INTELLIGENT ALGORITHM OPERATION AND DATA MANAGEMENT OF ELECTROMECHANICAL ENGINEERING POWER COMMUNICATION NETWORK BASED ON THE INTERNET OF THINGS

YING LI, WENJING QU, AND ZHENQIANG ZHANG

Abstract. With the rapid development and gradual popularization of Internet of Things technology, it is necessary to provide necessary Operation and Maintenance (O&M) services and data control for electromechanical engineering to maintain the normal function of the power system. This paper does the following research to carry out orderly and standardized management of related communication resources in power communication networks, conduct closed-loop and process-oriented management of communication O&M work, and ensure the safe, stable, and economical operation of power grids in electromechanical engineering. Firstly, the technology selection, algorithm implementation, solutions, and other related technologies that may be used in the platform design and implementation process are introduced and selected. Secondly, the research and call logic of the design and implementation of related algorithms for intelligent O&M are introduced. It includes the design and implementation of anomaly detection algorithms to monitor equipment health status and the design and construction of fault diagnosis algorithms for abnormal analysis. Finally, the simulation experiment of the proposed processing scheme is carried out on the Mininet simulation network to prove that the proposed scheme can provide a better anonymization effect when introducing low latency. The results show that the design of the gateway system realizes the module applications of the system, such as user, data storage, O&M, and fault management. Based on the technical selection, the algorithm is implemented and optimized, and the call logic of the algorithm is implemented in the O&M module. Simulation verifies that the anonymization algorithm can complete the mapping without introducing an additional delay of more than 3%.

Key words: Internet of things, electromechanical engineering, power communication networks, intelligent operation and maintenance, data control

1. Introduction. The safety and stability control system and dispatching automation system of the power communication network and power system are important supports for maintaining the safe production of the power system. Also, it is the basis for grid dispatching automation, marketization of power grid operation, and modernization of management and provides important maintenance means to ensure the safe, stable, and economic operation of the power grid. Orderly and standardized management of relevant communication resources in power communication networks, centralized and intelligent monitoring of communication equipment and communication services, closed-loop and process-oriented management of communication Operation and Maintenance (O&M) work, and the formation of intelligent, integrated, and automated intelligent full-process management and control network of power communication network are effective means to improve work efficiency and ensure the safe operation of communication equipment [1].

This paper confirms and analyzes the O&M requirements and O&M status of electromechanical equipment and studies the implementation direction and system architecture innovation of intelligent O&M and Internet of Things (IoT) systems. Then, the overall system is designed and implemented with the idea of microservices, and the architecture pattern of microservices is formed. Platform services are provided through the architecture pattern, forming a low-coupling structure to meet the needs of convenience and scalability. Next, the box-plot, time series anomaly detection algorithm, and isolated forest algorithm are studied and presented separately. After testing and evaluating the characteristics of the three algorithms, a hierarchical intelligent detection
mechanism is designed to cope with the needs of anomaly detection in different situations. After anomalies are detected, a Decision Tree (DT) algorithm is used to troubleshoot based on anomalous data. It realizes the intelligent O&M of the system.

Data control opens up some data center interfaces. In this case, the data center of the power communication network will be significantly more likely to be attacked, and the attack methods will be very different. Data center traffic needs to be dynamically processed to prevent grid data from being improperly stolen or leaked during transmission. Through the centralized dynamic management and control capability provided by Software Defined Network (SDN), the efficient and fine-grained anonymization of data transmission is completed to achieve the purpose of security protection of power communication networks.

2. Literature Review. The IoT is a new industrial concept based on the concept of the Internet. It consists of the Internet as the core of interaction and the basis of communication. IoT involves industrial, home, medical, and logistics aspects. It is a network that interconnects the entire society. According to Muteba’s research, the number of IoT devices worldwide was expected to reach 22 billion by 2025 [2]. Jalali said that the domestic output of the IoT industry was growing at a rapid rate of 20% per year, and the types and number of IoT devices were growing exponentially [3].

Algorithmic Intellectual Property Operations is an O&M with algorithms as the core. It is the latest O&M mode formed by O&M services with the continuous development of computing power, equipment, and concepts. At present, domestic and international intelligent O&M services are often integrated on large-scale platforms. For example, in the Amazon Web Services provided by Wei, the intelligent O&M function is only the overall edge function of the platform. The O&M function is relatively simple. The platform includes many functions not required by intelligent O&M services, which are prone to additional overhead [4]. You proposed a new architectural pattern that is different from traditional monolithic architecture. Its main feature is the division and treatment of system functions. The system’s overall function is split, and the single function of the split is realized in a monolithic architecture pattern [5].

Network traffic control includes private information such as IP addresses and network ports of network users. If maliciously detected and analyzed by the outside world, it will cause the leakage of this information, violating user privacy and even trade secrets. Hammoudeh proposed the idea of using the programmability of SDN for anonymous services. It realized the hiding of network resources through address mapping [6]. Later, Iordache proposed some other anonymization algorithms. These algorithms achieve the purpose of shielding private information from the outside world by hashing information such as IP addresses to varying degrees. Meanwhile, some algorithms reduce the impact of anonymization operations on performance by caching [7].

Compared with foreign network management research and network management system development, China started late. After more than ten years of construction, China’s network management standardization research has also achieved great results. However, the current intelligent O&M platform pays little attention to the emerging industry of the IoT. Besides, the current domestic and international intelligent O&M services are relatively simple O&M functions. The platform as a whole includes many functions that are not required by intelligent O&M services, which is prone to additional overhead. It can be seen that the intelligent O&M platforms on the market cannot meet the requirements. So, the innovation points here are as follows. First, the intelligent O&M mode is combined with the IoT industry to realize an intelligent O&M platform focusing on IoT devices’ monitoring and O&M services. Second, it is very necessary and meaningful research to ensure the lightweight and scalability of the management and control platform. At present, the lack of this part of the content on the market should be filled to meet the overall development trend of the IoT. Third, anomaly detection depends on the implementation and optimization of algorithms. However, no one algorithm is suitable for all anomaly detection scenarios. Therefore, this paper selects and combines algorithms to achieve the optimal detection function.

3. Methods and materials.

3.1. Technologies related to the intelligent O&M platform of the power grid.

3.1.1. Microservices architecture. The microservice architecture realizes convenient development and flexible deployment and improves the system’s scalability by redefining the separation and communication of services [8]. It is proposed that the power grid intelligent O&M platform has requirements for lightweight and
flexibility. The microservice architecture can realize the design of the system platform and further extend the system for this. As a result, the microservice system architecture is used to build the server-side system.

Spring Cloud is a mature framework that combines various mainstream frameworks. The implementation of individual services relies on the Spring Boot framework of the same company. Figure 1 shows the Spring Cloud framework architecture.

The Spring Cloud framework integrates with current mainstream third-party components or frameworks to provide microservices infrastructure capabilities. It encapsulates third-party components or frameworks during the integration process, shields the internal complex configuration and implementation principles, and forms a simple, reliable, and mainstream microservice system framework. There are specific requirements for lightweight deployment in the scenario of intelligent O&M for IoT devices. The non-intrusive Istio framework is not suitable for additional deployment tasks. The system studied here is designed and implemented based on the Java language, so there is no need for the Thrift system for microservice communication between multiple languages. Finally, Spring Cloud integrates many of the current mainstream components and frameworks. Additionally, it is widely used. As for stability and community considerations, Dubbo, as a recently restarted project, is far inferior to Spring Cloud.

3.1.2. Relational databases. Relational databases store data in a relational model. The data location is located by rows and columns inside the database. Several rows and columns form a table, and multiple tables form a database [9]. Figure 3.2 displays the database storage structure.

The basic data storage unit for Mongo (Humongous) is a document. Its structure uses a Binary Serialized Document Format similar to JavaScript Object Notation. As a mainstream document-based database, Mongo
Intelligent Algorithm Operation and Data Management of Electromechanical Engineering Power Communication Network

3.1.3. Intelligent O&M algorithm.
1. Isolation Forest (iForest) is a fast anomaly detection algorithm. The basic idea of the algorithm is ensemble learning, so it has certain advantages in linear time complexity and accuracy [10]. The rationale for iForest is to define anomalies as “outliers that are easily isolated.” Each data point is divided into a separate space by constantly cutting the data space from different characteristics. Multiple cuts are required, and sparsely distributed points can be divided into space by a few cuts [11].
2. DT is a commonly used classification method. Through the classification results of known data and the corresponding occurrence probability, the logical judgment of if-then-else is constructed by the effect of factoring. The classification method of sample data is obtained. DT is an algorithm based on probabilistic analysis [12]. In constructing if-then-else logic, binary trees are usually used as logical structures, so the resulting algorithm is called a DT.
3. Box-plot is a statistical map based on statistical principles to show data distribution. The box-plot can display the distribution characteristics of the data to judge the two data with a small probability of abnormalities to achieve anomaly detection [13]. The plot of the box-plot is achieved by quartiles, which reduce the influence of extreme values in the sample data with quartiles. The distance judges anomalies from the upper and lower quartiles.

3.2. Technologies related to power grid data management and control.
3.2.1. Data center software definition. The existing power communication network architecture is a three-layer architecture built on the extended tree protocol. The transmission of data packets is completed through various transmission protocols [14]. However, with the increasing scale of the IoT and massive data, the routing tables in routers are becoming increasingly complex, which brings many problems to the current network framework. When tuning network devices, network administrators must configure each switch or router one by one using the command line [15]. The SDN concept is designed to solve this type of problem. Figure 3 presents the SDN three-layer network architecture.

SDN is a network virtualization technology. It strips the control functions of traditional switches and separates the data plane and control plane. SDN entrusts the function of decision control to the control layer and uses the open interface provided by the control layer to complete the delivery of decisions [16]. The network layer data of data center traffic is processed through the SDN network, combined with pooled IP. The data center’s security is ensured by decoupling the internal organization of the data center from the outside world. Figure 4 shows the specific structure diagram of the communication data center.

The network layer mainly has two parts, and one is SDN-based data center control network feeding. The central node discovery and the message transmission are part of the Peer-to-Peer network composed of blockchain.
nodes responsible for node discovery, message propagation, and transmission protocol parsing in blockchain nodes [17]. The consensus layer and the data layer form the core of the blockchain data protection mechanism. The consensus layer improves the Delegated Proof of Stake (DPoS) algorithm. Based on the DPoS algorithm proxy system, the consensus layer combines the Practical Byzantine Fault Tolerance algorithm to improve the consensus efficiency and provide reliable node consensus.

3.2.2. Data copyright supervision mechanism. The center interacts with institutions in real time to fetch the returned data. These two methods are shown in Figure 3.5 for the data flow process.

The data itself has the characteristics of reproducibility and easy processing. In addition, the “see and own” nature of data makes it easier to own data than traditional goods. Agencies can move data at a lower cost [18].

4. Model design.

4.1. Simulation design of intelligent O&M service of power communication network. In the process of system implementation, for the consideration of applicability, the minimum available basic functions of anomaly detection should be provided for various types of IoT devices. As a classical algorithm with simple implementation, stable performance, and low requirements for data characteristics, box-plot has specific usability and wide applicability, which is suitable for the needs of this scenario. Therefore, the box-plot is selected as the basic algorithm for system anomaly detection. Table 4.1 records the relationship between the number of box-plot scenarios and the fluctuation of the detection accuracy of a single indicator.
Box-plot calculation formulas have inherent flaws. The data features cannot be effectively distinguished when the sample size is small. When the sample size is large, overfitting problems are prone to occur.

The data synchronization scheme in the data distribution scenario is designed, and the combination of synchronous calls and asynchronous callbacks is decided. Based on the message queuing transaction solution, the synchronous/asynchronous call method is improved to ensure that data can be successfully distributed to each node. Data is successfully synchronized to ensure data consistency. The specific implementation logic is demonstrated in Figure 4.1.

1. The first is the transmission of the O&M module. As the main functional module of the system, the O&M module has a higher priority than the data module. As data is the most critical dependency element of the O&M module, the real-time and accuracy of data transmission must be guaranteed.

2. The second is the transmission of data modules. The data module has a lower priority than the O&M module in system function priority. There are no requirements for data real-time. Therefore, the method of combining asynchronous call and callback confirmation is used to ensure the successful
transmission of data and data consistency.

3. The callback part receives an acknowledgment request after the data module receives the data. The acknowledgment request contains the device number and data timestamp. After the callback function receives the acknowledgment request, the corresponding data is deleted from the Redis cache. The data with the timestamp before the current callback data time cut-off is retrieved from the collection. It is the data that occurred before but still did not receive the callback acknowledgment.

The combination of synchronous transmission and asynchronous transmission ensures data consistency during data distribution between O&M modules and data modules, saving system overhead and shortening the time of data synchronization.

4.2. Data control center anonymization processing experimental simulation. This experiment studies the data anonymization guarantee scheme in the electromechanical engineering power system environment. The coupling relationship between network traffic is reduced by hashing and encrypting information such as account addresses and data payloads of data packets. The risk of improper in-depth analysis of data packets by the outside world is reduced to achieve the purpose of not leaking the internal information of the data control center.

The data anonymization solution is based on the SDN network. The pooled IP address is mapped in the egress switch of the data center to achieve the purpose of anonymization. Figure 4.2 shows the architecture of the anonymous service system.

SDN networks have the characteristics of centralized management and programmability, which can be used to deliver data at the data center egress gateway without the introduction of additional hardware devices. The topology and network information inside the data center can be completely shielded from the outside. The anonymization service component AnonyService is mounted on the SDN controller of the data center network. The main function of this component is to complete the mapping between Routable IP Address (RIPA) and Machine IP Address (MIPA) and the corresponding flow table delivery. MIPA is the actual address of the data center server, which is only perceived inside the data center and is shielded from the outside world. RIPA is an IP address that the outside world can access. After the controller calculates the mapping relationship between the two, it generates a flow table and delivers it.

When new traffic arrives, it determines whether the current anonymized flow table exceeds the given threshold. If it is exceeded, the timeout table is dropped. After that, determine whether the threshold has
Intelligent Algorithm Operation and Data Management of Electromechanical Engineering Power Communication Network

Fig. 4.2: Data center anonymous service system architecture

Table 4.2: Mapping algorithm pseudocode

Start
Input: Network traffic sequence S
Output: Anonymized flow table
Begin
If AnonyService.use >= max_capacity then
    AnonyService.delete_by_time();
    If AnonyService.use >= max_capacity then
        AnonyService.destroy();
        If flow_entry != null;
            AnonyService.destroy();
            AnonyService.init(max_capacity);
            index = (src_ip.hashcode + timestamp) % anonyIPs.current_size;
            new_src_ip = anonyIPs.get(index);
            flow_entry = AnonyService.generate_flow_entry(src_ip,
                new_src_ip, timestamp); insert_entry(flow_entry);
        End if
    End if
End if
End

been exceeded. If it is still exceeded, a new discarded current mapping is created. The flow table is emptied, and the process is repeated. Table 2 illustrates the mapping algorithm pseudocode.

The main operations of this mapping algorithm are as follows. The first is service initialization, which is responsible for creating the initial flow table and configuring the flow table threshold. The second is invalid stream table cleanup, which is responsible for clearing stream table entries that have not been matched within a period. Unlike the expiration time specified during the flow table configuration, this cleanup is an overall active cleanup to reduce the space footprint within the switch. The third is flow table emptying. If the switch space is still occupied after the invalid flow table cleanup, the flow table emptying operation will begin. The memory space occupied by the flow table is reclaimed, and a service initialization is triggered after recycling. The fourth is flow table lookup, which is mainly to find whether there is already a corresponding flow table.
entry based on the anonymous source IP address according to the need. If there is one, return it directly. The fifth is flow table insertion. This part is mainly for new traffic. It inserts the anonymized IP address into the switch flow table.

5. Results.

5.1. Intelligent operation platform algorithm implementation and optimization. The total number of trees in the forest affects the accuracy of the system’s detection. The detection results obtained by ensemble learning with enough isolated trees are more accurate. In addition, the total number of trees also affects the system’s performance. Too many trees cause additional overhead to the system and affect the system’s overall performance. Therefore, the number of trees in isolated forests needs to be studied to achieve optimal detection results and minimize the waste of system resources. The most stable quantitative boundary is obtained by studying the fluctuation of the detection accuracy of samples under different quantities. The results of the study after several tests are plotted in Figure 5.1.

From Figure 5.1, when the number of binary trees is less than 700, the variance of the accuracy rate is at a large value, indicating that the anomaly detection is relatively unstable. When the number of binary trees reaches 700, the accuracy variance is relatively small. The variance corresponding to the subsequent quantity also fluctuates within a small range, and the variance as a whole shows the nature of convergence and stability. When the number reaches 700, the detection effect of the system is stable. Adding more trees in the future does not significantly improve the detection effect, which will cause additional unnecessary overhead to the system. Therefore, 700 is selected as the number of binary trees in implementing the iForest algorithm.

Apache Jmeter is used to perform network stress tests on the system. The concurrent threads are created. Users are impersonated to access server ports. A stress test is conducted. The test environment is an i5 processor and 8G memory in a Windows environment. The final stress test results are revealed in Figure 5.2.

From Figure 5.2, when the number of concurrent requests exceeds 2,000, the proportion of error requests and system response time increase significantly. According to the test results, it is concluded that when the concurrent requests are less than 2,000, the system’s high availability can be guaranteed. Creating a cluster can improve system availability when requests are more than 2,000.

The test data set is used to detect anomaly detection performance and test the accuracy of different levels of anomaly detection function. The test results are shown in Figure 5.3.

5.2. Power communication network data analysis. Iperf simulated network traffic is tested to study the impact of anonymization schemes on response times for external requests. During the simulation, a delay of 1 ms between the client and the data center is set. Different traffic rates are sent through control flow
clients to test responses under various network stresses. For each flow rate, experiment one min and repeat the experiment ten times. The delay of the flow rate in each experiment is counted. Figure 5.4 shows the delay statistics.

From Figure 5.4, with or without anonymization, the response time is slightly over 1 ms. As data traffic grows, response times tend to rise. This shows that the network delay (1 ms) of the outgoing client access data center, and the internal response time of the data center is very low, about the order of 10μs. The experiment is relatively small, so the delay and jitter variation of data transmission are very small. However, it can be seen overall that the loss of response time caused by anonymization processing is very small. When the data flow rate is less than 50 Mbps, the internal latency of the data center without anonymization is about 0.039 ms, and the internal delay of the anonymized data center is about 0.041 ms. The response time increases by about 5%. This performance loss is very low and within the acceptable range. The test scale is relatively small,
so the internal latency of the data center is relatively low. Iperf statistics are not accurate enough, and the performance impact should be lower in practice.

6. Conclusion. This paper studies the power communication network supporting the IoT. Based on this research purpose, the research and design are carried out by combining the current relatively complete intelligent O&M and data management concepts. The research results are as follows: (1) When the binary trees reach 700, the detection effect of the system is stable. The subsequent addition of more trees does not significantly improve the detection effect, which will cause additional unnecessary overhead to the system. So, 700 is selected as the number of binary trees in implementing the iForest algorithm. (2) When the concurrent requests are less than 2,000, the system’s high availability can be guaranteed. When the number of requests exceeds 2,000, the system availability can be improved by adding servers and creating clusters. (3) When the algorithm data is increased to a specific value, the iForest algorithm improves the overall recognition rate of the system anomaly detection function. Therefore, it can be considered that the graded detection method adopted by the system is feasible and effective. (4) The internal latency of the anonymized data center is about 0.041 ms, and the response time increases by about 5%. This performance loss is low and within acceptable limits. However, due to the current status quo of software and hardware, there is still room for further optimization and improvement of the platform. First, data timing has a high dependence on the network state. Once network fluctuations occur, there will be a disorder in the data transmission process. Second, when providing O&M functions for different devices in the current system, it is necessary to redesign the algorithm and modify the source code. In the subsequent work, it is essential to improve the system and carry out additional design and implementation of functions such as data timing verification and retransmission start. Moreover, it is necessary to design and implement the algorithm training configuration module to realize the algorithm’s automatic training and configuration call algorithm functions to improve the intelligence and automation of the system.

REFERENCES


Edited by: B. Nagaraj M.E

Special issue on: Deep Learning-Based Advanced Research Trends in Scalable Computing

Received: Dec 27, 2023
Accepted: Mar 20, 2024