A Method for Anomaly Detection of Encrypted Traffic in Power IoT Base on Security Baseline Learning

Zhou Peng
State Grid Zhejiang Co., Ltd.
Information and Communication Branch
Hangzhou, China
zhou_peng@zj.sgcc.com.cn

Xu Qiang
China Electric Power Research Institute
Information and Communication Branch
Beijing, China
xuqiang@epri.sgcc.com.cn

Hu Lizhi
China Electric Power Research Institute
Information and Communication Branch
Beijing, China
hulizhi@epri.sgcc.com.cn

Li Menglin
China Electric Power Research Institute
Information and Communication Branch
Beijing, China
limenglin@epri.sgcc.com.cn

***Abstract*—As numerous power IoT terminals require secure access to the power grid, how to detect the encrypted malicious traffic in power IoT becomes a challenge. Following an in-depth investigation of potential security risks in power, IoT encrypted traffic, an unsupervised learning method is designed to establish a security baseline for normal business encrypted traffic and to identify the anomaly encrypted traffic. Firstly, a partially seeded K-means clustering algorithm is proposed for security baseline learning to construct a set of clusters for normal encrypted traffic. Secondly, an algorithm tailored for comparing similarity within a cluster is introduced to minimize the costs associated with anomaly detection. At last, experiments are performed to evaluate the effectiveness and performance of this method on real-power IoT encrypted traffic and open VPN encrypted traffic. The results demonstrate that this method has the ability not only to detect the anomaly in business encrypted traffic caused by terminal device error or manual wrong operation but also to identify unknown cyberattacks targeting internal applications to penetrate the secure access gateway by encrypted traffic**

Keywords— Power IoT, Encrypted traffic, Anomaly detection, Clustering

I. INTRODUCTION

With the rapid development of the Internet of Things(IoT), more industrial control systems(ICS) are connected to the Internet. However, it also allows cyber attacks to disrupt the real physical world directly [1-5]. ICS are highly interconnected and interdependent with national critical infrastructure[6]. Cyber attackers have increasingly realized that ICS attacks have low investment and high returns. Due to the physical isolation between traditional ICS and the Internet, existing intrusion detection techniques for ICS focus on system functional security without consideration for cyberattacks. At present, the common intrusion detection techniques for ICS can be divided into several categories: state-based [7], behavior-based [8], rule-based[9], feature-based[10], model-based[11], and machine learning-based [12][13].

In 2020, China proposed the "dual carbon" strategy, accelerating the construction of new power systems. Many new business power IoT terminals demand secure access to the power grid. For this purpose, secure access gateways have been widely deployed at the boundaries of power IoT, enabling the functions of encryption, authentication, and secure access. The encrypted communication between power IoT terminals and the secure access gateways provides a natural defence evasion way for attackers[14]. Still, traditional ICS intrusion detection techniques mainly deal with plain-text traffic. Therefore, the threats in encrypted traffic of power IoT become a blind spot.

How to effectively detect encrypted malicious traffic has become a challenge in the research field of cyber security[15]. At present, research on encrypted traffic detection can be divided into three categories [15], including distinguishing the encryption algorithms for encrypted traffic[16], identifying the encrypted malicious traffic based on machine learning[17][18], and identifying the encrypted malicious traffic based on cryptography[19][20].

Focusing on the problem of identifying the encrypted malicious traffic in the power IoT, this paper takes the machine learning approach. An unsupervised learning method of depicting the security baseline of encrypted business traffic, which bypasses the power IoT secure access gateway, is proposed to detect abnormal encrypted traffic. The main contributions of this paper are summarized as follows:

(1) A partially seeded K-means algorithm is proposed to address the issue of learning the security baseline of encrypted normal traffic. By utilizing the known application labels of some sessions in normal encrypted traffic, the security baseline is characterized by a set of clusters. The algorithm's effectiveness is experimentally evaluated on real power IoT and open VPN encrypted traffic.

(2) To address the problem of using clustering results as the security baseline to detect abnormal encrypted traffic, a detection algorithm based on a similarity comparison of nearest neighbours within a cluster is proposed. This reduces the computational complexity of the detection process. Two experiments also evaluate the permanence of this algorithm.

II. BACKGROUND

To support the needs of various terminals to access the power IoT management platform, a secure architecture with

the secure access gateway as the key device is used at the boundary of power IoT. The secure access gateway for power IoT carries many security functions, including communication encryption, authentication, secure access, and access control[21].The secure access gateway for the power IoT is a VPN device using TLCP protocol and cipher suites of the SM series. In this secure architecture, the IoT edge agent and the power IoT management platform are interconnected in a bidirectional manner. The research object of this work is the encrypted traffic of power IoT, which refers to the encrypted communication traffic between the secure access gateway and the IoT edge agent.

Transport Layer Cryptography Protocol(TLCP) is a kind of SSL VPN protocol, which is a Chinese national standard (GB/T38636-2020)[22] implemented on November 1, 2020. The design of the TLCP protocol refers to the international standard TLS protocol. The handshake and key exchange progress in TLCP is consistent with TLS protocol. However, there are still three differences. Firstly, the protocol version numbers are defined differently, and detailed differences exist in the packets of handshake and key exchange. Secondly, unlike the international cipher suits used in TLS protocol, TLCP mainly uses ciphers such as SM2/SM3/SM4 with GCM's suite, including ECC_SM4_GCM_SM3 and ECDHE_SM4_GCM_SM3. Thirdly, the TLCP protocol uses SM2 dual certificates, which separate the encryption certificate/private key and the signature certificate/private key. The client and server use two different sets of certificates for encryption and authentication. In recent years, domestic web servers, web browsers, and domestic VPN gateways have rapidly supported the TLCP protocol in China. In addition, OpenSSL libraries, released after the OpenEuler 22.09 version, also support the TLCP protocol..

III. Security Baseline Learning

Baseline refers to the normal behavior benchmark of ICS users, equipment, applications or traffic. Therefore, anomaly detection for encrypted traffic in power IoT based on security baseline learning establishes a benchmark for normal encrypted traffic in the past and compares the baseline with real-time generated encrypted traffic. An anomaly is detected when the measured traffic deviates from the baseline and exceeds the preset threshold. This method has two advantages. At first, it only requires collecting and learning the encrypted traffic of normal business to construct a secure baseline. The baseline is easy to re-construct when the business changes. Secondly, due to the whitelist-like strategy, this method can detect anomalies in business encrypted traffic caused by terminal device error or manual wrong operation and identify unknown cyberattacks targeting internal applications to penetrate the secure access gateway by encrypted traffic.

This paper proposes a security baseline learning and anomaly detection method for encrypted traffic in power IoT. Clustering-based methods are used to characterize the baseline of normal business encrypted traffic between the IoT edge agent and the secure access gateway. It is well known that the quality of training samples constrains machine learning-based technology. As for malicious or anomaly-encrypted network traffic, especially attack traffic targeting the IoT secure access gateway, it is even more difficult to obtain and accurately label. Therefore, the biggest advantage of applying unsupervised learning methods to detect anomaly traffic is the lower cost and convenience of obtaining and labelling training samples, considering that normal business encrypted traffic is much easier to capture.

A. Feature Extraction

It is necessary to design a feature extraction method for encrypted traffic to depict the security baseline of encrypted traffic by clustering in power IoT. Because the TLCP protocol is compatible with the TLS protocol, this paper improves the feature extraction method in MalDetect[23] to build a feature set for security baseline learning.

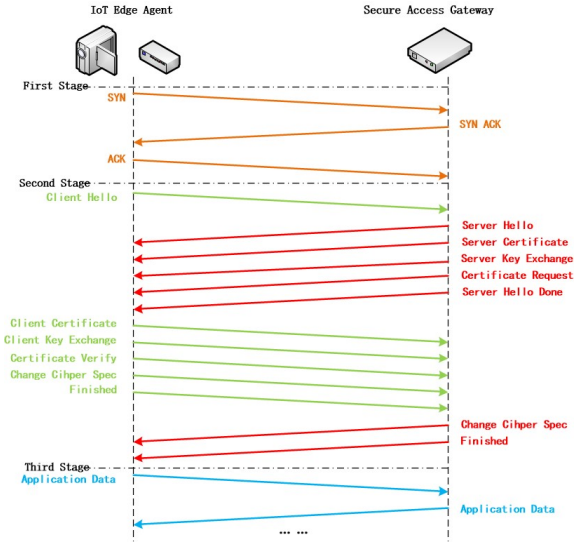

Fig. 1. Communication process between the IoT edge agent and the secure access gateway

In the TLCP handshake and key exchange process shown in Fig. 1, the IoT edge agent is a client to initiate a TCP session, while the secure access gateway is a server to responds to a TCP session. The encrypted session built can be divided into three stages. In the first stage, both parties establish a session through a three-way handshake of TCP. In the second stage, both parties follow the TLCP protocol to complete the authentication and the negotiation process. The client first sends a “Client Hello” packet, providing alternative server options, such as encryption suites and extensions. The server responds to a “Server Hello” to notify the selected option. Following the “Server Hello” packet are the “Server Certificate” and the “Server Key Exchange” packets, which contain server certificates and data for signature verification. Unlike the TLS protocol, TLCP contains two sets of SM2 certificates used for authentication and encryption by the server. Afterwards, the server sends the “Certificate Request” and the “Server Hello Done” packets. In power IoT, bidirectional authentication is a required option between the client and server. Therefore, the client will send the “Client Certificate”, the “Client Key Exchange”, and the “Certificate Verify” packets immediately, which also contain two sets of SM2 certificates used for authentication and encryption by the client, respectively. At last, the “Change Cipher Spec” and the “Finished” packets are delivered between the parties. In the third stage, both parties use the chosen cipher suite and the key to

encrypt the application data in the "Application Data" packets.

In MalDetect[23], the encrypted traffic features in the TLS protocol are all extracted from the packets in the first and the second stages. To be compatible with the TLCP protocol and dual certificate mechanism, the feature set is improved, and features of the "Application Data" packets in the third stage are introduced in this paper to characterize the communication behavior of both parties at the application layer. A feature set for security baseline learning of encrypted traffic in power IoT includes four types.

(1) **Package features**. Features of this type are consistent with MalDetect[23], which are not related to TLS protocol. They describe sessions in three dimensions: size, number, and time. Seven features are listed: inbound bytes, outbound bytes, inbound packets, outbound packets, duration, SPT (sequence of packet time), and SPL (sequence of packet length).

(2) **Encryption protocol features**. Based on MalDetect[23], features of this type are improved to be compatible with TLCP protocol and SM cipher suite by increasing the vector dimension, including protocol version, offered cipher suite, selected cipher suite, selected compression method, offered extensions, selected extensions, and TLS packet ratio.

(3) **Certificate features**. To support the dual certificate mechanism, certificate features are defined as server-side and client-side. Certificate count, certificate chain length, the issuer feature vector of the endpoint certificate, and the subject feature vector of the endpoint certificate are imported on the basis of MalDetect[23]. Other features include certificate number, bad certificate number, certificate version ratio, certificate extension ratio, certificate validity mean, certificate public key length mean, certificate public key algorithm ratio, and certificate signature algorithm ratio.

(4) **Application data features**. Although the application data has been encrypted, the communication behavior of the two parties can still be described from three dimensions: size, frequency, and time. Application data features proposed in this paper include upstream bytes, downstream bytes, upstream packets, downstream packets, upstream maximum length, downstream maximum length, upstream minimum length, downstream minimum length, upstream length mean, downstream length means, the most frequent length of upstream, the most frequent length of downstream, upstream interval means, and downstream interval mean.

Quantification is performed by referring to the method in MalDetect[23] to enable the features above to be calculated in a clustering algorithm. Their direct values are used for counters, while the appearance of field values is used for variables.

B. Clustering Algorithm

K-means[24] algorithm is one of the most classical clustering algorithms to cluster all samples around *K* center points in space and update the center points until the best results are obtained. However, how to choose the initial centers significantly affect the clustering results. Considering the performance can be improved by choosing the centers from labeled samples, Wagstaff, etal.[25] proposed the COP *K*-means and Basu, etal.[26] proposed seeded/constrained *K*-means algorithm. However, all these methods are either not used for the initial center selection or require that each cluster has a preexisting seed. In most of the occasions, only a portion of samples have known labels, so it cannot be guaranteed that each cluster has a seed. A partial seeded *K*-means algorithm (Table 1) was proposed in our early research on tracing the attacks of ICS[27]. Because clustering algorithms can use the known labels of some applications by which the encrypted traffic is generated, a partially seeded *K*-means algorithm is also applicable to depict the security baseline of normal encrypted traffic in power IoT.

TABLE I. PARTIAL SEEDED K-MEANS ALGORITHM

Input: given a sample set $D=\{x_1,x_2,...,x_m\}$, the clustering number *k*, the known clustering number *l*, $l \le k$, the sample subset with known labels $D'=\{x_1,x_2,...,x_n\}$, and the sample subset without known labels $D-D'$.

(1) Calculate the initial center $\mu_i(1 \le i \le l)$ of the known sample cluster $C_i(1 \le i \le l)$, $\mu_i = (1/|C_i|)\sum_{x \in C_i} x$.

(2) Calculate the minimum distance from each sample $x_j(1 \le j \le m-n)$ in $D-D'$ to the known centre $\mu_i(1 \le i \le l)$, and choose the sample which has the largest distance from $\mu_i(1 \le i \le l)$ as a new initial center.

(3) Repeat (2), until $k-l$ samples are selected as new initial centers, make $\mu_i(1 \le i \le l)$ and $\{\mu_{l+1},\mu_{l+2},...,\mu_k\}$ to be *k* initial centers $\mu_i(1 \le i \le k)$.

(4) Calculate the distance $d_{ji} = \| x_j - \mu_i \|_2$ which is from each sample $x_j(1 \le j \le m-n)$ in $D-D'$ to each center $\mu_i(1 \le i \le k)$.

(5) Choose the cluster label $\lambda_j = \arg\min_{i \in \{1,2,...,k-l\}} d_{ji}$ for the sample x_j according to the nearest center, and make x_j join into the cluster $C_{\lambda_j} = C_{\lambda_j} \bigcup \{x_j\}$.

(6) Calculate new center $\mu_i' = (1/|C_i'|)\sum_{x \in C_i'} x$, if $\mu_i' \ne \mu_i$ update μ_i to be μ_i'.

(7) Repeat (4) to (6), until no center is updated.

Output: clusters $C=\{C_1,C_2,...,C_k\}$.

The Partial seeded K-means algorithm utilizes some sample subsets with known labels as seeds to choose the initial centres. Considering there may be a variety of applications in power IoT, constraints on seed are not applied while adding a sample into the clusters. That

means a sample with a known label may be classified into the original cluster or a new cluster during the process. By adjusting the k in a training process and observing the clustering results, a set of k clusters with the best

performance can be obtained and used as the security baseline of normal encrypted traffic in power IoT.

IV. Anomaly Detection

If $C = \{C_1, C_2, ..., C_k\}$ is the output of the partially seeded *K*-means algorithm, and v_{new} is the feature vector of unknown encrypted traffic to be detected; re-clustering is not used as the anomaly detection method in this paper. An algorithm based on similarity comparison among the nearest neighbors in a cluster is proposed (Table 2). Compactness refers to the mean distance between a sample and its nearest δ neighbors in a cluster. Separation refers to the mean distance between a sample and other centers of the k-1cluster.

$$Compactness(v_i) = \sum_{j=1}^{\delta} d(v_i, v_j) / \delta \quad (1)$$

$$Seperation(v_i) = \sum_{j=1}^{k\text{-}1} d(v_i, \mu_j) / (k-1) \quad (2)$$

TABLE II. ANOMALY DETECTION ALGORITHM

Input: given the clusters $C = \{C_1, C_2, ..., C_k\}$, k centers of k clusters $\{\mu_1, \mu_2, ..., \mu_k\}$, and a feature vector v_{new} referring to the unknown encrypted traffic.

(1) Calculate the distance of v_{new} to all centers, choose the nearest cluster $C_{\lambda_{new}}$ as the candidate to join and the label is $\lambda_{new} = \arg\min_{i \in \{1,2,...,k\}} d(v_{new}, \mu_i)$.

(2) Find the δ nearest neighbors $V_{close} = \{v_i, (i = 1,2,...,\delta)\}$ of v_{new} in $C_{\lambda_{new}}$.

(3) Make v_{new} to join in $C_{\lambda_{new}}$, repeat (2) until find the nearest neighbors of each v_i($v_i \in V_{close}$), calculate Compactness of v_i and v_{new}; If $Compactness(v_{new}) \le \text{argmax}_{v_i \in V_{close}} Compactness(v_i)$ g o to(4); else $\lambda_{new} = None$ and go to output.

(4) Calculate Separation of v_i ($v_i \in V_{close}$) and v_{new}, If $Seperation(v_{new}) \ge \text{argmin}_{v_i \in V_{close}} Seperation(v_i)$ go to output; else $\lambda_{new} = None$.

Output: if $\lambda_{new} = None$, v_{new} refers to a abnormal encrypted traffic; else v_{new} refers to a normal encrypted traffic.

V. Experiment and Evaluation

The datasets used in the experiment are shown in Table 3. ERPI-1 and EPRI-2 are two batches of encrypted traffic captured from a real business environment of power IoT. Facebook, Email, and FTP are datasets of encrypted VPN traffic from ISCXVPN2016. Facebook datasets include encrypted chat and audio traffic. Email datasets are encrypted Gmail traffic with Google certificates. Encrypted FTP traffic from ISCXVPN2016 includes both control and data channels. However, traffic from the control channel involves the handshake process of Auth-TLS in FTP protocol, which significantly differs from TLS protocol. Therefore, only the encrypted traffic from the FTP data channel is used in the experiment. All the encrypted traffic is separated by session, and the features are extracted from sessions that contain handshake and key exchange progress. Finally, the features of all sessions are quantified as vectors.

TABLE III. DATASETS IN EXPERIMENT

Datasets	Details	Session Count
ERPI-1	TLCP protocol, SM suite	90
EPRI-2	TLCP protocol, SM suite	42
Facebook	TLS, Chat and Audio	34
Email	TLS, Gmail	70
FTP	TLS, FTP-Data	49

A. Effectiveness

Input the feature vectors in ERPI-1 into partial seeded K-means algorithm to get a set of clusters as the security baseline of normal encrypted traffic in power IoT. Because two different server certificates are found in encrypted traffic of ERPI-1, set l=2 in the clustering algorithm indicates two known clusters. Fig. 2 shows the number of clusters generated by the algorithm under different values of k. When k>17, the maximum number of clusters is six and remains stable, which means the algorithm has learned six patterns from ERPI-1 datasets. After manually analyzing the data from six clusters, it was found that the algorithm first divides session vectors with different server certificates into two clusters. Each cluster is further divided into three clusters. These three clusters show a clear pattern in terms of application data length. In the first pattern, the length of application data sent by the client is very short and only has a few fixed lengths, while the length of application data sent by the server is all 64 bytes. In the second pattern, the client sends a large amount of application data with a length of 1504 bytes, while the application data sent by the server is very short and only has a few fixed lengths; In the third pattern, the client sends a large amount of application data with a length of 1504 bytes, while the server sends a large amount of application data with lengths of 1424 and 1488 bytes. This experiment shows that the two parties in power IoT have relatively fixed communication behaviour patterns. Even if the communication traffic is encrypted, the baseline of these patterns can still be characterized by a certain set of features.

Use partial seeded *K*-means to cluster the feature vectors in other TLS VPN datasets, including Facebook, Email, and FTP. Set l=3 in the algorithm, indicating three known protocols. Fig. 3 illustrates the number of clusters the clustering algorithm generates for various k values. When k exceeds 19, the cluster count peaks at 7 and remains stable, suggesting that the algorithm has identified seven patterns from these datasets. As shown in Fig. 4, the data distribution of the seven clusters shows that Facebook data is split into two clusters, Email data into three, and FTP data into two. There is no data from one protocol assigned to another protocol's cluster. Analyzing the data from seven clusters reveals notable differences among the data of three protocols in encryption protocol features, certificate features, and application data features. This experiment shows that the partially seeded K-means algorithm also can learn a baseline for non-IoT encrypted traffic using the TLS protocol.

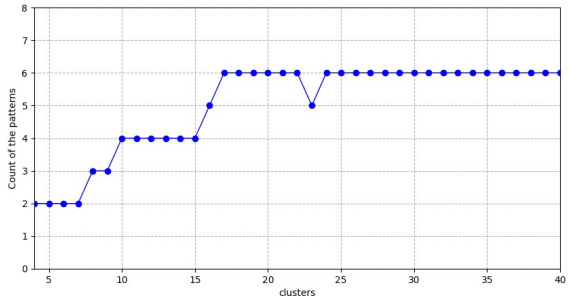

Fig.2 Clustering Results Of EPRI-1

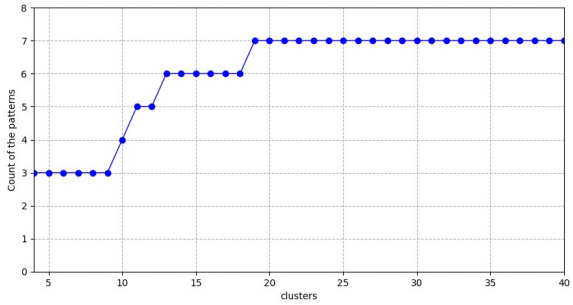

Fig.3 Clustering results of VPN datasets

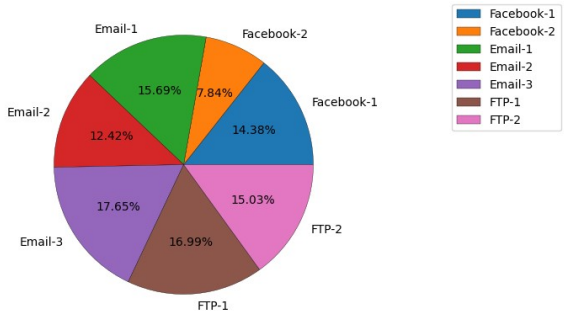

Fig. 4 Clustering result distribution of VPN datasets

B. *Performance*

Take the set of clusters of ERPI-1 as the baseline. Input the data in ERPI-2, Facebook, Email, and FTP as unknown encrypted traffic into the anomaly detection algorithm individually. Record the detection outputs. The statistical results under different values of δ are shown in Table 4. When $\delta = 3$ and $\delta = 4$, the Accuracy is 100%, while the FP (False Positives) rate and FN (False Negatives) rate are both 0. When $\delta = 5$, there is one false alarm in EPRI-2, with an accuracy of 99.49%, an FP rate of 2.38%, and an FN rate of 0.

TABLE IV. PERFORMANCE ON THE ORIGINAL DATASETS

Parameter	Datasets	Detected as abnormal	Detected as normal	Statistics
$\delta = 3$	EPRI-2	0	42	Accuracy=100% FP rate=0 FN rate=0
	Facebook	34	0	
	Email	70	0	
	FTP	49	0	
$\delta = 4$	EPRI-2	0	42	Accuracy=100% FP rate=0 FN rate=0
	Facebook	34	0	
	Email	70	0	
	FTP	49	0	
$\delta = 5$	EPRI-2	1	41	Accuracy=99.49% FP rate=2.38% FN rate=0
	Facebook	34	0	
	Email	70	0	
	FTP	49	0	

Considering the significant differences in encryption protocol features and certificate features between the encrypted VPN traffic and the power IoT encrypted traffic, the VPN datasets are modified to a series of new datasets. This new dataset simulates attack and abnormal traffic with higher similarity to IoT business traffic, in which the values of encryption protocol features and certificate features in Facebook, Email, and FTP vectors are randomly replaced with feature values from EPRI-1. However, the values of other features remain. Input the vectors in ERPI-2, Facebook_new, Email_new, and FTP_new into the anomaly detection algorithm. The statistical results under different δ values are shown in Table 5. When $\delta = 3$, there are three false negatives in Facebook and Email data, with an accuracy rate of 98.46%, an FP rate of 0, and an FN rate of 1.96%. When $\delta = 4$, there is still one false negative in Email data, with an accuracy of 99.49%, an FP rate of 0, and an FN rate of 0.07%. When $\delta = 5$, the false negative disappeared, but one false positive appears in EPRI-2, with an accuracy of 99.49%, the FP rate of 2.3%, and the FN rate of 0.

TABLE V. PERFORMANCE ON THE MODIFIED DATASETS

Parameter	Datasets	Detected as abnormal	Detected as normal	Statistics
$\delta = 3$	EPRI-2	0	42	Accuracy=98.46% FP rate=0 FN rate=1.96%
	Facebook	33	1	
	Email	68	2	
	FTP	49	0	
$\delta = 4$	EPRI-2	0	42	Accuracy=99.49% FP rate=0 FN rate=0.07%
	Facebook	34	0	
	Email	69	1	
	FTP	49	0	
$\delta = 5$	EPRI-2	1	41	Accuracy=99.49% FP rate=2.38% FN rate=0
	Facebook	34	0	
	Email	70	0	
	FTP	49	0	

These two experiments demonstrate that by employing the baseline learned from ERPI-1, the anomaly detection algorithm can effectively identify abnormal encrypted traffic with an Accuracy exceeding 98% overall. Moreover, by adjusting δ, the FP rate and the FN rate can be re-balanced. Therefore, δ can be optimized according to the needs of security operation requirements in practice.

VI. SUMMARY

Compared with the technology of identifying the encrypted malicious traffic based on supervised machine learning, the proposed method focuses on learning the encrypted traffic of normal business. It avoids obtaining and labelling the encrypted malicious traffic, with low difficulty and cost for incremental learning. The experiment results show that this method has the ability not only to detect anomaly in business encrypted traffic caused by terminal device error or manual wrong operation but also to identify unknown cyberattacks targeting internal applications to penetrate the secure access gateway by encrypted traffic.

In future work, more business types of encrypted traffic in power IoT will be used to test and improve the effectiveness of the partially seeded *K*-means algorithm. At the same time, new experiment datasets will be constructed by collecting abnormal encrypted traffic and attack traffic from real power IoT environments.

ACKNOWLEDGEMENT

This research is supported by the Project of State Grid Corporation of China (5700-202419250A- 1-1-ZN).

REFERENCES

[1] T. M. Chen, S. Abu-Nimeh, “Lessons from Stuxnet”. Computer, 2011, 44(4):91-93.

[2] D. Kushner, “The real story of Stuxnet”. IEEE Spectrum, 2013, 50(3):48-53.

[3] K. Zetter, “A cyberattack has caused confirmed physical damage for the second time ever”, 2015, https://www.wired.com/2015/01/german- steel-mill-hack-destruction/.

[4] K. Zetter, “Inside the cunning, unprecedented hack of Ukraine's power grid”. 2016, https://www.wired.com/2016/03/inside- cunning-unprecedented-hack-ukraines-power-grid/.

[5] ESET. ESET discovers dangerous malware designed to disrupt industrial control systems. 2017, https://www.eset.com/us/about/newsroom/press-releases/eset-discovers-dangerous-malware-designed-to-disrupt-industrial-control-systems/.

[6] K. Stouffer, V. Pillitteri, S. Lightman S, “Guide to industrial control systems (ICS) security”, NIST special publication 800-82 revision 2. NIST, 2023:2-2.

[7] A. Khalili, A. Sami, “SysDetect: a systematic approach to critical state determination for industrial intrusion detection systems using apriori algorithm,”. Journal of Process Control, 2015, 32:154-160.

[8] Y. J. Kwon, H. K. Kim, Y. H. Lim, “A behavior-based intrusion detection technique for smart grid infrastructure. Powertech”, IEEE Eindhoven. IEEE, 2015:1-6.

[9] Y. Yang, K. Mclaughlin, T. Littler, “Rule-based intrusion detection system for SCADA networks,”. 2013 IET Renewable Power Generation Conference, Beijing. IET, 2013:1-4.

[10] C. McParland, S. Peisert, A. Scaglione, “Monitoring security of networked control systems: it’s the physics”. IEEE Security & Privacy, 2014, 12 (6)32-39.

[11] Y. Mo, R. Chabukswar, B. Sinopoli, “Detecting integrity attacks on SCADA system,”. IEEE Transactions on Control Systems Technology, 2014, vol (4) 1396-1407.

[12] W. L. Shang, S. SZhang, M. Wan, “Modbus/TCP communication anomaly detection based on PSO-SVM”. Applied Mechanics and Materials, 2014, 490-491:1745-1753.

[13] C. Zhou, S. Huang, N. Xiong, “Design and analysis of multi-model-based anomaly intrusion detection systems in industrial process automation”. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2015, 45(10):1345-1360.

[14] MITRE, “Tactics of defense evasion in ATT&CK”, v15.1. 2024, https://attack.mitre.org/tactics/TA0005/.

[15] Y. Zeng, Y. WU, L.H, Dong, “Research on malicious traffic identification technology in encrypted traffic”, Journal of Xidian University, 2021, 48(3):170-187.

[16] L. T., Huang Z. C. Zhao, Y. Q. Zhao, “A two-stage cryptosystem recognition scheme based on random forest,”. Chinese Journal of Computers, 2018, 41(2):382-399 .

[17] J. Kim, H. S. Kim, “Intrusion detection based on spatiotemporal characterization of cyberattacks”. Electronics, 2020, 9(3):460.

[18] W. Bazuhair, W. Lee, “Detecting malign encrypted network traffic using perlin noise and convolutional neural network”. 2020 10th Annual Computing and Communication Workshop and Conference. IEEE, 2020:200-206.

[19] Y. L. Wang, “Research on conjunctive keyword search over encrypted data in cloud computing,”. Xi’an: Xidian University, 2019.

[20] Z. Q. Fan, Y. Zeng, X. Y. Zhu, “A group key agreement based encrypted traffic detection scheme for Internet of Things”. 2020 1st ACM International Workshop on Security and Safety for Intelligent Cyber-Physical Systems. ACM, 2020:19-26.

[21] T. Chen, “Research and practice on security protection of the electric power Internet of Things”. Network Security Technology & Application, 2020, 12:132-133.

[22] Q. Zheng, H. FMa, Z. B Wang, “Information security technology--transport layer cryptography protocol (TLCP):GB/T 38636—2020,” China Standard Publishing House, 2020.

[23] J. Y. Liu, Y. Z. Zeng, J. Y. Shi, “MalDetect: a structure of encrypted malware traffic detection”. Computers, Materials & Continua, 2019, 60(2):721-739.

[24] A. K. Jain, Data clustering: 50 years beyond k-means. Pattern recognition letters, 2010, 31(8):651-666.

[25] K. Wagstaff, C. Cardie, S.Rogers, “Constrained k-means clustering with background knowledge. 18th International Conference on Machine Learning”. Morgan Kaufmann Publishers Inc., 2001:577-584.

[26] S. Basu, A. Banerjee, R. Mooney, “Semi-supervised clustering by seeding”. 19th International Conference on Machine Learning. Morgan Kaufmann Publishers Inc., 2002:27-34.

[27] F. Xiao, E. H. Chen, Q. Xu Q, “ICSTrace: a malicious IP traceback model for attacking data of the industrial control system”, Security and Communication Networks, 2021:1-14.