A Hybrid-based Feature Selection Method for Intrusion Detection System

Xibin Sun[1], Heping Ye[1], and Xiaolin Liu[1],
(Corresponding author: Xibin Sun)
Guangdong Polytechnic of Science and Technology, Zhuhai, China[1]
Zhuhai 519090,China
Email: jacky5555@qq.com
(Received Aug. 11, 2022; Revised and Accepted Dec. 5, 2022; First Online Dec. 13, 2022)

Abstract

As we know, feature selection can improve the performance of machine learning algorithms for intrusion detection. This paper proposes a hybrid feature selection method, which ranks features according to two factors: relevancy and redundancy, and then adopts the forward search strategy to select the optimal feature subset from the ranked features. Experiments on the KDDCup'99 dataset showed that our proposed feature selection method could get better performance on the accuracy rate and false positive rate in intrusion detection compared with other feature selection approaches.

Keywords: Feature Selection; Hybrid; Intrusion Detection; Performance

1 Introduction

The intrusion detection system (IDS) as a part of network security infrastructure, can detect network attacks or abnormal behaviors. Traditional IDS can be categorized into two types: Signature-based IDS and Anomaly-based IDS. Signature-based IDS is good at detecting known network attacks and suffers from unknown or novel attacks. Anomaly-based IDS can detect novel attacks, yet it usually owns a high false positive rate. Therefore, there exist challenges in the traditional IDS when they are deployed into the real-world network environment. At the same time, as machine learning (ML) methods are applied to different fields successfully, many researchers introduce them into the intrusion detection domain for detecting network attacks. The authors [19] introduced ML algorithms to build classification models from the network datasets to predict the network attacks, such as the support vector machine (SVM) algorithms, the decision tree (DT) algorithms, the random forest (RF) algorithms, deep learning algorithms, and so on.

Though these built IDS by using ML methods can obtain better results in detecting network attacks, they often suffer from the high dimensional and massive network traffic data. Furthermore, the features of large-scale network traffic often contain redundancy, incompleteness, and irrelevance, which not only declines the performance of ML algorithms but also adds the time and complexity of building classification models. Therefore, it is significant to select the optimal feature subset from the initial network datasets before using ML algorithms to build classification models. In this paper, we have designed a hybrid-based feature selection method that consists of two phases. In the first phase, the filter method is used to sort the features from the original feature space of the network datasets according to their relevancy and redundancy. In the second phase, the wrapper feature selection approach is used to select the final feature subset from the sorted features of the first phase. To be specific, we start with the first feature of the sorted features and incrementally add a feature to the wrapper method one by one, and the feature subset that can make the classification model obtain the highest accuracy rate in detecting network attacks will retain the final selected features. The contributions of this paper are listed as follows:

1) We proposed a hybrid feature method that integrates the high efficiency of the filter feature selection method and the ability to extract optimal features of the wrapper feature selection method. The experimental results on the KDDCup'99 dataset showed our proposed method could get better performance than other relevant feature selection methods.

2) For other feature selection methods, such as MIFS [8], MIFS-U [16], MMIFS [7], and LCC-RF-RFEX [24], when they are used to select a feature subset from the initial feature space, the specific threshold or the number of final selected features need to be provided in advance. However, these values are often set based on the human experience, which is hard to set the optimal values when facing different network datasets. Our approach differs from the above feature selection methods, which don't provide a specific threshold or the selected fea-

ture number.

3) As we know, the traditional wrapper feature selection methods usually adopt a greedy search approach by evaluating all the possible combinations of features against the specific machine learning algorithm. Therefore, it is a time-consuming process when directly applying the wrapper methods to the initial features for feature selection. In the final phase of our approach, we use the wrapper method to select the final features from the sorted features rather than the initial features, which avoids the computational disaster of feature selection in the wrapper method and improves the efficiency of wrapper methods.

The rest of this paper is structured as follows. Section 2 presents the related works on feature selection approaches. Section 3 introduces our proposed feature selection method in detail. The analysis of experimental results shows in Section 4. Finally, Section 5 summarizes the works in this paper and points out the future works.

2 Related Works

The feature selection methods can reduce the time to build classification models or improve the accuracy of classification models. So far, we can divide the feature selection approaches into three categories: filter, wrapper, and hybrid. Filter methods mainly select feature subsets using the specific heuristic evaluation function to measure the relevance of features. Wrapper methods often adopt a classification algorithm to train a model to estimate the optimal feature subset. Therefore, filter methods are much faster than wrapper methods as they do not involve training models. However, the final selected features using the wrapper methods can often make the classification models obtain better performance than those using the filter methods. Hybrid methods often have the best performance by integrating the advantages of filter methods and wrapper methods.

The authors [5, 7, 8, 11, 14, 16, 20, 23, 24, 29] introduced the filter feature selection methods, such as [5, 7, 8, 16, 20, 24, 29] mainly presented the correlation-based feature selection (CFS) method for feature selection. At present, Linear Correlation Coefficient (LCC) and Mutual Information are the two main heuristic evaluation functions to evaluate the correlation between two random variables. For example, The authors [11] used the LCC function to measure the relevance between two features, then, ranked the features according to the calculated correlation values, and finally, selected the final feature subset by removing the features of which correlation values are below the specified thresholds. Similarly, The authors [5, 7, 8, 16, 20, 24, 29] introduced how to use Mutual Information (MI) methods to select features. The authors [8] first provided the mutual information method feature selection (MIFS) which maximizes the relevance between feature and class label and minimizes the redundancy of the selected features. MIFS-U [16], MMIFS [7], FMIFS [5], mRMR [20], and RPFMI [29] were all based on the improvement of MIFS. In MIFS-U [16], MMIFS [7], and mRMR [20], their feature selection algorithms need to provide the specific threshold as the input parameter before using them to select features. Though FMIFS [5] and RPFMI [29] overcome the above limitation, they belong to the filter method which mainly uses statistical techniques to evaluate the intrinsic relationship of features (i.e., the relevance and redundancy), and the final selected features are independent of the learning algorithm, which leads to the built IDS owning a lower detection accuracy.

The authors [1–4,9,10,13,21,22,25] introduced wrapper methods to select features. Such as, the authors [25] presented the wrapper method to select features, and SVM algorithm is used to build IDS based on the final selected features. Experimental results showed that the build IDS achieved 82.34% accuracy rate. The authors [22] used C4.5 tree and BN algorithms to select features, and got the higher accuracy rate and the lower false positive rate for the four types of attacks (Dos, Probe, R2L, and U2R) respectively by comparing the full 41 features. The authors [2] provided a new feature selection algorithm based on pigeon inspired optimizer(PIO) for IDS, and the experimental results showed that the PIO feature selection algorithm not only reduced the number of features of KDDCup'99, NSL-KDD, and UNSW-NB15 datasets respectively, but also maintained a high accuracy rate and reduced the required time for training the classification models significantly.

The authors [6, 17, 18, 27] demonstrated the hybrid feature selection methods to select features. Such as, the authors [6] used the filter method to eliminate the irrelevant and redundant features from the initial feature space and then, the remained features were fed to the wrapper method LS-SVM to select the final feature subset. Experiments showed that the proposed method could get the 98.9% accuracy rate classification accuracy. The authors [17] designed the hybrid method: FGLCC-CFA, which combined the filter FGLCC method and the wrapper method CFA. It first used the FGLCC to rank the initial features and select the opimal feature subset, and then, the feature subset was input to the CFA method to select the final features. Experimental results showed the FGLCC-CFA method got a higher accuracy rate and detection rate equal to 95.03% and 95.23%, respectively, and a lower false positive rate of 1.65% compared with the filter FGLCC method and the wrapper CFA method.

3 Proposed Feature Selection Method

Through the analysis of the above-related literate, we find the current feature methods may exist the following defects:

1) Many feature selection methods exist a limit that

needs to specify the threshold or the number of final selected features before using them to select the final feature subset. Such as, the authors [7,8,16,24] needs to set a specific value for the redundancy parameter in their feature selection algorithms. Yet, there is no empirical value for the parameter, and how to set an appropriate value for the parameter is still a vexing question to answer, especially when facing tasks in different domains and different datasets.

2) For the wrapper methods, evaluating all the possible combinations of features by using the machine learning algorithms from the initial feature space is often a time-consuming process, which is called an NP-complete problem [15], especially for high-dimension feature space.

To overcome the aforementioned problems, we proposed a hybrid feature selection method that contains two phases. In the first phase, we use mutual information to rank the features by comprehensively considering their relevance and redundancy. In the second phase, we adopt the forward search strategy (FSS) to incrementally select features from the sorted feature set and then feed them to the specific classification algorithm to count the accuracy rate (AR). The feature subset which gets the maximum AR will be retained as the final selected features. Different from other filter approaches, our approach only ranks the initial features rather than selects features, so there is no need for setting the specific threshold beforehand. Furthermore, in the second phase of our approach, we use a forward search strategy (FSS) to select a feature subset from the ranked feature space rather than the initial feature space, which effectively avoids the NP-complete problem of the feature combination in the wrapper methods. The workflow of our approach is shown in Figure 1.

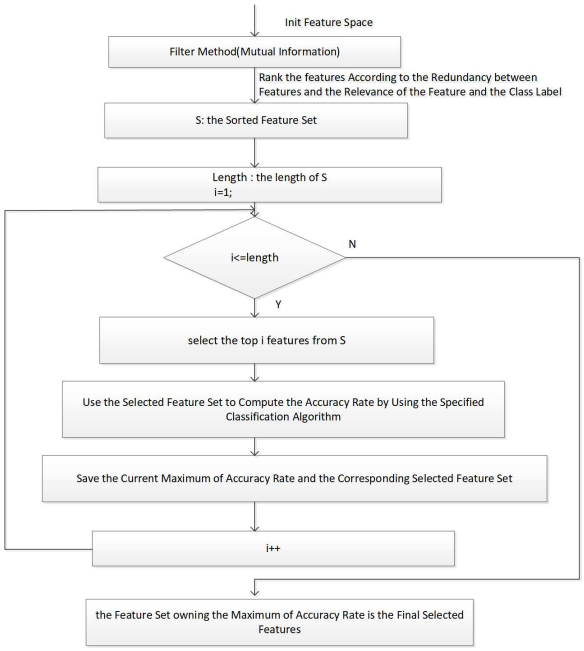

Figure 1: The workflow of our feature selection approach

3.1 Mutual Information

We use mutual information [8] as a heuristic evaluation function to rank the features in our proposed approach. As we know, mutual information is widely used to measure the relevance between random variables. If two random variables U=$\{u_1,u_2,...,u_n\}$ and V=$\{v_1,v_2,...,v_n\}$ belong to the discrete variables, where n is the total number of samples, the mutual information (MI) of the two variables is defined as shown in Equation (1) [8]:

$$I(U;V)=\sum_{u\in U}\sum_{v\in V}p(u,v)\log\frac{p(u,v)}{p(u)p(v)}. \quad (1)$$

Where p(u) and p(v) are the probability distribution of U and V separately. p(u,v) is a joint probability distribution. For continuous variables, the MI is defined as shown in Equation (2) [8]:

$$I(U;V)=\int_u\int_v log\frac{p(u,v)}{p(u)p(v)}dudv. \quad (2)$$

The MI value is larger, which presents that the two variables are closely related, and A zero value of MI indicates that the two variables are independent.

3.2 Proposed Feature Selection Algorithm

Our proposed feature selection method mainly contains two main phases: the first phase in which the filter method is used for feature ranking. Different from other filter methods aiming at feature selection, our approach mainly uses mutual information to rank the features according to the redundancy of features and the relevance of the feature and the class label. Inspired by [5], we use Equation (3) to decide the position of a feature in the final sorted feature subset.

$$G_{MI}=\text{argmax}_{f_i\in F}\left(I(C;f_i)-\frac{1}{|S|}\sum_{f_s\in S}MR\right) \quad (3)$$

Where F is the original feature set of datasets, f_i is the candidate feature, S is the sorted feature set from the original feature set F, f_s is the sorted feature in S, $|S|$ is the number of the final sorted features in S and C is the class label, MR in Equation (3) is the relative redundancy of feature f_i against feature f_s. MR is defined by Equation (4) [5]:

$$MR=\frac{I(f_i;f_s)}{I(C;f_i)} \quad (4)$$

Where I(C; f_i) is the mutual information value between the candidate feature f_i and the class C, and I(f_i; f_s) is

the mutual information value between the candidate feature f_i and the sorted feature f_s. Equation (3) is intended to select a feature f_i from the F that maximizes I(C; f_i) and minimizes the average of redundancy MR simultaneously.

In the second phase, we use the wrapper method to select the feature subset from the ordered feature set S which is coming from the first phase of our proposed approach. In the second phase, we adopt the forward search strategy (FSS) to incrementally select features from the sorted feature set S, and then feed them to the specific classification algorithm to count the accuracy rate. The final feature subset which can get the maximum accuracy rate will be retained. The pseudo-code of our proposed feature selection method is shown in Algorithm 1.

Algorithm 1 The Proposed Feature Selection Algorithm

Input:

F: *Feature set* $F = \{f_i | i = 1, .., n\}$

A: The Specific Classification Algorithm(e.g. Decision Tree, Naive Bayes, Support Vector Machine, etc)

Output:

$maxS$: The Final Selected Feature Subset

Begin
$S \leftarrow \varnothing$
Calculate I(C; f_i), for each feature f_i, i=1,..,n, C notes the class label.
if I(C;f_i)==0 **then**
  $F \leftarrow F \setminus \{f_i\}$
end if
Select the feature f_i: $f_i \in F$ that maximizes I(C; f_i).
$S \leftarrow S \cup \{f_i\}$
$F \leftarrow F \setminus \{f_i\}$
while $F \neq \varnothing$ **do**
  select the feature f_i using Equation (3)
  $S \leftarrow S \cup \{f_i\}$
  $F \leftarrow F \setminus \{f_i\}$
end while
$maxAR \leftarrow 0$
$maxS \leftarrow \varnothing$
$length \leftarrow$ the length of S
for $i = 1; i \leq length; i++$ **do**
  $S_{sub} \leftarrow$ select the top i features from S
  Count Accurate Rate (AR) of the classification algorithm A by using S_{sub}
  if $AR > maxAR$ **then**
    $maxAR \leftarrow AR$
    $maxS \leftarrow S_{sub}$
  end if
end for
return maxS
End

4 Experiments and Results

4.1 Datasets for Evaluation

KDDCup'99 dataset [26] is one of the datasets for evaluating intrusion detection. It contains 39 attack types divided into four categories: Dos, Probe, U2R, and R2L. Furthermore, it also provides the training and test datasets for evaluating the machine learning algorithms. The training dataset contains about five million connection records, and the test dataset includes around two million records. Each connection record that contains 41 features is labeled as either normal or an attack. Considering that there are a large number of redundant records and the imbalance of the distribution of attack records in the KDDCup'99 dataset, We selected partial data from the KDDCup'99 dataset to generate the corresponding training dataset and test dataset for each of the four attack categories. Details of the generated datasets are shown in Table 1.

4.2 Performance Metrics

In this paper, we mainly use two metrics to evaluate the performance of our proposed feature selection method, and the performance metrics are accuracy rate (AR) and false positive rate (FPR) separately. AR can be formally defined as:

$$AR = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FN} + \mathrm{FP}} \tag{5}$$

FPR is defined as:

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}} \tag{6}$$

4.3 Experimental Results and Analysis

Python language is used to realize our proposed feature selection approach, and all the experiments were performed on a Windows platform having configuration i5 core 4 CPU 2.3 GHz, 8GB RAM. Table 2 shows the final selected features of four types of attacks by using our approach based on the decision tree algorithm. Figure 2 and Figure 3 show the AR and FPR of the built IDS based on the selected features by using our proposed approach and full features(41) separately. The results demonstrate that the IDS based on the selected features can achieve better performances in DR and FPR metrics by comparing IDS constructed with the full features(41). Furthermore, we also test the total consuming time(training and test times) of the built IDS by using selected features and the full features(41) separately. Table 3 shows that the IDS based on selected features consumes less time than IDS based on the full features. This is principally because our proposed approach deletes the redundant and irrelevant features from the full features, which causes not only to reduce the total consuming time of classification models but also to improve their performance.

Table 1: Sample Distributions of Instances for Four Attack Types in Datasets

Attack Type	Attack Name	Training Data	Test Data
Dos	*normal*	20000	20000
	smurf	10000	10000
	neptune	5000	5000
	mailbomb	1500	1500
	back	500	500
	land	15	15
	teardrop	400	400
	processtable	350	350
	pod	100	100
	aparche2	250	250
	SubTotal	*training dataset:38115*	*test dataset:38115*
Probe	*normal*	10000	10000
	ipsweep	1247	306
	mscan	600	400
	nmap	130	100
	portsweep	540	500
	saint	400	300
	satan	800	600
	SubTotal	*training dataset:13717*	*test dataset:12206*
U2R	*normal*	10000	10000
	buffer_overflow	30	22
	httptunnel	158	158
	loadmodule	9	2
	perl	3	2
	rootkit	10	13
	SubTotal	*training dataset:10210*	*test dataset:10197*
R2L	*normal*	20000	20000
	ftp_write	8	3
	guess_passwd	53	4367
	imap	12	1
	multihop	8	19
	phf	6	3
	warezclient	1021	1021
	warezmaster	21	1603
	SubTotal	*training dataset:21129*	*test dataset:27017*

Table 2: Selected Features by Using Proposed Approach based on Decision Tree Algorithm

Attack Type	Selected Features
Dos	5, 37, 23, 3, 31, 12, 25, 36, 2, 6, 26, 16, 32, 13, 24, 39, 8
Probe	5
R2L	5,22,11
U2R	5, 14, 17, 13, 40, 18, 10, 11, 27, 15, 9, 16, 41

In addition, we also compared the performance of our approach with the filter methods, such as the linear correlation-based feature selection (LCFS) algorithm [7], the mutual information-based feature selection (MIFS) algorithm [8], and the wrapper methods, such as the Random Forest-Recursive Feature Elimination (RF-RFE) [12]. As shown in Figure 4, compared to other filter and wrapper methods, our approach has a higher AR, which indicates that the IDS based on the hybrid feature selection method has better performance than IDS based on a single filter method or wrapper method. Moreover, as mentioned in section III, unlike other wrapper methods based on initial features, we use the wrapper method on the ranked features to select the final feature subset in

Table 3: The Total Consume Time(training time and test time) of Decision Tree Algorithm Based on Selected Features and Full Features(41)

Attack Type	Total Consume Time(s)	
	Selected Features	Full Features(41)
Dos	0.21875	0.29688
Probe	0.03125	0.12500
R2L	0.09375	0.28125
U2R	0.03125	0.06250

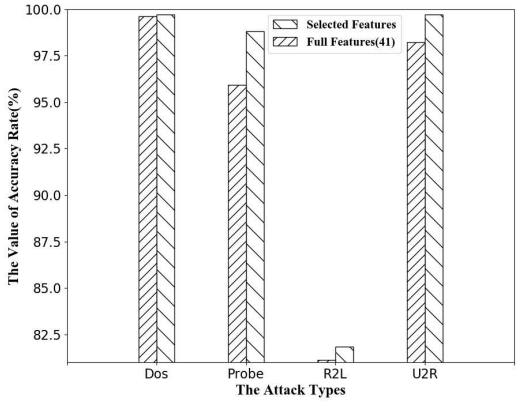

Figure 2: The Accuracy Rate(AR) of Decision Tree Algorithm With Selected Features and Full Features(41)

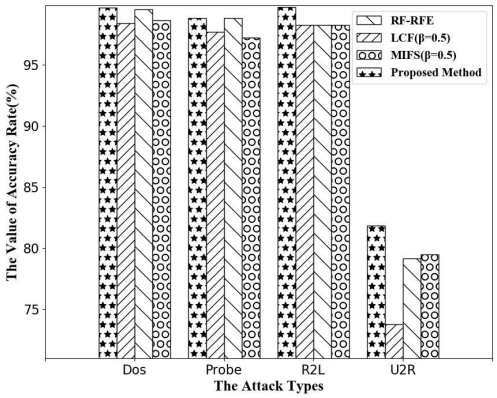

Figure 4: Accuracy Rate of Decision Tree Algorithm With non-hybrid Feature Selection Methods and Proposed Method

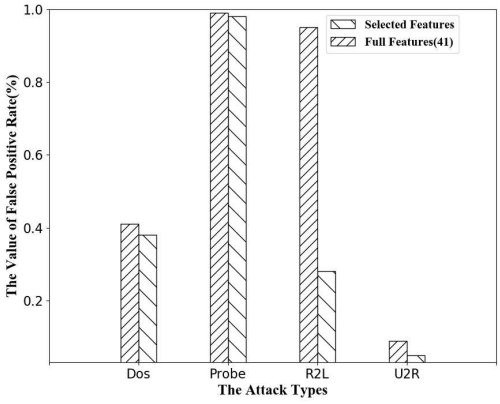

Figure 3: The False Positive Rate(FPR) of Decision Tree Algorithm With Selected Features and Full Features

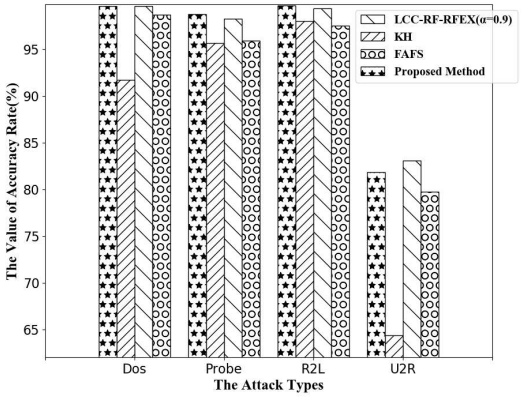

Figure 5: Accuracy Rate of the Classification Model based on Decision Tree Algorithm With Hybrid-based Feature Selection Methods

the second phase of our proposed method. Table 4 shows that wrapper methods based on sorted features have more efficient time performance than wrapper methods based on the original features.

Furthermore, we evaluated the IDS based on our feature approach and the recent hybrid feature selection methods, such as the LCC-RF-RFEX [24], KH [28], and FAFS [22] methods. Figure 5 shows the accuracy rate of classification models based on the decision tree algorithm with hybrid feature selection methods. Experimental results show that our proposed approach outperforms

Table 4: Consume Time of Selecting Feature Subset

Attack Type	**Consume Time(s)**	
	Wrapper Method based on Sorted Features **(Proposed Method)**	Wrapper Method based on Original Features **(RF-RFE)**
Dos	7.26562	7.56250
Probe	2.64062	4.54688
U2R	2.23438	2.90625
R2L	4.75000	7.75000

Table 5: Accuracy Rate of Our Feature Selection Method based On Different Classification Algorithms

Classification Algorithm	**Dos**	**Probe**	**U2R**	**R2L**
Random Forest	99.89%	98.25%	99.37%	83.11%
Naive Baye	90.99%	95.69%	98.03%	64.35%
Multi perceptron	98.61%	95.91%	97.54%	79.75%
Support Vector Machine	97.69%	97.65%	98.37%	77.29%
Decision Tree	99.62%	98.80%	99.70%	81.84%
Logistic Regression	93.73%	98.30%	98.77%	78.65%

these hybrid-based feature selection methods (except for the R2L attack type). This is mainly because we adopt the forward search strategy (FSS) to incrementally select features from the sorted feature set. The feature subset which gets the maximum AR will be saved as the final selected feature subset. Finally, we evaluate the performance of our feature selection method based on different classification algorithms. Table 5 shows that the IDS based on the Random Forest (RF) and Decision Tree (DT) can obtain better AR by comparing with the IDS based on other classification algorithms.

5 Conclusions and Future Work

This paper proposed a hybrid feature selection method for intrusion detection, which absorbs the advantages of the filter feature selection methods and the wrapper methods. Different from other filter feature selection methods, we use mutual information to rank the original features rather than select features. Unlike other wrapper approaches, we use the wrapper method to select a feature subset from the sorted features rather than initial features. Furthermore, we adopt the forward search strategy (FSS) to incrementally select features from the sorted feature set and then feed them to the specific classification algorithm to count the AR. The feature subset which gets the maximum AR will be retained as the final selected subset. Therefore, there is no need to specify the threshold or the number of the final selected features in advance when using our approach to select the final feature subset. Experimental results on the KDDCup'99 dataset showed that our approach could achieve better performance compared with other related feature selection methods.

So far, we only finish selecting the optimal feature subset from the labeled datasets. However, for unlabeled network traffic, how to use unsupervised technology to select the features of attacks will be considered in our future studies.

Acknowledgments

This study is mainly supported by the Guangzhou Science and Technology Program Project "Research on Feature Engineering Technology and Machine Learning algorithm Applied in the Intrusion Detection System" (Project No.: 202102080586). Furthermore, It is also supported in part by the Guangdong Province Educational Science Planning Project "Research on the Application Mode and Practice of VR Technology in Vocational Education" (Grant No.: 2018GXJK318).

References

[1] W. L. AI-Yaseen, A. K. Idrees, and F. H. Almasoudy, "Wrapper feature selection method based differential evolution and extreme learning machine for intrusion detection system", *Pattern Recognition*, vol. 132, no. 12, pp. 1–10, 2022.

[2] H. Alazzam, A. Sharieh, and K. E. Sabri. "A feature selection algorithm for intrusion detection system based on Pigeon Inspired Optimizer", *Expert Systems With Applications*, vol. 148, no. 15, pp. 1–13, 2020.

[3] F. H. Almasoudy, W. L. Al-Yaseen, and A. K. Idrees. "Differential Evolution Wrapper Feature Selection

for Intrusion Detection System", *Procedia Computer Science*, vol. 167, no. 2, pp. 1230–1239, 2020.

[4] O. Almomani. "A Feature Selection Model for Network Intrusion Detection System Based on PSO, GWO, FFA and GA Algorithms", *Symmetry*, vol. 12, no. 6, pp. 1–20, 2020.

[5] M. A. Ambusaidi, X. He, P. Nanda, and Z. Tan. "Building an intrusion detection system using a filter-based feature selection algorithm", *IEEE Transactions on Computers*, vol. 65, no. 10, pp. 2986–2998, 2016.

[6] M. A. Ambusaidi, X. He, Z. Tan, P. Nanda, L. F. Lu, and U. T. Nagar. "A Novel Feature Selection Approach for Intrusion Detection Data Classification", in *IEEE 13th International Conference on Trust, Security and Privacy in Computing and Communications*, pp. 82–89, Beijing, China, Sep 2014.

[7] F. Amiri, M. Mahdi, R. Yousefi, and A. Shakery. "Mutual information-based feature selection for intrusion detection systems", *Journal of Network and Computer Applications*, vol. 34, no. 4, pp. 1184–1199, 2011.

[8] R. Battiti. "Using mutual information for selecting features in supervised neural net learning", *IEEE Transactions on Neural Networks*, vol. 5, no. 5, pp. 537–550, 1994.

[9] M. S. Bonab, A. Ghaffari, F. S. Gharehchopogh, and P. Alemi. "A wrapper-based feature selection for improving performance of intrusion detection systems", *International Journal of Communication Systems*, vol. 33, no. 12, pp. 1–25, 2020.

[10] K. Bouzoubaa, B. Nsiri, and Y. Taher. "Predicting DOS-DDOS Attacks: Review and Evaluation Study of Feature Selection Methods based on Wrapper Process", *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 5, pp. 131–145, 2021.

[11] H. F. Eid, A. E. Hassanien, T. H. Kim, and S. Banerjee. "Linear correlation-based feature selection for network intrusion detection model", in *International Conference on Security of Information and Communication Networks*, pp. 240–248, Cairo, Egypt, Sep 2013.

[12] B. Gregorutti, B. Michel, and B.P. Saint-Pierre. "Correlation and variable importance in random forests", *Stat Comput*, vol. 27, no. 3, pp. 659–678, 2017.

[13] M. Hasan, M. Nasser, and K. Molla. "Feature Selection for Intrusion Detection Using Random Forest", *Journal of Information Security*, vol. 7, no. 3, pp. 129–140, 2016.

[14] S. M. Kasongo and Y. X. Sun. "Performance Analysis of Intrusion Detection Systems Using a Feature Selection Method on the UNSW-NB15 Dataset", *Journal of Big Data*, vol. 7, no. 105, pp. 1–20, 2020.

[15] B. Kumari and T. Swarnkar. "Filter versus wrapper feature subset selection in large dimensionality micro array: a review", *International Journal of Computer Science and Information Technologies*, vol. 2, no. 2, pp. 1048–1053, 2011.

[16] N. Kwak and C-H. Choi. "Input feature selection for classification problems", *IEEE Transactions on Neural Networks*, vol. 13, no. 1, pp. 143–159, 2002.

[17] S. Mohammadi and H. Mirvaziri. "Cyber intrusion detection by combined feature selection algorithm", *Journal of Information Security and Applications*, vol. 44, no. 8, pp. 80–88, 2019.

[18] H. Mohammadzadeh and F. S. Gharehchopogh. "A novel hybrid whale optimization algorithm with flower pollination algorithm for feature selection: Case study Email spam detection", *Computational Intelligence*, vol. 37, no. 1, pp. 176–209, 2021.

[19] U.S.Musa, M.Chhabra, A.Ali, and M.Kaur. "Intrusion Detection System using Machine Learning Techniques: A Review", in *2020 International Conference on Smart Electronics and Communication (ICOSEC)*, pp. 149–155, Trichy, India, Sep 2020.

[20] H. Peng, F. Long, and C. Ding. "Feature selection based on mutual information criteria of maxdependency, max-relevance, and min-redundancy", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.

[21] F.Salo, A. B. Nassif, and A. Essex. "Dimensionality reduction with IGPCA and ensemble classifier for network intrusion detection", *Computer Networks*, vol. 148, no. 15, pp. 164–175, 2019.

[22] B. Selvakumar and K. Muneeswaran. "Firefly algorithm based Feature Selection for Network Intrusion Detection", *Computers Security*, vol. 81, no. 2, pp. 148–155, 2019.

[23] M. A. Siddiqi and W. Pak. "Optimizing filter-based feature selection method flow for intrusion detection system", *Electronics*, vol. 9, no. 12, pp. 1–18, 2020.

[24] X. B. Sun, D. Zhang, H. O. Qin, and J.H.Tang. "Bridging the Last-Mile Gap in Network Security via Generating Intrusion-Specific Detection Patterns through Machine Learning", *Security and Communication Networks*, vol. 2022, no. 1, pp. 1–20, 2022.

[25] K. A. Taher, B. M. Y. Jisan, and M. M. Rahman. "Network intrusion detection using supervised machine learning technique with feature selection", in *2019 International Conference on Robotics,Electrical and Signal Processing Techniques (ICREST)*, pp. 643–646, Dhaka, Bangladesh, Jan 2019.

[26] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani. "A detailed analysis of the KDD CUP 99 data set", in *2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications*, pp. 1–6, Ottawa, ON, Canada, Jul 2009.

[27] N. A. Umar, Z. F. Chen, and Y. Liu. "Network Intrusion Detection Using Wrapper-based Decision Tree for Feature Selection", in *In Proceedings of the 2020 International Conference on Internet Computing for Science and Engineering*, pp. 5–13, Male, Maldives, Jan 2020.

[28] L. Xin, Y. Peng, J. Yiming, and L. Tian. "LNNLS-KH: A Feature Selection Method for Network Intrusion Detection", *Security and Communication Networks*, vol. 2021, no. 1, pp. 1–22, 2021.

[29] F. Zhao, J. Y. Zhao, X. X. Niu, and Y. Xin. "A Filter Feature Selection Algorithm Based on Mutual Information for Intrusion Detection", *Applied Sciences*, vol. 8, no. 9, pp. 15–35, 2018.

Biography

XiBin Sun received his Ph.D of Computer Technology and Application in 2022 from Faculty of Information Technology, Macau University of Science and Technology, Macau, China. He is a lecturer in the computer science department of Guangdong Polytechnic of Science and Technology. His current research involves the machine learning and network intrusion detection.

HePing Ye received his Ph.D. in Technology of Computer Application in 2012 from the South China University of Technology, China. He is a lecturer in the computer science department of Guangdong Polytechnic of Science and Technology. His current research involves data mining and network security.

XiaoLin Liu received his master's degree in Software Engineering in 2006 from the South China University of Technology, China. He is a senior engineer in the computer science department of Guangdong Polytechnic of Science and Technology. His current research involves information hiding and network security.