Non-Crypto Authentication for Smart Grid Based on Edge Computing

Wei Liu¹, Jianfang Song², Haitao Wu¹, Wenxin Lei², Fuquan Huang¹, Wenjing Hou¹, Honghao Liang¹ and Hong Wen²

¹Shenzhen Power Supply Bureau Co, Ltd., Shenzhen 510820, China
²School of Aeronautics and Astronautics, University of Electronic Science and Technology of China (UESTC), Chengdu 611731, China
Email: sunlike@uestc.edu.cn

Abstract. As a new computing architecture, edge computing has gradually penetrated into the smart grid and provide real time monitoring and control. However, as a new type of node arranged in the network, its own security protection has become an incredibly challenging task. Security authentication is the first line of defence for security protection. Due to its fast and light-weight characteristics, physical layer channel information access authentication technology for non-cryptographic packets is a potential identification method for the smart grid terminal access under the edge computing system, in which the real time response is incredibly important. This paper aims to investigate the physical layer authentication method compared with the traditional upper layer authentication technology, and an authentication scheme suitable for smart grid is proposed.

1. Introduction
A large number of smart terminals are deployed in the smart grid. They can realize the power grid status monitoring and power consumption interaction through communication, which makes the smart grid system vulnerable to message tampering and forgery, malicious node attacks, denial of service attacks (DoS), etc., so it is necessary to provide more effective security for the smart grid [1-3]. Message authentication can ensure the integrity and authenticity of messages [4-6]. But the traditional encryption based upper authentication mechanism has high computational complexity and time delay and is not suitable for real-time message authentication of intelligent terminal equipment with limited resources and energy.

Physical layer (PL) authentication technology is a kind of non-cryptographic authentication methods which can identify the uniqueness of the terminal through the characteristics of the physical layer and try to solve the security access problem of communication from the signal level [7-10]. It is not only lightweight, but also suitable for the access of massive heterogeneous terminals, which does not care about the upper layer protocol, or even do not need to know the detailed structure of the terminals’ physical frame. This paper aims to investigate the PL authentication method compared with the traditional upper layer authentication technology, and an authentication scheme suitable for smart grid is proposed.

2. Edge Computing and PL Authentication
As a new computing architecture, edge computing has gradually penetrated into many fields and plays an important role. Because of the proximity, edge computing network architecture can provide better
security than cloud computing architecture, but as a new type of node arranged in the network, its own security protection has become an incredibly challenging task.

When the security vulnerabilities in edge computing are exploited by hackers, the security of many access terminals and data can not be guaranteed, which will also threaten the security of the network. Therefore, the security protection system of the edge computing needs to be established and improved.

Security authentication is the first line of defence for security protection. Identity and authentication management functions run through the whole security service function level. The reliable identification of access terminal identity by edge devices is very important to protect the security of applications and data. In terms of its function, edge computing devices need to provide real-time response, so the secure access authentication methods need fast and light-weight. Therefore, fast and reliable secure access authentication for terminal devices on the edge side is a major security challenge faced by the edge computing system. At the same time, the computing resources of edge computing devices are much stronger than that of terminals, so it is very meaningful to make full use of the computing power of the edge side to develop a high security and fast authentication method suitable for edge computing architecture.

![Figure 1. Edge computing and physical layer authentication](image)

As the first entrance of data, edge computing has good real-time performance because of its natural distribution close to the terminal, which can well meet the needs of real-time response on the Internet of things and can also provide asymmetric resources. Specially, the identified accuracy of the PL authentication can be highly improved by the machine learning methods. However, training of these methods is cost resources. Under the edge computing system, such cost can be undertaken by the edge device and the terminal almost do nothing. Therefore, this authentication method is an authentication method with asymmetric resources. Compared with the terminal, the edge side can carry more complex computing and storage functions. When a terminal or node transmits data to the edge side, the edge side authenticates the data packet legally through the physical layer authentication. At the terminal, the terminal can extract channel information from the information packet for packet authentication without task redundancy, which makes the terminal or micro node hardly bear the computing load. At the same time, the computing resources of the edge side let the packet authentication to use machine learning and other methods for physical layer authentication.

### 3. The System Model of PL Authentication

In this section, the PL access authentication model is illustrated as Figure 2. The legitimate communication parties L and R conduct communication partners, and the illegal access node E also exists in the network. In this case, the edge device R can use the channel characteristics to identify whether the access node is consistent with its claim, that is, the physical layer authentication based on channel characteristics. In the following, we will use R for edge device, i.e. receiver, L for legal access node, and E for illegal access node Eve.
The specific steps of authentication are as follows. Firstly, suppose that the edge device R and the legal access node L have "handshake", the edge device R completes the first frame data packet preservation, extracts the channel estimation matrix $\hat{H}_{LR}$ on the legal path, and calculates the mean value $\hat{H}_{LR}^1$ of the channel estimation matrix in the time domain as

$$\hat{H}_{LR}^1 = \begin{bmatrix}
\hat{H}_{LR}^{1,(1)} & \cdots & \hat{H}_{LR}^{1,(N_s)} \\
\hat{H}_{LR}^{2,(1)} & \cdots & \hat{H}_{LR}^{2,(N_s)} \\
\vdots & \ddots & \vdots \\
\hat{H}_{LR}^{N,(1)} & \cdots & \hat{H}_{LR}^{N,(N_s)}
\end{bmatrix} \Rightarrow \hat{H}_{LR}^1 = \begin{bmatrix}
\hat{H}_{LR}^{1,(1)} \\
\vdots \\
\hat{H}_{LR}^{N,(N)}
\end{bmatrix}$$

Then, after receiving the second data frame, R also obtains the second frame $\hat{H}_{UR}^2$ from the unknown access node according to formula (1) and compares $\hat{H}_{UR}^2$ with $\hat{H}_{LR}^1$. Assuming that there is a strong correlation between consecutive data frames, the time interval between the two frames is within the coherent time. If the two are close, it means that the sender is a legal access node L; if the two are not close, it means that the intention access node is an illegal access node E, thus, one verification is completed. Finally, a lot of data frames are continuously verified. Suppose R has successfully authenticated that the $k$th data frame is sent by L, and the error free channel information is $H_{LR}^k$; the sender identity of the $(k+1)$th data frame is needed to verify, and the error free channel information is $H_{UR}^{k+1}$. We set threshold as $\eta$, the following assumptions can be made through the binary hypothesis test:

$$\mathcal{H}_0 : \text{diff} \left( \hat{H}_{UR}^{k+1}, \hat{H}_{LR}^k \right) < \eta$$
$$\mathcal{H}_1 : \text{diff} \left( \hat{H}_{UR}^{k+1}, \hat{H}_{LR}^k \right) > \eta$$

Zero hypothesis $\mathcal{H}_0$: the channel information without error in the $(k+1)$th frame is equal to that in the $k$th frame, and the node to be accessed is the legal access node L;

Alternative hypothesis $\mathcal{H}_1$: the channel information without error in the $(k+1)$th frame is not equal to that in the $k$th frame, and the node to be accessed is illegal access node E.

4. The Proposed PL Authentication

Machine learning, as one of the hot topics in academic and industry, aims to improve computer algorithm and performance through continuous data analysis and learning. Its classification algorithm is various and suitable for different data types and scenarios. Combining machine learning with
channel characteristics is a new exploration idea in the field of physical layer authentication. Decision tree, AdaBoost, support vector machine, $k$-nearest neighbour and ensemble learning all are belong to Machine learning, in which the training data dimension is one of the key problems.

The overall idea of the scheme is as follows: the edge device R collects and pre-processes the channel information of legal access node L and illegal access node E to form a binary data set. The training set is modelled by machine learning classification algorithm to train the classifier, and then the trained classifier is used to identify the access node. The details of the proposed scheme are presented as follows.

Step 1, the edge device R collects the channel state information (CSI) of transmission packets from the legal node L and illegal node E as following. Firstly, the edge device R collects and stores the initial CSI from legal node L as $\hat{H}_{LR}^{1,1}$ and stores the second frame CSI from legal node L as $\hat{H}_{LR}^{1,2}$, where the subscript 2,1 denotes the first collection of the 2$^{nd}$ data frame. Then, the edge device R does the first CSI collection to the 3$^{rd}$ data frame and get $\hat{H}_{LR}^{3,1}$ that will be labelled as $\{1, 1\}$, $Y = \{y_1, \ldots, y_\kappa, \ldots, y_\kappa\}$, in which if $y_\kappa = +1$ the data is sent by legal node L, otherwise, the data is sent by illegal node E:

$$
\begin{align*}
\hat{H}_{LR}^{3,1} &= \hat{H}_{LR}^{1,2}, y_\kappa = +1 \\
\hat{H}_{LR}^{3,1} &= \hat{H}_{LR}^{1,1}, y_\kappa = -1, k = 1, \ldots, \kappa
\end{align*}
$$

(3)

Final, the data set containing K acquisition results $\hat{H}_K$ is obtained by cycling the above acquisition 1, 2 and 3 frames

$$
\hat{H}_K = [(\hat{H}_{1,1}, \hat{H}_{2,1}, \hat{H}_{3,1}), \ldots, (\hat{H}_{1,\kappa}, \hat{H}_{2,\kappa}, \hat{H}_{3,\kappa})], k = 1, 2, \ldots, \kappa
$$

(4)

Step 2, edge device R pre-processes and extracts data sets $\hat{H}_K$ as

$$
\begin{align*}
T_{A,K} &= \left[\text{diff} (\hat{H}_{1,1}, \hat{H}_{2,1}, \hat{H}_{3,1}), \ldots, \text{diff} (\hat{H}_{1,\kappa}, \hat{H}_{2,\kappa}, \hat{H}_{3,\kappa})\right], k = 1, 2, \ldots, \kappa \\
T_{B,K} &= \left[\text{diff} (\hat{H}_{1,1}, \hat{H}_{2,1}, \hat{H}_{3,1}), \ldots, \text{diff} (\hat{H}_{1,\kappa}, \hat{H}_{2,\kappa}, \hat{H}_{3,\kappa})\right], k = 1, 2, \ldots, \kappa \\
T_{C,K} &= \left[\text{diff} (\hat{H}_{1,1}, \hat{H}_{2,1}, \hat{H}_{3,1}), \ldots, \text{diff} (\hat{H}_{1,\kappa}, \hat{H}_{2,\kappa}, \hat{H}_{3,\kappa})\right], k = 1, 2, \ldots, \kappa
\end{align*}
$$

(5)

Where $T_{A,K}$, $T_{B,K}$, $T_{C,K}$ denote three different sets and they are divided into training set $T_{A,t}$, $T_{B,t}$, $T_{C,t}$ and test set $T_{A,K+t}$, $T_{B,K+t}$, $T_{C,K+t}$.

Step 3, edge device R transfers the training set $T_{A,t}$, $T_{B,t}$, $T_{C,t}$ as two dimension binary training set $SS$:

$$
SS = [(x_{i1}, x_{i2}), y_i], \ldots, [(x_{i1}, x_{i2}), y_i], \ldots, [(x_{i1}, x_{i2}), y_i]
$$

(6)

In $SS$, we have
\[
\begin{align*}
x_{i,1} & \in \{T_{A,t}, T_{B,t}, T_{C,t}\} \\
x_{i,2} & \in \{T_{A,t}, T_{B,t}, T_{C,t}\}, i = 1, 2, \ldots, t \\
y_i & \in Y
\end{align*}
\]

(7)

Step 4, the edge device R uses machine learning classification algorithm to train the “machine” according to the training set \( SS \) of the second classification to generate the final classifier.

Step 5, the edge device R extracts the packet CSI of the unknown node to be accessed as \( \hat{H}_{3,k}^{LR} \), \( \hat{H}_{1,k}^{LR} \) and \( \hat{H}_{2,k}^{LR} \), and calculation \( T_A \), \( T_B \), \( T_C \) based on the difference of channel information in any two combinations of three consecutive frames. The classifier with qualified authentication rate is used to judge whether the received packet is a legal access node or a faked illegal access node, so as to realize the packet identification. The unknown access node needs to be a trained user, that is, its channel information is collected and stored, because the binary classification learning here is supervision study.

Using two-dimensional features can avoid the contingency of using only one-dimensional features, and can improve the authentication rate of the points to be classified in theory, which has a good advantage to the heterogeneous smart grid terminals.

5. Conclusions
The non-crypto PL authentication scheme for the smart grid edge computing system is proposed, in which the training set is modelled by machine learning classification algorithm to train the classifier, and then the trained classifier is used to identify the access terminal. Two-dimensional features of CSI are used as training and test sets in the machine learning algorithms armed to improve the performance of PL authentication. This scheme makes full use of the computing power of the edge side to develop a high security and fast authentication method suitable for edge computing architecture, in which terminals do almost nothing and can keep high security.

6. Acknowledgments
This work is supported by the National major R & D program (No. 2018YFB0904900, 2018YFB0904905).

7. References
[1] Xie Y, Wen H, Wu B, Jiang Y and Meng J 2019 A modified hierarchical attribute-based encryption access control method for mobile cloud computing, IEEE Transactions on Cloud Computing vol 7 p 383-391.
[2] McDaniel P and McLaughlin S, 2009 Security and Privacy Challenges in the Smart Grid, IEEE Security & Privacy, vol 7 p 75-77.
[3] Chen S, Wen H, Wu J S et al. 2019 Internet of Things Based Smart Grids Supported by Intelligent Edge Computing, IEEE Access, vol7 p74089-74102.
[4] Ji B, Li Y, Zhou B, Li C, Song K and Wen H 2019 Performance analysis of UAV relay assisted IoT communication network enhanced with energy harvesting, IEEE Access vol 7 p 38738-3874.
[5] G. R. Barai, S. Krishnan and B. Venkatesh, 2015 Smart metering and functionalities of smart meters in smart grid - a review, IEEE Electrical Power and Energy Conference (EPEC), London, ON, 2015, pp. 138-145.
[6] Pan F, Pang Z, Luvisotto M, Xiao M, Wen H 2018 Physical-Layer Security for Industrial Wireless Control Systems: Basics and Future Directions, IEEE Industrial Electronics Magazine, vol 12 p 18-27.
[7] Wen H, Tang J, Wu J, Song H, Wu T, Wu B and Sun L M 2014 A cross-layer secure communication model based on Discrete Fractional Fourier Transform (DFRFT), *IEEE Transactions on emerging topics in computing* vol 3 p 119-126

[8] Dan S, Kai Z, Xiang W, et al. 2013 PHY-CRAM: Physical Layer Challenge-Response Authentication Mechanism for Wireless Network, *IEEE Journal on Selected Areas in Communications*, vol 31 p 1817-1827.

[9] Wen H, Ho P H and Wu B 2014 Achieving secure communications over wiretap channels via security codes from resilient functions, *IEEE Wireless Communications Letters* vol 3 p 273-276

[10] Wen H, Ho P H, Qi C and Gong G 2010 Physical layer assisted authentication for distributed ad hoc wireless sensor networks, *IET information security* vol 4 p 390-396

[11] Wen H, Wang Y, Zhu X, Li J and Zhou L 2013 Physical layer assist authentication technique for smart meter system, *IET Communications* vol 7 p 189-197

[12] Hu L, Wen H, Wu B, Tang J, Pan F and Liao R F 2017 Cooperative-jamming-aided secrecy enhancement in wireless networks with passive eavesdroppers, *IEEE Transactions on Vehicular Technology* vol 67 p 2108-2117