International Journal on Semantic Web and Information Systems
Volume 18 • Issue 1

Intrusion Detection Using Normalized Mutual Information Feature Selection and Parallel Quantum Genetic Algorithm

Zhang Ling, Zhengzhou University of Light Industry, China*
Zhang Jia Hao, Zhengzhou University of Light Industry, China

ABSTRACT

This paper presents a detection algorithm using normalized mutual information feature selection and cooperative evolution of multiple operators based on adaptive parallel quantum genetic algorithm (NMIFS MOP-AQGA). The proposed algorithm is to address the problems that the intrusion detection system (IDS) has lower detection speed, less adaptability, and lower detection accuracy. In order to achieve an effective reduction for high-dimensional feature data, the NMIFS method is used to select the best feature combination. The best features are sent to the MOP-AQGA classifier for learning and training, and the intrusion detectors are obtained. The data are fed into the detection algorithm to ultimately generate accurate detection results. The experimental results on real abnormal data demonstrate that the NMIFS MOP-AQGA method has higher detection accuracy, lower false negative rate, and higher adaptive performance than the existing detection methods, especially for small samples sets.

KEYWORDS

Entropy, Features Reduction, Genetic Algorithm, Intrusion Detection, Mutual Information Feature Selection, Operator, Parallel Universe, Quantum

INTRODUCTION

IDS is proved to be an effective method of network security defense (Teng et al., 2020). Many researchers have used machine learning algorithms (Alyaseen et al., 2017; Kumar et al., 2019; Li et al., 2019) to research IDS, such as deep learning, support vector machine (SVM), fuzzy sets, outliers and random forest, and genetic algorithm, and have made many breakthroughs.

On the one hand, there are a large amount of network logs for IDS to detect, so an effective algorithm should be researched to delete the redundant features to improve the detection speed. There are some features selection algorithms used to reduce the redundant features, such as rough set, fuzzy set, and so on.

Feature selection algorithm (FSA) is introduced as a pretreatment to the anomaly detection to optimize existing classifiers. FSA can eliminate irrelevant and redundant features, reduce computational complexity, and improve the accuracy of the learning algorithms (Chunhui & Wenjuan, 2021; Ying- Wu et al., 2010).

Este´vez et al. (2009) designed a Mutual Information Feature Selection (MIFS) method. However, in the MIFS algorithm, the increase of the input features can easily lead to some irrelevant feature

DOI: 10.4018/IJSWIS.307324 *Corresponding Author

selections (Lashkia et al., 2004). Peng et al. (2014) proposed a minimal- Redundancy- Maximal-Relevance (mRMR) criteria, with which the impact of parameter β through the average of redundancy values was decreased. This criterion has a very low expense to give feature selection, but the entropy may vary considerably. Panigrahi (2021) gave an improved infinite feature selection for multiclass classification (IIFS-MC) to eliminate the superfluous attributes.

To increases the speed and deviation of mutual information among multi-valued attributes, the values of features are normalized in [0, 1]. The authors gave a NMIFS to reduce the algorithm complexity and obtain the optimal features. The experiment results showed that the NMIFS method has better performance on feature selection on several benchmark problems.

On the other hand, the classifier will directly affect the accuracy of anomaly detection (Yilei et al., 2021). JooHwa and KeeHyun (2019) designed an IDS with autoencoder - conditional, the generative adversarial networks and the random forest (AE - CGAN - RF), autoencoder-conditional method was adopted to reduce high-dimensional data dimension and to get a higher detection rate. Jiadong et al. (2019) gave a hybrid multilevel intrusion detection model. The outliers detection algorithm can effectively reduce some redundant attributes and improve the speed of detection. Alyaseen et al. (2017) used K - means algorithm to achieve training data set in a multilevel hybrid intrusion detection model, with which, they got better performance of classifiers. Yang et al. (2019) proposed an Effective IDS using the Modified Density Peak Clustering Algorithm and Deep Belief Networks (MDPCA-DBN). They used the Modified Density Peak Clustering Algorithm and Deep Networks to reduce the size of the training set, solve the imbalance of sample, and improve the efficiency of detection. Song et al. (2018) proposed an anti-adversarial hidden markov model for network-based intrusion detection (AA-HMM). However those algorithms had lower self-adaptability, lower detection rate, and higher false alert rate for small samples sets.

In order to improve the self-adaptability, the GA is used in IDS. Cheng et al. (2016) put forward an IDS using a new fuzzy rule-based classification system based on GA. A fuzzy rule-base classification system is used to find a compact set of fuzzy if-then classification rules. Genetic algorithm was used for rule weights specification (Varzaneh and Rafsanjani, 2021). Feng and Dou (2021) presented a weighted intrusion detection model of the dynamic selection (WIDMoDS) based on data features. The standards of classifier selection in dynamic selection were updated. Xi et al. (2021) introduced immune adaptive and feedback mechanism to build a multisource neighborhood immune detector adaptive model (MS-NIDAM). The detectors can be adaptively evolved in a more targeted search domain. But this method is greatly influenced by initial population distribution, and the crossover and mutation probability of genetic operators are fixed, which are not conducive to search the global optimal solution.

In the stage of generation of classifier, the quantum computing theory and cooperative evolution of multiple operators are applied to improve the detectors. The authors give an anomaly detection method based on a normalized mutual information and cooperative evolution of multiple operators based on adaptive parallel quantum genetic algorithm. The contributions of this paper include:

1. In the features selection stage, to improve the anomaly detection speed, the NMIFS method is used to select the optimum detection features from a given sample features set and achieve effective dimensionality reduction of high dimensional feature data.
2. In the stage of generation of classifier, combined with a strong learning generalization of genetic algorithm and the acceleration of the quantum computing, the authors give a MOP-AQGA algorithm. Synthesizing the two improved algorithms above, the authors give an anomaly detection method based on a normalized mutual information and cooperative evolution of multiple operators based on adaptive parallel quantum genetic algorithm.
3. Finally, with the benchmark of NSL-KDD and CICIDS2017, experiments are given to verify the reduction performances of NMIFS. Detecion experiments are to verify the anomaly and classification performances of MOP-AQGA, especially for small sample sets. The results on

anomaly data from real network traffic indicate that this method has a higher detection accuracy and lower false positive rate compared with other existed anomaly detection algorithms.

THE NORMALIZED MUTUAL INFORMATION FEATURE SELECTION ALGORITHM

Entropy is a mathematical measure for uncertain random variables and used to describe a measure of the random variables for the average amount of information. If *X*, *Y* is denoted as the two discrete random variables with joint probability mass function *p(x, y)* and marginal probability *p(x)*, *p(y)*, the information entropy of *X* is defined as (Wanwei et al., 2017):

$$H\left(X\right) = -\sum_{i=1}^{m} p(x_i) \log_2 p(x_i) \tag{1}$$

The combination entropy *H (X, Y)* of random variable *X* and *Y* is defined as:

$$H\left(X,Y\right) = -\sum_{i=1}^{m}\sum_{j=1}^{n} p(\mathrm{x}_i, \mathrm{y}_j) \log_2 p(\mathrm{x}_i, \mathrm{y}_j) \tag{2}$$

x_i and y_j denote all possible values of *X* and *Y* respectively.

The combination entropy *H (X, Y)* is the uncertainty of measurement between *X* and *Y*. The angle of *H (X, Y)* is $\max\{H\left(X\right), H\left(Y\right)\} \le H\left(X,Y\right) \le H\left(X\right) + H\left(Y\right)$. When *X* depends on *Y*, the value of *H (X, Y)* is the minimum, when *X* and *Y* are independent of each other, the value of *H (X, Y)* is the maximum.

The mutual information (MI) is the information measure of two random variables, namely common information measure of two random variables. The mutual information (MI) between *X* and *Y* is defined as:

$$I\left(X;Y\right) = \sum_{i=1}^{m}\sum_{j=1}^{n} p(\mathrm{x}_i, \mathrm{y}_j) \log_2 \frac{p(\mathrm{x}_i, \mathrm{y}_j)}{p(\mathrm{x}_i) p(\mathrm{y}_j)} \tag{3}$$

Two main properties which distinguish mutual information with other related measurements are: (a) the ability to measure any kind of variables relationship and (b) invariance under space transformation. The former property is from mutual information, which is based on the combination probability density function and edge probability density function of the variables without any statistic information of variable gradient. The second property is based on the fact that the number of independent variables in Eq. (3) is dimensionless, therefore, the integral value is irrelevant to the selected coordinate (feature space transform). This feature still remains in differentiable or reversible transformation. The relation among mutual information, entropy, and combination entropy is:

$$I\left(X;Y\right) = H(X) + H(Y) - H(X,Y) \tag{4}$$

The minimum value of entropy is the upper bound of mutual information of I(X;Y). The range of the value is: $0 \le I\left(X;Y\right) \le \min\{H(X), H(Y)\}$. FNMI algorithm adopts symmetrical uncertainty

to measure the correlation between features and categories, features *X* and features *Y*. The symmetric uncertainty SU (X, Y) between two random variables X and Y is defined as:

$$SU\left(X,Y\right)=2[\frac{H(X)-H(X\mid Y))}{H(X)+H(Y)}]=\frac{I(X;Y))}{[H(X)+H(Y)]/2} \tag{5}$$

Because $\min\{H(X),H(Y)\}\le[\mathrm{H(X)}+\mathrm{H(Y)}]/2$, the definition of entropy can be denoted $H(X)\le\log_2\sum_{\mathrm{i=1}}^{\mathrm{m}}\mathrm{p(x}_i)\frac{1}{\mathrm{p(x}_i)}=\log_2 m$ with Jensen inequation. So we can get $0\le H(X)\le\log_2 m$. *m* and *n* are the number of possible values of discrete random variables *X* and *Y* respectively. The fast-normalized mutual information is defined as:

$$\mathrm{NMI(X;Y)}=\frac{I(X;\mathrm{Y})}{\min\{\log_2 m,\log_2 \mathrm{n}\}} \tag{6}$$

By formula (6), the features values of anonymous logs are normalized in the range of [0, 1] before executing the feature selection algorithm. Standardized process of the fast normalized mutual information feature selection is shown in table 1.

Table 1. Standardized process of the fast normalized mutual information feature selection

Standardized process of the fast normalized mutual information feature selection
Initialization: Set $F=\{f_i,i=1,\ldots,N\}$,which contains *N* features. Make $S=\{\varnothing\}$
Calculate the mutual information between features and classes: For each $f_i\in F$, calculate $NMI\left(f_i;C\right)$
Select the first feature: find $\hat{f}_i=\max_{i=1,\ldots,N}\left\{I\left(f_i;C\right)\right\}$. Set $F\leftarrow F\setminus\left\{\hat{f}_i\right\},S\leftarrow\left\{\hat{f}_i\right\}$
Greedy choices: repeat these steps until $\left\|S\right\|=k$
Calculate the mutual information among features: for each pair $\left(f_i,f_s\right)$,calculate $NMI\left(f_i;f_{\mathrm{j}}\right)$,where $f_i\in F$ and $f_s\in S$ until the completion of the iteration
Turn to the next feature: choose feature $f_i\in F$ making Eq.(10)reach the maximum. Set $F\leftarrow F\setminus\left\{f_i\right\},S\leftarrow\left\{f_i\right\}$
Output set *S* with *K* selected features

THE COOPERATIVE EVOLUTION OF MULTIPLE OPERATORS BASED ADAPTIVE PARALLEL QUANTUM GENETIC ALGORITHM

GA has the adaptive mechanism, so it can be used in IDS to obtain higher detection performances. In order to generate better detectors and increase the diversity of population, the cooperative evolution of multiple operators, parallel multi-universe mechanism, and quantum theory are applied to improve the GA. GA has the ability to assign reasonable rotation Angle step length to individuals according to their fitness. GA can accelerate the algorithm convergence speed.

THE COOPERATIVE EVOLUTION OF MULTIPLE OPERATORS

This paper adopts the cooperative evolution of multiple operators to obtain the mutation probability of the individuals in the population according to the calculation values of more operator in every generation, with which we can increase the population diversity in the late period of population evolution. Hamming distance denotes the similarity of different individual species. Hamming distance is smaller, individuals have the higher similarity.

The cooperative evolution of multiple operators uses the individual similarity to measure the individual operators. The individual fitness evaluation operator and population mutation operator are applied to determine the mutation probability of the individuals in the current population.

Individual similarity evaluation operator x_{sim}: x_{sim} is used to calculate the individual differences in current population and is defined in formula (7) (Zhi-jian et al., 2019).

$$x_{sim} = \begin{cases} \dfrac{d_{avg} - d_{min}}{d_{max} - d_{min}}, & d_{max} \neq d_{min} \\ 0, & d_{max} = d_{min} \end{cases} \tag{7}$$

In formula (7), d_{max} is the individual which has the maximal hamming distance with the optimal ones in the current population, d_{min} is the individual which has the minimal hamming distance with the optimal ones in the current population, and d_{avg} is the average hamming distance of all individuals with the optimal ones in current population. The larger the individual similarity evaluation operator is, the more different the individual is in the current population, then we can use larger mutation probability to increase population diversity; On the contrary, the smaller the x_{sim} is, we can use smaller mutation probability to maintain the stability of the population.

The individual fitness evaluation operator y'_{fit}: y'_{fit} is adopted to evaluate the *i*th individual' fitness in current population, which is defined as formula (8):

$$y'_{fit} = \begin{cases} \dfrac{f_{max} - f_i}{f_{max} - f_{min}}, & f_{max} \neq f_{min} \\ 0, & f_{max} = f_{min} \end{cases} \tag{8}$$

In formula(8), f_{max} is the best fitness value, f_{min} is the worst fitness value, and f_i is the *i*th individual's fitness. If the value of y'_{fit} is higher, the fitness of the *i*th individual is closer to the worst individual in current population. We should select higher mutation probability when the individuals mutate.

The population variation adjustment operator $F_{acc}(n)$**:** $F_{acc}(n)$ is the function of the current evolution algebra *n*, we take the method by gradually increasing individual mutation probability to solve the problem of premature convergence. Population variation adjustment operators are defined as shown in formula (9):

$$\mathrm{F}_{\mathrm{acc}}(n)=\begin{cases}\mathrm{F}_{\mathrm{acc}}(n-1)+\mathrm{C}\times\dfrac{s-n}{s}\,,\ f_{\max}(n)=f_{\max}(n\text{-}T)\wedge \mathrm{n}>\mathrm{T}\\ \mathrm{F}_{\mathrm{acc}}(n-1)\,,\ f_{\max}(n)\neq f_{\max}(n\text{-}T)\wedge \mathrm{n}>\mathrm{T}\\ 0,\ n\le T\end{cases} \tag{9}$$

N is the current evolution algebra. s denotes the biggest evolutionary algebra. Constant *T* is the iterations number when optimum individuals don't change continuously in current population. Constant *C* (0 < *C*) is the adjusting parameter. $f_{\max}(n)$ is the optimal fitness value in the nth generation populations. When the optimal fitness value of the populations does not change for *T* consecutive generation, and the evolution algebra doesn't reach the maximum, the mutation probability is increased to adjust the variation probability, which is calculated by s, *n*, and *C*.

Individual similarity evaluation operator $\mathrm{x}_{\mathrm{sim}}$, individual fitness evaluation operator $\mathrm{y}'_{\mathrm{fit}}$, and population variation adjustment operator $\mathrm{F}_{\mathrm{acc}}(n)$ are used to calculate the current the mutation probability of the individuals in the population. The mutation probability is shown in formula (10).

$$p'_{\mathrm{n}}=\begin{cases}p_0\times \mathrm{y}^{n}_{fit}\times x_{\mathrm{sim}}+\mathrm{F}_{\mathrm{acc}}(n),\ f_{\max}\neq f_{\min}\\ 0,\ f_{\max}=f_{\min}\end{cases} \tag{10}$$

In formula (10), p'_{n} is the *i*th individual's mutation probability in the nth generation population and p_0 is the initial value of mutation probability.

THE MOP-AQGA

MOP-AQGA includes three mechanisms: adaptive adjustment of rotation angle, the cooperative evolution of multiple operators, and multi-universe mechanism. Adaptive adjustment mechanism of rotation angle adjusts rotation angle step length dynamically with the individual fitness. The rotation angles are shown in Table 2 (Ada et al., 2015; Zhi-jian et al., 2019).

Table 2. Adaptive adjustment mechanism of rotation angle adjusts rotation angle

x_i^j	b_{i}	$f(X^j)\ge f(X^t_{best})$	$\Delta\theta_i^j$	$S(\alpha_i^j,\beta_i^j)$			
				$\alpha_i^j\beta_i^j>0$	$\alpha_i^j\beta_i^j<0$	$\alpha_i^j=0$	$\beta_i^j=0$
0	0	false	$\theta_1^j=0$	-	-	-	-
0	0	true	$\theta_2^j=0$	-	-	-	-
0	1	false	$\theta_3^j=\theta^j$	+1	-1	0	±1
0	1	true	$\theta_4^j=\theta^j$	-1	+1	±1	0

Table 2 continued on next page

Table 2 continued

x_i^j	b_i	$f(X^j) \geq f(X_{best}^t)$	$\Delta\theta_i^j$	$S(\alpha_i^j, \beta_i^j)$			
				$\alpha_i^j \beta_i^j > 0$	$\alpha_i^j \beta_i^j < 0$	$\alpha_i^j = 0$	$\beta_i^j = 0$
1	0	false	$\theta_5^j = \theta^j$	-1	+1	± 1	0
1	0	true	$\theta_6^j = \theta^j$	+1	-1	0	± 1
1	1	false	$\theta_7^j = 0$	-	-	-	-
1	1	true	$\theta_8^j = 0$	-	-	-	-

In table 2, $f(X)$ denotes the fitness value of individual *x*; x_i^j is the *i*th gene of the *j*th individual; b_i is the *i*th value of the best individual in current population; $S(\alpha_i^j, \beta_i^j)$ is the rotation direction of rotation angle in polar coordinates; and θ^j is the rotation angle of step length of the *j*th individual. θ^j is defined in formula (11).

$$\theta^j = \begin{cases} \frac{f_j - f_{\min}}{f_{\max} - f_{\min}}(K_2 - K_1) + K_1, & f_{\max} \neq f_{\min} \\ K_1, & f_{\max} = f_{\min} \end{cases} \tag{11}$$

The relationship between rotation angle step length and the current individual fitness value is linear, individuals with higher fitness can be allocated larger rotation angle step length, and individuals with lower fitness can be allocated smaller rotation angle step length. The two methods above can make individuals change their status as soon as possible (Yuan et al., 2021; Zhang et al., 2021).

In the current population, the *i*th evolution of rotation angle step length and rotation direction of the *j*th individuals are shown in formula (12).

$$\Delta\theta_i^j = \theta^j \times S(\alpha_i^j, \beta_i^j) \tag{12}$$

In order to reduce the time cost of the extra calculation and improve the efficiency of the algorithm, the authors use multi-universe mechanism in MOP-AQGA. This paper uses the model of 4 universes to finish the parallel computing algorithm.

The main universe is responsible for the communication and collaboration with other auxiliary universes. Each universe executes population evolution independently. In certain periods, some excellent individuals migrate between main universe and auxiliary universe, so some poor individuals are replaced by excellent individuals. The scale of the universe immigration is usually set in the range 10% to 20% of the population size.

The cooperative evolution of multiple operators based adaptive parallel quantum genetic algorithm is shown in Table 3.

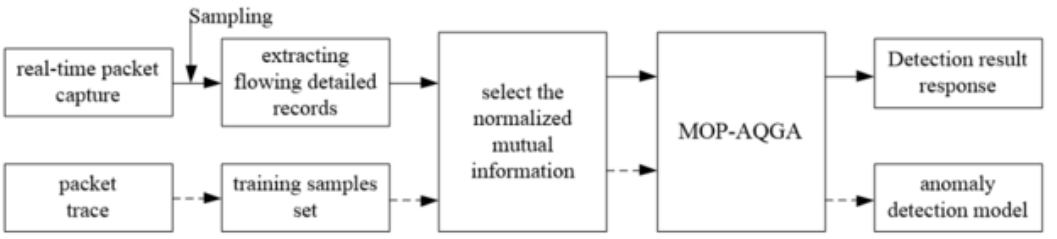
Table 3. MOP-AQGA

The Cooperative Evolution of Multiple Operators Based Adaptive Parallel Quantum Genetic Algorithm
Initialize the population:($t \leftarrow 0$); For each universe, execute following steps: Create new population $\mathrm{Q}(t)$; Observe $\mathrm{Q}(t)$ and obtain the observed state $\mathrm{P}(t)$; Calculate and evaluate the fitness of individuals of $\mathrm{P}(t)$; Select the optimal individuals into $\mathrm{B}(t)$; while(the individuals' lives are less than or equal to the maximum algebra evolution)do For each universe, execute the following steps: $t \leftarrow t+1$; Observe $\mathrm{Q}(t)$ and obtain the observed state $\mathrm{P}(t)$; Calculate and evaluate the new fitness of individuals of $\mathrm{P}(t)$; Calculate the angle rotation step with formula (11); Update $\mathrm{Q}(t)$ with angle rotation step; Select the optimal individuals into $\mathrm{B}(t)$; Calculate collective mutation operator with formula (7) - (9); Calculate individuals' variation probability calculation with formula (10); Individuals mutate with their mutation probability; The optimal individuals migrate among different universes; Select the optimal individuals into $\mathrm{B}(t)$; Save the optimal individuals of all individuals into global variable *b*.

THE ANOMALY DETECTION METHOD BASED ON NMIFS MOP-AQGA

The anomaly detection model based on NMIFS MOP-AQGA has two phases: off-line training and on-line detection. The detection flow based on NMIFS MOP-AQGA is shown in Figure 1.

Figure 1. The detection flow based on NMIFS MOP-AQGA



The flow of the training phase is as follows:

1. The offline cache data are as the input origins. NMIFS module extracts the given traffic feature data and creates the training samples set.
2. Execute the NMIFS algorithm, reduce high-dimensional feature data and get the effective data set with the method that we calculate the normalized mutual information between various features and different behavior categories in the sample data to get the best combination features set.
3. Plug the above best combination feature vector into the MOP-AQGA classifier. Learn, train, and acquire the anomaly detection individuals.

The main work of the detection phase is:

1) After sampling the received data packet, extract the flow detailed records with the pre-definition features in actual applications.
2) Select d-dimensional important features from the flow detailed record through the NMIFS module. Then, compare it with the detection model in the training phase and output the accurate detection results.

SIMULATIONS AND ANALYSIS FOR NS-KDD AND CICIDS2017

To verify the effectiveness of the proposed method NMIFS MOP-AQGA, in the experiments, the authors take the intrusion scenario correlation benchmarks NS-KDD and CICIDS2017. Therefore, they use some common performance indicators as parameters to detect antigens and present comparison analysis respectively according to NSL - KDD and CICIDS2017 data set. NSL - KDD and CICIDS2017 both include small attacks samples, with which the authors research the attack classification performances of NMIFS MOP - AQGA algorithm for small samples. With CICIDS2017 data set, the purpose is to research classification performances about new attacks.

DATA SET AND SIMULATION ENVIRONMENT SETUP

The minimum rotation angle step length is $K_1=0.001\pi$, the maximum rotation angle step length is $K_2=0.05\pi$, the initial mutation probability is $P_0=0.8$, the variation operator adjustments constant C is 0.08, and the parallel universe number is 4.

The anomaly detection algorithm NMIFS MOP-AQGA is implemented with C, simulation environment: CPU is the Intel Pentium of 4, 3.20 GHz, memory is 16 GB, and the operating system for Microsoft Windows 2016.

NSL-KDD

Lincoln laboratory provided NSL - KDD data set for experiment simulations of IDS. The training sample set KDDTrain + includes 125,973 records. The test sample set includes KDDTest - 21 and KDDTest +, which contain Normal, Dos, the Probe, and U2R and R2L five species of samples.

When the authors execute the experimental simulation, all the training samples are used to get the optimal features, and test samples KDDTest + are applied to test the NMIFS MOP - AQGA algorithm. Sample distributions of the data set are shown in Table 4.

Table 4. *Sample distribution of the data set NSL-KDD (He et al., 2017)*

No	Type	KDDTrain+	KDDTest+
1	Normal	67,343	9,710
2	DoS	45,927	7,458
3	Probe	11,656	2,422
4	U2R	52	67
5	R2L	995	2,887
Sum		125,973	22,544

As a result that U2R attacks samples are less than others, 67 U2R attacks of the KDDTest + data set are put as the experimental data also.

CICIDS2017

CICIDS2017 includes normal samples and 15 types of attacks, and contains 2,830,743 samples. In CICIDS2017, 60 percent of data are as the training samples, the rest of data are as testing ones. The distributions of samples are shown in Table 5.

Table 5. CICIDS2017' distribution

Type	Abbreviation	Number
Benign	Normal	2,273,097
Distributed Denial-of-service (DDos)	Ddos	128,027
Port Scan	PortS	158,930
Bot	Bot	1,966
Infiltration	Inf	36
Brute Force	XSS	2,180
Web Attack Structured Query Language (SQL)		
Injection Cross-site Scripting (XSS)		
File Transfer Protocol (FTP)- Patator	FTP	7,938
Secure Shell (SSH)-Patator	SSH	5,897
Denial-of-service (Dos) GoldenEye	DosGE	10,293
DoS Hulk	DoSHu	231,073
DoS Slowhttptest	DoSSH	5,499
DoS Slowloris	DoSSL	5,796
Heartbleed	Heart	11

THE DATA PREPROCESSING AND EVALUATION STANDARDS

There are four stages in the detection experiment: data normalization, data reduction, training, and testing. The effectiveness of some small features values may be easily ignored, since there are large

differences among features in the data set. The features need to be normalized before we apply NMIFS to select features. The authors assumed the data set contains *n* records and $f_j[i]$ represents the *i*th feature in the *j*th record. The authors first calculated its mean and standard deviation by adopting the following equations:

$$\overline{f}_j[i] = \frac{1}{n}\sum_{j=1}^{n} f_j[i] \tag{13}$$

$$s_j[i] = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}\left(f_j[i] - \overline{f}_j[i]\right)^2} \tag{14}$$

Where $\overline{f}_j[i]$ and $s_j[i]$ stand for the mean and standard deviation of the *i*th feature, respectively. Then, the authors normalized all features using (15) (Imamverdiyev & Abdullayeva, 2018):

$$\hat{f}_j[i] = \frac{f_j[i] - \overline{f}_j[i]}{s_j[i]} \tag{15}$$

The evaluation criteria of the test results are as follows (Moustafa, 2017):

$$DR = \frac{TP}{TP + FN} \tag{16}$$

$$FAR = \frac{FP}{TN + FP} \tag{17}$$

$$Pre = \frac{TP}{TP + FP} \tag{18}$$

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

$$F1\text{-}score = \frac{2 \times DR \times Pre}{DR + Pre} \tag{20}$$

TP denotes that the samples which belong to intrusions are correctly recognized; *TN* denotes the samples which belong to normal data are correctly recognized; *FP* denotes the samples which do not

belong to intrusions are wrongly recognized as behaviors which are belonging to intrusions; and FN denotes the samples which belong to normal data are wrongly recognized as behaviors which are not belonging to normal data (Tao et al., 2021; Xiaohui et al., 2019).

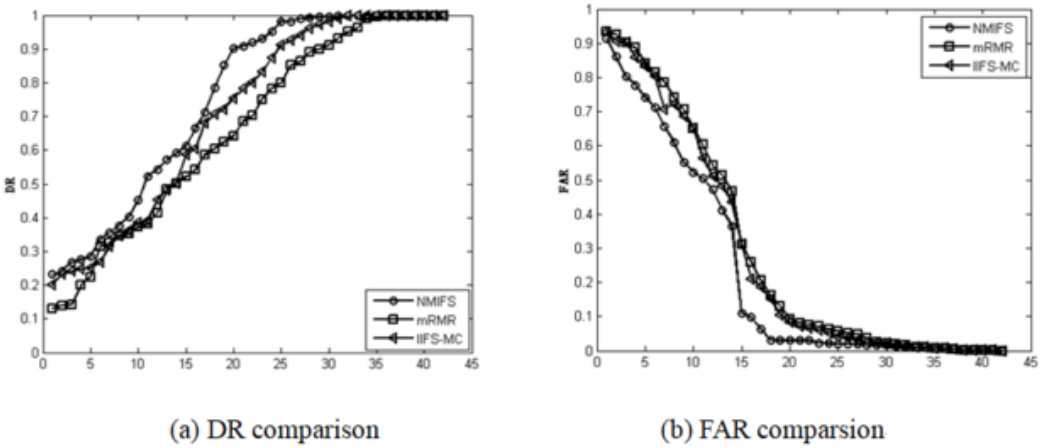
THE BEST FEATURES SET SELECTION BASED ON NMIFS

NMIFS algorithm is adopted to obtain the normalized mutual information between the statistical features and category labeled features. According to the normalized mutual information, the subsequent classifier combines these features by their importance. The NSL- KDD and CICIDS2017 are used to test the NMIFS The results and comparisons are shown in following subsections.

NSL-KDD

The authors sort the features according to the normalized mutual information, so that the subsequent classifier combines different numbers of features which can be tested though the C4.5 decision tree classifier, as seen in Figure 2. They also compared the proposed algorithm with the mRMR algorithm and IIFS-MC algorithm. The convergence rates of DR and FAR are obviously faster than the mRMR and IIFS-MC algorithms.

Figure 2. The comparisons on feature selection of FMIS, mRMR, and IIFS-MC



(a) DR comparison (b) FAR comparsion

It's concluded from the experiments: when the number of features is 20, the difference between the NMIFS algorithm's DR and FAR is greatest, meanwhile the DR is more than 90%. While the number of features is 25 and 20, the difference between the mRMR and IIFS-MC algorithm's (Panigrahi et al., 2021) DR and FAR are greatest respectively. Therefore, the authors choose the top 20 most important features as the best features in the NMIFS algorithm. The can get the conclusion that among the same type of classification algorithms, NMIFS has a better feature reduction performance than mRMR. Table 6 shows the best feature subset among 3 algorithms.

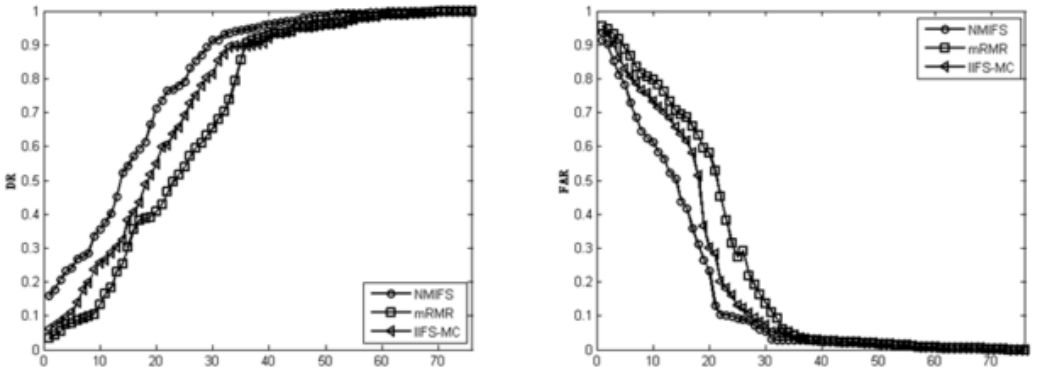
Table 6. The best feature subset

Algorithm	The Features Number	Selected Features (Label Number)
NMIFS	20	12,3,6,23,2,32,5,24,36,35,33,1,30,37,4,29,39,38,25
mRMR	25	32,27,23,5,3,12,13,22,11,2,9,37,28,38,1,4,6,14,29,40,39,35,33,41,30
IIFS-MC	20	2,3,17,4,5,18,12,6,15,16,8,19,11,13,14,23,28,10,22,36

CICIDS2017

The features selection process is the same with that of NSL- KDD above. The comparisons on feature selection of FMIS, mRMR, and IIFS-MC are shown in Figure 3.

Figure 3. The comparisons on feature selection of FMIS, mRMR, and IIFS-MC



(a) DR comparison (b) FAR comparsion

With the number of features increasing, the DR is rising, and the FAR is decreasing. The convergence rate of NMIS is the fastest one among the tree algorithms.

For NMIFS, when the number of features is 32, the difference between DR and FAR is greatest, meanwhile the DR is more than 93%. While the authors take the mRMR and IIFS-MC to simulate, the numbers of features is 37 and 34 and the differences between DR and FAR are greatest, meanwhile the DRs are about 90%. Table 7 shows the best feature subset through the two algorithms. There are 76 features in CICIDS2017, after features selection, the number of features is 32, the total time of intrusion detection is almost half of that before reduction.

Table 7. The best feature subset

Algorithm	The Features Number	Selected Features (Label Number)
NMIFS	32	1,75,73,18,33,67,11,35,25,55,56,57,58,20,59,32,49,23,45,50,43,74,76,15,19,31,14,30,44,71,16,38
mRMR	37	1,75,73,18,28,33,67,11,35,25,17,55,56,57,58,20,59,60,32,49,23,45,50,43,74,76,15,19,31,14,70,30,44,71,68,16,38
IIFS-MC	34	1,25,20,75,76,73,18,23,28,31,33,55,56,57,58,59,60,32,49,45,50,43,74,19,14,70,30,44,68,11,16,38,71

In conclusion, NMIFS can get the highest convergence speed among the three algorithms. So when the authors simulate with NSL-KDD, they use features collection {12, 3, 6, 23, 2, 32, 5, 24, 36, 35, 33, 1, 30, 37, 4, 29, 39, 38, 25} to generate the detectors; when they simulate with CICIDS2017, they use features collection {1, 75, 73, 18, 33, 67, 11, 35, 25, 55, 56, 57, 58, 20, 59, 32, 49, 23, 45, 50, 43, 74, 76, 15, 19, 31, 14, 30, 44, 71, 16, 38} to generate the detectors.

THE ANOMALY DETECTION PERFORMANCE RESULTS COMPARED WITH RELATED WORKS

In pre-section, the authors get the optimal features sets. With those features, the antibodies are put into the classifier. In the same way, NSL- KDD and CICIDS2017 are applied to verify the MOP-AQGA. The results are shown as follows.

NSL-KDD

The classification and anomaly detection performance results include the confusion matrix, the Receiver Operating Characteristic (ROC), classification results, and anomaly detection results.

The Confusion Matrix

The confusion matrix is adopted to evaluate the performances of NMIFS MOP-AQGA. The 22,544 samples in Table 4 are used to verify the NMIFS MOP-AQGA algorithm. The normal samples, Dos attacks, the Probe attacks, U2R attacks and R2L attacks are applied to test respectively, the results are shown in Table 8. The column data denote the actual classified results, the row data denote the predicted classified results.

Table 8. The confusion matrix of NMIFS MOP-AQGA

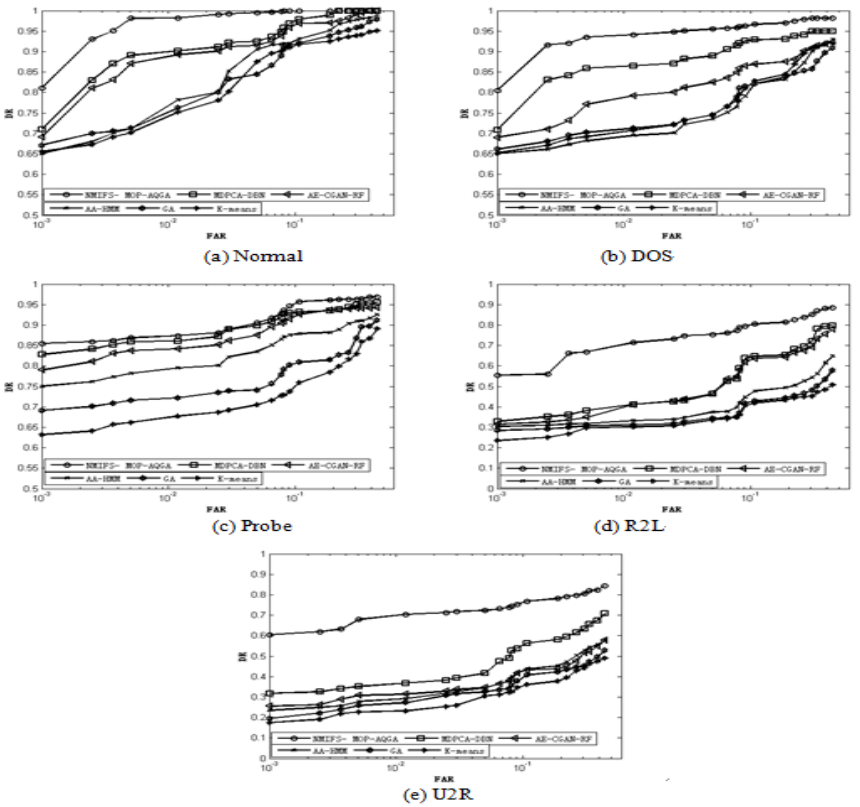
	Actual Values					
Predicted Values		**Norm**	**DoS**	**Probe**	**U2R**	**R2L**
	Norm	9,201	21	79	3	16
	DoS	110	7,360	58	8	112
	Probe	98	72	2,274	6	2
	U2R	121	0	1	45	6
	R2L	180	5	10	5	2,751

The ROC Curves

In order to verify the proposed algorithm, the authors compare the NMIFS MOP-AQGA algorithm with other classifiers. So they run the NMIFS MOP-AQGA algorithm with NSL-KDD and get the experiment results which are used in comparisons.

The K – means (Alyaseen et al., 2017), the AE - CGAN – RF (Jiadong et al., 2019), AA – HMM (Song et al., 2018), MDPCA- DBN (Yang et al., 2019), GA (Ying-Wu et al., 2010), and the NMIFS MOP - AQGA algorithm proposed in this paper are used to train and test with experimental data set. The ROC curves on five kinds of data set are shown in Figure 4. The results show: both with the normal data set, and the abnormal data set (DOS, the Probe, R2Land U2R), we can get a lower FAR a higher DR. Especially for R2L and U2R, the small samples set, we can get a higher detection rate and lower false detection rate. The detection rate of R2L is about 85.7% and the detection rate of U2R is about 76.3%.

Figure 4. The ROC curves of different types of data sets with various algorithms



From Figure 4, for the five types of samples sets, the NMIFS MOP-AQGA algorithm has higher DR and lower FAR than the other five algorithms.

*T*he Classification Detection Performance

In order to test the classification performance of the NMIFS MOP - AQGA algorithm, especially about the detection of small samples set, such as U2R and R2L, the NMIFS MOP - AQGA algorithm

is compared with other algorithms in table 9. Algorithm detection rates according to different attack types are shown in Table 9.

Table 9. The DR comparisons of different algorithms

Method	Norm	Dos	Prb	U2R	R2L
Mixed multilayer model (Alyaseen et al., 2017)	98.13	99.54	87.22	21.93	31.39
MDPCA-DBN (Yang et al., 2019)	97.38	81.09	73.94	17.25	6.50
Outlier+RF (Jiadong et al., 2019)	97.66	97.32	95.34	21.05	31.96
NMIFS MOP-AQGA	94.76	98.69	93.89	67.17	95.29

In Table 9, it is concluded that the proposed NMIFS MOP - AQGA algorithm has the highest DR among all the algorithms for the three types of attacks: Probe, U2R, and R2L attacks. For Dos attacks, the DR is lower than Mixed multilayer model (Alyaseen et al., 2017); For normal logs, the DR is lower. The NMIFS MOP - AQGA algorithm has a higher detection rate for all types of attacks; especially, it can improve the DR of the algorithm for small samples datasets.

Literature (Yang et al., 2019) gave the MDPCA – DBN's confusion matrix. With the results of the confusion matrix the authors calculated the DR and the Pre - and F1 scores. The results, compared with results of this paper, are shown in Table 10.

Table 10. The classification performances comparisons of different algorithms

Type	DR		Pre		F1-Score	
	MDPCA-DBN	NMIFS MOP-AQGA	MDPCA-DBN	NMIFS MOP-AQGA	MDPCA-DBN	NMIFS MOP-AQGA
Norm	71.42	94.76	97.38	98.72	82.40	96.69
Dos	96.34	98.69	81.09	96.23	88.06	97.44
Probe	85.85	93.88	73.94	92.74	79.45	93.31
U2R	11.82	67.16	6.50	26.01	8.39	37.50
R2L	57.30	95.29	17.25	93.22	26.51	94.24
average	64.54	89.96	55.23	81.39	56.96	83.84

The conclusions are shown in Table 10: the NMIFS MOP - AQGA algorithm has higher DR, Pre, and F1-Score than MDPCA-DBN. Especially for the U2R and R2L, the proposed algorithm has better classification performances to small samples sets.

*T*he Abnormal Detection Performance

Last, the authors compared the abnormal detection performance of the NMIFS MOP - AQGA algorithm with the algorithms in this paper. In order to finish the abnormal detection, all the attacks are as abnormal samples for testing. The testing samples are for testing; each group of data are run 10 times and the average values are calculated. The comparisons of the abnormal detection performance are shown in Table 11.

Table 11. The comparisons of the abnormal detection performance(N/A denotes the results are unknown)

Algorithm	DR	Acc	FAR	Pre	F1-score
K-means (Alyaseen et al., 2017)	95.17	95.75	1.87	N/A	N/A
AE-CGAN-RF (JooHwa & KeeHyun, 2019)	61.57	66.18	13.06	95.51	74.87
AA-HMM (Song et al., 2018)	91.06	93.48	N/A	93.63	92.33
MDPCA-DBN (Yang et al., 2019)	93.55	94.36	2.34	N/A	N/A
GA+Fuzzy (Varzaneh & Rafsanjani, 2021)	95.33	N/A	0.18	N/A	N/A
WIDMoDS (Feng & Dou, 2021)	N/A	99.60	N/A	N/A	96.00
MS-NIDAM (Xi et al., 2021)	about 92.00	N/A	about 5.00	N/A	N/A
NMIFS MOP-AQGA	96.76	98.72	0.89	98.72	96.69

From table 11, the results showed that the proposed NMIFS MOP - AQGA algorithm's false positives are 0.71% higher and detection rate is 1.43% higher than the GA + Fuzzy algorithm in Varzaneh and Rafsanjani (2021); all the detection indicators are better than other algorithms, and there is a better balance between the DR and the FAR.

CICIDS2017

Because CICIDS2017 has more and newer types of attacks, we take it as simulation dataset. The detection performance results include the confusion matrix, the classification results comparisons.

The Confusion Matrix

In Table 12, 1,132,296 testing samples are applied to simulate with the NMIFS MOP - AQGA algorithm. The testing set includes 909,239 normal samples, 51,211 samples of Ddos, 63,572 PortS samples, 786 Bot samples, 14 Inf attack samples, 872 XSS attack samples, 3,175 FTP attack samples, 2,359 SSH attack samples, 4,117 DosGE attack samples, 92,429 DosHu attack samples, 2,200 attack samples, 2,318 DoSSL attack samples, and 4 Heart attacks samples. The results are shown in Table 12.

Table 12. The confusion matrix for CICIDS2017

		Actual Values												
Predict Values		**Normal**	**Ddos**	**PortS**	**Bot**	**Inf**	**XSS**	**FTP**	**SSH**	**Dos GE**	**DoS Hu**	**DoS SH**	**DoS SL**	**Heart**
	Normal	908,527	32	5	136	1	13	2	3	3	193	6	5	0
	Ddos	1	51,165	0	0	1	0	0	0	0	0	0	0	0
	PortS	307	0	63,558	0	0	0	0	0	0	0	1	1	0
	Bot	79	0	0	647	0	0	0	0	0	0	0	0	0
	Inf	0	0	0	0	12	0	0	0	0	0	0	0	0
	XSS	0	0	0	3	0	857	0	0	0	0	0	0	0
	FTP	0	0	1	0	0	0	3170	0	0	0	0	0	0
	SSH	0	0	0	0	0	0	0	2354	0	0	0	0	0
	DosGE	15	6	0	0	0	0	0	0	4106	9	1	1	0
	DoSHu	295	8	8	0	0	0	0	0	7	92227	1	0	0
	DoSSH	13	0	0	0	0	2	0	2	1	0	2187	3	0
	DoSSL	2	0	0	0	0	0	3	0	0	0	4	2308	0
	Heart	0	0	0	0	0	0	0	0	0	0	0	0	4

*T*HE CLASSIFICATION DETECTION PERFORMANCE

In order to verify the proposed algorithm, classification performance comparison between NMIFS MOP - AQGA and other algorithms are shown in this section.

Performance comparisons are given for the two data sets respectively. The algorithm performance indexes in this section are calculated by confusion matrix. The detection performances of Single - RF, the RAFID, and AE - CGAN - RF (JooHwa & KeeHyun, 2019) are given, so the authors compare the NMIFS MOP - AQGA algorithm's performances with these three algorithms for CICIDS2017 data sets. The results are shown in Figure 5, Figure 6, and Figure 7.

Figure 5. Comparisons of classification performances for CICIDS2017

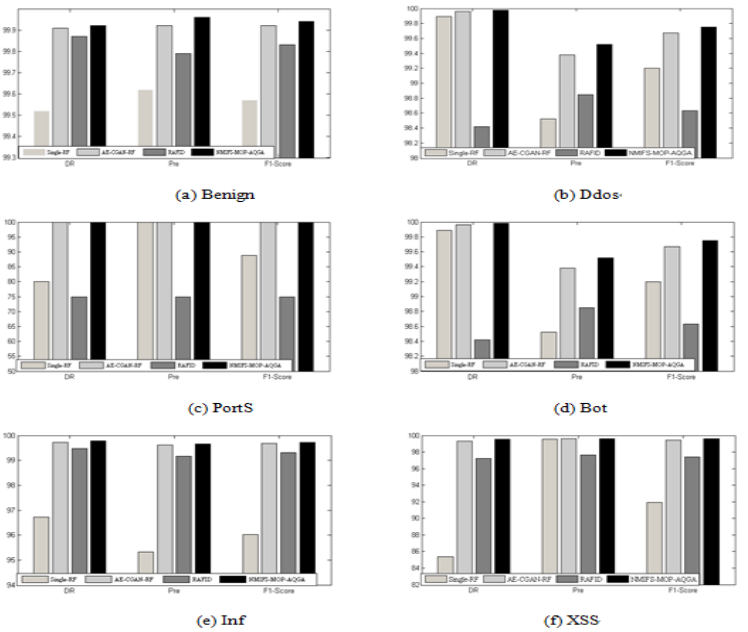


Figure 6. Comparisons of classification performances for CICIDS2017

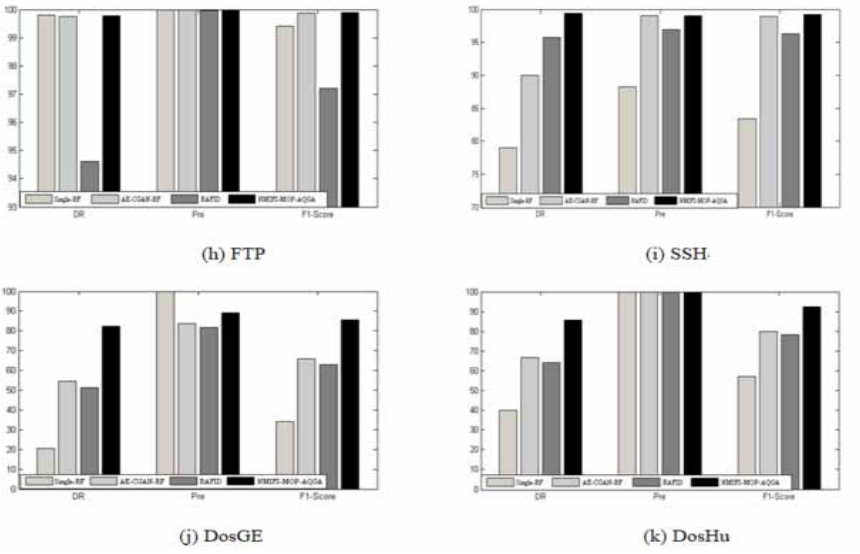
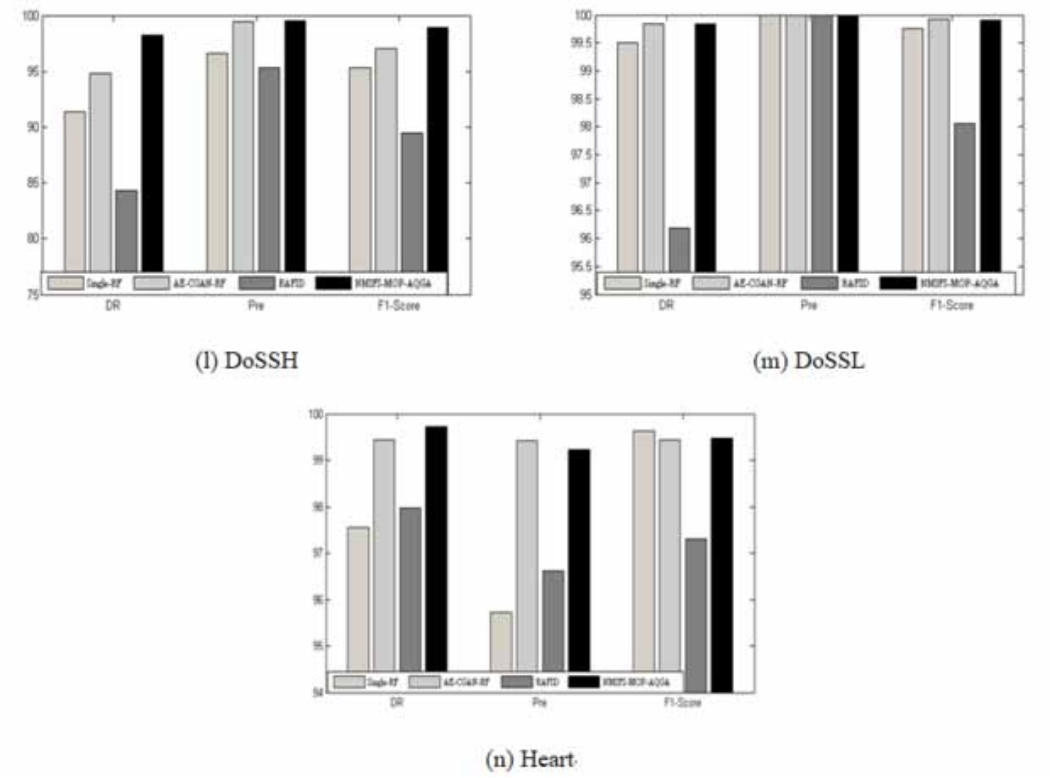


Figure 7. Comparisons of classification performances for CICIDS2017



(l) DoSSH (m) DoSSL

(n) Heart

For DR, Pre, and F1-Score the three detection indicators, the NMIFS MOP – AQGA algorithm, are superior to single random forest algorithm, RAFID algorithm, and similar to the AE - CGAN - RF algorithm (JooHwa & KeeHyun, 2019). It has better classification performances for the 13 datasets. And for small samples sets, such as Bot and Infiltration attacks, the detection performances are improved. The synthetic detection performances of the NMIFS MOP - AQGA algorithm proposed in this paper are better than that of the Single - RF RAFID and AE - RF algorithm.

In conclusion, NSL - KDD and CICIDS2017 are obtained from the simulated network environment which is similar to the actual one. With the NMIFS MOP - AQGA algorithm, classic intrusions can be classified, new attacks can be identified, the algorithm is feasible, and the algorithm has better detection performance. The MOP - AQGA algorithm can optimize antibodies and improve the detection performance. Through the above experiments, it is concluded that the proposed algorithm is practical to the actual network intrusion detection. With the MOP - AQGA algorithm, better detectors can be generated, so it has better adaptivity.

CONCLUSION

This paper applied the NMIFS algorithm to obtain the best features for sample features sets. With NMIFS, the authors got an effective dimension reduction from high dimensional features, so that the detection speed was improved. Also, the authors developed a cooperative evolution of multiple operators based adaptive parallel quantum genetic algorithm (MOP-AQGA), in which they used individual similarity evaluation operator, individual fitness evaluation operator, and population variation adjustment operator to get individuals' mutation probability and, at the same time, applied

adaptive parallel quantum genetic algorithm to ensure the diversity of individuals. Experimental results on anomaly data from real network traffic showed that, the proposed method, NMIFS MOP-AQGA, has a higher detection accuracy, lower false positive rate, and better adaptability, especial according to small samples sets compared with existing anomaly detection algorithms. While the defect of algorithm is for unknown attacks research. The future work is to research the zero day attacks and how to detect the unknown attacks and deploy the algorithm in the Internet-of-Things (Tewari & Gupta., 2020).

CONFLICT OF INTEREST

The authors of this publication declare there is no conflict of interest.

FUNDING AGENCY

This research was supported by the National Natural Science Foundation of China [61502436]; and the Project of Science and technology tackling key problems in Henan Province [202102210149].

ACKNOWLEDGMENT

The authors would like to thank the Faculty of software, University of Zhengzhou University of Light Industry, for their professional advice about this study.

REFERENCES

Ada, C., Wu, P., & Chu, F. (2015). Improved quantum-inspired evolutionary algorithm for large-size lan reservation. *IEEE Transactions on Systems, Man, and Cybernetics. Systems*, *45*(12), 1535–1548. doi:10.1109/TSMC.2015.2417509

Alyaseen, W. L., Othman, Z. A., Nazri, M. Z. A., & Nazri, M. Z. A. (2017). Multi-level hybrid support vector machine and extreme learning machine based on modified K-means for intrusion detection system. *Expert Systems with Applications*, *67*(1), 296–303. doi:10.1016/j.eswa.2016.09.041

Cheng, Y. F., Shao, W., Zhang, S. J., & Li, Y. P. (2016). An improved multi-objective genetic algorithm for large planar array thinning. *IEEE Transactions on Magnetics*, *3*(52), 1–4. Advance online publication. doi:10.1109/TMAG.2015.2481883

Chunhui, W., & Wenjuan, L. (2021). Enhancing intrusion detection with feature selection and neural network. *International Journal of Intelligent Systems*, *7*(36), 3087–3105. doi:10.1002/int.22397

Este'vez, P. A., Tesmer, M., Perez, C. A., & Zurada, J. M. (2009). Normalized mutual information feature selection. *IEEE Transactions on Neural Networks*, *20*(2), 189–201. doi:10.1109/TNN.2008.2005601 PMID:19150792

Feng, T., & Dou, M. F. (2021). A weighted intrusion detection model of dynamic selection. *Applied Intelligence*, *51*(7), 4860–4873. doi:10.1007/s10489-020-02090-8

He, Y. B., Mendis, G. J., & Wei, J. (2017). Real-time detection of false data injection attacks in smart grid: A deep learning-based intelligent mechanism. *IEEE Transactions on Smart Grid*, *5*(8), 2505–2516. doi:10.1109/TSG.2017.2703842

Imamverdiyev, Y., & Abdullayeva, F. (2018). Deep learning method for denial of service attack detection based on restricted Boltzmann machine. *Big Data*, *2*(6), 159–169. doi:10.1089/big.2018.0023 PMID:29924649

Jiadong, R., Xinqian, L., Qian, W., Haitao, H., & Xiaolin, Z. (2019). An multi-level intrusion detection method based on KNN outlier detection and random forests. *Journal of Computer Research and Development, 56*(3), 566-575. 10.7544/issn1000-1239.2019.20180063

Kumar, G. K., Kumar, R. R., Basha, M. S., & Reddy, K. N. (2019). Intrusion detection using an ensemble of support vector machines. *Advances in Engineering, Management and Sciences, 3*(S), 266-275. 10.26782/jmcms.spl.3/2019.09.00020

Lashkia, G. V., & Anthony, L. (2004). Relevant irredundant feature selection and noisy example elimination. *IEEE Transactions on Systems, Man, and Cybernetics. Part B, Cybernetics*, *34*(2), 888–897. doi:10.1109/TSMCB.2003.817106 PMID:15376837

Lee, J. H., & Park, K. H. (2019). AE-CGAN-RF model based high performance network intrusion detection system. *Applied Sciences-Basel*, *20*(9), 1–14. doi:10.3390/app9204221

Li, D., Deng, L., Gupta, B. B., Wang, H., & Choi, C. (2019). A novel CNN based security guaranteed image watermarking generation scenario for smart city applications. *Information Sciences*, *479*, 432–447. doi:10.1016/j.ins.2018.02.060

Moustafa, N. (2017). *Designing an online and reliable statistical anomaly detection framework for dealing with large high-speed network traffic*. University of New South Wales.

Panigrahi, R., Samarjeet, B., Akash, K. B., Muhammad, F. I., & Moumita, P. (2021). A consolidated decision tree-based intrusion detection system for binary and multiclass imbalanced datasets. *Mathematics*, *9*(7), 751. Advance online publication. doi:10.3390/math9070751

Peng, L. Z., Zhang, H. L., Yang, B., & Chen, Y. H. (2014). Feature evaluation for early stage internet traffic identification. *Algorithms and Architectures for Parallel Processing*, *8630*, 511–525. doi:10.1007/978-3-319-11197-1_39

Song, C. Y., Pons, A., & Yen, K. (2018). AA-HMM: An Anti-Adversarial Hidden Markov Model for network-based intrusion detection. *Applied Sciences-Basel*, *12*(8), 1–25. doi:10.3390/app8122421

Tao, L., Zhaojie, W., Yuling, C., Chunmei, L., Yanling, J., & Yixian, Y. (2021). Is semi-selfish mining available without being detected? *International Journal of Intelligent Systems*, ●●●, 1–22. doi:10.1002/int.22656

Teng, H., Qixiang, Z., Jiabao, L., Ruitao, H., Xianmin, W., & Ya, L. (2020). Adversarial attacks on deep-learning-based SAR image target recognition. *Journal of Network and Computer Applications*, *2020*(162), 1–12. doi:10.1016/j.jnca.2020.102632

Tewari, A., & Gupta, B. B. (2020). Security, privacy and trust of different layers in Internet-of-Things (IoTs) framework. *Future Generation Computer Systems*, *108*, 909–920. doi:10.1016/j.future.2018.04.027

Varzaneh, Z. A., & Rafsanjani, M. K. (2021). Intrusion detection system using a new fuzzy rule-based classification system based on genetic algorithm. *Intelligent Decision Technologies*, *15*(2), 231–237. doi:10.3233/IDT-200036

Wanwei, H., Jianwei, Z., Haiyan, S., Huan, M., & Zengyu, C. (2017). An anomaly detection method based on normalized mutual information feature selection and quantum wavelet neural network. *Wireless Personal Communications*, *96*(2), 2693–2713. doi:10.1007/s11277-017-4320-2

Xi, L., Wang, R. D., Yao, Z. Y., & Zhang, F. B. (2021). Multisource neighborhood immune detector adaptive model for anomaly detection. *IEEE Transactions on Evolutionary Computation*, *3*(25), 582–594. doi:10.1109/TEVC.2021.3058687

Xiaohui, K., Ming, Z., Hu, L., Guang, Z., Huayang, C., & Zhendong, Z. et al.. (2019). DeepWAF: Detecting web attacks based on CNN and LSTM models. *Cyberspace Safety and Security*, *11983*(PT II), 121–136. doi:10.1007/978-3-030-37352-8_11

Yang, Y. Q., Zheng, K. F., Wu, C. H., Niu, X. X., & Yang, Y. X. (2019). Building an effective intrusion detection system using the modified density peak clustering algorithm and deep belief networks. *Applied Sciences-Basel*, *2*(9), 238–262. doi:10.3390/app9020238

Yilei, W., Guoyu, Y., Tao, L., Lifeng, Z., Yanli, W., Lishan, K., & Yi, D. (2021). Optimal mixed block withholding attacks based on reinforcement learning. *International Journal of Intelligent Systems*, *12*(35), 2032–2048. doi:10.1002/int.22282

Ying-Wu, Z., Jia-Hai, Y., & Jin-Xiang, Z. (2010). Anomaly detection based on traffic information structure. *Journal of Software*, *21*(10), 2573–2583. doi:10.3724/SP.J.1001.2010.03698

Yuan, F., Chen, S., Liang, K., & Xu, L. (2021). *Research on the coordination mechanism of traditional Chinese medicine medical record data standardization and characteristic protection under big data environment*. Shandong People's Publishing House.

Zhang, X., Wang, Y., Geng, G. & Yu, J. (2021). Delay-Optimized Multicast Tree Packing in Software-Defined Networks *IEEE Transactions on Services Computin*g, 1-14. 10.1109/TSC.2021.3106264

Zhi-jian, Q., Yu-hang, C., Pan-jing, L., Xiao-hong, L., & Cai-hong, L. (2019). Cooperative evolution of multiple operators based adaptive parallel quantum genetic algorithm. *Acta Electonica Sinica*, *47*(2), 266–273. doi:10.3969/j.issn.0372-2112.2019.02.002

APPENDIX A.

Table 13. Abbreviations table

Abbreviation	Full Name
AE – CGAN – RF	autoencoder - conditional, the generative adversarial networks and the random forest
NMIFS MOP- AQGA	normalized mutual information feature selection and cooperative evolution of multiple operators based on adaptive parallel quantum genetic algorithm
IDS	intrusion detection system
NMIFS	normalized mutual information feature selection
MOP- AQGA	cooperative evolution of multiple operators based on adaptive parallel quantum genetic algorithm
SVM	support vector machine
FSA	Feature selection algorithm
mRMR	minimal- Redundancy- Maximal- Relevance
MDPCA-DBN	Modified Density Peak Clustering Algorithm and Deep Belief Networks
AA-HMM	anti-adversarial hidden markov model for network-based intrusion detection
GA	Generation algorithm
MS-NIDAM	multisource neighborhood immune detector adaptive model
DDos	Distributed Denial-of-service
SQL	Web Attack Structured Query Language
XSS	Injection Cross-site Scripting
FTP	File Transfer Protocol
SSH	Secure Shell
Dos	Denial-of-service
IIFS-MC	Improved Infinite Feature Selection for Multiclass Classification (IIFS-MC)
ROC	Receiver Operating Characteristic