Deep Learning Based Network Intrusion Detection

Jun Yu(✉), Jiwei Hu, and Yong Zeng

Wuhan Fiberhome Technical Services Co., Ltd., Wuhan, China
yujun@fiberhome.com

Abstract. With the advancement of the times, the network has become an important part of people's daily life, and its connection with our daily life of clothing, food, housing, transportation, medical education has become increasingly close. However, while the network brings us a richer and faster life, the network security problem is also becoming more and more prominent. Network security risks are posing new challenges to the economy, politics, ecology, national security, science and technology development and other fields, and the network security problem has received wide attention, and network intrusion detection, as an important part of the network security field, needs more attention from us. To further improve the performance of feature extraction for network intrusion data, a combined model based on convolutional neural networks and long short-term memory units is proposed for the problems of gradient disappearance and gradient explosion of ordinary neural networks, and also the improved seagull optimization algorithm is applied to the optimization of the model parameters, and Batch Normalization and Adam optimizer, thus constructing an efficient model.

Keywords: network intrusion detection · convolutional neural network (CNN) · long short-term memory (LSTM)

1 Introduction

While the network has brought us a richer and more convenient life, its security problems have become increasingly prominent. The risk of network security is posing new challenges to politics, economy, culture, national defense, ecology and other fields [1]. Therefore, it has become an urgent problem to improve the reliability of information, purify the network security environment, maintain the computer system from damage and improve the security and reliability of the network.

With the development of the times and the progress of technology, machine learning technology has attracted the attention of many scholars, and a large number of scholars have participated in the research on how to apply machine learning to network intrusion detection. Traditional machine learning methods have been widely used in network intrusion detection [3]. The intrusion detection system designed by Saleh et al. Combines many machine learning methods, has strong real-time performance, and can solve multi classification problems. Dimensionality reduction is an important operation in data processing. This system uses the basic feature selection (NBFS) technology to process the

W. Hong and G. Kanaparan (Eds.): ICCSE 2023, CCIS 2023, pp. 125–136, 2024.
https://doi.org/10.1007/978-981-97-0730-0_12

data samples, and then eliminates the samples with high dispersion through the optimized support vector machine (OSVM). Finally, PKNN (an improved version of the classic KNN) is used to detect the attack, and experiments are carried out on the classic KDD 99, NSL-KDD and kyoto2006 + data sets. It shows the real-time performance of detection and achieves good detection effect [6–8]. Literature [9] uses support vector machine (SVM) combined with compressed sampling for intrusion detection. Compressed sampling method is adopted for the processing of network data stream. A model based on SVM is established to output the results, and good experimental results are achieved in training efficiency and detection speed. Reference [10] proposed a clever data processing method, which uses Fisher discriminant ratio (FDR) method for feature extraction and denoising. In the subsequent processing, in order to preserve the characteristics of the data, the probabilistic self-organizing maps (PSOM) is used for modeling, which has achieved good results in the classification accuracy. The training time of classifier is also an indicator worthy of attention. Al-Yaseen and others constructed a multi-level model in order to improve the training efficiency and the performance of the model. Firstly, in terms of data processing, the improved k-means algorithm is used to purify and sample the original training set. Starting from the specific data itself, the training efficiency of the optimized data is greatly improved. In the final result output stage of the model, it uses the combination of limit learning machine (ELM) and support vector machine to carry out multi classification, and uses the classic KDD99 data set for experimental evaluation. In terms of the accuracy index, its effect reaches 95.75% [11]. From the above research results, it can be seen that when applying machine learning methods to solve the problem of network intrusion, scholars generally first use some efficient methods in data processing (such as sampling), and then use machine learning methods to extract features, construct modeling to realize classification or prediction, and output the final results. In terms of model selection, we can not only stack the basic models, but also adopt the strategy of improving the basic models. Of course, we can also combine the two, and its ultimate goal is to build an efficient model. However, today's network attack traffic is growing, the data dimension is increasing, the attack methods are constantly changing, and the severity of data imbalance is becoming more and more prominent. The traditional machine learning methods are also unable to resist these complex situations. First of all, its deficiency is the extraction of features. Manual extraction plays an important role in the process of extracting features. However, it is difficult to explore the internal features of data manually, which mainly depends on experience. It is unable to well distinguish the correlation and temporal and spatial characteristics of data, and it is difficult to explore the internal relations and laws of data. Furthermore, from the perspective of data volume, such methods are obviously inefficient in the face of big data. Therefore, the intrusion detection technology of traditional machine learning method is facing many difficulties, and new technologies and ideas need to be developed urgently.

2 Related Work

Long short term memory (LSTM) was proposed by Hochreiter [2] and others to better realize these functions. Each LSTM unit includes a memory cell, an input gate, a forget gate and an output gate. The typical structure is shown in the (Fig. 1).

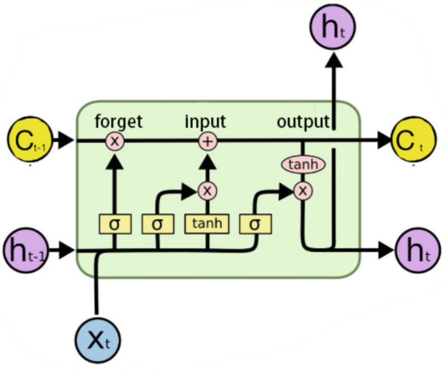

Fig. 1. Typical structure diagram of LSTM

The unit state determines the selection and updating of input data [3]. The updating of time correlation helps to classify and predict time series signals. Its principle and description are as follows:

Corresponding to the structure in the figure, the learnable weights of LSTM are input weights, recursive weights and bias. Matrices W, R and B are series of input weights, recursive weights and deviations of each component, respectively. The connection formula of these matrices is as follows:

$$W = \begin{bmatrix} W_i \\ W_f \\ W_g \\ W_o \end{bmatrix}, \ R = \begin{bmatrix} R_i \\ R_f \\ R_g \\ R_o \end{bmatrix}, \ b = \begin{bmatrix} b_i \\ b_f \\ b_g \\ b_o \end{bmatrix} \tag{1}$$

The cell state expression for time *t* is as follows:

$$h_t = o_t \otimes \sigma_c(c_t) \tag{2}$$

where σ_c represents the activation function. Generally, the LSTM layer function uses the hyperbolic tangent function (tanh) as the activation function, where $\otimes$ represents the element product of the vector.

Although the structure of forgetting gate in the structural diagram is concise, it plays a decisive role in discarding the input information. The expression is as follows:

$$f_t = \sigma_g\left(W_f x_t + R_f h_{t-1} + b_f\right) \tag{3}$$

where f_t represents the forgetting threshold, h_{t-1} is the last output information, and x_t is the input data information. The input data saves the useful data and forgets the useless data through the activation function σ_g.

The function of the input gate is to handle the input of the current sequence position, and its expression is as follows:

$$i_t = \sigma_g(W_i x_t + R_i h_{t-1} + b_i) \tag{4}$$

The function of the output gate is to determine the output value, which also depends on its state unit, and its expression is:

$$o_t = \sigma_g(W_o x_t + R_i h_{t-1} + b_o) \tag{5}$$

The function of the state unit is that the information of the past time runs directly on the whole chain, with only a small amount of linear interaction. Its expression is:

$$g_t = \sigma_c\big(W_g x_t + R_g h_{t-1} + b_g\big) \tag{6}$$

The process of network traffic data classification by LSTM is roughly as follows: first, convert the network traffic sequence after conversion to standardization into F and input it to the state unit, then filter and update F through the state unit to form a new state F1, the softmax classifier acts on the data output corresponding to the state in the previous stage, and finally output the training results.

3 Network Intrusion Detection Method Based on CNN and LSTM

The design of the model structure has a decisive impact on the experimental results. If pooling and convolution stacking are used in the classical convolutional neural network (CNN), it is easy to cause over fitting, and the extracted features are not necessarily representative and are very limited. In view of the above problems, this chapter adopts the method of CNN and long-term and short-term memory unit (LSTM) neural network, starts with the data characteristics, uses the time-series correlation of the detection data, adopts LSTM unit, compresses the network parameters and reduces the amount of calculation through CNN, and better extracts the temporal and spatial characteristics of the data and retains the correlation between the data through the combination of nonlinear modules. The super parameters of the ISOA optimization model are used, and the ReLU function is used as the activation function to make the neurons more robust, reduce the risk of over fitting, and build an efficient model.

The setting of model parameters is optimized through ISOA, and the optimization process is consistent with the previous chapter. Here we focus on the model structure of this chapter, which is shown in the figure below. The model processing steps are as follows: the first step is data preprocessing. The original NSL-KDD data sample contains 41 features. We expand it to multi-dimensional by transposing the matrix, and then input it to the convolutional neural network in the subsequent module for processing. The second step is feature extraction. The main process of feature extraction is shown in the figure below. Each small module plays a different role. Convolution and Max pooling play the role of feature purification. In the model, conv is convolution operation, and then the maximum pooling operation is carried out. The feature is down sampled to a certain size (according to the setting of convolution kernel) as the input of the next step. Batch normalization is mainly used to adjust the data relationship and reduce the impact of data covariance offset. On the whole, batch normalization, activation function and dropout algorithm are nested in each convolution layer, cooperate with each other and constantly adjust and optimize. Of course, the model of this step is stackable, and the number of stacking layers and settings depend on parameter optimization and experimental experience. Step 3: LSTM layer. The convolution and pooling in the previous stage are aimed at the spatial characteristics of the input data, which can maintain the internal feature correlation of the data. However, as we mentioned earlier, the input data has a great correlation in time sequence, and the adoption of LSTM gives the whole further feature extraction ability, fully exploring the main internal space-time correlation of the data. The fourth step is classified output. Finally, through the full connection layer connection, the data dimension transformation is completed, and the softmax classifier is used to detect different attack types and identify normal traffic, so as to complete the output of multi classification results (Fig. 2).

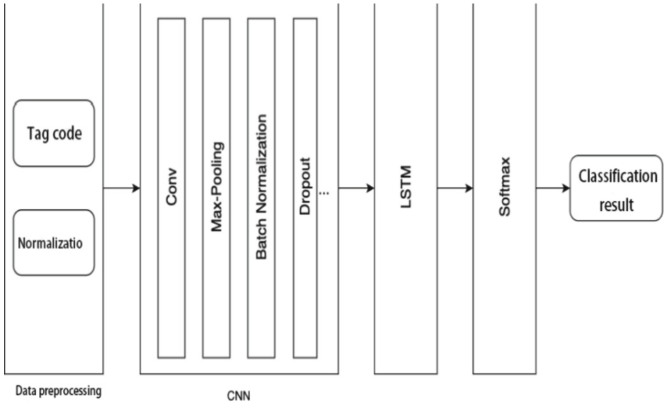

Fig. 2. Structure diagram of CNN + LSTM intrusion detection model

The CNN + LSTM network intrusion detection model proposed in this section takes advantage of the advantages of CNN spatial feature extraction and LSTM time series feature extraction. The combination of the two can fully extract the features of the data. In the parameter setting of the model, we use ISOA to optimize the parameters of the model, so that the parameter setting of the model is more reasonable and the performance of the model is stronger. In terms of the overall steps, after optimizing and setting the parameters, we first preprocess the data, including the conversion of label symbols and normalization processing, and then convert the processed data into the data type received by the neural network and send it to the network model for repeated iterative training. In each round of training, the same amount of data will be randomly selected and put into our model for training. The local features will be extracted and stored through the convolution layer, and then further feature filtering and dimensionality reduction will be carried out in the pooling layer. Then, the BN algorithm and Adam optimizer are used to update the data distribution and network weight, and dropout is used to reduce the impact of over fitting. The whole process can be combined and stacked. The basic features of the original data are continuously extracted and transformed under the action of CNN, LSTM and many optimization methods, forming higher-level abstract features, retaining the temporal and spatial correlation of the data, and feature extraction is more sufficient. The final category judgment is output through the softmax layer, the overall loss value will be calculated, and can be iteratively optimized through back propagation.

The weight and bias of network parameters at each layer are constantly updated, the loss value is continuously reduced, the model is more robust, and the effect obtained in the test set is better.

4 Experiment and Analysis

4.1 Data Sets

International Knowledge Discovery and Data Mining Competition (KDD cup) is the highest level international competition organized by ACM data exploration and data discovery expert group (SIGKDD). It has absolute authority in the field of data retrieval research. The English name of KDD is knowledge discovery and data mining, that is, knowledge discovery and data retrieval competition. The theme of KDD competition in 1999 is computer network intrusion detection. After years of development and verification, KDD99 data set has become an authoritative standard for evaluating performance in network intrusion detection. Many scientific research and experiments use KDD99 as dataset [4]. The KDD99 data set was constructed based on the data captured by MIT Lincoln Laboratory in the DARPA ‘ 98 evaluation program. It includes 7-week training data and 2-week test data. These data are actually network packets, which are related to network connection. A network connection definition is: the sequence of a TCP packet from the beginning to the end, and the information data transmitted from the source IP address to the target IP address under predefined protocols (such as TCP and UDP). These data include 39 types of attacks, which can be divided into four categories, as described below (Table 1):

Table 1. Dataset attack type

Type of attack	Annotation
DOS	Denial of service, the purpose of which is to prevent users from accessing services normally, such as smurf flood, ping flood, syn flood, teardrop attacks, http halfconnect, dns flood, etc.
R2L	Unauthorized remote access, such as guessing password, ftp write, imap, multihop, phf, worm, etc.
U2R	Unauthorized access to local administrator privileges, such as buffer overflow, perl, load module, root kit, sql attack, etc.
PROBING	Monitoring or other probing activities such as portsweep, ipsweep, satan, mscan, saint, etc.

The KDD99 dataset consists of a total of 5 million records, each of which has 41 features, examples of which are as follows (Table 2):

Among them, the first 41 features can be divided into four types of attributes, and the last one is label. There are 42 attributes in total. The basic contents of various attributes are as follows: 1. Basic features of TCP connection (features 1–9, 9 in total). These

Table 2. Sample dataset

Number	Sample
1	0,tcp,http,SF,215,45076,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,1,1,0.00,0.00,0.00,0.00,1.00,0.00,0.00,0,0,0.00,0.00,0.00,0.00,0.00,0.00,0.00,0.00,normal
2	0, tcp, private, REJ, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 38, 1, 0.00, 0.00, 1.00, 1.00, 0.03, 0.55, 0.00, 208, 1, 0.00, 0.11, 0.18, 0.00, 0.01, 0.00, 0.42, 1.00, portsweep
3	2, tcp, smtp, SF, 1684, 363, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0.00, 0.00, 0.00, 0.00, 1.00, 0.00, 0.00, 104, 66, 0.63, 0.03, 0.01, 0.00, 0.00, 0.00, 0.00, 0.00, normal
4	0, tcp, smtp, SF, 787, 329, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0.00, 0.00, 0.00, 0.00, 1.00, 0.00, 0.00, 76, 117, 0.49, 0.08, 0.01, 0.02, 0.00, 0.00, 0.00, 0.00, normal

attributes include the protocol type of the current connection (TCP / UDP / ICMP), the network service type of the destination host, the number of urgent packets, the connection duration, the number of bytes from the source host to the destination host, the status bit indicating whether the connection is normal or not, etc. they are an accurate description of a TCP connection. 2. The content characteristics of TCP connection (No. 10 ~ 22 features, 13 kinds in total). For attacks such as r2l and u2r, due to the particularity of their attacks, they are not frequently recorded in the system like DoS attacks, so their characteristics look no different from normal TCP connection, which is also the main factor for the imbalance of this data set. In order to ensure that such attacks can be detected, the experimenters of KDD99 only extract some feature content from the data content, such as the number of unsuccessful login verification and the frequency of root user access. It is only possible to reflect the intrusion behavior, but it may not be able to detect it accurately. 3. The statistical characteristics of network traffic on time series (No. 23 ~ 31 features, a total of 9 kinds) express the temporal correlation of various attack modes. Therefore, a period of time between connections can be divided into independent time slices, so as to calculate the relationship between current connections and connections in earlier time slices. This is also an important reason for us to adopt the cyclic neural network model that can make use of time correlation. 4. Host based network traffic statistics (features 32–41, 10 in total), that is, the aforementioned time slice based traffic statistics, is the relationship statistics between the current connection and the previous time slice connection. The fixed time period in the data set is 2 s.

In 2009, Tavalllaee et al. proposed a revised version of the KDD99 dataset and named it NSL-KDD [4]. The NSL-KDD dataset overcomes some shortcomings of the KDD99 dataset. For example, NSL-KDD deletes some of the duplicate records in KDD99, reducing the impact of some frequent records on the experimental results; the number of entries in the dataset It has been simplified, and a relatively uniform dataset of different proportions is artificially selected, which improves the comparison and fairness of the research results. The filtered records contain different classification difficulty levels, and the number is more balanced, which makes the evaluation more effective and fair. The total number of records is kept within a reasonable size, allowing the algorithm to be applied to the entire dataset, rather than a randomly selected small subset, thus making it easier to compare different studies.

In this paper, the experimental part will use the improved data set NSL-KDD to evaluate the actual performance of the system. We use the most commonly used data files to test, so as to maintain the fairness of the comparison as much as possible.

4.2 Experimental Results and Analysis

In order to objectively evaluate the performance of the model, after sorting out the data of NSL-KDD data set, we randomly selected 80% of the data as the training set and the remaining 20% of the data as the test set. The number of runs is 20, and the experimental results are as follows (Table 3):

It can be seen from the above results that in the NSL-KDD data set, the result of ISOA + CNN + LSTM is the best. In terms of average accuracy, the performance ranking of the model is as follows: ISOA + CNN + LSTM > CNN + LSTM > CNN > LSTM. Among them, the average accuracy of ISOA + CNN + LSTM is 98.43%,

Table 3. The experimental results

	Experimental results of nsl-kdd dataset				
Dataset	Algorithm model	Average accuracy(ACC)	Best accuracy (BACC)	Detection rate(DR)	False alarm rate (FAR)
NSL-KDD	LSTM	94.68%	96.19%	91.08%	6.58%
	CNN	96.53%	97.64%	94.22%	3.90%
	CNN + LSTM	97.92%	98.47%	96.36%	4.05%
	ISOA + CNN + LSTM	98.43%	99.17%	98.74%	1.49%

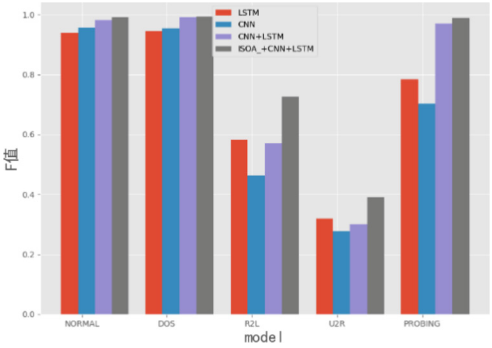

Fig. 3. NSL-KDD dataset F value result

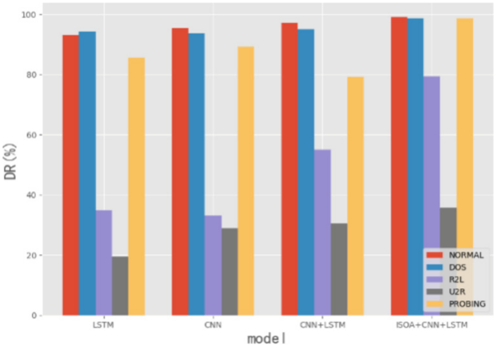

Fig. 4. NSL-KDD dataset DR value result

the best accuracy is 99.17%, the average accuracy is nearly 0.51% higher than CNN + LSTM, and 0.84% higher than the ISOA + Gru + MLP method proposed in Sect. 3, The detection rate is also the highest among these models, and the false alarm rate is as low as 1.49%, which is the lowest among the models proposed in Sects. 3 and 4, which also proves that the model we designed has excellent performance. As can be seen from Figs. 3 and 4, in terms of attack types, the comparison of F values shows that the detection effect of DoS attack and pro attack is significantly better than that of R2l and U2r attack. Through analysis, this is also in line with the characteristics of uneven data in our NSL-KDD data set (Fig. 5).

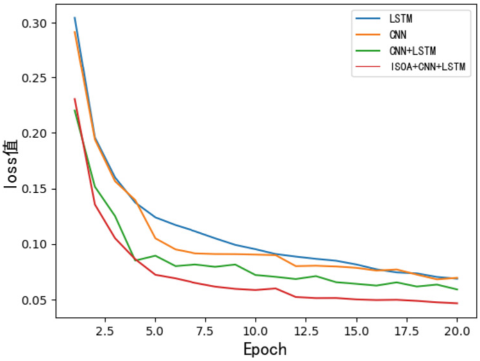

Fig. 5. Variation of loss value in different models

Figure 3 shows the variation curve of the average loss value of the experiment, and compares the convergence of different algorithms. It can be seen that the proposed ISOA + CNN + LSTM model has fast convergence speed and lower loss value. The model converges almost after 13 epochs. Combined with the previous results, we can see that using the ISOA + CNN + LSTM model not only has advantages in accuracy, but also has significantly faster convergence speed.

In order to better analyze the experimental results, we compared the experimental results with some existing research results ("–" means the indicator is not explained in the paper). The results are as follows:

Table 4. NSL-KDD dataset performance comparison

Model	Average accuracy(ACC)	Detection rate(DR)	False alarm rate (FAR)
VELM [12]	97.58%	97.69%	2.22%
DBN + SVM [13]	92.87%	–	–
SVM + ELM [14]	94.85%	–	–
LSTM + MLP [14]	96.41%	97.33%	3.24%
ISOA + GRU + MLP(this article)	97.59%	97.94%	2.01%
ISOA + CNN + LSTM(this article)	98.43%	98.74%	1.49%

It can be seen from Table 4 that ISOA + CNN + LSTM performs well on NSL-KDD data set, achieving the lowest false alarm rate (FPR): 1.49% and the highest accuracy (ACC): 98.43%.

It is worth noting that the above comparison is only a reference rather than an absolute distinction between advantages and disadvantages. In fact, different network intrusion detection methods respond differently to different kinds of intrusion, so it is difficult to find a method that can achieve the best performance in any case. In addition, due to slight differences in evaluation methods, such as different processing methods of data sets and different random sampling, the final results will also be different. Nevertheless, compared with other recent works, our proposed method still has clear advantages in accuracy, detection rate, convergence speed and so on.

5 Conclusion

In order to better extract the temporal and spatial characteristics of the data set, this chapter uses the combined model based on ISOA, CNN and LSTM to detect network intrusion. Firstly, the appropriate model parameter setting is obtained through ISOA, and then in the process of model training, make full use of the locality and spatial correlation of data features, use convolution neural network to extract and filter features, retain the main features of data and eliminate redundant information, so as to reduce the amount of data calculation and avoid over fitting. Then the LSTM layer is added. Using the characteristics of mining and using the temporal correlation of data, LSTM further extracts the high-level abstract features in the data, modifies the distribution of data and updates the weight of the network through BN and Adam algorithms, and continuously optimizes the iterative model through back propagation, so as to train the model more powerful. Finally, the NSL-KDD data set is used to verify the performance of our combined model. The comparative experiment shows that the detection model proposed in this chapter has excellent performance indicators in the experimental process, in which the highest detection accuracy is 99.17% and the average accuracy is 98.43%, which has obvious advantages over other research methods in recent years.

Acknowledgment. This work is funded by the National Natural Science Foundation of China under Grant No. 61772180, the Key R & D plan of Hubei Province(2020BHB004, 2020BAB012).

References

1. Ots, K.: Network Security Azure Security Handbook, pp. 59–76. Apress, Berkeley, CA (2021)
2. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)
3. Xiong, G., Przystupa, K., Teng, Y., et al.: Online measurement error detection for the electronictransformer in a smart grid. Energies **14**(12), 3551 (2021)
4. Sommer, R., Paxson, V.: Outside the closed world: on using machine learning for network intrusion detection. In: 2010 IEEE Symposium on Security and Privacy, pp. 305–316. IEEE (2010)

5. Xu, H., Przystupa, K., Fang, C., Marciniak, A., et al.: A combination strategy of feature selection based on an integrated optimization algorithm and weighted k-nearest neighbor to improve the performance of network intrusion detection. Electronics **9**(8), 1206 (2020)
6. Song, W., Beshley, M., Przystupa, K., et al.: A software deep packet inspection system for network traffic analysis and anomaly detection. Sensors **20**(6), 1637 (2020)
7. Kayacik, H.G., Zincir-Heywood, A.N., Heywood, M.I.: Selecting features for intrusion detection: a feature relevance analysis on KDD 99 intrusion detection datasets. In: Proceedings of the Third Annual Conference on Privacy, Security and Trust, vol. 94, pp. 1723–1722 (2005)
8. Saleh, A.I., Talaat, F.M., Labib, L.M.: A hybrid intrusion detection system (HIDS) based on prioritized k-nearest neighbors and optimized SVM classifiers. Artif. Intell. Rev. **51**(3), 403–443 (2019)
9. Chen, S., Peng, M., Xiong, H., Yu, X.: SVM intrusion detection model based on compressed sampling. J. Electr. Comput. Eng. **2016**, 1–6 (2016)
10. De La Hoz, E., Ortiz, A., Ortega, J., Prieto, B.: PCA filtering and probabilistic SOM for network intrusion detection. Neurocomputing **164**, 71–81 (2015)
11. Al-Yaseen, W.L., Othman, Z.A., Nazri, M.Z.A.: Multi-level hybrid support vector machine and extreme learning machine based on modified K-means for intrusion detection system. Expert Syst. Appl. **67**, 296–303 (2017)
12. Xu, C., Shen, J., Du, X., Zhang, F.: An intrusion detection system using a deep neural network with gated recurrent units. IEEE Access **6**, 48697–48707 (2018)
13. Shen, Y., Zheng, K., Wu, C., et al.: An ensemble method based on selection using bat algorithm for intrusion detection. Comput. J. **61**(4), 526–538 (2018)
14. Kun-peng, Y.: An intrusion detection model based on deep belief networks. Mod. Comput. **02**, 10–14 (2015)