Abnormal network access detection of power Internet of Things terminal layer equipment based on equipment portrait

Dongkai Yang¹, Xiaoliang Zhang*, Kehe Wu¹ and Lipeng Lu²

¹ School of Control and Computer Engineering, North China Electric Power University, Beijing, Beijing, 102206, China
² State Grid Handan Power Supply Company, Handan, Hebei, 056000, China

* Corresponding author’s e-mail: wkh@ncepu.edu.cn

Abstract. In the environment of the power Internet of Things, equipment network security is mainly analyzed through the physical signal characteristics of the equipment or a single flow characteristic, so as to realize the equipment network abnormality detection. Therefore, a method for detecting network abnormalities of power Internet of Things terminal equipment based on device fingerprints is proposed. Aiming at the problem of incomplete detection methods, a multi-dimensional matching device fingerprint model is established. Firstly, the basic information of the device is collected, such as IP, MAC, etc.; then the network traffic information of the device is analyzed to extract the characteristics of network traffic; then, the frequency of keyword and keyword group of service protocol is counted. The device fingerprint model is established by using the traffic and service characteristics. When the network anomaly occurs, the device fingerprint model can be found effectively. The experimental results show that the device fingerprint model can detect the abnormal behavior of the device.

1. Introduction

With the development of sensor technology and network construction, the field perception capabilities of various professions have been deepened and broadened, and the system's operational status sampling, fault identification, and real-time control have been increasingly refined. Similarly, network security perception monitoring and threat prevention The breaking capacity also needs to be further extended to the field side.

Due to the characteristics of network-specific and protocol-specific features, the power Internet of Things urgently needs to further research on the security monitoring technical means suitable for power Internet of Things terminals based on the traditional analysis of network message attack characteristics, and improve the power Internet of Things to resist various The ability to attack threats[1]. It is an important task for all countries to create a power Internet of Things with comprehensive status perception, efficient information processing, and convenient and flexible applications[2]. At present, according to the comprehensive perception and software defined IOT architecture proposed in reference [3], the application of power grid terminal equipment supervision method based on IP can effectively reduce the complexity of operation and maintenance. Literature [4] established an information access specification for the perception layer terminal of the power Internet of Things, which enables data to interact with a unified information model in the network layer to improve interoperability. Literature [5] proposed an innovative idea and method to extract system-level behavior characteristics from the ICS system industrial control protocol interaction mode as an
ICS scene fingerprint. In [6], non-parametric Bayesian methods are used to detect the number of devices and to classify multiple devices in an unsupervised passive manner, which can be used for intrusion detection.

Based on the above research, this article will take all the traffic generated by the terminal device as the object of investigation, comprehensively consider the physical layer, network traffic and protocol behavior characteristics of the terminal device, and establish the device portrait of the terminal based on historical data to reflect the terminal's network behavior status; And further based on specific attack scenarios, this paper research the abnormal network access behavior analysis technology of terminal equipment, and realize the monitoring and identification of abnormal network access behavior of the terminal.

2. Device fingerprint technology
At present, device fingerprint technology is mainly divided into three categories: based on transient characteristics, based on modulated signals, and based on internal sensing. For the hidden features of high recognition accuracy and good stability, the corresponding technology still has some drawbacks. Based on the transient feature refers to the use of the radio signal on/off transient feature to realize the identification of the device. The related research reflects that the robustness of the device fingerprint is weak, and the device fingerprint is prone to significant changes; based on the modulation signal mainly It extracts features from modulated signals to generate device fingerprints to identify devices; based on internal sensors, it is mainly used for unique identification of smart terminals by application software. The related research reflects that the accuracy rate is easily affected by environmental noise. The elimination of environmental interference requires further research.

In this study, the device fingerprint is constructed through the basic attributes and access behavior attributes of the device, and the phishing and malicious terminal devices can be accurately identified in combination with specific attack scenarios, so as to realize the monitoring and identification of abnormal network access behavior of devices.

3. Terminal abnormal network access detection based on device fingerprint

3.1. Terminal network flow order construction
The fingerprint technology of power industrial control system equipment is based on the deep packet analysis technology, which analyzes the quintuple of the communication protocol and the content of the deep packet, and builds the fingerprint space of the industrial control system based on a multi-dimensional information system. The basic properties of the network extracted in this article include: IP address, traffic type, message length, message time slot, and message direction.

Among them, the IP address is the address of the remote connection end, and the traffic type is the specific protocol type of the network traffic, such as TCP, UDP, etc. The message length is the size value of a single message, the message slot is the time interval between adjacent messages, and the message direction is divided into inflow and outflow. Through further preprocessing of the above-mentioned basic attributes, the final network flow order can be obtained.

The specific process is as follows:

The text of your paper should be formatted as follows:

- Obtain time domain characteristics of traffic data packets. Think of a traffic instance as two time series: a sequence of incoming packets of different sizes with timestamps and a similar sequence of outgoing packets. For these two time series, calculate the number of bytes transmitted per time unit in a fixed time period. Therefore, each data stream is transformed into two distributions of the same size.
- Obtain frequency domain characteristics of traffic data packets. For the original message length, message time slot and other data, Fourier transform is used to transform the timing information into frequency domain information, and the first k values are selected as the network flow order characteristics.
3.2. Business protocol message analysis

According to the obtained message, it is judged whether the message content and protocol are abnormal from two levels of syntax and semantics. At the grammatical level, the protocol parallel analysis technology is used to quickly extract protocol key fields and their corresponding field values from the message, and then compare the protocol key fields or corresponding field values with the set whitelist. The ones on the list are business security exceptions. For the power Internet of Things business protocol, the message contains multiple key fields. The main principle of parallel analysis is to analyze the key fields 1, field 2, and field 3 in the message at the same time. Parallel processing can significantly increase the speed of message parsing, thereby speeding up the detection of anomalies in the basic whitelist matching business protocol. The first is the parallel analysis when the position of the protocol keyword is fixed. In the parallel parsing process of the power Internet of Things business message, if the position of the field is fixed, the pointer can be shifted to quickly locate the different field, which is easier to implement. Second, parallel analysis when the position of the protocol keyword is variable. If the position where the field appears is not fixed and the length of the field content is variable, in order to realize the parallel analysis of the protocol message, it is necessary to solve the problem of rapid positioning of the key fields of the protocol. To this end, the message can be divided into p equal parts, and each equal part can be scanned for keywords in parallel (partial matching is sufficient), so as to quickly find the specific position of the key field in the original message, and then realize the parallel analysis of protocol messages. After parsing the message, you can compare the value of the field or the content of the field with the preset whitelist. If the parsed value is not in the whitelist, it is judged that the protocol packet is abnormal.

Attackers can use normal protocols to perform malicious attacks, and cannot find anomalies at the protocol syntax level. Therefore, a deeper anomaly detection at the protocol semantic level is required. At the semantic level, use the frequency and sequence of the agreement keywords to establish the behavior model of the agreement. Then based on the model, abnormal business operation is detected. However, the operating logic of the protocol is complex, and different protocols have different design ideas, so it is necessary to propose a practical protocol behavior modeling method. This article takes the key fields of the agreement as the object of investigation, and establishes the agreement behavior model based on the frequency and sequence of the key fields of the agreement. First, in normal traffic, for each protocol keyword, count its frequency of occurrence. Define the two keywords that appear before and after as a keyword combination. In normal traffic, for each keyword combination, count its frequency of occurrence. Statistics afterwards. For each keyword combination, count the time difference between the two keywords before and after. Finally, a protocol behavior model is constructed to detect abnormal protocol messages. Based on the above statistical results, the behavior model of the protocol can be established, and abnormal protocol message detection can be performed. The detection process is as follows:

- Establish model nodes, and node attributes include key field names and frequency of occurrence. Edge attributes include time interval and frequency of keyword combinations.
- Traverse each key node, and if the frequency of a key field exceeds the set threshold, it is judged abnormal;
- If the frequency of the keyword combination exceeds the set threshold, it will be judged as abnormal;
- If the time interval between the before and after keywords deviates from the normal distribution, it is judged abnormal.
4. Experimental results and analysis
The experimental environment is the intelligent substation industrial control network system. The intelligent substation industrial control system can be divided into three layers: station control layer, bay layer, and process layer. The communication between bay layer equipment and station control layer equipment is through MMS message format; the data characteristics between process layer equipment and bay layer equipment mainly include: the sampling value SV message sent by the process layer electronic transformer merging unit; smart switch The GOOSE message sent by the protection unit; the control command message sent by the bay layer device. The configuration of this type of communication message data is mostly sent periodically, so it is relatively stable, but there will be an obvious cycle when an abnormality or network intrusion occurs Sexual change.

The use of sections to divide the text of the paper is optional and left as a decision for the author. Where the author wishes to divide the paper into sections the formatting shown in table 2 should be used.

4.1. Construction of Flow Fingerprint of Substation Equipment

4.1.1. Network flow order characteristics
According to the characteristics of the IEC61850 network communication protocol for smart substations, this paper selects the number of outbound/inbound packets of SV/GOOSE messages, the average packet interval, the amount of outbound/inbound bytes, and the flow direction as the network flow order characteristics, as shown in Table 1.

| attributes  | network order flow characteristics | description                                      |
|------------|-----------------------------------|--------------------------------------------------|
| \( x_1 \)  | Flow direction                    | Equipment flows into or out of the terminal       |
| \( x_2 \)  | Incoming bytes                    | The amount of message data received by the terminal |
| \( x_3 \)  | Bytes out                         | The amount of message data sent by the terminal   |
| \( x_4 \)  | Number of packages                | Number of packets received by the terminal        |
| \( x_5 \)  | Number of packages out            | Number of packets sent by the terminal            |
| \( x_6 \)  | Average packet interval           | Average message interval                          |

4.1.2. Features of business agreement
This paper selects keyword frequency, keyword combination frequency, and time interval between keywords as the characteristics of the business agreement, as shown in Table 2.

| attributes  | network order flow characteristics | description                                                  |
|------------|-----------------------------------|--------------------------------------------------------------|
| \( x_7 \)  | Keyword frequency                 | Frequency of key fields of agreement                          |
| \( x_8 \)  | Keyword combination frequency     | Frequency of the two keywords before and after               |
| \( x_9 \)  | Keyword interval                  | Time difference between keyword combinations                 |

The content of the SV message includes protection current IP, neutral current Im, phase voltage VP, measurement current Ig, and measurement voltage V. GOOSE message content includes key value X and switch value Y. As the keyword features of the business agreement above, Ig and V are selected as keyword phrases. Through the wireshark tool to analyze the data within 10s, the keyword frequency
\[ P = \{ I_{P}, I_{m}, V_{P}, I_{g}, V, X, Y \} \]  

(1)

can be obtained as \{976,1048,997,1060,1060,933,984\}.

The keyword group frequency \( I_{g}, V \) is consistent with the keyword frequency. The time interval between keywords is \( t=100\text{ms} \).

### 4.1.3. Device fingerprint model

According to the protocol type feature extraction method in the previous article, set the relevant equipment: SV sending sampling points 80, sampling frequency 50, single frame ASDU1, serial number synchronization second pulse synchronization, quality factor 01100000010000, manual increase interval 100ms; GOOSE sending interval \( T=2 \), the quality factor is 0100000000000. Through analysis, the 9 characteristic attributes can be divided into three categories: flow characteristics, data characteristics, and keyword characteristics. The typical protocol characteristic vectors of terminal interaction in the smart substation network can be obtained as follows:

\[
X_{\text{GOOSE}} = \begin{bmatrix} x_1, x_2, x_3 \\ x_4, x_5, x_6 \\ x_7, x_8, x_9 \end{bmatrix} = \begin{bmatrix} 1 & 114 & 208 \\ 3702 & 4381 & 23 \\ 0.13 & 0.1 & 1.183 \end{bmatrix}
\]

(2)

\[
X_{\text{SV}} = \begin{bmatrix} 1 & 174 & 281 \\ 4800 & 5901 & 17 \\ 0.014 & 0.001 & 1.229 \end{bmatrix}
\]

(3)

Applying neural network algorithm, using equations (2) and (3) as the algorithm input to train neurons, the protocol type characteristic output can be obtained. Under normal operating conditions, with the real-time flow data of the intelligent substation process communication as input, the output can be obtained: \( Y = (0,1,1,0) \)

That is, in a safe state, the process layer protocol of the smart substation is SV and GOOSE.

By analyzing the process layer network traffic communication link, the wireshark tool can be used to obtain the set of all information sources and the set of message lengths at the process layer, so that the data volume characteristics can be obtained as \( f = 528364 \).

From this, the device fingerprint network monitoring model matrix can be established:

\[
\Delta = \begin{bmatrix} Y \\ P \\ f \\ t \end{bmatrix}
\]

(4)

According to \( Y \) to detect abnormal protocol type, according to \( P \), to detect abnormal protocol content, according to detect abnormal network traffic. The detection step is to collect real-time flow data, set the rated fault tolerance value \( \varepsilon \), count the flow data within 100s and calculate the characteristic value. Make the difference with the eigenvector in the normal state, if the result is less than or equal to the tolerance value, it is in a safe state.
4.1.4. Simulation and results

In order to verify the effectiveness of the device fingerprint model in this article for network security monitoring, UDMView is used to simulate the SV and GOOSE messages sent by the merging unit and the intelligent protection unit in the process layer network. The analysis of the simulation results in this paper is based on the wireshark tool and obtained through C++ compilation and analysis scripts.

The data amount of the specified fault tolerance value is 100. The protocol data packet with abnormal behavior is simulated through UDMView. Get vector:

\[ \Delta = \begin{bmatrix} (0,1,1,0) \\ (1077,1150,1099,1065,1038,1084) \\ 528471 \\ 100 \end{bmatrix} \]

Calculating that the third line exceeds the fault tolerance value, it can be obtained that the device network is abnormal.

5. Conclusion

This article introduces deep learning of the Internet of Things into the edge computing environment to optimize network performance. The edge computing architecture allows edge nodes to reduce intermediate data for uploading data, reducing network traffic from IoT devices to cloud servers. For the limited service capabilities of edge nodes, an offload scheduling algorithm is proposed to maximize the number of tasks in the edge computing environment. The experimental results show that the IoT deep learning application and the offload scheduling algorithm for edge computing can increase the number of tasks deployed in the edge server while ensuring the quality of service requirements.

References

[1] Liao, HX., Zhen, P., Li, L. (2019) Overview of Ubiquitous Power Internet of Things Information Security. Electric Power Information and Communication Technology., 17: 18-23.
[2] Fu, ZX., Li, XY., Yuan, Y. (2019) Discussion on Key Technologies of Ubiquitous Power Internet of Things . Electric Power Construction., 40:1-12.
[3] Lu, J.L., Liu, R.L., Wang, P., Lin, J.Y. (2018) Distribution Internet of Things Architecture Based on Comprehensive Perception and Software Definition . Power System Technology., 42: 3108-3115.
[4] Jing, M.C., Wang, J.Y., Cheng, Z.H. (2014) Research and application of sensor information model for power Internet of Things . Power System Technology., 38: 532-537.
[5] Peng, Y., Xiang, L., Zhang, M. (2016) Scene fingerprint and anomaly detection in industrial control system. Journal of Tsinghua University (Natural Science Edition)., 56: 14-21.
[6] Efstathopoulos, G., Grammatikis, P.R., Sarigiannidis, P. 2019. Operational Data Based Intrusion Detection System for Smart Grid. In: Computer Aided Modeling and Design of Communication Links and Networks. Limassol. pp.11 -13.