problem of DDoS attack detection in this paper, and because the scheme uses a simple genetic algorithm, the convergence effect is poor. Therefore, an improved genetic algorithm is proposed in this paper to optimize the LSTM model. We add the elite retention strategy, use two-class cross entropy loss to replace RMSE, and use an improved adaptive strategy to perform crossover and mutation operations for further optimization of the algorithm.

First, the elite retention strategy joins. In a simple genetic algorithm, the generation of the new generation population is used to completely replace the parent population. Sometimes, the optimal individuals of the current population are lost and the genetic algorithm cannot converge to the global optimal solution. Therefore, the elite retention strategy is added so that the optimal individuals appearing in the evolution process will not be destroyed due to crossover and mutation. The steps are as follows:

Supposing the evolution to the t–th generation, the best individual in the group is $N(t)$ and $N(t+1)$ is the new generation population; $N(t)$ is directly added to the new population $N(t+1)$ without crossing and mutation. Then n_{pop} individuals are taken out from it to maintain the population size. In this way, the purpose of retaining elite individuals can be achieved and the global convergence ability of the genetic algorithm can be greatly improved.

Because RMSE of the fitness function is relatively poor for the classification of the binary classification problem of the DDoS attack detection in this paper, the cross-entropy loss function of the binary classification is used as the fitness function. When a sample has a true label $y_n = 1$, the greater the classifier prediction probability P, the smaller the loss; when a sample has a true label $y_n = 0$, the greater the classifier prediction probability P, the greater the loss. The formula is as follows:

$$L_B = -[\, y_n \log(P) + (1 - y_n)\log(1 - P)\,]. \tag{8}$$

The roulette is used in Ref. [10] for selection, but the selection error of the selection method is also large due to the random operation and even individuals with large fitness may be missed. In this paper, the tournament selection model is used to randomly select N individuals from the group for fitness comparison. The highest fitness individual is inherited to the next generation. The number of individuals for each fitness comparison is the league size N.

Then adaptive strategies are used to improve the previous algorithm. The improvement of the adaptive genetic algorithm lies in the adaptive adjustment of genetic parameters to maintain the diversity of the population and ensure the convergence of the algorithm.

For the basic genetic algorithm, the probability of crossover and mutation is fixed. However, the adaptive strategy needs adaptive adjustment during the evolution process: choosing a larger probability of crossover and mutation at the beginning. Such a rough search process is conducive to maintaining population diversity. In the later stage, it can be adjusted to a smaller value for detailed search to prevent the optimal solution from being destroyed and accelerate the convergence speed. However, in order to stabilize the population in future evolutions, the impact of mutations needs to be reduced. The corresponding measure is to reduce the possibility or degree of mutation.

The calculation method of the cross probability P_c in the Srinvivas adaptive genetic algorithm is shown in Eq. (9).

$$P_c = \begin{cases} P_{c1} - \dfrac{(P_{c1} - P_{c2})(f' - f_{avg})}{(f_{max} - f_{avg})}, & f' \geqslant f_{avg} \\ P_{c1}, & f' < f_{avg} \end{cases}, \tag{9}$$

where P_c is the crossover probability, f_{max} is the largest fitness

value in the group, f_{avg} is the average fitness value of each generation group, and f' is the larger fitness value of the two individuals to be crossed.

In this paper, we improve the calculation method of P_m in the Srinvivas adaptive genetic algorithm by introducing the Cauchy distribution function as a function of the degree of variation. In this way, as the number of population evolution increases, the value of the degree of variation decreases gradually. Because of the distribution characteristics of the Cauchy distribution function, the tails at both ends are long and the values at the center are suitable, so that it meets the requirements of the adaptive mutation and has a good ability to adjust mutations. A reasonable mutation value can be obtained in the early stage of evolution, and the probability of mutation can be greatly reduced in the later stage of evolution to avoid destroying the optimal solution. The calculation method of P_m is shown in Eq. (10).

$$P_m = k \times \frac{1}{\pi \times (1 + gen^2)}, \tag{10}$$

where P_m is the mutation probability, gen refers to the generation number, and k is the coefficient.

In summary, the algorithm steps to optimize the LSTM model using the improved genetic algorithm are as follows:

• Step 1: Gene coding is performed in binary, the first four bits of the binary string indicate the size of the time window, and the last six bits indicate the number of neurons;

• Step 2: LSTM neural network training is carried out and the binary classification cross-entropy loss function is used as the individual fitness value evaluation;

• Step 3: The tournament selection model method is used to determine the parent individuals involved in replication.

Fig. 3 shows the results of the original algorithm applied to the data set training. It can be seen that the loss function RMSE performs poorly on the {0, 1} binary classification, which makes it difficult for the genetic algorithm to converge.
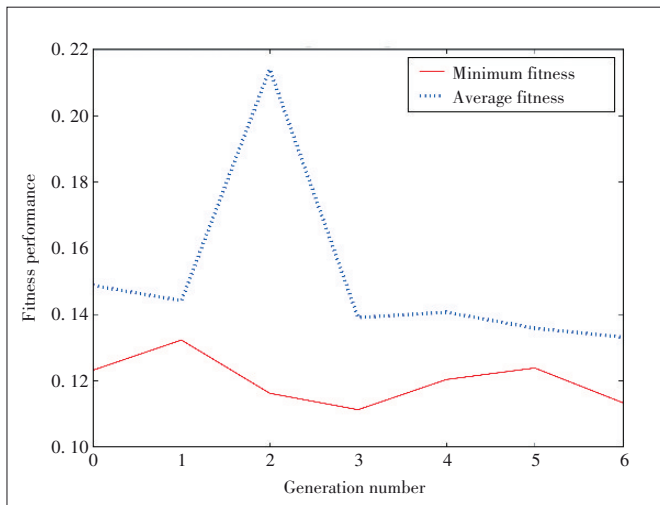
Fig. 4 shows the training results obtained by using the binary classification cross-entropy loss function instead of RMSE as the fitness function. Compared with Fig. 3, the training results have been improved, but it can be seen that the convergence speed of the simple genetic algorithm is still slow and the convergence effect is not good.

Fig. 5 shows the training results obtained by using the improved genetic algorithm in this paper. It can be seen from the figure that the convergence speed is faster and the convergence effect is better.

The optimal solution based on the training results obtained by the improved genetic algorithm is: the optimal time window length is 15, the number of neurons in the corresponding LSTM model is 40, and the corresponding two-class cross-entropy loss function value is 0.10876369144052776 at this time.
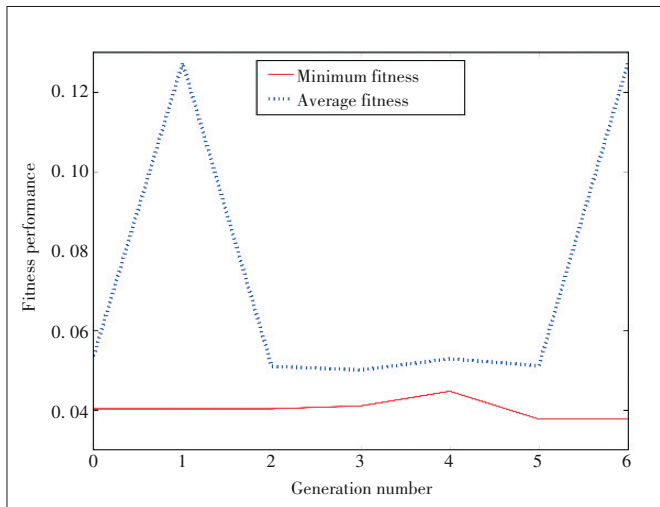
▲Figure 3. Results of the original algorithm applied to the data set training.


▲ Figure 4. Training results obtained by using the binary classification cross-entropy loss function instead of root mean square error (RMSE) as the fitness function (still in the state of simple genetic algorithm).
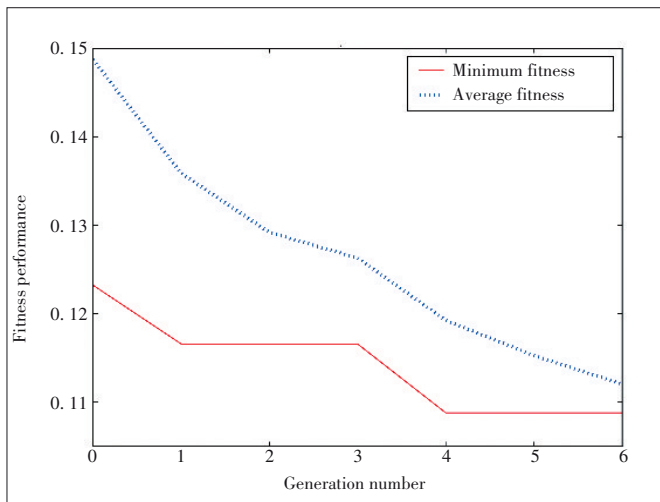

▲Figure 5. Training results obtained by using the improved genetic algorithm.

The data set is {*Seq*}, the data training environment is Linux system Ubuntu 16.04, the processor is Intel Core i7, and the GPU is GeForce 940MX. A total of 30 training cycles are conducted and the classification accuracy rate is 99.77%. The training model is accurate and high and can realize the time series detection of the flow characteristic information. The final training results of the LSTM model optimized by the improved genetic algorithm are shown in **Figs. 6**, **7** and **8**, which are the accuracy curve, error curve and confusion matrix, respectively.

In addition, the judgment time for a set of traffic of the LSTM depth model is optimized with the time spent in the single experiment more than 400 μs and the total time more than 2 s. In order to solve the problem of misjudgment caused by the initial unstable network traffic, reduce the detection time of the system and reduce the burden on the system, SVM is in-
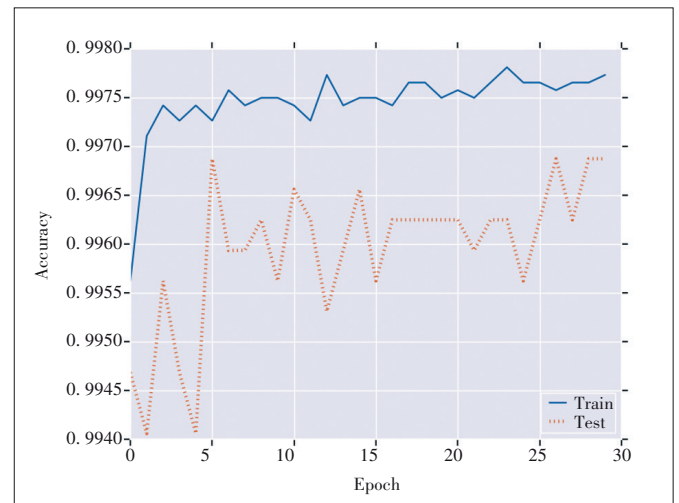
▲Figure 6. Accuracy curve of the optimized Long Short-Term Memory (LSTM) model.
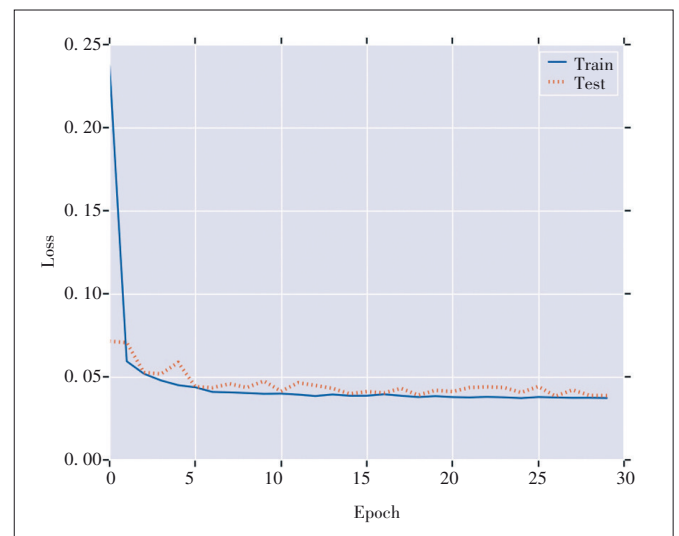
▲ Figure 7. Error curve of the optimized Long Short-Term Memory (LSTM) model.

troduced to classify the flow table feature vectors at a certain time. A normal feature vector shows that the network has not received an attack. If the vector is judged to be abnormal, a time series is formed together with several flow table feature vectors of the previous time to perform data preprocessing, and then sent to the LSTM deep learning model for secondary judgment. It has been experimentally determined that the hybrid detection mechanism can solve the aforementioned problem of misjudgment of normal traffic at the initial stage of the network.

The data training environment is Linux system and the processor is Intel Core i7. The data set is {*data*}, which is trained using the SVM model and the Random Forest model. **Table 1** shows the training results.

Both classifier models show good accuracy, but the detection time of the SVM model is shorter, which is beneficial to reduce the burden on the system, so SVM is selected as the front-end classifier of the hybrid model.

5 Experimnetal Verification and Simulation

A space-based network based on SDN architecture with Mininet software was built to verify the effectiveness of the DDoS attack detection mechanism proposed in this paper.

The experiment was completed in Linux environment, using OpenFlow1.3 protocol. As shown in **Fig. 9**, the space-based network topology consists of the Ground station SDN controller, satellites (as OpenFlow Switches) and virtual hosts connected to satellites.
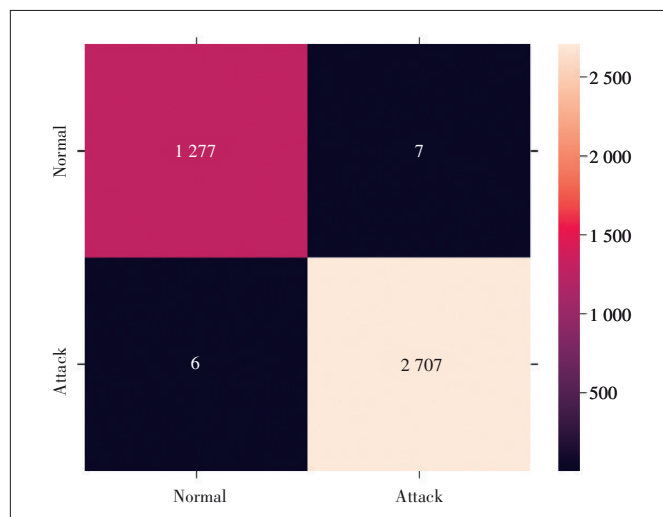
▲ Figure 8. The confusion matrix of the optimized Long Short-Term Memory (LSTM) model.
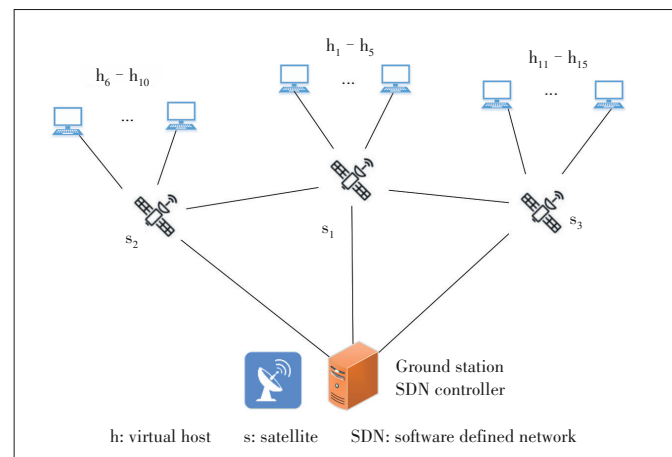
▼Table 1. Training results

Model	Training Set Accuracy/%	Test Set Accuracy/%	Single Test Time/ms
SVM	99.85	99.86	0.116
Random forest	99.90	99.82	118.966

SVM: Support Vector Machine

The network background traffic (Normal) was first simulated in the SDN environment, and then the Hping3 tool was used to launch DDoS attacks. As shown in **Fig. 10**, this DDoS attack detection method for space-based network based on SDN architecture could successfully detect DDoS attacks.

6 Conclusions

A DDoS attack detection method is proposed for space-based network based on SDN architecture in this paper. This method combines the optimized LSTM deep learning model and SVM. First of all, the OpenFlow flow table feature extraction algorithm is used and the time series is pre-processed. Then an improved genetic algorithm is used to optimize the LSTM neural network model to better evaluate the time series prediction problem. Finally, SVM is introduced to solve the misjudgment caused by the sensitivity of the LSTM model to data during the network startup phase. Such an attack detection mechanism can not only make classification judgments on the time series, but also detect and judge the traffic characteristics through a period of time, so as to reduce the false alarm

▲Figure 9. Space-based network topology.

▲ Figure 10. Successful detection of the Distributed Denial of Service (DDoS) attack detection method for space-based network based on software defined network (SDN) architecture.

problem caused by a single machine learning classifier for individual abnormal traffic. In addition, the detection mechanism can also reduce the misjudgment rate of the network startup stage and further reduce the detection time and the system burden. Finally, an SDN space-based network experimental simulation platform is built to verify the feasibility of the DDoS attack detection method for space-based network based on SDN architecture.

References

[1] GIOTIS K, ARGYROPOULOS C, ANDROULIDAKIS G, et al. Combining openflow and sflow for an effective and scalable anomaly detection and mitigation mechanism on SDN environments [J]. Computer networks, 2014, 62: 122 – 136. DOI: 10.1016/j.bjp.2013.10.014

[2] BRAGA R, MOTA E, PASSITO A. Lightweight DDoS flooding attack detection using NOX/OpenFlow [C]//35th Annual IEEE Conference on Local Computer Networks. Denver, USA: IEEE, 2010: 408 – 415. DOI: 10.1109/LCN.2010.5735752

[3] DEEPA V, SUDAR K M, DEEPALAKSHMI P. Detection of DDoS attack on SDN control plane using hybrid machine learning techniques [C]//International Conference on Smart Systems and Inventive Technology (ICSSIT). Tirunelveli, India: IEEE, 2018: 299 – 303. DOI: 10.1109/ICSSIT.2018.8748836

[4] XU Y H, SUN H T, XIANG F, et al. Efficient DDoS detection based on K-FKNN in software defined networks [J]. IEEE access, 2019, 7(160536 – 160545). DOI: 10.1109/ACCESS.2019.2950945

[5] MCKEOWN N. Software-defined networking [Z]. Infocom Keynote Talk, Rio de Janeiro, Brazil, 2009

[6] ZHENG S J. Research on SDN-based IoT security architecture model [C]//IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC). Chongqing, China: IEEE, 2019: 575 – 579. DOI: 10.1109/ITAIC.2019.8785456

[7] SHIN S, XU L, HONG S, et al. Enhancing network security through software defined networking (SDN) [C]//25th International Conference on Computer Communication and Networks (ICCCN). Waikoloa, USA: IEEE, 2016: 1 – 9. DOI: 10.1109/ICCCN.2016.7568520

[8] YANG L, ZHAO H. DDoS attack identification and detection using SDN based on machine learning method [C]//15th International Symposium on Pervasive Systems, Algorithms and Networks (I-SPAN). Yichang, China: IEEE, 2018: 174 – 178. DOI: 10.1109/I-SPAN.2018.00036

[9] TATANG D, QUINKERT F, FRANK J, et al. SDN-guard: protecting SDN controllers against SDN rootkits [C]//2017 IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN). Berlin, Germany: IEEE, 2017: 297 – 302. DOI: 10.1109/NFV-SDN.2017.8169856

[10] VAITHEESWARAN S S, VENTRAPRAGADA V R. Wind power pattern prediction in time series measuremnt data for wind energy prediction modelling using LSTM-GA networks [C]//10th International Conference on Computing, Communication and Networking Technologies (ICCCNT). Kanpur, India: IEEE, 2019: 1 – 5

Biographies

JIA Min (jiamin@hit.edu.cn) received her M.Sc degree in information and communication engineering from Harbin Institute of Technology (HIT), China in 2006 and Ph.D. degree from SungKyungKwan University of Korea and HIT in 2010. She is currently a professor and Ph.D. supervisor with the School of Electronic and Information Engineering, HIT. Her research interests include advanced digital signal processing and integrated satellite and terrestrial communication systems.

SHU Yuejie received her B.S. and M.Sc degrees from Harbin Institute of Technology, China in 2018 and 2020, respectively. Her research interest focuses on software defined networking applications.

GUO Qing received his M.Sc. degree from Beijing University of Post and Telecommunications, China in 1985 and and Ph.D. degree from Harbin Institute of Technology (HIT), China in 1998. He is currently a professor and the dean of the School of Electronics and Information Engineering, HIT. His research interest focuses on satellite communications.

GAO Zihe received the B.S. degree in electronics and information engineering, the M.E. degree in signal and information processing, and the Ph.D. degree in information and communication engineering from Harbin Institute of Technology, China in 2005, 2007 and 2011, respectively. At present, he is mainly engaged in the research of satellite communication technology at the Insititute of Telecommunication Satellite, China Academy of Space Technology.

XIE Suofei received his B.S. and M.Sc degrees in communication and information system from Chongqing University, China. He is an engineer with the Institute of Telecommunication Satellite, China Academy of Space Technology. His research interest focuses on satellite communications.