Research on Intelligent Configuration Method of Mine IoT Communication Resources Based on Data Flow Behavior

YIFAN MENG (Student Member, IEEE), AND JINGZHAO LI (Senior Member, IEEE)
College of Electrical and Information Engineering, Anhui University of Science and Technology, Huainan 232000, China
Corresponding author: Jingzhao Li (ljzaust@outlook.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 51874010, and in part by the Key Technology Research & Innovation Team Project under Grant 201950ZX003.

ABSTRACT In order to ensure production safety and improve production efficiency, mining enterprises are constantly accelerating the construction of mine Internet of Things systems. In the context of a substantial increase in the number of devices with network communication capabilities in the mine, the mine network communication facilities are under tremendous pressure. We propose a device business classifier based on convolutional neural networks to improve the service quality of mine network communication infrastructure. The classifier uses wavelet transform to extract the data flow and construct behavior characteristics to classify device business categories. According to the classification results, the system flexibly adjusts the parameters of network services provided to the terminal equipment. In this way, the network resources of the system can be allocated reasonably. We evaluate the performance of the classifier model through the test data set. The performance evaluation results show that the comprehensive recognition rate of the classifier model reaches 97.2%. We optimize and adjust the classifier model according to the hardware environment in which the classifier is actually deployed.

INDEX TERMS Convolutional neural network, data flow classification, wavelet transform mine Internet of Things.

I. INTRODUCTION
In the past ten years, the continuous improvement of computing device performance and the rapid iteration of communication network technology has driven the rapid development of Internet of Things technology in combination with the needs of various industries [1]. Due to the harsh mining environment, mining companies need to strengthen their ability to perceive and control the mining system [2]. These have accelerated the construction of the mine IoT system [3]. In the mine IoT system, a large number of devices with network communication capabilities are deployed. The huge amount of data generated by them has caused tremendous pressure on the mine’s network communication infrastructure. With the increasing demand of managers for mine condition perception, more types of equipment will enter the mine IoT system. By then, the pressure on the mine’s network communication infrastructure will further increase. To ensure the stable operation of the mine IoT system, the network communication infrastructure in the mine needs to optimize the network load, adjust the network traffic, and accurately allocate network resources. This situation requires the quality of service (QoS) capability of the mine communication network.

In the mine IoT control system, because it is difficult to balance the performance and size of the device, a large number of terminal devices cannot process data on-site [4]. These devices need to upload data to the data computing facilities in the mine IoT system for analysis and processing. Due to the different functions of mine IoT devices, various types of mine IoT devices have diverse requirements for communication network parameters and bandwidth.

For some mine IoT devices, they only generate data flows when triggered by specific events. These specific events may be sensitive and critical. Therefore, when data flow is generated, it needs to be forwarded as soon as possible. As a real-time monitoring class of mine IoT devices, such as environmental monitoring stations, there will be small...
data packets uploaded periodically. It is not necessary to give such devices high priority, as long as the network delay is within a reasonable range. At the same time, there are also a large number of surveillance cameras in the mine, and the data flow generated by them is massive. These large and dense data flows will cause greater load pressure on the network facilities at the edge. This requires that the mine network infrastructure needs to optimize the scheduling of network communications for different types of equipment.

A major factor restricting the development of mine Internet of Things is how to improve the service quality of mine network infrastructure. One of the solutions is to classify data flow services at the network communication infrastructure level. According to the category of equipment, it can allocate network resources reasonably to achieve the improvement of mine network service quality. The terminal equipment in the mine IoT system comes from different manufacturers and follows different technical specifications. By actively optimizing the network of mine IoT terminal equipment at the network communication facility level, it is not necessary for the mine IoT equipment manufacturers to carry out targeted optimization design for the equipment.

In some special environments, IoT devices cannot obtain a stable external power supply, and battery power is a relatively common form. Researchers have done a lot of research work on the power consumption control of the Internet of Things system. Venkatraman Balasubramanian and others defined a unified resource allocation architecture for edge computing devices with limited resources, maximizing the survival time of edge devices [5] Mobile devices in mine IoT systems are often battery-powered, and it is also necessary to avoid increasing their power consumption burden. The network parameters of the equipment are optimized from the level of the mine Internet of Things infrastructure to avoid the additional power loss of terminal equipment. The gateways, switches, routers, base stations and other equipment that constitute the mine’s Internet of Things infrastructure are generally provided with electricity by the mine’s power supply system. At the same time, an uninterruptible power supply (UPS) will be installed at the key nodes of the mine’s Internet of Things to ensure the reliability of power supply. This power supply system often uses 110V alternating current in coal mines in China.

In this paper, we use the network data flow collected from the mine IoT system to design a classifier that runs in the mine IoT communication infrastructure. The classifier can analyze the communication behavior of the device based on the data flow generated by the device, and identify the type of mine IoT device. Therefore, the network communication infrastructure can realize network parameter tuning for different types of mine IoT devices. In order to implement this classifier, we chose to conduct behavior analysis for data flows instead of analyzing individual data packets based on the characteristics of mine IoT devices. We use wavelet transform to analyze the time series consisting of the length of the data packet in the time-frequency domain. Based on the wavelet transform results, we construct a set of feature data that contains both individual packet information and time series information. Then we designed a classifier based on the convolutional neural network (CNN) [6] and trained it using the data set obtained from the mine IoT system. Finally, the test results show that the classifier can complete the classification of equipment business categories within the time required, and the recognition accuracy can reach 97.2%. Because network devices are often designed based on low-power embedded platforms, we optimize the classifier model for the hardware and software environment in which the classifier is actually deployed.

The rest of this paper is structured as follows. Related research work is summarized in the second part. In the third part we discussed the collection of training data, feature extraction and model training. Model testing, result analysis, and model optimization and adjustment are given in fourth part. The fifth part gives the summary and prospect of this article.

II. RELATED WORKS
In the field of network data classification, researchers have studied many classification methods based on different ideas. These methods can be summarized into two technical routes: Deep Packet Inspection (DPI) and Deep Flow Inspection (DFI).

DPI is a traffic detection and control technology based on the application layer [7]. When TCP or UDP data flows pass through a management system based on DPI technology, the system reads the contents of the IP packet payload in depth. The management system reorganizes the application layer information to obtain the content of the entire application, and then shapes the traffic according to the management strategy defined by the system. M. Al-hisnawi and M. Ahmadi use Quotient Filter to detect the data packets in the network to avoid the consumption of a lot of memory and CPU resources and achieve higher throughput [8]. To solve the problem of packet inspection occupying computing resources, Ruxia Sun’s team proposed a method to improve DPI by using regular expressions [9].

DFI is a recognition technology based on traffic behavior. Different services have different states on session connection or data flow [10]. A. X. Liu et al. proposed a framework for automated online application protocol field extraction of DFI, which can effectively reduce the resources occupied by DFI and improve the running speed of the classifier [11].

Because DPI needs to read the payload content of IP data packets in-depth, and identify and classify at the application layer, this makes DPI technology difficult to efficiently classify IoT devices which used encrypted transmission [12]. DFI technology does not focus on individual data packets but analyzes the overall behavior of the data stream composed of multiple data packets. For our classifier to have wide compatibility in the mine Internet of Things, we chose to design the classifier based on DFI technology. Based on clarifying the
basic technical route, we will analyze the research status of related technologies.

The first thing we need to consider is the data interaction mode of mine IoT devices. Mine Internet of Things is a heterogeneous fusion network, a large number of terminal devices are connected to the mine LAN through various communication gateways [13]. The mine LAN is a high-speed industrial Ethernet composed of optical fibers and follows the Open System Interconnection (OSI) network standard protocol [14]. Mine Internet of Things has inherited web services in representing the transmission state architecture [15]. Devices use Constrained Application Protocol (CoAP) [16], MQTT [17], and other communication protocols designed for IoT devices. The IoT devices in mine can access standard IP networks through these protocols and use standard IP routing technology to communicate with existing IP devices in the network. At the same time, M. Z. Shafiq and others found that the upload stream data traffic of M2M devices is more than the download stream traffic by studying the data flow in the network [18].

Second, we need to consider how to extract features from the time series composed of data flows and classify the device to which the data packet belongs. How to construct a classifier for data flows is a hot field that researchers have been paying more attention to. Researchers have done a lot of work for different application backgrounds. Because the parameters and performance indicators of different communication networks are very different, it is difficult for us to compare the performance of various network traffic classifiers. Therefore, when we investigate related work, we mainly analyze from the perspective of research ideas.

A. Sivanathan et al. collected a large amount of data from an IoT system containing 28 smart devices and trained a second-order classifier from these data [19]. They analyzed the characteristics of domain names, port numbers, etc. in the data set through a native Bayesian polynomial classifier. The classifier’s recognition accuracy for specific IoT devices exceeds 99%. Manuel Lopez-Martín et al. proposed a network traffic classifier (NTC) based on the combination of recurrent neural network and convolutional neural network [20]. The classifier obtains packet header data from the IP layer of the network for analysis. Antunes Mário and others proposed to use semantic analysis to analyze the IoT data. In recent years, low-power wide area networks (LPWAN) have been increasingly used to build wireless data exchange channels for IoT systems. W. Yang et al. analyzed the technical characteristics of LoRa and NB-IoT that belong to LPWAN [25]. Researchers have also done a lot of work in the collaborative power control of IoT systems. Venkatraman Balasubramanian et al. proposed an RL-based Droplet framework for autonomous energy management [26].

In summary, the work of the above researchers has made important contributions. On the basis of their experience, we analyzed the behavior of the data stream of mine IoT terminal equipment, and designed a business data flow classification model for the mine IoT system.

III. CONSTRUCTION CLASSIFIER FOR MINE IoT DEVICES

To deploy the classifier model to the communication facilities of the mine Internet of Things system, we need to try to avoid significantly increasing the computational expenses of the network communication facilities. The classifier does not classify and identify a single data packet but analyzes the overall network data flow interaction behavior of the terminal device. Based on the wavelet transform, we construct a set of new features that can describe the behavior of data in both time and frequency domains from a time series composed of data flows. Then train a mine IoT device traffic classifier based on a convolutional neural network based on the data set collected in the actual mine IoT system.

A. DATA PREPROCESSING

We collect data from the mine IoT system for feature analysis and train the classifier. After obtaining the consent of the mining enterprise, our team used port mirroring to capture data packets from routers located at the edge of the ring network in the underground coal mine for research. Because the data body sent by the Internet of Things device may be encrypted, the classifier cannot parse the body of the data packet, so we do not pay attention to the content of the body, and only collect the header information of the data packet. The packet capture work is carried out in the network layer of the seven-layer OSI network model. We have collected nearly two million pieces of valid data. As shown in Table 1, there are 14 terminal devices connected to the router, including 4 general network devices and 10 mine IoT devices. The data distribution of each device is shown in Figure 1.

The original format of the data set collected from the router is a .pcap file, and we use the scapy software library to analyze it. Since we analyze does not focus on the body content of the data packets, we only extract the frame header information of each data packet when parsing the .pcap file. Then, according to the source MAC address in the frame...
TABLE 1. Device list of our data set.

| ID | Device            | Packets | Type            |
|----|------------------|---------|-----------------|
| 1  | Voice Broadcast  | 310307  | Non-IoT Device  |
| 2  | EP Camera-A      | 265700  | IoT Device      |
| 3  | EP Camera-B      | 789411  | IoT Device      |
| 4  | Turnout Controller| 86715   | IoT Device      |
| 5  | Pedestrian sensor| 6288    | IoT Device      |
| 6  | Gas Sensor       | 22263   | IoT Device      |
| 7  | Environmental Monitor | 140203 | IoT Device    |
| 8  | Floor Scale      | 29195   | IoT Device      |
| 9  | Semaphore-A       | 70728   | IoT Device      |
| 10 | Air Door Sensor  | 254778  | IoT Device      |
| 11 | Semaphore-B      | 6433    | IoT Device      |
| 12 | PDA-A            | 100622  | Non-IoT Device  |
| 13 | PDA-B            | 103210  | Non-IoT Device  |
| 14 | PDA-C            | 16128   | Non-IoT Device  |

FIGURE 1. Data flow distribution of each device in the data set.

header, the upload data flow of each device in the subnetwork of the router is obtained; and then according to the target MAC address in the frame header, the download data flow of each device is obtained.

FIGURE 2. Schematic diagram of window sequence movement.

The length of the wavelet transform result is related to the length of the input time series. We cannot simply divide the data flow according to the fixed number of data packets. But we can split the data flow by constructing a fixed-width time window. As shown in Figure 2, we first segment the upload data flow of the device according to a fixed length of time, write these pieces of data as $FU = \{fu_i | i \in \{1, 2, \ldots, n\}\}$.

In this way, the starting time $ts_i$ and the ending time $te_i$ of each segment are determined. Split the download data flow of the device according to the $ts_i$ and $te_i$. Ensure one-to-one correspondence between upload data flow segments and download data flow segments. Through this data segmentation method, we ensure that the length of the input wavelet transform time series is consistent, thereby ensuring the consistency of the data format for subsequent processing.

B. FEATURE EXTRACTION BASED ON WAVELET TRANSFORM

It is difficult for the network devices in the mine IoT system to parse the body of each data packet. This is because various IoT devices follow different encryption methods. Therefore, we need to construct a set of features that do not involve the content of the data body based on the device data flow information. In the mine IoT system, some IoT devices need to periodically send keep-alive heartbeat packets to the server to maintain a long connection. Other data packets containing payloads follow different rules. Some data packets are transmitted periodically, while others are event-driven. Therefore, we can construct a set of time signals from the data flow payload size. This set of time signals can be regarded as a composite of sub-data flows of different services.

C. Liu et al. consider using Fast Fourier Transform (FFT) to analyze the frequency spectrum of the packet length sequence to classify the data traffic [27] However, due to the Gibbs Phenomenon, the Fourier transform is difficult to fit the abrupt signal. The wavelet transform [28] can analyze the signal from the time-frequency domain, and its good multi-resolution characteristics are especially suitable for analyzing non-stationary signals to extract feature information.

We use Morlet continuous wavelet transform to process the data flow signal, zero-fill the data flow segments at fixed time intervals, and expand to a uniformly sampled sequence. The Morlet wavelet signal is shown in (1), and the wavelet base is shown in (2).

$$\psi(t) = \exp \left( -\frac{t^2}{\sigma^2} \right) \exp \left( i \omega_0 t \right)$$

$$\psi_{a,b}(t) = \exp \left( -\frac{(t-b)^2}{a^2} \right) \exp \left( i \omega_0 \frac{a}{\gamma} (t-b) \right)$$

The formula of Morlet wavelet transform is:

$$\Psi_f(a, b) = \int_{-\infty}^{+\infty} f(t) \psi_{a,b}(t) dt$$

$$= \int_{-\infty}^{+\infty} f(t) \exp \left( -\frac{(t-b)^2}{a^2} \right) \exp \left( -i \omega_0 \frac{a}{\gamma} (t-b) \right) dt$$

We use PyWavelets [29] software library to calculate wavelet transform, which provides rich wavelet transform API. As shown in Figure 3, in order to expand the divided data segment into a time series with a fixed sampling period, the original data segment is filled with zeros at a fixed time interval. Then perform wavelet transform feature extraction on each group of upload and download data flow fragments, and save all the calculation results as a .json file.

When intercepting data flow fragments, it is necessary to
count the number of effective communications that occur within the fragments. If in this time period, the times of communication of the upload and download data flows is 0, the time segment is discarded. The reason for this step is that some devices may have no upload data flow or download data flow for a long time.

In order to visually analyze the special rules of the device data flow, data visualization operations are performed on the data flow of each device. As shown in Fig. 4 to Fig. 11, it is the result of wavelet transform of upload data flow and download data flow of several typical devices.

Figure 4 and Figure 5 are the upload and download data of the real-time surveillance camera, which represents the Internet of Things equipment for video surveillance in the mine. It can be seen from the figure that the upload data flow of the video surveillance equipment has a large number of data packets of uniform size. Figures 6 and 7 show the upload and download data flow of the gas sensor. It belongs to real-time status monitoring equipment in the mine Internet of Things system. Such devices need to send monitoring data to the server in real time. The upload data flow of such devices often has a certain periodicity. Figures 8 and 9 represent various types of controlled actuators in the mine, which receive control commands sent by the management personnel from the remote end and then perform corresponding actions. Figures 10 and 11 are the upload and download data flows of pedestrian monitoring sensors. Pedestrian monitoring sensor is an event-driven device. Only when a valid event triggers the device, the data flow will be generated.

Through data visualization, it can be found that the data interaction behaviors of different kind of mine IoT devices have obvious differences. It is possible to construct features based on the device data flow behavior. In the next section, we will build a CNN-based classifier model based on these behavioral features.

C. CNN-BASED CLASSIFIER TRAINING

After extracting features through wavelet transform, the output is a matrix of size $m \times n$. Where $m$ is the number of sampling points for wavelet transform; $n$ is determined by the length of the data flow segment. In the feature extraction process, we set the number of sampling points to 64, and the length of the data flow sequence after zero-padding expansion is 2048. After combining the upload and download feature matrices of the same interval, it can be regarded as a single-channel picture with a resolution of $128 \times 2048$. The classifier needs to match the characteristic spectrum of the device data flow in the matrix to identify the
device category. We analyze the business characteristics of various devices in the mine IoT system and divide the devices into five categories, as shown in Table 2.

The convolutional layer of the convolutional neural network can scan the entire matrix by moving the filter, so no matter where the spectral feature appears in the matrix, it can be effectively identified. The classifier is designed based on LeNet-5 [30] model. The classifier structure is shown in Figure 12.

The construction of a neural network is only the first step. The choice of hyperparameters plays a decisive role in the success of neural networks. Reasonable setting of network
IV. CLASSIFIER TEST AND APPLICATION TUNING

After the classifier model training is complete, we use the test data set to evaluate the performance of the classifier.

TABLE 2. Classification of equipment in mine IoT system.

| Number | Device type                      |
|--------|----------------------------------|
| 0      | Video surveillance equipment     |
| 1      | Real-time condition monitoring equipment |
| 2      | Remote-Controlled actuator       |
| 3      | Event-driven equipment           |
| 4      | Non-IoT equipment                |

Hyperparameters can better train neural network models. Setting parameters such as learning rate, momentum coefficient, number of iterations, etc. can optimize hyperparameters and allow the neural network to achieve the best prediction effect. After many experiments, this paper selects the best set of network training results as the final hyperparameters. The initial learning rate of the model during training is set to 0.01, the weight attenuation coefficient is 0.001, the momentum coefficient is 0.9, and the batch size is 30.
And optimize the hardware platform characteristics of the classifier in practical applications.

A. MODEL EVALUATION

In this section, we use the test data set to evaluate the classifier on the hardware platform where the classifier is trained. The computer hardware and software configuration parameters used to train the classifier are shown in Table 3. Classifier model is built and trained based on the Tensorflow [31] library.

The classifier is to identify the device category based on the overall behavior of the device’s data flow over a period of time, and then to optimize the network communication parameters of different types of devices. Based on this application mode, the system has no strict requirements for the real-time processing capability of the classifier. In the performance evaluation, the main consideration is the accuracy of classification. General quantitative classifier performance indicators are True Positive (TP), True Negative (TN), False Negative (FN), False Positive (FP). We also quantify the accuracy of the classifier from two indicators: Precision and Recall. Precision is calculated by dividing the number of samples of this type with correct prediction by the ratio of all predictions of the model to the number of samples. The calculation formula is shown in (4). The meaning of Precision is to predict how many of them are really positive. Recall represents the ratio of the number of correctly identified targets to the number of all targets in the test set, and its definition is shown in (5).

\[
\text{Precision} = \frac{TP}{TP + FP} \quad (4)
\]

\[
\text{Recall} = \frac{TP}{TP + FN} \quad (5)
\]

Table 5 shows the classification of the classifier for each device in the test data set. The table mainly reflects the classification of the classifier for different devices. Among them, the recall rate for non-IoT devices is lower than other types of devices, which is consistent with the situation reflected in Table 4.

B. MODEL TUNING FOR EMBEDDED PLATFORM

However, it is generally difficult for network infrastructure at the edge of the network to have the performance as our model training platform. These network devices are often designed based on low-power embedded chipsets, such as Intel’s low-power Celeron processors, or other ARM Cortex-A series core processors. We hope that when the edge network infrastructure runs the classifier model, there is no need to significantly improve the performance of the hardware. Since the classifier model needs to run on a low-power embedded platform, we need to streamline the model. In the classifier model compression process, we mainly use the methods of pruning and weight sharing.

According to the classification accuracy of each category shown in Table 4, the performance of the classifier is generally satisfactory. It is more difficult to identify non-IoT devices. Since non-IoT devices may have data flows of multiple business types, and the types of services carried out over time are different, it is difficult to generalize obvious characteristics according to the behavior of data flows.
When training the test classifier model based on the training platform, we directly input the $128 \times 2048$ matrix output by the wavelet transform into the classifier for training and testing. However, the matrix size of the input classifier directly affects the speed of the model operation. Therefore, we need to compress the input data to reduce the amount of calculation of the classifier. Here we use principal component analysis (PCA) [32] to compress the feature data matrix. The size of the feature data matrix after PCA compression is $64 \times 1024$. We use compressed data to reconstitute the classifier model and compress the classifier model. The test results of the compressed classifier on the low-power embedded platform are shown in Table 7. The control experimental group is the operation of the original classifier model on the embedded low-power test platform.

### TABLE 7. Test results of classifier based on low power embedded platform.

| Device type                  | Compression | Uncompressed |
|------------------------------|-------------|--------------|
| Video surveillance equipment | 91.7%       | 97.9%        |
| Real-time monitoring equipment| 92.1%       | 98.1%        |
| Remote controlled actuator   | 90.5%       | 97.2%        |
| Event-driven equipment       | 92.4%       | 98.2%        |
| Non-IoT equipment            | 88.9%       | 92.7%        |
| Processing time              | 971s        | 1659s        |

As can be seen from data shown in Table 7, the accuracy of the classifier model constructed with compressed data is lower than that of the original classifier model, but the accuracy of all types of equipment is still maintained at 90%. After double compression of the model and feature data, the processing speed of the classifier model on the low-power embedded platform is greatly improved. Although the workflow of the classifier does not require real-time, the faster processing speed will reduce the occupation time of the hardware and prevent the high-load communication services that block the network infrastructure.

#### FIGURE 13. Partial structure of a typical mine network.

**C. APPLICABILITY TEST OF RESOURCE INTELLIGENT ALLOCATION METHOD**

Figure 13 shows a typical mine IoT structure. Mine communication networks are often built on the basis of high-speed optical fiber ring networks. As the backbone communication network of the mine, the current rate of the optical fiber ring network has far exceeded 10Gbps. Therefore, the main bottleneck of the mine Internet of Things lies in the edge of the network. By deploying the device classifier model in the basic communication facilities at the edge of the network, identify and mark the category of the device to which the data flow belongs. After the device category is determined, the network communication device can use different communication resource scheduling strategies to configure the device’s network parameters. In this way, the mine communication infrastructure can dispatch network resources for terminal equipment in an orderly manner to improve network service quality.

Different types of equipment play different roles in mine IoT systems, and their response priorities are often different. Therefore, we can construct a suitable scheduling strategy according to the priority of each category. In order to verify the effect of the mine Internet of Things communication resource allocation method, we built an experimental environment containing five real devices and 25 virtual devices. The main purpose of deploying twenty-five virtual devices in the experimental environment is to increase the load of the test network equipment and simulate the network environment of the system under high load. Since the embedded hardware platform we used for testing has only one RJ45 port, we matched it with a layer 2 network switch in the experimental environment to achieve single-arm routing. The test results are shown in Table 8.

### TABLE 8. Test results of intelligent resource allocation.

| Device              | Network delay | Packet loss rate |
|---------------------|---------------|------------------|
| Surveillance camera | Before 14ms   | After 20ms       | Before 6.32%    | After 1.28% |
| Remote switch       | Before 15ms   | After 7ms        | Before 3.17%    | After 0.35% |
| Cable temperature   | Before 14ms   | After 14ms       | Before 4.73%    | After 2.14% |
| Monitor             | Before 14ms   | After 14ms       | Before 4.73%    | After 2.14% |
| Gas detector        | Before 13ms   | After 6ms        | Before 5.41%    | After 0.24% |
| PDA                 | Before 15ms   | After 21ms       | Before 4.37%    | After 2.74% |

The test result shows that after the routing equipment recognizes the equipment category, it starts to actively adjust the network parameters of each equipment and intelligently configure the network communication resources. After the optimization is completed, the network delay of the Remote Switch as a Remote-Controlled Actuator device is significantly reduced. At the same time, the data flow of Gas Monitor, which is an event-triggered device, is also forwarded preferentially. This is because these two types of devices have higher priority in the scheduling strategy of our experimental environment.

### V. CONCLUSION

In this paper, we designed a classifier that runs on the network infrastructure to improve the communication service quality of the mine Internet of Things. The classifier analyzes the overall behavior of the data flow of the device and identifies...
the device type according to the data interaction characteristics of the device. Then optimize the network parameters of the equipment according to the classification results, and allocate network resources reasonably. Our contributions can be summarized as follows:

1. Separate the data flow through the network infrastructure of the mine IoT system according to the MAC address, and construct an upload data flow and a download data flow for each device. And we use the wavelet transform to extract the behavior characteristics of the device data flow.

2. Based on the feature data extracted by wavelet transform, construct a device classification model for the mine IoT system. The test results show that the classifier model can effectively identify devices of different service types, which provides support for subsequent network optimization for devices of different service types.

3. To deploy the classifier model in the mine IoT communication infrastructure, we compress the classifier model and compress the feature matrix based on PCA. Tests show that the running speed of the compressed classifier model has been greatly improved and the classification accuracy meets the requirements.

The classifier model proposed in this paper provides a network optimization strategy for the future development and construction of the mine Internet of Things system, and the classification effect is verified through experiments. In the next step, we will continue to study the packet classification model for the mine IoT system, and provide network optimization support for the mine IoT system from the data packet level.

REFERENCES

[1] J. A. Stankovic, “Research directions for the Internet of Things,” IEEE Internet Things J., vol. 1, no. 1, pp. 3–9, Feb. 2014, doi: 10.1109/JIOT.2014.2312291.

[2] G. Kortuem, F. Kawsar, V. Sundramoorthy, and D. Fitton, “Smart objects as building blocks for the Internet of Things,” IEEE Internet Comput., vol. 14, no. 1, pp. 44–51, Jan. 2010, doi: 10.1109/MIC.2009.143.

[3] B. Jo and R. Khan, “An Internet of Things system for underground mine communication infrastructure, we compress the classifier model and compress the feature matrix based on PCA. Tests show that the running speed of the compressed classifier model has been greatly improved and the classification accuracy meets the requirements.

[4] C. De-You, “Optical fiber layout and communication of coal mine safety network comprehensive monitoring system,” coal technol., vol. 33, no. 2, pp. 124–125, Feb. 2014.

[5] Z. Shen, H. Qingsong, and W. Gang, “Consideration about construction of standards for mine Internet of Things,” Ind. Mine Autom., vol. 44, no. 1, pp. 1–5, Jan. 2018.

[6] S.-C. Son, N.-W. Kim, B.-T. Lee, C. H. Cho, and J. W. Chong, “A time synchronization technique for coap-based home automation systems,” IEEE Trans. Consum. Electron., vol. 62, no. 1, pp. 10–16, Feb. 2016, doi: 10.1109/TCE.2016.7485557.

[7] H. W. Chen and F. J. Lin, “Converging MQRTI resources in ETSI standards based M2M platform,” in Proc. IEEE Int. Conf. Internet Things, Taipei, Japan, Sep. 2014, pp. 292–295, doi: 10.1109/IIOT.2014.52.

[8] M. Z. Shafiq, L. Ji, A. X. Liu, J. Pang, and J. Wang, “Large-scale measurement and characterization of cellular machine-to-machine traffic,” IEEE/ACM Trans. Netw., vol. 21, no. 6, pp. 1960–1973, Dec. 2013, doi: 10.1109/TNET.2013.2256431.

[9] A. Sivanathan, H. H. Gharakhani, F. Lai, A. Radford, C. Wijenayake, A. Vishwamath, and V. Sivaraman, “Classifying IoT devices in smart environments using network traffic characteristics,” IEEE Trans. Mobile Comput., vol. 18, no. 8, pp. 1745–1759, Aug. 2019, doi: 10.1109/2866249.

[10] M. Lopez-Martin, B. Carro, A. Sanchez-Esguevillas, and J. Lloret, “Network traffic classifier with convolutional and recurrent neural networks for Internet of Things,” IEEE Access, vol. 5, pp. 18042–18050, 2017, doi: 10.1109/ACCESS.2017.2745560.

[11] A. Marno, D. Gomez, and R. Aguar, “Towards IoT data classification through semantic features,” Future Gener. Comput. Syst., pp. 792–798, Dec. 2017, doi: 10.1016/j.future.2017.11.045.

[12] E. J. Palomo, J. North, D. Elizondo, M. R. Luque, and T. Watson, “Application of growing hierarchical SOM for visualisation of network forensics traffic data,” Neural Netw., vol. 32, pp. 275–284, Aug. 2012.

[13] T. Ito and A. S. M. Sajeev, “Boosted probabilistic neural network for IoT data classification,” in Proc. IEEE 16th Intl Conf Dependable, Autonomic Secure Comput., Aug. 2018, pp. 408–411, doi: 10.1109/DASC/PCom/DataCom/CyberSciTech.2018.00082.

[14] M. A. Ambusaidi, X. He, P. Nanda, and Z. Tan, “Building an intrusion detection system using a filter-based feature selection algorithm,” IEEE Trans. Comput., vol. 65, no. 10, pp. 2986–2998, Oct. 2016, doi: 10.1109/TC.2016.2519914.

[15] W. Yang, M. Wang, J. Zhang, J. Zou, M. Hua, T. Xia, and X. You, “Narrowband wireless access for low-power massive Internet of Things: A bandwidth perspective,” IEEE Wireless Commun., vol. 24, no. 3, pp. 138–145, Jun. 2017, doi: 10.1109/MWC.2017.1600298.

[16] V. Balasubramanian, F. Zaman, M. Alaqilay, S. A. M. Sajeev, and M. R. Luque, “Reinforcing the edge: Autonomous energy management for mobile device clouds,” in Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS), Apr. 2019, pp. 44–49, doi: 10.1109/INFOCOMW.2019.8845263.

[17] C. Liu, Z. Cao, Z. Li, and G. Xiong, “LaFFT: Length-aware FFT based fingerprinting for encrypted network traffic classification,” in Proc. IEEE Symp. Comput. Commun. (ISCC), Jun. 2018, pp. 1–6, doi: 10.1109/ISCC.2018.8583732.

[18] H.-H. Chang, K.-L. Lian, Y.-C. Su, and W.-J. Lee, “Power-Spectrum-based wavelet transform for nonintrusive demand monitoring and load identification,” IEEE Trans. Ind. Appl., vol. 50, no. 3, pp. 2081–2089, May 2014, doi: 10.1109/TIA.2013.2253318.
[29] R. Gregory Lee, R. Gommers, F. Wasilewski, and K. Wohlfahrt, “PyWavelets: A Python package for wavelet analysis,” Journal of Open Source Software, vol. 4, no. 36, p. 1237, Apr. 2019.

[30] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” Proc. IEEE, vol. 86, no. 11, pp. 2278–2324, Dec. 1998.

[31] A. Martin, A. Ashish, B. Paul, B. Eugene, C. Zhifeng, C. Craig, G. S. Corrado, D. Andy, D. Jeffrey, D. Matthieu, G. Sanjay, G. Ian, H. Andrew, I. Geoffrey, I. Michael, J. Yangqing, K. Lukasz, K. Manjunath, L. Josh, and Z. Xiaoqiang, “TensorFlow: Large-scale machine learning on heterogeneous distributed systems,” Google Res., Mountain View, CA, USA, White Paper, 2019. [Online]. Available: https://arxiv.org/abs/1603.04467

[32] A. H. Monahan, “Nonlinear principal component analysis by neural networks: Theory and application to the lorenz system,” J. Climate, vol. 13, no. 4, pp. 821–835, Feb. 2000.

YIFAN MENG (Student Member, IEEE) was born in Huaibei, Anhui, China, in 1995. He received the bachelor’s degree in electrical engineering and automation from Henan Polytechnic University, in 2016, and the master’s degree in electrical engineering from the Anhui University of Science and Technology, in 2019, where he is currently pursuing the Ph.D. degree. His research interests include the application of the Internet of Things, blockchain, and embedded technology in industrial production.

JINGZHAO LI was born in Huainan, Anhui, China, in 1964. He received the M.A. degree from the China University of Mining and Technology, in 1992, and the Ph.D. degree from the Hefei University of Science and Technology, in 2003. He is currently a Professor with the School of Electrical and Information Engineering, Anhui University of Science and Technology, China. His research interests include CPS, the Internet of Things (IoT) technology, and embedded systems.

* * *

YIFAN MENG

JINGZHAO LI