IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON ARTIFICIAL INTELLIGENCE FOR PHYSICAL-LAYER WIRELESS COMMUNICATIONS

Received February 22, 2019, accepted March 11, 2019, date of publication March 20, 2019, date of current version April 5, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2905754

A Method for Guaranteeing Wireless Communication Based on a Combination of Deep and Shallow Learning

QIAO TIAN, JINGMEI LI, AND HAIBO LIU

College of Computer Science and Technology, Harbin Engineering University, Harbin 150001, China

Corresponding author: Haibo Liu (liuhaibo@hrbeu.edu.cn)

This work was supported by the Natural Science Foundation of Heilongjiang Province of China under Grant F2018011.

ABSTRACT Wireless communication has changed and improved people's lives and society, especially with the arrival of the Internet of Things (IoT) era. Despite the maturity of wireless communication, the security issue of communication remains the most stubborn and troublesome problem due to the increasingly complex and large amounts of data. An intrusion detection system is the guarantee of secure communication. However, variable protocols and drastic growth in data volume make intrusion detection a difficult task. In this paper, we proposed a framework of anomaly-based network intrusion detection system to finish the detection job. First, UNSW-NB15 is selected as the research object. Based on this new dataset, we built a detection model combining a deep learning method and a shallow learning approach. The former one is a deep auto-encoder used for feature learning, which can discover important representations of data and accelerate detection. The latter one is a powerful support vector machine (SVM), where the artificial bee colony (ABC) algorithm is used to find optimal parameters for SVM with five-fold cross validation (5FCV). Various experiments are conducted and the simulation results prove that the proposed method performs quite better than some of state-of-the-art intrusion detection approaches, including the method based on the principal component analysis (PCA) and some other machine learning strategies.

INDEX TERMS Wireless communication, big data, intrusion detection system, deep auto-encoder.

ACRONYMS AND ABBREVIATIONS

5FCV	5-fold cross validation
ABC	Artificial Bee Colony
ARM	Association Rule Mining
EM	Expectation Maximization
FAR	False Alarm Rate
MSE	Mean Squared Error
NIDS	Network Intrusion Detection System
RBF	Radial Basis Function
ReLU	Rectified Linear Unit
ROC	Receiver Operating Characteristic
SI	Swarm Intelligence

I. INTRODUCTION

Upgrade and cooperation of different techniques such as sensors, positioning system, artificial intelligence, pervasive computing, broadband access, and cloud services have given birth to a new technology named the Internet of Things [1], [2]. An important foundation of IoT technology is wireless communication. Wireless communication is the bridge for information exchange [3]. The development of wireless communication has greatly improved the improvement of other technology [4]. The Internet of things can provide a wide range of services in different domains such as industrial process, home intelligence, intelligent transportation, health-care, government work, environmental monitoring without human intervention [5], [6]. It has a huge impact on people's production and lifestyle and creates significant benefits and convenience.

However, due to the increase in the number and variety of wireless communications, the channel environment has become more and more complex [7]–[9]. The normal communication is often mixed with a lot of bad data [10]. Owing to self-characteristics including complex structure and large volume, IoT network is easy to suffer from various

The associate editor coordinating the review of this manuscript and approving it for publication was Guan Gui.

intrusion attacks which aim to destroy the network [11]. For more widespread adoption of IoT, security issues should be addressed perfectly enough so that users might have confidence in its ability to protect information. The physics-based method is one of the basic aspects of security [12]–[14]. Some research is about the physical security [15]. And Network Intrusion Detection System is another one of the most effective and most primary protection tools. Generally speaking, there are two types of NIDS. The misuse-based approach uses the data stored in its database to accurately perform specific attack types of detection [16]. But it is hard to adapt to the current network environment because of the rapidly growing cases and types of vulnerability attacks [17]–[19]. Another one is the anomaly-based method. It usually judges an attack by comparing the current activities of a system with a historical normal behavior profile [20]. It will generate a warning whenever attack occurs. In fact, the latter method has an important application prospect in the environment where the type of attack is variable. And it is more suitable to detect unknown malicious cases [21].

In the existing state-of-the-art in anomaly-based NIDS, performances regarding the feasibility and reliability are receiving substantial interest [22]. High false alarm rate still restricts the application of this method. Main limitations include the drastic growth volume of network data and different protocols [23]. Inspired by the success of deep learning in the classification problem, we employ an auto-encoder model to carry the task of representation learning by leveraging the abundance of training data. It has many advantages including the ability of dimensionality reduction and capability of dealing with imbalanced data [24]. Above characteristics are of great help to the building of intrusion detection models.

The contribution of our work contains:

1) A deep auto-encoder technique is used for the feature learning process, which can provide service to facilitate a comprehensive analysis of data and reduce data dimensionality so as to accelerate the identification process. The result will be compared with that using PCA method.

2) We adopt a deep learning method combined with a shallow learning engine to construct a NIDS. Here, we adopt deep auto-encoder to reduce the dimension of the dataset. The compressed data will input into a SVM classifier, where the Artificial bee colony algorithm will optimize the parameter within SVM classifier.

The rest of this paper is organized as follows. Section II provides a summary of related work in NIDS. Section III describes some basic theory of deep auto-encoder, Support Vector Machine and Artificial Bee Colony algorithm. In Section IV, the details of the proposed work will be explained. Section V gives a simulation and we conclude in Section VI.

II. RELATED WORK

Over recent decades, many scholars have studied intrusion detection problems. Dataset is of great importance for the practice of the Intrusion Detection System. In this paper, a new dataset UNSW-NB15 is chosen to do research. So the following related work is related to this dataset.

In [25], a hybrid method is used to implement IDS. The value of features' center point is first calculated and Association Rule Mining (ARM) is used to select subset less than 11 attributes. The Expectation Maximization (EM) clustering, Logistic Regression, and Naïve Bayes are used as the decision engine for intrusion detection with the accuracy of 77.2%, 83.0% and 79.5% respectively. The results can reflect the complexity of the data and the difficulty of detection. Moustafa and Slay [26] compared UNSW-NB15 with KDD'99 and demonstrated the complexity and effectiveness of their proposed dataset in three different perspectives, which can show that UNSW-NB15 can represent modern attack more properly. After that, five different classifiers are used to finish the model building task but the highest accuracy can only reach 85.56% while the FAR is 15.78%. In [27], a framework that contains 5 different feature-selection strategies including Wrapper, Filtering, Merged and Union methods are proposed. Besides, J48 classifier and Naïve classifier are combined with the above five strategies separately. The result shows that J48 with GR filter performs best with 88% accuracy. Bhamare *et al.* [28] adopted SVM classifiers with different kernel function to finish detection job. However, the result is not that good. In [29], an Association Rule Mining technique helps generate the strongest feature subset. For each attack type, a subset is selected. And a three-layer IDS framework is proposed where Naive Bayes and EM clustering are used as the decision engine to classify. In [30], a two-stage classifier method is proposed. At first stage, according to protocol type, classify incoming data into three categories that is TCP, UDP and other. Based on the previous work, use Information Gain method to select features which will create 3 different subsets. Finally, REPTree is used as decision machine and the accuracy is 88.95%. But test dataset is not the whole UNSW-NB15, which makes the result incomplete.

In summary, the current IDS model has the problem that the detection accuracy is not very high, but the false alarm rate is high. Our work is to improve the performance of IDS.

III. BACKGROUND

A. DEEP AUTO-ENCODER

An auto-encoder is an unsupervised neural network, which has an input layer, an output layer (with the same dimension as the input layer) and a hidden layer, shown in Fig. 1. Deep auto-encoder is a neural network with more than one hidden layer. It extracts the inner relationship of data through learning the optimal network parameters which will result in output similar to input as much as possible [31], [32].

A dataset $X = (x_1, x_2, \ldots, x_m)$ has m samples and x_i is high-dimensional. The encoding layer maps the raw data to a hidden layer representation as given in (1):

$$h_i = f_\theta(x_i) = s(Wx_i + b) \tag{1}$$

where $\theta = \{W, b\}$ is the mapping parameters. In this phase, the data is transformed in some form, and the result is that

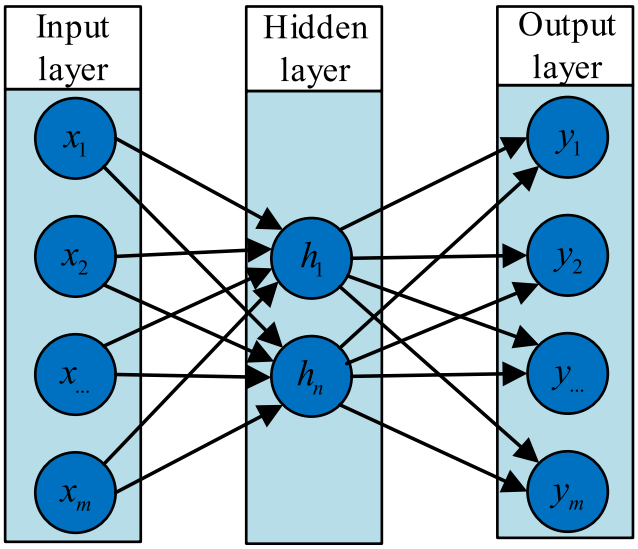


FIGURE 1. A simple structure of auto-encoder.

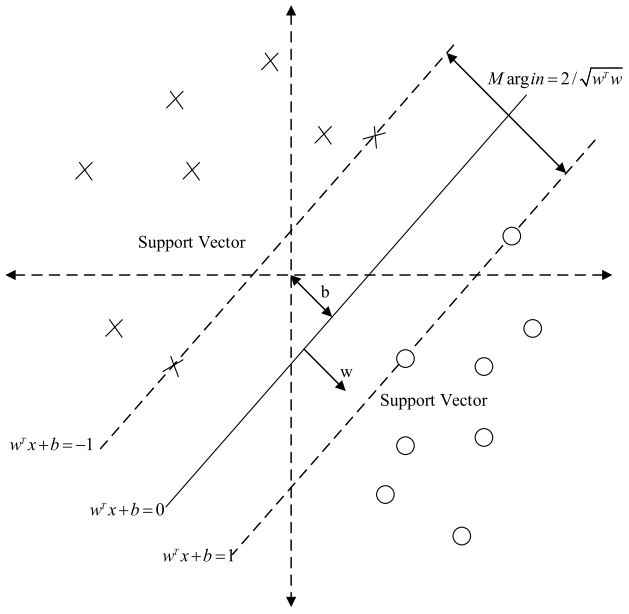


FIGURE 2. Basic concept of support vector machine.

the dimension can be reduced. s is the activity function. In this paper, we adopt sigmoid function and ReLU function as shown in (2) and (3):

$$sigmoid(t) = \frac{1}{1+\exp^{-t}} \tag{2}$$

$$\mathrm{ReLU}(t) = \begin{cases} t, & if\ \mathrm{t} > 0 \\ 0, & if\ \mathrm{t} \le 0 \end{cases} \tag{3}$$

The process of model training will narrow the gap between output and input, which is a key factor to evaluate a model. Loss function plays such a role to penalize the dissimilarity between y and x. In this paper, Mean Squared Error (MSE), shown in (4), is adopted

$$\theta(\omega) = \max_{\alpha_i \ge 0} L(\omega, b, \alpha) = \frac{1}{2}\|\omega\|^2$$

as a loss function during the feature reduction process.

$$L(x, y) = \frac{1}{m}\sum_{i=1}^{m}\|x_i - y_i\|^2 \tag{4}$$

B. SUPPORT VECTOR MACHINE

Support vector machine, shown in Fig. 2, is one of the strongest and most powerful machine learning algorithms [33]. It can find a balance between model complexity and classification ability given limited sample information [34]. SVM has many advantages compared with other machine learning methods. It can overcome the impact of noise and work without any prior knowledge [35].

Consider a perfectly linearly separable problem with the dataset $\{X, Y\} = \{(x_1, y_1), \ldots, (x_N, y_N)\}$, where $x_i \in R^n$ is a n dimensional vector, $y_i \in \{-1, 1\}$ is label and N is the number of samples. A function named separating hyperplane can be written as $f(x) = w \cdot x + b$, where w is the weight vector and b is the bias. A new object x can be classified as follow:

$$\begin{cases} (\omega \cdot x) + b \ge 0, & y = +1 \\ (\omega \cdot x) + b < 0, & y = -1 \end{cases} \tag{5}$$

So from another perspective, classification can be regarded as a process of finding the optimal hyperplane through quadratic programming, that is:

$$\min(\frac{1}{2}\|\omega\|^2), \quad s.t.\ y_i\,(\omega \cdot x_i + b) \ge 1, \quad i = 1, 2, \ldots, N \tag{6}$$

This constrained optimization problem is solved using the following Lagrangian form:

$$L\,(\omega, b, \alpha) = \frac{1}{2}\|\omega\|^2 - \sum_{i=1}^{N}\alpha_i(y_i(\omega^T \cdot x + b) - 1)$$
$$s.t.\ y_i(\omega^T \cdot x + b) \ge 1 \tag{7}$$

where α_i is the Lagrange factor and $\alpha_i \ge 0$. When all constraints are met, The optimal solution is:

$$\theta(\omega) = \max_{\alpha_i \ge 0} L(\omega, b, \alpha) = \frac{1}{2}\|\omega\|^2 \tag{8}$$

The nonnegative relaxation factor ξ_i and castigate factor C are introduced to solve the linearly nonseparable problem, where the constraints become:

$$\min(\frac{1}{2}\|\omega\|^2) + C\sum_{i=1}^{N}\xi_i$$
$$s.t.\ y_i\,(\omega \cdot x_i + b) \ge 1 - \xi_i, \xi_i \ge 0, \quad i = 1, 2, \ldots, N \tag{9}$$

To solve this problem, kernel functions are used, such as Polynomial Kernel, Radial Basis Function (RBF) and so on. The use of kernel function is to map data from a

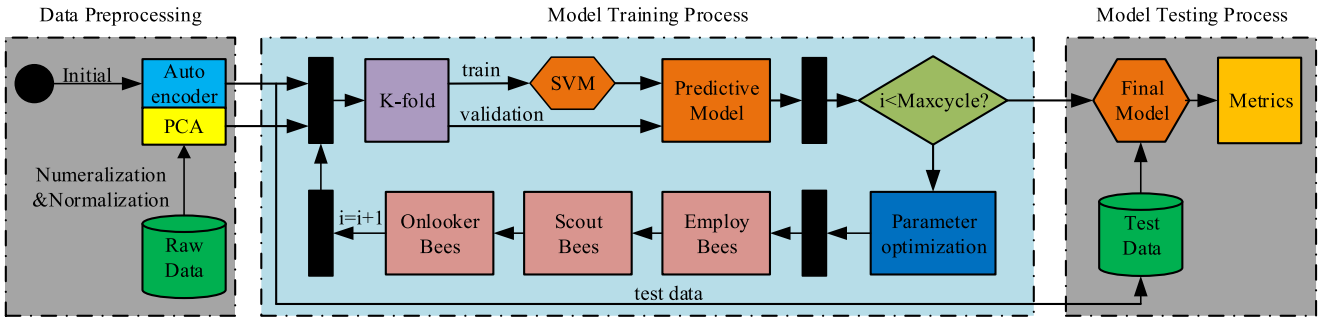


FIGURE 3. Proposed framework for anomaly intrusion detection.

low dimension to a high dimension, and this mapping can make nonlinear classification linearly separable. In this paper, we adopt the RBF kernel as below:

$$K(x, x_i) = \exp(-\frac{\|x - x_i\|^2}{2\sigma^2}) \tag{10}$$

So the SVM with RBF kernel has two parameters, castigate factor C and kernel parameter σ. The combination of two parameters influences the performance of SVM, and many methods are adopted to get the optimal values.

C. ARTIFICIAL BEE COLONY

The Artificial Bee Colony, proposed by Karaboga, is one of the most popular Swarm Intelligence (SI) approaches [36]. The inspiration for this algorithm comes from the behavior of the honey bee.

In nature, there are three different types of individuals in bee colony: employ bees, scout bees, and onlooker bees. The process of using the artificial bee colony algorithm to find the optimal value is like bees collecting honey as follows:

First, randomly generate an initial population as food sources.

$$x_{m,i} = l_i + rand(0, 1) \times (u_i - l_i) \tag{11}$$

where $x_{m,i}$ means food source m. u_i and l_i respectively represent the upper and lower limits of $x_{m,i}$. And number i is an index of parameters to be optimized.

Next, the employed bees mine the food source $v_{m,i}$.

$$v_{m,i} = x_{m,i} + \phi_{m,i}(x_{m,i} - x_{k,i}) \tag{12}$$

where $x_{k,i}$ is a randomly selected food source. And $\phi_{m,i}$ is a random value between -1 and 1. At the same time, calculate the fitness of v_m and compare it with that of old x_m.

$$fit(x_m) = \begin{cases} \dfrac{1}{1 + f(x_m)}, & f(x_m) \geq 0 \\ 1 + abs(f(x_m)), & f(x_m) < 0 \end{cases} \tag{13}$$

where $f(x_m)$ is the objective function value of solution x_m.

Then, the employed bee will share the best solution with the onlooker bees who will evaluate the nectar information and pick a solution for further work with a probability p_i.

$$p_i = \frac{fit(x_m)}{\sum_{m=1}^{SN} fit(x_m)} \tag{14}$$

For those selected food sources, the onlooker bees will search continuously using (12).

During the literation, if the solution stays still the same after specified trials, called 'limit', the employed bees could become scouts and the corresponding solution will be abandoned.

Finally, after a certain number of cycles, an optimal solution will be obtained.

IV. PROPOSED INTRUSION DETECTION SYSTEM

In this section, a framework for intrusion detection is proposed as shown in Fig. 3. We divide the model into three parts, data preprocessing, model training process, and testing process.

In the first phase, raw data will successively experience feature numeralization, class numeralization, and feature normalization. After that, we respectively adopt PCA and deep auto-encoder model to learn a better representation of original data. Besides, the preprocessed results will be taken apart into training set and testing set.

In the next part, a strong classifier is used to classify and the optimal parameters of the SVM will be found with the help of the ABC algorithm through 5-fold cross-validation.

Finally, input the test set into the optimal model and evaluate the performance using relevant metrics. The details of the proposed work are as follows:

A. DATASET

A suitable data set is critical for machine learning problems. A complex dataset that reflects the state of contemporary cyber attacks is obviously very important [37].

A lot of work have been done on the dataset of network intrusion detection. Among them, KDD'99 and its updated version NSL-KDD are classical [38], [39]. However, both of them can not represent the network nowadays, because of redundancy and data imbalance as well as old types of attacks. A complex and balanced data set is especially important for

TABLE 1. UNSW-NB15 dataset.

Number	Description	Number	Description
1	dur	23	dwin
2	proto	24	tcprtt
3	service	25	synack
4	state	26	ackdat
5	spkts	27	smean
6	dpkts	28	dmean
7	sbytes	29	trans_depth
8	dbytes	30	response_body_len
9	rate	31	ct_srv_src
10	sttl	32	ct_state_ttl
11	dttl	33	ct_dst_ltm
12	sload	34	ct_src_dport_ltm
13	dload	35	ct_dst_sport_ltm
14	sloss	36	ct_dst_src_ltm
15	dloss	37	is_ftp_login
16	sinpkt	38	ct_ftp_cmd
17	dinpkt	39	ct_flw_http_mthd
18	sjit	40	ct_src_ltm
19	djit	41	ct_srv_dst
20	swin	42	is_sm_ips_ports
21	stcpb	43	attack_cat
22	dtcpb	44	label

intrusion detection. A comprehensive network-based dataset, UNSW-NB15, created by Moustafa et al, can be a microcosm of the modern network.

UNSW-NB15 data set has the following advantages such as rich data types. It has 9 different kinds of attack type, which is more in line with the characteristics of modern network traffic data; Besides, data distribution in the training and testing dataset is similar. In training dataset, the normal percentage is 45% and the anomaly is 55%. In the testing set, the normal percentage is 32% and abnormal is 68%.

B. DATA PREPROCESSING

In general, dataset contains too many features that are not only redundant and noisy but also harmful to the detection accuracy and running speed.

The more features in the dataset, the more difficult the problem is to solve in modeling and testing. There are 42 features in UNSW-NB15. Obviously, it is necessary to streamline it to simplify the construction of the model and increase the speed of the operation.

In this paper, data preprocessing includes the following parts.

The first one is numerailzation. Convert the nominal attribute values into continuous numerical values. For example, the feature *state* has 4 different nominal values, replace them with 1, 2, 3, 4 respectively.

Then, get standard format attributes. The purpose of normalization is to eliminate the impact of different data units. Use a Min-Max technique to normalize the data denoted in (15) as:

$$x^* = \frac{x - \min}{\max - \min} \tag{15}$$

Finally, use the deep auto-encoder and PCA algorithm to finish the feature reduction job. The proposed deep auto-encoder architecture is shown as Fig. 4.

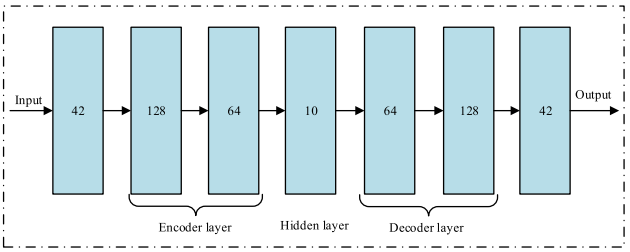


FIGURE 4. The architecture of proposed deep auto-encoder.

UNSW-NB15 has 42 features, so the number of neurons in the input and output layer is 42 respectively. To obtain satisfactory reconstruction performance, multiple groups of encoder and decoder layer were used for the experiment. As for the hidden layer, we will illustrate below. Besides, activity function has a certain effect on the performance of the model, which is used in encoder and decoder layer. As we all know, ReLU function has the advantage of solving the problem of convergence, which makes it easier to optimize the neural network. While the sigmoid function has good performance on the expression of the output layer. In this paper, the last decoder layer adopts sigmoid function and the rest employ ReLU function.

C. OPTIMIZATION OF RBF KERNEL PARAMETERS

Different parameters will influence the performance of the same classifier. At this stage, in order to obtain a better training model, the Artificial Bee Colony algorithm is used to optimize the parameters in the SVM model. The process is shown in Fig. 5.

In this paper, the ABC algorithm adopted is improved by Karaboga and Gorkemli [40], where the behavior of onlooker bees is different from that in standard ABC as shown below:

$$v_{N_m,i}^{best} = x_{N_m,i}^{best} + \phi_{m,i}(x_{N_m,i}^{best} - x_{k,i}) \tag{16}$$

$x_{N_m}^{best}$ will be obtained within the neighborhood of source m. To define a neighborhood for source x_m, mean Euclidean distance md_m denoted in (17) is used:

$$md_m = \frac{\sum_{j=1}^{SN} d(m,j)}{SN - 1} \tag{17}$$

If the distance between solution j and m is less than md_m, then j is regarded as a neighbor of m.

$$fit(x_{N_m}^{best}) = \max(fit(x_{N_m}^1), fit(x_{N_m}^2), \ldots, fit(x_{N_m}^S)) \tag{18}$$

Here, the fitness function is the average accuracy of the 5-fold cross validation for SVM classifier. After calculate each fitness, select the biggest one as the best one. Then stop the program when the cycle reaches the specified iterations or the result can reach a quite high score.

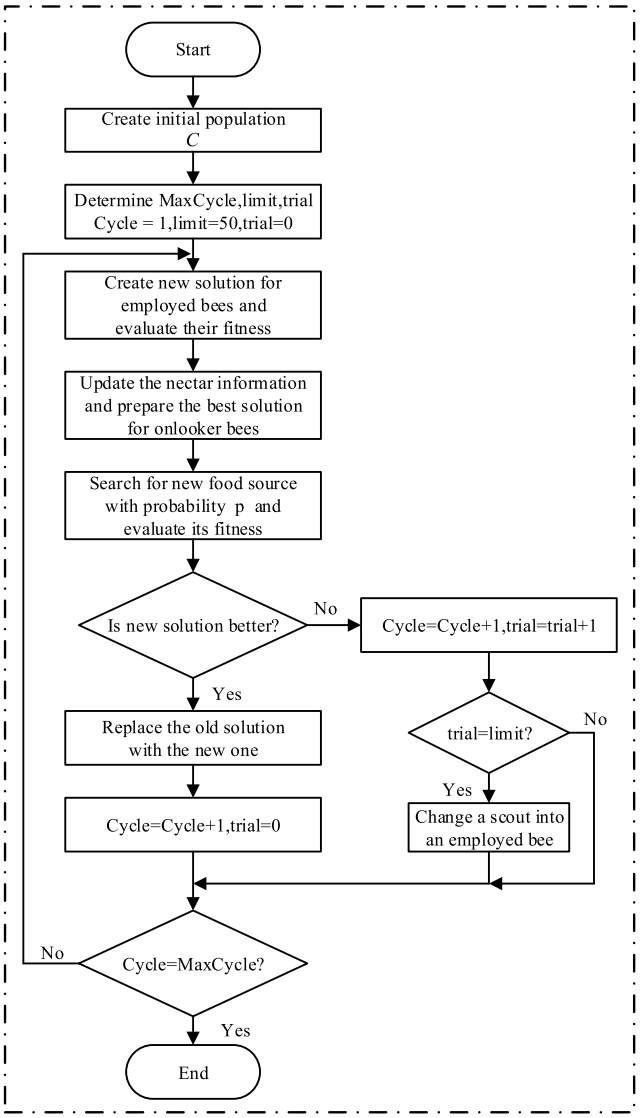


FIGURE 5. Framework for parameter optimization of SVM using the ABC algorithm.

V. SIMULATION

In this section, the UNSW-NB15 dataset is selected to do the experiment and we adopt different metrics to evaluate the performance of the model.

In the data preprocessing stage, we first use the PCA method to deal with normalized data as a benchmark for the following test. We will get a cumulative principal component contribution rate curve shown in the upper part of Fig. 6. It is obviously that the first 10 components can get a very high contribution rate. PCA is based on the theory of maximum variance which means that that the most easily distinguishable data in the projection direction will be separated. Generally speaking, when cumulative contribution rate reaches 90% or more, the features after dimensionality reduction can contain most information of the original data features. But it is not that proper to determine only according to the contribution rate.

In order to determine the size of the data dimension used in the subsequent classification experiments, we selected different numbers of principal components to perform experiments with different parameter $C = \{0.01, 0.1, 1, 10\}$. Then we get a detection rate curve in the lower half of Fig. 6.

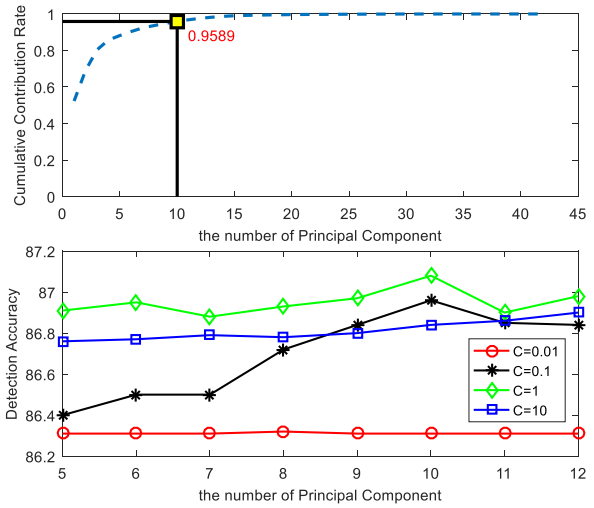


FIGURE 6. Cumulative contribution rate and detection accuracy with different number of PC under different *C*.

With the help of the experiment result, we can know that 10 dimensions might get a better performance. So we choose 10 dimensions to complete the next work. The structure of DAE indicates that the number of neurons in the hidden layer will be equal to the dimension after reduction. Therefore, the number of neurons in the hidden layer is also determined to be 10.

We separately put two dimensionality reduction data into SVM classifier and use ABC algorithm to find the optimal parameters. Finally, input the test data into the SVM model with the optimal parameter. Here to evaluate the performance of the proposed work, we adopt different indicators such as Accuracy, False Alarm Rate (FAR) and Receiver Operating Characteristic (ROC) curve.

For the anomaly detection problem, we first get the confusion matrix of test dataset shown in table 2, where TN is true negative while FN means false negative and TP/FP respectively stands for true positive and false positive.

TABLE 2. The confusion matrix of classification result.

test / predict	Normal	Attack
Normal	TN = 52742	FN = 14930
Attack	FP = 3258	TP = 104411

With the definition above, to evaluate the performance of a classifier, we adopt metrics such as the False Alarm Rate and the Accuracy.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{19}$$

$$FAR = \frac{1}{2}\left(\frac{FP}{FP + TN} + \frac{FN}{FN + TP}\right) \quad (20)$$

The comparison results of classification accuracy and FAR between the proposed deep learning framework and those in [27] and [28] as well as the PCA-based method are separately shown in Fig. 7. It can be seen from Fig. 7 that the accuracy in the proposed work is the highest and the FAR is the lowest.

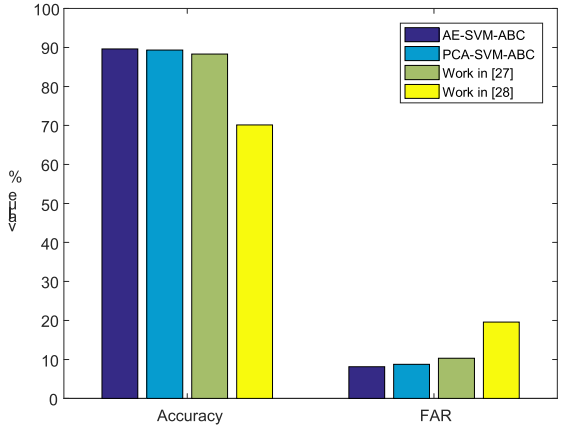


FIGURE 7. Comparison of accuracy and FAR.

Fig. 8 indicates the Receiver Operating Characteristic curve.

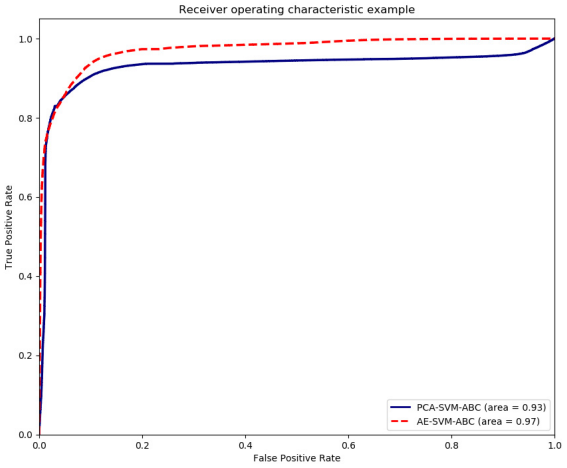


FIGURE 8. Comparison of receiver operating characteristic curve.

It shows superiority of proposed deep and shallow learning strategy in anomaly detection compared PCA-based method. The Area Under the Curve (AUC) in the proposed framework can reach 0.97, which indicates that the model has a good ability to predict.

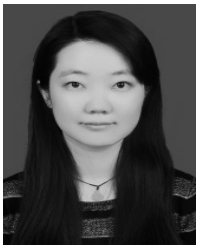
VI. CONCLUSION

The rapid development of the Internet of Things technology has had a profound impact on people's lives. However, during its development process, security issues remain the top priority for people's concerns. In this paper, we have studied the problem of network intrusion detection. The UNSW-NB15 dataset is adopted as a research object. In data preprocessing phase, a deep learning method auto-encoder is used to reduce features. SVM is used as decision engine with the help of the ABC algorithm for finding the best parameter. The simulation results tell that the proposed work is better than the PCA-based method and other machine learning strategies. But there are also many shortcomings such as high training time. Besides, the parameters of auto-encoder and SVM might not be the best pair. The next step will still focus on the promotion of accuracy and reduction of false alarm rates through improving the structure of the proposed work.

REFERENCES

[1] B. B. Zarpelão, R. S Miani, C. T. Kawakani, and S. C. de Alvarenga, "A survey of intrusion detection in Internet of Things," *J. Netw. Comput. Appl.*, vol. 84, pp. 25–37, Apr. 2017.
[2] F. A. Alaba, M. Othman, I. A. T. Hashem, and F. Alotaibi, "Internet of Things security: A survey," *J. Netw. Comput. Appl.*, vol. 88, pp. 10–28, Jun. 2017.
[3] G. Gui, H. Huang, Y. Song, and H. Sari, "Deep learning for an effective nonorthogonal multiple access scheme," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8440–8450, Sep. 2018.
[4] Y. Lin, C. Wang, J. Wang, and Z. Dou, "A novel dynamic spectrum access framework based on reinforcement learning for cognitive radio sensor networks," *Sensors*, vol. 16, no. 10, p. 1675, 2016.
[5] M. Aazam, M. St-Hilaire, C.-H. Lung, and I. Lambadaris, "PRE-Fog: IoT trace based probabilistic resource estimation at fog," in *Proc. 13th IEEE Annu. Consum. Commun. Netw. Conf. (CCNC)*, Jan. 2016, pp. 12–17.
[6] X. Ma, T. Wang, Y. Lin, and S. Jin, "Parallel iterative inter-carrier interference cancellation in underwater acoustic orthogonal frequency division multiplexing," *Wireless Pers. Commun.*, vol. 102, no. 2, pp. 1603–1616, 2018.
[7] Y. Li *et al.*, "MUSAI-$L_{1/2}$: Multiple sub-wavelet-dictionaries-based adaptively-weighted iterative half thresholding algorithm for compressive imaging," *IEEE Access*, vol. 6, pp. 16795–16805, 2018.
[8] H. Huang, J. Yang, H. Huang, Y. Song, and G. Gui, "Deep learning for super-resolution channel estimation and DOA estimation based massive MIMO system," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8549–8560, Sep. 2018.
[9] H. Wang, L. Guo, Z. Dou, and Y. Lin, "A new method of cognitive signal recognition based on hybrid information entropy and D-S evidence theory," *Mobile Netw. Appl.*, vol. 23, no. 4, pp. 677–685, 2018.
[10] A. Riahi, Y. Challal, E. Natalizio, Z. Chtourou, and A. Bouab, "A systemic approach for IoT security," in *Proc. IEEE Int. Conf. Distrib. Comput. Sensor Syst.*, May 2013, pp. 351–355.
[11] Y. Lin, X. Zhu, Z. Zheng, Z. Dou, and R. Zhou, "The individual identification method of wireless device based on dimensionality reduction and machine learning," *J. Supercomput.*, pp. 1–18, Dec. 2017. [Online]. Available: https://link.springer.com/article/10.1007/s11227-017-2216-2#citeas. doi: 10.1007/s11227-017-2216-2.
[12] Y. Tu, Y. Lin, J. Wang, and J.-U. Kim, "Semi-supervised learning with generative adversarial networks on digital signal modulation classification," *CMC-Comput. Mater. Continua*, vol. 55, no. 2, pp. 243–254, 2018.
[13] N. Zhao, F. R. Yu, M. Li, Q. Yan, and V. C. Leung, "Physical layer security issues in interference-alignment-based wireless networks," *IEEE Commun. Mag.*, vol. 54, no. 8, pp. 162–168, Aug. 2016.
[14] B. Tang, Y. Tu, Z. Zhang, and Y. Lin, "Digital signal modulation classification with data augmentation using generative adversarial nets in cognitive radio networks," *IEEE Access*, vol. 6, pp. 15713–15722, 2018.
[15] N. Ye, Y. Zhang, and C. M. Borror, "Robustness of the Markov-chain model for cyber-attack detection," *IEEE Trans. Rel.*, vol. 53, no. 1, pp. 116–123, Mar. 2004.
[16] Z. Zhang, X. Guo, and Y. Lin, "Trust management method of D2D communication based on RF fingerprint identification," *IEEE Access*, vol. 6, pp. 66082–66087, 2018.
[17] S. T. Ikram and A. K. Cherukuri, "Improving accuracy of intrusion detection model using PCA and optimized SVM," *J. Comput. Inf. Technol.*, vol. 24, no. 2, pp. 133–148, 2016.

[18] G. Ding, Q. Wu, L. Zhang, T. A. Tsiftsis, and Y. Yao, "An amateur drone surveillance system based on cognitive Internet of Things," *IEEE Commun. Mag.*, vol. 56, no. 1, pp. 29–35, Jan. 2018.
[19] Z. Xue, J. Wang, G. Ding, Q. Wu, Y. Lin , and T. A. Tsiftsis, "Device-to-device communications underlying UAV-supported social networking," *IEEE Access*, vol. 6, pp. 34488–34502, 2018.
[20] I. S. Thaseen and C. A. Kumar, "Intrusion detection model using fusion of chi-square feature selection and multi class SVM," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 29, no. 4, pp. 462–472, 2016.
[21] A. L. Buczak and E. Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 1153–1176, 2nd Quart., 2016.
[22] M. S. Islam, W. Khreich, and A. Hamou-Lhadj, "Anomaly detection techniques based on kappa-pruned ensembles," *IEEE Trans. Rel.*, vol. 67, no. 1, pp. 212–229, Mar. 2018.
[23] N. Shone, T. N. Ngoc, V. D. Phai, and Q. Shi, "A deep learning approach to network intrusion detection," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 2, no. 1, pp. 41–50, Feb. 2018.
[24] F. Farahnakian and J. Heikkonen, "A deep auto-encoder based approach for intrusion detection system," in *Proc. 20th Int. Conf. Adv. Commun. Technol. (ICACT)*, Feb. 2018, pp. 178–183.
[25] N. Moustafa and J. Slay. (2017). "A hybrid feature selection for network intrusion detection systems: Central points." [Online]. Available: https://arxiv.org/abs/1707.05505
[26] N. Moustafa and J. Slay, "The evaluation of network anomaly detection systems: Statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set," *Inf. Secur. J., Global Perspective*, vol. 25, nos. 1–3, pp. 18–31, 2016.
[27] H. M. Anwer, M. Farouk, and A. Abdel-Hamid, "A framework for efficient network anomaly intrusion detection with features selection," in *Proc. 9th Int. Conf. Inf. Commun. Syst. (ICICS)*, Apr. 2018, pp. 157–162.
[28] D. Bhamare, T. Salman, M. Samaka, A. Erbad, and R. Jain, "Feasibility of supervised machine learning for cloud security," in *Proc. Int. Conf. Inf. Sci. Secur. (ICISS)*, Dec. 2016, pp. 1–5.
[29] N. Moustafa and J. Slay, "The significant features of the UNSW-NB15 and the KDD99 data sets for network intrusion detection systems," in *Proc. 4th Int. Workshop Building Anal. Datasets Gathering Exper. Returns Secur.*, Nov. 2015, pp. 25–31.
[30] M. Belouch, S. El Hadaj, and M. Idhammad, "A two-stage classifier approach using reptree algorithm for network intrusion detection," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 6, pp. 389–394, 2017.
[31] F. Fogelman-Soulie, *Automata Networks in Computer Science: Theory and Applications*. Princeton, NJ, USA: Princeton Univ. Press, 1987.
[32] J. T. Zhou, H. Zhao, X. Peng, M. Fang, Z. Qin, and R. S. M. Goh, "Transfer hashing: From shallow to deep," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 12, pp. 6191–6201, Dec. 2018.
[33] B. M. Aslahi-Shahri *et al.*, "A hybrid method consisting of GA and SVM for intrusion detection system," *Neural Comput. Appl.*, vol. 27, no. 6, pp. 1669–1676, 2016.
[34] A.-C. Enache and V. Sgârciu, "Anomaly intrusions detection based on support vector machines with an improved bat algorithm," in *Proc. 20th Int. Conf. Control Syst. Comput. Sci.*, May 2015, pp. 317–321.
[35] A.-C. Enache and V. V. Patriciu, "Intrusions detection based on support vector machine optimized with swarm intelligence," in *Proc. 9th IEEE Int. Symp. Appl. Comput. Intell. Inform. (SACI)*, May 2014, pp. 153–158.
[36] D. Karaboga, "An idea based on honey bee swarm for numerical optimization," Dept. Comput. Eng., Erciyes Univ., Tech. Rep. TR06, 2005.
[37] G. Ding, Q. Wu, Y.-D. Yao, J. Wang, and Y. Chen, "Kernel-based learning for statistical signal processing in cognitive radio networks: Theoretical foundations, example applications, and future directions," *IEEE Signal Process. Mag.*, vol. 30, no. 4, pp. 126–136, Jul. 2013.
[38] P. Gogoi, M. H. Bhuyan, D K. Bhattacharyya, and J. K. Kalita, "Packet and flow based network intrusion dataset," in *Proc. Int. Conf. Contemp. Comput.*, 2012, pp. 322–334.
[39] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, "A detailed analysis of the KDD CUP 99 data set," in *Proc. IEEE Symp. Comput. Intell. Secur. Defense Appl.*, Jul. 2009, pp. 1–6.
[40] D. Karaboga and B. Gorkemli, "A quick artificial bee colony-qABC-algorithm for optimization problems," in *Proc. Int. Symp. Innov. Intell. Syst. Appl.*, Jul. 2012, pp. 1–5.



QIAO TIAN was born in Harbin, Heilongjiang, China, in 1991. She received the B.S. degree from the College of Computer Science and Technology, Harbin Engineering University, Harbin, China, where she is currently pursuing the Ph.D. degree. Her main research interests include system parallel optimization, computer architecture, and communication.

JINGMEI LI was born in 1964. She received the M.S. and Ph.D. degrees from the College of Computer Science and Technology, Harbin Engineering University, Harbin, China.

From 1992 to 2000, she was a Lecturer. From 2000 to 2006, she was an Associate Professor. Since 2006, she has been a Professor with the College of Computer Science and Technology, Harbin Engineering University. She has published more than 70 papers. Her research interests include computer architecture performance optimization, big data and cloud computing, network and information security, and embedded technology.

HAIBO LIU was born in 1976. He received the B.S, M.S., and Ph.D. degrees from the College of Computer Science and Technology, Harbin Engineering University, Harbin, China, where he has been an Associate Professor, since 2006. He has published more than 100 papers and ten books. His research interests include machine learning, information security, and computer vision.