Network Intrusion Detection Based on Directed Acyclic Graph and Belief Rule Base

Bang-Cheng Zhang, Guan-Yu Hu, Zhi-Jie Zhou, You-Min Zhang, Pei-Li Qiao, and Lei-Lei Chang

Intrusion detection is very important for network situation awareness. While a few methods have been proposed to detect network intrusion, they cannot directly and effectively utilize semi-quantitative information consisting of expert knowledge and quantitative data. Hence, this paper proposes a new detection model based on a directed acyclic graph (DAG) and a belief rule base (BRB). In the proposed model, called DAG-BRB, the DAG is employed to construct a multi-layered BRB model that can avoid explosion of combinations of rule number because of a large number of types of intrusion. To obtain the optimal parameters of the DAG-BRB model, an improved constraint covariance matrix adaption evolution strategy (CMA-ES) is developed that can effectively solve the constraint problem in the BRB. A case study was used to test the efficiency of the proposed DAG-BRB. The results showed that compared with other detection models, the DAG-BRB model has a higher detection rate and can be used in real networks.

Keywords: Network intrusion detection, Belief rule base, Directed acyclic graph, Covariance matrix adaption evolution strategy, Evidential reasoning rule.

I. Introduction

Network intrusion detection is a critical problem in network situation awareness [1] because different attacks pose different threats to networks. Network intrusion detection is a complex multi-classification problem used to identify attacks on a network, and it is a key factor for assessing the state of network security. Currently available methods for network intrusion detection can be divided into two types. The first consists of direct methods such as the neural network-based model [2]–[4], the support vector machine (SVM)-based model, and belief-based pattern classification models [5]–[7]. The second type of network intrusion detection methods consists of combination methods, such as the one-against-one SVM [8] the one-against-all SVM [8], and the DAG-SVM [9].

As a popular direct method, the back propagation (BP) neural network uses the principle of empirical risk minimization to detect network intrusion [10]. However, the BP neural network is a black box model, which makes it difficult to integrate human expert knowledge into its learning process. Moreover, the number of parameters and the training time of the BP neural network increase with the number of dimensions of data, and its detection accuracy declines as the complexity of the problem grows. The SVM model is a mainstream detection model that can use the structural risk minimization principle for network intrusion detection. Because an SVM can only distinguish two types of network attacks, combinatorial SVMs have been proposed, such as the DAG-SVM, which contains SVMs combined by using the DAG structure [11].

Manuscript received May 12, 2016; revised Feb. 11, 2017; accepted May 8, 2017.

Bang-Cheng Zhang (zhangbangcheng@mail.ccut.edu.cn) is with the School of Mechatronic Engineering, Changchun University of Technology, China.

Guan-Yu Hu (corresponding author, huguanyu0708@163.com) is with the School of Information Science and Technology, Hainan Normal University, Haikou, China.

Zhi-Jie Zhou (corresponding author, zhouzj04@mails.tsinghua.edu.cn) and Lei-Lei Chang (leiliechang@hotmail.com) are with the High-Tech Institute of Xi'an, China.

You-Min Zhang (Youmin.Zhang@concordia.ca) is with the Department of Information and Control Engineering, Xi'an University of Technology, China.

Pei-Li Qiao (qiaopeili@163.com) is with the School of Computer Science and Technology, Harbin University of Science and Technology, Harbin, China.

The above-mentioned models are data driven methods that cannot effectively utilize semi-quantitative information containing expert knowledge and quantitative data. Moreover, these models cannot deal with uncertain information, whereas a great deal of it needs to be considered in real network systems.

The belief rule base (BRB) is an expert system that can utilize semi-quantitative information [12], [13], and has been applied to many fields [14]–[18]. Therefore, the BRB can be used as detection model for network attacks. However, the following problems need to be considered:

(1) The original BRB model needs to be adjusted for network intrusion detection. The excessively large number of referenced values of the BRB model cause a combinatorial explosion in classification because the number of referenced values determine the rule number in BRB [19]. Therefore, a new hybrid BRB model is proposed in this paper, where a few BRB models establish a multi-layered model by using a DAG structure.

(2) To detect network attacks, it is necessary to collect network data, which is voluminous and contains protocol type, duration, features of TCP connect, features of hosts, and so on. Only a part of network data is associated with network attacks, but this is challenging for experts to identify and locate. Thus, principal component analysis (PCA) is used in this paper. However, the network data lose their meaning following dimension reduction by using PCA. Therefore, when these related network data processed by PCA are used as the antecedent attributes of the BRB model, the experts cannot determine the referenced values of these attributes. This paper shows how to automatically determine the optimally referenced values.

(3) Although the initial values of the belief degrees, and the rule weights and attributes of the DAG-BRB model can be provided by experts, these parameters should be optimized to obtain more accurate results.

From the above, it is clear that the parameters to be optimized in the DAG-BRB model are the referenced values, the belief degrees, and the rule weights and attributes. Thus, an objective function is constructed, and an improved constraint covariance matrix adaptation evolution strategy (CMA-ES) [20], [21], [22], [23] is developed to train the parameters.

The remainder of this paper is organized as follows: In Section II, based on the BRB model, the problem of network intrusion detection is formulated. In Section III, a new DAG-BRB classification model is constructed. In Section IV, based on the improved CMA-ES algorithm, the optimized process of the proposed model is developed. A case study to detect network intrusion is presented to test the proposed method in Section V, and the conclusions of this paper are provided in Section VI.

II. BRB Model for Network Intrusion Detection

1. Problem Formulation

Network intrusion detection is a complex multi-classification problem. Network attacks can be broadly divided into four types [24], [25]: the denial-of-service (DoS) attack, which crashes a server by sending massive amounts of network data; the surveillance or probe (Probe) attack, which compromises the privacy of other hosts using the scanning technique; the remote-to-local (R2L) attack, which logs in through remote computers by exploiting security loopholes; and the user-to-root (U2R) attack, which executes unauthorized operations using the highest privileges of a server. Different attacks pose different threats. For example, the DoS attack is a serious threat to a Web server, which can crash as a result. As a prelude to the hacker attack, the Probe attack is a lightweight threat to the network.

In light of the above, to more accurately assess of a network's security situation, network attacks need to be classified accurately. In this paper, they are divided into five classes: normal data, DoS, Probe, U2R, and R2L attacks, and are denoted by 1, 2, 3, 4, and 5, respectively. The purpose of this paper is to distinguish classes of network attacks using the trained DAG-BRB model, which uses both expert knowledge and observable data. The basic principles of the BRB model and the ER rule are detailed below.

2. Background of the BRB Model

Assume that x_i denotes the ith antecedent attribute (i.e., the input of BRB) and M denotes the number of antecedent attributes. Then, a belief rule is described as in [17]:

$$\begin{aligned} &R_k: \text{If}\left(\mathrm{x}_1 \text{ is } A_1^k\right) \wedge \left(\mathrm{x}_2 \text{ is } A_2^k\right) \wedge \cdots \wedge \left(\mathrm{x}_M \text{ is } A_M^k\right), \\ &\text{Then } \left\{\left(D_1, \beta_{1,k}\right), \left(D_2, \beta_{2,k}\right), \ldots, \left(D_N, \beta_{N,k}\right)\right\} \\ &\text{With a rule weight } \theta_k \text{ and attribute weight } \delta_1, \delta_2, \ldots, \delta_M \end{aligned} \tag{1}$$

where $R_k (k = 1, 2, \ldots, L)$ denotes the kth rule of the BRB model, and $A_i^k (i = 1, 2, \ldots, M)$ denotes the referential value of the ith input to the kth rule. $D_j (j = 1, 2, \ldots, N)$ denotes the jth consequent, $\beta_{j,k}$ denotes the belief degree of the jth consequent of the

kth rule, θ_k denotes rule weight, and δ_i denotes attribute weight.

A BRB model is composed of a number of belief rules as shown in (1). When the input data are available, the evidential reasoning (ER) approach [26], [27], [28] is used to aggregate the belief rules to generate the results of the final assessment [13].

3. Background of the ER Rule

As the inference tool of the BRB model, the ER rule can deal with multiple items of uncertain information and integrate qualitative knowledge with quantitative data.

Assume that there are some basic attributes of a general attribute in a two-level hierarchy. The ER rule can integrate these basic attributes to obtain an evaluation of the general attribute in terms of assessment grades. The ER rule first uses the initial weights of the basic attributes given by experts and converts them into a basic probability mass. The basic combination principle of the ER rule is introduced in (12)–(19).

III. DAG-BRB Model for Network Intrusion Detection

1. Framework of the DAG-BRB Model

As mentioned above, network intrusion detection is a complex multi-classification problem. Its practical applications have shown that the detection rate is particularly low, and certain special types of attacks cannot be detected. This is because meaning of network data is lost after reducing their number of dimensions using PCA, and this is more evident in complex multi-classification problems. Another reason is that the differences between normal data and those used for network attacks are minor.

To solve the above problems, a DAG-BRB model is proposed where PCA is used to reduce the number of dimensions of the input network data, and the referenced values of the antecedent attributes are used as the parameters to be optimized. Several BRB models are then combined by using a directed acyclic graph (DAG) structure, where a single BRB model is mainly used to distinguish two types of network attacks. The framework of the DAG-BRB model is shown in Fig. 1.

In Fig. 1, the network dataset contains the attributes and the labels of network data, such as the IP address, the TCP fields, and the connection time. It is necessary to remove unrelated data. Thus, the network's dataset is first processed using PCA.

In this paper, the outputs processed by PCA contain only five attributes. The dataset is then divided into a testing dataset and a training dataset. In the sorter operation, the training dataset is divided into 10 subsets. Each contains only two classes of network attacks. Every BRB model needs to be trained by using the corresponding subset. For example, a subset containing normal data and DoS attack data, called the Normal-DoS BRB model, is used to train the BRB model by using the constrained CMA-ES algorithm and the ER approach. The detailed structure of these BRB models is described below.

2. Combination BRB Model Based on DAG

In Fig. 1, several BRB models are used to form a combination BRB model, as shown in Fig. 2, where every node in the combination model is a BRB model that only distinguishes two types of network attacks. When all BRB models have been trained by using the corresponding training subset, the testing dataset is entered into the combination BRB model as shown in Fig. 2. The detailed procedure is as follows:

Step 1. The testing dataset is first placed in the top node (Normal-U2R BRB model), they are determined to be normal data or U2R attack data.

Step 2. The testing data are entered into the other BRB model in the second layer according to type. If the testing data are normal, they are entered into the Normal-R2L

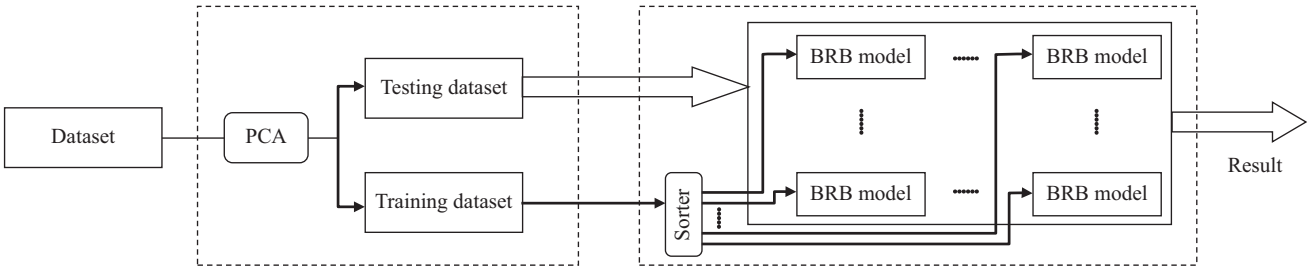


Fig. 1. Framework of the DAG-BRB model.

https://doi.org/10.4218/etrij.17.0116.0305

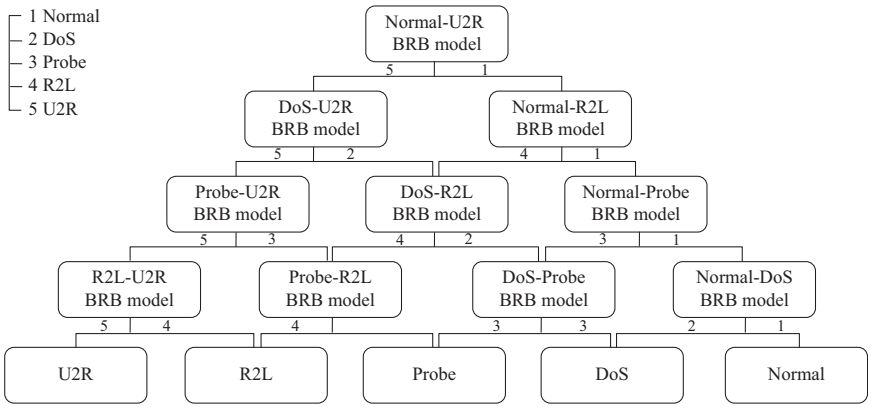


Fig. 2. Combination BRB model using directed acyclic graph.

BRB model, the right node in the second layer, and are judged to be normal or R2L attack data.

Step 3. By repeating the above steps, the final result can be obtained in the bottom layer. For example, when Probe attack data are in the combination model, the procedure can be described as shown in Fig. 3, where it is evident that network data processing by PCA requires four BRB models in the combination model. If the network data do not belong to any type of BRB model, they are assigned to the most similar type. In Fig. 3, the network data are first handled by the Normal-U2R model and then judged as belonging to the U2R attack type. The final result can be obtained in the final layer.

The disadvantage of the DAG structure is that errors in the top layer are transferred down (passed) to the other layers. A mistake in the upper model may cause an error in the final result. Therefore, it is important to improve the accuracy of every BRB model.

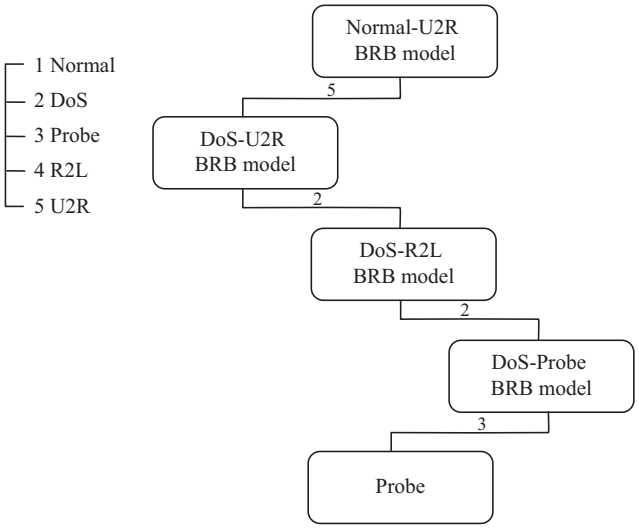


Fig. 3. Classification path of Probe attack data.

IV. Improved CMA-ES Algorithm to Train the DAG-BRB Model

As mentioned above, every BRB model in the Fig. 2 needs to be trained independently. The training process of these models is described below.

1. PCA Process of the DAG-BRB Model

Since most attributes of network data are unimportant in this context, it is necessary to reduce their number of dimensions. In this paper, PCA [29] was used for this. The basic process is as follows:

1) Calculating the mean value of network data. Assume that **X** denotes network data:

$$\mathbf{X} = \begin{pmatrix} x_{11} & \cdots & x_{1M} \\ \vdots & \ddots & \vdots \\ x_{T1} & \cdots & x_{TM} \end{pmatrix}, \tag{4}$$

where M is the number of the attributes that denote the dimensionality of the network data and T denotes the number of the network data samples. The mean value can be calculated as:

$$mean_h = \frac{1}{M}\sum_{i=1}^{M} x_{hi} \quad (h = 1, \ldots, T), \tag{5}$$

where $h \in (1, \ldots, T)$ denotes the *hth* network data sample and $i \in (1, \ldots, M)$ denotes the *ith* dimension of a sample.

2) **X** can be transformed into the following form:

$$\mathbf{X} = \begin{pmatrix} x_{11} - mean_1 & \cdots & x_{1M} - mean_1 \\ \vdots & \ddots & \vdots \\ x_{T1} - mean_T & \cdots & x_{TM} - mean_T \end{pmatrix}. \quad (6)$$

3) Calculating the covariance matrix:

$$\text{cov} = \frac{\mathbf{X}' \times \mathbf{X}}{T}, \quad (7)$$

where $\mathbf{X}^{'}$ denotes the transport matrix of $\mathbf{X}$.

4) Calculating the eigenvalue $\mathbf{e}$ of the covariance matrix in descending order.

5) Generating transformational matrix $\mathbf{E}$ using the previous eigenvalue $M^{'}$ in $\mathbf{e}$. The new network data $\tilde{\mathbf{X}}$ can then be calculated as:

$$\tilde{\mathbf{X}} = \mathbf{E} \times \mathbf{X}, \quad (8)$$

where $M^{'}$ denotes the number of attributes after reducing the dimensionality of the data using PCA. In this paper, $\tilde{\mathbf{X}}$ always denotes the training dataset processed by PCA.

2. Sorting Process

The sorting process is used to separate the training dataset into different subsets according to the corresponding BRB model. As shown in Fig. 1, following PCA, the training dataset $\tilde{\mathbf{X}}$ is fed into the sorter. $\tilde{\mathbf{X}}$ is divided into 10 groups as shown in Fig. 2, and each contains two types of attacks. $\tilde{\mathbf{X}}_r$ is described as

$$\tilde{\mathbf{X}}_r = \begin{pmatrix} \tilde{x}^r_{11} & \cdots & \tilde{x}^r_{1M} \\ \vdots & \ddots & \vdots \\ \tilde{x}^r_{T'1} & \cdots & \tilde{x}^r_{T'M} \end{pmatrix}, \quad (9)$$

where $\tilde{\mathbf{X}}_r \in \tilde{\mathbf{X}}$, and r denotes the rth BRB model in the DAG-BRB model, $r = (1, \ldots, 10)$. The order of r is top to bottom and left to right. $T^{'}$ denotes the number of network data items of the rth BRB model.

However, one problem remains outstanding: when PCA is applied to the BRB model, the meaning of $\tilde{\mathbf{X}}_r$ is lost, and experts can no longer determine the referenced values of the antecedent attributes. In the next section, the referenced values are treated as parameters to be optimized.

3. Optimizing Objective Function of the DAG-BRB

When the processed data $\tilde{\mathbf{X}}_r$ are received, they are first entered into the rth BRB model. The inference process of the $r$thBRB model can be described as follows:

1) The matching degree of a network data item $(\tilde{x}^r_{h1}, \ldots, \tilde{x}^r_{hM})$ is calculated, where h denotes the hth network data item of $\tilde{\mathbf{X}}_r$, and $h = (1, \ldots, T^{'})$. The matching degree a^k_i of the ith antecedent attribute of $(\tilde{x}^r_{h1}, \ldots, \tilde{x}^r_{hM})$ in the kth rule can be calculated by:

$$a^k_i = \begin{cases} \frac{A^{l+1}_i - \tilde{x}^r_{hi}}{A^{l+1}_i - A^l_i} & k = l\left(A^l_i \le \tilde{x}^r_{hi} \le A^{l+1}_i\right) \\ 1 - a^k_i & k = l+1 \\ 0 & k = 1, \ldots, N(k \ne l, l+1) \end{cases}, \quad (10)$$

where A^l_i and A^{l+1}_i denote the referential values.

2) The activated weight w_k of the kth rule is calculated according to the following equation:

$$w_k = \frac{\theta_k \prod_{i=1}^{M} \left(a^k_i\right)}{\sum_{l=1}^{L} \theta_l \prod_{i=1}^{M} \left(\alpha^l_i\right)}, \quad (11)$$

where θ denotes the initial weight of the rules.

3) The belief degrees of the outputs are calculated by using the ER rule. When the activated weight is calculated by (3), a belief rule can be activated if it is not 0. Then, the belief degrees can be converted into the following basic probability masses:

$$m_{j,k} = w_k \beta_{j,k}, \quad (12)$$

$$m_{D,k} = 1 - w_k \sum_{j=1}^{N} \beta_{j,k}, \quad (13)$$

$$\bar{m}_{D,k} = 1 - w_k, \quad (14)$$

$$\tilde{m}_{D,k} = w_k \left(1 - \sum_{j=1}^{N} \beta_{j,k}\right), \quad (15)$$

where $m_{j,k}$ denotes the probability mass of the jth class D_j in the kth rule, $m_{D,k}$ denotes the rest probability mass that is not distributed to any other class, and $\bar{m}_{D,k}$ denotes the unimportant degree of the kth rule. If the kth rule is completely important, $\bar{m}_{D,k} = 0$. $\tilde{m}_{D,k}$ denotes the degree of incompleteness of the kth rule. If the result of the evaluation of the kth rule is complete, $\tilde{m}_{D,k} = 0$.

Using the ER approach to combine the first k rules, the probability mass of the jth class D_j can be calculated as:

https://doi.org/10.4218/etrij.17.0116.0305

$$m_{j,I(k+1)} = K_{I(k+1)}\begin{bmatrix} m_{j,I(k)}m_{j,k+1} + m_{j,I(k)}m_{D,k+1} \\ +m_{D,I(k)}m_{j,k+1} \end{bmatrix}, \quad (16)$$

$$m_{D,I(k)} = \bar{m}_{D,I(k)} + \tilde{m}_{D,I(k)}, \quad (17)$$

$$\tilde{m}_{D,I(k+1)} = K_{I(k+1)}\begin{bmatrix} \tilde{m}_{D,I(k)}\tilde{m}_{D,k+1} + \tilde{m}_{D,I(k)}\bar{m}_{D,k+1} \\ +\bar{m}_{D,I(k)}\tilde{m}_{D,k+1} \end{bmatrix}, \quad (18)$$

$$\bar{m}_{j,I(k+1)} = K_{I(k+1)}\left[\bar{m}_{D,I(k)}\bar{m}_{D,k+1}\right], \quad (19)$$

where $I(k)$ denotes the combination of the first k rules and $m_{j,I(k)}$ denotes the probability mass of the jth class D_j having combined the first k rules using Dempster's combination rule. $K_{I(k+1)}$ can be calculated as:

$$K_{I(k+1)} = \frac{1}{1 - \sum_{j=1}^{N}\sum_{\substack{q=1 \\ q\neq j}}^{N} m_{j,I(k)}m_{q,k+1}}. \quad (20)$$

The belief degree of D_j can be calculated as:

$$\widehat{\beta}_j = \frac{m_{j,I(L)}}{1 - \bar{m}_{D,I(L)}} \quad (j = 1, 2, \ldots, N). \quad (21)$$

Through the above description, the belief degree can be calculated by the following analytical expression:

$$\widehat{\beta}_j = \frac{\mu \times \left[\prod_{k=1}^{L}\left(w_k\beta_{j,k} + 1 - w_k\sum_{i=1}^{N}\beta_{i,k}\right) - \prod_{k=1}^{N}\left(1 - w_k\sum_{i=1}^{N}\beta_{i,k}\right)\right]}{1 - \mu \times \left[\prod_{k=1}^{L}(1 - w_k)\right]}. \quad (22)$$

$$\mu = \left[\sum_{j=1}^{N}\prod_{k=1}^{L}\left(w_k\beta_{j,k} + 1 - w_k\sum_{i=1}^{N}\beta_{i,k}\right) - (N-1)\prod_{k=1}^{L}\left(1 - w_k\sum_{i=1}^{N}\beta_{i,k}\right)\right]^{-1}. \quad (23)$$

The result with the highest belief degree is the final classification, which is represented as $\widehat{j}$. Assume that E_h denotes the error in classification. If $\widehat{j} = j$, $E_h = 0$; if $\widehat{j} \neq j$, $E_h = 1$, where j denotes the original classification of the network data. Then, the objective function can be described as follows:

$$f(\mathbf{P}) = \frac{1}{T'}\sum_{h=1}^{T'} {E_h}^2, \quad (24)$$

where $\mathbf{P}$ denotes the optimal parameter vector of the BRB model that can be represented as:

$$\mathbf{P} = [\theta_1, \ldots, \theta_L, \beta_{1,1}, \ldots, \beta_{N,L}, A_1^1, \ldots, A_M^L]', \quad (25)$$

where $(\theta_1, \ldots, \theta_L)$ denotes the weights of L rules, $(\beta_{1,1}, \ldots, \beta_{N,L})$ denotes the belief degree of the outputs, where they are $N \times L$ in number, $(A_1^1, \ldots, A_M^L)$ denotes the referenced values of the antecedent attributes, and they are $L \times M$ in number.

Thus, the optimal problem can be described as:

$$\begin{aligned} &\min\{f(\mathbf{P})\} \\ &\text{s.t. } 0 \le \theta_k \le 1, k = 1, \ldots, L \\ &0 \le \beta_{j,k} \le 1, j = 1, \ldots, N, k = 1, \ldots, L \\ &\sum_{j=1}^{N}\beta_{j,k} = 1 \\ &lb_i \le A_i^k \le ub_i, i = 1, \ldots, M, k = 1, \ldots, L \\ &A_i^1 = lb_i \\ &A_i^L = ub_i, \end{aligned} \quad (26)$$

where lb_i denotes the lower bound and ub_i the upper bound of the ith attribute. The bound of the antecedent attributes should normally be set to the extreme value of the data. Since the network data are normalized before classification, the lower bounds are all set to 0 and the upper bounds to 1.

Remark 1: From (25), it is clear that the referenced values are treated as optimization parameters and experts cannot determine their values. However, the DAG-BRB model is still semi-quantitative because the other initial parameters are determined by experts, such as the weight of the rules and the belief degrees.

Remark 2: Note that the above training process is only for one model in the combination model, as shown in Fig. 1.

4. Optimization of the DAG-BRB Model

It is important to select an appropriate optimization algorithm to solve the problem described as (26). The following problems need to be considered:

1) From (26), it can be seen that the problem is a constrained optimization problem.

2) The objective function of the optimization problem is a complex, non-linear, and multi-peak function that needs a global optimization algorithm.

The covariance matrix adaption evolution strategy (CMA-ES) was proposed by Hansen [20]. It is a non-linear, non-convex optimization algorithm considered to be the state of the art [20]. The CMA-ES can quickly search for the global optimal solution in a small population but is an unconstrained optimization algorithm. An improved

constrained CMA-ES algorithm is proposed to solve the constraint optimization problem using constraint conditions in this paper.

The improved CMA-ES algorithm can be divided into four parts:

1) The sampling operation, which is used to an generate initial population with a multivariate normal distribution.

$$\mathbf{P}_q^{g+1} \sim \omega^g + \sigma^g \mathbb{N}(0, C^g)\ (q = 1, \ldots, \lambda), \tag{27}$$

where $\mathbf{P}_q^{g+1}$ denotes the qth solution of the population in the $(g + 1)$th generation; it also denotes the optimal parameter vector as described above. λ denotes the population size, ω denotes the mean of the population, and the initial value of ω^0 is equal to that of the initial parameter vector. σ denotes the step size, $\mathbb{N}$ denotes the normal distribution, and C denotes the covariance matrix of the population.

2) Selection operation. This operation is used to select τ solutions from the population according to the fitness values.

3) Recombination operation. This operation is used to update the mean value of the population:

$$\omega^{g+1} = \sum_{i=1}^{\tau} \eta_i \mathbf{P}_{i:\lambda}^{g+1} \quad \left(\sum_{i=1}^{\tau} \eta_i = 1\right), \tag{28}$$

where τ denotes the size of the offspring population, η denotes the weight coefficients, and $\mathbf{P}_{i:\lambda}^{g}$ denotes the ith individual of λ individuals in the gth generation.

4) Adjustment operation. This operation is used to update the covariance matrix of the population. Assume that the initial covariance matrix C^0 is a symmetric and positive-definite matrix; then, the new covariance matrix can be calculated as:

$$C^{g+1} = (1 - c_1 - c_2)C^g + c_1 p_c^{g+1} \left(p_c^{g+1}\right)^T + c_2 \sum_{i=1}^{\tau} \eta_i \Psi_{i:\lambda}^{g+1} \left(\Psi_{i:\lambda}^{g+1}\right)^T, \tag{29}$$

where c_1 and c_2 are the learning rates for updating the covariance matrix. Ψ in (29) can be calculated by

$$\Psi_{i:\lambda}^{g+1} = \frac{\left(\mathbf{P}_{i:\lambda}^{g+1} - \omega^g\right)}{\sigma^g}, \tag{30}$$

p_c in (29) denotes the path of evolution, where the initial $p_c = 0$. p_c can be updated as:

$$p_c^{g+1} = (1 - c_c)p_c^g + \sqrt{c_c(2 - c_c)\tau_{\text{eff}}} \frac{\omega^{g+1} - \omega^g}{\sigma^g}, \tag{31}$$

where $c_c \leq 1$ denotes the backward time horizon of the path of evolution. τ_{eff} is a variance-effective selection mass, and can be calculated as:

$$\tau_{\text{eff}} = \left(\frac{\|\eta\|_1}{\|\eta\|_2}\right)^2 = \frac{\|\eta\|_1^2}{\|\eta\|_2^2} = \frac{1}{\|\eta\|_2^2} = \left(\sum_{i=1}^{\tau} \eta_i^2\right)^{-1}. \tag{32}$$

The step size σ can be updated as:

$$\sigma^{g+1} = \sigma^g \exp\left(\frac{c_\sigma}{d_\sigma}\left(\frac{\|p_\sigma^{g+1}\|}{\mathbf{E}\|\mathbb{N}(0, \mathbf{I})\|} - 1\right)\right), \tag{33}$$

where d_σ denotes a damping parameter, $\mathbf{E}\|\mathbb{N}(0, \mathbf{I})\|$ denotes the expectation of the Euclidean norm of $\mathbb{N}(0, \mathbf{I})$, c_σ denotes the backward time horizon, and denotes the conjugate evolution path, with the initial $p_\sigma = 0$, p_σ can be updated by

$$p_\sigma^{g+1} = (1 - c_c)p_\sigma^g + \sqrt{c_\sigma(2 - c_\sigma)\tau_{\text{eff}}} C^{(g)-\frac{1}{2}} \frac{\omega^{g+1} - \omega^g}{\sigma^g}. \tag{34}$$

5) Constraint operation. This operation is used to satisfy the constraint conditions shown in (26) using constraint processing technology. In this paper, the direct modification method is used to process the constraints, where the solutions generated by selection operation need to be modified.

The above operations are executed in cycles. When the constraint condition is satisfied, the best solution $\mathbf{P}_{\text{best}}$ is obtained, and the rth BRB model is trained.

V. Case Study

The network intrusion detection problem was investigated using the KDD'99 dataset to test the efficiency of the DAG-BRB model, where the inputs were the network data and the outputs were types of network attacks.

A total of 3,156 samples representing five attack types were selected from the dataset, and the numbers of training and testing data items are shown in Table 1.

The training and testing data were selected randomly according to the above distribution. A total of 30 independent runs using the improved CMA-ES algorithm were performed for each instance. The case was

https://doi.org/10.4218/etrij.17.0116.0305

implemented in MATLAB R2013a on a Core (TM) i7-3632QM CPU 2.20 GHz with Windows 7 OS.

1. Results of the Trained DAG-BRB Model

There were nine rules for each model shown in Fig. 2, and the results of each only contained two types of attacks: 1 denotes one type and 2 the other. These models used the same initial parameters as shown in Table 2, where A_i denotes the initially referenced values.

Every BRB model in the combination model needed to be trained with the corresponding training data by using the improved CMA-ES algorithm. The initial mean of the CMA-ES algorithm was equal to the initial parameter vector in Table 2, and the improved CMA-ES algorithm was used to optimize the parameter vector of each BRB model. The optimization process is shown in Fig. 4, where the vertical axis denotes the average fitness values and the horizontal axis the number of iterations of the improved CMA-ES algorithm.

The final output $\mathbf{P}_{\text{best},r}$ was used to establish the BRB models and then the combination model. The results generated by the trained model are shown in Fig. 5 and Table 3.

As shown in Fig. 5 and Table 3, the results of the trained DAG-BRB model were satisfactory and, in some cases, the detection rate was at or close to 100%. It was very difficult to obtain good results for R2L and U2R because the sample size was too small, but the proposed model performed well.

Table 1. Distribution of the samples.

Training Data	Normal	DoS	Probe	R2L	U2R
	1,000	1,000	310	120	200
Testing Data	Normal	DoS	Probe	R2L	U2R
	200	200	62	24	40

Table 2. Initial parameters of belief rules.

Rule	Rule weight	Reference points		Belief degrees
1	1	0	0	0.5, 0.5
2	1	0	0.5	0.5, 0.5
3	1	0	1	0.5, 0.5
4	1	0.5	0	0.5, 0.5
5	1	0.5	0.5	0.5, 0.5
6	1	0.5	1	0.5, 0.5
7	1	1	0	0.5, 0.5
8	1	1	0.5	0.5, 0.5
9	1	1	1	0.5, 0.5

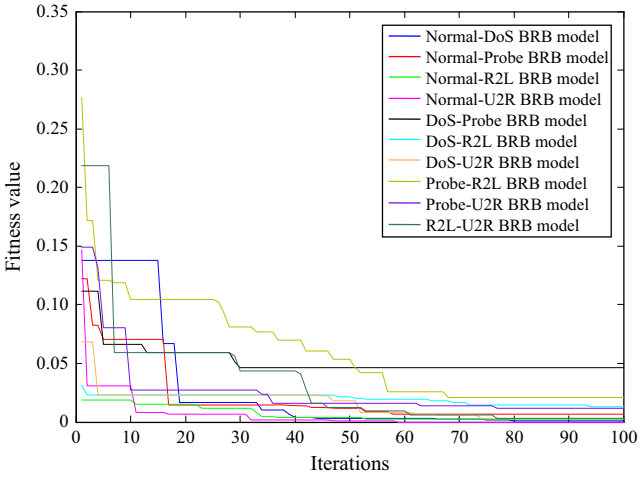


Fig. 4. Optimization process of the DAG-BRB model.

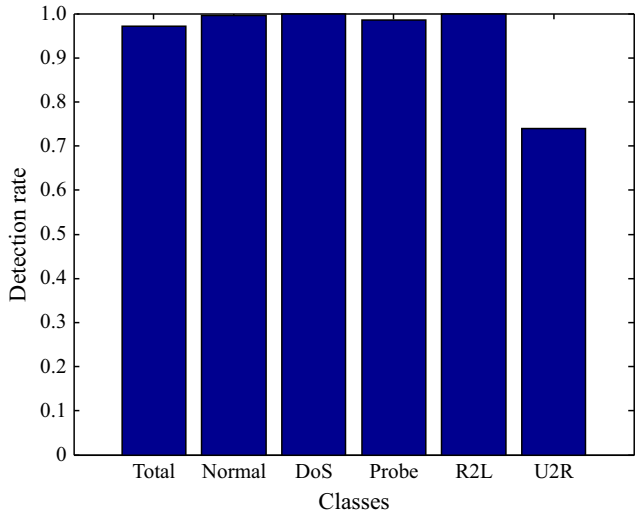


Fig. 5. Results generated by the trained DAG-BRB model.

Table 3. Detection rate of the model in Fig. 5.

Total	Normal	DoS	Probe	R2L	U2R
97.22%	99.5%	100%	98.46%	100%	74.00%

2. Comparison between DAG-BRB and BRB Models

To show that the combination BRB model can be applied to network intrusion detection, a comparison between the DAG-BRB model and the BRB model was conducted. In the BRB model, the attacks were divided into five classes, and the other sets of parameters were the same as in

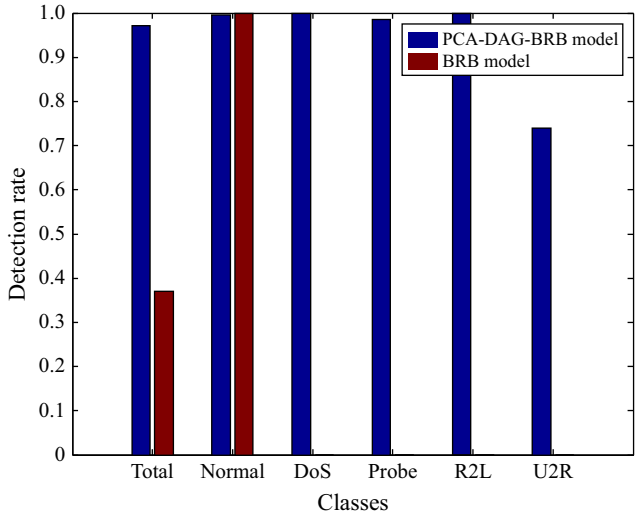


Fig. 6. Comparative results of the DAG-BRB and BRB models.

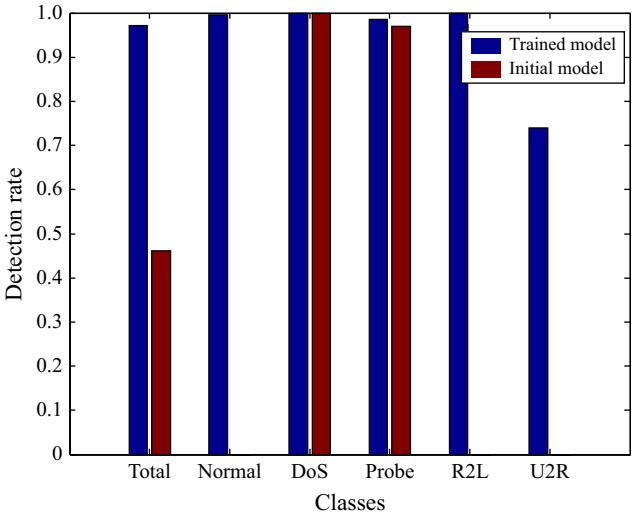


Fig. 7. Comparative results of the initial and trained DAG-BRB.

Table 4. Detection rate of the model in Fig. 6.

DAG-BRB	Total	Normal	DoS	Probe	R2L	U2R
	97.22%	99.5%	100%	98.46%	100%	74%
BRB	Total	Normal	DoS	Probe	R2L	U2R
	37.04%	100%	0%	0%	0%	0%

Table 5. Detection rate of the model in Fig. 7.

Trained model	Total	Normal	DoS	Probe	R2L	U2R
	97.2%	99.5%	100%	98.5%	100%	74%
Initial model	Total	Normal	DoS	Probe	R2L	U2R
	46.1%	0%	100%	96.9%	0%	0%

DAG-BRB. The comparative results are shown in Fig. 6 and Table 4.

From Fig. 6 and Table 4, it is clear that the results generated by the original BRB model were poor, which shows that it is unsuitable for network intrusion detection.

3. Comparison between Initial and Trained DAG-BRB Models

As mentioned above, the initial parameters of the DAG-BRB model were provided by an expert. To prove the need for the optimization algorithm in the DAG-BRB model, a comparison between the initial and the trained models was conducted using the improved CMA-ES algorithm, as shown in Fig. 7 and Table 5.

From Fig. 7 and Table 5, it is clear that the detection rates of the trained DAG-BRB model with the CMA-ES algorithm were more accurate than those of the initial DAG-BRB model.

4. Comparison between Improved CMA-ES and other Algorithms

Some other algorithms can be used to train the parameters of the DAG-BRB. Sequential quadratic programming (SQP) [30] and the constrained differential evolution algorithm (DE) [31] were chosen for this. SQP can solve a constraint problem by using a sequential quadratic programming sub-problem. DE is an effective algorithm with a simple mutation operation based on a differential strategy. The parameters of the algorithms are shown in Tables 6-7 and the comparative results in Fig. 8 and Table 8.

5. Comparison between DAG-BRB and other Models

DAG-SVM and BP were also used in the comparative study, where PCA was used to reduce the dimensionality

Table 6. Parameters of the CMA-ES algorithm.

Individuals	Iterations	Step size	c1	c2	cc
8	100	0.5	0.0473	0.3651	0.4502

Table 7. Parameters of the DE algorithm.

Individuals	Iterations	Step size
16	100	0.5

https://doi.org/10.4218/etrij.17.0116.0305

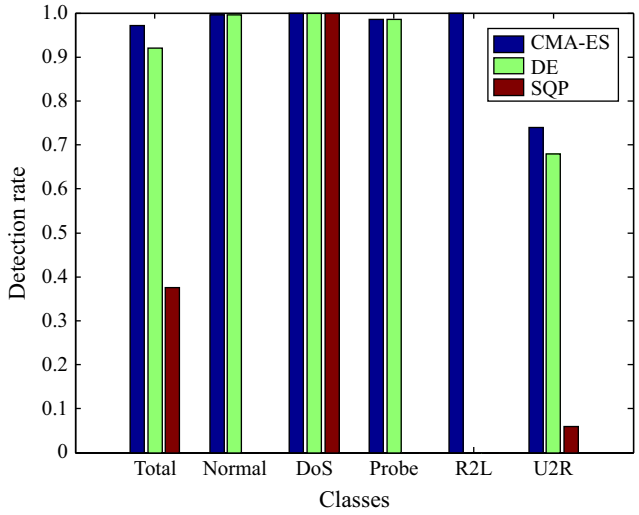


Fig. 8. Comparative results of CMA-ES and other algorithms.

Table 8. Detection rate of the model in Fig. 8.

CMA-ES	Total	Normal	DoS	Probe	R2L	U2R
	97.2%	99.5%	100%	98.5%	100%	74%
DE	Total	Normal	DoS	Probe	R2L	U2R
	92.0%	99.5%	100%	98.4%	0%	68%
SQP	Total	Normal	DoS	Probe	R2L	U2R
	37.6%	0%	100%	0%	0%	6%

of the samples and improved CMA-ES to train the parameters of the models. The DAG-SVM model is a multi-layered model combining several SVM classification models using DAG. The parameters of the DAG-SVM model and the BP model are shown in Tables 9-10. Their results in comparison with the DAG-BRB model are shown in Table 11. The performance index, in terms of accuracy, specificity, sensitivity, AUC, and precision and recall curves, is shown in Table 12 and Fig. 9.

Table 9. Parameters of the DAG-SVM model.

Models	Kernel	Type	Gamma	Cost
10	RBF	C-SVC	1	1

Table 10. Parameters of the BP model.

Input neuron no.	Hidden neuron no.	Output neuron no.
5	8	5

Table 11. Detection rates of the models.

DAG-BRB	Total	Normal	DoS	Probe	R2L	U2R
	97.2%	99.5%	100%	98.5%	100%	74%
DAG-SVM	Total	Normal	DoS	Probe	R2L	U2R
	95.7%	99.5%	100%	95.4%	100%	62%
BP	Total	Normal	DoS	Probe	R2L	U2R
	77.0%	100%	100%	24.6%	0%	0%

Table 12. Performance index of the models.

	Accuracy	Specificity	Sensitivity	AUC
DAG-BRB	99.47%	100%	100%	0.93
DAG-SVM	98.18%	100%	100%	0.89
BP	96.51%	97.35%	99.00%	0.86

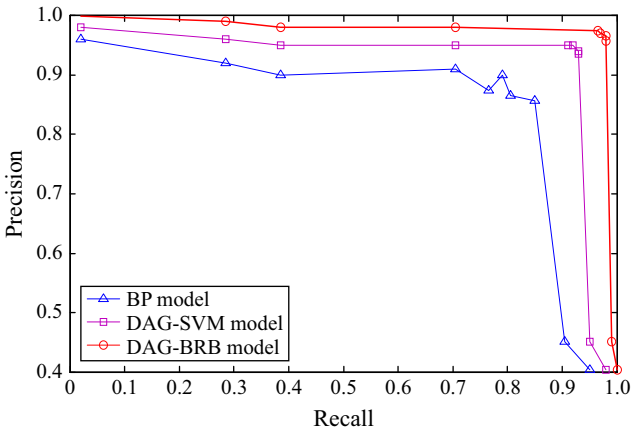


Fig. 9. Precision and recall curves of the models.

The above results show that compared with other models, the DAG-BRB model had better intrusion detection capability in the network.

VI. Conclusion

In this paper, a new DAG-BRB model was proposed for network intrusion detection and a constraint CMA-ES algorithm was also developed to train the parameters of this model. The contributions of this paper can be summarized as follows:

1) Directed acyclic graphs (DAGs) are used in the DAG-BRB model, where several BRB models form a combination model, and the referenced values of each BRB model are considered part of the parameters to be optimized.

2) The ER rule and the constraint CMA-ES algorithms are used to train the parameters of each BRB model in the DAG-BRB.

A case study was conducted to demonstrate how the DAG-BRB model can be used for network intrusion detection. The results showed that compared with the original BRB model, the initial DAG-BRB, as well as other models and optimization algorithms shown in Figs. 8 and 9, the trained DAG-BRB had superior intrusion detection capability in the network. One reason for the impressive results of by the DAG-BRB model might be that the belief rules have a stronger ability to describe uncertain network information. Another might be that the CMA-ES algorithm can better deal with optimization problems in the proposed model.

In future work, the authors will investigate reducing the computational complexity of the DAG-BRB model and finding more practical applications.

Acknowledgement

Bang-Cheng Zhang thanks the NSFC under Grant 61374138 for partial support. Guan-Yu Hu thanks the NSFC under Grant 61362016 and the Natural Science Foundation of Hainan Province under Grants 617120 and 617121 for partial support. Zhi-Jie Zhou thanks the NSFC under Grant 61370031, the Postdoctoral Science Foundation of China under Grants 2015M570847 and 2016T90938, and the Natural Science Foundation of Shaanxi under Grant 2015JM6354 for partial support.

References

[1] T. BASS, "Intrusion Detection System and Multi-sensor Data Fusion: Creating Cyberspace Situation Awareness," *Commun. ACM*, vol. 43, no. 4, Apr. 2000, pp. 99–105.

[2] Y.H. Liu, D.X. Tian, and A.N. Wang, "ANNIDS: Intrusion Detection System Based on Artificial Neural Network," In *IEEE Int. Conf. Mach. Learning Cybern.*, Xi'an, China, Nov. 5, 2003, pp. 1337–1342.

[3] A.K. Ghosh and A. Schwartzbard, "A Study in Using Neural Networks for Anomaly and Misuse Detection," In *Proc. USENIX Security Symp.*, Washington, D.C., USA, Aug. 23–26, 1999, pp. 141–152.

[4] J.M. Bonifacio et al., "Neural Networks Applied in Intrusion Detection," In *Proc. Int. Joint Conf. Neural Netw.*, Anchorage, AK, USA, May 4–9, pp. 205–210.

[5] P. Xu et al., "Evidential Calibration of Binary SVM Classifiers," *Int. J. Approximate Reasoning*, vol. 72, May 2016, pp. 55–70.

[6] Z.G. Liu et al., "Hybrid Classification System for Uncertain Data," *IEEE Trans. Syst., Man, Cybern.: Syst.*, no. 99, Nov. 2016, pp. 1–8.

[7] Z.G. Liu et al., "Credal Classification Rule for Uncertain Data Based on Belief Functions," *Pattern Recogn.*, vol. 47, no. 7, July 2014, pp. 2532–2541.

[8] C. Angdo and L. Gonzalez, "1-v-1 Tri-Class SV Machine," In *Proc. Eur. Symp. Artif. Neural Netw.*, Bruges, Belgium, Apr. 23–25, 2003, pp. 355–360.

[9] J.C. Platt, N. Cristianini, and J. Shawetaylor, "Large Margin DAGs for Multiclass Classification," In *Advances in Neural Information Processing Systems 12*, MIT Press, 2000, pp. 547–553.

[10] B. Widrow et al., "Neural Network Application in Industry, Business and Science," *Commun. ACM*, vol. 37, no. 3, Mar. 1994, pp. 93–105.

[11] C. Cortes and V. Vapnik, "Support Vector Networks," *Mach. Learn.*, vol. 20, no. 3, Sept. 1995, pp. 273–295.

[12] J.B. Yang and D.L. Xu, "Evidential Reasoning Rule for Evidence Combination," *Artif. Intell.*, vol. 205, Dec. 2013, pp. 1–29.

[13] F.J. Zhao et al., "A New Evidential Reasoning-Based Method for Online Safety Assessment of Complex Systems," *IEEE Trans. Syst., Man Cybern.: Syst.*, no. 99, Dec. 2016, pp. 1–13.

[14] Z.J. Zhou et al., "Hidden Behavior Prediction of Complex Systems Under Testing Influence Based on Semi-quantitative Information and Belief Rule Base" *IEEE Trans. Fuzzy Syst.*, vol. 23, no. 6, Dec. 2015, pp. 2371–2386.

[15] Z.J Zhou et al., "A New BRB-ER Based Model for Assessing the Life of Product Using Data Under Various Environments," *IEEE Trans. Syst., Man Cybern.: Syst.*, Nov. 2016, vol. 46, no. 11, pp. 1529–1543.

[16] Z.G Zhou et al., "A Bi-Level Belief Rule Based Decision Support System for Diagnosis of Lymph Node Metastasis in Gastric caNcer," *Knowl-Based Syst.*, vol. 54, Dec. 2013, pp. 128–136.

[17] Y.W. Chen et al., "Identification of Uncertain Nonlinear Systems: Constructing Belief Rule-Based Models," *Knowl-Based Syst.*, vol. 73, Jan. 2015, pp. 124–133.

[18] G. Li et al., "A New Safety Assessment Model for Complex System Based on the Conditional Generalized Minimum Variance and the Belief Rule Base," *Safety Sci.*, vol. 93, Mar. 2017, pp.108–120.

[19] J.B. Yang and D.L. Xu, "Introduction to the ER Rule for Evidence Combination," in *Lecture Notes in Computer Science*, vol. 7027, Springer, 2011, pp. 7–15.

[20] N. Hansen, "The CMA Evolution Strategy: a Comparing Review," In *Advances on Estimation of Distribution Algorithms*, vol. 192, Springer, 2006, pp. 75–102.

[21] N. Hansen and S. Kern, "Evaluating the CMA Evolution Strategy on Multimodal Test Functions," In *Parallel Problem Solving from Nature - PPSN VIII*, Springer, 2004, pp. 282–291.

[22] N. Hansen, S.D. Müller, and P. Koumoutsakos, "Reducing the Time Complexity of the Deran-Domized Evolution Strategy with Covariance Matrix Adaptation (CMA-ES)," *Evolutionary Comput*, vol. 11, no. 1, Mar. 2003, pp. 1–18.

[23] A. Auger and N. Hansen, "Benchmarking the (1+1)-CMA-ES on the BBOB-2009 Function Tested," In *Proc. Genetic Evolutionary Comput. Conf.*, Montreal, Canada, July 8–12, 2009, pp. 2389–2396.

[24] K. Wang and J.S. Salvatore, "Anomalous Payload Based Network Intrusion Detection," In *Proc. Int. Symp. Recent Adv. Intrusion Detection*, Sophia Antipolis, France, Sept. 15–17, pp. 203–222.

[25] S.J. Stolfo, L. Wenke, and P.K. Chan, "Data Mining-Based Intrusion Detectors: An Overview of the Columbia IDS Project," *ACM SIGMOD Record*, vol. 30, no. 4, Dec. 2001, pp. 5–14.

[26] Z.J. Zhou et al., "Online Updating Belief-Rule-Base Using the RIMER Approach," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 41, no. 6, Nov. 2011, pp. 1225–1243.

[27] X. Xu and X.N. Wang, "An Adaptive Network Intrusion Detection Method Based on PCA and support Vector Machines," In *Adv. Data Mining Applicat., Second Int. Conf.*, ADMA 2006, China, 2006, pp. 696–703.

[28] Z.G. Liu et al., "Hybrid Classification System for Uncertain Data," *IEEE Trans. Syst., Man, Cybern.: Syst.*, no. 99, Nov. 2016, pp. 1–8.

[29] Z.G. Liu et al., "Credal c-means Clustering Method Based on Belief Functions," *Knowl.-Based Syst.*, vol. 74, Jan. 2015, pp. 119–132.

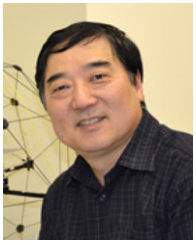
[30] J.B. Jian, "A Superlinearly and Quadratically Convergent SQP Type Feasible Method for Constrained Optimization," *Appl. Math. J. Chinese Univ. (B)*, vol. 15, 2000, pp. 319–332.

[31] S. Das and P.N. Suganthan, "Differential Evolution: A Survey of the State-of-the-Art," *IEEE Trans. Evolut. Comput*, vol. 15, 2011, pp. 4–31.

Bang-Cheng Zhang received his B.Eng. and M.Eng. from the Changchun University of Technology, China, in 1995 and 2004, respectively, and his Ph.D. in electrical engineering and automation from Jilin University, Changchun, China, in 2011. He is currently a professor at Changchun University of Technology. He also served as an academic visitor at Tsinghua University in Beijing, China, in 2007. He has published 15 articles. His research interests include mechatronics measurement techniques, fault diagnosis, and complex systems' modeling.

Guan-Yu Hu received his B.Eng. from the Harbin University of Science And Technology in 2005 in Harbin, China, his M.Eng. from Changchun University of Technology in 2010 in Changchun, China, and his Ph.D. from the Harbin University of Science And Technology in 2005. He is an associate professor at Hainan Normal University. He has published 15 articles. His research interests include intelligent computing, optimization algorithm, network security, and complex systems' modeling.

Zhi-Jie Zhou received his B.Eng and M. Eng from the High-Tech Institute of Xi'an, China in 2001 and 2004, respectively, and his Ph.D. from Tsinghua University, Beijing, in 2010. He is an associate professor at High-Tech Institute of Xi'an. He was a visiting scholar at the University of Manchester from March to August, 2009. He has published 70 articles. His research interests include belief rule base, dynamic system modeling, hybrid quantitative and qualitative decision modelling, and fault prognosis and optimal maintenance of dynamic systems.



You-Min Zhang received his B.S., M.S., and Ph.D. with a specialization in automatic controls from Northwestern Polytechnical University, Xi'an, in 1983, 1986, and 1995. He is a professor at the Department of Mechanical and Industrial Engineering and the Concordia Institute of Aerospace Design and Innovation, Faculty of Engineering and Computer Science, Concordia University, Montreal, Canada. He is also a guest professor at Xi'an University of Technology under the Shaanxi Province "100 Talents Plan." He has published four books, over 380 journal and conference papers (including 110 refereed journal papers since 1992), and book chapters. Dr. Zhang is a senior member of AIAA and IEEE, and a member of AUVSI/USC, CASI, and CSME.

Pei-Li Qiao graduated from Fudan University in 1974. He has been a professor at Harbin University of Science And Technology since 2005, Harbin, China. He has published five books, and over 70 articles in the last five years. He was managing director of the Computer Society in Heilongjiang, China. His research interests include network and information security, production scheduling and intelligent algorithms, database applications, enterprise information systems, and software engineering.

Lei-Lei Chang received his M.S. from the National University of Defense Technology. He is currently a lecturer at the High-tech Institute of Xi'an. His research interests include belief rule base, evidential reasoning, and complex systems' modeling.

https://doi.org/10.4218/etrij.17.0116.0305