Complex & Intelligent Systems
https://doi.org/10.1007/s40747-024-01388-1

ORIGINAL ARTICLE

Meta learning-based few-shot intrusion detection for 5G-enabled industrial internet

Yu Yan[1] · Yu Yang[1] · Fang Shen[1] · Minna Gao[2] · Yuheng Gu[1]

Received: 23 September 2023 / Accepted: 12 February 2024

Abstract
With the formation and popularization of the 5G-enabled industrial internet, cybersecurity risks are increasing, and the limited number of attack samples, such as zero-day, leaves a short response time for security protectors, making it substantially more difficult to protect industrial control systems from new types of malicious attacks. Traditional supervised intrusion detection models rely on a large number of samples for training and their performance needs to be improved. Therefore, there is an urgent need for few-shot intrusion detection. Aiming at the above problems, this paper proposes a detection model based on a meta-learning framework, which aims to effectively improve the accuracy and real-time performance of intrusion detection, and designs a meta-learning intrusion detection model containing a sample generation module, a feature mapping module and a feature metric module. Among them, the sample generation module introduces the residual block into the Natural GAN and proposes a new method to generate high-quality antagonistic samples—Res-Natural GAN, which is used to enhance the antagonism of the generated samples and the feature mining degree, to improve the accuracy of malicious traffic detection; the feature mapping module proposes a new attention mechanism, the multi-head fast attention mechanism, which is applied to the encoder structure of the transformer and combined with a parameter optimization algorithm based on particle swarm mutation to shorten the mapping time and improve the real-time performance of the model while mapping the features effectively; the feature metric module proposes a prototype structure based on a prototype storage update algorithm and combines it with a prototype network to achieve correct classification by measuring the Euclidean distance between the detected samples and the class of prototypes, and to shorten the inference time while ensuring the detection accuracy; finally, the three modules are combined to form a real-time meta-learning intrusion detection model. To evaluate the proposed model, five different types of experiments are conducted on multiple public datasets. The experimental results show that the model has higher detection accuracy than the traditional model for both few-shot and zero-shot malicious attacks, and is not only applicable to 5G-enabled industrial internet, but also generalized to different network environments and attack types.

Keywords 5G-enabled industrial internet · Cyberspace security · Intrusion detection system · Meta-learning · Few-shot detection

Introduction

The fourth revolution of science and technology [1] is sweeping the world, and the technological revolution mainly based on artificial intelligence [2], virtual reality [3], quantum technology [4], and biotechnology [5] is deeply affecting human life. With the advent of Industry 4.0, the Internet and industrial production are increasingly integrated, and the closed environment has been broken, forming a new industrial Internet. 3G and 4G technologies, which have shortcomings such as poor mobility, inflexible networking, high latency, and unreliability, are gradually being replaced by 5G technology [6]. Therefore, 5G-enabled industrial Internet [7] with high capacity, high speed, low power consumption, and good reliability has come into being (Fig. 1).

In recent years, by utilizing the unreliability of information transmission on the industrial Internet [8], numerous malicious attacks have been launched against specific industrial systems, institutions, and regions, which are often new, rare,

✉ Yu Yang
aa18634816079@163.com

1 College of Information Engineering, Chinese People's Armed Police Force Engineering University, Xi'an 710086, ShanXi, China

2 College of Missile Engineering, Rocket Military Engineering University, Xi'an 710086, ShanXi, China

Published online: 25 March 2024

and extremely harmful. The "Stuxnet"virus [9], detected in June 2010, proved to be the world's first "cyber weapon" to be put into actual combat, and was extremely virulent and destructive. The WannaCry "worm-like" ransomware virus [10] outbreak in 2017 has had a huge impact on industrial control systems in hundreds of countries around the world; the"zero-day exploit" [11] was launched on the same day the vulnerability was discovered, and it had a small sample size and a short response time, leaving security decision-makers with far less time to respond. There are a large number of cyberattacks with the same characteristics as the above security threats, which seriously affect the normal functioning of the industrial internet.

To address the above problems, access control [12], data encryption protection, network isolation protection, firewall [13], intrusion detection [14] and other security protection technologies are widely used. Among them, the intrusion detection system can realize the identification and detection of external attacks, internal attacks and misoperation, which is a proactive protection technology. However, the current intrusion detection technology for 5G-enabled industrial internet has more shortcomings, such as the poor stability and robustness of the model, and the low adaptability to the adversarial industrial Internet environment; the model has poor timeliness, and the detection rate of the type of attack that appears for the first time and occurs infrequently, but is extremely harmful is low, and it is not possible to realize the correct identification of few-shot and zero-shot new types of attacks in a short period of time. Existing supervised methods require a large amount of labeled data, which are difficult to obtain in practice; therefore, to address the above problems, a method that can accomplish efficient detection using only a small number of samples is urgently needed.

Due to the consideration of the above problems, the motivation of this paper mainly contains the following four aspects: first, to overcome the difficulties of traditional supervised algorithms that are greatly affected by the number of labelled samples, and for attacks that are difficult to obtain samples, e.g., zero-day, Stuxnet, etc., to propose an effective few-shot intrusion detection method, and to achieve the purpose of correctly identifying the type of malicious attacks in the situation where few-shot or even zero-shot data are available for learning and training; second, to adapt the intrusion detection model to the current real network environment where adversarial attacks are frequent, a method is proposed to enable the model to effectively detect adversarial attacks, to improve the robustness of the model, and to increase the degree of fit with the network environment of dynamic games; third, to reduce the adverse effects of network attacks on 5G industrial internet and shorten the security response time, the constructed model should have a faster inference speed and higher detection efficiency, and shorten the inference time as much as possible while ensuring the detection accuracy, to minimise the network damage; fourthly, to solve the problem of frequent 5G-enabled industrial Internet network attacks and the difficulty of detection and protection, the model should be experimentally verified for effectiveness on the 5G-enabled industrial Internet intrusion detection dataset after the model construction is finished to confirm the fit with the application scenarios.

Inspired by the metric-based meta-learning algorithm strategy [15], we propose a few-shot intrusion detection model based on a meta-learning framework, which includes a sample generation module based on Res-Natural GAN, a feature mapping module based on transformer encoder, and a feature metric module based on a prototype network and a prototype, and is capable of realizing the identification and detection of few-shot and zero-shot attacks, and accomplishing the recognition and detection of attacks within and across the task space. detection, and accomplishes generalization within and across the task space.The main contributions of this paper are as follows:

1. Propose a Natural GAN model based on residual neural network, which generates higher quality of antagonistic samples than the traditional model, and helps to improve the adaptability of intrusion detection system to 5G-enabled industrial internet environment.
2. Propose a transformer model based on multi-head fast attention mechanism, and combine it with a parameter optimisation algorithm based on particle swarm mutation for parameter tuning, which is faster than the self-attention mechanism for training, and is conducive to building a real-time intrusion detection system in 5G-enabled industrial internet environment.
3. Propose a prototype model based on a new prototype storage updating algorithm, and form a closed-loop experimental machine link to complete the updating and broadcasting of the prototype table, shorten the feature metric time, and improve the efficiency of the real-time intrusion detection system.

The paper is organized as follows. The next section presents the related research work. The subsequent section constructs an intrusion detection model based on the meta-learning framework. The penultimate section evaluates the results of five types of experiments. The final section summarizes the existing work in this paper and provides an outlook for future work.

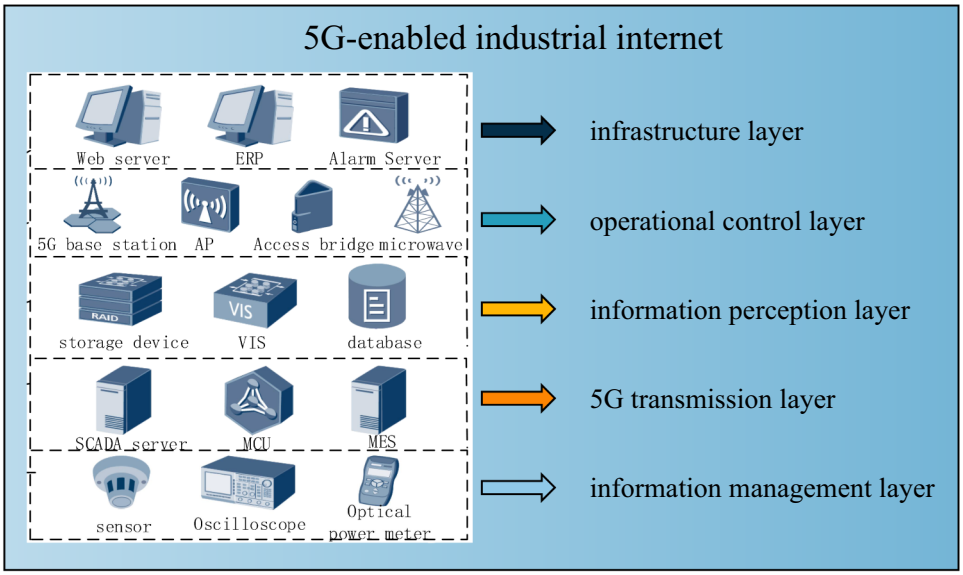

Fig. 1 The architecture of 5G-enabled industrial internet

Background and related works

Generating adversarial samples

The 5G-enabled industrial internet has always been an arena where attackers and defenders confront each other. To adapt to the adversarial environment and simulate the attacker's transformation strategy in a dynamic network environment, adversarial samples can be generated from real samples and fed into the model for training. As shown in Fig. 2, the existing research work can be broadly classified into three categories: (i) simulating the known network structure of a semi-white-box attack [16] in and setting the adversarial objective function against an intrusion detection model or an approximate model stand-in. The research work [17] divided the objective function into two parts, namely the output error value and the sample perturbation value. Then, these two parts were approximated as Taylor first-order expansions before a single-step gradient descent method was employed to quickly generate adversarial samples. However, this FGSM method cannot obtain an optimal solution. Therefore, the research work [18] proposed the BIM method, which iterates the multi-step FGM and strictly restricts the value domain to search for the optimal solution. (ii) Simulating the known input–output mapping part of the semi-black box attack. The mapping metric is converted into a Jacobi matrix, and perturbations are added to the input features that have the greatest impact on the output. The research work [19] proposed the JSMA method to generate the Jacobi matrix between the input and output and construct a saliency map that defines the importance of each pixel in the input image to generate the adversarial samples in a targeted manner. (iii) Simulating the known training data part of the white-box attack by assuming that the model is in a white-box, i.e., the attacker is fully aware of the network structure and training data, which increases the difficulty of classifying the adversarial samples. Research work [20] used GAN to generate new virtual samples with a similar distribution to the real samples to solve the data imbalance problem; combined with a random forest classification model, ablation experiments demonstrated that the use of GAN greatly improved the accuracy of the model detection.

However, adversarial attacks designed specifically to evade intrusion detection may undermine the effectiveness of the proposed model, and it is highly likely that adversarial attacks that have been modified, spoofed, and intelligently operated will bypass the recognition of intrusion detection systems, thus increasing the model false alarm and miss alarm rates, which is not conducive to the security and stability of the network environment. To address the above problems, research work [21] creatively proposes a new method to generate high-quality adversarial samples based on the neural style transfer (NST) algorithm steganography, which employs conditional generative adversarial networks (cGANs) to learn the embedded secret information to achieve the embedding and extraction of secret images, and experiments have proved that the algorithm generates images that achieve a high peak signal-to-noise ratio (PSNR), structural similarity index (SSIM) and visual information fidelity (VIF), which provides a new high-quality method for generating adversarial samples; research work [22] provides a new and effective method for detecting adversarial samples using points of interest in digital video frames to estimate the camera position and orientation, and adopting the Euclidean distance metric feature vectors, and at the same time, using the RANSAC method to remove the camera noise and the environmental noise, experimentally proving the effectiveness of this aspect on different types of videos, which is

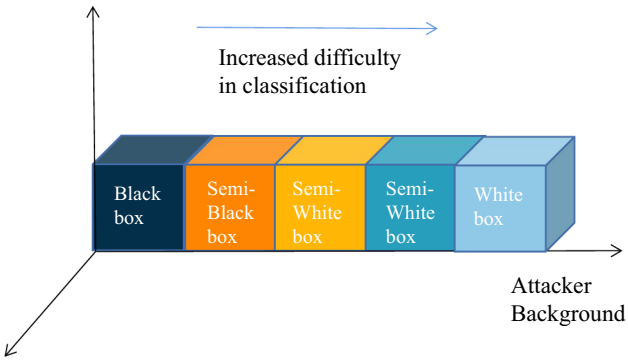

Fig. 2 The schematic diagram of the attack model

highly relevant to fight against malicious attacks on different data types; research work [23] is able to generate high-quality images with high spatio-temporal resolution, and proposes the STF model UAV-Net that contains an improved ResNet (MResNetet), a feature pyramid network (FPN), and a decoder module, which experimentally demonstrates the practicality and effectiveness of the proposed model in generating images on the dataset, providing an efficient and innovative approach for generating adversarial samples in the field of intrusion detection, and thus improving the stability of the model.

In addition, the emergence of adversarial attacks may bring more difficulties for security decision response, making it difficult to formulate decisions to adapt to the real network environment of dynamic games. Research work [24] starts from the purpose of making the strategy results comprehensive and accurate, and facilitating the decision-maker to analyse the problem in depth, introduces the concept of using Einstein operations as aggregation operators (AOs), proposes the definitions and procedures related to the basic bipolar neutral nerves (BN) and the neutral neural set (NS), provides decision-making schemes, and proves that the proposed schemes are effective and practical using the robot selection as an example to explore the problem, a practical fit to the decision-making purpose, providing a valuable reference for decision-making response to adversarial attacks in the field of intrusion response.

Feature extract and comparative classification

Traditional machine learning and deep learning algorithms are still important methods for intrusion detection for 5G-enabled industrial Internet applications. To address the optimization problem of intrusion detection for 5G industrial Internet, the research work [25] proposed a lightweight intrusion detection model based on density-solving fuzzy clustering, and on the ICX dataset, it was experimentally demonstrated that the algorithm greatly improves the detection accuracy and can be applied to 5G industrial Internet. The research work [26] addressed the problem of high complexity of the industrial Internet and the low detection rate and poor real-time performance of existing intrusion detection algorithms. An intrusion detection method based on multi-feature data clustering was proposed, and a unified standard was constructed to classify network nodes and select clustering centres. It was experimentally proven that the accuracy of the proposed method is as high as 97.8% in the industrial Internet. However, the above research work cannot achieve correct classification of few-shot malicious attacks. For this reason, a metric-based meta-learning strategy was used, which can be divided into feature extraction and comparative classification, which focuses on comparative classification.

The commonly used comparative classification networks include Siamese network, matching network, prototype network, and relational network. Research work [27] uses matching networks in the field of natural language query, focuses on the hot issue of code search, creatively proposes a deep neural solution for semantic code retrieval using neural graph matching and a search method to retrieve the best matching code fragments through rich semantic features, experimentally proves that the retrieval effect is substantially improved compared with the baseline model on six representative programming languages and the retrieval accuracy can be as high as 97%, which provides a higher value of reference to the use of matching networks in the field of intrusion detection; research work [28] used nine different meta-learning methods such as relational networks and conducted cross-sectional comparisons on 21 datasets, respectively. It was proven through experiments that relational networks have strong detection accuracy in such meta-learning strategies, but they perform dynamic learning and suffer from long training time and poor training capability for complex samples. To address this issue, research work [29] proposed an end-to-end learning framework called Multi-View Prototype Network (MVPN) to characterize class prototypes of each view of 3D shapes and perform metric analysis. It was experimentally demonstrated on ModelNet and ShapeNet Core55 datasets that the framework achieved higher detection accuracy than traditional deep learning neural networks such as CNN. Meanwhile, the method has low complexity and requires relatively less computational effort, and classification can be achieved by a simple distance calculation between the class prototype and the sample vector. For the sake of model timeliness and the detection rate of new classes, the prototype network in the meta-learning algorithm is chosen in this paper and improved by adding a new discriminator and a prototype machine to improve the usability of the constructed model in the 5G-enabled industrial internet.

As this study involves more deep learning models, and the parameter selection directly affects the classification prediction. Therefore, the study is dedicated to solving the

parameter optimisation problem. Parameter optimisation has been a key task in the fields of machine learning, pattern recognition [30], computer vision, optimisation algorithms [31], etc. Many parameter optimisation algorithms have been proposed, such as gradient-based methods, evolutionary computation methods and Bayesian optimisation methods. However, these methods have limitations, such as slow convergence, tendency to fall into local optimum and sensitivity to initial parameters. To overcome these problems, research work [32] proposed a new algorithm DDMPEA based on selection, crossover and mutation search operations to effectively explore the search space and find the optimal solution, which was tested by tuning four parameters and then proved to be able to give the tuning parameters for effectively exploring the globally optimal solution; and research work [33], for the design of a digital IIR filter that does not fall into local optimum. Diversity Driven Multi-Parent Evolutionary Algorithm with Adaptive Non-Uniform Mutation (DDMPEA-ANUM) is used, and the developed algorithm is tested on 23 benchmark functions to prove that it has the smallest error, which verifies its validity; research work [34] addresses the problems of premature convergence and easy to fall into local optimum solution of evolutionary algorithms using different mutation schemes, such as wavelet mutation, Levy flight mutation, particle swarm optimization based mutation, chaotic mutation, and non-uniform mutation, which are effective in exploring the global optimum solution. Different mutation schemes such as wavelet mutation, Levy flight mutation, particle swarm optimisation based mutation, chaotic mutation and non-uniform mutation are used to improve the evolutionary algorithm, and compared and validated on WSN application scenarios, and the result confirms that the optimal strategy for WSN area coverage optimisation problem—DDMPEA with chaotic mutation—is more effective and practical compared to other techniques. All of the above three research works provide efficient and novel approaches for the parameter optimisation task, which are extremely valuable references for deep learning model parameter tuning and selection in this paper.

Problem formulation

The problem of classification is similar in nature, and it is drawn on the ideas of machine learning algorithms involving feature extraction and classification. If an algorithm can distinguish between different types of data, it can effectively identify and amplify the "differences" regardless of the number of data samples available, thus enabling the model to work in few-shot and zero-shot situations.

Meta-learning can migrate from known tasks to unknown tasks and focuses on the properties of unknown samples. Zero-shot detection is difficult to achieve with machine learning and deep learning, and it can be regarded as a mapping between the feature space and attribute space, where the attribute space is artificially labeled, and the feature space contains extracted features. Based on the above considerations, the best approach for zero-shot learning is to map the data in the feature space and the attribute space to a unified embedding space and then adopt certain metrics to compare the vectors of the same dimension in the embedding space to make classifications.

Traditional machine learning algorithms and deep learning algorithms face large dataset objects. They focus on training and testing for a specific task and only guarantee the correct output for the input in a defined scenario. Consequently, the large dataset is usually divided into a training set and a test set. The training data are used to train the model, including tuning of parameters, etc., while the test data are used to make classifications. It has been proven in research work that this simple data partitioning strategy is suitable for large datasets with a high concentration of samples. However,for few-shot of malicious attacks in networks, this paper introduces a meta-learning algorithm. Meta-learning aims to 'learn to learn' by focusing on the partitioning and training of the task space, and the nature of few-shot indicates that they differ significantly from the way machine learning datasets are partitioned (Fig. 3). The meta-learning algorithm divides the samples into a training set and a test set, where the training set includes the sample set and the query set, and the test set includes the support set and the test set. This division can effectively establish the correspondence between task subsets and is consistent with the working mechanism.

In the field of network intrusion detection, the focus is on classifying normal and attack samples. Traditional machine learning algorithms are often regarded as a classical simple binary classification problem, but the type of attacks corresponding to different scenarios may vary. Here, the application scenario of the meta-learning algorithm is described in this paper. Suppose that there are five types of network traffic: 1, 2, 3, 4, and 5, with 1 representing normal traffic and 2, 3, 4, and 5 representing attack traffic. Specifically, classes 2, 3, and 4 correspond to known malicious attack types with sufficient labeled samples, and class 5 corresponds to novel malicious attack types that have not been trained on the training set. The purpose of the model functions as follows: firstly, to correctly detect the attack traffic of classes 2, 3, and 4 when few-shot are input; secondly, to effectively classify the new malicious attack class 5 with zero -shot. Therefore, the test set task space should contain two corresponding sub-task sets. From the perspective of the N-Way K-shot model, the sample set usually contains N classes of data with K samples in each class, while a certain number of samples are randomly selected from the N classes to form the query set. To implement the model function in this

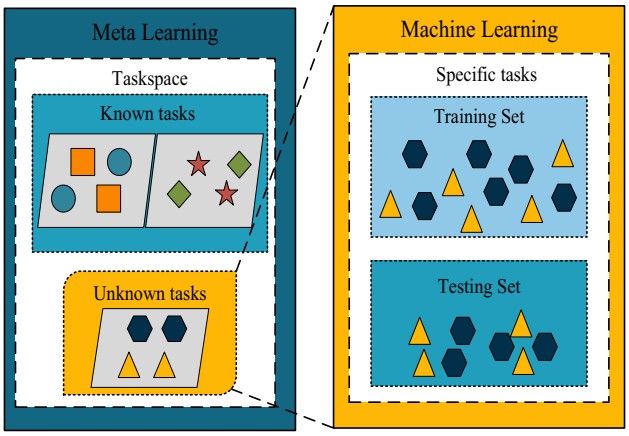

Fig. 3 Machine learning and meta-learning examples of binary classification

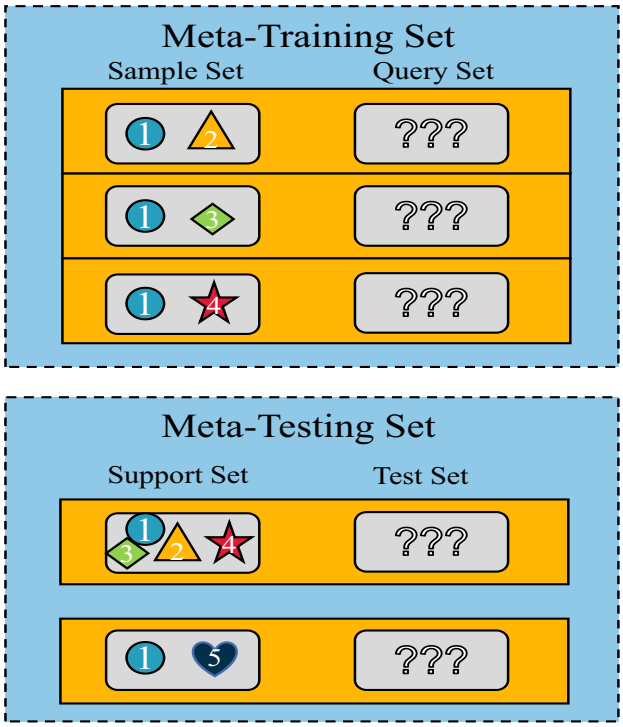

Fig. 4 The schematic diagram of meta-learning dataset partitioning

scenario, the task space and the data set are divided in this paper as shown in Fig. 4.

Methodology

To solve the problems of few-shot and zero-shot attack detection difficulties and poor real-time performance existing in 5G-enabled industrial internet, this paper adopts the improved Res-Natural model, the transformer model based on the fast attention mechanism, and the improved prototype network and prototype, and designs a few-shot intrusion detection model consisting of a sample generation module, a feature mapping module, and a feature metric module based on meta-learning framework. The specific components are illustrated in Fig. 5.

Sample generation module

Aiming at the characteristics of fast updating speed of industrial Internet network attacks and the emergence of new types of malicious attacks in the 5G era, this paper is committed to be closer to the real network environment and to realise the function of generating samples in different networks, so as to achieve the purpose of the expansion of the number of few-shot and the improvement of confrontation. Therefore, the sample generation module consists of a dataset division part and an adversarial sample generation part, to obtain higher quality adversarial samples and improve the fit with the real industrial Internet environment. However, higher quality adversarial samples are not enough, adversarial attacks designed to evade intrusion detection through modifications, spoofing operations, noise interference, etc., may destroy the effectiveness and robustness of the proposed model. Therefore, inspired by [21–23], the superiority of Generative Adversarial Networks (GANs) in generating adversarial samples is exploited, and residual blocks are introduced into the Natural GAN model to propose the improved Res-Natural GAN method, and during the training process, different proportions of the adversarial samples (0%, 20%, 40%, 60%, 80%, 100%) are mixed with the original samples and adjusted the parameters to the best effect to reduce the damage of the adversarial samples to the model effectiveness and improve the robustness of the intrusion detection model.

Dataset definition and partitioning

The dataset is defined and partitioned in "Problem formulation" and the partitioning algorithm is shown in Algorithm 1.

Res-Natural GAN to generate adversarial samples

The GAN model [35] belongs to the deep learning model and consists of a generator and a discriminator. To address the underfitting or overfitting problems that may occur when using machine learning methods to expand few-shot, GAN can expand the number of few-shot and improve the quality of the generated adversarial samples by learning the distributional features of the real data and generating high-quality brand-new samples in different hidden spaces. In the process of model training, the generator and the discriminator are always in the game state, in the minimum-maximum competition between the two sides, and improve the effect of outputting antagonistic samples in the mutual antagonistic relationship. The specific process is that the generator tries

to generate confrontation samples similar to the real samples to deceive the discriminator, while the discriminator is committed to distinguishing the real samples from the confrontation samples; when it is difficult for the discriminator to judge the authenticity of the samples, the deception samples generated by the generator are used as the final output confrontation samples to realize the specific function of the GAN model. However, the original GAN model uses JS scatter to measure the degree of similarity between the distribution of real samples and adversarial samples, which has the disadvantages of a single distribution of generated data, pattern collapse, and the convergence situation cannot be measured. WGAN [36] effectively avoids the problem of the gradient being unable to be updated in the case of non-overlapping distributions by adopting Wasserstein distance to measure the distance between two distributions P_r, P_g.

Algorithm 1 Generating a Few-Shot Task From Dataset

Input: Label $\dot{L}=\{0, 1, 2, 3, 4, 5, 6, 7\}$. Dataset $\dot{D}=\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}(xi \in R^d, y_i \in \dot{L})$. The number of sample sets(S_a) and query sets (Q) is the same, with N categories, each containing K samples. There are two subsets in the test set, (Test1 and Test2) one is small sample detection with N-1 classification and the other is zero sample detection with two classifications.

Output: Few-shot task $\hat{T} = \{Sa, Q, K, Test1, Test2\}$

Require: Random decimation ($\dot{L}$, K, $\dot{D}$) abbreviated as Rd($\dot{L}$, K, $\dot{D}$). stands for K sample sets labeled L randomly selected from the dataset

① Generate Sample Set
$Sa < 1 > \leftarrow Rd(\dot{L}(1), K, D)$
$Sa < 2 > \leftarrow Rd(\dot{L}(2), K, D)$
$Sa \leftarrow Sa < 1 > \cup Sa < 2 >$
② Generate Query Set
$Q < 1 > \leftarrow Rd(\dot{L}(1), K, D)$
$Q < 2 > \leftarrow Rd(\dot{L}(2), K, D)$
$Q \leftarrow Q < 1 > \cup Q < 2 >$
③ Generate training set labels
for S in $\{Sa, Q\}$ **do**
  for (xi,yi) in D **do**
    if yi≠0 **then**
      $yi \leftarrow yi + 1$
    end if
  end for
end for
④ generate Test1,Test2
$Test1 < 1 \sim 4 > \leftarrow Rd(\dot{L}(1 \sim 4), D)$
$Test2 < 1, 5 > \leftarrow Rd(\dot{L}(1, 5), D)$
$\hat{T} \leftarrow \{Sa, Q, K, Test1, Test2\}$

However, the WGAN model has the problem of generating adversarial samples with weak targeting. To address this problem, drawing on the generation of image samples, the Natural GAN model [37] is used, based on the WGAN framework, which searches for the hidden vectors of the adversarial samples in the hidden feature space by establishing the correspondence between the sample space and the hidden feature space, making it more difficult to detect the perturbation factors that are added in the deeper space, which is disorienting. It includes the following two stages:

(1) First train the generator, discriminator and reverser with real data, in which the generator realizes the mapping from the hidden vector space z to the sample space x (Eq. 1); construct a reverser that realizes the mapping from the sample space x to the hidden vector space z (Eq. 2), and establishes the correspondence relationship (Eq. 3) to realize the inter-transformation of the two spatial domains.

$$x = G(z) \tag{1}$$

$$z = G(x) \tag{2}$$

$$z = G(I(x)) \quad x = I(G(z)) \tag{3}$$

$$\min E_{x\sim p(x)}\left(\left\|G_\theta\left(I_\gamma(x)\right) - x\right\|\right) + \lambda \bullet E_{z\sim p(z)}\left(L\left(z, I_\gamma\left(G_\theta(z)\right)\right)\right) \tag{4}$$

(2) Utilize the network in (1) to search for the adversarial sample in the hidden vector space z. Specifically, the reverser is utilized to map the real sample x in the hidden vector space z, and then a random perturbation is added to obtain the adversarial hidden vector; and then the generator is utilized to complete the transformation of the adversarial hidden vector to the adversarial sample (Fig. 6).

Natural GAN model provides an idea of generating high-quality adversarial samples in a dynamic game environment, where the generator and the discriminator usually adopt deep neural networks. However, adversarial networks suffer from the problems of unstable training and excessive freedom, and the Res-Natural GAN model is proposed to overcome the disadvantage that the front-end gradient of deep neural networks is vulnerable to the back-end, causing the vanishing gradient problem. The residual structure is introduced into the neural network, and the residual neural network is used to construct generators and discriminators to improve the stability and robustness of the model.

The basic principle of the residual neural network is to establish straight-through channels between the input and output so that the model only needs to learn the difference between the input and output, thus simplifying the network learning objective and intensity. Unlike VGGNET, the residual neural network consists of residual blocks (Fig. 7) with multiple straight-through bypasses from the input to the back layer, which can alleviate the vanishing gradient and gradient explosion problems.

Feature mapping module

The main component of the feature mapping module is the transformer model based on the multi-head fast attention mechanism, which is a mature feature mapping network to map the input one-dimensional feature information and attribute information into a small embedding space of the

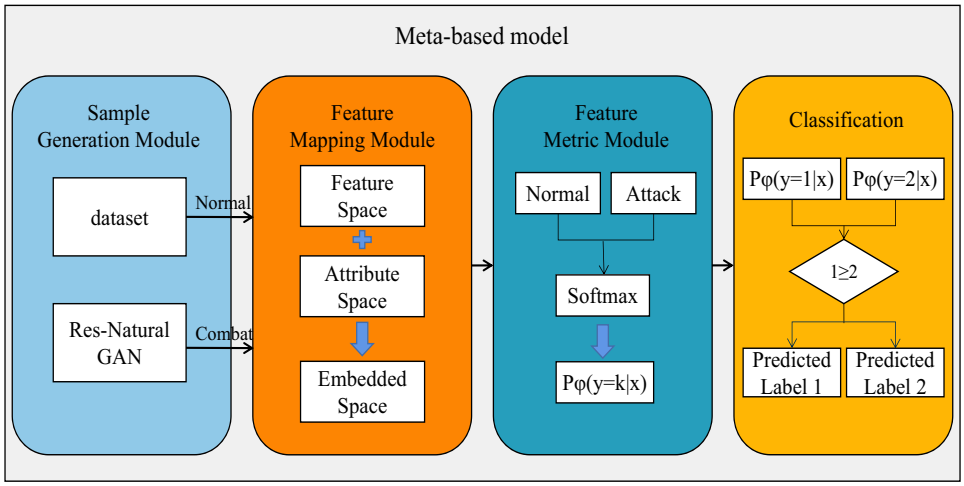

Fig. 5 The schematic diagram of the model structure

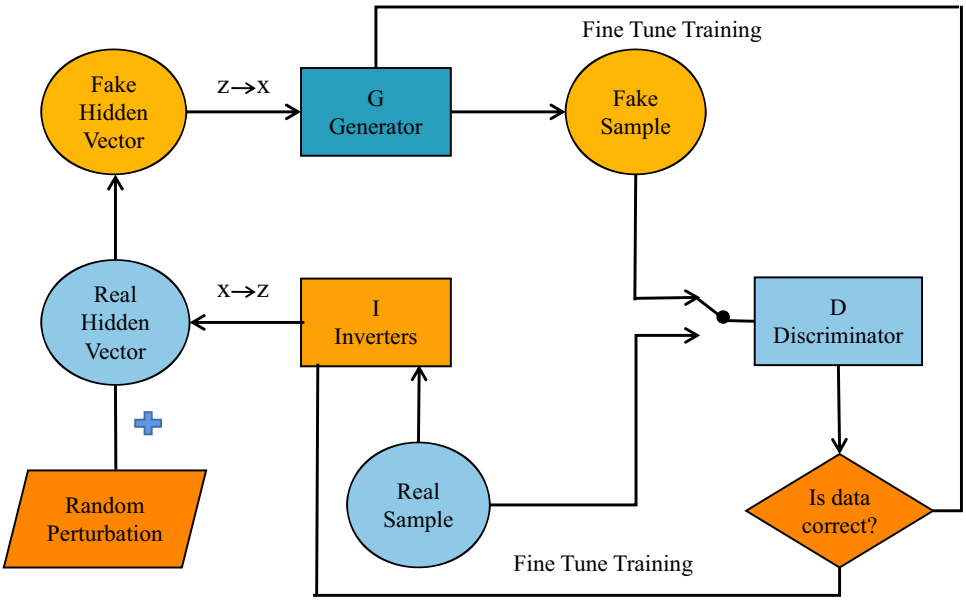

Fig. 6 The schematic diagram of Natural GAN structure

same dimension, thus making the features more distinguishable. The transformer model is chosen in this paper to complete the mapping of network traffic features because it introduces a positional embedding and a multi-headed self-attention mechanism that can identify the positions between elements and derive attention weights, thus allowing parallel computation and reducing the training time significantly. In addition, to make the improved model have better performance, inspired by [34], the parameter optimisation method based on particle swarm mutation is selected after comparing with other methods; to overcome the problems of premature convergence and local optimum stagnation of existing evolutionary optimisation algorithms, the adopted method is more suitable for 5G-enabled industrial internet intrusion detection application scenarios. By adopting an efficient parameter optimisation method, the feature mapping module can more accurately and rapidly mine deep-level features, mitigate the risk of overfitting, and improve stability and reliability to a certain extent.

Improved transformer encoder

The transformer model consists of an encoder and a decoder [38]. The encoder maps the sequence of network traffic into a hidden layer, i.e., a mathematical representation containing the sequence information; meanwhile, the decoder maps this hidden layer into a different form of sequence again. Based on the perspective of this paper, only the encoder of the transformer is used. The original encoder adopts a multi-headed self-attentive mechanism, and the memory consumption and encoding time is greatly affected by the number of features, which is not conducive to building a real-time intrusion detection system. For this reason, this paper focuses on improving the attention mechanism and proposes an encoder based on

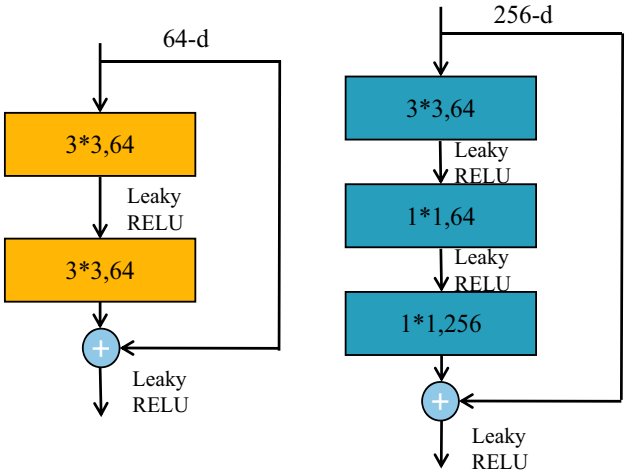

Fig. 7 The schematic diagram of the residual block structure

the multi-head fast attention mechanism, which can effectively improve the speed of feature mapping (Fig. 8).

Attentional mechanisms [39] are derived from the visual mechanisms of the human brain for long sequences containing a large amount of information, which allows focusing on a small amount of relatively important information in a large amount of information while ignoring the rest of the secondary information. To summarize, the attentional weights of feature information at each position during the encoding process are computed by some arithmetic method, and the whole feature information implicit vector is then represented as a weight sum. In this process, when the encoding proceeds to each position, it will excessively focus the attention to the current position and ignore the correlation between other positions and the current position.

The self-attention mechanism [40] is an improvement of the attention mechanism and focuses more on the internal correlation of feature information. Firstly, multiply the embedding vector with the random initialization matrix to get three vectors of Query, Key, and Value; and after that, multiply the Q and K vectors pointwise, divide by a scaling factor for normalization, and go through softmax to get the correlation magnitude of the rest of the specified features with respect to the features at the current location; finally, multiply the Value vector with the feature correlation magnitude to get the weight value of the self-attention mechanism at weight value of each node. The multi-attention mechanism, on the other hand, is a parallelization of multiple self-attention mechanisms, i.e., the attention weights of multiple locations are computed at the same time. However, since most of this mechanism is multiplicative, when the dimension of feature information increases, the occupied memory and training time increase exponentially by square, which is not favourable for building a real-time intrusion detection system.

$$Q = \text{Linear}\left(X_{\text{embedding}}\right) = X_{\text{embedding}} W_Q \tag{5}$$

$$K = \text{Linear}\left(X_{\text{embedding}}\right) = X_{\text{embedding}} W_K \tag{6}$$

$$V = \text{Linear}\left(X_{\text{embedding}}\right) = X_{\text{embedding}} W_V \tag{7}$$

$$\text{Attention}\left(Q, K, V\right) = \text{soft max}\left(\frac{QK^T}{\sqrt{d_k}}\right) V \tag{8}$$

The multiple fast attention mechanism is an improvement for this problem in this paper. Focusing on the repetitiveness of the dot product of Q and K vectors, it reduces the frequency of the internal correlation calculation of features, strengthens the attention to the current feature, and at the same time weakens the attention to the rest of the unimportant features (Fig. 9). For features with large correlation coefficients with the current location feature, the attention weight calculation is performed in accordance with the steps of the self-attention mechanism; on the contrary, for cases with small correlation coefficients, the correlation magnitude is regarded as 0 for the calculation, thus shortening the training time.

At the same time, the combination of the multiple attention mechanism allows the encoders to jointly pay attention to the information from the representation subspaces at different locations, compute h sets of Q, K, and V vectors in parallel, and finally splice the h sets of attentional node values for the transformation.

$$h_i = f\left(W_i^{(q)} q, W_i^{(k)} k, W_i^{(v)} v\right) \in R^{Pv} \tag{9}$$

$$W_0 \begin{bmatrix} h_1 \\ \cdot \\ \cdot \\ h_h \end{bmatrix} \in R^{p_0}, \tag{10}$$

where $W_i(q) \in R^{p_q \times dq}$, $W_i(k) \in R^{p_k \times d_k}$, $W_i(v) \in R^{p_v \times d_v}$ is the learnable parameter and f is the attention pooling function, the splicing result of h-heads is obtained after the linear transformation of Eq. 10. For a feature, only the value in the top 50% of its correlation coefficient size is calculated, and the value in the rest of the positions is regarded as 0, which is then calculated with the V vector to obtain the attention weight. The detailed working process is shown in Algorithm 2.

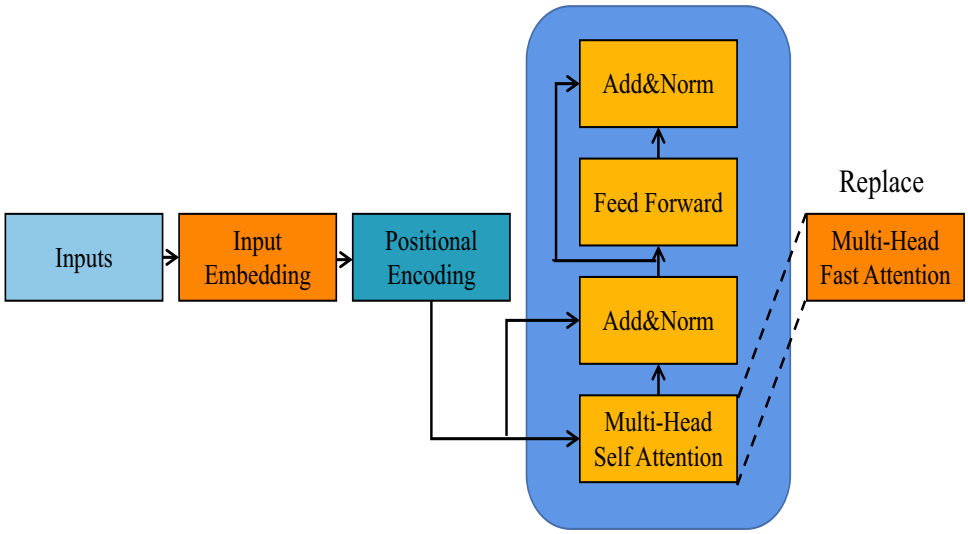

Fig. 8 The diagram of the improved attention mechanism

Algorithm 2 Multi-Headed Fast Attention Mechanism

Input: Feature data $F = \{F_1, F_2 \ldots, F_{94}\}$, Embedded Vector $E_M = \{E_1, E_2 \ldots, E_{94}\}$, Vector Query=Q, Vector Key=K, Vector Value=V.

Output: $Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)_{F_1 \ldots F_{94}}$

Require: Corr(i,i+1)represents the correlation coefficient between the ith and i+1th eigenvectors; $S_{(F_i, F_{i+n})} = \left(\frac{QK^T}{\sqrt{d_k}}\right)_{F_i, F_{i+n}}$ represents the correlation of eigenvectors.

① Calculate the value of correlation coefficient between eigenvectors.
for i in [1, 94), $n = (0, 1, 2, 3 \ldots)$ **do**
  if i+n≤94 **then**
    $(F_i, F_{i+n}) \leftarrow corr(i, i+n)$
  end if
end for
② Compare the correlation coefficient values and select the top 50
for i in [1, 94) **do**
  if corr(i, i+1)<corr(i,i+2) **then**
    Mark corr $(i, i+1)$ the first 1
  end if
  Mark $S(F_i, F_{i+1}) = 0$
  return result
end for
Calculate S value
$S_{(F_i, F_{i+n})} \leftarrow \left(\frac{QK^T}{\sqrt{d_k}}\right)_{F_i, F_{i+n}}$
$Attention(Q, K, V)_{F_i} \leftarrow softmax(S_{F_i})V$

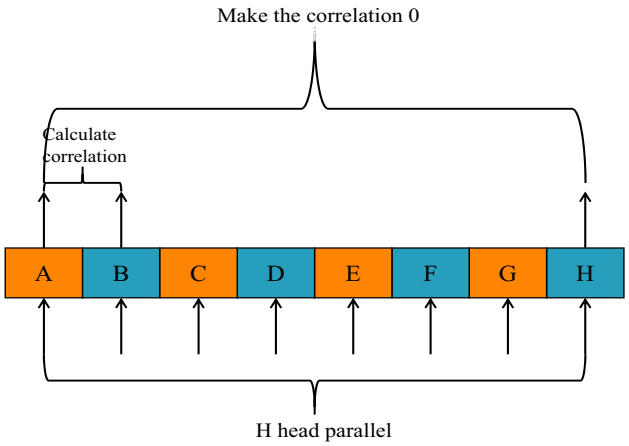

Fig. 9 The specific principles for calculating the correlation coefficient

Mutation based on PSO (APSO)

Traditional evolutionary optimisation algorithms often suffer from the problems of premature convergence and easy to fall into local optimal solutions, to solve the above problems, a parameter optimisation method based on particle swarm mutation was proposed in [34], which integrates the genetic algorithm (GA) method into the original PSO algorithm to form a mutation, to overcome the shortcomings of premature stagnation. In this paper, this improved method is applied to the parameter optimisation of feature mapping module and good results are achieved.

The original PSO method searches for optimal solutions by simulating the movement of individual particles in the solution space and information sharing, including initialising the particle swarm, evaluating the fitness, updating the individual optimal and global optimal solutions, as well as updating the velocity and position during the iteration process. In the initialisation phase, randomly generated particles are given initial positions and velocities; the fitness function is used to calculate the fitness of each particle and evaluate the quality of the solution; and the individual optimal solution is later updated by comparing the current fitness with the individual historical best fitness. At the same time, the particle with the best fitness is found in the whole particle swarm, and its corresponding solution is taken as the global optimal solution. This process fully simulates the searching behaviour of the particles in the solution space by adjusting the speed and position to find a better solution. The iterative

process repeats the above steps until the stopping condition is satisfied or a predefined number of iterations is reached.

Adaptive ASO mutation (APSO) introduces Genetic Algorithm (GA) into the original PSO method, which introduces variability in the search space by randomly selecting individuals among the offspring produced by crossover operations, with the help of diversity-driven multi-parent evolutionary algorithms of different mutation operations to facilitate deep exploration.

$$\text{mut}\left(\text{OC}_{i,j}^{t}\right)=\begin{cases}\text{OC}_{i,j}^{t}-\omega; & \text{rm}<\text{Prob}_{m}\\ \text{OC}_{i,j}^{t}+\omega; & \text{rm}\geq\text{Prob}_{m}\end{cases}, \tag{11}$$

where $\text{OC}_{i,j}^{t}$ denotes the offspring produced by the mutation operation, ω denotes a randomly generated parameter from within a tenth of the length of the search space, and rm denotes a random number between $[-1, +1]$.

Feature metric module

The main components of the feature metric module include a prototype network [41] and a prototype machine based on a novel prototype storage update algorithm. The core idea is to obtain the class prototype by calculating the average feature vector of each class of samples in the sample set and then measure the distance between the samples in the query and test sets and the class prototype to predict the probability of obtaining the correct classification. The specific implementation steps are as follows.

(1) Constructing class prototypes: Generate embeddings for all data points in the sample set and average the embeddings over similar samples to construct class prototypes.

$$C_k=\frac{1}{\mid S_k\mid}\sum_{(xi,yi)\in S_k}f_{\varphi}(x_i), \tag{12}$$

where x_i and y_i represent a sample of the sample set and its corresponding category, respectively; f_{φ} represents the feature embedding function; $|S_k|$ is the total number of samples in category K; φ is the learnable parameter, and C_k is the prototype of each category.

(2) Calculating the query point embedding: generating an embedding $f_{\varphi}(x)$ for the query point using the embedding function.

(3) Metric Euclidean distance: calculate the Euclidean distance d between the query point embedding and the class prototype.

$$d\left(f_{\varphi}(x),C_k\right)=\sqrt{\left(f_{\varphi}(x)-C_k\right)^2}, \tag{13}$$

where $f_{\varphi}(x)$ refers to the embedding of the query point, C_k refers to the class prototype, and d refers to the distance function between the two.

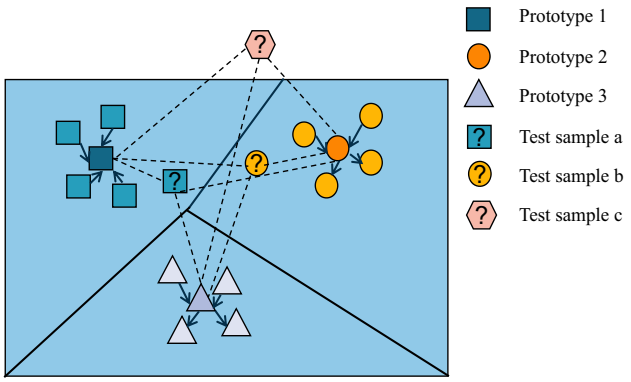

Fig. 10 The schematic representation of metric classification

(4) Predicted classification: performing softmax on the d obtained in (3),the probability of classification of the query set samples is obtained.

$$P_{\varphi}(y=k\mid x)=\frac{\exp\left(-d\left(f_{\varphi}(x),C_k\right)\right)}{\sum_k\exp\left(-d\left(f_{\varphi}(x),C_k\right)\right)}. \tag{14}$$

(5) Building a new class discriminator: One of the great challenges facing the current prototype network is the ability of the meta-learner to learn to master the meta-knowledge of the base learner, i.e., whether new types that do not appear in the sample set can be correctly classified. Considering this, a probability threshold is set as an interval criterion for determining the type of new attacks. The new class discriminator follows the principle of "Small determinations and large judgments", i.e., once the predicted value of the classification probability in (4) is smaller than the probability of any known type, the new attack type is identified; otherwise, the final classification is determined by the magnitude of the classification probability.

$$\begin{cases}P_{\varphi}(y=k\mid x)\geq c & x=k\\ P_{\varphi}(y=k\mid x)<c & x=a\end{cases}, \tag{15}$$

where $p_{\varphi}(y=k\mid x)$ refers to the predicted probability in (4), k refers to the type in the sample set, and a refers to the new attack type. A graphical representation of the metric classification is presented in Fig. 10.

(6)Minimizing loss: choose a negative log probability loss function and use stochastic gradient descent to minimize loss.

$$J(\varphi)=-\log\left(P_{\varphi}(y=k\mid x)\right). \tag{16}$$

However, since this paper aims to solve the few-shot and zero-shot classification problem, there is little prior knowledge available in the sample set; meanwhile, considering real-time requirements, based on the prototype network, a

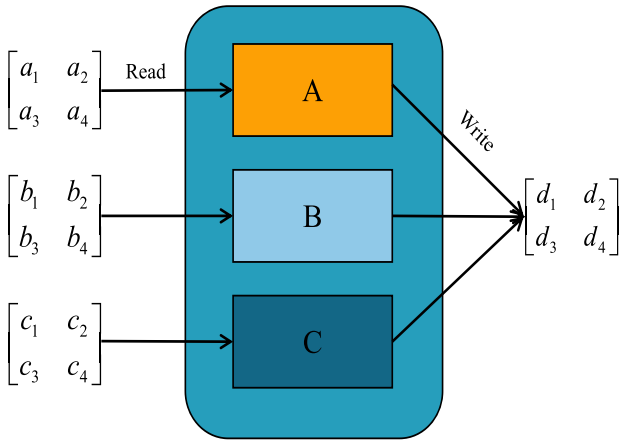

Fig. 11 The schematic diagram of the principle of the prototype

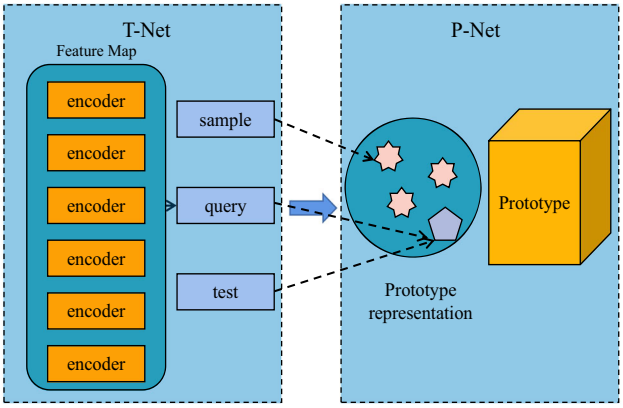

Fig. 12 The diagram of the TP-Net architecture

prototype machine is constructed that can provide the following functions.

(1) Fast metrics: after the prototype network has calculated the class prototypes of the sample set, a write operation is implemented to write the class prototypes to the prototype machine; after the query and test sets have calculated the embedding, a read operation is implemented to compare directly with the class prototypes in the prototype machine to achieve fast metrics(Fig. 11).

(2) Prototype update: To address the problem of inaccurate calculation of class prototypes due to the small number of samples, a real-time update is adopted. That is, after the sample labels in the query set and the test set are compared to achieve correct classification, the matching class prototypes in the prototype are averaged to increase the number of samples and enhance the accuracy (Algorithm 3).

Algorithm 3 New Prototype Storage Update Algorithm

Input: Class prototype vector P_n, Query Set Vector Q_x, Test Set Vector T_y.
Output: Euclidean distance of similarity measure and Classification probability
Require: $P_{(x,n)}$ Indicates the probability that sample x is predicted as type n.

① Storage class prototype vector.
$P < n >= P_n$
② Compare the sample with the class prototype vector and update the prototype.
for i ≤n **do**
  if $P_{x,i} \leq P_{x,i+1}$ **then**
    $P_{x,i} \leftarrow (P_{i+1} + Q_x)/2$
  end if
end for
③ Output classification
$Q_x, T_x \leftarrow P_{x,i+1}$

(3) Weighted selection: To avoid traversal operations, the class prototypes corresponding to samples with a high frequency in the query and test sets are given higher weights and prioritized as the object to be measured, thus reducing the inference time.

(4) Broadcast prototype: Under the wide coverage and high complexity of the 5G-enabled industrial Internet, drawing from the routing table idea, a small LAN containing four experimental machines is constructed, and this intrusion detection model is deployed. The prototype contains a prototype table, and whenever a new class prototype appears, it is broadcast to the rest of the prototypes in the LAN and updated accordingly to improve detection efficiency.

During the model construction process, a cascading network TP-Net is proposed for end-to-end implementation of feature mapping and metric roles(Fig. 12). In this figure, T-Net refers to a transformer model based on a multi-headed fast attention mechanism, and P-Net refers to the combination of a prototype network and a prototype machine. The TP-Net calculation flow is illustrated in Algorithm 4.

Algorithm 4 The Computational Flow of the TP-Net

Input: Sample Set,Query Set,Testing Set1,Testing Set2
Output: Test task 1 ACC,Test task 2 ACC,average ACC.

① Feature mapping.
Sample Vector←Sample Set+T-net
Query Vector←Query Set+T-net
Testing Vector←Testing Set+T-net
② Feature measurement
for i i in SV{1, 2 . . . k} **do**
  Class prototype←P-net(Sa< k >)
  for j in QV,TV{1, 2 . . . n} **do**
    P(j,i)←P-net(softmax)
  end for
end for
if P(j,i)≤P(j,i+1) **then**
  $QV_n, TVn \leftarrow SV_{i+1}$
end if
Calculate ACC

Evaluation

Implementation

Data used in the experiments

The data used to evaluate the model in this paper was based on the following three principles.

(1) Considering the need for uniformity in the embedding space for the model metric module, the data input should have both feature space and attribute space, i.e., each sample should have a corresponding label.

(2) Since the model is constructed for the 5G-enabled industrial internet,the experimental dataset should be within the industrial domain to verify the validity within the application scenario. There are public datasets for the intrusion detection domain such as NSL-KDD, UNSW-NB15, KDD CUP 99, ADFA,Kyoto, CICIDS2018, ICS, etc., but no datasets are available specifically for evaluating few-shot detection. Datasets that conform to the above principles include CICIDS2018 [42] and ICS, etc. Among them, the CICIDS2018 dataset is a collaborative project between the Communications Security Establishment (CSE) and the Canadian Institute for Cybersecurity (CIC), which collects real network traffic and contains the latest benign and malicious attacks; the ICS dataset is proposed by the Centre for Infrastructure Protection at Mississippi State University in 2014. It is obtained by injecting attacks into natural gas pipeline control systems, and it can be used for industrial control system intrusion detection assessment. The above two datasets were sampled to construct a few-shot detection dataset that satisfies the above three principles, as shown in Table 1, where the coded attack types starting with c are from the CICIDS2018 dataset and the coded attack types starting with i are from the ICS dataset.

Table 1 The abbreviations of attack types

Code	Attack types
C_1	Normal
C_2	DoS
C_3	DDoS
C_4	Bruteforce
C_5	Bortnet
C_6	Infiltration
I_1	Normal
I_2	NMRI
I_3	CMRI
I_4	MFCI
I_5	MPCI

Metrics

The purpose of this paper is to design a meta-learning model for the few-shot classification problem. Therefore, a good measure of the performance of the model is to evaluate its completeness and accuracy for the classification task. To facilitate the comparison of the validity of the research work, this study adopts four types of evaluation metrics commonly used in the field of intrusion detection: accuracy, precision, recall and F1 Score [43]. Among them, the meanings of the sub-metrics involved are as follows:

TP(True Positive): indicates the number of positive samples predicted as positive by the model (correct).

FP(False Positive): indicates the number of negative samples that the model predicts as positive (incorrect).

TN(True Negative): indicates the number of negative samples that the model predicts as negative (correct).

FN(False Negative): indicates the number of positive samples predicted as negative samples by the model (incorrect).

Based on the above sub-indicators, the following four categories are obtained:

Accuracy is the ratio of the number of samples correctly classified by the model to the total number of samples:

$$\text{Accuracy} = \frac{TP + TN}{FP + FN + TP + TN} \tag{17}$$

Precision is the proportion of all data predicted to be a positive sample, to the number of all true positive samples:

$$\text{Precision} = \frac{TP}{FP + TP} \tag{18}$$

Recall measures the ability of the model to predict all positive samples:

$$\text{Recall} = \frac{TP}{FN + TP} \tag{19}$$

F1 Score combines precision and recall and is a reconciled average of the two to measure the balanced performance of the model:

$$F_1 - \text{score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{20}$$

Since the task space contains two subsets of test tasks, their performance is calculated as Test Task 1 ACC/Pre/Re/F1 Score and Test Task 2 ACC/Pre/Re/F1 Score, respectively; the samples in the test task space are merged into a single task set for testing, and the performance evaluation metrics Average ACC/Pre/Re/F1 Score are obtained for evaluating the effectiveness of the model. In few-shot testing, the experimental time is relatively short due to the limited number of samples and the small number of statistical indicators for a

Table 2 The schematic of the experimental setup

Project	Environmental parameters
CPU	Intel Core i7-8550u CPU
GPU	NVIDIA RTX2060
Python version	3.7.2
Pytorch version	1.12.1
Keras version	2.5.0
PC	Windows10 (x64)

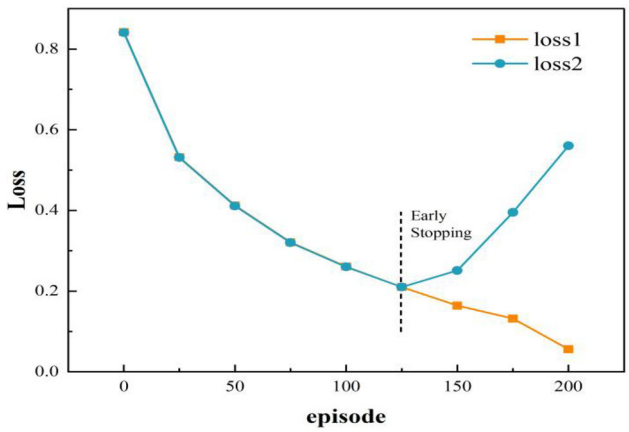

Fig. 13 The schematic diagram of the principle of the early stop method

specific type of task. Therefore, this paper repeats the experiment several times to take the average value to ensure the credibility of the experiment.

Experimental environment and hyperparameter settings

The specific experimental setup of this paper is shown in Table 2:

During the experiment, the concept of the early stop method [44] was introduced. The early stop method (Fig. 13) mainly solves the problem of difficulty in setting the number of stereotyped cycles in model training. By comparing the detection accuracy on the validation set after a certain number of training rounds, the final value of this parameter was decided. In this way, the underfitting and overfitting problems were effectively avoided, and the generalization ability of the model across the task space was improved.

At the same time, when the parameter optimisation algorithm APSO is used for the feature mapping module, the set of hyperparameters to be optimised is set as (learning rate, batch size, number of heads); since the model performs feature mapping for few-shot, the batch size per class ranges fluctuates in [10,15]; it is found that the performance of the model is significantly better than the other heads when the number of heads is 2 and 6 in the process of seeking optimisation. To verify the effectiveness of the adopted parameter optimisation algorithm and to further illustrate the scientific and reasonable nature of the experimental hyper-parameter settings, some of the solutions in the particle optimisation process are shown (Table 3):

Table 3 shows the six groups of particle positions and hyperparameter settings for head number 2 and 6, and it can be seen that the global optimal position is [1.00000000e−04,1.00000000+01,6.00000000+00], and at this time, the corresponding optimal parameters are: the learning rate is 0.0001, the batch size is 10, and the number of heads is 6. Therefore, the hyperparameters of the feature mapping module are set as above. In addition, the hyperparameters of the feature metric module are set as follows: the number of categories in the training set is N, the number of samples is K, a total of 3,000 rounds of training are performed in three times, and the test results are output once every 200 rounds. Among them, the data set is divided according to the division of the sample generation module, Zero-Shot refers to zero-shot type, Test Task 1 refers to test subset 1, Test Task 2 refers to test subset 2, and Average refers to the merging of Test Task 1 and Test Task 2.

Experimental setup

To fully measure the performance of the real-time intrusion detection system constructed in this paper, five types of experiments were set up by varying the experimental variables.

K-value change experiments

As the system adopts a meta-learning algorithm for few-shot detection, it follows the N-Way K-Shot principle. By varying the K values of the number of samples of each type in the sample set and the query set and comparing the effect of different K values on the experimental results, the scenario of a few malicious attacks in a real network environment with only a finite number of samples of different sizes was simulated. For a subset of tasks in the test task space, K=5, K=10, and K=15 were set, and three sets of comparison experiments were conducted(The division of the dataset is shown in Table. 4, the experimental results are shown in Figs. 14 and 15).

The above experimental results show that for a subset of the two test task spaces, the accuracy increases with the increase of K value. This indicates that the meta-learning framework can successfully detect zero -shot and few-shot, and the effect of the number of samples on the detection results is not obvious when few-shot are detected; a longitudinal comparison between test task 1 and test task 2 shows that the accuracy rate increases instead of decreases to a certain extent with the increase in the number of sample categories N. This is due to the fact that, in general, the higher the num-

Table 3 APSO partial experimental results

Particle Position	Learning rate	Batch size	Number of attention heads	Loss	Acc
[1.00000000e−04,1.50000000e+01,6.00000000e+00]	0.0001	15	6	0.07589	92.79%
[1.00000000e−04,1.10000000e+01,6.00000000e+00]	0.0001	11	6	0.07394	92.83%
[1.00000000e−04,1.00000000e+01,6.00000000e+00]	0.0001	10	6	0.07379	92.86%
[6.92506322e−02,1.40000000e+01,2.00000000e+00]	0.06925	14	2	0.47843	83.51%
[1.00000000e−01,1.30000000e+01,2.00000000e+00]	0.1	13	2	0.53913	83.43%
[8.83000000e−02,1.20000000e+01,2.00000000e+00]	0.083	12	2	0.50157	83.46%

Table 4 K-value change Experimental setup table

Zero-shot type	C_2	C_3	C_4	C_5
Test Task 1	C_1, C_2	C_1, C_3	C_1, C_4	C_1, C_5
Test Task 2	C_1, C_3, C_4, C_5	C_1, C_2, C_4, C_5	C_1, C_2, C_3, C_5	C_1, C_2, C_3, C_4

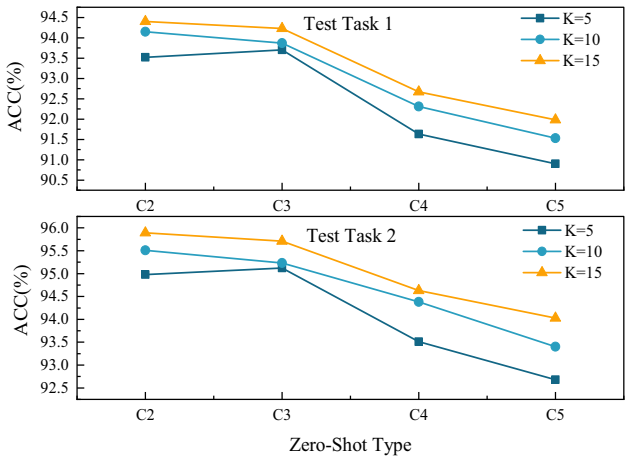

Fig. 14 The accuracy of zero-shot detection at different K values

ber of sample data categories is, the more rich the features to be learnt by the model and the more difficult the classification prediction is; while in the above experimental setup, the second classification corresponds to the zero-shot detection task and the fourth classification corresponds to the few-shot detection task, the zero-shot detection requires the model to make inferences in the face of the unknown categories, while the few-shot detection can be learnt and inferred from the existing sample data. Therefore, within a given sample range, the experimental results show that test task 2, which focuses on few-shot detection, has higher detection accuracy compared to test task 1, which focuses on zero-shot detection.

Comparison of the detection performance of the average task set shows that in the experiments under K=5, K=10 and K=15, the detection accuracy of few-shot and zero-shot increases with the increase of the number of samples, but the improvement is not significant. This indicates that the meta-learning framework constructed in this paper can mine distributions and learn features from data, but the detection function is more effective with a smaller number of samples. In addition, comparing the detection accuracy of the average task set with the detection accuracy of test task 1 and test task 2, it can be seen that the detection performance of the average task set is not as good as that of the sub-test task set when the samples have the same few-shot and zero-shot data types, this is due to two reasons: firstly, the increase in the number of sample types N leads to the increase in the difficulty of the feature learning and training, and the accuracy of the classification detection is affected; secondly, the average task set contains two types of tasks, few-shot detection and zero-shot detection, which means that the model needs to have the ability to judge unknown categories and be able to make classification predictions from limited sample data, which puts higher requirements on the model's detection performance, resulting in a detection accuracy that is not as desirable as that of the testing of the single-class task. Meanwhile, by comparing the detection accuracies of different few-shot types under the same K-value, it can be seen that the implementation of the model is not limited to a specific type, and the C_2 type (DoS) has the highest detection accuracy, which can be used as the experimental data for the subsequent work, and the results are more convincing and comparable.

Module replacement experiments

In this paper, the sample generation module, the feature mapping module, and the feature metric module of the framework are designed and improved. To demonstrate the effectiveness of the improvements in each module, the models used in each module are replaced, and the experimental results are compared. Meanwhile, machine learning and deep learning algorithms are used to classify few-shot outside the meta-learning framework to illustrate the effectiveness of the constructed framework.

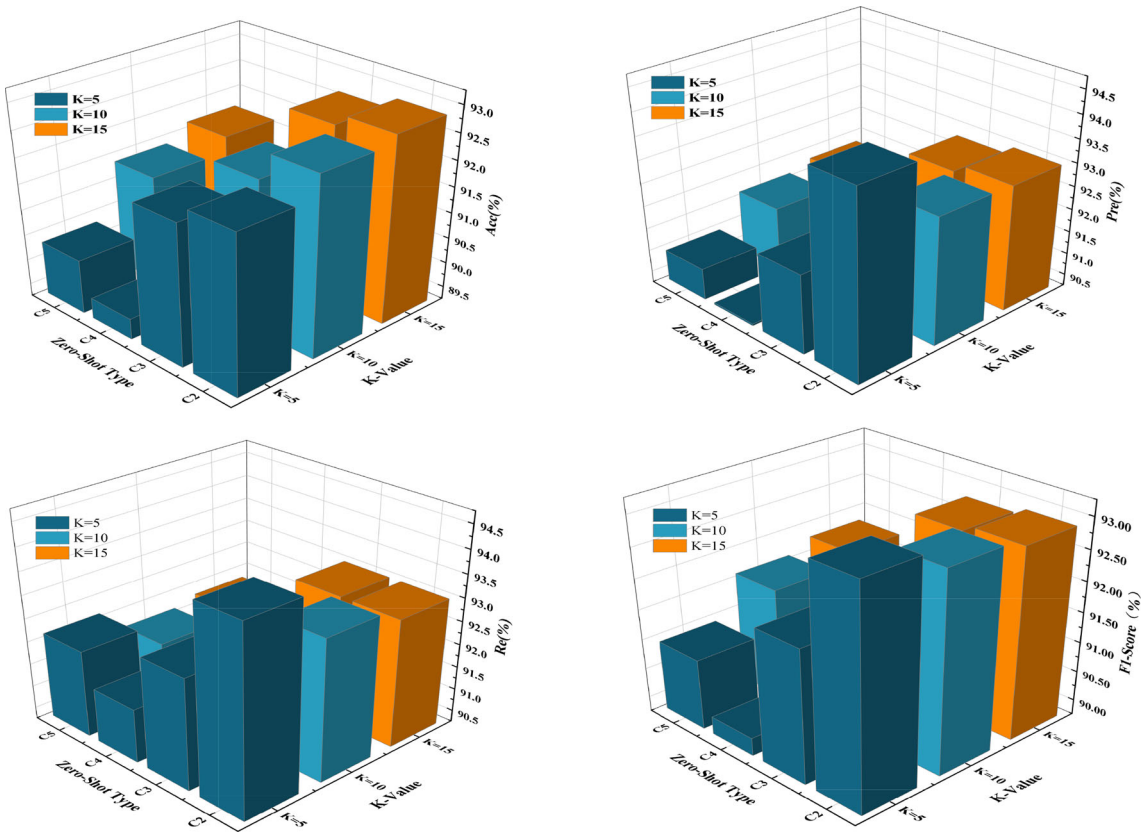

Fig. 15 Experimental results of zero-shot detection on average test task at different K values

As can be seen in Fig. 16, the experiment replaces the models in each module and changes the number of different K-values to verify the effectiveness of the proposed method with four evaluation metrics: accuracy, precision, recall, and F1-score on three types of settings: K=5,K=10 and K=15, and the experimental results are more comprehensive and convincing. Replacing the Res-Natural GAN part of the sample generation module with Res-GAN and Res-WGAN models is obviously not as effective as the Res-Natural GAN model in adversarial training, and thus the evaluation metrics of the four categories have been reduced to a certain extent; at the same time, experimental comparison of the proposed method with the method of converting one-dimensional network traffic into two-dimensional RGB images, combined with CNN and RNN, shows that the proposed method is more comprehensive and effective. The experimental comparison of the method shows that this paper overcomes the bottleneck of the existing research work to a certain extent, and achieves good detection results by directly mapping the features of the 1D network traffic; by replacing the proposed feature metric module with relational, matching and twin networks, the experimental results show that the prototype network based on the new determinants has a higher detection accuracy. Replacing the proposed meta-learning framework, classical machine learning algorithms such as SVM, LR, DT, and RF are used for comparison experiments, and the detection accuracy is greatly reduced, indicating that the meta-learning algorithms perform significantly better than the traditional machine learning algorithms in detecting few-shot. In summary, the performance of the traditional original algorithms is not as desirable as the meta-learning framework, whether replacing the models in each module or the overall framework, further illustrating the effectiveness and practicality of the proposed work.

Adversarial sample experiments

To adapt the new model to the gaming adversarial environment, this paper uses the Res-Natural GAN part of the sample generation module to generate adversarial samples to effectively handle the variability of attacks launched by attackers with known training data, and the difference between the DoS attack and the real sample distribution is presented in Fig. 17.

To validate the effectiveness of the Res-Natural GAN model in the sample generation module, the test set was used as an invariant covariate to generate adversarial samples on a type-by-type basis, and six different proportional compari-

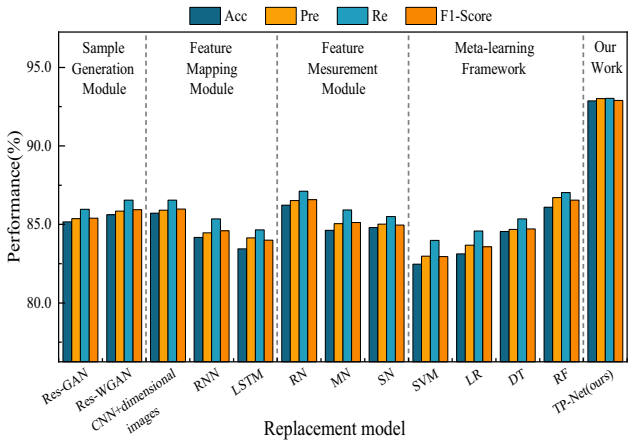

Fig. 16 The histogram of results of model replacement experiments

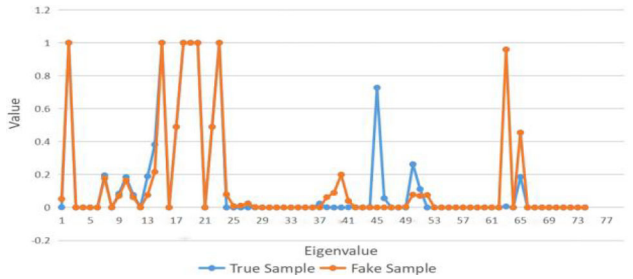

Fig. 17 The distribution of real samples vs adversarial samples

son training groups were set up. The specific settings and the experimental results are shown in Table 5.

As shown in Fig. 18, the accuracy of the fusion task set improves as the proportion of adversarial samples rises, with the test set samples remaining unchanged. It indicates that the adversarial sample input training effectively enhances the robustness of the model in the dynamic gaming environment and ensures the detection accuracy of the model when it encounters dynamic attacks again.

Data alteration experiments

The above three sets of experiments involve four attack types, which are typical of the five classification problems. However, it cannot illustrate the effectiveness of the model for 5G-enabled industrial internet, as well as its adaptability in different network environments and its applicability to different attack types. Therefore, this paper further constructed eight sets of comparative experiments for different unknown attack types with the data in Sect. 4.1.1 (Table. 6).

As can be seen from the table, the experiments based on the publicly available datasets CICIDS2018 and ICS, i.e., the first five sets of experiments, have an accuracy rate of up to 93.20%, which indicates that the effectiveness of the constructed meta-learning framework in 5G-enabled industrial internet is not confined to specific application contexts and types of attacks.

Time comparison experiment

This paper aims to build a real-time intrusion detection system for the 5G-enabled industrial internet, so it requires extremely high real-time and detection speed. Considering this, time-comparison experiments were conducted on two speed-boosting components: the prototype, and the transformer model based on a multi-head fast attention mechanism for ablation(Table 7).

The experimental results demonstrate that both speed-boosting components reduce the inference time to a certain extent, meeting the requirements of real-time intrusion detection systems.

Conclusion and future work

To solve the problem of few-shot attacks occurring with low frequency and small number of samples, traditional supervised detection algorithms rely on a large amount of labelled data and cannot effectively detect few-shot or even zero-shot attacks. At the same time, considering the rapid occurrence of network attacks and short response time, order to improve few-shot attack detection accuracy and inference speed to a certain extent, this paper constructs a meta-learning-based intrusion detection framework that contains a sample generation module, a feature mapping module and a feature metric module. In this framework, the residual block is creatively introduced into the Natural GAN model, and a new method of generating high-quality antagonistic samples, Res-Natural GAN, is proposed to increase the number of samples while improving the attack antagonism to enhance the detection accuracy; the encoder structure of the transformer is improved to some extent, and a fast multi-head based attention mechanism of the transformer structure, through the introduction of multiple independent attention heads with weight matrices to achieve parallel computation, which greatly shortens the training and inference time of the model while mapping the effective feature information, and improves the detection efficiency; a new class discriminator for identifying zero-shot attacks and a prototype for prototype storage updating are proposed in the prototype network, which improves the model's accuracy and enhance the timeliness of the model.

To evaluate the proposed meta-learning framework, five types of experiments are conducted in this paper. The validation by changing the number of samples, adding antagonistic samples, changing the data source, module ablation experiments and model comparison experiments show that the framework not only has high detection accuracy and fast

Table 5 Adversarial sample experiment setup table

Proportion	Changed numbers	Changed type	
0%	0	None	
20%	1	C_2	K = 5;
40%	2	C_1, C_2	Zero-shot type: C_2;
60%	3	C_1, C_2, C_3	Test Task 1: C_1, C_2;
80%	4	C_1, C_2, C_3, C_4	Test Task 2: C_1, C_3, C_4, C_5.
100%	5	C_1, C_2, C_3, C_4, C_5	

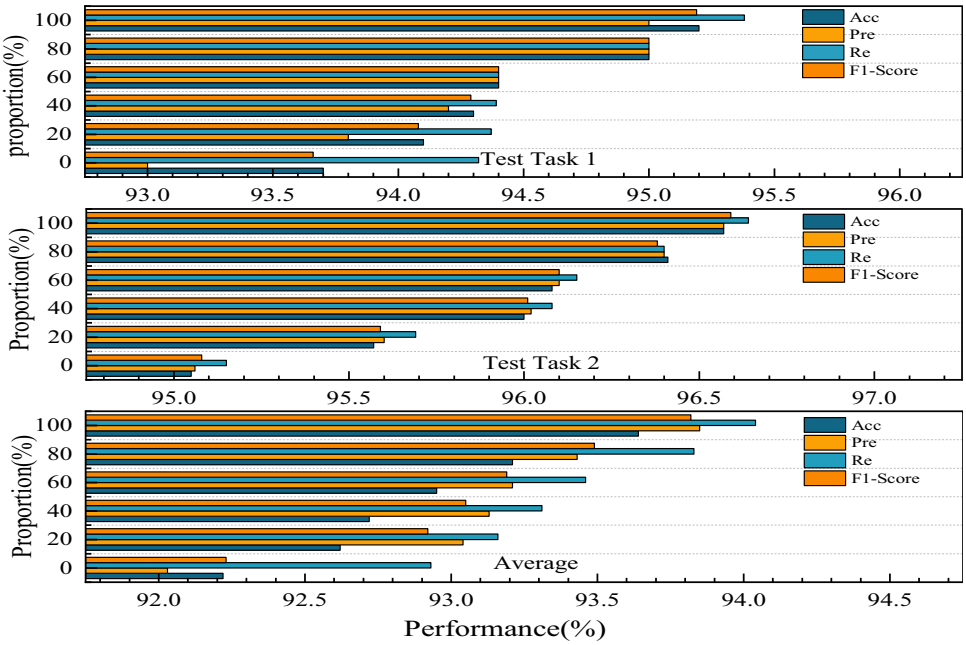

Fig. 18 The results of the experiment on the change in the proportion of adversarial samples

Table 6 The results of data alteration experiments

Group	Zero-shot type	Test Task 1	Test Task 2	Accuracy		
				1	2	Average
Group1	C_6	C_6, C_1	C_1, C_2, C_3, C_4, C_5	90.45%	92.56%	89.55%
Group2	I_2	I_2, I_1	I_1, I_3, I_4, I_5	91.00%	93.40%	90.25%
Group3	I_3	I_3, I_1	I_1, I_2, I_4, I_5	89.96%	91.25%	88.50%
Group4	I_4	I_4,I_1	I_1, I_2, I_3, I_5	90.65%	92.15%	89.69%
Group5	I_5	I_5,I_1	I_1, I_2, I_3, I_4	88.97%	90.60%	87.40%

detection speed, but also is equally applicable to different network environments and variable attack types, and still has high accuracy in 5G-enabled industrial Internet application scenarios. In addition, this paper also finds that the meta-learning algorithm is less sensitive to the number of samples in the experimental process, and the detection accuracy for different numbers of samples varies but not much, which further confirms that the meta-learning algorithm is suitable for few-shot detection, and still has a high detection accuracy for unlabelled zero-shot attacks, and for the most accurately categorised DoS category, the detection accuracy can be as high as 92.86%.

In our future work, we will continue our research to further improve the detection accuracy and efficiency of the model; at the same time, we will start from three aspects, namely, model lightweighting, the use of different parameter optimisation algorithms with the further reduction of the impact of adversarial attacks on the intrusion detection model, to improve the effectiveness of the work done. First, the meta-learning intrusion detection framework constructed in this paper consists of three modules with more parameters to achieve higher detection accuracy, and the next step is to design a lighter intrusion detection model and make it better applied to the actual network environment; second,

Table 7 The results of ablation experiments

Prototype machine	Transformer based on multi-fast attention	Infer time
✓	✓	34ms
✓	×	39ms
×	✓	37ms
×	×	42ms

the optimisation algorithm selected for parameter optimisation of the feature mapping module has played a good role, and the next step is to try other parameter optimisation methods to be applied in the samples. The next step is to try other parameter optimization methods applied to the sample generation module and the feature metric module to further improve the model effectiveness. Finally, to reduce the impact of adversarial attacks on the intrusion detection model, this paper chooses to generate different proportions of adversarial samples and add them to the normal samples for model training, combining the standard training process with the process of adding adversarial samples to improve the model robustness, and we intend to try the integration of the defence mechanism, monitoring and feedback loop strategy and other methods by integrating multiple defence mechanisms, monitoring and feedback loop strategy, which can be applied in real network environments. loop strategy and other methods to monitor the model performance in real time by integrating multiple intrusion detection systems or establishing a monitoring system to detect the impact caused by adversarial attacks in a timely manner.

Data availability The relevant data and material in this article are available from the corresponding author.

Declarations

Conflict of interest The authors state that there is no conflict of interest.

References

1. Iammartino R, Bischoff J, Willy C, Shapiro P (2016) Emergence in the us science, technology, engineering, and mathematics (stem) workforce: an agent-based model of worker attrition and group size in high-density stem organizations. Complex Intell Syst 2:23–34
2. Zhang Q, Lu J, Jin Y (2021) Artificial intelligence in recommender systems. Complex Intell Syst 7:439–457
3. Li Q, Kumar P, Alazab M (2022) Iot-assisted physical education training network virtualization and resource management using a deep reinforcement learning system. Complex & Intelligent Systems, 1–14
4. Palittapongarnpim P, Wittek P, Zahedinejad E, Vedaie S, Sanders BC (2017) Learning in quantum control: High-dimensional global optimization for noisy quantum dynamics. Neurocomputing 268:116–126
5. Yu L-P, Wu F-Q, Chen G-Q (2019) Next-generation industrial biotechnology-transforming the current industrial biotechnology into competitive processes. Biotechnol J 14(9):1800437
6. Xie C, Hua Q, Zhao J, Guo R, Yao H, Guo L (2022) Research on energy saving technology at mobile edge networks of iots based on big data analysis. Complex Intell Syst 8(5):3943–3952
7. Khan BS, Jangsher S, Ahmed A, Al-Dweik A (2022) Urllc and embb in 5g industrial iot: a survey. IEEE Open J Commun Soc 3:1134–1163
8. Ding D, Han Q-L, Xiang Y, Ge X, Zhang X-M (2018) A survey on security control and attack detection for industrial cyber-physical systems. Neurocomputing 275:1674–1683
9. Stevens C (2020) Assembling cybersecurity: The politics and materiality of technical malware reports and the case of stuxnet. Contemporary Security Policy 41(1):129–152
10. Zimba A, Mulenga M (2018) A dive into the deep: demystifying wannacry crypto ransomware network attacks via digital forensics. Int J Inform Technol Secur 10(2):57–68
11. Roumani Y (2021) Patching zero-day vulnerabilities: an empirical analysis. J Cybersecur 7(1):023
12. Atlam HF, Azad MA, Alassafi MO, Alshdadi AA, Alenezi A (2020) Risk-based access control model: A systematic literature review. Future Internet 12(6):103
13. Alicea M, Alsmadi I (2021) Misconfiguration in firewalls and network access controls: Literature review. Future Internet 13(11):283
14. Yan Y, Yang Y, Shen F, Gao M, Gu Y (2023) Gde model: a variable intrusion detection model for few-shot attack. J King Saud Univ-Comput Inform Sci 35(10):101796
15. Li X, Sun Z, Xue J-H, Ma Z (2021) A concise review of recent few-shot meta-learning methods. Neurocomputing 456:463–468
16. Khosravy M, Nakamura K, Hirose Y, Nitta N, Babaguchi N (2021) Model inversion attack: Analysis under gray-box scenario on deep learning based face recognition system. KSII Transactions on Internet & Information Systems **15**(3)
17. Goodfellow IJ, Shlens J, Szegedy C (2014) Explaining and harnessing adversarial examples. arXiv preprint arXiv:1412.6572
18. Kurakin A, Goodfellow IJ, Bengio S (2018) Adversarial examples in the physical world. In: Artificial Intelligence Safety and Security, pp. 99–112. Chapman and Hall/CRC, ???
19. Papernot N, McDaniel P, Jha S, Fredrikson M, Celik ZB, Swami A (2016) The limitations of deep learning in adversarial settings. In: 2016 IEEE European Symposium on Security and Privacy (EuroS&P), pp. 372–387. IEEE
20. Kou L, Ding S, Rao Y, Xu W, Zhang J (2022) A lightweight intrusion detection model for 5g-enabled industrial internet. Mobile Netw Appl 27(6):2449–2458
21. Garg M, Ubhi JS, Aggarwal AK (2023) Neural style transfer for image steganography and destylization with supervised image to image translation. Multimedia Tools Appl 82(4):6271–6288

22. Maini D, Aggarwal AK (2018) Camera position estimation using 2d image dataset. Int J Innov Eng Technol 10:199–203
23. Xiao J, Aggarwal AK, Rage UK, Katiyar V, Avtar R (2023) Deep learning-based spatiotemporal fusion of unmanned aerial vehicle and satellite reflectance images for crop monitoring. IEEE Access
24. Jamil M, Afzal F, Maqbool A, Abdullah S, Akgül A, Bariq A (2023) Multiple attribute group decision making approach for selection of robot under induced bipolar neutrosophic aggregation operators. Complex & Intelligent Systems, 1–15
25. Liang W, Li K-C, Long J, Kui X, Zomaya AY (2019) An industrial network intrusion detection algorithm based on multifeature data clustering optimization model. IEEE Trans Ind Inform 16(3):2063–2071
26. Li Y, Xu Y, Liu Z, Hou H, Zheng Y, Xin Y, Zhao Y, Cui L (2020) Robust detection for network intrusion of industrial iot based on multi-cnn fusion. Measurement 154:107450
27. Bibi N, Maqbool A, Rana T, Afzal F, Akgül A, El Din SM (2023) Enhancing semantic code search with deep graph matching. IEEE Access
28. Wu Z, Yang P, Wang Y (2019) Mvpn: multi-view prototype network for 3d shape recognition. IEEE Access 7:130363–130372
29. Li J, Chiu B, Feng S, Wang H (2020) Few-shot named entity recognition via meta-learning. IEEE Trans Knowl Data Eng 34(9):4245–4256
30. Aggarwal A (2020) Enhancement of gps position accuracy using machine vision and deep learning techniques. J Comput Sci 16(5):651–659
31. Chauhan S, Singh M, Agarwal AK (2019) Crisscross optimization algorithm for the designing of quadrature mirror filter bank. In: 2019 2nd International Conference on Intelligent Communication and Computational Techniques (ICCT), pp. 124–130 . IEEE
32. Chauhan S, Singh M, Aggarwal AK (2021) Experimental analysis of effect of tuning parameters on the performance of diversity-driven multi-parent evolutionary algorithm. In: 2021 IEEE 2Nd International Conference on Electrical Power and Energy Systems (ICEPES), pp. 1–6. IEEE
33. Chauhan S, Singh M, Aggarwal AK (2023) Designing of optimal digital iir filter in the multi-objective framework using an evolutionary algorithm. Eng Appl Artificial Intell 119:105803
34. Chauhan S, Singh M, Aggarwal AK (2023) Investigative analysis of different mutation on diversity-driven multi-parent evolutionary algorithm and its application in area coverage optimization of wsn. Soft Computing, 1–27
35. Yinka-Banjo C, Ugot O-A (2020) A review of generative adversarial networks and its application in cybersecurity. Artificial Intell Rev 53:1721–1736
36. Arjovsky M, Chintala S, Bottou L (2017) Wasserstein generative adversarial networks. In: International Conference on Machine Learning, pp. 214–223. PMLR
37. Zhao Z, Dua D, Singh S (2017) Generating natural adversarial examples. arXiv preprint arXiv:1710.11342
38. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I (2017) Attention is all you need. Advances in neural information processing systems **30**
39. Shen G, Chen Z, Wang H, Chen H, Wang S (2022) Feature fusion-based malicious code detection with dual attention mechanism and bilstm. Comput Secur 119:102761
40. Hernández A, Amigó JM (2021) Attention mechanisms and their applications to complex systems. Entropy 23(3):283
41. Liu X, Zhou F, Liu J, Jiang L (2020) Meta-learning based prototype-relation network for few-shot classification. Neurocomputing 383:224–234
42. D'hooge L, Wauters T, Volckaert B, De Turck F (2019) Classification hardness for supervised learners on 20 years of intrusion detection data. IEEE Access 7:167455–167469
43. Fawcett T (2006) An introduction to roc analysis. Pattern Recognit Lett 27(8):861–874
44. Prechelt L (1998) Automatic early stopping using cross validation: quantifying the criteria. Neural Netw 11(4):761–767