IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received February 20, 2019, accepted March 8, 2019, date of publication March 13, 2019, date of current version April 1, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2904600

Health Evaluation of MVB Based on SVDD and Sample Reduction

ZHAOZHAO LI, LIDE WANG, YUEYI YANG, XIAOMIN DU, AND HUI SONG
School of Electrical Engineering, Beijing Jiaotong University, Beijing 100044, China

Corresponding author: Zhaozhao Li (14117379@bjtu.edu.cn)

This work was supported in part by the Fundamental Research Funds for the Central Universities under Grant 2017YJS181, and in part by the Beijing Municipal Natural Science Foundation under Grant L171009.

ABSTRACT Multifunction vehicle bus (MVB) is the most widely used train communication network whose performance degradation and anomaly will heavily affect the train's safe and stable operation. However, current scheduled maintenance and post-failure maintenance of MVB cannot detect the early anomaly and evaluate the health condition of the network in time. This paper provides a method to detect the anomaly and evaluate the health condition of MVB based on a one-class classification (OCC) algorithm called density-based sample reduction for support vector data description (DBSRSVDD). First, network features are extracted from physical layer waveform parameters. In order to reduce the computational complexity of SVDD, a sample reduction operation is conducted to screen out the edge samples as support vector candidates. Then, the SVDD models representing the normal patterns of a single MVB node are trained based on the support vector candidates. Performance degradation of the node is quantified by the distance between the tested sample and the trained hyper sphere. The whole network's health condition is the linear weighted sum of the nodes' scores based on their bandwidth occupancy. The experimental results show that the proposed method can detect the anomaly and degradation of MVB successfully, improve accuracy, and reduce training time compared with the existing methods.

INDEX TERMS Support vector domain description, sample reduction, health evaluation, anomaly detection, multifunction vehicle bus.

I. INTRODUCTION

MVB is the central nervous of the train which transmits control data and monitoring data. Its health condition is closely related with the train's safe and stable operation. However, current means of MVB maintenance remain in simple schedule maintenance and post-failure maintenance, leaving the incipient fault undetectable until it develops into major failure and causes an emergency train stop. Therefore, the anomaly detection and health evaluation of MVB are of great importance to eliminate potential risk.

The approach of MVB anomaly detection and health evaluation is to train a model representing the normal behaviors of MVB and use the model to identify anomalies. Feature extraction and health evaluation method are the key issues to address. For the first issue, a MVB protocol analyzer is utilized in our work to extract features from physical waveforms of MVB which characterize the network status. For the second issue, performance deterioration and health evaluation are conducted to find and quantize patterns that do not conform to the expected behaviors. Common methods of anomaly detection and health evaluation include classification-based techniques, cluster-based techniques, statistical techniques.

Classification-based techniques operate in two steps. In the first step a classifier is learned based on both positive class data and negative class data. In the second step, tested instances are classified into normal class or anomalous class by the trained classifier. Wen and Meng [1] have applied Support Vector Machine (SVM) trained by both normal and abnormal data to quantify the health condition of ethernet. Features are extracted from the quality of service (QOS) indicators of ethernet. Xie *et al.* [2] has presented a fast support vector data description (SVDD) algorithm to detect outliers in wireless sensor networks. Rätsch *et al.* [18] have presented an anomaly detection method called One-Class

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

Support Vector Machine (OCSVM) by separate the origin from normal data with a maximum distance classification hyper plane. Replicator Neural Networks method has been introduced in detecting anomaly by Hawkins *et al.* [3] and Williams *et al.* [40]. Wang *et al.* [16] have introduced the hierarchical temporal memory (HTM) network to detect anomaly of In-vehicle network.

Clustering algorithms are used to group similar data patterns into one cluster [22], [23]. The principle of cluster based anomaly detection is declaring any data instance that does not belong to any cluster as an outlier. Clustering algorithms which will not force every data to be grouped into a cluster such as DBSCAN [24], Robust Clustering using Links (ROCK) [25], Shared Nearest Neighbor (SNN) [26] can be used in anomaly detection. The distance between the tested data and its closest cluster centroid is calculated as its anomaly score [5]–[7]. Self-Organizing Maps (SOM) algorithm has been also introduced into anomaly detection and fault diagnostic [27]–[29]. Wang *et al.* [41] have presented a Service Feature Clustering method in telecommunication network performance anomaly detection.

Statistical techniques assume that the normal samples distribute in high probability region of stochastic model while anomalies lie in the low probability regions of stochastic model. Boxplot is the simplest method in anomaly detection. Solberg and Lahti [10] have applied boxplot method in digital health and Guttormsson *et al.* [11] have introduced it in steam turbine rotor maintenance. Jia *et al.* [4] have introduced Gaussian Mixture Model to fit the power curve of wind turbine based on its power and wind speed data. The deviation between current condition and normal condition is used to evaluate the degradation of the wind turbine.

The mentioned methods above have their advantages and disadvantages respectively. Statistical techniques need to assume that abnormal samples are located in high probability regions of stochastic model while anomalies lie in the low probability regions of stochastic model. Large amount of normal and abnormal samples are indispensable in probabilistic model estimation. The lack of negative samples in practical use limits its application. Clustering algorithms' effectiveness for anomaly detection depends on the density distribution in feature space. The Classification-based techniques test fast in the testing phase since they use learned models for classification but both labeled normal and abnormal data is essential in the training phase. Normal data is sampled easily in practical use while sampling abnormal data is often costly, resulting in the training data set unbalanced and class skew problem.

One-Class classification (OCC) algorithms need only one class data and can avoid this problem. SVDD is one of the most popular OCC algorithms and it aims at training a minimum hyper sphere which contains most of the samples in the feature space [12]. By calculating the Euclidean distance between the tested sample and hyper sphere centroid, the health condition of a system is quantified. However, model training of SVDD is a very slow process when the sample size is large, since the quadratic programming (QP) problem implies high training time complexity $O(n^3)$ and space complexity $O(n^2)$. Besides, parameter optimization of SVDD will also consumes large amount of time in cross validation process. Therefore, speeding up the model training of SVDD is a notable and significant issue. Dong *et al.* [31] and Platt [32] have divided the original QP problem into small pieces to reduce the size of the whole QP problem. In their work the memory requirement in huge data set has been largely reduced while the training time is still long. Another group of approaches are scaling down the training set by selecting support vector candidates. Because hyper sphere of SVDD only relies on a fraction of samples called support vectors, non-support vectors can be removed without affecting the classification accuracy. Hence, sample reduction methods are proved to be an effective way for SVDD to solve large scale classification problems. Liu *et al.* [13] have presented a geometry-based method and built a reduction sphere to remove the inner samples. Since the data set is decreasing and changing, the reduction sphere is also moving in the feature space and finally the edge points are left behind. However, it only works on convex set and fails to solve irregularly distributed set. Information entropy has been introduced to present the similarity between samples by Li *et al.* [14]. Samples distributed in edge region usually have higher entropy than the inner ones since they are distributed sparsely and less similar to each other according to their definition of entropy. The samples in original data set whose entropy are greater than a threshold θ are selected as support vector candidates and put into the training set of SVDD. And θ is adjusted by a parameter v. The bigger v is, the less θ is and the more support vector candidates there are. But the proposed information entropy based sample reduction for SVDD (IESRSVDD) method takes a lot of time to calculate the entropy values of all samples. Sample reduction has been applied to speed up SVM in binary classification without reducing its accuracy by Li *et al.* [15]. The sample whose neighborhood contains both positive and negative data is identified as edge sample in their work. However, in health evaluation and anomaly detection problem there is only one class data.

The local density of SVDD support vectors is sparse since they are located in edge areas. Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is sensitive to the local density of the data set and can screen out the noisy data at the same time [24]. Inspired by this advantage of DBSCAN we can relax the screening criteria and the "noisy data" distributed in edge areas can be selected as the support vector candidates. In this paper we call the proposed algorithm DBSRSVDD. MVB health evaluation and anomaly detection have three steps. Firstly, a model is trained by the support vector candidates to represent the normal condition of nodes. Secondly, samples to be tested are classified by the model and the health conditions of MVB nodes are quantified by the distances between the hyper sphere centroid and them. Finally, the health condition of whole MVB network is calculated as the weighted linear sum of the node scores.

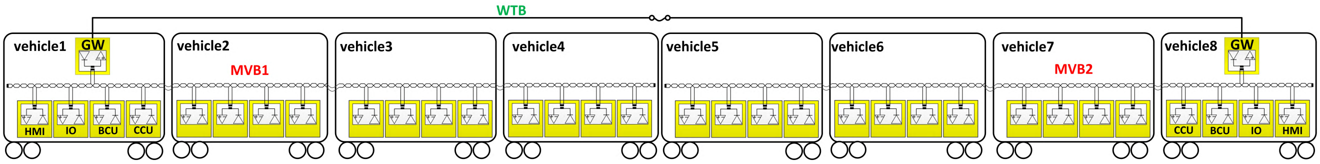

FIGURE 1. Typical application of multifunction vehicle bus (MVB).

We have made the following contributions:

1) A feature extraction method has been proposed which samples data from physical layer waveforms to represent the operating condition of MVB. Most current researches on control network focus on improvement of its real-time capability. Few studies on network health evaluation issues have been carried by academe.
2) Unlike some other evaluation methods, the proposed evaluation method is based on trained models without any human subjective factors of experts. The models are trained from objective data. Abnormal condition will be detected once the monitoring data deviates from the models. Health condition of MVB is measured at both device level and network level. Performance degradation is quantified by the distance between the hyper sphere centroid and the monitoring data. The whole network's health condition is the linear weighted sum of the devices' scores according to the bandwidth occupancy of the devices. By this way, the maintenance method of MVB can be improved to condition based maintenance (CBM) based on the evaluation result and potential risks can be eliminated in the bud.
3) In order to accelerate model training and reduce computational complexity, a new sample reduction method improved from DBSCAN is proposed. When v is chosen properly, the proposed method can significantly reduce sample size and computational complexity. DBSRSVDD takes much less time and space in model training than IESRSVDD and SVDD and maintains the classifier performance at the same time. Parameter optimization process of SVDD can also be eliminated in DBSRSVDD.

The rest of the paper is organized as follow. Section 2 introduces some basic knowledge of MVB and Section 3 explains the features extraction process. Our proposed sample reduction method is elaborated in Section 4. Section 5 gives an introduction of SVDD evaluation process. Experiments and results are presented and discussed in Section 6. Finally, conclusion and further work are given in Section 7.

II. MVB FRAMEWORK AND HEALTH EVALUATION SYSTEM

In modern electrical train, there are many control devices and supervising devices in each vehicle, including Central Control Units (CCU), Brake Control Units (BCU), Human Machine Interfaces (HMI), Remote Input Output Modules (IO) and so on. All of them are integrated to a networked control system by control network. Multifunction Vehicle Bus (MVB) is the vehicle-level control network and can be reckoned as the subnet of the train-level network which is called the Wired Train Bus (WTB). MVB is a fixed topology network while WTB can be changed dynamically according to the demand. Through the connection and disconnection of WTB, the train's length can be adjusted in a flexible way. Figure 1 gives a typical example of the train communication network. MVB1 connects the devices in vehicle1 to vehicle4 and MVB2 connects the devices in vehicle5 to vehicle8. WTB integrates the two subnets to a whole through gateways (GW).

From the perspective of controllability, MVB is a multi-agent network and can be divided into some subnetworks in terms of time scales. Different control tasks have different time scales according to their priority. Long *et al.* [9], [19] have studied the group controllability of multi-agent networks with two-time-scale features and some easy-to-use criteria have been proposed for them. From the perspective of communication, MVB is a time division multiplexing (TDM) mode master-slave network [35]. Some researches have been done on the real-time features of MVB to improve the control performance of the whole networked system by Yan *et al.* [20] and Wang *et al.* [21]. Synchronization is another issue about the train control network. However, almost all of the current investigations on networked systems are interested in the nodes of networked systems. Su *et al.* [8], [17] have focused their attention on edge synchronization of networked system. They have used Line graph to represent the communication topology of the edges and obtained the feedback gain matrix and observer matrix by resolving linear programming problems. The health condition of MVB will heavily affect the controllability of the train networked system. In this paper, we mainly focus on the health evaluation method of MVB to ensure the stability and safety of the train control system. The physical layer electrical connection of MVB is shown in Fig.2. Baud rate of MVB is 1.5 M/bps and its frame is encoded as Manchester code between 1.5V and 5V. The logic states are defined as rising edge ('0') and falling edge ('1'). Fig.3 shows the typical MVB frame format which consists of start delimiter, frame data and end delimiter. In a MVB frame, the format of start delimiter and end delimiter is fixed.

The main task of MVB is transmitting the frame data correctly in the harsh and complex operating environment. MVB failure causes vary from situation to situation. Typical MVB failures include communication latency, packet loss, node shutdown fault and so on. The main causes of these failures are electromagnetic interference, impedance mismatch, cable terminal fault, degradation of medium. According to the

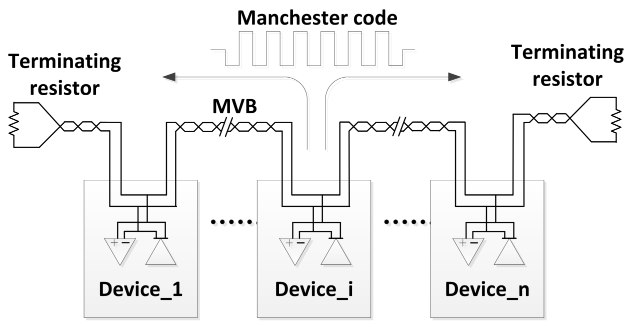

FIGURE 2. MVB electrical connection.

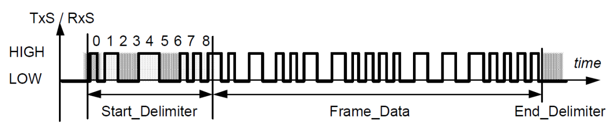

FIGURE 3. MVB frame format.

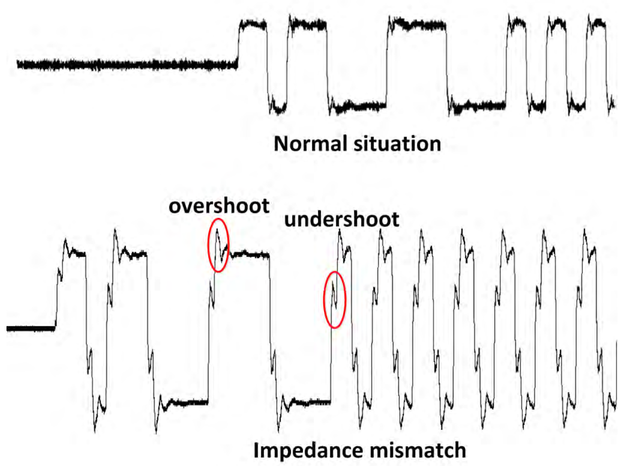

FIGURE 4. Waveform distortion.

industrial network experts, the majority of the failure modes commonly occurring on MVB are due to one or more the following factors:

Impedance mismatch. It is suggested in IEC 61375-1 that characteristic impedance of the cable is 120 Ω and each end of the network needs to be terminated with a 120 Ω resistor [35]. Missing terminating resistors caused by inappropriate maintenance or contact resistance increasing of cable connectors may lead to this kind of failure. Impedance mismatch will lead to channel reflection and physical waveform distortion. As a consequence, bit error rate and packet loss rate dramatically increase. Besides, impedance mismatch will make MVB vulnerable to electromagnetic interference. Waveform distortion caused by impedance mismatch is shown in Fig.4. The overshoot may breakdown electronic components on the network card. The undershoot may cause decoding error and result in the performance degradation of MVB.

Degradation of medium. The network performance will deteriorate with the degradation of medium. Network cables and cable connectors can be considered as durable components in the vibrating, humid, low or high temperature train operating environment. Hostile environment and improper maintenance can change the electronic properties of the cables and connectors. For example, the contact resistance will increase by vibration and corrosion. The broken cable shield will also make the network vulnerable to electromagnetic interference.

However, current situation is that all these problems cannot be detected and maintained in time until they deteriorate into significant failures. Therefore, we have presented a health evaluation and anomaly detection method of MVB based on physical waveform features. The proposed health evaluation system is illustrated in Fig.5.

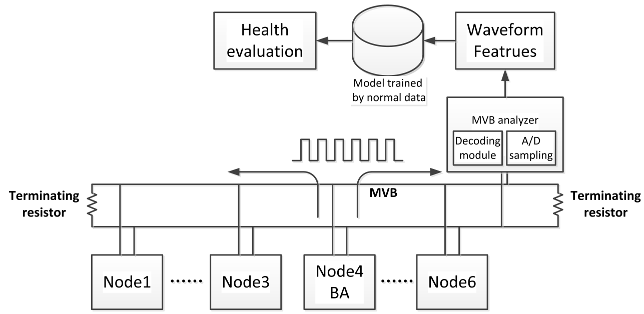

FIGURE 5. Illustration of health evaluation system.

III. FEATURE EXTRACTION

Feature extraction is the base of MVB health evaluation. Reasonably extracted features can effectively represent the state of the network. Failures mentioned above will affect MVB physical waveforms notably. Based on the high speed analog signal acquisition of MVB analyzer, waveform parameters are extracted from each bit of the start delimiter and the network features are the averages of these parameters. These physical waveform features are shown as follow:

Steady-state positive voltage Vp: the signal level when waveforms reach the positive steady state.

Steady-state negative voltage Vn: the signal level when waveforms reach the negative steady state.

Overshoot Vos: the difference between the maximum voltage and steady-state voltage.

Slew rate S: the slope of the waveform at zero crossing point.

Rising slope Kp: the slope when the frame signal rises from 10 percent of Vp to 90 percent of Vp.

Falling slope Kn: the slope when the frame signal descends from 10 percent of Vn to 90 percent of Vn.

Feature extraction process is shown in Fig.6.

IV. DBSCAN AND SAMPLE REDUCTION

A. DBSCAN

DBSCAN views clusters as areas of high density separated by areas of low density. There are two parameters in DBSCAN, *MinPts* and ϵ, which define the density of the data set in feature space. ϵ is the radius of the defined neighborhood. When the sample size in ϵ neighborhood of x_j is no less than *MinPts*,

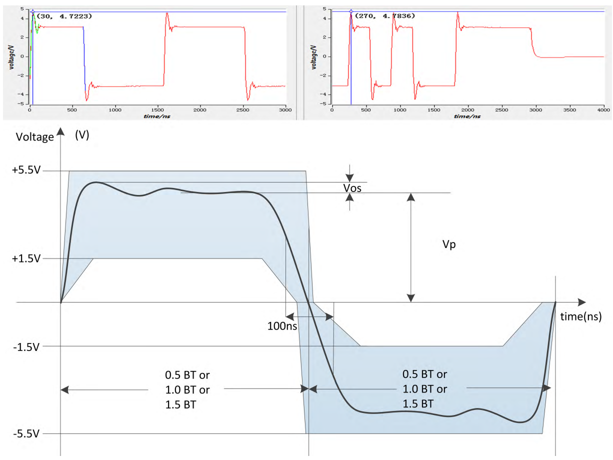

FIGURE 6. **Feature extraction.**

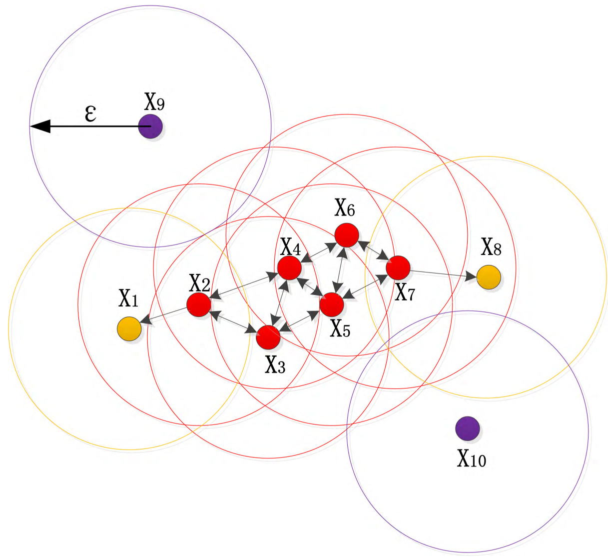

FIGURE 7. **DBSCAN.**

we call x_j a core sample. Large *MinPts* and small ϵ indicate that the density of samples should be high to form a cluster. A cluster in DBSCAN consists of core samples and non-core samples close to the core samples. The formation of clusters in DBSCAN is related to the connectivity between samples. Given a data set $D = \{x_1, x_2, x_3 \ldots, x_j \ldots, x_n\}$, the main concepts of connectivity in DBSCAN are defined as follow:

Directly density-reachable: If x_j lies in the ϵ neighborhood of x_i and x_i is a core sample, x_j is directly density-reachable from x_i.

Density-reachable: If there is a sample sequence $x_i, x_{i+1}, x_{i+2} \ldots, x_{i+k} \ldots, x_{m-1}, x_m$ and x_{i+k} is directly density-reachable from x_{i+k+1}, we call that x_i is density-reachable from x_m.

Density-connected: If there is a sample x_k both density-reachable from x_i and x_j, we call x_i is density-connected from x_j.

As shown in Fig.7, $MinPts = 3$, $x_2, x_3, x_4, x_5, x_6, x_7$ are core samples. x_1, x_3, x_4 are directly density-reachable from x_2. x_1 is density-reachable from x_3 and x_8. x_9, x_{10} are noisy points.

B. SAMPLE REDUCTION BASED ON DBSCAN

A sample in sparse region that belongs to no cluster in DBSCAN is identified as a noisy point. For OCC problem, the samples located in margin area are likely support vectors and their local densities are usually sparse. Inspired by the noise elimination function of DBSCAN, the support vector candidates in margin area may be screened out as ''noisy data'' if the noisy data identification criterion of the algorithm are relaxed. As a consequence, a reduced subset consist of support vector candidates can be used to speed up training the SVDD model in network health evaluation. *MinPts* can be defined as fixed values such as 5,6,7,8,9,10. ϵ is defined as follow:

$$\epsilon = v\left(\frac{1}{n^2}\sum_{i,j=1}^{n} dist(x_i, x_j)\right) \tag{1}$$

n is the sample size of the training data set and $dist(x_i, x_j)$ is the distance between x_i and x_j in feature space. ϵ is proportional to the mean distance between the samples. Coefficient v can be adjusted to change the noisy data identification criterion and the size of support vector candidates. The larger v is, the less samples will be selected into subset since the ϵ neighborhood is larger so that we can easily find at least *MinPts* other samples in it. The density-based sample reduction (DBSR) process can be summarized as the following 4 steps. A synthetic banana-shape data set is used to verify the effectiveness of the proposed sample reduction algorithm as shown in Fig.8.

1) Initialize the subset X with $\{\varnothing\}$, *MinPts* and v with a fixed value. Adjustment of parameter v can change the subset size. As a rule of thumb, its value between 0.5 and 1 is preferable.
2) Calculate ϵ according to (1).
3) Based on parameters ϵ and *MinPts*, screen out the samples in sparse regions of the original data set. For x_j, the sample size in its ϵ neighborhood is defined as $N_\epsilon(x_j) = \{x_i \in D | dist(x_i, x_j) \leq \epsilon\}$. If $N_\epsilon(x_j) < MinPts$, x_j is labeled as a support vector candidate and put into X. If $N_\epsilon(x_j) \geq MinPts$, x_j and all samples which are directly density-reachable, density-reachable, density-connected from x_j are labeled as inner points.
4) Algorithm stops when all samples in original data set are labeled as inner points or support vector candidates.

V. SVDD AND NETWORK HEALTH EVALUATION

A. SVDD

SVDD is a typical OCC algorithm. It aims at obtaining a minimum hyper sphere which contains most data samples in the feature space. The objective function of SVDD is defined as:

$$F(R, a) = R^2 + C\sum_{i=1}^{n}\xi_i$$
$$s.t. \ \|\Phi(x_i) - a\|^2 \leq R^2 + \xi_i, \quad \xi_i \geq 0, \ \forall i \tag{2}$$

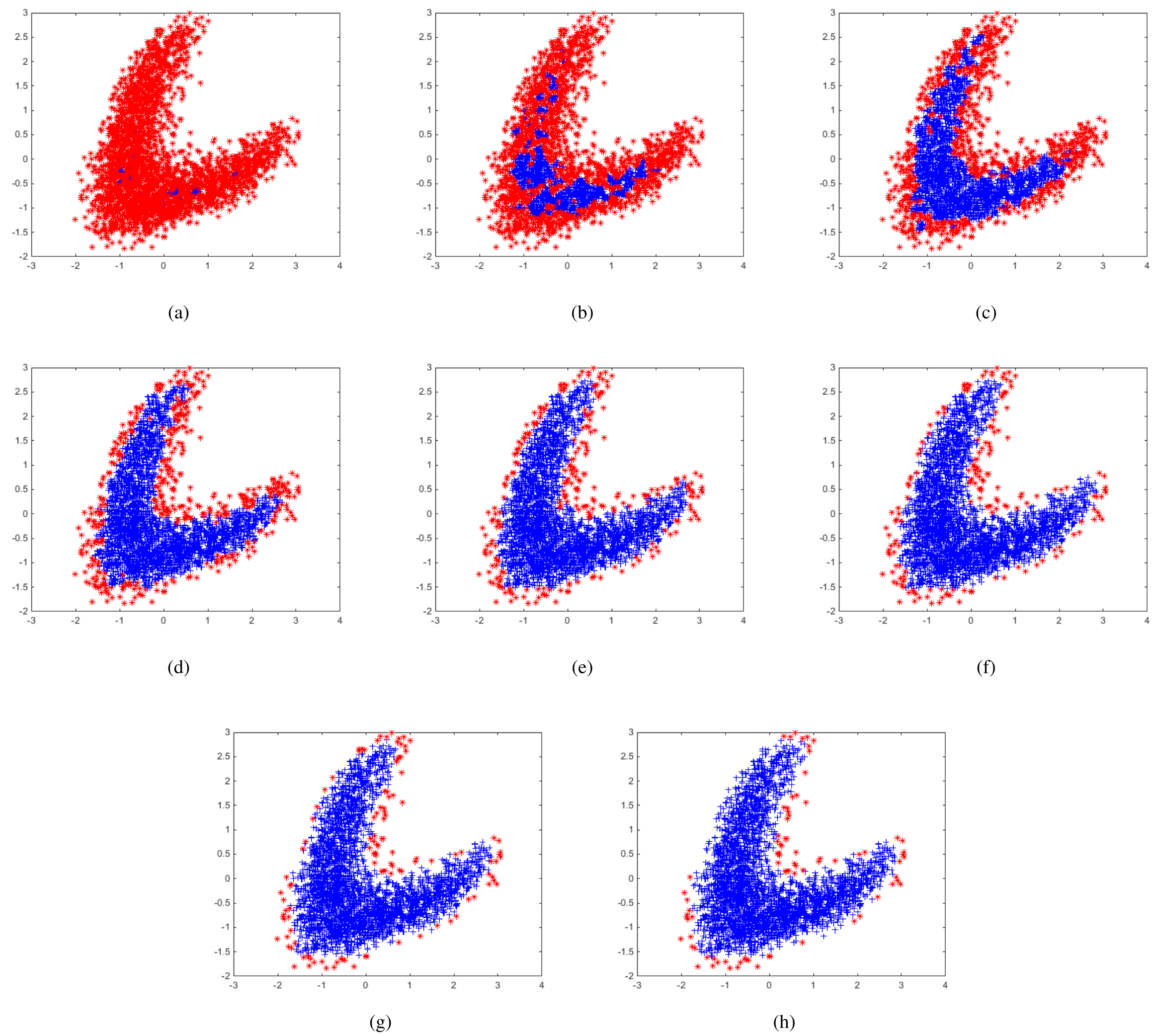

FIGURE 8. **The sample reduction with different v of DBSR in banana data set(red * stands for support vector candidates and blue + stands for inner points). (a) v = 0.3. (b) v = 0.4. (c) v = 0.5. (d) v = 0.6. (e) v = 0.7. (f) v = 0.8. (g) v = 0.9. (h) v = 1.0.**

R is the radius of the SVDD hyper sphere and a is the center. ξ_i is the slack variable. C is the penalty factor. Φ maps x_i to high-dimensional space. As the same with SVM, Lagrange multiplier method can be used to solve (2).

$$\begin{aligned} &L(R, a, \alpha_i, \gamma_i, \xi_i) \\ &\quad = R^2 + C\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\gamma_i\xi_i - \sum_{i=1}^{n}\alpha_i\{R^2 + \xi_i \\ &\qquad - (\|\Phi(x_i)\|^2 - 2a\Phi(x_i) + \|a\|^2)\} \end{aligned} \tag{3}$$

α, γ are Lagrange multipliers and $\alpha, \gamma \geq 0$. Solve the Lagrange equation and make the partial differential of R,a,ξ_i we can get (4).

$$\begin{aligned} &\frac{\partial L}{\partial R} = 0 : \sum_{i=1}^{n}\alpha_i = 1 \\ &\frac{\partial L}{\partial a} = 0 : a = \frac{\sum_{i=1}^{n}\alpha_i\Phi(x_i)}{\sum_{i=1}^{n}\alpha_i} = \sum_{i=1}^{n}\alpha_i\Phi(x_i) \\ &\frac{\partial L}{\partial \xi_i} = 0 : C - \alpha_i - \gamma_i = 0 \end{aligned} \tag{4}$$

Substitute (4) into (3):

$$\begin{aligned} L = \sum_{i=1}^{n}\alpha_i K(x_i, x_j) &- \sum_{i,j=1}^{n}\alpha_i\alpha_j K(x_i, x_j) \\ s.t.\ 0 \leq \alpha_i \geq C, &\quad \forall i \\ \sum_{i=1}^{n}\alpha_i = 1 & \end{aligned} \tag{5}$$

$K(x_i, x_j)$ is the kernel function equal to the inner product of $\Phi(x_i)$ and $\Phi(x_j)$. When x_i satisfies the inequality $\|\Phi(x_i) - a\|^2 < R^2 + \xi_i$, the constraint is satisfied and the corresponding Lagrange multiplier will be zero($\alpha_i = 0$). If a data satisfies the equality $\|\Phi(x_i) - a\|^2 = R^2 + \xi_i$,its Lagrange multiplier will be unequal to zero($\alpha_i > 0$).Thus, we have the following conditions.

$$\|\Phi(x_i)-a\|^2 < R^2+\xi_i, \qquad \alpha_i=0,\ \gamma_i=C,\quad \xi_i = 0 \tag{6}$$
$$\|\Phi(x_i)-a\|^2 = R^2+\xi_i, \quad 0<\alpha_i < C,\ \gamma_i=0,\ \xi_i = 0 \tag{7}$$
$$\|\Phi(x_i)-a\|^2 > R^2+\xi_i, \qquad \alpha_i=C, \gamma_i > 0,\quad \xi_i > 0 \tag{8}$$

(4) shows that the center of the SVDD hyper sphere is a linear combination of the objects. Only samples x_i whose α_i is larger than zero, namely the support vectors, have an effect on SVDD hyper sphere. That's the theoretical foundation of the sample reduction in SVDD. By definition, R is the distance from the center of the hyper sphere a to the hyper sphere boundary and it can be calculated with any support vectors x_k that satisfies (7).

$$R^2 = K(x_k, x_k) - 2\sum_{i=1}^{n}\alpha_i K(x_i, x_k) + \sum_{i,j=1}^{n}\alpha_i\alpha_j K(x_i, x_j) \quad (9)$$

If the decision function is satisfied in (10), the new tested sample x_z will be accepted.

$$\|\Phi(x_z) - a\|^2 = K(x_z, x_z) - 2\sum_{i=1}^{n}\alpha_i K(x_i, x_z) + \sum_{i,j=1}^{n}\alpha_i\alpha_j K(x_i, x_j) \le R^2 \quad (10)$$

The used kernel function in health evaluation is Gaussian kernel function and the parameter of kernel function is defined as (12). $C = 1/n$. n is the size of the training set [14], [38].

$$K(x_i, x_j) = e^{-\gamma\|x_i - x_j\|^2} \quad (11)$$

$$\gamma = 1/\frac{1}{n^2}\sum_{i,j=1}^{n}\|x_i - x_j\|^2. \quad (12)$$

B. HEALTH EVALUATION BASED ON SVDD

In industrial application, pauta criterion has been adopted to detect abnormal value. When the tested data is more than 3 times of the standard deviation, it is considered as an outlier. According to the pauta criterion, we define the health evaluation function as (13).

$$score(x_z) = \begin{cases} 100, d \le R \\ 100 - 40\dfrac{d - \bar{d}}{3\sigma}, & R < d < \bar{d} + 3\sigma \\ 60\dfrac{\bar{d} + 3\sigma}{d}, & d \ge \bar{d} + 3\sigma \end{cases} \quad (13)$$

d is the distance between the tested sample and the center a in the high dimensional feature space. $\bar{d}$ is the mean value of d. σ is the standard deviation of d. $\bar{d} = \|\Phi(\frac{\sum_{i=1}^{n} x_i}{n}) - a\|$, $d = \|\Phi(x_z) - a\|$.

VI. EXPERIMENTS

A. DBSRSVDD

The MVB network health evaluation process based on DBSRSVDD is as follow.

1) Physical waveforms of each node are sampled in normal situation.
2) Extract features from physical layer waveforms and form the training set X. Principal component analysis (PCA) algorithm is conducted to reduce dimension of the training samples.

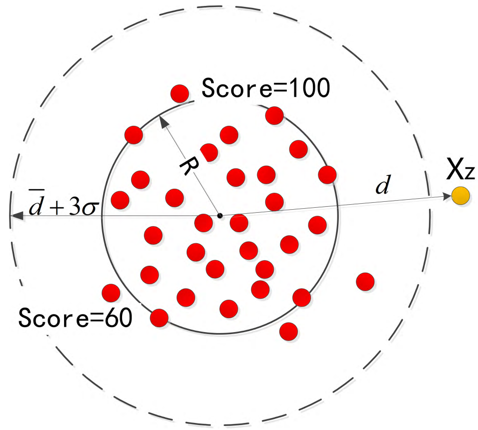

FIGURE 9. Evaluation function.

3) Initialize the parameter v and $MinPts$.
4) Calculate the neighborhood parameter ϵ according to (1).
5) Conduct the sample reduction algorithm on X. The connectivity between samples is detected to screen out the sparse points as "noisy data". These so-called noisy points are support vector candidates and form the training subset X_s. In fact the true noisy points have been also put into X_s in this step.
6) There is a small amount of noisy data in it during the feature extraction process and sample reduction process. Hence, the training data must be cleaned after sample reduction operation. Otherwise, the trained hyper sphere will be seriously affected by the noisy data since the noisy data's proportion in the training set become larger after the sample reduction operation. The left figure of Fig.10 is the boxplot of the subset and the right figure shows the feature space before data cleaning. As is shown in Fig.11, the hyper sphere can't enclose the training data well because the data set is polluted by noisy data.
7) Train SVDD model based on subset X_s.
8) Collect waveforms data and extract features in operation situation. The node operation condition is quantized with the trained model and evaluated according to (13). Each MVB node is tested 20 times continuously and the average of them is reckoned as the evaluation result.
9) The health condition of MVB is the linear weighted sum of the node scores. Weight of each node is calculated based on their bandwidth occupancy.

B. EXPERIMENTAL SETTINGS

In order to verify the effectiveness of the proposed DBSRSVDD, experiments under laboratory conditions have been conducted on a test bed consisting of 6 MVB devices (5 slave devices and 1 master device) as illustrated in Fig.12. Fig.13 shows the experiments performed in normal condition and four kinds of fault conditions. The sample size of normal condition is 5000 for each network node and the sample size is 1000 in fault conditions.

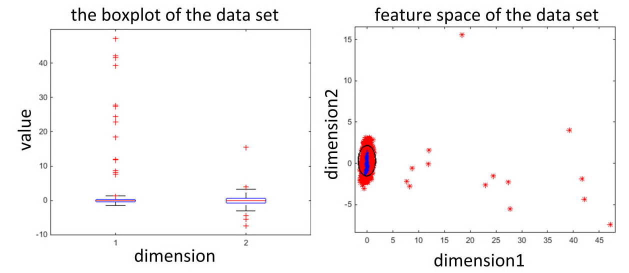

FIGURE 10. **The boxplot and feature space of the data set before data cleaning.**

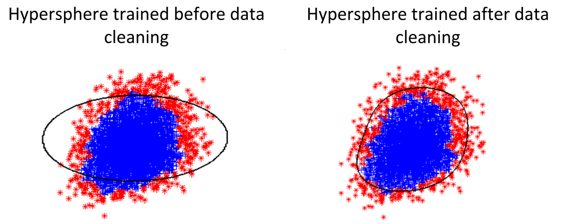

FIGURE 11. **Hypersphere.**

FIGURE 12. **Test bed in laboratory.**

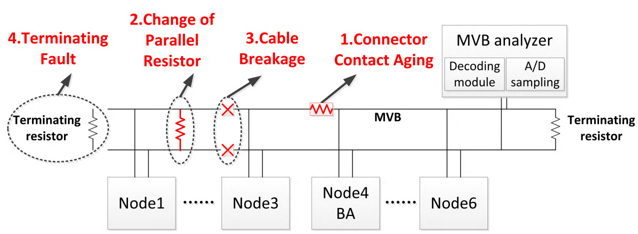

FIGURE 13. **Schematic layout of MVB health evaluation.**

1) Fault1:Series impedance increases when there is loose connection or cable connectors are aging.
2) Fault2:Parallel impedance increases when the cable or circuit components degrade.
3) Fault3:Cable disconnected. The vibration operation environment of the train make it vulnerable to disconnection.
4) Fault4:Terminating resistor mismatch. Improper maintenance and vibration operation environment may lead to the mismatch of terminating resistors.

The classification performance is measured with $G-mean = \sqrt{TPR * TNR}$ where TPR is the classification accuracy of positive class and TNR is the classification accuracy of negative class [14]. The higher G-mean is the better is the prediction and the hyper sphere. The training set consists of 80% of the samples randomly chosen without replacement from the normal data, while the rest of the normal data and the whole fault data are utilized for testing. SVDD and IESRSVDD methods are also conducted to compare with DBSRSVDD. Parameter optimization process of penalty factor C and kernel parameter γ will take large amount of time and slow down the training process. This contradicts our aim to accelerate training process. So all of the three methods are trained on fixed penalty factor C and kernel parameter γ according to equation (11)(12). Sample reduction and the trained hyper spheres of DBSRSVDD are shown in Fig.15($v = 0.8$ and the size of the reduced data set/the size of the original data set).

C. HEALTH EVALUATION OF SINGLE NODE

G-mean values, the numbers of the support vectors, training time of SVDD, IESRSVDD, DBSRSVDD are shown in Fig.17, Fig.18, Fig.19. In order to compare the performance of the proposed method with IESRSVDD, the sample reduction parameter v vary from 0.1 to 1. The trends of changing with parameter v are opposite between DBSRSVDD method and IESRSVDD method.

In terms of G-mean, DBSRSVDD achieves better performance compared with IESRSVDD and the conventional SVDD method except in Fault1 condition. The positive class samples overlap with the negative class samples in Fault1 condition because the performance of MVB decrease slightly. The density of negative class data in the overlapping region is larger than the one of positive class. DBSRSVDD and IESRSVDD have trained their models based on the edge samples so that their hyper spheres lie closer to the edge region than SVDD. When the hyper sphere encloses the overlapping region, the positive samples in this region are classified into positive class but there are more negative samples enclosed into hyper sphere at the same time. TPR increases slightly but TNR decreases significantly, resulting in the decrease of G-mean. Even so DBSRSVDD and IESRSVDD can sustain their accuracy and significantly decrease the computational complexity at the same time when the parameter v is selected properly.

As is shown in Fig14, hyper spheres of SVDD don't enclose all positive samples well on fixed parameter C and γ. Many positive samples of testing dataset are divided out of the hyper spheres, resulting in low TPR and G-mean. For example, TPR of SVDD is 51.96% and TNR is 100% in Fault4 leading to that G-mean of SVDD is about 70 %. When v is selected improperly in DBSRSVDD and IESRSVDD, samples are not reduced significantly and we will get bad hyper spheres as the same with SVDD. That's why we get low G-mean values in Fault2, Fault3, Fault4 even they are divided from normal data clearly. But if proper v is chosen, we can reduce samples effectively and the hyper spheres are well trained based on edge samples, as is shown in Fig15. Take Fault4 for example, the hyper spheres of DBSRSVDD

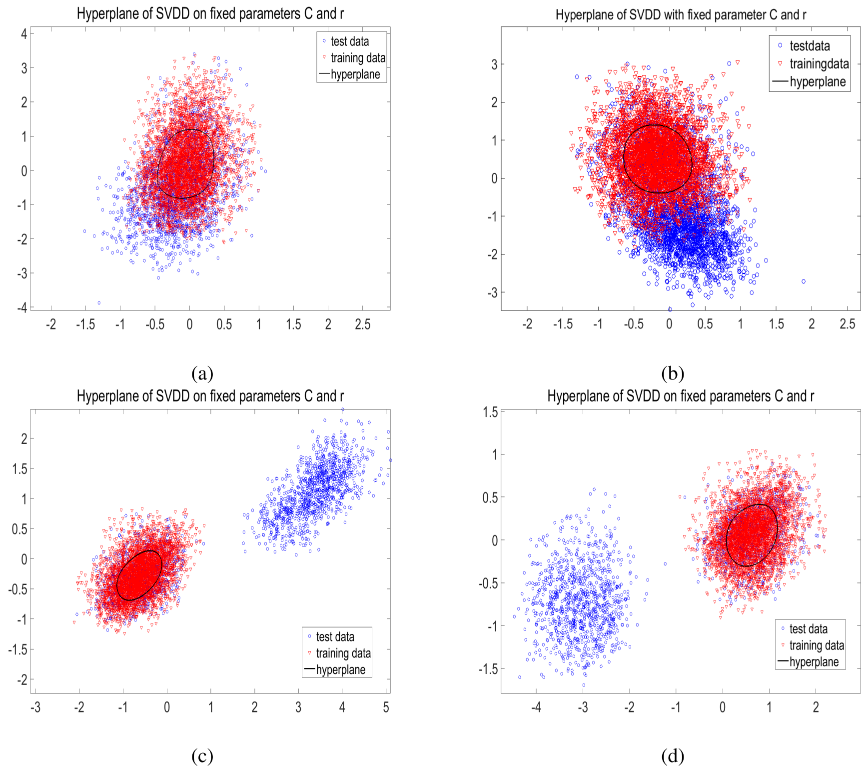

FIGURE 14. **Hyperplanes of SVDD with fixed parameter C and γ. (a) Fault1. (b) Fault2. (c) Fault3. (d) Fault4.**

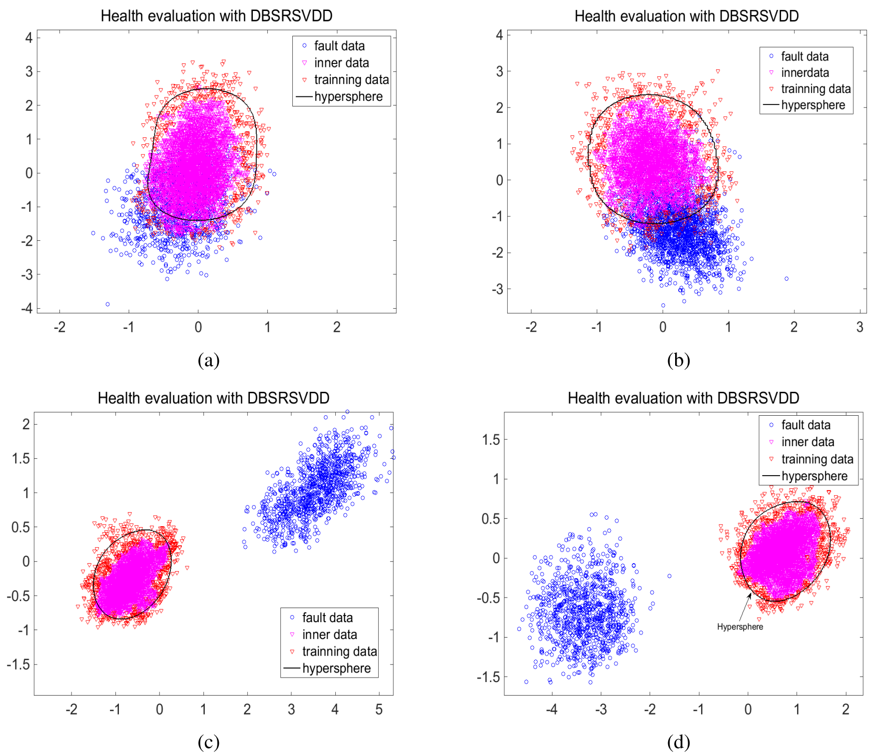

FIGURE 15. **Hyperplanes of DBSRSVDD with fixed parameter C and $\gamma(v = 0.8)$. (a) Fault1(496/5000). (b) Fault2(501/5000). (c) Fault3(532/5000). (d) Fault4(545/5000).**

is enlarging with the increase of v as shown in Fig16. Since most data samples are screened out as edge samples when v is less than 0.6, the hyper spheres don't enclose all normal data well, resulting in bad detection performance as shown in Fig17. The detection performance of DBSRSVDD gets better with the increase of v as well. When v is bigger than 0.8, G-means of Fault3 and Fault4 are almost 100% and we can get proper hyper spheres for network health evaluation.

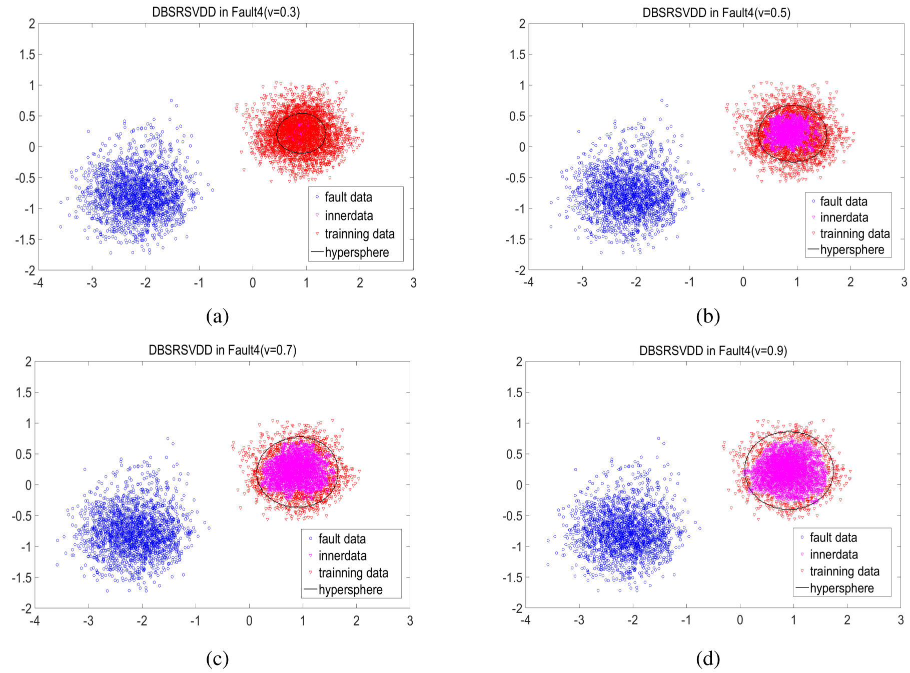

FIGURE 16. **Hyperplanes of DBSRSVDD enlarge with the increase of v. (a) Fault4. (b) Fault4. (c) Fault4. (d) Fault4.**

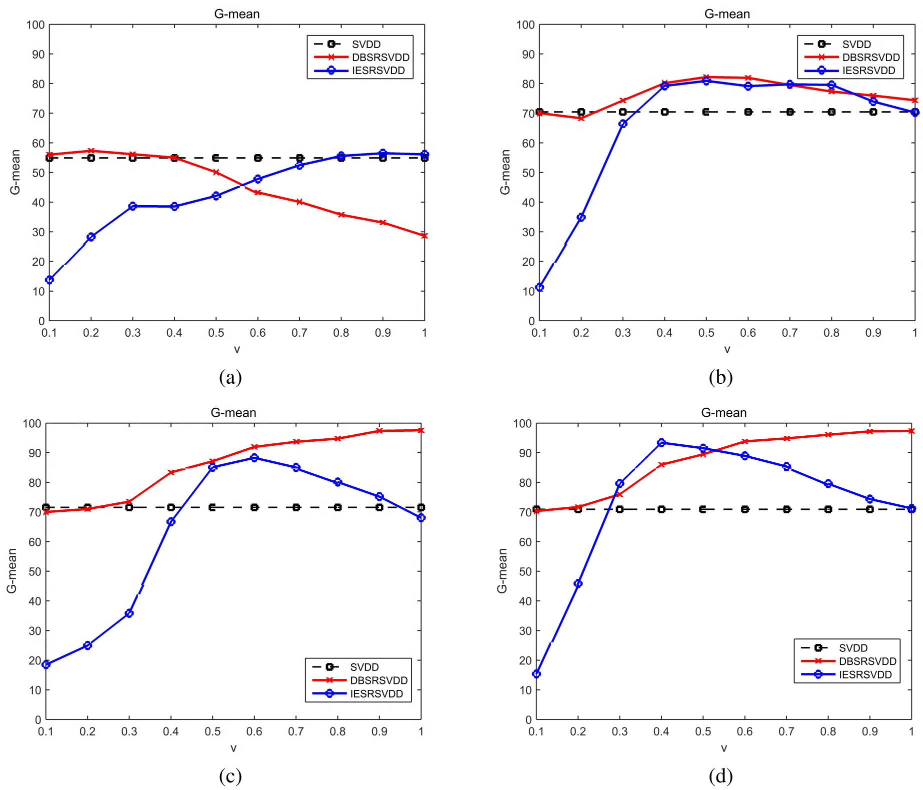

FIGURE 17. **G-mean of SVDD, IESRSVDD, DBSRSVDD in 4 fault conditions (v = 0.1-1.0). (a) Fault1. (b) Fault2. (c) Fault3. (d) Fault4.**

Both DBSRSVDD and IESRSVDD can effectively reduce the sample size and the support vector size as shown in Fig.18. However, from the perspective of training time, DBSRSVDD is much better than the conventional SVDD method and IESRSVDD. The value of time includes the sample reduction time and all the training and testing time. It takes at least 30 seconds to train the model with SVDD on fixed C and γ since the sample size have not been reduced. IESRSVDD

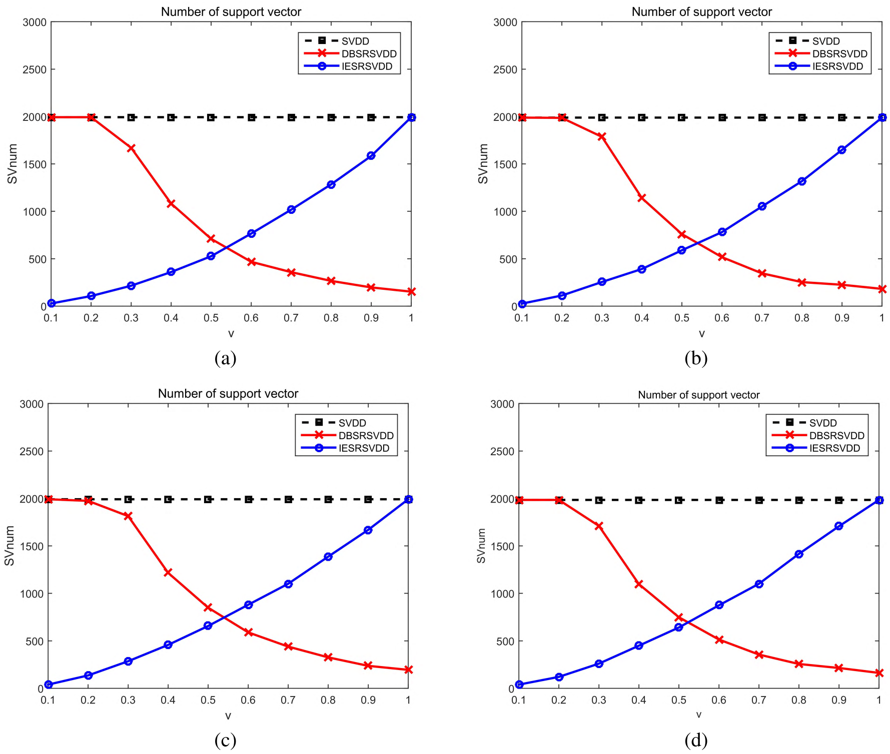

FIGURE 18. Numbers of support vectors of SVDD, IESRSVDD, DBSRSVDD in 4 fault conditions (v = 0.1-1.0). (a) Fault1. (b) Fault2. (c) Fault3. (d) Fault4.

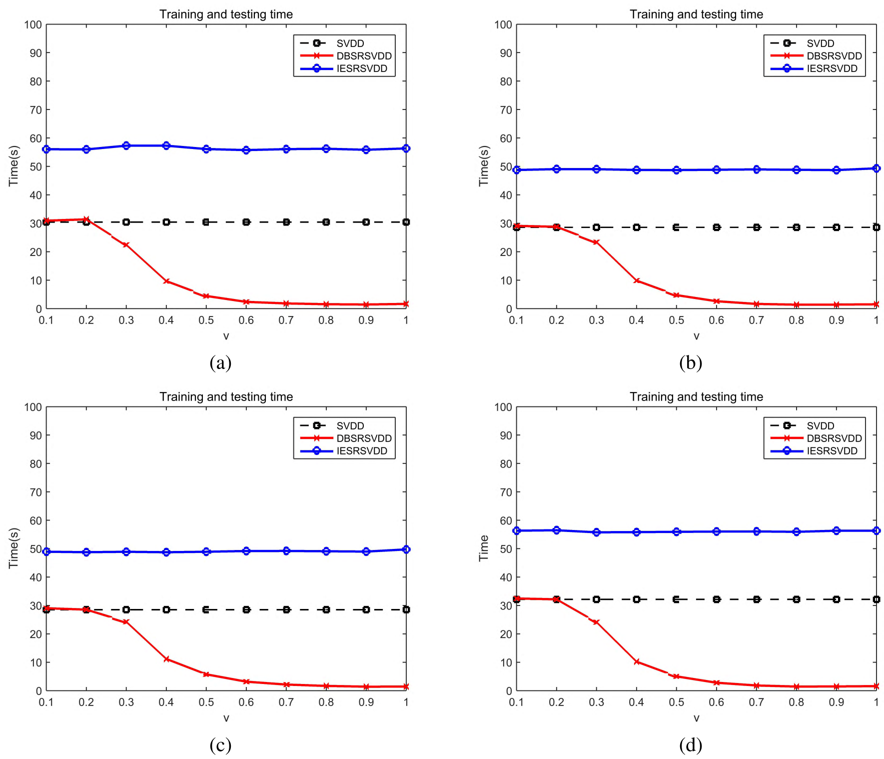

FIGURE 19. Time consumption of SVDD, IESRSVDD, DBSRSVDD in 4 fault conditions (v = 0.1-1.0). (a) Fault1. (b) Fault2. (c) Fault3. (d) Fault4.

needs longer time which is at least 50 seconds because the Euclidean distance between each other, probability of uncertainty between two samples and entropy of each sample need to be calculated. Probability of uncertainty between two samples is defined as $p_{ij} = dist_{ij}/\sum_{k=1}^{n} dist_{ik}$ and entropy of each sample is defined as $H_i = -\sum_{k=1}^{n} p_{ik} \log p_{ik}$(n is the sample size of the original data set) [14]. Plenty of time is spent on calculating information entropy before

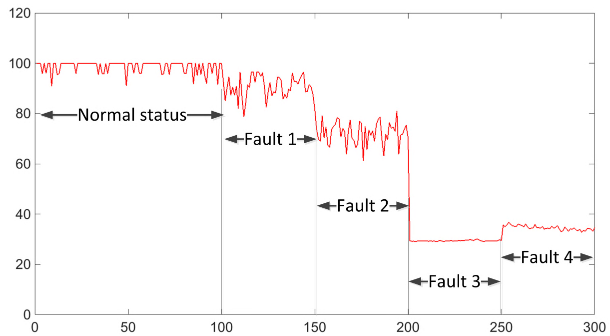

FIGURE 20. Health condition score of Node4.

TABLE 1. Health condition of each node.

Node	Normal	Fault1	Fault2	Fault3	Fault4
Node4	100	95	71	29	35
Node1	100	92	75	0	41
Node2	100	96	80	0	39
Node3	100	95	78	30	42
Node5	100	96	83	46	54
Node6	100	90	77	35	39

sample reduction. In the sample reduction phase of DBSRSVDD algorithm, besides the Euclidean distance calculation between each other, there are only $(n + n_1 + n_2 + \ldots + n_i + \ldots n_p)$ times of comparison operation. p is the number of loops and n_i is sample size in the i_{th} loop. In DBSRSVDD n_{i-1} are smaller than n_i. That's why DBSRSVDD takes less time than IESRSVDD in sample reduction operation. When parameter v is bigger than 0.4 the sample reduction time is usually less than 5 seconds and the training set size is less than one-tenth of the original sample size. Therefore, the convex optimization process of SVDD only consumes at most 0.1 second which is negligible due to the significant reduction in sample size. In conclusion, DBSRSVDD perform better than SVDD and IESRSVDD.

Health condition of each node in MVB is quantized by DBSRSVDD in normal condition and four fault conditions. Take Node4 for example, its health condition result is shown in Fig.20(average value in each stage: 100/95/71/29/35). Each node's health condition in four faults is shown in table1. Because Node1 and Node2 are off-line in Fault3, their evaluation scores are 0.

D. HEALTH EVALUATION OF THE WHOLE NETWORK

The whole network's evaluation result is the linear weighted sum of the node scores. Their weights are determined by the bandwidth occupancy of each node. Fig 21 and table2 show the bandwidth occupancy information of each node. According to [35], the main communication mode of MVB is process data (PD) port communication and its medium access control method is time division multiplexing. PD ports communication can be seen as the basic transmitting unit of MVB nodes. It takes two steps to complete one time PD communication. A bus administrator (BA, Node4 in test bed) polls the periodic scheduling list for addressing the logical ports, and broadcasts main frames containing logical port addresses to MVB. The slave node that matches to the logical addresses responds to the bus administrator and broadcasts slave frames to MVB. 1 millisecond is a basic period and consists of several times of ports communication. The macro period is the polling cycle in which all ports will be polled at least once [35]. A macro period consists of several basic periods, as illustrated in Fig.21. In the experiments we set the macro period with 3 basic periods. Each node is configured with some PD ports and gets access to the bus when it is polled by the master frame. There are 7 times of PD ports communication in period1, 5 times of PD ports communication in period2 and 7 times of PD ports communication in period3. Each time of ports communication is made up of a master frame which is transmitted by node4 and a slave frame which is transmitted by the slave node. The length of master frame is fixed to 16 bits and the length of slave frame is 16/32/64/128/256 bits depending on requirement.

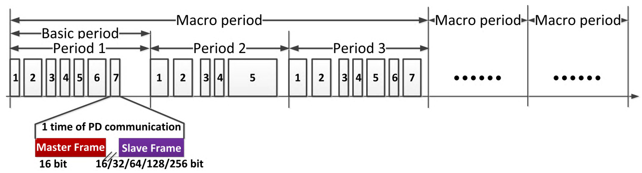

FIGURE 21. MVB scheduling list in the experiments.

TABLE 2. The sizes of the communication data volume of each node.

Node	Period1(bits)	Period2(bits)	Period3(bits)	total(bits)
Node4(BA)	16	32	32	384
Node1	32	32	32	96
Node2	16*2	16	16*2	80
Node3	16	16	32	64
Node5	32	0	16	48
Node6	16	64	32	112

Table2 shows the length of slave frames transmitted by MVB nodes in each basic period. For node4, we can know that it transmits 19(7 + 5 + 7) master frames, 1 slave frame (16bits) in period1, 1 slave frame (32bits) in period2, 1 slave frame (32bits) in period3 totally. So we can calculate that there are 384 bits data transmitted by node4 in a macro period: $(7 + 5 + 7) * 16 + 16 + 32 + 32 = 384$ bits. As shown in equation (14), bits number of node1 is 96bits, node2 is 80bits, node3 is 64bits, node5 is 48bits, node6 is 112bits.

$$
\begin{aligned}
W(4) &= \frac{384}{384+96+80+64+48+112} = 0.4898 \\
W(1) &= \frac{96}{384+96+80+64+48+112} = 0.1224 \\
W(2) &= \frac{80}{384+96+80+64+48+112} = 0.1020 \\
W(3) &= \frac{64}{384+96+80+64+48+112} = 0.0816 \\
W(5) &= \frac{48}{384+96+80+64+48+112} = 0.0612 \\
W(6) &= \frac{112}{384+96+80+64+48+112} = 0.1429 \quad (14)
\end{aligned}
$$

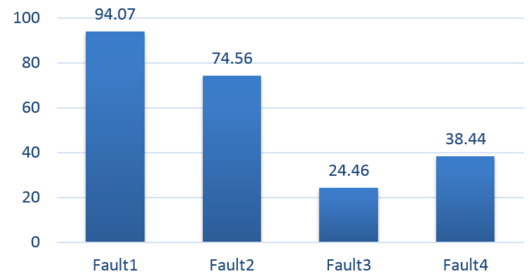

FIGURE 22. Health condition of the whole network in 4 fault conditions.

Based on table1 and table2, evaluation result of the whole network can be calculated as equation (15). W(k) is the weight of node in equation (14) and score (k) is the node's evaluation result in Table1. The health conditions of the whole network in 4 fault conditions are shown in Fig.22.

$$F = \sum_{k=1}^{6} W(k) * score(k) \tag{15}$$

VII. CONCLUSION

In this paper, a novel method has been proposed to evaluate the network performance and detect the abnormal condition of MVB. A health evaluation method combined SVDD with a sample reduction strategy is proposed to reduce the training time and computational complexity. Samples lie in the edge area of the positive class are selected as support vector candidates by an density-based sample reduction method. According to the results of experiments, the effectiveness of the proposed method has been validated. The proposed method can significantly reduce sample size and accelerates model training. Without any optimization process of penalty parameter and kernel function parameter, DBSRSVDD takes much less time in training model than IESRSVDD and SVDD while sustains the classification performance at the same time. Moreover, parameter v in DBSRSVDD decides the proportion of the selected samples. The effect of different v values on sample reduction has been investigated and the trend is opposite to IESRSVDD. When v is bigger than 0.6, a good sample reduction result is achieved. Since fault diagnose and location of MVB are still problems that plagues engineers, further researches will focus on these two issues.

REFERENCES

[1] X. X. Wen and X. R. Meng, "Network health evaluation based on SVM and cloud model," *J. Beijing Univ. Posts Telecommun.*, vol. 35, no. 1, pp. 10–14, Jan. 2012.
[2] Y. Xie, X. Chen, X. Yu, B. Yue, and J. Guo, "Fast SVDD-based outlier detection approach in wireless sensor networks," *Chin. J. Sci. Instrum.*, vol. 32, no. 1, pp. 46–51, May 2011.
[3] S. Hawkins, H. He, G. Williams, and R. Baxter, "Outlier detection using replicator neural networks," in *Proc. Springer Int. Conf. Data Warehousing Knowl. Discovery*, Aix-en-Provence, France: Springer, Sep. 2002, pp. 170–180.
[4] X. Jia, C. Jin, M. Buzza, W. Wang, and J. Lee, "Wind turbine performance degradation assessment based on a novel similarity metric for machine performance curves," *Renew. Energy*, vol. 99, no. 1, pp. A1191–A1201, Dec. 2016.
[5] D. J. Marchette, "A statistical method for profiling network traffic," in *Proc. Workshop Intrusion Detection Netw. Monitor.*, Apr. 1999, vol. 9, no. 12, pp. 119–128.
[6] N. Wu and J. Zhang, "Factor-analysis based anomaly detection and clustering," *Decis. Support Syst.*, vol. 42, no. 1, pp. 375–389, Oct. 2006.
[7] Z. Xue, Y. Shang, and A. Feng, "Semi-supervised outlier detection based on fuzzy rough C-means clustering," *Math. Comput. Simul.*, vol. 80, no. 9, pp. 1911–1921, May 2010.
[8] H. Su, H. Wu, X. Chen, and M. Z. Q. Chen, "Positive edge consensus of complex networks," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 48, no. 12, pp. 2242–2250, Dec. 2018.
[9] M. Long, H. Su, and B. Liu, "Second-order controllability of two-time-scale multi-agent systems," *Appl. Math. Comput.*, vol. 343, pp. 299–313, Feb. 2019.
[10] H. E. Solberg and A. Lahti, "Detection of outliers in reference distributions: Performance of Horn's algorithm," *Clin. Chem.*, vol. 51, no. 12, pp. 2326–2332, Dec. 2005.
[11] S. E. Guttormsson, R. J. Marks, II, M. A. El-Sharkawi, and I. Kerszenbaum, "Elliptical novelty grouping for on-line short-turn detection of excited running rotors," *IEEE Trans. Energy Convers.*, vol. 14, no. 1, pp. 16–22, Mar. 1999.
[12] D. M. J. Tax and R. P. W. Duin, "Support vector data description," *Mach. Learn.*, vol. 54, no. 1, pp. 45–66, Jan. 2004.
[13] C. Liu, W. Wang, M. Wang, F. Lv, and M. Konan, "An efficient instance selection algorithm to reconstruct training set for support vector machine," *Knowl.-Based Syst.*, vol. 116, no. 15, pp. 58–73, Jan. 2017.
[14] D. Li, Z. Wang, C. Cao, and Y. Liu, "Information entropy based sample reduction for support vector data description," *Appl. Soft Comput.*, vol. 71, pp. 1153–1160, Oct. 2018.
[15] B. Li, Q. Wang, and J. Hu, "A fast SVM training method for very large datasets," in *Proc. Int. Joint Conf. Neural Netw.*, Jun. 2009, pp. 1784–1789. [Online]. Available: https://ieeexplore.ieee.org/document/5178618?arnumber=5178618
[16] C. Wang, Z. Zhao, L. Gong, L. Zhu, Z. Liu, and X. Cheng, "A distributed anomaly detection system for in-vehicle network using HTM," *IEEE Access*, vol. 6, pp. 9091–9098, Jan. 2018.
[17] H. S. Su, H. Wu, and X. Chen, "Observer-based discrete-time nonnegative edge synchronization of networked systems," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2446–2455, Oct. 2018.
[18] G. Rätsch, S. Mika, B. Schölkopf, and K. Müller, "Constructing boosting algorithms from svms: An application to one-class classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 9, pp. 1184–1199, Sep. 2002.
[19] M. K. Long, H. S. Su, and B. Liu, "Group controllability of two-time-scale multi-agent networks," *J. Franklin Inst.*, vol. 355, no. 13, pp. 6045–6061, Sep. 2018.
[20] X. Yan, H. Li, L. Wang, and P. Shen, "Optimal bandwidth scheduling of networked learning control system based on Nash theory and auction mechanism," *Math. Problems Eng.*, vol. 2013, no. 1, pp. 1–8, Jun. 2013.
[21] Y. Z. Wang *et al.*, "Fuzzy immune particle swarm optimization algorithm and its application in scheduling of MVB periodic information," *J. Intell. Fuzzy Syst.*, vol. 32, no. 6, pp. 3797–3807, May 2017.
[22] A. Rodriguez and A. L. Dubes, "Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492–1496, Jun. 2014.
[23] G. A. Marcoulides, "Discovering knowledge in data: An introduction to data mining," *J. Amer. Stat. Assoc.*, vol. 100, no. 472, p. 1465, Dec. 2005.
[24] M. Ester, H. Kriegel, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. 2nd Int. Conf. Knowl. Discovery Data Mining (KDD)*, Portland, OR, USA, 1996, pp. 226–231.
[25] S. Guha, R. Rastogi, and S. K. Rock, "A robust clustering algorithm for categorical attributes," *Inf. Syst.*, vol. 25, no. 5, pp. 345–366, Feb. 2000.
[26] L. Ertz, M. Steinbach, and V. Kumar, "Finding topics in collections of documents: A shared nearest neighbor approach," in *Clustering and Information Retrieval*. Boston, MA, USA: Springer, 2004, pp. 88–103.
[27] A. Ypma, E. Ypma, and R. P. W. Duin, "Novelty detection using self-organizing maps," *Prog. Connectionist-Based Inf. Syst.*, vol. 2, no. 1, pp. 1322–1325, Apr. 1997.
[28] V. Emamian, M. Kaveh, and A. H. Tewfik, "Robust clustering of acoustic emission signals using the Kohonen network," in *Proc. IEEE Int. Conf. Acoust.*, Istanbul, Turkey, Jun. 2002, pp. 3891–3894.
[29] X. Jiang, K. Liu, J. Yan, and W. Chen, "Application of improved SOM neural network in anomaly detection," *Phys. Procedia*, vol. 33, pp. 1093–1099, Jun. 2012.

[30] F. Kuang, W. Xu, and S. Zhang, "A novel hybrid KPCA and SVM with GA model for intrusion detection," *Appl. Soft Comput.*, vol. 18, pp. 178–184, May 2014.

[31] J.-X. Dong, A. Krzyzak, and C. Y. Suen, "Fast SVM training algorithm with decomposition on very large data sets," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 4, pp. 603–618, Apr. 2005.

[32] J. C. Platt, "Fast training of support vector machines using sequential minimal optimization," in *Support Vector Learning*. Cambridge, MA, USA: MIT Press, 1999, pp. 185–208.

[33] X. Yang, J. Lu, and G. Zhang, "Adaptive pruning algorithm for least squares support vector machine classifier," *Soft Comput.*, vol. 14, no. 7, pp. 667–680, May 2010.

[34] J. Huang, X. Hu, and F. Yang, "Support vector machine with genetic algorithm for machinery fault diagnosis of high voltage circuit breaker," *Measurement*, vol. 44, no. 6, pp. 1018–1027, Jul. 2011.

[35] *The IEC/IEEE Train Communication Network*, IEEE Standard 61375-2, 2003.

[36] V. Chandola, A. Banerjee, and V. Kuamr, "Anomaly detection: A survey," *ACM Comput. Surv.*, vol. 41, no. 3, pp. 11–23, Jul. 2009.

[37] W. Lawrenz, "CAN system engineering," in *From Theory to Practical Applications*. London, U.K.: Springer, 2013, pp. 41–130.

[38] I. W. Tsang, A. Kocsor, and J. T. Kwor, "Efficient kernel feature extraction for massive data sets," in *Proc. 12th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)* Philadelphia, PA, USA, 2006, pp. 724–729.

[39] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Comput.*, vol. 13, no. 7, pp. 1443–1471, Jul. 2014.

[40] G. Williams, R. Baxter, H. He, S. Hawkins, and L. Gu, "A comparative study of RNN for outlier detection in data mining," in *Proc. IEEE Int. Conf. Data Mining*, Dec. 2002, pp. 709–712. [Online]. Available: https://ieeexplore.ieee.org/document/1184035

[41] Y. Wang, Z. Wu, Q. Li, and Y. Zhu, "A model of telecommunication network performance anomaly detection based on service features clustering," *IEEE Access*, vol. 5, pp. 17589–17596, Mar. 2017.

ZHAOZHAO LI received the B.Eng. degree from the Hebei University of Technology, China, in 2012. He is currently pursuing the Ph.D. degree in electrical engineering with the School of Beijing Jiaotong University, Beijing, China. His research interests include fault diagnosis and anomaly detection of train communication networks and industrial control networks.

LIDE WANG received the M.Eng. degree from Southwest Jiaotong University, China, in 1986. He is currently a Professor with the School of Electrical Engineering, Beijing Jiaotong University. His research interests include the control of electric traction systems, computer control network technique, and embedded systems and application. He is a Senior Member of the China Railway Society.

YUEYI YANG received the B.Eng. degree in electrical engineering from Zhongyuan University of Technology, in 2011, and the M.Sc. degree in electrical engineering from Southwest Jiaotong University, in 2014. He is currently pursuing the Ph.D. degree in electrical engineering with Beijing Jiaotong University, Beijing, China. His research interests include networked and distributed systems, and industrial cyber-physical systems.

XIAOMIN DU received the B.Eng. degree in electrical engineering and automation from the China University of Mining and Technology, Beijing, China, in 2017. She is currently pursuing the master's degree in electrical engineering with Beijing Jiaotong University, Beijing. Her research interests include fault diagnosis and fault prognostics.

HUI SONG received the B.Eng. degree in electronic and information engineering from Hebei University, in 2017. He is currently pursuing the master's degree in electrical engineering with Beijing Jiaotong University, Beijing, China. His research interests include fault injection and fault diagnosis.