Neural Computing and Applications (2021) 33:897–907
https://doi.org/10.1007/s00521-020-05272-0

S.I. : ATCI 2020

Confidential information protection method of commercial information physical system based on edge computing

Xiaheng Zhang[1] · Jiazhong Lu[2] · Doudou Li[3]

Received: 23 May 2020 / Accepted: 29 July 2020 / Published online: 17 August 2020

Abstract
With the rapid integration and wide application of the enterprise Internet of Things, big data and 5G-class large networks, traditional enterprise cloud computing systems cannot timely process various mass information data generated by connecting with network edge electronic devices. There are obvious technical disadvantages. In order to effectively solve this complex problem, edge mobile computing came into being. The purpose of this article is to study the protection methods of commercial confidential information, using the security relationship between data information in edge technology computing and the technical characteristics of data privacy information protection. This paper proposes a theoretical system and technical architecture centered on the use of data security technology. The three key technologies of access control, identity authentication and information privacy security protection are researched on the privacy protection processing methods of commercial information security physical systems. The experimental data show that the recognition errors mainly occur in identifying non-anomalous data as abnormal data. The analysis and identification of abnormal information data are basically accurate, and it can quickly complete various processing tasks to eliminate abnormal information data to meet the requirements. The experimental data show that the abnormal data can be used to monitor the security of the commercial information physical system, and it is also a good security protection method for commercial confidential information. It has guiding significance for the confidential information protection of commercial information physical systems. In the next few years, more than 50% of major data applications will need to be analyzed, processed and data stored at the edge of the network. Cloud computing technologies at the edge are widely used.

Keywords Edge computing · Data security · Business information physical system · Identity authentication · Confidential protection

1 Introduction

In recent years, business information technology has developed rapidly, and the amount of data in their respective business management systems is getting larger and larger, the content is more and more complex, and the types are increasing. Commercial information also responded to this situation and immediately relied on the commercial data center to successfully establish a business big data information summary and comprehensive analysis service platform [1]. However, how to effectively protect the security of information and data related to trade secrets under the current severe network security situation has gradually become a key issue affecting the healthy and rapid development of China's business.

With the rapid and in-depth development of intelligent technologies such as the mobile Internet of Things and 5G,

✉ Jiazhong Lu
ljz@cuit.edu.cn

Xiaheng Zhang
rainchang0811@163.com

Doudou Li
601340635@qq.com

[1] School of Economics and Management, Chuzhou University, Chuzhou 239000, Anhui, China

[2] School of Cybersecurity, Chengdu University of Information Technology, Chengdu 610225, Sichuan, China

[3] Business School, Northwest University of Political Science and Law, Xi'an 710122, Shaanxi, China

the development trend of the mobile Internet of Everything continues to deepen. The Internet of Things smart technology and terminal smart mobile devices are increasingly penetrating into the daily lives of modern people. New mobile business development models such as smart mobile grids, smart cities and autonomous driving continue to emerge. The number of smart mobile device users will continue to show explosive rapid growth. According to IDC's forecast, by 2020, at least one terminal intelligent mobile device with a capacity of up to 50 billion will be able to access the mobile Internet. The following problem is the "massive" quality data generated by industrial terminal video. Taking an industrial scene video as an example, a single video camera 1080 ps format industrial video will simultaneously produce 330g of industry at a high bit rate of 4 mbps per day. Massive data terminal communication equipment resource interconnection management mode puts forward higher technical requirements for accurate response, duration and data security performance of data resource management requests. Cloud computing's "pay-as-you-go" interconnection model has completely freed traditional enterprises and network end users from the constraints of many key details in resource ownership and secure management of large amounts of data, such as saving storage resources, computing time constraints and reducing network communication occupation cost. However, in many industrial sites and other high-latency-sensitive application environments, when millions of traditional smart home devices begin to request smart services, the current smart cloud computing infrastructure is difficult to fully meet the requirements of traditional smart home devices for high-speed mobile networks. There is a great demand for support, location information awareness and low latency. As a result, a new model of data processing on the edge side of the object has been spawned, that is, the use of edge computing, and thus has received widespread attention in the information academia and information industry [2, 3]. Edge data computing reduces the load of cloud-based network computing systems by directly empowering massive intelligent computing devices located on the edge side to perform data calculations and network data processing calculations in combination with current intelligent cloud computing and centralized network data processing system models. It alleviates the transmission pressure on network data bandwidth and improves the data collection and processing transmission efficiency for massive intelligent devices [4].

Pasupuleti and other companies have proposed a public key outsourcing cloud platform data privacy security protection solution (esppa) for large mobile terminal devices. The solution mainly uses high-probability public key outsourcing encryption technology (ppke) and Baidu keyword outsourcing ranking. Search algorithm (rks), an outsourced ranking data query, can implement data privacy security protection on large mobile device terminals with limited data resource traffic [5]. First, mobile users need to generate an index of file data and encrypt the file data and other index files before uploading. Second, in order to access the department's ciphertext data stored in the password in the cloud in real time, the user generates a trapdoor password for each keyword password and publishes it to the user cloud. In the end, the cloud computing server traps according to the search results of the user and sorts all users to return the data that are stored based on the correlation of the data and the sorted results, and then automatically decrypts the original number obtained by the user [6]. Bahrami et al. proposed a lightweight cloud encryption storage method for us in the current environment of mobile computing, multi-cloud and cloud computing, which was originally used to store the encrypted data in the cloud [7]. This encryption method mainly uses a pseudo-random encryption permutation (prpm) algorithm based on the chaotic encryption system to directly implement lightweight replacement encryption. The replacement encryption operation is usually performed directly on a mobile communication device, rather than in the entire cloud.

To effectively protect the privacy of user databases [8, 9], this article first introduces the environmental background of big data and the security requirements of trade secret protection and then discusses the security management of data in a big data environment and the technical protection platform. Based on the edge computing, this paper proposes a method for protecting confidential information of commercial information physical systems. Finally, experiments are performed based on distance to determine abnormal data. The identification of abnormal data is basically accurate, and the task of eliminating abnormal data can be completed [10]. Through data acquisition experiments in human behavior recognition scenarios, it was found that recognition errors mainly occurred in identifying non-anomalous data as abnormal data, and monitoring business information from identifying the abnormal data can provide a good protection for commercial information physical system method.

2 The methods of classroom research

2.1 Edge computing architecture

The "edge" in marginalized cloud computing networks is actually a relatively narrow concept. It refers to any computing network resource and other network computing resources transmitted from one data source to the entire cloud computing data center. Edge terminal computing

capabilities allow terminal devices to simultaneously migrate data storage and application computing capabilities tasks to multiple network edge terminal nodes, such as wired base stations (bs), wireless network access points (wap) and edge terminal servers. It meets the requirements of higher computing processing capacity and scalability of network terminal equipment, and at the same time, it can effectively save computing power tasks to complete the data transmission between the network cloud terminal server and the network terminal equipment and the link in the link. The basic architecture of edge core computing is mainly divided into four main functional architecture levels: edge core computing infrastructure, edge computing data center, edge computing network and terminal mobile application terminal [11].

2.2 Data security and privacy protection

The main content of data security and privacy protection refers to the security of outsourced business data, including the security and confidentiality of external data and secure data sharing, integrity data auditing and secure and reliable search data encryption [12, 13]. Data sharing confidentiality and secure data sharing usually need to be implemented using data encryption technology. Usually, the data are encrypted first and then sent back online. The traditional data encryption processing methods mainly include symmetric password encryption and other asymmetric. The password is encrypted, but the encryption process of the user's subsequent shared data will also provide some difficulties. Currently, the widely used file attribute-based information encryption technology is a threshold access policy based on information attributes. When a user needs to possess certain attribute-based information, it can access and decrypt these information attributes. This kind of access threshold from an original monotonous hierarchical access threshold tree gradually evolved to form a multi-level access feature. The file attribute ciphertext encryption solution based on the file ciphertext protection strategy is widely used in enterprise cloud computing and database storage and information sharing security. The repeated encryption key algorithm of the data proxy is usually a semi-trusted data agent. It can directly convert the ciphertext into the proxy ciphertext for multiple data proxy users at the same time through the repeated encryption key. Encrypted plain text cannot be obtained simultaneously between data agents. Message data proxy reconstruction and encryption are widely used in cloud security system applications for large data message forwarding, file data distribution and multi-user data sharing [14]. The problem that complete information audit needs to solve is how to completely determine the information integrity and data availability of audit data, and the audit functions include dynamic management audit, batch management audit, privacy protection audit and low work complexity. Programmable search data encryption technology refers to the effective search and query of encrypted edge data information in the mobile phone cloud. One of the main problems currently solved for cloud edge data calculation is that the complexity of its processing algorithm is too high. During the calculation, a large amount of calculation power is generated. At present, the support of elastic search code encryption is mainly divided into security and ranking support, elastic search code encryption based on agent attribute, elastic search code encryption supporting agent dynamic code update and agent importance encryption supporting elastic search encryption.

Privacy security protection can be subdivided into user data privacy security protection, location information privacy security protection and user identity privacy security protection. The main applications are concentrated in cloud computing for mobile phones and other cloud computing applications. The protection of location data privacy information is mainly to ensure that host users can flexibly use the existing data while ensuring the security of the data. The location information privacy data protection scheme is based on the anonymous location server. The data use the *k*-server's anonymous data algorithm, but because the k-server's anonymous data algorithm usually consumes a lot of data bandwidth in practice, it is particularly important to design a lightweight and efficient location privacy data protection solution [15]. The protection of identity privacy right now is only explored in various mobile and cloud computing application environments. The edge computing data security and privacy protection research system diagram is shown in Fig. 1.

2.3 Authentication

Because the current edge information calculation process is a multi-function entity's real-time calculation processing paradigm, which itself contains many functional domain-based entities, the functional entities between different trust domains in edge management calculations currently need to switch to each other in real time. Verification, including internal identity information authentication technology in a single domain, is used to mutually resolve the internal identity information allocation authentication problem of each functional entity, and the entity has certain self-privacy information protection features; cross-domain identity information authentication technology is used for mutual authentication. The value of identity authentication between entities in different trust domains is extreme, and the current technology is in the initial stage of development; real-time switching identity authentication technology is used for various mobile terminals and handover

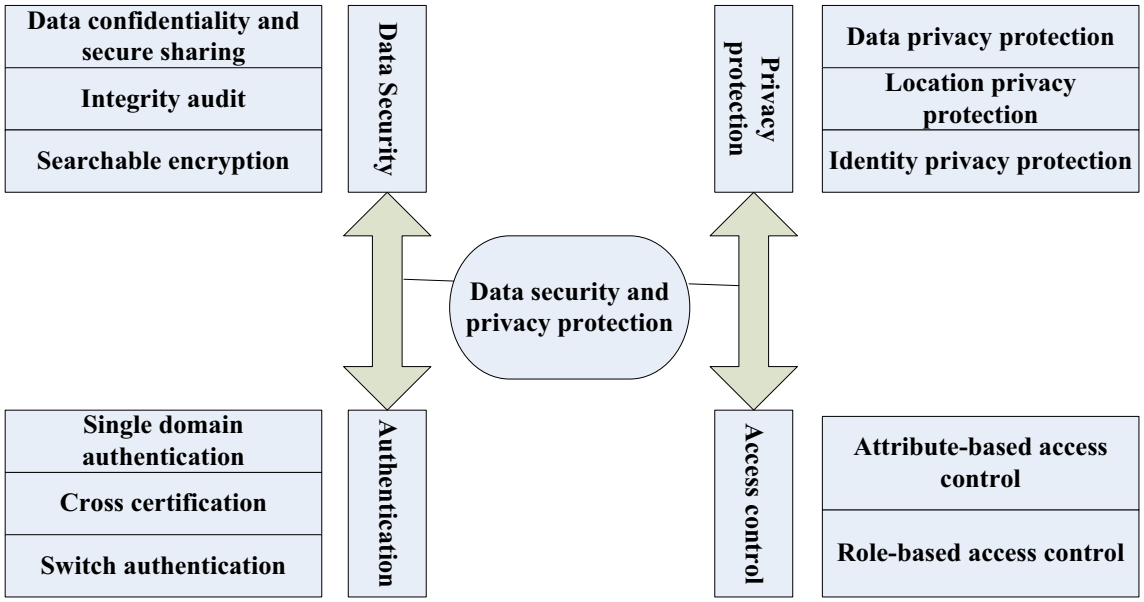

Fig. 1 Research system chart of computing data security and privacy protection

devices with high-precision mobile performance and other characteristics in current edge management computing. A tailor-made real-time handover management technology based on user internal identity information authentication ensures the real-time handover accuracy of the device, but switching authentication is important for the privacy protection of user IDs during the handover task.

There are several types of identity authentication:

1. Identity authentication in a single domain

The security identity information authentication in a single authorized trust domain is mainly used to solve the security identity information distribution security problem of each network entity. First, the respective entities must first pass the security identity authentication of a single authorized data center to be able to directly access the information store and then data computing.

2. Cross-domain authentication

At present, the theoretical research on communication authentication mechanisms applicable to different communication domain authentication entities is still in the initial stage of theory, and a relatively complete theoretical context of authentication research and specific theoretical research methods have not been basically formed. In the research of cross domain identity authentication management based on cloud computing, the multi cloud identity management and authentication between different cloud computing service providers can be regarded as or a new cross domain identity authentication management form, which can make some authentication management machines suitable for single multi cloud authentication management standards and multi cloud single sign on (SSO). It is hoped that the system can be widely applied to multi cloud identity management authentication between multiple trusted domains [16].

3. Switch authentication

Due to the high mobility of mobile terminal devices in mobile edge network computing, the geographic location of mobile users often changes, making traditional mobile centralized user identity information authentication network protocols no longer applicable in such situations. Authentication switching identity authentication is an application authentication handover application technology to effectively solve the user's personal identity information authentication of high-edge mobile performance. Therefore, in-depth research on authentication switching identity authentication handover technology can not only provide a strong technical guarantee for real-time accurate information authentication handover of users using edge mobile devices in edge mobile computing, but also solve the user's personal identity in the practice of applying authentication handover technology. The issue of authentication privacy protection has also been the focus of academic research [17].

Because edge computing is an open dynamic system in which multiple entities and multiple trust domains coexist, the identity authentication protocol must consider the correspondence between entities and trust domains. Specific research contents include cross-domain authentication and switch authentication of the same entity between different trust domains; identity authentication and mutual authentication of different entities in the same trust

domain; and finally, while achieving lightweight identity authentication, taking into account anonymity and integrity. Features such as reliability, traceability and batch certification are also important research points.

2.4 Access control

Remembering the real-time control of the resource access process based on predetermined models and policies is the key technology and important method to ensure system security and protect user privacy in edge computing. Based on the system architecture of edge computing, multiple entities in different trust domains put forward attribute-based access control and role-based access control. Access control is divided into the following types:

1. Attribute-based access control

Because edge computing is a data-dominated computing model, access control for edge computing is usually implemented using cryptographic techniques. Traditional cryptographic techniques are not suitable for distributed parallel computing environments, and attribute encryption (ABE) works well. It is suitable for distributed architectures to achieve fine-grained data sharing and access control [18].

2. Role-based access control

Role-based access control provides flexible control and management through a dual permission mapping mechanism, that is, user-to-role and role-to-data object permission mapping.

2.5 Data preprocessing module

According to the design of the data preprocessing module for edge computing, the data type and local computing power are considered. After selection, this paper uses a distance-based method to determine abnormal data. For the processing of abnormal data, this article will completely delete the found abnormal data and then fill it with the missing data.

Considering the characteristics of the timeliness of environmental data and the limitations of processor performance, this paper will use the nearest neighbor padding of hot card padding when performing data padding on edge devices, combined with Lagrange interpolation. The essence of Lagrangian interpolation is to construct an interpolation function that is simple and has sufficient accuracy based on known node data or data from some known points on the line graph and use this interpolation function to quickly obtain the location data; this method of using several nodes to construct the interpolation function is called Lagrange interpolation [19].

The mathematical expression is that the number of interpolation nodes on the original function $f(x)$ is $n + 1$, which are (x^0, y^0), (x^1, y^1), (x^2, y^2),…, (x^n, y^n), substituted into the following formula:

$$y = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots + a_{n-1} x^{n-1} \tag{1}$$

The expression of the simultaneous available Lagrange interpolation polynomial is:

$$y = p(x) = \sum_{i=0}^{n} y_i \prod_{i=0, i \neq j}^{n} \frac{x - x_i}{x_i - x_j} \tag{2}$$

After collecting the user's face information, the authority distribution authority will send the user's authority vector Li to the edge computing node; Li is an n-dimensional zero-bit string representing the user's authority information for the building where the n edge computing nodes are:

$$L_i[j] = 1 \quad (1 \leq j \leq n) \tag{3}$$

On behalf of the user has the right to enter the building where the jth edge computing device is located, otherwise if:

$$L_i[j] = 0 \quad (i \leq j \leq n) \tag{4}$$

it means that the user does not have the permission.

After calculating the nearest points from the missing data by the above formula and substituting them into the above expression, a corresponding polynomial can be obtained. Then, the point X corresponding to the exact function value is substituted into the interpolation polynomial to obtain the approximate value of the missing value.

For the extraction of human behavior recognition feature vectors, the edge computing device uses the human behavior feature extraction SDK provided by the cloud service provider to perform feature extraction on the user's human behavior picture and extracts a 160-dimensional feature vector from the face picture, which is recorded as:

$$f_i = \left(f_{i,1}, f_{i,2}, \ldots, f_{i,160}\right)^{\mathrm{T}} \tag{5}$$

Make:

$$\hat{f_{ia}}[j] = \hat{f_{ib}}[j] = \hat{f_i}[j]; \quad S[j] = 1 \tag{6}$$

Calculate the final encryption result:

$$\left(M_1^T \ \hat{f_{ia}}, M_2^T \ \hat{f_{ib}}\right) \tag{7}$$

First t edge computing devices ($1 \leq j \leq t$) calculate:

$$U_j^i[k] = Z_{ij}[k] \prod_{l=1, l \neq j}^{t} \frac{-x_l}{x_j - x_l} (\mathrm{mod}\ p) \tag{8}$$

The transformation of the feature vector rewrites the feature vector as:

$$\hat{f_q} = \left(\frac{f_{q,1}}{\|f_q\|}, \frac{f_{q,2}}{\|f_q\|}, \ldots, \frac{f_{q,160}}{\|f_q\|} \right) \tag{9}$$

2.6 Data acquisition for human behavior recognition scenarios

Human behavior recognition is a technology that collects specific information about human behavior and then analyzes and judges human behavior status [20]. Based on this definition, the workflow of human behavior recognition is to obtain data information of human behavior through various channels, analyze the original data, extract features that have the ability to distinguish between different behavior types, model the sample data and finally realize the classification and recognition of human behavior [21].

However, in traditional human behavior recognition systems, the analyzed data are generally existing data that have been collected, so they do not include a data acquisition module for real-time data transfer. On the one hand, this architecture is not conducive to correcting the model after the model is completed due to the lack of real-time and effective feedback information and universality [22]. When edge devices are introduced, on the one hand, edge devices and their derivatives can serve as data collection modules to avoid the problems mentioned above. At the same time, the edge device can be used to test the new model that has been established. By comparing with the data collected by itself, it can timely feedback the prediction results to modify the existing model, so as to build a behavior recognition model for the current individual and improve the accuracy of model prediction degrees [3, 23].

Therefore, the solution determined in this article is to use the relevant sensors connected to the edge device to complete the real-time collection of data required for human body recognition and then preprocess the data accordingly and send the data to the cloud. The cloud computing center uses related machine learning algorithms to complete the modeling work and sends the model back to the edge device. The edge device uses the trained model to predict the pose behavior of the current surveillance object. When the model prediction result does not match the current status, it can also notify the cloud computing center of the data and the correct result to modify the relevant model [24].

3 The process

3.1 Experimental settings

Because the collected human behavior data are physically accelerated and angular velocity transmitted by the data pre-processing module, such stream data have no specific physical meaning when analyzed independently, so model training cannot be performed. Therefore, in order to obtain training samples suitable for machine learning models, the stream data need to be converted into sub-sequences divided by a certain size before the formatted data are analyzed.

The algorithms under this module are implemented using Python programming. The experimental test runs are divided into two parts: the cloud computing center and the edge device. The specific configurations are as follows:

Cloud computing center: Intel (R) Core (TM) i5-4570; CPU dual core@3.20 GHz; 4.00 GB memory; Windows 7 64-bit operating system.

Edge device: BCM2863; ARM Cortex-A7 CPU; 1 GB memory; RASPBIAN JESSIE WITH PIXEL operating system.

In the current scenario, the system's training for human behavior recognition is done using data stored in the relevant database on the UCI database. The data collection method in this database is similar to the method described earlier in this chapter. All three IMU devices are bound to the subject's chest, arms and ankles to collect three-dimensional data of the human body. In addition, the database also contains a record of the real-time heartbeat frequency of the detected object. The collection frequency of all data is 100 Hz.

3.2 Experimental data and parameters

During the monitoring period, the behavior of the monitored objects was subdivided into 19 items by the UCI database. However, due to the needs of the test in this experiment, these behaviors are summarized into four main behavioral characteristics: sitting, walking, running and standing. Behaviors that are not monitored temporarily such as cycling and skipping will not be added to training among the data.

After sorting the data according to the above requirements, a total of 34,091 experimental data were obtained, which included relevant data for a total of 8 men and women monitored. As mentioned above, there are a total of 20 variables in this data set as input to the model.

After processing the processed UCI data set, the parameters obtained are shown in Table 1.

Because Xgboost comes with cross-validation, we can call the corresponding library function to complete the test of the model. We perform a 50-fold cross-validation method on the existing data, as shown in Table 2.

4 Interpretation of result

4.1 Xgboost model

1. The parameter interface is provided by Xgboost developers under python. Users can modify the relevant parameters of the model according to their own needs, which has achieved the ideal model training effect.

An important parameter that needs to be adjusted in a decision tree is the height of the decision tree. Due to the characteristics of Xgboost's algorithm, the shape of the basic classifier is required to generate the model, that is, the number of leaves and the vertical depth of the decision tree play a very important role in the process of model building. Therefore, when doing model training, we need to find a suitable value to ensure that the final classifier can meet the needs of users on the one hand, and it will not affect its applicability on other related data.

Therefore, in this experiment, we need to control the maximum depth of the final decision tree by modifying the max_depth parameter under the xgb.train () function. The obtained parameters are shown in Table 3.

For the current model, because the classification result of the data set is four categories, the soft function multi-classifier is used as the loss function. Generally speaking, when the learning rate is small, the generalization ability of the model will be higher than that without the learning rate. The model improves a lot. Generally, the learning rate is inversely proportional to the number of iterations. The lower the learning rate, the more iterations are required. In this experiment, in order to make each learning more accurate, the learning step is set to 0.15, during the iterative process. The deeper the maximum depth of each tree, the more features are learned, but it will also reduce the generalization ability of the model. Therefore, the depth is set to 8; the multi-class error rate is used as the data measurement method. The data graph is shown in Fig. 2.

2. Cross-validation: The per-judgment accuracy of the existing model for the training data is about 98.3%, and the accuracy will continue to increase as the cycle deepens [25–27]. However, because the decision tree itself is a greedy algorithm, theoretically, if the number of iterations is infinite, the accuracy of the prediction will reach 100%. However, in this case, over-fitting generally occurs, so you need to adjust the values of the two parameters min_child_weight and max_depth in Xgboost to prevent over-fitting. By setting the parameter mun_round that controls the number of iterations, the accuracy of classification can also be further improved in the model obtained after the above training, as shown in Fig. 3.

Table 1 Comparison of algorithms in UCI data sets

Algorithm	*F*-measure	Accuracy	Running time
Decision tree (C4.5)	0.9334	0.9324	28.95
GBDT	0.9446	0.9468	35.34
Random forest	0.9235	0.9366	44.24
Naive Bayes	0.9236	0.9483	36.45
Xgboost	0.9642	0.9854	18.67

4.2 Comparison test

1. Compared with other basic algorithms, Xgboost runs faster and has greater advantages in accuracy and recall. Compared to weak classifiers such as C4.5 decision trees, the advantages of integrated learning algorithms are more obvious. Since Xgboost has more regularization of self-models than other models, this type of model has stronger generalization ability. In actual work, as the amount of data continues to increase, the accuracy advantage of Xgboost will be correspondingly more obvious [28, 29].

However, it should also be pointed out that comparing the model performance during parameter optimization, it will be found that it is impossible to greatly improve the performance of the model only by adjusting the parameters and small optimization of the model. To achieve the overall qualitative leap in classification accuracy, we also need to rely on other means, such as feature engineering and model combination. For the design of the data analysis function in this edge system, in addition to completing the training of the classification model in the cloud computing center, it is also necessary to use the trained model in the edge device to implement the classification test on real-time data.

First, you need to build a software environment in the edge device that can run the classification algorithm. Because Xgboost has a corresponding optimization algorithm for memory, and the calculation amount of the algorithm itself is mainly focused on the construction of the model, the edge device selected in the current environment, the Raspberry P_i, can fully bear the human behavior recognition model based on Xgboost Testing works. The comparison test is shown in Fig. 4.

Table 2 Fold cross-validation method data

S. no.	Test-merror-mean	Test-merror-std	Test-merror-mean	Test-merror-std
0	0.022881	0.001195	0.019118	0.001647
1	0.020006	0.000853	0.015430	0.001645
2	0.019507	0.000932	0.014916	0.001223
3	0.018510	0.000795	0.014029	0.001349
4	0.018305	0.001218	0.013853	0.001215

Table 3 Xgboost model parameters

Parameter	Value
Objective	0
Eta	0.15
Max_depth	8
Nth-read	4
Num_class	4
Eval_metric	0
Min_child_weight	1

2. Analysis of accuracy and maximum depth: As the depth increases, the classification effect of the Xgboost model becomes better. However, when the depth reaches 8, the increase of this effect becomes less obvious, which indicates that if the depth of the decision tree continues to be increased, the current model cannot effectively improve the classification ability, but will cause the model to over-fit the current data and over-fitting. Therefore, it is finally determined that when the value of the parameter max_depth is set to 8 under this experiment, the best classification effect can be obtained.

Another parameter that needs attention is eta. This parameter is equivalent to another parameter learning_rate under the more popular machine term library learn. This parameter can improve the robustness of the model by reducing the weight of each step, that is, reducing the weight of features to make the promotion calculation process more conservative. In actual use, it will be set between 0 and 1 according to the actual situation; usually, the range is [0.01–0.2], as shown in Fig. 5.

5 Conclusions

This article mainly analyzes the information security requirements of edge computing from the perspective of identity authentication and privacy protection. In edge computing, due to different deployment options, edge computing service providers can be located in different places. In this context of open interconnection. The identity authentication and management functions are spread across all functional levels of the edge computing reference architecture. The user's identity authentication is the first line of defense for edge computing. Identification is critical to ensuring the security and confidentiality of applications and data. In this paper, human body recognition technology in commercial information protection methods is studied through data collection and preprocessing, mainly to enable data mining. Regardless of whether you use existing models or algorithms designed by yourself, you need a hardware environment that can be implemented. This includes not only the basic computing power required by the algorithm, but also appropriate storage space for data and model storage and scheduling. When multiple sensors

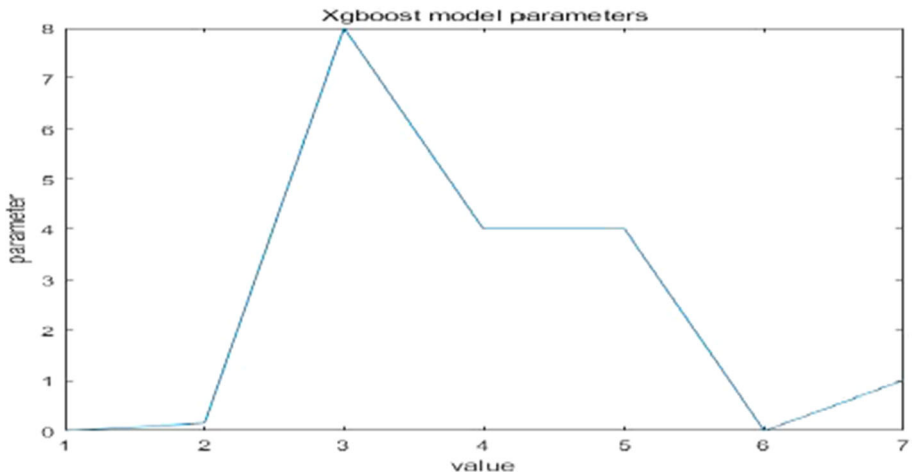

Fig. 2 Xgboost model parameters

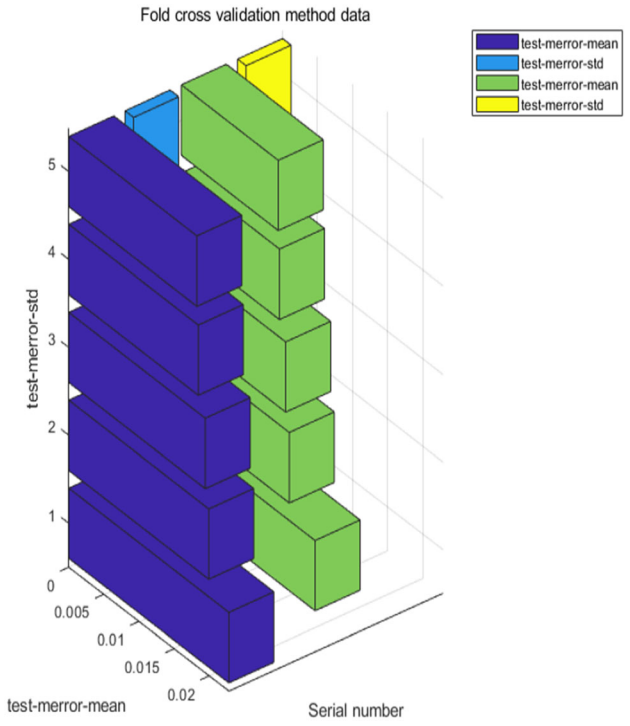

Fig. 3 Fold cross-validation method data

send data to the edge device, each sensor uploads its respective data to the edge device through parallel transmission. For network data, the edge device itself needs to access the target web page through the network module, download the required data and complete the data collection task. The data exchange between the edge device connected to the Internet and the cloud computing center is completed by wireless transmission, so no other device is needed for data collection.

This article is aimed at the research of confidential protection of commercial information physical systems, and encryption algorithms can be used. Existing data confidentiality and secure data sharing solutions are usually implemented using encryption technology. The conventional process is that the data owner encrypts and uploads the outsourced data in advance and decrypts the data when needed. Data encryption technology provides effective solutions to ensure data security in various computing modes. In an open edge computing environment, traditional encryption solutions can be combined with features such as parallel distributed architecture in edge computing, limited terminal resources, edge big data processing and highly dynamic environments to achieve lightweight and distributed data security protection system.

Human behavior recognition technology has been applied in various fields of people's clothing, food and shelter. Compared with traditional password-based authentication, human behavior recognition technology has the advantages of being directly friendly, not easy to be stolen and noninvasive and has become an application in the field of identity authentication mainstream. A typical solution is to use a third-party human behavior recognition technology provider to achieve accurate face recognition. In this paper, xgboost model is used to do data preprocessing module experiments to determine the abnormal data. The identification of abnormal data is basically accurate, and it can complete the task requirements of eliminating abnormal data. It is applied to a popular business information confidentiality protection method human behavior recognition scene data acquisition. The

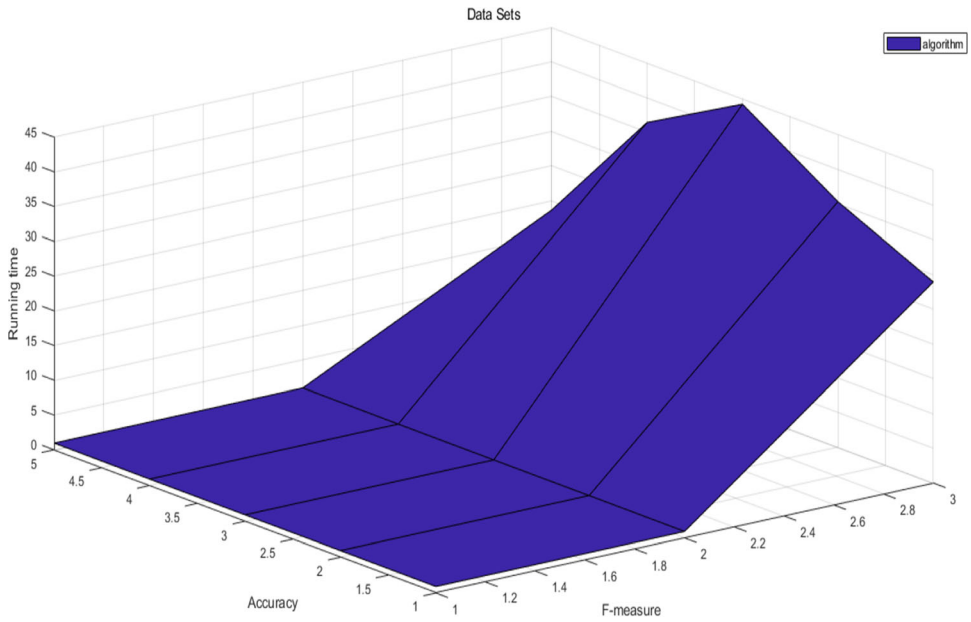

Fig. 4 Comparison of algorithms in UCI data sets

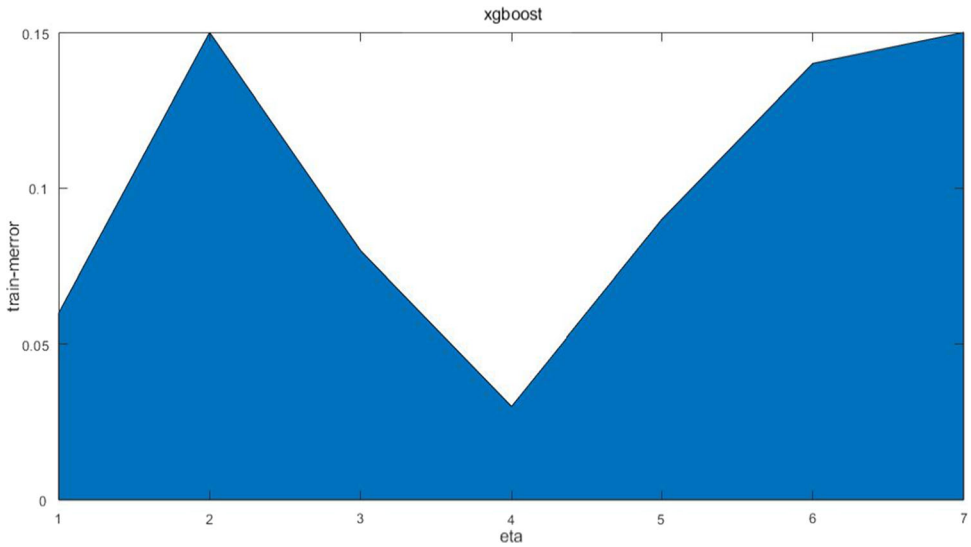

Fig. 5 Relationship between accuracy and learning rate

experimental data show that the identification error mainly occurs in identifying non-anomalous data as abnormal data. By identifying the abnormal data, it can play a monitoring role in the security of the commercial information physical system, and it is also a good security protection method for commercial confidential information. The experimental data show that the comparison of data after identifying anomalies can physically monitor the security of business information and has guiding significance for the confidential information protection of business information physical systems.

Acknowledgements This work was supported by Social Science Foundation of Shaanxi Province "Sustainable trade promotion of cross border e-commerce in China's Silk Road Economic Belt" (2019S037) and the Young Academic Innovation Team of Northwest University of Political Science and Law.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interests in this work.

References

1. Newman R, Chang V, Walters RJ (2016) Model and experimental development for business data science. Int J Inf Manag 36(4):607–617
2. Satyanarayanan M (2017) The emergence of edge computing. Computer 50(1):30–39
3. Shi W, Dustdar S (2016) The promise of edge computing. Computer 49(5):78–81
4. Rankothge W, Le F, Russo A (2017) Optimizing resource allocation for virtualized network functions in a cloud center using genetic algorithms. IEEE Trans Netw Serv Manag 14(2):343–356
5. Huang T (2019) Research on cluster mining algorithms for personal privacy protection in the background of big data. J Phys Conf Ser 1314(1):012153
6. Xie PS, Fu TX, Fan HJ (2019) An algorithm of the privacy security protection based on location service in the internet of vehicles. Int J Netw Secur 21(4):556–565
7. O'Brien LL, Guo Q, Bahrami-Samani E (2018) Transcriptional regulatory control of mammalian nephron progenitors revealed by multi-factor cistromic analysis and genetic studies. PLoS Genet 14(1):e1007181
8. Martin KD, Borah A, Palmatier RW (2017) Data privacy: effects on customer and firm performance. J Mark 81(1):36–58
9. He Y, Zhang Y, Wang X (2020) A new image encryption algorithm based on two-dimensional spatiotemporal chaotic system. Neural Comput Appl 32:247–260
10. Cheng J, Xu R, Tang X (2018) An abnormal network flow feature sequence prediction approach for DDoS attacks detection in big data environment. Comput Mater Contin 55(1):95–119
11. Sabella D, Vaillant A, Kuure P (2016) Mobile-edge computing architecture: the role of MEC in the Internet of Things. IEEE Consum Electron Mag 5(4):84–91
12. Li Y, Yu Y, Min G (2017) Fuzzy identity-based data integrity auditing for reliable cloud storage systems. IEEE Trans Dependable Secure Comput 16(1):72–83
13. Al-Hazaimeh OM, Al-Jamal MF, Alhindawi N et al (2019) Image encryption algorithm based on Lorenz chaotic map with dynamic secret keys. Neural Comput Appl 31:2395–2405
14. Kalaiprasath R, Elankavi R, Udayakumar DR (2017) Cloud security and compliance—a semantic approach in end to end security. Int J Mech Eng Technol 8(5):987–994
15. Cohen J, Lefevre J, Maâmra K (2016) A self-stabilizing algorithm for maximal matching in anonymous networks. Parallel Process Lett 26(04):1650016
16. Castiglione A, Palmieri F, Chen CL (2016) A blind signature-based approach for cross-domain authentication in the cloud environment. Int J Data Wareh Min (IJDWM) 12(1):34–48
17. Mattos DMF, Duarte OCMB (2016) AuthFlow: authentication and access control mechanism for software defined networking. Ann Telecommun 71(11–12):607–615
18. Goulas KA, Gunbas G, Dietrich PJ (2017) ABE condensation over monometallic catalysts: catalyst characterization and kinetics. ChemCatChem 9(4):677–684
19. Jantsch P, Webster CG, Zhang G (2019) On the Lebesgue constant of weighted Leja points for Lagrange interpolation on unbounded domains. IMA J Numer Anal 39(2):1039–1057

20. Chen L, Chen X, Ni L (2017) Human behavior recognition using Wi-Fi CSI: challenges and opportunities. IEEE Commun Mag 55(10):112–117
21. Kuo CFJ, Juang Y (2016) A study on the recognition and classification of embroidered textile defects in manufacturing. Text Res J 86(4):393–408
22. Liu Y, Kuang Y, Xiao Y (2017) SDN-based data transfer security for Internet of Things. IEEE Internet Things J 5(1):257–268
23. Cheng B, Li TY, Wei PC (2018) Layer-edge device of two-dimensional hybrid perovskites. Nat Commun 9(1):1–7
24. Wang S, Sun S, Li Z (2017) Accurate de novo prediction of protein contact map by ultra-deep learning model. PLoS Comput Biol 13(1):e1005324
25. Liu H, Kou H, Yan C (2019) Link prediction in paper citation network to construct paper correlation graph. EURASIP J Wirel Commun Netw 2019(1):1–12
26. Wenwen G, Lianyong Q, Yanwei X (2018) Privacy-aware multidimensional mobile service quality prediction and recommendation in distributed fog environment. Wirel Commun Mob Comput 2018(4):1–8
27. Chai X, Fu X, Gan Z et al (2020) An efficient chaos-based image compression and encryption scheme using block compressive sensing and elementary cellular automata. Neural Comput Appl 32:4961–4988
28. Zhang X, Zhou S, Fang J, Ni Y (2020) Pattern recognition of construction bidding system based on image processing. Comput Syst Sci Eng 35(4):247–256
29. Song H, Srinivasan R, Sookoor T et al (2017) Smart cities: foundations, principles, and applications. Wiley, New York