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

An Intrusion Detection System Based on Normalized Mutual Information Antibodies Feature Selection and Adaptive Quantum Artificial Immune System

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

ABSTRACT

The intrusion detection system (IDS) has lower speed, less adaptability, and lower detection accuracy especially for small samples sets. This paper presents a detection model based on normalized mutual antibodies information feature selection and adaptive quantum artificial immune with cooperative evolution of multiple operators (NMAIFS MOP-AQAI). First, for a high intrusion speed, the NMAIFS is used to achieve an effective reduction for high-dimensional features. Then, the best feature vectors are sent to the MOP-AQAI classifier, in which vaccination strategy, the quantum computing, and cooperative evolution of multiple operators are adopted to generate excellent detectors. Lastly, the data is fed into NMAIFS MOP-AQAI which ultimately generates accurate detection results. The experimental results on real abnormal data demonstrate that the NMAIFS MOP-AQAI has higher detection accuracy, lower false negative rate, and a higher adaptive performance than the existing anomaly detection methods, especially for small samples sets.

KEYWORDS

Antibody, Antibody Information Entropy, Antigen, Artificial Immune, Mutual Information Antibodies Features Selection, Operator, Quantum, Vaccination

INTRODUCTION

The intrusion detection system (IDS) is one of the most crucial techniques proposed for data integrity and confidentiality (Sahar et al., 2020). Several modern techniques (Kumar et al., 2019) existing in the literature address these issues, such as deep learning, support vector machines, k-means, clustering, outliers, random forest, aggregation, genetic algorithm (GA), and artificial immune (AI) systems (Castillo-Zúñiga et al., 2020; Tewari & Gupta, 2020).

Redundant attributes are bound to affect the rate of IDS when facing substantial data volumes with multiple attributes (Sahoo & Gupta, 2021). Therefore, the features selection algorithm (FSA) plays an important role and is the key phase in data preprocessing (Anupama et al., 2021; Lv et al., 2020; Zhang et al., 2021).

To solve these problems, the K-means algorithm was used to develop a training data set, and then a multi-layer hybrid intrusion detection model improved intrusion classification (Alyaseen et al., 2017). Wu et al. (2020) proposed a network intrusion detection method based on semantic re-encoding

DOI: 10.4018/IJSWIS.308469 *Corresponding Author

(SR) and deep learning to improve the detection speed. Chou et al. (2020) adopted an incremental approach to choose the minimal Redundancy-Maximal Relevance (mRMR) criterion, which is used to calculate the mean value of redundant attributes to reduce the effects of β. The advantage of the mRMR criterion is that with lower computational resources, we can get the best features; the drawback is that there are more differences in information entropy. Fatemeh et al. (2011) improved the MIFS, MIFS -u, and mRMR algorithms to reduce the effects among attributes due to the mutual information deviation. They proposed a normalized mutual antibodies information entropy feature selection (NMIFS) algorithm, which had a higher performance for feature selection. Nguyen et al. (2017) designed a mutual information feature selection (MIFS) algorithm; however, with the number of features increasing, MIFS may choose some redundant features. Huang et al. (2017) adopted FMIFS and quantum wavelet neural network (QWNN) to reduce network logs' redundant attributes, improving the algorithm's speed. So we adopted the NMAIFS to reduce the redundant attributes of network logs to increase the speed of IDS. On the other hand, the appropriate classifier greatly influences the anomaly detection performance (Alshdadi et al., 2021; Madan & Bhatia, 2021).

Lee and Park (2019) designed an auto-encoder-conditional and the generative adversarial networks and random forest (AE-CGAN-RF), auto-encoder-conditional method was used to reduce the redundant attributes, and a random forest was used to classify the intrusion. Feng and Dou (2021) gave an intrusion detection model based on dynamic weighted values (WIDMoDS). The hierarchical clustering algorithm with evaluation indexes was adopted to get the weight values, and the voting algorithm with weight values was used to classify the detected data.

To improve the adaptive performance, resistance to competition, and online learning ability of IDS, AI is used in intrusion detection research. Naila et al. (2020) designed a negative selection for network anomaly detection (NSNAD), and they improved the negative selection algorithm to get the anonymous detection classifier. Sahar et al. gave an internet of things intrusion detection system based on AI using deep learning (DL) and dendritic cells algorithm (DCA) to identify internet invasions and reduce the false positives rate. Yang et al. (2019) proposed an effective IDS using the Modified Density Peak Clustering Algorithm and Deep Belief Networks (MDPCA-DBN); they used the MDPCA and A-DBN to reduce the size of the training set, solve the imbalance of samples, and therefore improve the detection efficiency. Song et al. (2018) proposed an anti-adversarial hidden Markov model for network-based intrusion detection (AA-HMM). Ehsan et al. (2021) proposed a new complex mixed artificial immune intrusion detection system; the system integrated the negative selection algorithm (NSA) and the DCA for detectors. Chou et al. (2020) used AI and the parallel automaton (PA) method to design a high adaptive hybrid intrusion detection algorithm; the state automaton theory was used to define the different data states; the artificial immune algorithm was used to convert the states. Xi et al. (2021) introduced immune adaptive and feedback mechanism to build a multi-source neighborhood immune detector adaptive model (MS-NIDAM). As a result, the detectors can be adaptively evolved in a more targeted search domain. These algorithms can generally improve the adaptive performance of intrusion detection, but for the limitations of antibodies, which may lead to local convergence (Yilei et al., 2021).

The quantum computing theory and cooperative evolution of multiple operators are applied to generate effective detectors to improve the AI. We design a detection method based on a normalized mutual information and cooperative evolution of multiple operators based on adaptive parallel quantum artificial immune. There are three primary contributions in this paper are:

1. In the stage of feature selection, to improve the detection speed, the NMAIFS method is applied to select the optimum detection features from a given feature set and achieve effective dimensionality reduction of high dimensional features.
2. In the stage of generation of classifier, with a strong learning generalization of AI algorithm, vaccination strategy, and the acceleration of the quantum computing, the artificial immune algorithm is improved, we give an MOP-AQAI algorithm. Furthermore, synthesizing the

two improved algorithms, we give an anomaly detection method based on normalized mutual information and the cooperative evolution of multiple operators based on adaptive parallel quantum artificial immune.

3. Finally, with the benchmark of KDD99 and UNSW-NB15, experiments are presented to verify the reduction performances of NMAIFS; the anomaly and classification performances of MOP-AQAI, especially for small sample sets. The anomaly data results from real network traffic indicate that this method has a higher detection accuracy and a lower false positive rate than existing anomaly detection algorithms.

NMAIFS

In the reduction stage, NMAIFS is applied to omit the redundant attributes for high detection speed. First, all the training logs are normalized and transformed into antibodies. Next, the NMAIFS algorithm selects the optimal features of antibodies (Fatemidokht et al., 2021; Mishra et al., 2021).

Entropy is a mathematical measure for the uncertainty of random variables and describes a measure of the random variables of the average amount of information. For example, antibodies attribute information entropy (Feng et al., 2021; Xi et al., 2021; Zhi-jian et al., 2019), and antibodies combination entropy are given in the following paragraphs.

Antigen: $ag \in Ag$, $Ag \subset D$, $D = \{0,1\}^l$, ($l \in N, l > 0$),*Ag* denote antigen set, *D* is the binary character string with length *l*, and the value of antigen *ag* represents the behavior characteristics of the binary string.

Antibody: $ab \in Ab$, $Ab\left\{\left\langle d, s, age, count\right\rangle\right\}$, $d \in D$, $s \in \left\{00, 01, 10\right\}$, $age \in N$. *Ab* is antibody set, *s* is the state of antibody, whose value is 00,01 or 10; *age* denotes the age of antibody, *count* is the matching number of antibody and antigen; *N* is a positive integer collection.

$Ab = AbI \cup AbT \cup AbM$, AbI denotes the collection of immature antibodies, $AbI = \left\{Ib \middle| Ib \in Ab, Ib.s = 00\right\}$; AbT denotes the collection of mature antibodies, $AbT = \left\{Tb \middle| Tb \in Ab, Tb.s = 01\right\}$; AbM denotes the collection of memory antibodies,and $AbM = \left\{Mb \middle| Mb \in Ab, Mb.s = 10\right\}$.

Self denotes the collection of normal behaviors. Nonself denotes the collection of abnormal behaviors, $Self \cap Nonself = \phi$.

The function fit (*ab*, *ag*) is on behalf of the affinity between antigen *ag* and antibody *ab*.

We calculate the affinity between antigen and antibody with the Euclidean formula, shown in formula (1).

$$fit\left(ab, ag\right) = \sqrt{\sum_{i=1}^{L} (agi - abi)^2} \tag{1}$$

ag_i is the *i*th character of antigen, ab_i is the *i*th character of antibody.

The antibody information entropy *H (X)*: We denote *X*, *Y* as the two discrete random variables, the joint probability mass function *p* (*x*, *y*) and marginal probability *p*(x), *p*(y), and the antibody information entropy of *X* is defined as (Feng & Dou, 2021):

$$H\left(X\right) = -\sum_{i=1}^{m} p\left(x_i\right) \log_2 p\left(x_i\right) \tag{2}$$

In formula (2), x_i is the possible value of the antibody attribute.

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

$$H\left(X,Y\right) = -\sum_{i=1}^{m}\sum_{j=1}^{n} p(x_i, y_j) \log_2 p(x_i, y_j) \tag{3}$$

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*, H (*X*, *Y*) has the minimum value, when *X* and *Y* are independent of each other, H (*X*, *Y*) has the maximum value.

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

$$MI\left(X;Y\right) = \sum_{i=1}^{m}\sum_{j=1}^{n} p(x_i, y_j) \log_2 \frac{p(x_i, y_j)}{p(x_i)p(y_j)} \tag{4}$$

The number of independent variables in Eq. (4) is dimensionless; therefore, the integral value is irrelevant to the selected coordinate (feature space transform). This feature remains in differentiable or reversible transformation. The relations among mutual information, entropy, and combination entropy are:

$$MI\left(X;Y\right) = H\left(X\right) + H\left(Y\right) - H\left(X,Y\right) \tag{5}$$

The minimum value of entropy is MI(X;Y), the range of the value is: $0 \le MI\left(X;Y\right) \le \min\{H\left(X\right), H\left(Y\right)\}$. NMAIFS algorithm adopts symmetrical uncertainty to measure the correlation between features and categories, features and features. The symmetric uncertainty SU (*X*, *Y*) between two random antibody variables *X* and *Y* is defined as:

$$SU\left(X,Y\right) = 2\left[\frac{H\left(X\right) - H(X \mid Y))}{H\left(X\right) + H\left(Y\right)}\right] = \frac{MI\left(X;Y\right))}{\left[H\left(X\right) + H\left(Y\right)\right] / 2} \tag{6}$$

For $\min\{H\left(X\right), H\left(Y\right)\} \le \left[H(X) + H(Y)\right] / 2$, we denote the definition of antibody entropy as $H\left(X\right) \le \log_2 \sum_{i=1}^{m} p(x_i) \frac{1}{p(x_i)} = \log_2 m$ with Jensen inequality. So we can get $0 \le H\left(X\right) \le \log_2 m$, where m and n are the numbers of possible values of discrete random variables X and Y, respectively. We define the normalized mutual information as:

$$\mathrm{NMAI}\left(\mathrm{X};\mathrm{Y}\right)=\frac{MI(X;Y)}{\min\{\log_2 m \quad log_2 n\}} \tag{7}$$

By formula (6), the feature values of anonymous logs are normalized between [0, 1] before executing the feature selection algorithm. The standardized process of the fast normalized mutual information feature selection is shown in Table 1.

Table 1. The standardized NMAIFS process

Standardized process of the NMAIFS
Initialization: Set $F=\left\{f_i, i=1,\ldots,N\right\}$,which contains N features. Make $S=\left\{\varnothing\right\}$
Calculate the antibody mutual information between features and classes: For each $f_i \in F$, calculate $NMAI\left(f_i;C\right)$
Select the first feature: find $\hat{f}_i=\max_{i=1,\ldots,N}\left\{MI\left(f_i;C\right)\right\}$. 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 antibody mutual information among features: for each pair $\left(f_i, f_s\right)$,calculate $NMAI\left(f_i;f_j\right)$,where $f_i \in F$ and $f_s \in S$ until the completion of the iteration
Turn to the next feature: choose antibody feature $f_i \in F$ making Eq.(3)reach the maximum. Set $F \leftarrow F \setminus \left\{f_i\right\}, S \leftarrow \left\{f_i\right\}$
Output set *S* with *K* antibody selected features

MOP-AQAI

To improve the adaptability of IDS, AI is improved in this paper. First, the cooperative evolution of multiple operators is used to accelerate the convergence process of IDS; second, the vaccination strategy is to increase the fitness of antibodies; last, we used the quantum rotation angle step length generate the best antibodies for high detection performance, especially for small samples sets.

THE COOPERATIVE EVOLUTION MECHANISM OF MULTIPLE OPERATORS

The cooperative evolution mechanism of multiple operators includes antibody similarity evaluation operator, antibody fitness evaluation operator and population variation adjustment operator to determine the current populationism of multiple operators includes anti (Feng & Dou, 2021; Xi et al., 2021).

Antibody similarity evaluation operator $x_{\mathrm{sim}}: x_{\mathrm{sim}}$ to calculate the antibody differences in current population, defined in formula (8):

$$x_{\text{sim}} = \begin{cases} \dfrac{d_{\text{avg}} - \mathrm{d}_{\text{min}}}{d_{\text{max}} - \mathrm{d}_{\text{min}}}, d_{\text{max}} \neq d_{\text{min}} \\ 0, d_{\text{max}} = \mathrm{d}_{\text{min}} \end{cases} \tag{8}$$

In formula (8), d_{max} is the maximal Euclidean distance of the antibody with the optimal ones in the current population, d_{min} is the minimal Euclidean distance of the antibody with the optimal ones in the current population, d_{avg} is the average Euclidean distance of all antibodies with the optimal ones in the current population. The larger the antibody similarity evaluation operator is, the more different the antibodies are in the current population, then we can use a larger mutation probability to increase population diversity. On the contrary, the smaller the x_{sim} is, we can use smaller mutation probabilities to maintain the population stability.

Antibody fitness evaluation operator y'_{fit}**:** y'_{fit} is adopted to evaluate the *i*th antibody fitness in the current population, which is defined as formula(9):

$$\mathrm{y}'_{\text{fit}} = \begin{cases} \dfrac{f_{\text{max}} - f_i}{f_{\text{max}} - f_{\text{min}}}, f_{\text{max}} \neq f_{\text{min}} \\ 0, f_{\text{max}} = f_{\text{min}} \end{cases} \tag{9}$$

In the current population, f_{max} is the maximum value of fitness and f_{min} is the minimum value of fitness, f_i is the *i*th antibody fitness. If the value of y'_{fit} is high, the fitness of the *i*th antibody is closer to the worst individual in the current population. Therefore, we should select a higher mutation probability when the antibodies mutate.

Population variation adjustment operator $F_{\text{acc}}\left(n\right)$**:** $F_{\text{acc}}\left(n\right)$ is the function of the current evolution algebra *n*, this operator is used to increase antibody mutation probability against premature convergence gradually. The population variation adjustment operator is defined as shown in formula (10):

$$F_{\text{acc}}\left(n\right) = \begin{cases} F_{\text{acc}}\left(n-1\right) + C \times \dfrac{s-n}{s}, f_{\text{max}}(n = f_{\text{max}}(n-T \ \wedge n > T \\ F_{\text{acc}}\left(n-1\right), f_{\text{max}}(n \neq f_{\text{max}}(n-T \ \wedge n > T \\ 0, n \leq T \end{cases} \tag{10}$$

The current evolution algebra is *n*, *s* is the maximum evolutionary algebra, and constant *T* is the iteration number if antibodies do not change continuously in the current population. Constant *C* (0 < *C*) is the adjusting parameter, and $f_{\text{max}}(n$ is the optimal fitness value in the *n*th generation populations. When the optimal fitness value of the populations did not change for consecutive *T*

generations, and the evolution algebra does not reach the maximum, the mutation probability increases to adjust the variation probability, which is calculated by *s*, *n,* and *C*.

Antibody similarity evaluation operator x_{sim}, antibody fitness evaluation operator y'_{fit} and population variation adjustment operator $F_{\text{acc}}\left(n\right)$ are used to calculate the current mutation probability of the antibodies in the population; the mutation probability is calculated with formula (11).

$$p'_{n} = \begin{cases} p_0 \times y_{fit}^{n} \times x_{\text{sim}} + F_{\text{acc}}\left(n\right), f_{\max} \neq f_{\min} \\ 0, f_{\max} = f_{\min} \end{cases} \tag{11}$$

In formula (11), p'_{n} is the *i*th antibody(1 bability is calculated in*n*th generation population, p_0 is the initial value of mutation probability.

VACCINATION

Vaccination: $va \in S$, $S = \{0,1,*\}^{l}$, ($l \in N$, $l > 0$),where *va* is defined as a string with 0, 1, and *, whose length is l, va_k denotes the *k*th code of va. The antibody population $Ab = \left\{ab1, ab2, \cdots abn\right\}$, ab_i^k is the *k*th gene code of the *i*th antibody.

Vaccine extraction operator: $ab1, ab2, \cdots abs$ are the optimal antibodies with high fitness in antibody population $Ab = \left\{ab1, ab2, \cdots abn\right\}$, ab_i^k is the value of gene, vaccine extraction operator is defined with formula (12):

$$va^k = \begin{cases} 1, & \left(1/s\right)\sum_{i=1}^{s} ab_i^k) > \alpha \\ 0, & \left(1/s\right)\sum_{i=1}^{s} ab_i^k) < \beta \\ *, & other \end{cases} \tag{12}$$

The values of parameters α, β are $\alpha \geq 0.8$, $\beta \leq 0.2$.

Vaccination: Vaccination is the process that optimal gene of vaccination takes place with the alleleocess that optimal gene of vaccination with formulaallycan use eval.

Vaccination operator: Let us suppose that *a* is an antibody, *va* is a vaccination, and the vaccination operator is $\hat{a} = a\Theta va$, $\hat{a}$ is the antibody after vaccination. The vaccination operator is defined in the following formula:

$$\hat{a}^k = a^k \Theta va^k = \begin{cases} va^k, va^k = 0\text{or}1 \\ a^k \quad va^k = * \end{cases} \tag{13}$$

EVALUATION METHOD OF ANTIBODY AFTER VACCINATION:

Let us set *va* as the vaccine of antibody population *A*, the individual is a_i, and the evaluation method of antibody after vaccination is shown in formula (14):

$$E\left(va\right) = E'\left(va\right) + \sum\nolimits_{i=1}^{n}\left(fit\left(\hat{a}i, ag\right) - fit\left(ai, ag\right)\right) \tag{14}$$

In formula (14), $E'\left(va\right)$ is the effect of antibody before vaccination, fit is the affinity function between antibody and antigen, $\hat{a}i$ is the individual after vaccination of antibody a_i.

THE ADAPTIVE QUANTUM ARTIFICIAL IMMUNE ALGORITHM WITH COOPERATIVE EVOLUTION OF MULTIPLE OPERATORS

Adaptive adjustment of rotation angle and the cooperative evolution of multiple operators are adopted to design the MOP-AQGA algorithm to improve the adaptability and detection performance. Adaptive adjustment mechanisms of the rotation angle dynamically adjust the rotation angle step length according to the individual step length nd detection performance, a ato Table 2.

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

x_i^j	b_i	$f(X^j \geq f(X_{best}^t$	”, $_i^j$	$S\left(\pm_i^j, {}^2{}_i^j\right)$			
				$\pm_i^j {}^2{}_i^j > 0$	$\pm_i^j {}^2{}_i^j < 0$	$\pm_i^j = 0$	${}^2{}_i^j = 0$
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
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, $f(X$ is the fitness value of antibody *x*; x_i^j is the *i*th gene value of the *j*th antibody; b_i is the *i*th value of the best antibody in the current population; $S\left(\alpha_i^j, \beta_i^j\right)$ denotes the rotation

direction of rotation angle in polar coordinates; θ^j is the rotation angle of step length of the *j*th antibody. θ^j is defined in formula (15).

$$\theta^j = \begin{cases} \dfrac{f_j - f_{\min}}{f_{\max} - f_{\min}}\left(K_2 - K_1\right) + K_1, f_{\max} \neq f_{\min} \\ K_1, f_{\max} = f_{\min} \end{cases} \tag{15}$$

The relation between θ^j and the fitness of the current antibody is a linear function; with the increasing antibody fitness, we allocate the greater rotation angle step. On the other hand, with decreasing antibody fitness, we allocate the smaller rotation angle step.

In the current antibody population, the *i*th rotation angle step $\Delta\theta_i^j$ of the *j*th antibody is calculated with the formula (16).

$$\Delta\theta_i^j = \theta^j \times S\left(\alpha_i^j, \beta_i^j\right) \tag{16}$$

The MOP-AQAI is shown in Table 3.

Table 3. The MOP-AQAI algorithm

MOP-AQAI algorithm
Initialize the memory antibodies population:($t \leftarrow 0$);
Create new population $Q(t)$;
Observe $Q(t)$ and obtain the observed state $P(t)$;
Calculate and evaluate the fitness of antibodies of $P(t)$;
Select the optimal individuals into $ME(t)$;
Save the antibodies which have higher fitness into memory antibody collection $ME(t)$;
while($n \leq s$)
do
$t \leftarrow t+1$;
Obtain the new observed state $P(t)$ according to memory antibodies;
Calculate and evaluate the fitness of antibodies of $P(t)$;
Select the optimal individuals into $ME(t)$;
Calculate the angle of rotation step with formula (12);
Save those antibodies into immature antibodies set $IM(t)$;
Calculate collective mutation operator with formula (7) - (9);
Calculate individuals' variation probability with formula (10);
Individuals mutate with their mutation probability;
Calculate and evaluate the fitness of antibodies of $IM(t)$;
Give these antibodies according to their fitness in descending order;
Select the antibodies with the highest fitness as vaccination;
Select the antibodies which own lower fitness, and vaccine them with formula (13);
Evaluate the antibodies after vaccination with formula (14);
If the fitness of the antibody is lower than it before vaccination, the antibody after vaccination is deleted, otherwise save it;
Judge if the period of immature antibody exceeds the threshold value *T1*, if exceeding, then kill the immature antibody,
otherwise, use it to match the self- antigens, save the immature antibodies into mature antibodies set $MA(t)$;
If the mature antibody match anyone, then kill it;
Else turn it to be a mature antibody;
Judge if immature antibody exceeds the threshold value *T2* in its period, if exceeding, then kill the immature antibody, otherwise, activate it;
Calculate the fitness of antibody and obtain new $ME(t)$.

THE DETECTION MODE OF NMAIFS MOP-AQAI

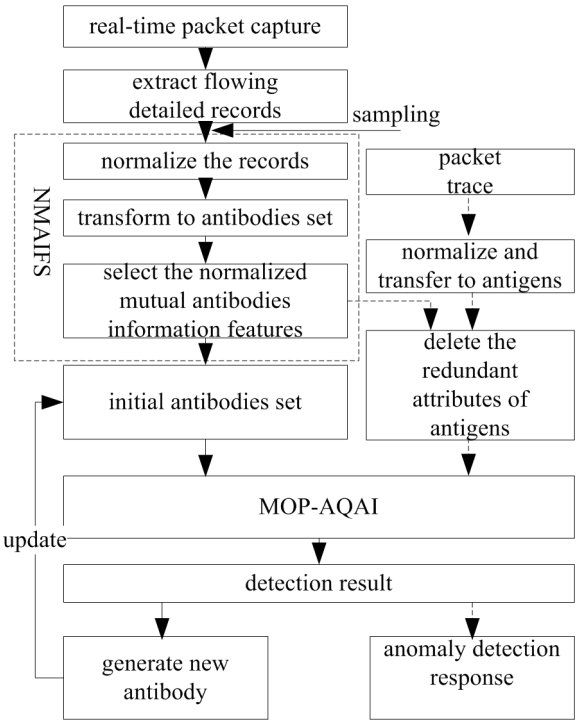
The detection of NMAIFS MOP-AQAI has three stages:

In the preprocessing stage, the real-time packet capture such as *snort* obtains the flowing logs and extracts them as flowing detailed records.

In the training stage, NMAIFS and MOP-AQAI are two core sub-modules. NMAIFS is applied to reduce the redundant attributes and obtain the initial memory antibodies set with optimal features. MOP-AQAI is used to dynamically generate excellent antibodies with the cooperative evolution of multiple operators, vaccination strategy, and the quantum theory.

In the detection stage, the packet trace records are normalized and transferred as antigens; simultaneously, the redundant features are omitted and detected by the memory antibodies. If the detection results are anomalous, the IDS gives an alert, transforming the antigens into antibodies. Meanwhile, the antibodies are put into initial memory antibodies set for the next evolution. The NMAIFS MOP-AQAI intrusion detection model is shown in Figure 1.

Figure 1. NMAIFS -MOQAI detection module



The NMAIFS MOP-AQAI is shown in Table 4.

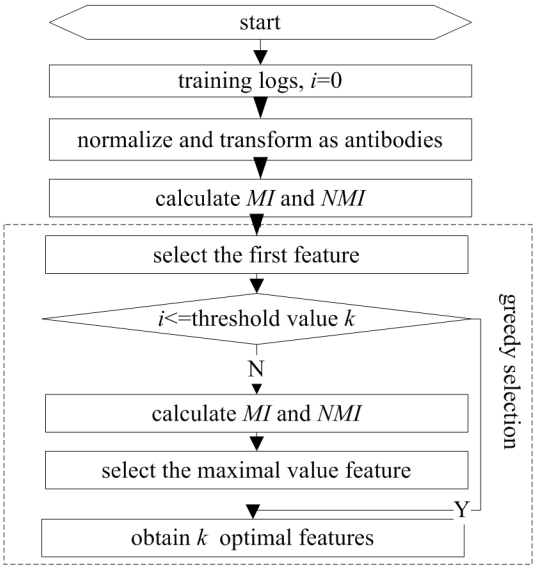
Table 4. The NMAIFS MOP-AQAI

NMAIFS MOP-AQAI
The offline cache data are the original input origins. NMAIFS sub-module extracts the traffic feature data and creates the initial antibody set; Execute the NMAIFS algorithm, reduce high-dimensional feature data and get the effective data set with the method that we calculate the normalized antibody mutual information between various features and different behavior categories in the sample data to get the best combination feature set; The optimal antibodies collection with the best combination of characteristics is solved by a quantum immune algorithm, adaptive learning to get memory antibody collection; After sampling the received data packet, extract the flow detailed record with the pre-definition features in actual applications; Select d-dimensional important features from the flow detailed record through the NMAIFS module. Then, compare it with the detection model in the training phase and output the accurate detection results; The antigen after detection are transferred antibody, and put MOP-AQAI, update the MOP-AQAI module.

The two core sub-modules, NMAIFS and MOP-AQAI, are used in the training stage of the detection to generate excellent antibodies.

The sampling records are fed into the NMAIFS sub-module and normalized by the relationship of different attributes. Then, antibodies' optimal features are selected with the greedy choices algorithm. The flowchart for the NMAIFS sub-module is shown in Figure 2.

Figure 2. The flowchart of NMAIFS



Then all the training antibodies are inserted into the MOP-AQAI sub-module for dynamic evolution. The antibodies, after reduction, are placed into an initial antibody set. The antibodies possessing higher fitness are initial memory antibodies; the memory antibodies experience executed crossover, mutation, and vaccination to develop immature antibodies. Some immature antibodies are transferred into mature antibodies, and part mature antibodies in their life cycles are activated to become memory antibodies. Finally, memory antibodies are adopted to detect the antigens; the

detection results are based on the new antibodies in a fresh evolutionary round. The flowchart of the MOP-AQAI sub-module is shown in Figure 3.

SIMULATIONS AND ANALYSIS

To verify the effectiveness of the proposed method, NMAIFS MOP-AQAI, in the experiments, we take the intrusion scenario correlation benchmarks KDD99 (KDD, 2010) and UNSW-NB15 to test the algorithm proposed in this paper. Therefore, we use some common performance indicators as parameters to detect antigens and present comparison analysis, respectively, according to KDD99 and UNSW-NB15 data sets (Moustafa, 2017). KDD99 and UNSW-NB15 both include small attack samples, with which we research the attack classification performances of the NMAIFS MOP-AQAI algorithm for small samples. With the UNSW-NB15 data set, the purpose is to research the classification performances of novel attacks.

DATA SET AND SIMULATION ENVIRONMENT SETUP

KDD99 Dataset

The intrusion scenario correlation benchmark KDD99 was used in the experiments to verify the effectiveness of the proposed NMAIFS MOP-AQAI detection method.

Lincoln laboratory provided KDD99 for experiment simulations of intrusion detection; the training sample set KDD99 includes 494,021 records. The test sample set includes 311,029 records. The data set contains Normal, Dos, the Probe, U2R, and R2L five species of samples. When we execute the experimental simulation, we adopted all the training samples to train and test the NMAIFS MOP-AQAI algorithm. We show the sample distribution of the data set in Table 5.

Table 5. *Sample distribution of the data set KDD99*

No	Type	Training set	Testing set
1	Normal	97,278	60,593
2	Dos	391,458	229,853
3	Probe	4,107	4,166
4	U2R	52	228
5	R2L	1,126	16,189
Sum		494,021	311,029

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 $P_0 = 0.8$, and the variation operator adjustments constant C is 0.08.

The anomaly detection algorithm NMAIFS MOP-AQAI was implemented with C as the simulation environment. The CPU is an Intel Pentium of 4, 3.20 GHz, memory is 16 GB, and the operating system was Microsoft Windows 2016.

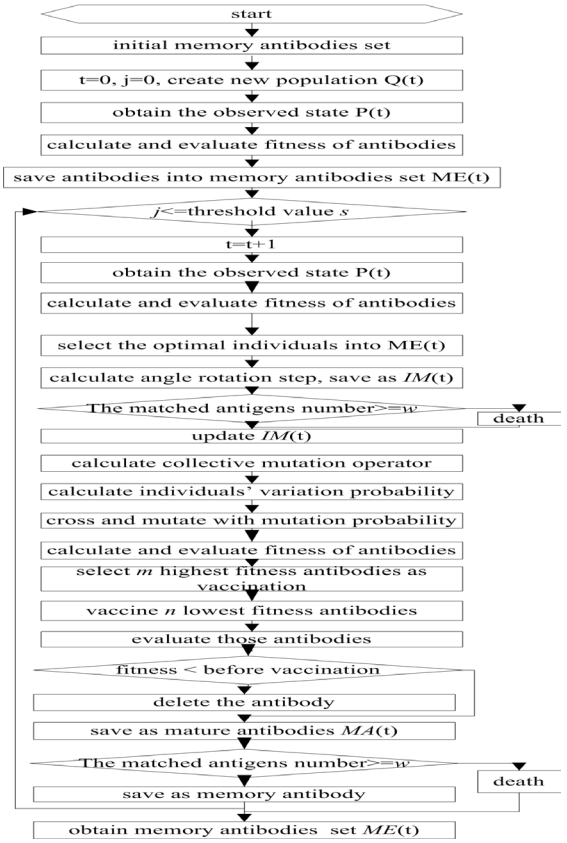
UNSW-NB15 DATASET

The UNSW-NB15 datasets include normal samples and ten types of attacks and contain 257,673 records (He et al., 2017). The sample distribution of the UNSW-NB15 data set is shown in Table 6.

Table 6. Sample distribution of the data set UNSW-NB15

No	Type	Training set	Testing set
1	Normal	56,000	37,000
2	Generic	40,000	18,871
3	Exploits	33,393	11,132
4	Fuzzers	18,184	6,062
5	Dos	12,264	4,089
6	Reconnaissance	10,491	3,496
7	Analysis	2,000	677
8	Backdoor	1,746	583
9	Shellcode	1,133	378
10	Worms	130	44
Sum		175,341	82,332

Figure 3. The flowchart of MOP-AQAI



THE DATA PRE-PROCESSING AND EVALUATION STANDARD

There are four steps in the testing experiment: normalization, reduction, training, and testing. We may easily ignore the effectiveness of some features owing small values, since there are substantial differences among features in the data set. Therefore, the feature data must be normalized before NMAIFS selects optimal features. For example, let us suppose that the antibodies set contains *n* records, $f_j[i]$ represents the *i*th feature of the *j*th record. The mean and standard deviation are calculated with the following equations:

$$\overline{f}_j\left[i\right] = \frac{1}{n}\sum\nolimits_{j=1}^{n} f_j\left[i\right] \tag{17}$$

$$s_j\left[i\right] = \sqrt{\frac{1}{n-1}\sum\nolimits_{j=1}^{n}\left(f_j\left[i\right] - \overline{f}_j\left[i\right]\right)^2} \tag{18}$$

Where $\overline{f}_j\left[i\right]$ and $s_j\left[i\right]$ denote the mean and standard deviation of the *i*th feature, respectively. Then, all features are normalized with formulae (19):

$$\hat{f}_j\left[i\right] = \frac{f_j\left[i\right] - \overline{f}_j\left[i\right]}{s_j\left[i\right]} \tag{19}$$

We show the evaluation criteria of the test results in the following:

$$DR = \frac{TP}{TP + FN} \tag{20}$$

$$FAR = \frac{FP}{TN + FP} \tag{21}$$

$$Pre = \frac{TP}{TP + FP} \tag{22}$$

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{23}$$

TP denotes that the samples which belong to the category *C* are correctly recognized as the category *C*; *TN* denotes the samples which do not belong to the category *C* are correctly recognized

as other categories; *FP* denotes the samples which do not belong to the category *C* are wrongly recognized as the category *C*; and *FN* denotes the samples which belong to the category are wrongly recognized as not belonging to the category.

THE BEST FEATURE SET SELECTION WITH NMAIFS

In the NMAIFS sub-module, the optimal features are selected by calculating the normalized mutual antibodies information of the statistical characteristic and features with marked category. We select the normalized mutual information values in descending order according to their characteristics, which are the preparations for MOP-AQGA based on attributes and significance of features. The C4.5 decision tree algorithm was adopted to test the antigens with a varying number of characteristics. The mRMR (Fatemeh et al., 2011) and NMIFS (Huang et al., 2017) algorithms were compared with NMAIFS.

There are two principles of the optimal features subset. First, when the difference between DR and FAR of NMAIFS, NMIFS, and mRMR is greatest. Second, when the average detection rate is higher than 90%, those features are selected.

THE NMAIFS FOR KDD99 DATASETS

There are 41 conditional attributes in KDD99 sets. If all the features are used in intrusion detection, much more time will be consumed to deal with the redundant attributes. With NMAIFS, the optimal features are selected, which are shown in Table 7.

Table 7. The optimal attributes set of KDD99

Type	NMAIFS	NMIFS	mRMR
Normal	30, 3, 13, 33, 14, 15, 10, 17, 31, 36	30, 3, 13, 33, 14, 15, 10, 17, 31, 36	13, 15, 3, 31, 10, 14, 33, 17, 18, 21, 30, 36
Dos	3, 2, 4, 5, 26, 30, 32, 25, 38, 39, 37	4, 3, 2, 5, 26, 30, 32, 25, 38, 37, 39	25, 38, 2, 5, 26, 30, 32, 35, 3, 4, 37, 39, 40
Probe	16, 7, 10, 12, 8, 23, 24, 6, 36, 11, 39	16, 7, 23, 10, 12, 8, 11, 24, 6, 36, 39	36, 7, 39, 8, 11, 12, 16, 24, 10, 29, 31, 23, 34, 6
U2R	3, 14, 10, 13, 33, 15, 30, 17, 31, 36	3, 14, 33, 10, 15, 30, 17, 13, 31, 36	14, 8, 10, 13, 17, 18, 3, 5, 30, 31, 15, 33, 36
R2L	31, 3, 6, 8, 10, 21, 7, 27, 9,38, 1	31, 3, 10, 8, 21, 7, 27, 6, 9,38, 1	6, 3, 7,, 10, 12, 1,18, 21, 27, 8, 9 31, 38

We conclude from Table 7 that for the five types of data sets, the subsets of NMAIFS and NMIFS are similar; the numbers for are generally lower than those of mRMR.

THE NMAIFS FOR UNSW-NB15

There are 43 conditional attributes in UNSW-NB15 sets. With the same process and principles of KDD99, NMAIFS is adopted to obtain the best features collection. We show the optimal features subset in Table 8.

Table 8. The optimal attributes set of UNSW-NB15

Type	NMAIFS	NMIFS	mRMR
Normal	19, 36, 10, 20, 11, 21, 6, 34	19, 10, 36, 11, 20, 21, 6, 34	6, 10, 11, 15, 17, 18, 19, 20, 21, 24, 26, 34, 36
Generic	10, 18, 15, 6, 11, 13, 16, 9, 17, 12, 20	15, 10, 18, 6, 11, 16, 13, 9, 17, 12, 20	3, 5, 6, 9, 10, 11, 12, 13, 15, 16, 17, 18, 20, 28, 29
Exploits	6, 5, 10, 19, 37, 11, 41, 42, 36	10, 6, 5, 37, 11, 41, 19, 42, 36	2, 5, 6, 10, 11, 13, 15, 16, 19, 22, 36, 37, 41, 42
Fuzzers	40, 41, 6, 15, 14, 16, 36, 11, 37, 39, 42	40, 6, 41, 15, 16, 36,, 14 11, 37, 39, 42	4, 6, 11, 14, 15, 16, 17, 18, 26, 36, 37, 39, 40, 41, 42
Dos	15, 42, 11, 6, 16, 36, 39, 37, 40	15, 42, 6, 16, 11, 36, 37, 39, 40	6, 8, 9, 11, 15, 16, 18, 20, 25, 27, 36, 37, 39, 40, 42
Reconnaissance	41, 42, 37, 9, 14, 16, 10, 17, 28	41, 37, 42, 14, 16, 9, 10, 17, 28	4, 5, 8, 9, 10, 14, 16, 17, 20, 21, 25, 28, 30, 37, 41, 42
Analysis	6, 11, 34, 35, 12, 10, 15, 13, 16, 14, 37	6, 34, 11, 12, 35, 10, 13, 15, 16, 14, 37	6, 7, 9, 10, 11, 12, 13, 14, 15, 16, 24, 26, 27, 34, 35, 37
Backdoor	41, 10, 6, 14, 11, 16, 15, 37, 42	10, 41, 6, 11, 16, 14, 37, 15, 42	2, 3, 5, 6, 10, 14, 15, 16, 25, 26, 28, 30, 37, 41, 42
Shellcode	18, 23, 9, 12, 10, 14, 16, 15, 13, 17, 6	18, 9, 10, 23, 14, 16, 12, 15, 13, 6, 17	6, 8, 9, 10, 12, 13, 14, 15, 16, 17, 18, 23, 26, 39, 40
Worms	10, 5, 9, 11, 14, 13, 17, 37, 41, 23	5, 9, 10, 11, 13, 17, 14, 37, 41, 23	1, 5, 8, 9, 10, 11, 13, 14, 17, 19, 21, 23, 26, 27, 32, 37, 41

For the ten types of samples, the optimal features subsets of NMAIFS are similar to NMIFS, the numbers of NMAIFS are almost less than those of mRMR.

In conclusion, for either KDD99, or UNSW-NB15, with a C4.5 decision tree algorithm, NMAIFS has similar feature subsets to NMIFS; NMAIFS gets a more concise feature set than mRMR. So with fewer features, the intrusion detection speed can be improved.

THE ANOMALY DETECTION PERFORMANCE RESULTS COMPARED WITH RELATED WORKS

KDD99 and UNSW-NB15 are respectively used to demonstrate the effectiveness of the NMAIFS MOP-AQAI algorithm. We executed respectively attacks classification and anomaly detection with those two sets.

THE ANOMALY DETECTION PERFORMANCE RESULTS COMPARISON OF KDD99

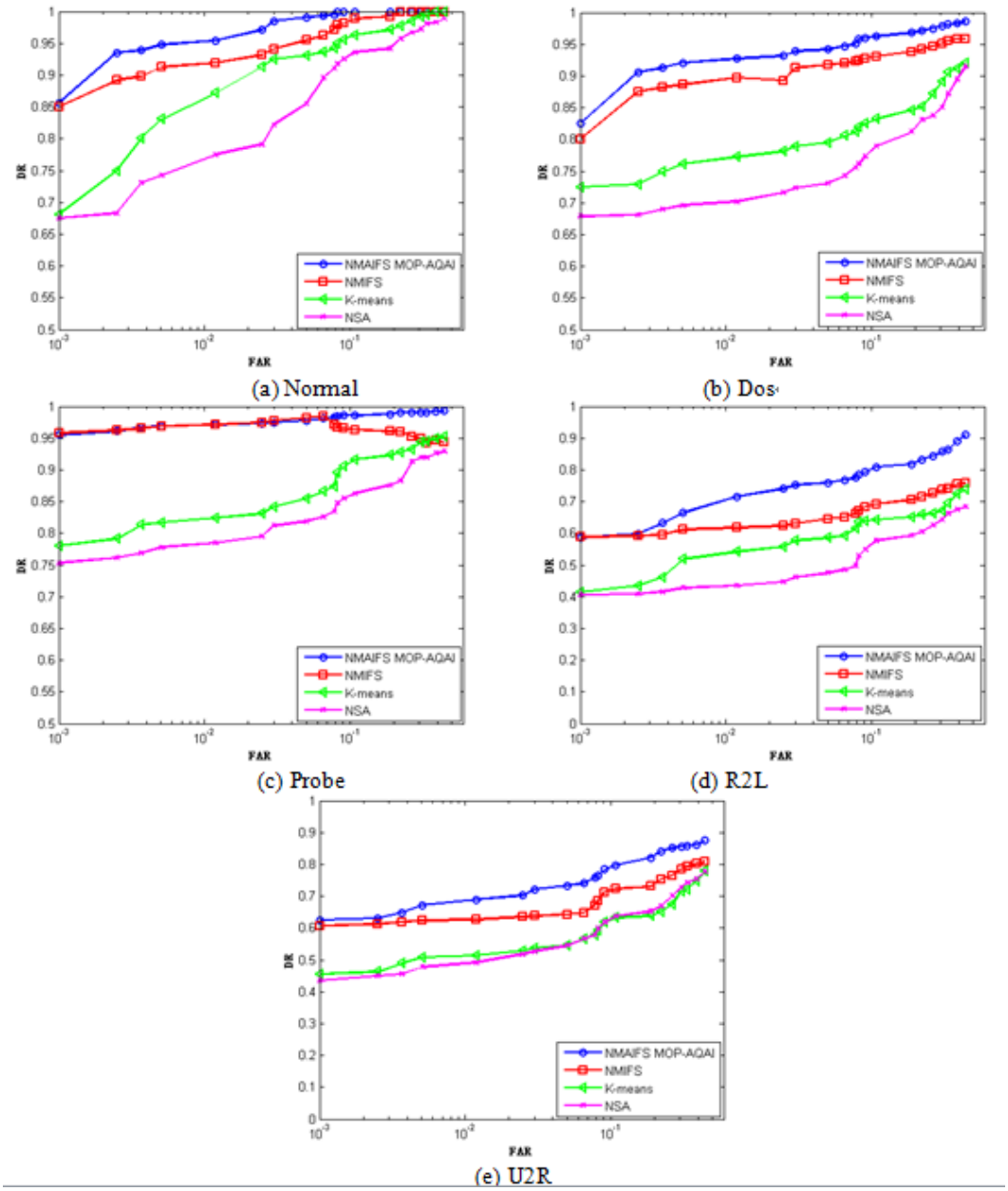
With KDD99, the Receiver Operating Characteristic (ROC), the classification of intrusions, and anomaly detection are applied to verify the performances of the NMAIFS MOP-AQAI algorithm.

ROC

The NSA (Naila et al., 2020), K-means (Alyaseen et al., 2017), NMIFS (Huang et al., 2017), and NMAIFS MOP-AQAI algorithm proposed in this paper were used to train and test with KDD99.

Normal, Dos, the Probe, U2R, and R2L are adopted to obtain the ROC curves. With different threshold values, we get the DRs and FARs, respectively, and the ROC on five sample types shown in Figure 4.

Figure 4. The ROC curves of different types of data sets with various algorithms for KDD99



In Figure 4, the results show how with the increase of DR, the FAR in decreasing; compared with the existing common detection methods, both in the normal data set and in the abnormal data

sets (DOS, the Probe, R2Land U2R), we can get a lower FAR a higher DR. Meanwhile, NMAIFS MOP-AQAI can guarantee at a relatively low rate of false positives and higher detection rate for all the samples. When considering the small sample sets U2R and R2L, we can get higher DR and lower FAR, and we have a better balance between DR and FAR.

THE CLASSIFICATION OF INTRUSIONS

Five different categories of datasets are applied to simulate the algorithm's classification ability to verify its effectiveness. Each dataset runs ten times; the average values are taken as the test results. Finally, the MDPCA–DBN (Yang et al., 2019), NMIFS + QWNN (Huang et al., 2017), and the NMAIFS MOP-AQAI algorithm proposed in this paper are compared. The results are shown in Table 9.

Table 9. Detection performance comparisons of classification on KDD99

Type	Methods	DR	Acc	FAR	Pre
Normal	**NMAIFS MOP-AQAI**	**98.90**	**99.41**	**0.45**	**98.12**
	MDPCA-DBN	71.42	N/A	N/A	97.38
	NMIFS+QWNN	99.92	99.84	0.08	N/A
DOS	**NMAIFS MOP-AQAI**	**99.49**	**99.48**	**0.49**	**99.82**
	MDPCA-DBN	96.34	N/A	N/A	81.09
	NMIFS+QWNN	98.83	98.88	0.10	N/A
Probe	**NMAIFS MOP-AQAI**	**96.35**	**99.82**	**0.12**	**91.10**
	MDPCA-DBN	85.85	N/A	N/A	73.94
	NMIFS+QWNN	84.85	88.57	0.56	N/A
U2R	**NMAIFS MOP-AQAI**	**85.70**	**99.81**	**0.01**	**37.40**
	MDPCA-DBN	11.82	N/A	N/A	6.50
	NMIFS+QWNN	84.85	88.57	0.56	N/A
R2L	**NMAIFS MOP-AQAI**	**96.83**	**99.75**	**0.01**	**98.26**
	MDPCA-DBN	57.30	N/A	N/A	17.25
	NMIFS+QWNN	79.00	88.79	88.79	N/A

As it can be seen from Table 9, the detection results of NMAIFS MOP-AQAI are obviously better than MDPCA – DBN, which has higher DR, Acc, Pre, and lower FAR. Particularly for small samples sets U2R and R2L, the entire performances are more superior.

Compared with NMIFS + QWNN, according to the large sample dataset, the NMAIFS MOP-AQAI's DR is a little lower than NMIFS + QWNN, but the DR reaches over 96%. Furthermore, because of the small data samples, such as U2R and R2L, detection performances are higher than NMIFS + QWNN. So NMAIFS MOP-AQAI's classification ability is better than the other two algorithms on the five attack types, particularly for the U2R and R2L small sample sets.

THE ANOMALY DETECTION

The third comparison concerns the anomaly detection of NMAIFS MOP-AQAI, which comprises longitudinal comparisons, such as immune algorithms, and horizontal comparisons with other machine learning algorithms. All the attacks were abnormal samples for simulations; we adopted the testing sets for detection, each dataset group runs ten times, and we get the average values. Different algorithms' anomaly detection performance comparisons are shown in Table 10.

Table 10. The comparisons for KDD99 (the N/A reflects an unknown value)

Method	DR	Acc	FAR	Pre
AI+PA (Chou et al., 2020)	98.72	95.90	1.56	89.70
MDPCA-DBN(Yang et al., 2019)	61.57	66.18	13.06	95.51
WIDMoDS (Feng et al., 2021)	N/A	98.8	N/A	N/A
SR+DL (Wu et al., 2020)	N/A	94.2	N/A	N/A
MS-NIDAM (Xi et al., 2021)	about 92.00	N/A	about 5.00	N/A
NMAIFS MOP-AQAI	**99.18**	**99.68**	**0.25**	84.94

The results in table 10 show that the NMAIFS MOP-AQAI algorithm has a higher DR, and Acc than the other five algorithms and the FAR is lower than the other five algorithms. At the same time, the Pre of NMAIFS MOP-AQAI is lower than MDPCA-DBN but higher than AI+PI.

THE COMPARISONS FOR UNSW-NB15

With UNSW-NB15, the ROC and anomaly detection were adopted to verify the performances of NMAIFS MOP-AQAI algorithm.

ROC

With the same process and principles of KDD99, the NSA (Naila et al., 2020), K-means (Alyaseen et al., 2017), NMIFS (Huang et al., 2017), and the NMAIFS MOP-AQAI algorithm proposed in this paper were used to train and test with UNSW-NB15 set, the ROC curves on ten data set types are shown in Figures 5 and 6.

Figure 5. The ROC curves of different types of data sets with various algorithms for UNSW-NB15

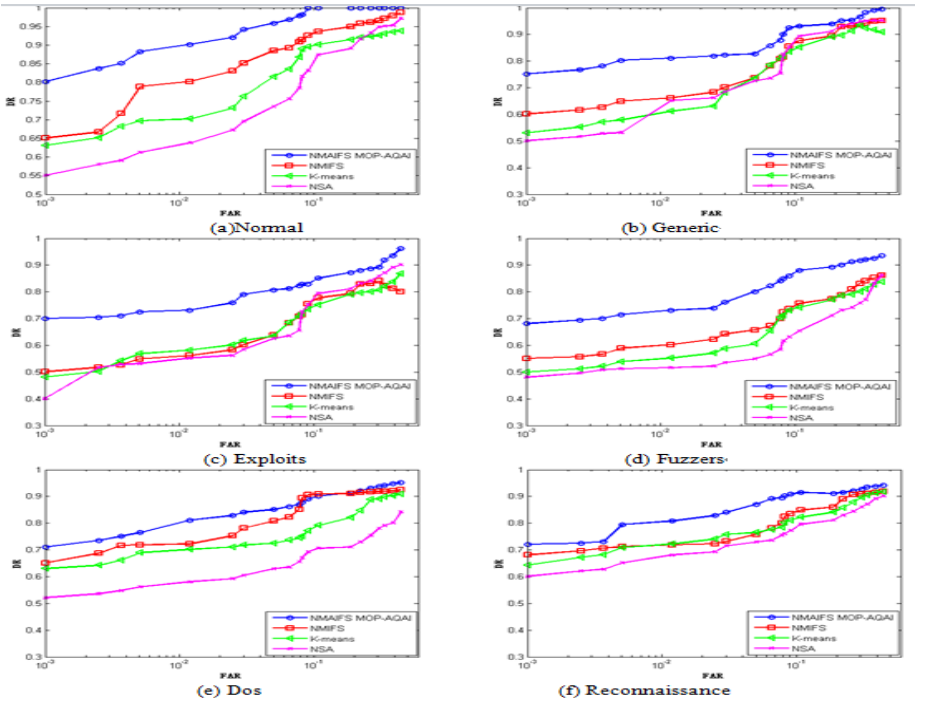
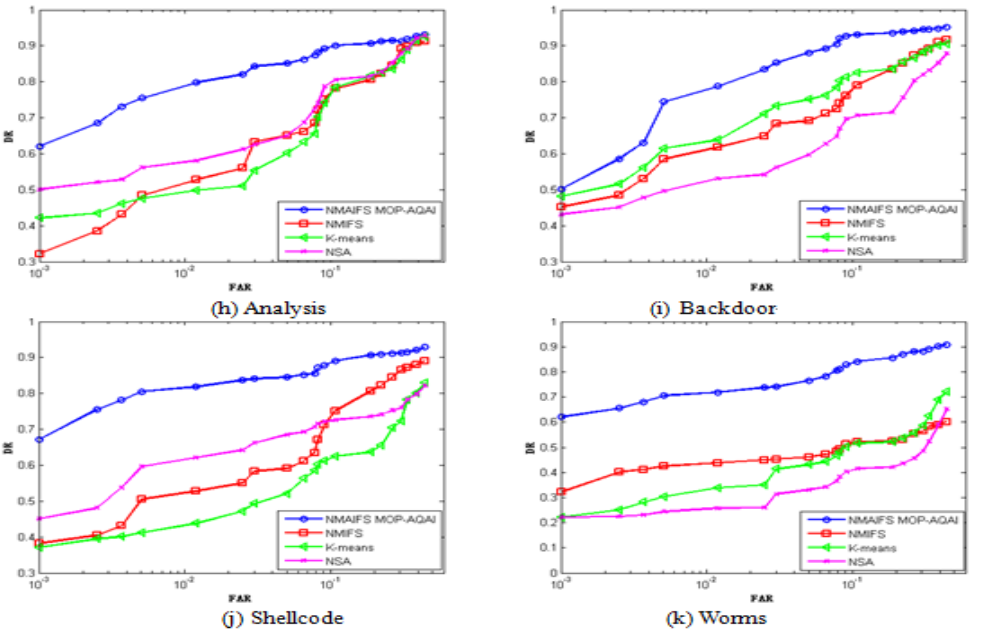


Figure 6. The ROC curves of different types of data sets with various algorithms for UNSW-NB15



Figures 5 and 6 show that NMAIFS MOP-AQAI performs better than NSA, K-means, and NMIFS. In UNSW-NB15, Analysis, Backdoor, Shellcode, and Worms are small sample sets, the DR is the highest, and the FAR is the lowest in the four algorithms. Meanwhile, NMAIFS MOP-AQAI has a better balance between DR and FAR.

THE ANOMALY DETECTION

Similarly, the anomaly detection is used to analysis the intrusion detection indexes of NMAIFS MOP-AQAI.

All nine types of attacks are abnormal samples for training; the detection sets are used for detection, and each dataset group runs ten times and calculates the average values. Different algorithms' anomaly detection performance comparisons are shown in Table 11.

Table 11. Detection performance comparisons of classification on UNSW-NB15(the N/A reflects an unknown value)

Method	DR	Acc	FAR	Pre
SL+DCA (Wu et al., 2020)	N/A	98.73	N/A	99.17
NSNAD (Naila et al., 2020)	91.34	92.00	9.76	95.00
DCA (Farzadnia et al., 2021)	95.90	78.70	59.20	61.83
DCA+NSA (Ehsan et al., 2021)	99.10	97.30	15.30	86.63
NMAIFS MOP-AQAI	99.58	99.26	0.46	85.32

Results in Table 11 show that the detection rate and accuracy of the NMAIFS MOP-AQAI algorithm proposed in this paper are higher than other algorithms. The Acc is 13.85% lower than SL+DCA, 9.68% less than NSNAD, and 1.31% less than DCA+NSA. Therefore, the NMAIFS MOP-AQAI algorithm strikes a better balance between DR and FAR.

In conclusion, whether for KDD99 or UNSW-NB15, NMAIFS MOP-AQAI has a higher DR, Acc, Pre, lower FAR, a better balance between DR and FAR, and a better classification ability and anomaly detection ability. In particular, for small sample sets, it has superior detection performance.

CONCLUSION

This paper applied the NMAIFS algorithm to obtain the best features collection, by which we got an effective dimension reduction of the multiple dimensional features to improve the detection speed. Then, we developed a cooperative evolution of multiple operators based adaptive parallel quantum artificial immune algorithm. In which, we used individual similarity evaluation operator, individual fitness evaluation operator, and individual similarity evaluation operator to update individual's mutation probability; simultaneously, we applied parallel quantum and vaccination strategies to improve the adaptive artificial immune algorithm for the diversity of individuals. The goal is to obtain effective classifiers for high DR and low FAR. At last, with KDD99 and UNSW-NB15, experiment results on anomaly data from real network traffic showed that the proposed method NMAIFS MOP-AQAI has a higher DR, lower FAR, and better adaptivity, especially for small samples. While the drawback of the algorithm is that for unknown attacks, the performance needs further research. The future jobs are to research the zero-day attacks and how to detect the unknown attacks and deploy the algorithm in the Internet-of-Things (Ivan et al., 2021; Yuan et al., 2021).

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

Alshdadi, A. A., Alghamdi, A. S., Daud, A., & Hussain, S. (2021). Blog backlinks malicious domain name detection via supervised learning. *International Journal on Semantic Web and Information Systems*, *17*(3), 1–17. doi:10.4018/IJSWIS.2021070101

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

Anupama, M., Neena, G., & Gupta, B. B. (2021). Defense mechanisms against DDoS attack based on entropy in SDN-cloud using POX controller. *Telecommunication Systems*, *77*(1), 47–62. doi:10.1007/s11235-020-00747-w

Castillo-Zúñiga, I., Luna-Rosas, F. J., Rodríguez-Martínez, L. C., Muñoz-Arteaga, J., López-Veyna, J. I., & Rodríguez-Díaz, M. A. (2020). Internet data analysis methodology for cyberterrorism vocabulary detection, combining techniques of big data analytics, NLP and semantic web. *International Journal on Semantic Web and Information Systems*, *16*(1), 69–86. doi:10.4018/IJSWIS.2020010104

Chou, F.-I., Ho, W.-H., Chen, Y. J., Tsai, J.-T., & Chang, C.-W. (2020). Detecting mixed-type intrusion in high adaptability using artificial immune system and parallelized automata. *Applied Sciences (Basel, Switzerland)*, *10*(5), 1566. doi:10.3390/app10051566

Ehsan, F., Hossein, S., & Alireza, N. (2021). A novel sophisticated hybrid method for intrusion detection using the artificial immune system. *Journal of Information Security and Applications*, *58*, 102721. doi:10.1016/j.jisa.2020.102721

Farzadnia, E., Shirazi, H., & Nowroozi, A. (2021). A new intrusion detection system using the improved dendritic cell algorithm. *The Computer Journal*, *8*(64), 1193–1214. doi:10.1093/comjnl/bxaa140

Fatemeh, A. (2011). Mutual information-based feature selection for intrusion detection systems. *Journal of Network and Computer Applications*, *34*(4), 1184–1199. doi:10.1016/j.jnca.2011.01.002

Fatemidokht, H., Rafsanjani, M. K., Gupta, B. B., & Hsu, C. H. (2021). Efficient and secure routing protocol based on artificial intelligence algorithms with UAV-assisted for vehicular ad hoc networks in intelligent transportation systems. *IEEE Transactions on Intelligent Transportation Systems*, *22*(7), 4757–4769. doi:10.1109/TITS.2020.3041746

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

Huang, W. W., Zhang, J. W., & Sun, H. Y. (2017). An anomaly detection method based on normalized mutual information feature selection and quantum wavelet neural network. *Wireless Personal Communications, 96*, 2693-2713. 10.1007/s11277-017-4320-2

Ivan, C., Dragan, P., Marko, P., & Brij, G. (2021). Ensemble machine learning approach for classification of IoT devices in smart home. *International Journal of Machine Learning and Cybernetics*, *12*(11), 3179–3202. doi:10.1007/s13042-020-01241-0

KDD. (2010). *KDD Cup 1999 Data*. http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

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 Science*, *3*(3), 266–275. doi:10.26782/jmcms.spl.3/2019.09.00020

Lee, J., & Park, K. (2019). AE-CGAN Model based high performance network intrusion detection system. *Applied Sciences (Basel, Switzerland)*, *20*(9), 4221. doi:10.3390/app9204221

Lv, X., Hou, H., You, X., Zhang, X., & Han, J. (2020). Distant supervised relation extraction via DiSAN-2CNN on a feature level. *International Journal on Semantic Web and Information Systems*, *16*(2), 1–17. doi:10.4018/IJSWIS.2020040101

Madan, K., & Bhatia, R. K. (2021). Ranked deep web page detection using reinforcement learning and query optimization. *International Journal on Semantic Web and Information Systems*, *17*(4), 99–121. doi:10.4018/IJSWIS.2021100106

Mishra, A., Gupta, N., & Gupta, B. B. (2021). Defense mechanisms against DDoS attack based on entropy in SDN-cloud using POX controller. *Telecommunication Systems*, *77*(1), 47–62. doi:10.1007/s11235-020-00747-w

Moustafa, N. (2017). *Designing an online and reliable statistical anomaly detection framework for dealing with large high-speed network traffic* [Doctoral thesis, University of New South Wales]. UNSW Open Access institutional repository. http://hdl.handle.net/1959.4/58748

Naila, B. A., Mohamed, G., & Abdelouahid, D. (2020). NSNAD: Negative selection-based network anomaly detection approach with relevant feature subset. *Neural Computing & Applications*, *32*(8), 3475–3501. doi:10.1007/s00521-019-04396-2

Nguyen, H. B., Xue, B., Andreae, P., & Zhang, M. (2017). Particle swarm optimisation with genetic operators for feature selection. *2017 IEEE Congress on Evolutionary Computation (CEC)*, 286-293. doi:10.1109/CEC.2017.7969325

Sahar, A., Daniyal, A., & Li, C. (2020). DeepDCA: Novel network-based detection of IoT attacks using artificial immune system. *Applied Sciences (Basel, Switzerland)*, *10*(6), 1909. doi:10.3390/app10061909

Sahoo, S. R., & Gupta, B. B. (2021). Multiple features based approach for automatic fake news detection on social networks using deep learning. *Applied Soft Computing*, *100*, 106983. doi:10.1016/j.asoc.2020.106983

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

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

Tewari, A., & Gupta, B. B. (2020). Secure timestamp-based mutual authentication protocol for IoT devices using RFID tags. *International Journal on Semantic Web and Information Systems*, *16*(3), 20–34. doi:10.4018/IJSWIS.2020070102

Wu, Z., Wang, J., Hu, L., Zhang, Z., & Wu, H. (2020). A network intrusion detection method based on semantic re-encoding and deep learning. *Journal of Network and Computer Applications*, *164*, 102688. doi:10.1016/j.jnca.2020.102688

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

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

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 Computing*, 1. Advance online publication. doi: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

Table 12. The abbreviations used in this research

Abbreviation	The full name
IDS	Intrusion detection system
NMAIFS MOP- AQAI	Normalized mutual antibodies information feature selection and adaptive quantum artificial immune with cooperative evolution of multiple operators
NMAIFS	Normalized mutual antibodies information feature selection
MOP- AQAI	Adaptive quantum artificial immune with cooperative evolution of multiple operators
GA	Genetic algorithm
AI	Artificial immune
FSA	Feature selection algorithm
mRMR	Minimal- Redundancy- Maximal- Relevance
NMIFS	Normalized mutual information feature selection
MIFS	Mutual information feature selection
AE - CGAN – RF	Autoencoder - conditional, the generative adversarial networks and the random forest
WIDMoDS	An intrusion detection model based on dynamic weighted values
NSA	Negative selection algorithm
NSNAD	Negative selection for network anomaly detection
DCA	Dendritic cells algorithm
DL	Deep learning
PA	Parallel automaton
MDPCA-DBN	Modified Density Peak Clustering Algorithm and Deep Belief Networks
AA-HMM	Anti-adversarial hidden markov model for network-based intrusion detection
MS-NIDAM	Multisource neighborhood immune detector adaptive model
MI	The mutual information
QWNN	Quantum wavelet neural network
SR	Semantic re-encoding
ROC	Receiver operating characteristic