Computers & Security 137 (2024) 103619

Contents lists available at ScienceDirect

Computers & Security

journal homepage: www.elsevier.com/locate/cose

Anomaly detection of aviation data bus based on SAE and IMD

Huang Li [a], Yiqin Sang [a], Hongjuan Ge [a,*], Jie Yan [b], Shijia Li [a]

[a] College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing, China
[b] The 54th Research Institute of China Electronics Technology Group Corporation, Shijiazhuang, China

ARTICLE INFO

Keywords:
SAE
IMD
CAS
Heuristic multi-threshold selection method
Anomaly detection

ABSTRACT

To detect remote terminal (RT) spoofing attacks on MIL-STD-1553B data bus and prevent the network paralysis of integrated avionics system (IAS) caused by misjudgment, an anomaly detection method of aviation data bus based on the combination of sparse autoencoder (SAE) and integrated mahalanobis distance (IMD) is proposed. Aiming at the communication traffic training set with only normal data, an unsupervised learning algorithm SAE is used to train a model that only represents normal behavior. To combine the feature information of each layer within SAE, the IMD, which can measure the similarity between data characteristics, is used to obtain the anomaly score of test data, and the comprehensive anomaly score (CAS) is obtained by considering the reconstruction error between SAE input and output. To solve the problem that data distribution and detection requirements were not considered in a single threshold, a heuristic multi-threshold selection method is proposed, which maximizes the performance of the classifier by considering the accuracy, youden index (YI), and F1. The experimental results demonstrate the effectiveness and feasibility of the method.

1. Introduction

In order to improve the operational efficiency and safety of civil airliners, reduce operating costs, and improve the convenience of passengers' communication, Boeing and Airbus have successively put forward the strategies of “e-Enable” and “e-Solution”, aiming at integrating all key aviation applications and services into a highly networked communication system, including wireless sensor networks, automatic dependent surveillance-broadcast (ADS-B), global navigation satellite system (GNSS), next-generation satellites, IP based networks, and so on. However, in different aviation environments, integrating all these different technologies will inevitably lead to complex infrastructure and increase vulnerability to various network threats. Among them, due to the lack of security measures in air traffic control system, GNSS and ADS-B are vulnerable to various attacks, which may threaten the safety of aircraft, passengers and staff. To address this problem, Habler et al. (2021) proposed a software-based security solution to detect abnormal traffic conditions without adding external sensors. A stacked long short-term memory (LSTM) encoder-decoder model is proposed for non-local ADS-B messages, which can learn the flight mode of aircraft in the surveillance airspace and accurately detect common attack types. The civil navigation messages (CNAV) of GNSS is transmitted in an open channel, lacks security authentication mechanism, faces the threat of interception and tampering, and is susceptible to spoofing attacks. Wu et al. (2023) constructed an anti-spoofing architecture for CNAV based on blockchain, and proposed a blockchain-based authentication scheme for CNAV using domestic cryptographic algorithms, and implemented authentication protocols to prevent CNAV tampering in GNSS. Considering the low bandwidth and limited available data bits of ADS-B systems, Wu et al. (2020) proposed an ADS-B message authentication method based on short certificate-less signatures. The method reduces the signature length by 3/4 during the signature stage and improves overall performance to a certain extent.

IAS is an integrated system of multiple avionics subsystems cross-linked together by a multiplexed transmission data bus. The system is flexible and highly reliable, and has gradually been applied in some advanced fighters and high-performance civil aircraft (Qiao et al., 2020; He et al., 2020a). The most mature data bus in IAS is MIL-STD-1553B, abbreviated as 1553B, which has the characteristics of strong reliability and fast computing speed (Li et al., 2018). However, the 1553B bus is a control network isolated from other networks, belonging to a legacy system. At the beginning of its design, it was not designed safely, and there was a lack of access authentication mechanisms in the application layer protocol, making the whole bus system vulnerable to modern network threats (Santo et al., 2021), such as denial of service (DoS) attacks, data leakage, and integrity damage (e.g. spoofing

[*] Corresponding author.
E-mail address: gehongjuan1101a@nuaa.edu.cn (H. Ge).

https://doi.org/10.1016/j.cose.2023.103619
Received 3 July 2023; Received in revised form 30 September 2023; Accepted 20 November 2023
Available online 29 November 2023

attacks), which may compromise the confidentiality, integrity, and availability of the 1553B bus (Stan et al., 2018). Once an intruder successfully exploits the vulnerability of the 1553B bus, it may lead to the complete collapse of the network in the IAS. So the research of 1553B aviation data bus communication traffic anomaly detection is crucial to the network security of the IAS and even the entire aircraft, and has received increasing attention from scientific research institutions and universities in recent years (Santo et al., 2021; He et al., 2020a, 2020b, 2020c; Yahalom et al., 2019a; Stan et al., 2017, 2020; Losier et al., 2019; Genereux et al., 2019; Onodueze et al., 2020).

The 1553B bus is mainly composed of three key components, namely, bus controller (BC), remote terminal (RT), and bus monitor (BM), and only BC and RT are authorized to transmit data on the bus. Therefore, the attacker who injects malicious messages into the IAS usually faked as BC or RT to launch integrity or availability attacks (He et al., 2020a). The attack source is faked as BC can send data to the 1553B bus, and RT spoofing attacks can cause BC to receive false data, resulting in the inability to accurately determine the state of the IAS, which is easier and more frequent than the BC spoofing attack (He et al., 2020b). So we focus on anomaly detection of the RT spoofing attack traffic dataset generated by scholars from Ben-Gurion University on the 1553B data bus (Yahalom et al., 2019a). However, the dataset is in a high-dimensional space, sparsely distribution, and difficult to estimate density, which is not conducive to data characteristics extraction. In addition, the training set in this dataset only has normal data, but lacks abnormal data with real labels. The commonly used anomaly detection algorithms based on machine learning can be divided into supervised learning and unsupervised learning. If the traditional supervised learning algorithms are used to learn the features of the training set, the detection cost and difficulty are high, and it may not be possible to accurately identify intrusions behavior in the testing set. In contrast, unsupervised learning is more suitable for anomaly detection in the whole training process, without any labeled data and prior information.

Therefore, we use SAE, a variant of classical unsupervised learning algorithm autoencoder (AE), as the basic algorithm, which cannot only extract the most important features of the original input samples, reduce the number of parameters that need to be updated in the whole model, and improve the nonlinear ability and training speed of the model, but also establish a normal behavior model based on unlabeled normal data, and calculate the reconstruction error between the input test data and the output as the anomaly score. However, only by using reconstruction-based algorithms to detect the RT spoofing attacks on the 1553B bus, we may not be able to capture anomalies with low reconstruction errors and close to the manifold with potential dimensions of internal samples (Denouden et al., 2018). It has been proved that by diversifying the error sources in anomaly score calculation, the distribution of mean square error anomaly scores on normal and abnormal is farther and farther, which can improve the performance of anomaly detection (Ryu et al., 2022). Aiming at the problem that there are only normal data in the training set used in this paper, the metric learning algorithms IMD is introduced into the unsupervised learning algorithm SAE to quantify the similarity between the characteristics of each layer of SAE, which is helpful for the model to learn the data distribution better and extract data characteristics effectively, and lays the foundation for anomaly detection of data bus network.

The main contributions are summarized as follows:

- To comprehensively consider the input-output reconstruction error of SAE and the characteristic information of the each layer output in SAE, an anomaly detection method based on the combination of SAE and IMD, and the concept of CAS are proposed to detect RT spoofing attack. The weighted sum of the reconstruction errors between the IMD containing all layer characteristics of SAE and the input-output of SAE is used to obtain CAS, which can better distinguish the heterogeneity between normal and abnormal data and improve the accuracy of anomaly detection.
- Threshold is crucial for anomaly detection, and the current traditional threshold setting method is based on a single performance metric, without considering data distribution and the requirements of anomaly detection, which is one-sided. We proposed a heuristic multi-threshold selection method, which combines accuracy, YI, and F1 to find the optimal threshold, maximize the performance of the classifier and effectively identify abnormal traffic in complex network environments.
- The proposed method is compared with the traditional Machine Learning (ML) algorithms, AE variants, state-of-the-art methods, ablation studies, the latest solutions, and the anomaly discrimination analysis between the proposed method and the basic algorithms is performed based on heuristic multi-threshold selection methods.

The rest of this paper is as follows: the theoretical background and achievements of existing networks or bus security analysis and anomaly detection are discussed in Section 2. Section 3 summarizes the basic theories, including SAE and Mahalanobis distance (MD). In addition, the anomaly detection process of our proposed method is described. In Section 4, a comparative experimental analysis based on the 1553B bus RT spoofing attacks dataset is performed to describe the superiority of our proposed method. In Section 5, results and discussion are given. Section 6 concludes the paper and offers future work directions.

2. Related works

2.1. 1553B security analysis

Given the network threats that the 1553B bus usually faces, Stan et al. (2017) built a simulation platform for launching DoS and spoofing attacks on the 1553B bus. According to the periodic characteristics of messages on the bus, the Markov chain was used to analyze the temporal interval of messages to detect network attacks on the bus, but the influence of non-periodic messages was not considered. Later, Stan et al. (2020) built a RT-based authentication detection module and proposed an anomaly detection module that can identify illegal connections and command word operations with a high detection rate.

Temporal characteristics are an effective means to identify threats, Losier et al. (2019) analyzed the size of the time interval of a benign data transmission process according to the histogram of each time series characteristic value and considered the 1553B bus as abnormal if its average percentage difference from the normal baseline exceeded a user-defined anomaly threshold. Genereux et al. (2019) proposed an anomaly detection system for spoofing attacks and faked data injection attacks by automatically optimizing the sliding size of the time interval based on detected traffic. However, this method can only detect time windows with periodic anomalies, but cannot identify specific abnormal messages. Onodueze et al. (2020) used various ML algorithms to detect aperiodic faked messages on the 1553B bus generated by Yahalom et al. (2019a), however, due to the highly unbalanced distribution of the training set, the test results deviated greatly from the normal data, which was seriously inconsistent with the actual results.

At present, the security analysis of the 1553B bus is mainly based on the temporal features of messages on the bus, and a time series model with thresholds is adopted for anomaly detection. When only the ML algorithms are used to detect the anomaly of 1553B bus communication traffic, the expected results may be unsatisfactory due to the unbalanced data distribution.

2.2. Anomaly detection based on reconstruction error

Rumelhart et al. (1986) first put forward the concept and basic principles of AE. The reconstruction method is used to detect the network traffic with a small abnormal proportion, and the threshold is used to distinguish whether the detected data is normal or not. In the past few years, the methods of obtaining the reconstruction error thresholds for AE and its variants have achieved remarkable results in network anomaly detection (Chen et al., 2018; Nguimbos et al., 2019; Tun et al., 2020; Choi et al., 2019; Salahuddin et al., 2021; Yokkampon et al., 2022).

Chen et al. (2018) used AE and its variants for anomaly detection with the threshold set to the 25th epoch of training loss for each model. Nguimbos et al. (2019) obtained satisfactory detection results on AE using KDD CUP99 training set reconstruction error percentile. Tun et al. (2020) automatically selected from percentile and recall thresholds, and based on the SAE model, the accuracy, precision (P), and recall (R) on the NSL-KDD dataset were all above 94 %. Choi et al. (2019) based on AE detection for distributed denial of service (DDoS) attacks, the threshold of reconstruction error is set by using the percentage of abnormal data, and the threshold is designed by using the reconstruction error θ_α corresponding to the upper limit α% of the distribution of reconstruction error so that α% abnormal data can be detected. Salahuddin et al. (2021) developed a threshold selection heuristic algorithm to maximize the F1 of DDoS attacks. Yokkampon et al. (2022) proposed a multi-scale convolution variational autoencoder and a threshold setting strategy based on error rate, which were used to optimize the anomaly detection performance of multivariate time series data with time correlation, instead of relying on receiver operator characteristic (ROC) based threshold.

In summary, the threshold setting of reconstruction error of AE and its variants is usually based on statistical distribution, training loss, or maximization of a single performance metric, without a comprehensive consideration of various performance metrics and data distribution, which leads to the one-sidedness of threshold setting. In addition, if only the reconstruction error between the AE input-output and its variants is calculated without considering the features of each layer, feature information extraction is not comprehensive enough, which may lead to poor test results.

2.3. MD-based anomaly detection

MD is a common distance index in metric learning, and many achievements have been made in anomaly detection based on MD, for example, Wang et al. (2021) used the dissimilarity metric of MD to identify related variables affecting clustering performance, applied shadowed rough-fuzzy clustering to obtain real values close to the prototype based on iterations, and confirmed the effectiveness of its clustering division on the NSL-KDD dataset. Park et al. (2018) detect network anomalies based on probability analysis, with characteristics highly related to multivariate normal distribution as input, and MD was used to detect network traffic data anomalies. Zheng et al. (2022) proposed an MD anomaly detection method based on rescaling transformation to reduce the difference between the reconstructed and the original distribution, improving the feature extraction ability of AE. Vilaça et al. (2019) proposed a semi-supervised anomaly detection algorithm for botnets by using robust principal component analysis and MD, which can detect potential attacks.

The existing anomaly detection methods combining metric learning algorithm with AE and its variants mostly use them to extract characteristics and reduce dimensions, and then use MD to determine the threshold of normal data distribution in low-dimensional space. If it exceeds this threshold, it is considered as anomaly (Zheng et al., 2022; Utkin et al., 2017; Ryu et al., 2022). Alternatively, MD is used as a coarse-grained division, and AE and its variants are combined with classifiers to achieve fine-grained anomaly identification (Li et al., 2022). Guo et al. (2018) used KNN distance to improve the detection performance of AE out-of-distribution, instead of just using AE input-output to reconstruct the error, which has a good detection effect. However, this technique requires iterations over the training set at inference time, which has high computational loss and occupies a lot of memory cost. The above approaches do not fully consider the MD of each layer containing all characteristic information of AE and its variants, which can fully reflect the anomaly, resulting in some limitations of detection results.

At present, the anomaly detection works based on AE and its variants, MD, and the combination of the two types of models has not been found on the 1553B data bus, so the method proposed in this paper provides a new idea for bus intrusion identification and anomaly detection in IAS.

3. Proposed method

In response to the problem of only normal data in the training set of 1553B communication traffic used in this paper, the test data are input into the trained SAE model, and the reconstruction error between input and output is calculated and then scaled to the range of [0, 1], which is called the anomaly score, to predict the possibility of sample anomalies. The anomaly score of the test data is obtained using IMD, which can measure the characteristic similarity of all layers of SAE, and CAS is obtained by weighted summation with the reconstruction error of SAE input and output, which makes the sources of the anomaly score diversified and better distinguishes the distribution heterogeneity among data, and improves the detection accuracy of RT spoofing attacks. In addition, a heuristic multi-threshold selection method is proposed, which compares the numerical value with the CAS. When the CAS is lower than the threshold, it is regarded as normal data, otherwise, it is regarded as abnormal data, which realizes the flexible distinction of abnormal data. The specific method is described below.

3.1. SAE training process

SAE is a variant model proposed by Ng (2011) to solve the problem of AE overfitting with strong nonlinear generalization ability. The input layers and hidden layers constitute the encoder layers, while the hidden layers and output layers constitute the decoder layers.

In the encoding stage, high-dimensional data is mapped into low dimensional, capturing the hidden patterns of the data, and the input is reproduced in the decoding stage. Compared with AE, SAE has more neurons in the hidden layer than in the input layer, allowing fewer units to be activated at the same time in the hidden layer. The L1 norm or KL divergence regularization term is adopted to sparsity constraints, forcing SAE to learn more characteristics (Yan et al., 2018).

First, the difference between the input and the reconstruction vector (i.e., the output) is minimized on the basis of AE, and the parameters are trained using the input x to find the appropriate parameters $L(x, g(f(x)))$ so that the decoded output $g(f(x))$ approach the input x to the maximum extent. The approximation degree is expressed by the reconstruction error $J(W, b)$, which is a multidimensional variable and generally refers to the mean square error (MSE) between the output and the original input value. Therefore, the anomaly score is the MSE loss function of input and output, and is expressed as follows:

$$\begin{aligned} J(W,b) &= \frac{1}{N}\sum_{x\in S} L\{x, g(f(x))\} \\ &= \frac{1}{N}\sum_{x\in S} L\{x, S_g(W_2 f(x) + b_2)\} \\ &= \frac{1}{N}\sum_{x\in S} L\{x, S_g\left(W_2\left(S_f(W_1 x + b_1)\right) + b_2\right)\} \\ &= \frac{1}{2N}\sum_{x\in S}\left(x - S_g\left(W_2\left(S_f(W_1 x + b_1)\right) + b_2\right)\right)^2 \end{aligned} \tag{1}$$

Where W is the weight parameter, b is the bias vector, S_f and S_g represent the Sigmoid activation functions of the encoder and decoder (Tun et al., 2020), respectively.

Afterward, adding sparsity constraint to AE. Given the input x, $h_j(x)$ represents the activation value of the hidden neuron j, and m is the number of hidden neurons. The average activation value of each hidden unit $\widehat{\rho}_j$ is:

$$\widehat{\rho}_j = \frac{1}{m}\sum_{i=1}^{m}\left[h_j(x)\right] \tag{2}$$

To get $\widehat{\rho}_j$ close to 0, an additional penalty factor is added to the loss function for cases where the penalty $\widehat{\rho}_j$ and the sparsity parameter ρ (usually a small value close to 0) are largely different. The KL divergence regularization term is used as a penalty factor:

$$\sum_{j=1}^{S_l} KL(\rho \| \widehat{\rho}_j) = \sum_{j=1}^{S_l}\left(\rho\log\frac{\rho}{\widehat{\rho}_j} + (1-\rho)\log\frac{1-\rho}{1-\widehat{\rho}_j}\right) \tag{3}$$

According to $J(W, b)$ obtained from Eq. (3) and Eq. (1), the MSE loss function $J_{sparse}(W, b)$ of SAE can be determined as follows:

$$\begin{aligned} J_{sparse}(W,b) &= J(W,b) + \beta\sum_{j=1}^{S_l} KL(\rho \| \widehat{\rho}_j) \\ &= \frac{1}{2N}\sum_{x\in S}\left(x - S_g\left(W_2\left(S_f\left(W_1 x + b_1\right)\right) + b_2\right)\right)^2 \\ &\quad + \beta\sum_{j=1}^{S_l}\left(\rho\log\frac{\rho}{\widehat{\rho}_j} + (1-\rho)\log\frac{1-\rho}{1-\widehat{\rho}_j}\right) \end{aligned} \tag{4}$$

Where β is the weight of the penalty factor controlling sparsity. Therefore, the last term of this function is a sparse penalty term, which is used to impose sparse constraints on the hidden unit, and S_l is the number of all neurons in the hidden layer.

The gradient descent method is used to find the optimal solution of the parameters by iteratively updating the connection weights W and the bias matrix b, and the updating rules are as follows:

$$W = W - \alpha\frac{\partial}{\partial w}J_{sparse}(W,b) \tag{5}$$

$$b = b - \alpha\frac{\partial}{\partial b}J_{sparse}(W,b) \tag{6}$$

Where α represents the learning rate, to ensure that the parameter update during backpropagation can reach convergence.

The flow chart for training SAE with preprocessed 1553B normal traffic data is shown in Fig. 1.

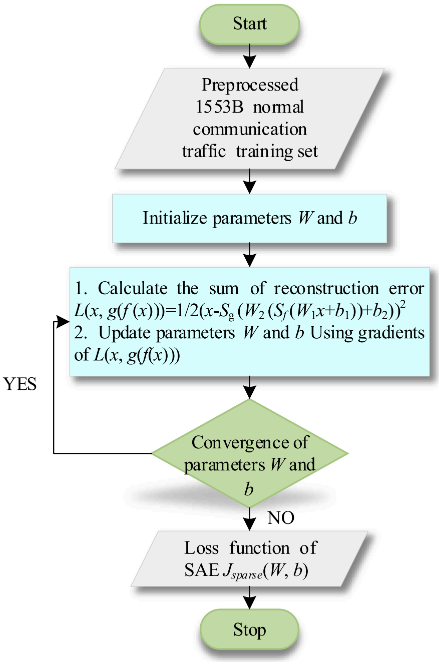

Start
Preprocessed 1553B normal communication traffic training set
Initialize parameters W and b
1. Calculate the sum of reconstruction error $L(x, g(f(x)))$=1/2(x-S_g(W_2(S_f(W_1x+b_1))+b_2))^2
2. Update parameters W and b Using gradients of $L(x, g(f(x)))$
YES
Convergence of parameters W and b
NO
Loss function of SAE $J_{sparse}(W, b)$
Stop

Fig. 1. Flow chart of the SAE training process.

3.2. Mahalanobis distance

MD is not affected by dimensionality, and the sample with multiple dimensions can be converted to one dimension by the covariance matrix, thus eliminating the influence of dimensionality and improving computational efficiency. Afterward, the one-dimensional sample are standardized and the distance between this standardized sample and the average value of the test sample is measured, that is, the similarity between the samples is obtained, and then their correlation is judged, correcting the problem of inconsistent and related scales for each dimension in the Euclidean distance (Zhang et al., 2022).

For a multivariate representation with n data, each of which is an m-dimensional column vector denoted as $X = (X_1, X_2,...., X_n)^T$, where the mean of X is $\mu = (\mu_1, \mu_2, \mu_m)^T$, the covariance matrix is $\sum$, and $\sum^{-1}$ is the inverse matrix of $\sum$. Therefore, the MD from a point $x_i = (x_1, x_2, \ldots, x_m)^T$ in X to the average value μ is expressed as (Imani et al., 2019):

$$M(x_i) = \sqrt{(x_i - \mu)^T \sum{}^{-1}(x_i - \mu)} \tag{7}$$

The element $\sum_{ij}$ of the i row and the j column of the covariance matrix $\sum$ is expressed as:

$$\begin{aligned} \sum_{ij} &= \operatorname{cov}\left(x_i, x_j\right) = E\left[\left(x_i - \mu_i\right)\left(x_j - \mu_j\right)\right] \\ &= \frac{\sum_{k=1}^{n}\left(\left(X_{k,i} - avg_i\right)\cdot\left(X_{k,j} - avg_j\right)\right)}{n-1} \\ &= \frac{\sum_{k=1}^{n}\left(\left(X_{k,i} - \sum_{k=1}^{n} X_{k,i} \Big/ n\right)\cdot\left(X_{k,j} - \sum_{k=1}^{n} X_{k,j} \Big/ n\right)\right)}{n-1} \end{aligned} \tag{8}$$

Where $i = [1, n], j = [1, n], k = [1, n]$. μ_i and μ_j represent the average values of vectors x_i and x_j, respectively, and the covariance matrix $\sum$ can be expressed as:

$$\sum = \begin{bmatrix} \operatorname{cov}(x_1, y_1) & \cdots & \operatorname{cov}(x_1, y_n) \\ \vdots & \ddots & \vdots \\ \operatorname{cov}(x_n, y_1) & \cdots & \operatorname{cov}(x_n, y_n) \end{bmatrix} \tag{9}$$

The inverse matrix $\sum^{-1}$ of the covariance matrix is calculated as follows:

$$\sum{}^{-1} = \frac{1}{\left|\sum\right|}\sum{}^{*} \tag{10}$$

Where $|\sum|$ is the rank of the covariance matrix $\sum$, and $\sum^{*}$ represents the adjoint matrix of $\sum$.

If a test data point is abnormal, then the calculated MD is obviously different from that of the normal data, with a significant similarity difference. The value of MD is the abnormal score of the test sample, and the larger the value, the higher the abnormal score. Combined with the set anomaly evaluation threshold, it indicates whether the sample point is normal or not.

3.3. The proposed anomaly detection method

The overall structure of the proposed anomaly detection method combining SAE and IMD is shown in Fig. 2.

SAE uses the 1553B traffic training set containing only normal data, inputting the testing set into the trained SAE so that the characteristics of each layer can be extracted from the test sample. The input layer characteristics of the extracted 1553B traffic testing set is $f_I(x)$, the hidden layer characteristics $f_H(x)$ and the output layer characteristics $f_O(x)$, and the each layer characteristics can be expressed as $f_l(x)$, and l is the number of layers.

The MD in this paper only uses the average value of the 1553B normal traffic data in the calculation process, so it will not be affected by

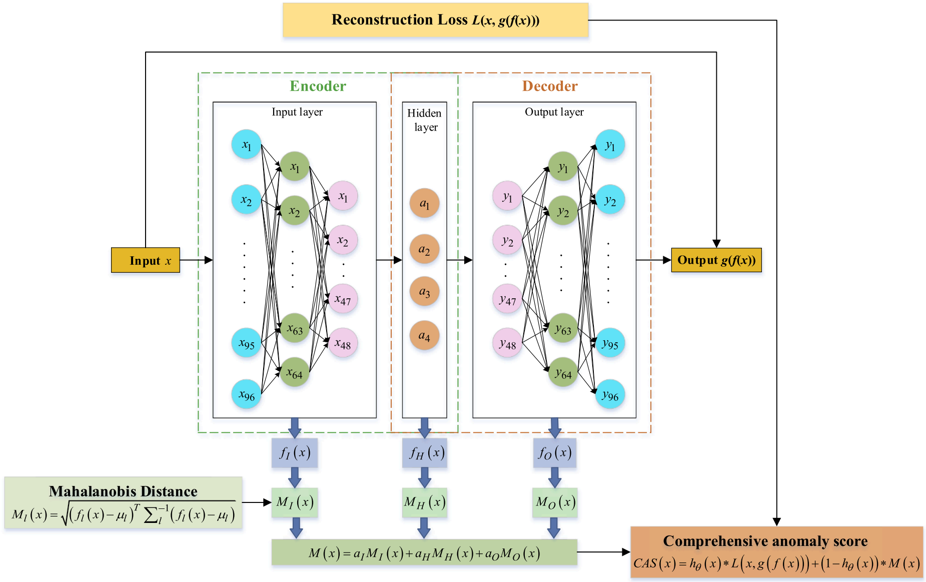


Fig. 2. The overall structure of our proposed method.

the intrusion data (Imani et al., 2019). Let the average value of the test set be μ_I and the covariance matrix be $\sum_I$. According to Eq. (7), the MD from the characteristics of each layer of SAE to the 1553B bus traffic test set μ_I is expressed as follows:

$$M_I(x) = \sqrt{(f_I(x) - \mu_I)^T \sum\nolimits_I^{-1} (f_I(x) - \mu_I)} \tag{11}$$

Therefore, the MDs of the input layer characteristics, hidden layer characteristics, and output layer characteristics of the test set are expressed as $M_I(x)$, $M_H(x)$, and $M_O(x)$ respectively. According to the composition of SAE, the number of input layers is a_I, the number of hidden layers is a_H and the number of output layers is a_O, which leads to the following IMD expression for all the layers of SAE in the 1553B bus communication traffic testing set:

$$M(x) = a_I M_I(x) + a_H M_H(x) + a_O M_O(x) \tag{12}$$

Logistic regression (LR) cannot only predict the result of binary classification but also get the probability of belonging to this category without assuming the data distribution in advance. Therefore, considering the reconstruction error of SAE input and output and IMD containing the characteristic similarity information of SAE layers, given a test sample input value x, the weight coefficient of reconstruction error $h_\theta(x)$ and the weight coefficient 1-$h_\theta(x)$ of IMD corresponding to the input sample are predicted by using LR classifier. It can comprehensively extract the characteristics of the test samples containing abnormal information, laying a solid foundation for the next step of accurately identifying abnormal traffic. Among them, the posterior probability estimates for output categories 1 and 0 are as follows, taking values in the range of [0, 1]:

$$P(y=1|x,\theta) = h_\theta(x) = \frac{1}{1+\exp(-x\theta)} \tag{13}$$

$$P(y=0|x,\theta) = 1 - h_\theta(x) = \frac{\exp(-x\theta)}{1+\exp(-x\theta)} \tag{14}$$

From Eq. (1), the reconstruction error between the input and output is $L(x, g(f(x)))$. The inverse matrix of each layer of SAE can be obtained according to Eq. (9) and Eq. (10). Therefore, the $CAS(x)$ includes the characteristic information of each layer in SAE with the input and output reconstruction errors, on the 1553B bus traffic testing set is shown as follows:

$$\begin{aligned} CAS(x) &= h_\theta(x) * L(x, g(f(x))) + (1 - h_\theta(x)) * M(x) \\ &= \frac{1}{1+\exp(-x\theta)} * \frac{1}{2}\left(x - S_g\left(W_2\left(S_f(W_1 x + b_1)\right) + b_2\right)\right)^2 \\ &\quad + \frac{\exp(-x\theta)}{1+\exp(-x\theta)} * \left(\begin{array}{l} a_I\sqrt{(f_I(x)-\mu_I)^T\sum_I^{-1}(f_I(x)-\mu_I)} \\ +a_H\sqrt{(f_H(x)-\mu_H)^T\sum_H^{-1}(f_H(x)-\mu_H)} \\ +a_O\sqrt{(f_O(x)-\mu_O)^T\sum_O^{-1}(f_O(x)-\mu_O)} \end{array} \right) \end{aligned} \tag{15}$$

Where, $CAS(x)\in[0,1]$, the greater $CAS(x)$ is, the greater the possibility that data X is abnormal.

3.4. Threshold selection method and detection process

In the anomaly detection of the 1553B traffic dataset, comparing $CAS(x)$ with the set threshold. When the threshold is set too high, it is helpful to distinguish all anomalies, but some abnormal samples may be regarded as normal, resulting in a higher false negative rate (FNR). Lower thresholds may mistakenly identify some normal samples as abnormal, resulting in a higher false positive rate (FPR). Therefore, determining the optimal threshold will directly affect the detection results.

The current commonly used threshold settings for anomaly detection are based on maximizing a single performance metric, without considering various performance metrics comprehensively. In addition, the data distribution is not considered, and a threshold of 0.5 is not the optimal decision boundary when the model is trained using an unbalanced training set, leading to a certain one-sidedness in threshold classification. Therefore, we proposed a heuristic multi-threshold selection method of combining accuracy, YI, and F1, which is used to identify abnormal traffic of the 1553B bus.

Accuracy, abbreviated as AC, is usually applied to scenarios where the proportion of normal and abnormal data is relatively balanced. It is the proportion of the sum of true positive (TP) samples and true negative (TN) samples to all samples. AC is proportional to the effect of anomaly classification, and the threshold for anomaly detection that can be set at this time is the threshold corresponding to the maximum accuracy. Among them, the accuracy is expressed as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{16}$$

Where TP is the number of samples correctly classified as attacks and TN is the number of samples correctly classified as normal. FN is the most critical metric in the 1553B bus anomaly detection. FP, also known as false positive, is the number of normal samples that are wrongly classified as attacks, which may lead to wrong decisions of the aircraft network security managers.

F1 is more concerned with rare classes under unbalanced data, it is a weighted harmonic average of precision and recall, which is conducive to the correct classification of abnormal traffic samples.

$$F1 = \frac{2 \times P \times R}{P + R} = \frac{2 \times P \times TPR}{P + TPR} \tag{17}$$

YI is the difference between true positive rate (TPR) (also called recall) and false positive rate (FPR), that is, TPR-FPR, which is the key metric to evaluate anomaly detection models. FPR and TPR are expressed as follows:

$$FPR = \frac{FP}{FP + TN} \tag{18}$$

$$TPR = \frac{TP}{TP + FN} \tag{19}$$

Therefore, considering the YI and F1 score, the threshold corresponding to the maximum value of the weighted harmonic average ω_β is solved, which is more conducive to the accurate identification of malicious traffic than the comprehensive anomaly score $CAS(x)$. Where β is the weighting factor, the expression of ω_β is:

$$\omega_\beta = \frac{\left(1+\beta^2\right) \times YI \times F1}{\beta^2 \times YI + F1} = \frac{1}{\dfrac{1}{\left(1+\dfrac{1}{\beta^2}\right) \times F1} + \dfrac{1}{\left(1+\beta^2\right) \times YI}} \tag{20}$$

According to Eq. (17) and Eq. (20), the detailed expression of ω_β can be deduced as follows:

$$\begin{aligned}\omega_\beta &= \frac{1}{\dfrac{\beta^2 \times (PR + TPR)}{\left(1+\beta^2\right) \times 2 \times PR \times TPR} + \dfrac{1}{\left(1+\beta^2\right) \times \max(TPR - FPR)}} \\ &= \frac{1}{\dfrac{\beta^2 \times (PR + TPR)}{\left(1+\beta^2\right) \times 2 \times PR \times TPR} + \dfrac{1}{\left(1+\beta^2\right) \times \max\left(\dfrac{TP \times TN - FP \times FN}{(TP + FN) \times (FP + TN)}\right)}}\end{aligned} \tag{21}$$

When $\beta = 1$, the weights representing YI and F1 are the same, that is, the two performance metrics are equally important. When $\beta\langle$ 1, it means that YI has more weight than F1, and YI is more important. When $\beta\rangle$ 1, the weight of F1 is greater than YI, and F1 is more important.

Fig. 3 shows the proposed method for anomaly detection of RT spoofing attack on the preprocessed 1553B traffic testing dataset.

In the testing stage, the test samples are input into the trained SAE model, and the reconstruction error $L(x, g(f(x))$ between SAE input and output is calculated using the parameters W and b obtained in the training stage. The SAE of the input test sample is extracted for each layer of characteristics $f_l(x)$, and the integrated Mahalanobis distance $M_l(x)$ containing each layer of characteristic information is calculated. $CAS(x)$ is obtained by a weighted sum of $L(x, g(f(x))$ and $M_l(x)$, and malicious traffic data is detected by the heuristic multi-threshold selection method. When the data distribution is relatively balanced, the threshold corresponding to the maximum accuracy is set as the optimal threshold. When the data distribution is unbalanced, the optimal threshold is set to the one corresponding to the maximum value of the weighted sum average ω_β. When there is no data distribution problem, the maximum value of the thresholds corresponding to the three performance metrics is used as the final threshold. For a given test sample including normal and abnormal communication traffic, it is determined whether its $CAS(x)$ exceeds the threshold, and if it does not, it is judged as normal, and vice versa.

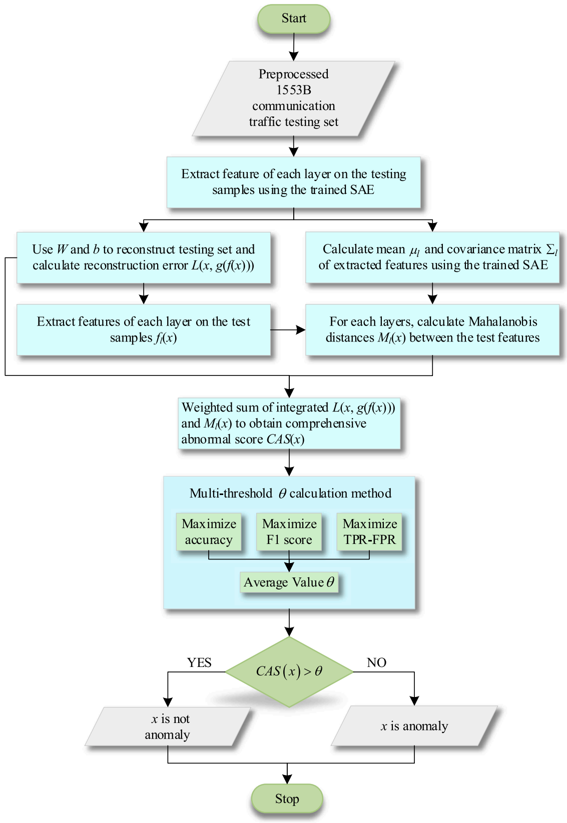

Start
Preprocessed 1553B communication traffic testing set
Extract feature of each layer on the testing samples using the trained SAE
Use W and b to reconstruct testing set and calculate reconstruction error $L(x, g(f(x)))$
Calculate mean μ_l and covariance matrix Σ_l of extracted features using the trained SAE
Extract features of each layer on the test samples $f_l(x)$
For each layers, calculate Mahalanobis distances $M_l(x)$ between the test features
Weighted sum of integrated $L(x, g(f(x)))$ and $M_l(x)$ to obtain comprehensive abnormal score $CAS(x)$
Multi-threshold θ calculation method
Maximize accuracy
Maximize F1 score
Maximize TPR-FPR
Average Value θ
YES
$CAS(x) > \theta$
NO
x is not anomaly
x is anomaly
Stop

Fig. 3. The anomaly detection flow of the proposed method in the testing stage.

3.5. Performance assessment metrics

To detect malicious traffic more effectively, besides the performance evaluation metrics mentioned in subsection 3.4, PR, G-mean, Matthews Correlation Coefficient (MCC), area under the curve (AUC), and average precision (AP) can also evaluate the performance of the models, as described below (Krueger et al., 2016).

PR is the proportion of true positive sample to predict positive sample, which is shown as follows:

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

The G-mean combines TPR and True Negative Rate (TNR), which applies to the case of little difference in data balance and is expressed as:

$$Gmean = \sqrt{TPR * TNR} \tag{23}$$

MCC balances all four detection results of the confusion matrix in a single metric, which is particularly useful when the data is unbalanced. It is bounded at $[-1, 1]$, where 1 indicates the best performance and -1 indicates a complete inconsistency between the test results and the true labels.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{24}$$

The ROC curve can be used to provide the relative importance of TPR and FPR, which is considered as an important metric of the performance for visualizing anomaly detection models. The closer the AUC is to 1, the more perfect classification performance is. The AUC expression is as follows:

$$AUC = \int_0^1 \frac{TP}{TP + FN} d \frac{FP}{TN + FP} \tag{25}$$

PR curve is composed of precision and recall, and the area surrounded by the curve is called AP. The higher the value of AP, the better the performance. Under extremely unbalanced data, the PR curve may be more practical than the ROC curve.

4. Experiment

All experiments were conducted on a personal computer with the following hardware configurations: Intel Core i5-9400F@2.90 GHz, NVIDIA GeForce GT 730, 256GB RAM, and Windows 10 operating system. The model was implemented in Python 3.6. Tensorflow, PyTorch1.2.0, and Keras frameworks are used to build the model, Pandas and Numpy libraries for data analysis and preprocessing, and Scikit Learn library for preprocessing and implementing metrics evaluation.

4.1. Experimental scenario description

In 2019, scholars from Ben-Gurion University simulated a simplified 1553B bus architecture for a flight control system (Yahalom et al., 2019a). The structure of the testing system (as shown in Fig. 4) consists of three PCs, which are used to simulate various components of the simplified IAS. The system provides a BC and several RTs, including a RT for the flight control computer, an RT for the aileron controller, and a malicious RT. Each component is connected to the bus through a 1553 interface card. The subsystems are distributed on different PCs, so that they can be physically placed at different distances from the bus, thus simulating the actual bus topology setting. The main components of the test system simulation are:

BC. This component is used to initiate all communication on the bus, where BC is implemented in BC/attacker PC (item 3 in Fig. 4) and connected to the bus through DDC BU-67114Hx interface.

BM. BM is implemented in the monitor PC (item 1 in Fig. 4) and connected to the bus through DDC BU-67114Hx interface, which is used for online monitoring and detecting abnormal messages transmitted through the bus.

RTs. RTs program is implemented on RTs PC (item 2 in Fig. 4). The RTs PC is connected to the bus through two interfaces: a DDC BU-67114Hx interface and an Excalibur ex-4000pcie card, which logically allows up to 32 connections.

Attacker component. It has the function of BC or RT, and is responsible for executing various attacks as a fake RT/BC (i.e., illegally connected to the bus) or damaged RT/BC. The attacker component is controlled manually through GUI, and its software is implemented in the same PC as the BC (item 3 in Fig. 4), and connected to the bus through another DDC BU-67114Hx interface.

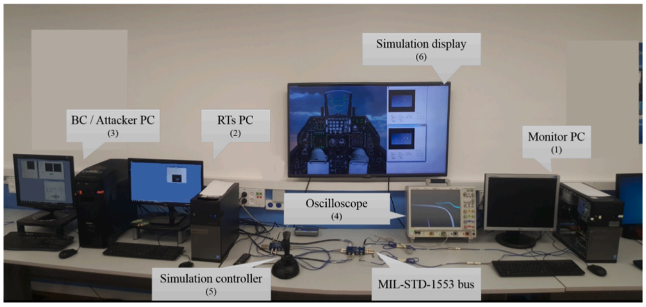


Fig. 4. Testing system architecture.

The testing system also contains: (1) an oscilloscope (item 4 in Fig. 4) for visualizing electric signals transmitted over the bus, (2) a controller (item 5 in Fig. 4) for simulating user operations, and (3) a display (item 6 in Fig. 4) for visualizing the physical impact on the simulated operations on the simulated system.

4.2. 1553B bus traffic datasets

Most of the traffic data are normal data that generated by the testing system mentioned in Section 4.1, the malicious RT spoofing attacks on the 1553B system are simulated by injecting data packets.

In Fig. 5, the eleven blue packets demonstrate a normal set request packet sequence, which is generated when RT@1 sends a request to set the value of property #0 for RT@2. Due to the lack of authentication in 1553B bus, malicious RT can exploit such a set request by impersonating RT@1 and setting a malicious value to property #0, which will alter the state of the aircraft. Assuming that RT@1 is the flight control computer, RT@2 is the aileron controller, and property #0 is the required roll value, the four orange packets (8M-11M) indicate that the malicious RT sets the malicious value to the required roll value by simulating the aircraft control calculator, and then the malicious RT replays the first seven packets (1M-7M) to be valid by masking the leftover packets (8–11). The normal data and abnormal data are available in Mendeley Data at https://doi.org/10.17632/jvgdrmjvs3.3 (Yahalom et al., 2019b).

The generated RT spoofing attack data includes three non-sequential datasets and three sequential datasets, and the recording form is shown in Fig. 6. The non-sequential datasets are message streams generated based on the data packets recorded on the bus simulator, and the sequential datasets are sequences of non-sequential message streams split into ten consecutive messages.

The distribution of the 1553B sequential dataset is given in Table 1, and it can be seen that the training dataset is all normal data and the testing dataset contains abnormal classes, while it can be seen that the incidence of attack messages in dataset1, dataset2, and dataset3 is 11 %, 1 %, and 0.1 %, respectively. The main challenge in studying this dataset is that the ratio of malicious samples to normal is about 1:10 to 1:1000. At this time, using traditional ML algorithms to train unprepared data with severely imbalanced distribution may lead to test results being too focused on normal behavior, and it is difficult to learn the characteristics of abnormal samples, making the detection results seriously distorted.

4.3. 1553B traffic datasets preprocessing

In this paper, dataset1 with the highest abnormal percentage in 1553B sequential dataset as the research object. Firstly, to overcome the unbalanced traffic data classes, the SMOTE oversampling algorithm generates more new samples for the minority class based on the idea of K nearest neighbors, which overcomes the overfitting problem in random oversampling. Assuming that the minority class is $\overline{X}$, X_i is a randomly selected sample from the K nearest neighbors of $\overline{X}$, a new synthetic sample X can be expressed as (Chen et al., 2018):

$$X = \overline{X} + rand(0,1) * (X_i - \overline{X}) \tag{26}$$

Where $rand(0,1)$ denotes a random number in the range of (0,1), $i = 1,2,\ldots,k$.

To eliminate the variance in characteristic quantification, the training set and testing set of the new synthetic sample X are normalized by maximum-minimum normalization to achieve data normalization, and the data is limited to the interval of [0,1]. The expression is as follows:

$$\overline{X} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{27}$$

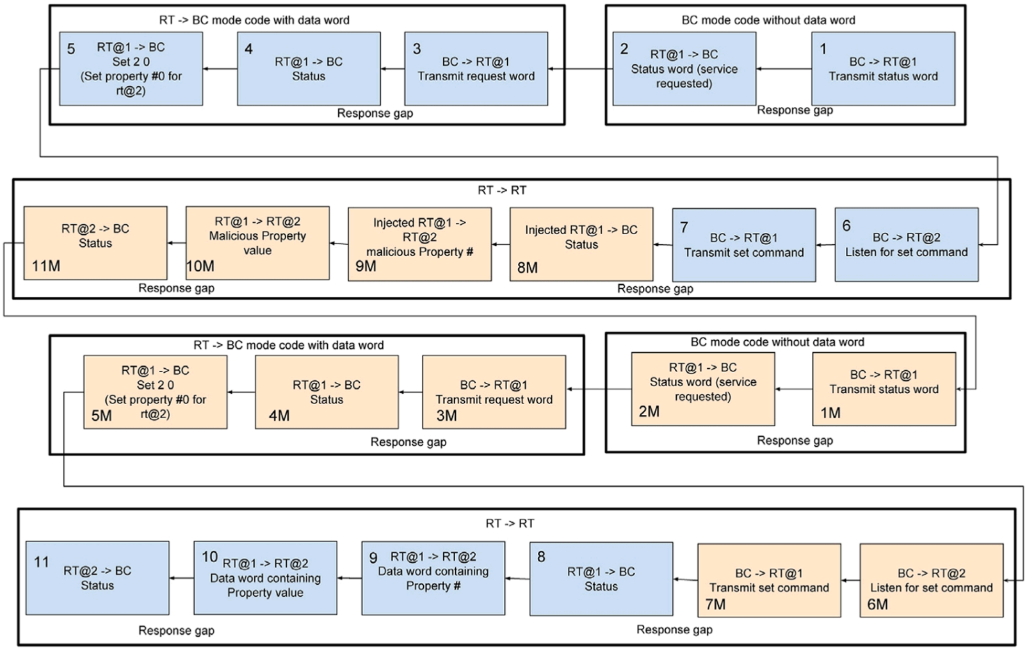
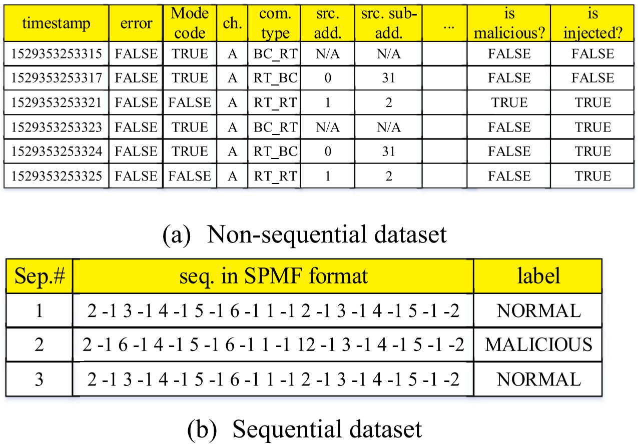

RT -> BC mode code with data word
5 RT@1 -> BC Set 2 0 (Set property #0 for rt@2)
4 RT@1 -> BC Status
3 BC -> RT@1 Transmit request word
Response gap
BC mode code without data word
2 RT@1 -> BC Status word (service requested)
1 BC -> RT@1 Transmit status word
Response gap
RT -> RT
RT@2 -> BC Status 11M
RT@1 -> RT@2 Malicious Property value 10M
Injected RT@1 -> RT@2 malicious Property # 9M
Injected RT@1 -> BC Status 8M
7 BC -> RT@1 Transmit set command
6 BC -> RT@2 Listen for set command
Response gap
Response gap
RT -> BC mode code with data word
RT@1 -> BC Set 2 0 (Set property #0 for rt@2) 5M
RT@1 -> BC Status 4M
BC -> RT@1 Transmit request word 3M
Response gap
BC mode code without data word
RT@1 -> BC Status word (service requested) 2M
BC -> RT@1 Transmit status word 1M
Response gap
RT -> RT
11 RT@2 -> BC Status
10 RT@1 -> RT@2 Data word containing Property value
9 RT@1 -> RT@2 Data word containing Property #
8 RT@1 -> BC Status
BC -> RT@1 Transmit set command 7M
BC -> RT@2 Listen for set command 6M
Response gap
Response gap

Fig. 5. RT spoofing attack via set request.



timestamp	error	Mode code	ch.	com. type	src. add.	src. sub-add.	...	is malicious?	is injected?
1529353253315	FALSE	TRUE	A	BC_RT	N/A	N/A		FALSE	FALSE
1529353253317	FALSE	TRUE	A	RT_BC	0	31		FALSE	FALSE
1529353253321	FALSE	FALSE	A	RT_RT	1	2		TRUE	TRUE
1529353253323	FALSE	TRUE	A	BC_RT	N/A	N/A		FALSE	TRUE
1529353253324	FALSE	TRUE	A	RT_BC	0	31		FALSE	TRUE
1529353253325	FALSE	FALSE	A	RT_RT	1	2		FALSE	TRUE

(a) Non-sequential dataset

Sep.#	seq. in SPMF format	label
1	2 -1 3 -1 4 -1 5 -1 6 -1 1 -1 2 -1 3 -1 4 -1 5 -1 -2	NORMAL
2	2 -1 6 -1 4 -1 5 -1 6 -1 1 -1 12 -1 3 -1 4 -1 5 -1 -2	MALICIOUS
3	2 -1 3 -1 4 -1 5 -1 6 -1 1 -1 2 -1 3 -1 4 -1 5 -1 -2	NORMAL

(b) Sequential dataset

Fig. 6. 1553B spoofing attack datasets record form.

Table 1
Distribution of 1553B sequential datasets.

Dataset	Training set		Testing set	
	Normal	Malicious	Normal	Malicious
Dataset1	1278	0	13,664	1690
Dataset2	3029	0	14,945	190
Dataset3	6013	0	14,289	20

Where $\overline{X}$ is the normalized data, X_{min} and X_{max} are the minimum and maximum values of X respectively.

Based on the above preprocessing steps, this paper uses the SMOTE oversampling method to generate 100 samples of malicious data in the dataset1 training set, so that the ratio of normal to malicious data in the training set and the testing set is close to avoid the generation of biased models with low attack detection rate. Then it is normalized to remove the influence of different dimensions.

4.4. Comparison methods description and settings

To verify the effectiveness of the proposed method, we will experimentally compare it with ML algorithms, the AE variants, the state-of-the-art methods, and the ablation research on the preprocessed 1553B bus RT spoofing attack traffic dataset. To ensure the fairness of the experiments, the parameters of the compared algorithms and the proposed method should be as consistent as possible, and the environment also be consistent.

The proposed method consists of MD, SAE, and encoder-decoder module, using LR prediction model to obtain the weight of each component of the anomaly score. The SAE module consists of 5 encoder layers and 5 decoder layers, which is an additional penalty by adding a KL divergence regularization term to the MSE loss function. Using the ReLU activation function enhances the sparsity of the network and improves the robustness of data. Dropout prevents the network from overfitting. Each group of hyper-parameters is in the parameter space by grid search, and evaluated by 10-fold cross-validation, to get the optimal hyper-parameter settings, as shown in Table 2.

4.4.1. ML algorithms

Experiment I: Traditional ML anomaly detection algorithms can be roughly divided into similarity-based methods, such as K-nearest neighbor (KNN) and local outlier factor (LOF). Probability models-based methods, such as Gaussian naive Bayes (GNB). Tree-based methods, such as random forest (RF), decision tree (DT), and extreme gradient

Table 2
The optimal values of the hyper-parameters.

Hyper-parameters	Optimal values
Encoder/Decoder layer	5
Encoder layer size (neuron)	96,64,48,16,4
Decoder layer size (neuron)	4,16,48,64,96
optimizer	Adam
Loss function	MSE
Activation function	ReLU
Dropout	0.1
Batch size	10
Learning rate	0.0001
Number of epoch	100

boosting (XGB). Linear learning method, linear discriminant analysis (LDA), and quadratic discriminant analysis (QDA). Methods based on artificial neural networks, such as multi-layer perceptron (MLP), etc.

KNN: The anomaly score is usually calculated according to the average distance from each data to its k neighbors. In this paper, $k = 5$, weight = uniform, algorithm = kd_tree, and leaf size = 30.

LOF: The sample whose density is much lower than that of its neighbors is considered as an outlier. In this paper, nearest neighbor = 20, algorithm = auto, contamination = 0.1 and novelty = True.

GNB: Assuming that the conditional probability of each characteristic dimension of the sample obeys Gaussian distribution, the posterior probability of the new sample belonging to each category under a certain characteristic distribution is calculated according to the Bayesian formula, and the sample category is determined by maximizing the posterior probability. The parameters are set to the default value.

RF: It is a classifier that uses multiple decision trees. In this paper, n_estimators = 1000.

DT: The selected parameters are entropy for impurity selection, random_state = 30 in the branch, and splitter = random.

XGB: It is an optimized distributed gradient enhancement library. In this paper, the parameters of XGB are max_depth = 5, eta = 2, gamma = 4, min_child_weight = 6, subsample = 0.8, silent = 0, num_round = 50, logistic = 50.

LDA: Project the training samples on a line, with the projection points of similar samples as close as possible and the projection points of heterogeneous samples as far as possible. The parameters are set to default value.

QDA: It is a variant of LDA that allows the nonlinear separation of data. Floating point number = 0.0, store_covariance = false, and tol = 1e−4.

MLP: The number of hidden layer neurons is 15, solver = lbfgs, and L2 regularization = 1e−5.

4.4.2. AE variants

Experiment II: To compare the anomaly detection performance of different AE variants with the proposed model on the 1553B bus RT spoofing attack testing dataset. AE variants include denoising autoencoder (DAE), SAE, stacked autoencoder (Stacked AE), variational autoencoder (VAE), etc. More complex models are derived from these variants, such as deep autoencoder (Deep AE), sparse deep autoencoder (SDAE), sparse deep denoising autoencoder (SDDAE), etc.

DAE: The input is Gaussian noise or damaged data of the dropout layer. The structure of the DAE is consistent with that of the SAE module mentioned in this paper, but a noise factor of 0.5 is added to the DAE.

SAE: It is consistent with that of the SAE module proposed in this paper, and the sparse part is added with L1 regularization of 10e−5.

Stacked AE: Using layer-by-layer greedy training, multiple AEs with the same structure are connected in sequence. It is different from the SAE module proposed in this paper in that it has no sparse parts.

VAE: The loss of VAE includes the sum of MSE reconstruction loss and KL loss. The input layer includes $x = 12$, z_mean = 2, z_log_var = 2, and the other hyper-parameters are consistent with the SAE hyper-parameters proposed in this paper.

Deep AE: Multiple AEs are superimposed, and the hyper-parameters are consistent with the SAE in this paper, but it lacks a sparse part.

SDAE: The sparse part is added based on the deep autoencoder through an L1 regularization = 10e−5.

SDDAE: The noise part is added on the basis of SDAE, which is realized by adding noise_factor = 0.5.

4.4.3. The state-of-the-art methods

Experiment III: To demonstrate the superiority of the proposed method, it is compared with several state-of-the-art anomaly detection methods. DAGMM is combining AE with the Gaussian mixture model (Zong et al., 2018), LSTM-AE combined with a LSTM network for processing time series data (Elsayed et al., 2020), multi-stage globally trainable convolutional neural network (CNN) (Krizhevsky et al., 2017), and the derivative models of CNN combined with AE: convolutional autoencoder (CAE) (Aytekin et al., 2018), convolutional denoising autoencoder (CDAE) (Du et al., 2016).

DAGMM uses AE to generate a low-dimensional embedding vector of the input and a reconstruction error vector and then uses a Gaussian mixture model to estimate the density of this vector. The neurons in the hidden layer of the compression network are set to [60, 30, 10, 1], the neurons in the hidden layer of the estimation network are [10, 4], the dropout is 0.1, and the random number is 1111.

LSTM-AE is to encode LSTM into compressed value, and then LSTM decodes and reconstructs the input. Timestep = 1, the input shape is [timesteps, 12], and the other structures are same to the proposed method.

CNN includes convolution layers, dropout layers, flatten layers, and dense layers. Convolution layer = 128, kernel = 2, dropout = 0.05, and dense layer units = 2.

CAE: The encoder layers consist of convolution layers, max-pooling layers, and dense layers, in which convolution layer = 128, kernel = 2, holes = 1, padding = cause, convolution step = 1, and L2 regularization is used to prevent overfitting. Max-pooling = 2, and convolution timestep = 2. The unit of the dense layer is 1. The decoder layer is composed of upper sampling layers, convolution layers, flatten layers, and dense layers. The units of the upper sampling layer are 3 and 2. The corresponding parameters of the convolution layer and dense layer are consistent with those of the convolution layer and dense layer in the encoder layer.

The structure of the CDAE model is consistent with the CAE model, noise = 0.5. Filter = 128, dilation_rate = 1, kernel = 2, convolution step = 1, and the padding is causal.

4.4.4. Ablation research

The complete model proposed in this paper includes the reconstruction error of SAE and the IMD of the output characteristics of each layer of its encoder layer and decoder layer. To measure the effect of each component on the complete model, it is necessary to remove the components of each stage from the proposed model sequentially and compare them with the complete model, name these models as follows.

- Mahal: Model only has Mahalanobis distance.
- SAE: Only sparse autoencoder model is available.
- Encoder: The model only includes the characteristic information of each layer of the SAE encoder layer, which abbreviated is enc.
- Decoder: The model that only contains the characteristic information of each layer of the SAE decoder layer, which abbreviated is dec.
- IMD_enc: IMD of the output characteristics of each encoder layer of the SAE.
- IMD_dec: IMD of the output characteristics of each decoder layer of the SAE.
- MD_SAE: The model combines the reconstruction error of SAE and MD.
- Eec_dec: Model of SAE encoding and decoding output characteristics of each layer.
- SAE_enc: A model combining the reconstruction error of SAE with the output characteristics of each layer of its encoder layer.
- SAE_dec: A model combining the reconstruction error of SAE with the output characteristics of each layer of its decoder layer.
- IMD_enc_dec: IMD of output characteristics of SAE encoder layer and decoder layer.
- IMD_SAE_enc: IMD combines the reconstruction error of SAE with the output characteristics of each layer of its encoder layer.
- IMD_SAE_dec: IMD combines the reconstruction error of SAE with the output characteristics of each layer of its decoder layer.
- SAE_enc_dec: It combines the reconstruction error of SAE with the output characteristics of each layer of its encoder layer and decoder layer.

- IMD_SAE_enc_dec: It combines the SAE reconstruction error with the IMD of each output characteristics of its encoder and decoder layer.

5. Results and discussion

5.1. Experimental comparison results

After Experiment I, our proposed method and the traditional ML algorithms anomaly detection on the traffic data of the preprocessed RT spoofing attack testing dataset of 1553B bus respectively, and the bar comparison of performance evaluation metrics are shown in Fig. 7.

As can be seen from the above figure, compared to other ML algorithms, our proposed method is competitive in AC, AUC, AP, and MCC. In contrast, XGB achieves the optimal values for PR, F1, and Gmean, because XGB is less affected by data distribution, and has relatively better classification results due to its robust treatment of data relationships, distribution, and the number of hyper-parameters that can be fine-tuned. LOF has the worst performance in all metrics, which may be because LOF relies too much on the characteristic selection and design of researchers and is sensitive to parameter selection. All the other shallow ML algorithms get AUC less than or equal to 0.5, AP less than 0.1, and MCC less than 0.5 when detecting the anomaly of RT spoofing attacks on the 1553B bus, indicating their poor ability to identify malicious data for such high-dimensional traffic. The lower the FPR the better, and the higher the TPR the better, and our proposed method has reached the best value in both performance metrics. It can be observed that, compared with all metrics, the overall performance of our proposed method is significantly superior to other ML algorithms.

The results of comparing the performance of our proposed method with that of various AE variants of the folding diagram can be obtained from experiment II, as shown in Fig. 8.

As can be seen, compared to other variants of AE, SDAE has the highest AC, PR, and TPR for detecting RT spoofing attacks on the 1553B bus, and the FPR reaches the minimum value of 0.1579 among the variants of AE, indicating that SDAE has the best detection performance, and realizes the accurate classification of normal and malicious data. In addition, SAE has achieved the highest values of F1, AUC, AP, Gmean, and MCC, indicating that SAE is the best at identifying intrusions regardless of whether the data distribution is balanced or not. The detection performance of SAE and SDAE is comparable because they both add sparsity constraints to AE, which is effective in detecting malicious traffic data.

Although these variants of AE show outstanding ability in the most basic metrics, it is noteworthy that our proposed method shows higher AC, F1, AUC, AP, Gmean, and FPR is far lower than other models, presenting an incomparable advantage in anomaly detection abilities. This is because our proposed model not only takes into account the reconstruction error of SAE input and output but also combines the IMD of characteristic information of each layer of SAE, which makes the score distribution of normal and abnormal samples more distant, which can solve the problem of high similarity between some malicious and normal traffic and detects malicious traffic data more accurately.

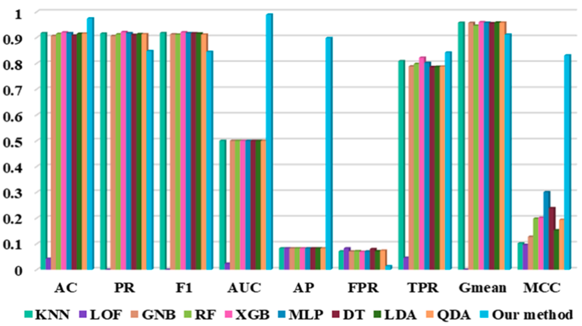


Fig. 7. Comparison of performance metrics between various ML algorithms and the proposed method.

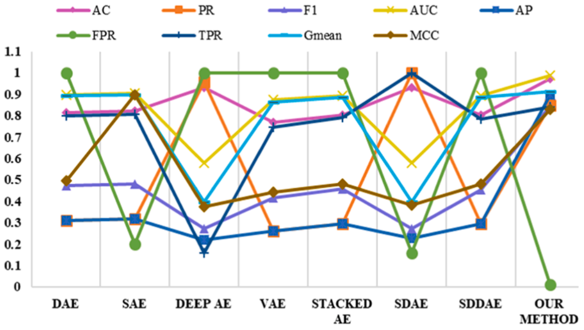


Fig. 8. Comparison of performance metrics between various AE variants and the proposed method.

After the implementation of experiment III, the bar chart of anomaly detection results between the state-of-the-art methods and our proposed method is shown in Fig. 9.

As can be seen from the figure, the PR, F1, Gmean, and TPR obtained by CAE and CDAE reach the best values among all models, which may be because the CNN with AE and its variants combines the advantages of both models, it is able to extract the two-dimensional spatial structure characteristics in the communication traffic data using the convolution kernel, and can unsupervised encoding and decoding of the characteristics using AE and its variants, making the detection performance to be significantly improved. However, the AUC and AP are only 0.5, indicating a poor actual binary classification ability. CNN is relatively difficult to obtain useful characteristics from communication traffic data containing complex nonlinear relationships compared with these two models, and the PR, F1, and TPR are slightly inferior, and the FPR is too high to meet our expectations. LSTM-AE can combine the memory ability of long-term time series dependence and the advantage of AE in characteristic extraction to obtain higher AC, PR, and TPR. However, compared with the method proposed in this paper, F1, AUC, AP, Gmean, and MCC are all inferior and their advantages are not obvious. DAGMM is different to keep the original characteristic space topology, and it is highly dependent on tuning parameters and prior probability, resulting in high FPR. In summary, our proposed method has the best comprehensive performance, the smallest FPR, and a competitiveness that cannot be ignored.

According to experimental IV to carry out the ablation research, the anomaly scores of all components are input into the logistic regression classifier, and the accuracy of predicting the corresponding categories of input samples can be explained by various performance evaluation metrics. The evaluation results of the anomaly detection performance of the proposed method and all components in the preprocessed 1553B bus RT spoofing attack testing dataset are shown in Table 3, and the

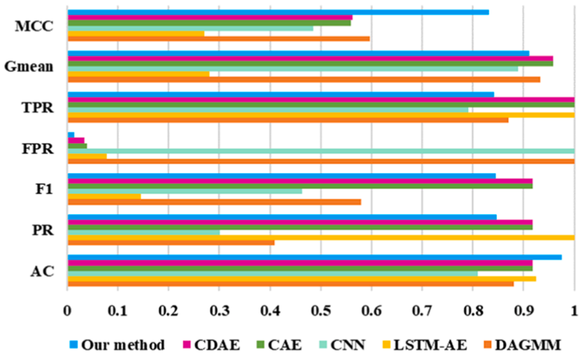


Fig. 9. Comparison of performance metrics between the state-of-the-art methods and our method.

Table 3
Comparison of performance metrics between our proposed method and its ablation research.

Methods	AC	PR	F1	Gmean	MCC	FPR	TPR
MD	0.9153	0.4811	0.3953	0.5697	0.3580	0.0325	0.3355
SAE	0.8230	0.3180	0.4825	0.8984	0.8984	0.2017	0.8071
Encoder	0.9696	0.8692	0.8014	0.8579	0.7879	0.0181	0.7434
Decoder	0.9175	0.0401	0.0358	0.0473	0.0275	0.0192	0.0429
IMD_enc	0.9745	0.8477	0.8449	0.9114	0.8310	0.0146	0.8421
IMD_dec	0.9153	0.4811	0.3953	0.5697	0.3580	0.0325	0.3355
MD_SAE	0.9153	0.4811	0.3953	0.5697	0.3580	0.0325	0.3355
Enc_dec	0.9701	0.8760	0.8043	0.8581	0.7913	0.0195	0.7434
SAE_enc	0.9696	0.8692	0.8014	0.8579	0.7879	0.0167	0.7434
SAE_dec	0.9175	0.0147	0.0258	0.0314	0.0165	0.0158	0.0752
IMD_enc_dec	0.9745	0.8477	0.8449	0.9114	0.8310	0.0139	0.8421
IMD_SAE_enc	0.9745	0.8477	0.8449	0.9114	0.8310	0.0139	0.8421
IMD_SAE_dec	0.9153	0.4811	0.3953	0.5697	0.3580	0.0325	0.3355
SAE_enc_dec	0.9701	0.8760	0.8043	0.8581	0.7913	0.0195	0.7434
IMD_SAE_enc_dec	0.9756	0.8477	0.9058	0.9503	0.8310	0.0136	0.9211

numbers in bold are the best performance values. The comparison chart of the ROC curve and PR curve are shown in Figs. 10 and 11.

From Table 3, Figs. 10, and 11, it can be seen that if the MD model is simply used, the results of all the metrics except AC and AUC are poor, which is because MD-based anomaly discrimination usually assumes that the data conform to a normal distribution, and when the test data deviate more from the normal distribution, the normal points will not all be concentrated near the data center, leading to an increase in the FPR. The highest FPR is obtained using only the SAE anomaly detection model, which shows that using the input-output error of the unsupervised learning algorithm as the anomaly score for anomaly discrimination is not necessarily optimal. Compared with the MD_SAE module, the IMD_SAE_enc module shows a significant improvement in all performance metrics, for example, the AC and TPR are improved by nearly 6.47 % and 1.5 times, respectively, while the performance metrics of the IMD_SAE_dec module and the MD_SAE module are similar, thus showing that the IMD_SAE_enc module demonstrates a higher contribution. In contrast, our proposed method has achieved very encouraging results in terms of AC, F1, AUC, AP, Gmean, FPR, and TPR. Compared with only using the SAE model, our proposed method improves about 18.54 %, 87.73 %, and 14.12 % in each of the AC, F1, and TPR respectively, while the FPR is reduced by nearly 93.26 %. It shows that the combined SAE input and output reconstruction error and IMD with each layer of characteristic information can better improve the detection accuracy of a few attacks and the performance is substantially improved.

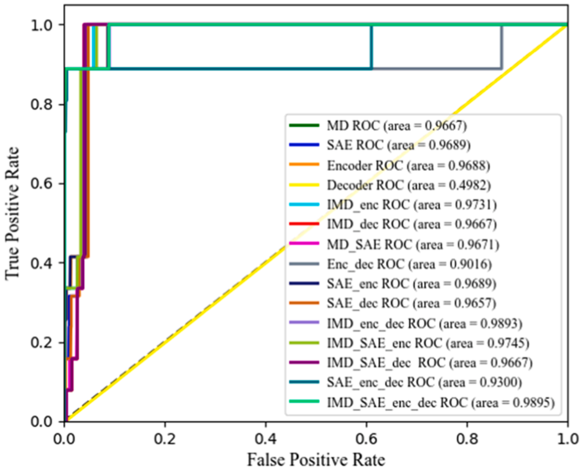


Fig. 10. Comparison of ROC curves between the proposed method and its ablation research.

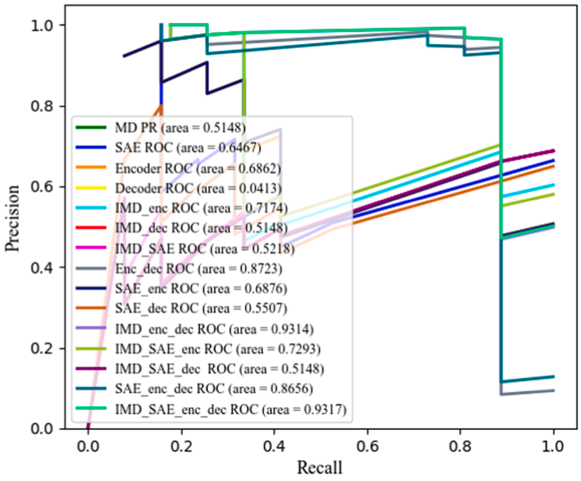


Fig. 11. Comparison of PR curves between the proposed method and its ablation research.

5.2. *Comparison with the latest solutions*

Onodueze et al. (2020) researched anomaly detection in 2020 based on the 1553B bus RT spoofing attack dataset used in this paper, using ML algorithms including bi-directional long short-term memory (BiLSTM) network for identifying context anomalies of time series data. OCSVM, is a kind of incremental semi-supervised algorithm based on the maximum edge method. Isolated forest (IF) based on decision trees without calculating the distance and density, the minimum covariance determinant (MCD) of the correlation between different characteristics can be calculated. The comparison of each performance evaluation result obtained with the method proposed in this paper is shown in Table 4, and the numbers in bold are the best performance values.

As can be seen from Table 4, Onodueze et al. (2020) did not preprocess the non-periodic fake messages on the 1553B bus, and the skewed class distribution caused difficulties in data training, resulting in test results that were biased toward a larger proportion of normal data, which was seriously inconsistent with the actual results. Although BiLSTM obtained the highest F1 and TPR, but Gmean, MCC, FPR, and AUC are the worst, making it difficult to accurately detect abnormal messages. The performance of IF and MCD was comparable, and poor results are obtained in the three metrics of Gmean, MCC, and AUC, indicating that both models were unable to classify severely imbalanced data. In contrast, OCSVM performs moderately well, but it is still slightly inferior to the method proposed in this paper. It can be seen that if the unbalanced data are not preprocessed, the intrusion cannot be correctly identified when the model is trained using only benign messages.

Table 4
Performance comparison of the proposed 1553B bus anomaly detection with the latest solutions.

Author(s), Ref.	Methods	AC	PR	F1	Gmean	MCC	FPR	TPR	AUC
Onodueze et al. (2020)	BiLSTM	0.8899	0.8899	0.9417	0.0000	0.0000	1.0000	1.0000	0.5000
	One-Class SVM	0.8231	1.0000	0.8896	0.8951	0.5544	0.0000	0.8012	0.9000
	Isolation Forest	0.8269	0.8825	0.9052	0.0000	−0.0912	1.0000	0.9292	0.4600
	MCD	0.8163	0.8800	0.8982	0.0000	0.0991	1.0000	0.9172	0.4600
Our proposed method	IMD_SAE_enc_dec	0.9756	0.8477	0.9058	0.9503	0.8310	0.0136	0.9211	0.9895

Table 5
Heuristic multi-threshold calculation results of different modules of IMD and SAE.

Methods	$F1_{best}$	TH of the $F1_{best}$	$Youden_{best}$	TH of the $Youden_{best}$	AC_{best}	TH of the AC_{best}	TH of the averages
SAE	0.8150	0.2587	0.9592	0.2587	0.9625	0.2603	0.2592
MD	0.7979	0.5778	0.9544	0.5778	0.9555	0.5572	0.5709
IMD_enc	0.8920	0.7565	0.9041	0.0500	0.9826	0.6921	0.4995
IMD_dec	0.8150	0.3124	0.9592	0.3124	0.9625	0.3011	0.3086
IMD_SAE	0.8150	0.3124	0.9592	0.3124	0.9625	0.2925	0.3058
IMD_enc_dec	0.8920	0.7614	0.9130	0.0575	0.9832	0.7591	0.5260
IMD_SAE_enc	0.8920	0.7586	0.8992	0.2382	0.9826	0.6787	0.5585
IMD_SAE_dec	0.8150	0.3122	0.9592	0.3122	0.9625	0.3007	0.3084
IMD_SAE_enc_dec	0.8920	0.6503	0.9166	0.6001	0.9832	0.6503	0.6336

5.3. Anomaly discrimination analysis

Based on the heuristic multi-threshold selection method proposed in section 3.4, when the distribution of normal and abnormal data is relatively balanced, the threshold for abnormal discrimination should be set to the threshold corresponding to the maximum accuracy. When the data distribution is unbalanced, the YI and F1 are considered together, and the weights of these two metrics depend on the decision-makers, and the results of the anomaly discrimination threshold with β=1 are given in this paper. Based on the preprocessed RT spoofing attack data of the 1553B bus, the combination of IMD with SAE's internal components and its input-output reconstruction errors, respectively, as well as the best F1, YI, accuracy, and the corresponding thresholds obtained by SAE and MD. In the case of balanced distribution of normal and abnormal data, the accuracy, YI, and F1 are equally important, so the average values of the thresholds corresponding to the best performance metrics can be obtained, as shown in Table 5.

From the above table, we can see that the average thresholds of SAE, MD, and IMD_SAE_enc_dec are 0.2592, 0.5709, and 0.6336, respectively. According to the criterion of anomaly score and threshold, if the anomaly score is greater than the threshold, the sample is regarded as an anomaly. Therefore, in order to verify the accuracy of the heuristic multi-threshold selection method proposed in this paper on the anomaly discrimination of the proposed method, the preprocessed 1553B spoofing attack testing dataset is input into SAE, MD, and the proposed method, the anomaly scores of SAE, MD and the comprehensive anomaly score of the method proposed in this paper can be obtained respectively, compared with their respective average thresholds, and the anomaly discrimination results are shown in Fig. 12. The blue circle and orange triangle are normal and abnormal data points respectively, and the red lines are the set threshold.

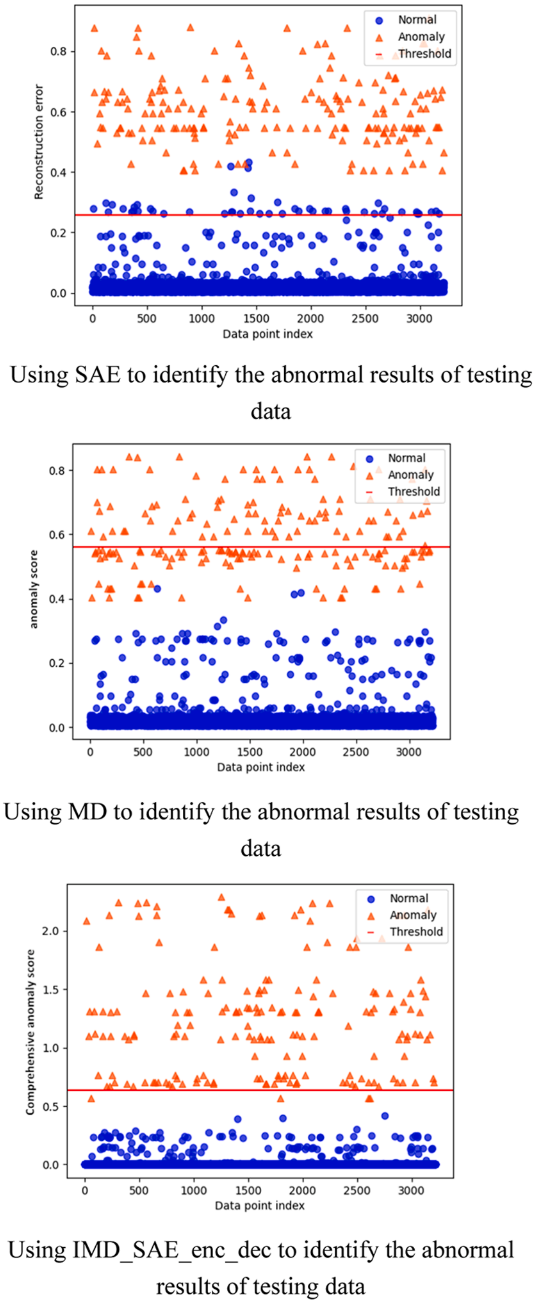
In Fig. 12(a), the anomaly detection of test data using the SAE model has a high reconstruction error for abnormal data and a low reconstruction error for normal data, which is consistent with the theory, but a small amount of normal data are still confused with abnormal data and cannot be significantly separated, which may be caused by the high similarity between some abnormal data and normal data. According to the red line of the threshold, some normal data exceed the selected threshold, which leads to the misjudgment of normal data and affects the accuracy of the detection results. From Fig. 12(b), if only MD is used for anomaly detection, the anomaly scores of many abnormal data are lower than the threshold, and the abnormal is classified as normal, which will interfere with the judgment of the airborne network security administrator, and may cause the paralysis of the aircraft bus system, seriously affect the aircraft safety.

It can be seen that the distance metric method by assessing characteristic similarity alone is more important than the characteristic extraction method for evaluating the reconstruction error. In Fig. 12(c), the CAS in the combined method of SAE and IMD proposed in this paper is higher for abnormal data and lower for normal data, and only a few abnormal data are classified as normal for the method selection corresponding to the optimal threshold. In summary, based on the heuristic multi-threshold selection method, compared with SAE and MD, the thresholds in the proposed method can accurately judge the CAS of normal and malicious data, with low FPR and excellent detection effects.

6. Conclusion and future work

We proposed an anomaly detection method based on a combination of SAE and IMD and a heuristic multi-threshold selection method to identify malicious communication traffic on the MIL-STD-1553B bus in IAS. Not only takes into account the input and output reconstruction error of SAE but also the characteristic information contained in each layer of SAE is considered, which improves the detection effect of RT spoofing attacks. Compared with many methods, the FPR is low, only 1.36 %, and multiple detection metrics are above 90 %, with the highest detection accuracy. Moreover, the multi-threshold selection method has outstanding flexibility and can effectively distinguish malicious traffic, providing a new idea for anomaly detection of aviation bus systems. Our method can be applied to anomaly detection in aviation bus network, vehicle network, and industrial Internet.

The future research directions are: attacks based on bus communication packets include spoofing attacks, replay attacks, and DoS attacks, among which spoofing attacks include faked RT spoofing attacks and compromised RT spoofing attacks, while we only study RT spoofing attacks. Therefore, in the future, we will begin to study the anomaly



(a) Using SAE to identify the abnormal results of testing data

(b) Using MD to identify the abnormal results of testing data

(c) Using IMD_SAE_enc_dec to identify the abnormal results of testing data

Fig. 12. Comparison of anomaly discrimination results.

detection methods for other attack methods and make risk assessment and analysis of the threats that the aviation bus systems may face.

CRediT authorship contribution statement

Huang Li: Conceptualization, Methodology, Software, Data curation, Writing – original draft. **Yiqin Sang:** Conceptualization, Methodology, Software, Data curation, Writing – original draft. **Hongjuan Ge:** Supervision. **Jie Yan:** Supervision, Writing – review & editing. **Shijia Li:** Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (NSFC) under Grant [U2133203] and [U2233205].

References

Aytekin, C., Ni, X., Cricri, F., Aksu, E., 2018. Clustering and unsupervised anomaly detection with l2 normalized deep auto-encoder representations. In: 2018 International Joint Conference on Neural Networks (IJCNN). IEEE, pp. 1–6.

Chen, Z., Chai, K., Lee, B., Lau, C., 2018a. Autoencoder-based network anomaly detection. In: 2018 Wireless Telecommunications Symposium (WTS). IEEE, pp. 1–5.

Chen, Z., Yan, Q., Han, H., Wang, S., Peng, L., et al., 2018b. Machine learning based mobile malware detection using highly imbalanced network traffic. Inf. Sci. 433-434, 346–364.

Choi, H., Kim, M., Lee, G., Kim, W., 2019. Unsupervised learning approach for network intrusion detection system using autoencoders. J. Supercomput. 75 (9), 5597–5621.

Denouden, T., Salay, R., Czarnecki, K., Abdelzad, V., Phan, B., Vernekar, S., 2018. Improving reconstruction autoencoder outof-distribution detection with mahalanobis distance. arXiv preprint arXiv:1812.02765 1–9.

Du, B., Xiong, W., Wu, J., Zhang, L., Zhang, L., Tao, D., 2016. Stacked convolutional denoising auto-encoders for characteristic representation. IEEE Trans. Cybernet. 47 (4), 1017–1027.

Elsayed, M., Le-Khac, N., Dev, S., Jurcut, A., 2020. Network anomaly detection using LSTM based autoencoder. In: Proceedings of the 16th ACM Symposium on QoS and Security for Wireless and Mobile Networks, pp. 37–45.

Genereux, S., Lai, A., Fowles, C., Roberge, V., Vigeant, G., Paquet, J., 2019. Maidens: mil-std-1553 anomaly-based intrusion detection system using time-based histogram comparison. IEEE Trans. Aerosp. Electron. Syst. 56 (1), 276–284.

Guo, J., Liu, G., Zuo, Y., Wu, J., 2018. An anomaly detection framework based on autoencoder and nearest neighbor. In: 2018 15th International Conference on Service Systems and Service Management (ICSSSM). IEEE, pp. 1–6.

Habler, E., Shabtai, A., 2021. Analyzing sequences of airspace states to detect anomalous traffic conditions. IEEE Trans. Aerosp. Electron. Syst. 58 (3), 1843–1857.

He, D., Liu, X., Zheng, J., Chan, S., Zhu, S., Min, W., et al., 2020a. A lightweight and intelligent intrusion detection system for integrated electronic systems. IEEE Netw. 34 (4), 173–179.

He, D., Qiao, Q., Gao, J., Chan, S., Zheng, K., Guizani, N., 2020b. Simulation design for security testing of integrated electronic systems. IEEE Netw. 34 (1), 159–165.

He, D., Gao, Y., Liu, X., Chan, S., Cheng, Y., Liu, X., et al., 2020c. Design of attack and defense framework for 1553B-based integrated electronic systems. IEEE Netw. 35 (4), 234–240.

Imani, M., 2019. Difference-based target detection using mahalanobis distance and spectral angle. Int. J. Remote Sens. 40 (3), 811–831.

Krizhevsky, A., Sutskever, I., Hinton, G., 2017. ImageNet classification with deep convolutional neural networks. Commun. ACM. 60 (6), 84–90.

Krueger, D., Maharaj, T., Kramár, J., Pezeshki, M., Ballas, N., Ke, N., 2016. Zoneout: Regularizing rnns by randomly preserving hidden activations. arXiv:1606.01305. [Online] Available: https://arxiv.org/abs/1606.01305.

Li, X., Meng, C., Wang, C., Chen, Z., 2018. Research on equipment status and operation information acquisition based on equipment control bus. In: International Conference on Frontier Computing: Theory, Technologies and Applications, pp. 864–871.

Li, B., Peng, L., Dai, F., 2022. Abnormal network traffic detection method combining mahalanobis distance and autoencoder. Comput. Eng. 48 (04), 133–142.

Losier, B., Smith, R., Roberge, V., 2019. Design of a Time-based Intrusion Detection Algorithm for the mil-std-1553. Royal Military College of Canada, Kingston, p. 2102. Project number DTAES-8.

NG, 2011. Sparse autoencoder. CS294A Lect. Notes 72 (1), 1–19.

Nguimbous, Y., Ksantini, R., Bouhoula, A., 2019. Anomaly-based intrusion detection using autoencoder. In: 2019 International Conference on Software, Telecommunications and Computer Networks (SoftCOM). IEEE, pp. 1–5.

Onodueze, F., Josyula, D., 2020. Anomaly detection on MIL-STD-1553 dataset using machine learning algorithms. In: 2020 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom). IEEE, pp. 592–598.

Park, J., Choi, D., Jeon, Y., Nam, Y., Hong, M., Park, D., 2018. Network anomaly detection based on probabilistic analysis. Soft Comput. 22 (20), 6621–6627.

Qiao, Q., He, D.J., Gao, Y., Zhu, S., Gao, J., Chan, S., 2020. Hybrid intrusion detection mechanisms for integrated electronic systems. In: 17th Annual International Conference on Sensing, Communication, and Networking (SECON). IEEE, pp. 1–9.

Rumelhart, D., Hinton, G., Williams, R., 1986. Learning representations by back-propagating errors. Nature. 323 (6088), 533–536.

Ryu, S., Yim, J., Seo, J., Yu, Y., Seo, H., 2022. Quantile autoencoder with abnormality accumulation for anomaly detection of multivariate sensor data. IEEE Access. 10, 70428–70439.

Salahuddin, M., Pourahmadi, V., Alameddine, H., Bari, M., Chronos, B.R., 2021. Ddos attack detection using time-based autoencoder. IEEE Trans. Netw. Serv. Manag. 19 (1), 627–641.

Santo, D., Malavenda, C.S., Romano, S.P., Vechio, C., 2021. Exploiting the mil-std-1553 avionic data bus with an active cyber device. Comput. Secur. 100, 102097.

Stan, O., Elovici, Y., Shabtai, A., Shugol, G., Tikochinski, R., Kur, S., 2017. Protecting military avionics platforms from attacks on mil-std-1553 communication bus. arXiv preprint arXiv:1707.05032 1–15.

Stan, O., Cohen, A., Elovici, Y., Shabtai, A., 2018. On the security of mil-std-1553 communication bus. In: Security and Safety Interplay of Intelligent Software Systems: ESORIC 2018 International Workshops, ISSA, pp. 153–171.

Stan, O., Cohen, A., Elovici, Y., Shabtai, A., 2020. Intrusion detection system for the MIL-STD-1553 communication bus. IEEE Trans. Aerosp. Electron. Syst. 56 (4), 3010–3027.

Tun, M., Nyaung, D., Phyu, M., 2020. Network anomaly detection using threshold-based sparse autoencoder. In: Proceedings of the 11th International Conference on Advances in Information Technology, pp. 1–8.

Utkin, L., Zaborovsky, V., Lukashin, A., Popov, S., Podolskaja, A., 2017. A Siamese autoencoder preserving distances for anomaly detection in multi-robot systems. In: 2017 International Conference on Control, Artificial Intelligence, Robotics & Optimization (ICCAIRO). IEEE, pp. 39–44.

Vilaça, E., Vieira, T., Sousa, R., Costa, J., 2019. Botnet traffic detection using RPCA and mahalanobis distance. In: 2019 Workshop on Communication Networks and Power Systems (WCNPS). IEEE, pp. 1–6.

Wang, L., Wang, J., Ren, Y., Xing, Z., Li, T., Xia, J., 2021. A shadowed rough-fuzzy clustering algorithm based on mahalanobis distance for intrusion detection. Intell. Autom. Soft Comput. 30 (1), 31–47.

Wu, Z., Guo, A., Yue, M., Liu, L., 2020. An ADS-B message authentication method based on certificateless short signature. IEEE Trans. Aerosp. Electron. Syst. 56 (3), 1742–1753.

Wu, Z., Liang, C., Zhang, Y., 2023. Blockchain-based authentication of GNSS civil navigation message. IEEE Trans. Aerosp. Electron. Syst. 59 (4), 4380–4392.

Yahalom, R., Barishev, D., Steren, A., Nameri, Y., Elovici, Y., 2019a. Datasets of RT spoofing attacks on MIL-STD-1553 communication traffic. Data Brief. 23, 103863.

Yahalom, R., Barishev, D., Steren, A., Nameri, Y., 2019b. RT spoofing attacks on MIL-STD-1553 communication traffic. Mendeley Data 3. https://doi.org/10.17632/jvgdrmjvs3.3.

Yan, B., Han, G., 2018. Effective characteristic extraction via stacked sparse autoencoder to improve intrusion detection system. IEEE Access. 6, 41238–41248.

Yokkampon, U., Mowshowitz, A., Chumkamon, S., Hayashi, E., 2022. Robust unsupervised anomaly detection with variational autoencoder in multivariate time series data. IEEE Access. 10, 57835–57849.

Zhang, Y., Han, Y., Wang, C., Wang, J., Zhao, Q., 2022. A dynamic threshold method for wind turbine fault detection based on spatial-temporal neural network. J. Renew. Sustain. Energy. 14 (5), 053304.

Zheng, J., Qu, H., Li, Z., Li, L., Tang, X., Guo, F., 2022. A novel autoencoder approach to characteristic extraction with linear separability for high-dimensional data. PeerJ Comput. Sci. 8, e1061.

Zong, B., Song, Q., Min, M., Cheng, W., Lumezanu, C., Cho, D., et al., 2018. Deep autoencoding Gaussian mixture model for unsupervised anomaly detection. In: International Conference on Learning Representations, pp. 1–19.

Huang Li was born in Shanxi, China, in 1992. She received the B.S. degree in electrical engineering and its automation from the North China University of Science and Technology, China, in 2016, and the M.S. degree in control engineering from Shanxi University, China, in 2018. She is currently pursuing the Ph.D. degree in airworthiness technology and management with the Nanjing University of Aeronautics and Astronautics, China. Her current research interests include deep learning, machine learning, airborne network intrusion detection, and safety risk assessment of airborne network systems.

Yiqin Sang was born in Jiangsu, China, in 1997. She received the B.S. degree in aircraft airworthiness technology from the Nanjing University of Aeronautics and Astronautics, China, in 2019, where she is currently pursuing the Ph.D. degree in airworthiness technology and management. Her current research interest includes safety analysis of airborne power systems.

Hongjuan Ge was born in Jiangsu, China, in 1966. She received the B.S. and M.S. degrees in electrical engineering from Southeast University, Nanjing, China, in 1985 and 1988, respectively, and the Ph.D. degree in electric machines and electric apparatus from the Nanjing University of Aeronautics and Astronautics, China, in 2006. She is currently a Full Professor with the College of Civil Aviation and the Department of Automation. Her current research interests include the space-vector control of PWM, AC-AC converters, and airworthiness technology.

Jie Yan received the B.S. and M.S. degree in aircraft airworthiness technology from the Nanjing University of Aeronautics and Astronautics, China, in 2020 and 2023, now she works in the 54th Research Institute of China Electronics Technology Group Corporation. Her current research interest includes safety analysis technology of airborne network.

Shijia Li received the B.S. degree in traffic information Control engineering from the Tongji University, China, in 2022, where she is currently pursuing the M.S. degree in airworthiness technology and management in the Nanjing University of Aeronautics and Astronautics, China. Her current research interest includes safety analysis technology of airborne network.