symmetry

MDPI

Article

Deep-Reinforcement-Learning-Based Wireless IoT Device Identification Using Channel State Information

Yuanlong Li [1], Yiyang Wang [2], Xuewen Liu [2], Peiliang Zuo [2,*], Haoliang Li [2] and Hua Jiang [2]

[1] Certification Management Division, The Ministry of Information and Network Security of the State Information Center, Beijing 100045, China; lyl@sic.gov.cn
[2] Department of Electronic and Communication Engineering, Beijing Institute of Electronic Science and Technology (BESTI), Beijing 100070, China; besti.wang@foxmail.com (Y.W.); xuewen_liu1990@bupt.edu.cn (X.L.); lhl2018@bupt.edu.cn (H.L.); jhbesti@126.com (H.J.)
* Correspondence: zplzpl88@bupt.cn

Abstract: Internet of Things (IoT) technology has permeated into all aspects of today's society and is playing an increasingly important role. Identity authentication is crucial for IoT devices to access the network, because the open wireless transmission environment of the IoT may suffer from various forms of network attacks. The asymmetry in the comprehensive capabilities of gateways and terminals in the IoT poses significant challenges to reliability and security. Traditional encryption-based identity authentication methods are difficult to apply to IoT terminals with limited capabilities due to high algorithm complexity and low computational efficiency. This paper explores physical layer identity identification based on channel state information (CSI) and proposes an intelligent identification method based on deep reinforcement learning (DRL). Specifically, by analyzing and extracting the features of the real received CSI information and a setting low-complexity state, as well as action and reward parameters for the deep neural network of deep reinforcement learning oriented to the scenario, we obtained an authentication method that can efficiently identify identities. The validation of the proposed method using collected CSI data demonstrates that it has good convergence properties. Compared with several existing machine-learning-based identity recognition methods, the proposed method has higher recognition accuracy.

Keywords: IoT; channel state information; encryption; reinforcement learning; identity recognition

Citation: Li, Y.; Wang, Y.; Liu, X.; Zuo, P.; Li, H.; Jiang, H. Deep-Reinforcement-Learning-Based Wireless IoT Device Identification Using Channel State Information. *Symmetry* **2023**, *15*, 1404. https://doi.org/10.3390/sym15071404

Academic Editor: Christos Volos

Received: 17 May 2023
Revised: 4 July 2023
Accepted: 7 July 2023
Published: 12 July 2023

1. Introduction

The past few decades have witnessed the rapid development of computer technology and internet technology, as well as the great convenience they have brought to social life. As an important technology and application extension, the Internet of Things (IoT) is significantly bettering the human condition, augmenting enterprises, and benefiting society through broader automation, connectivity, efficiency, security, data insights, and asset optimization [1,2]. The IoT, which is closely related to 5G and 6G, is indispensable for continued progress and expansion of the digital economy into all realms of life and industry [3,4]. Additionally, the world is becoming highly integrated, automated, and data-driven thanks to the power and promise of IoT [5–7].

Thanks to the large number of widely deployed and diverse sensing terminal devices, the IoT can timely and comprehensively obtain the dynamic situation of the target area, and can provide managers with good decision-making and feedback interaction. However, the IoT also faces greater security risks at the same time, as the large number of terminals poses challenges for the access management of the IoT gateway [8–11]. Specifically, if the terminals cannot be authenticated in a timely and reliable manner, the IoT may face problems of low efficiency and network attacks. Eventually, the network will be paralyzed, and sensitive data will be lost. For the IoT, due to the defects in identity authentication, the types of network attacks have become diverse, such as eavesdropping, man-in-the-middle

attacks, denial of service attacks, data loss, and data tampering, etc. These problems show that efficient and reliable identity authentication and real-time terminal status monitoring are important for the IoT [12–14].

In general, cryptographic methods can reliably authenticate user terminals. Specifically, the IoT gateway and terminals can rely on symmetric encryption or asymmetric encryption methods to achieve one-way or two-way authentication. However, these related methods generally have the characteristics of complex protocol processes, pre-shared keys, and high computational complexity, which make the methods difficult to apply to IoT scenarios where the computing and storage capabilities of terminals are weak. By comparison, identity authentication methods founded in physical layer attributes tend to be lightweight and streamlined. These characteristics are expected to make them an important support for the identity authentication process in IoT scenarios [15].

1.1. Related Work

Compared with identity authentication methods based on cryptographic algorithms, authentication methods created on the basis of physical layer characteristics generally have the advantages of low complexity and high execution efficiency. Common physical layer characteristics include, but are not limited to, receive signal strength (RSS), channel impulse response (CIR), carrier frequency offset (CFO) and channel state information (CSI). These features or measurements are generally difficult for attackers to imitate or predict because they have obvious differences in time and space.

By analyzing the CSI of signals received from Wi-Fi access points, human activities can be detected in [16], as CSI could characterize how Wi-Fi signals propagate through the surrounding environment. Similarly, by capturing the subtle impact of people on Wi-Fi signal properties as evidenced through transitions in CSI, artificial intelligence can gain acute insight into the presence, movements, and behaviors of inhabitants within an environment [17]. CSI presents a conduit for the contactless sensing of human activity with far-reaching possibilities. To improve the efficiency of CSI-based device authentication and enhance the accuracy of such methods for authenticating mobile devices, ref. [18] proposed using small-scale autoencoders to authenticate individual packet data. In ref. [19], the mapping relationship between CSI information and devices was learned by deep neural networks to ensure that the authenticator had a high accuracy and could be independent of encryption-based authentication methods. By observing the subtle effects of user keyboard tapping on Wi-Fi CSI, a user identifier based on multiple classifiers was proposed in [20], and experiments based on real data showed that its accuracy was satisfactory.

Based on CSI fingerprint extraction, ref. [21] proved that its proposed intelligent recognition algorithm for the IoT could achieve an authentication cycle of 1.5 ms, which has significantly higher efficiency than cryptographic authentication methods. The authors in [22] divided the CSI-based user identification process into login and interaction stages. The former realizes the learning of CSI features through deep learning, while the latter realizes the recognition of user finger gestures through classifiers. A framework for user identification in both static and dynamic scenarios was established in [23]. The CSI-based method could accurately distinguish different users and could determine the identity of dynamic users through the time–domain correlation of CSI measurements. In ref. [24], the spatio-temporal features of the environment contained in the CSI information were extracted by convolutional neural networks (CNN) and bound to the device for identity authentication. The method still had a high recognition accuracy at low signal-to-noise ratios (such as 0 dB).

The existing identification methods using physical layer information are gradually improving in performance. However, due to the complexity of authentication methods or the limitations of authentication accuracy, they are still difficult to apply to authentication environments with a large number of wireless terminals. Especially for the identity authentication of numerous terminals in the IoT scenario, lightweight and high accuracy have always been the goals pursued by authentication methods. Existing methods are difficult to

balance well between the two. Cryptographic authentication methods have high accuracy but low efficiency, while physical layer feature-based authentication methods have room for improvement in accuracy.

1.2. Contributions of This Paper

To further improve the accuracy and efficiency of identity authentication methods for scenarios such as the IoT, an intelligent authentication method with low complexity and high accuracy based on CSI information has been proposed. The key contributions of this paper are threefold:

- By setting up a real wireless communication environment, we completed the collection of CSI data for Wi-Fi and preliminarily demonstrated the ability of CSI data to support device identity discrimination through data processing and analysis.
- Furthermore, an identity authentication method on the basis of deep Q-network (DQN) has been proposed. By processing CSI and setting parameters including state space as well as action space, the method realizes lightweight mapping of the CSI features and device identity.
- Finally, we verified the performance of the proposed method and the comparative methods based on the collected CSI data. A large number of experiments showed that the proposed method has better authentication accuracy compared with the comparative methods.

This paper is organized as follows. The system model and CSI data collection process are introduced in Section 2. Section 3 covers the preliminary knowledge related to the proposed method in this paper. Section 4 provides the proposed intelligent identification method using CSI. Section 5 presents the experimental results and performance analysis, and Section 6 provides the conclusions.

Notation: In this paper, scalars are expressed by a non-boldface type, while matrices, as well as vectors, are expressed by a boldface type. $E\{\cdot\}$ and $(\cdot)^T$ signify the statistical expectation and matrix transpose, respectively. Finally, ω_i denotes the ith entry of the vector $\boldsymbol{\omega}$.

2. System Model and CSI Collection

In this section, we first introduce the model of wireless channel CSI, which helps to demonstrate the generation principle of observational data for identity authentication methods based on physical layer information. Then, we introduce the process of obtaining CSI data, which is used for CSI feature analysis and performance verification of the proposed method in this paper.

2.1. CSI Model

Unlike wired channels, wireless channels may face a series of attenuation processes during signal transmission, such as large-scale fading, shadowing, multipath effects, and distortion. CSI is used to describe the state attributes of a communication link. It can reflect the attenuation of signals in the transmission link. Mathematically, it corresponds to the value of each element in the channel gain matrix $\boldsymbol{H}$, such as distance attenuation, environmental attenuation, and signal scattering information. By using channel state information, communication systems can adaptively adjust relevant transmission and reception processes to adapt to the time–domain dynamics of the channel. CSI information can also guide the adjustment process of multi-antenna communication systems to improve the overall performance of the system. The signal reception process of a wireless fading channel can be modeled as follows:

$$\mathbf{y} = \mathrm{H}\mathbf{x} + \mathbf{w}, \tag{1}$$

where $\boldsymbol{y}$ denotes the received signal vector, $\mathbf{w} \sim N(0, \sigma_w^2 \mathbf{I_n})$ is the additive Gaussian white noise vector, and $\mathbf{x} = [x_1, \ldots, x_M]$ represents the transmitted signal vector. Con-

sidering the characteristics of the normal distribution, the channel coefficient matrix $\mathbf{H} = diag\{h_1, \ldots, h_N\}$ also follows the same logarithmic normal distribution. Without a loss of generality, the probability density function of h can be expressed as follows [25,26]:

$$p_h(h) = \frac{1}{\sqrt{2\pi}\,\sigma_h\, h} \exp[-\frac{(\ln h - \mu_h)^2}{2\,\sigma_h^2}]. \tag{2}$$

σ_h^2 and μ_h correspond to the variance and mean of $\ln h$, respectively, and $\ln h$ follows a Gaussian distribution. We standardize h, i.e., $E[h] = 1$, where $E[.]$ represents the expectation calculation to ensure that the average power of the wireless signal is constant. Thus, the probability density function of h can be further expressed as follows:

$$p_h(h) = \frac{1}{\sqrt{2\pi}\,\sigma_h\, h} \exp[-\frac{(\ln h + \sigma_h^2/2)^2}{2\,\sigma_h^2}]. \tag{3}$$

For the symbol detection process of the receiver, a simple and low complexity linear detector minimum mean square error (MMSE) is usually used for processing to achieve channel estimation and equalization, which can be represented as follows:

$$\hat{\mathbf{x}} = \arg\min_{x \in \chi} \left\| \mathbf{y} - \hat{\mathbf{H}}\mathbf{x} \right\|_2^2, \tag{4}$$

where $\hat{\mathbf{H}}$ represents the estimated value of the channel state information, and $\|.\|_2$ represents the binomial operation. According to references [27,28], it can be assumed that each constituent element in $\hat{\mathbf{H}}$ follows a Gaussian distribution; then, we have $\hat{\mathbf{H}} \sim N(\mathbf{H}, \sigma_h^2\, \mathbf{I_n})$.

From the above derivation, it can be seen that the channel state information can be calculated by referring to the estimation process of the channel. The complex sequence corresponding to CSI values can be extracted as follows:

$$\mathbf{Y} \in C^{N_{sa} \times N_{sc} \times N_{tx} \times N_{rx}}, \tag{5}$$

where N_{sa} represents the number of sampled network packets, N_{sc} represents the number of subcarriers, and N_{tx} and N_{rx} represent the number of transmitted and received antennas, respectively.

It is worth noting that the CSI information of wireless channels has two unique properties. On the one hand, the interchangeability of channel status indicates that the channel state information detected by the transmitting and receiving ends at the same time is nearly the same. On the other hand, there is negligible spatial correlation; that is, when the distance between the channel transmitting and receiving ends is more than half a wavelength, the detected CSI information is not correlated with the information of the transmitting and receiving ends. The unique characteristics of the wireless channel CSI have made important preparations for its application as one of the physical layer security means. Since the representation parameters of channel state information are numerous and have heterogeneity in space but only show short-term correlation in time, and because the channel state information is jointly determined by the specific transmission environment, transmitter, and spatio-temporal characteristics, the transmitting and receiving ends can judge the identity change of the communication counterpart based on the CSI information.

2.2. CSI Data Collection

We adopted the Intel 5300 network card, which contains three antennas with omni-directional signal reception characteristics to collect Wi-Fi protocol data (as shown in Figure 1). By replacing the network interface card (NIC) of the Intel 5300 network card on the computer and relying on the network cable for external connection, the uninterrupted observation of the data collection process could be realized. The data collection processing computer was installed with the Linux 802.11n CSI tool [29], which can obtain signal strength and phase information for Wi-Fi data packets at the subcarrier level of the channel of the multi-antenna receiving device and transmit the related data to the computer for processing.

The CSI data collection process was conducted in an un-manned laboratory during the process. Considering that the Intel 5300 network card cannot connect multiple Wi-Fi APs at the same time, during the data collection process, multiple sets of CSI data were obtained by changing the data collection time and access point (AP) position to analyze the differences in the channel state information in both spatio-temporal and temporal dimensions. The product model of the AP is a TP-LINK AC1900. During the data collection process, we set the usage frequency band of the AP to 2.4 GHz and set the channel to randomly selected, which could well achieve the randomness of the collected data.

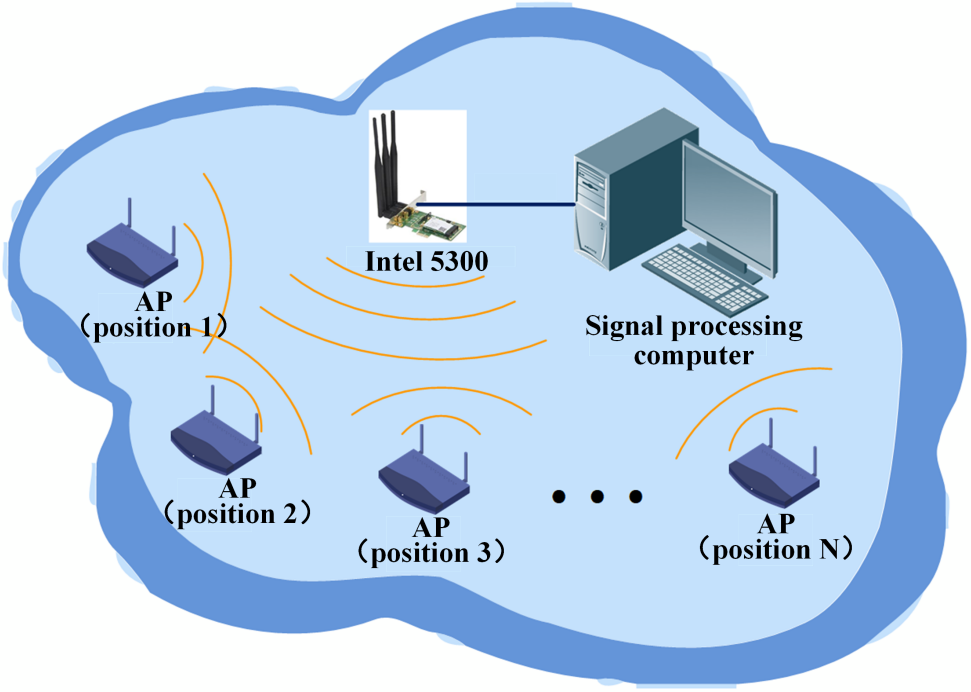

Figure 1. Schematic diagram of CSI data collection process in this paper.

The CSI tool used in this paper can achieve uniform sampling of the 20 MHz bandwidth channel where Wi-Fi is located. For each receiving antenna, 30 uniform frequency domain CSI samples can be obtained. Therefore, 90 sample values were obtained in total for each sampling during the collection process. To fully reflect the dynamic differences of channel state information in the time domain and spatial domain, this paper collected a total of 10 groups of sample data, with 10,000 samples in each group, corresponding to a data collection duration of 100 s, i.e., the sampling frequency of the sample data was 10 ms each time. The data size of each group was around 7.2 MBytes.

3. Preliminaries

To support the intelligent authentication method proposed in this paper, we first introduced deep reinforcement learning and then analyzed the characteristics of the collected CSI data.

3.1. Deep Reinforcement Learning

Reinforcement learning is a technique that is used to enhance the model's coping ability through interaction with the environment. It iteratively trains the model by selecting actions in a constantly changing environment so as to achieve the model's ability to choose high-scoring actions [30–35]. In general, the Markov decision process (MDP) has been utilized to describe the reinforcement learning process. Specifically, in a certain environmental state, the agent of the reinforcement learning makes action choices in a certain way or probability, thus promoting the transformation of the environmental state and obtaining corresponding rewards. Mathematically, we denote the environmental state at time t as

s_t, and the selected action as a_t; therefore, a numerical reward r_t can be acquired, and the subsequent state of the environment would be transformed into s_{t+1} at the next moment. The quadruple $\langle \boldsymbol{S}, \boldsymbol{A}, \boldsymbol{P}, \boldsymbol{R} \rangle$ with $\boldsymbol{S}, \boldsymbol{A}, \boldsymbol{P}, \boldsymbol{R}$ respectively represent the state space, action space, reward set, and state transition probability matrix, and they are usually adopted to describe the learning process of the reinforcement learning. The decision-making strategy of the agent $\pi_t(s, a)$ that indicates the action choice recommendation for state s at time t can be updated once the experience sequence $\{(s_t, a_t, r_t, s_{t+1}), \ldots\}$ has been obtained during the agent's ongoing exploratory learning. The target of reinforcement learning is to ensure that the agent has the ability to obtain the maximum expected cumulative reward in the decision-making process. Then, we can phrase it as follows:

$$R_t = \sum_{l=0}^{\infty} \vartheta^l r_{t+l}, \tag{6}$$

in which $\vartheta \in [0, 1]$ denotes the discount ratio.

Q-learning is a popular approach among reinforcement learning methods [34,35], whereby the Q-values of the model are updated via interactions with the environment. The Q-value mirrors the effectiveness evaluation of the appropriating action a under policies π in circumstantial state s. The environment can be expressed as follows:

$$Q^{\pi}(s, a) = E[R_t \mid s_t = s, a_t = a]. \tag{7}$$

We denote $Q^*(s, a) = \max_{\pi} Q^{\pi}(s, a)$ as the optimal action value function and, according to the Bellman optimality equation, we can acquire the following formula:

$$Q^*(s, a) = E\left[r_{t+1} + \vartheta \max_{a'} Q^*\left(s', a'\right) \mid s_t = s, a_t = a\right], \tag{8}$$

in which s' denotes the next state after the action a was selected and utilized by the agent. By utilizing the experience sequence acquired through the interaction with the environment, the agent of Q-learning can continuously update the optimal action value function. By denoting $q(s_t, a_t)$ as the estimated Q-value of the iterative process, we can express the update process of Q-learning using the following formula:

$$q(s_t, a_t) \leftarrow q(s_t, a_t) + \xi(r_t + \vartheta \max_{a'} q(s_{t+1}, a_{t+1}) - q(s_t, a_t)), \tag{9}$$

in which $\xi \in [0, 1]$ represents the learning rate.

Exploration and exploitation are two important processes in reinforcement learning, which should be properly handled to achieve a reasonable balance. The exploitation process signifies the agent adopting the learned decision-making strategies, i.e., selecting actions solely based on the model's output. Meanwhile, the exploration process means that the agent selects actions in a random manner in order to explore the decision space. It should be noted that an agent should not solely favor one process to preclude scenarios in which model training is absent or model decisions are overly random; these are detrimental to the convergence or performance improvement of the model. To avoid this situation, the ε-greedy algorithm is usually adopted, which can be expressed as follows [31–35]:

$$a = \begin{cases} \arg\max_a q(s, a), & \text{with probability } 1-\varepsilon \\ \text{a random action}, & \text{with probability } \varepsilon \end{cases}. \tag{10}$$

In general, the reinforcement learning method can effectively train the decision model and guide the agent's action selection, and the reward value of the decision is satisfactory. However, the challenge is that when the number of elements in the state space or action space increases sharply, reinforcement learning will face the dilemma of both training efficiency and of decision-making efficiency declining, due to the huge database corresponding to the Q-table. The deep reinforcement learning (DRL) method integrating reinforcement learning and deep learning was proposed to overcome the problems faced by reinforcement learning [36]. DRL aims to use deep neural networks (DNNs) to map between states and actions instead of directly recording the correspondence between state elements and action elements in reinforcement learning. In other words, the efficiency of the DRL method has

been improved, because the data table in reinforcement learning has been replaced by a practical neural network that can evaluate the function of action value. Due to the ability of deep neural networks to fit complex nonlinear relationships, DRL can generally handle the problem where a large number of elements are generated by the large-scale state space and action space well.

As a common method of DRL, the change in the DQN relative to Q-learning in mathematical expression is reflected as $Q^*(s,a) \approx Q(s,a|\boldsymbol{\theta})$. The input of the DNNs in the DQN is one state s, while the output is the Q-value of each action in the action space. $q(s,a|\boldsymbol{\theta})$ represents the output of the neural networks, and it is determined by the weights $\boldsymbol{\theta}$ of the neural networks. The backpropagation process is generally applied to the update learning process of the model parameters (i.e., $\boldsymbol{\theta}$). On the basis of equality (8), the following formula should be utilized to gain the target in the learning model that adopts the architecture of the dual DNNs:

$$L(\boldsymbol{\theta},\boldsymbol{\theta}') = \mathrm{E}\left[\left(r(s,a)+\vartheta \max_{a'} Q(s',a' \mid \boldsymbol{\theta}') - Q(s,a \mid \boldsymbol{\theta})\right)^2\right], \tag{11}$$

where $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$ separately denote the weights of the main DNN and the target DNN, respectively.

3.2. CSI Data Feature Analysis

The data of the three antennas obtained using the CSI tool are shown in Figures 2, 3, and 4, respectively. To intuitively observe the changes of the data in the time domain, the 10 groups of data were respectively smoothed by a sliding average using a window of size five (Smoothing data can enhance its observability. We find that the larger the smoothing window size, the better the smoothing effect. We chose a smooth window of size five in this paper, because when the smooth window reached that size, we could clearly see the differences between different data groups). It can be seen from the three figures that the CSI data of different groups showed different amplitude levels, respectively, and the oscillation trends of each group of data in the time domain were also somewhat different, which preliminarily reflects that there are obvious differences between IoT devices in the time domain and spatial domain. Meanwhile, some interesting phenomena can be observed: for example, the sampling data of some groups showed obvious oscillation characteristics on some antennas, but were relatively stable on other antennas, such as sampling groups 1, 3, and 10. This proves to a certain extent the incoherent characteristics of signals in space, even for very small spatial distances between several receiving antennas. It should also be noted that the CSI values received by different antennas for the same (group) signal were obviously different, thus indicating that the transmission of the signals in space will experience different attenuation processes, which also provides a prerequisite for using CSI data to further authenticate terminal device fingerprints (identity status).

In order to show the statistical situation of the obtained CSI data, we calculated the average value and variance of 10 groups of data respectively. The results are shown in Figures 5 and 6, respectively. From Figure 5, it can be seen that, due to the differences in the position of the AP and the sampling time, the CSI average value of the same antenna showed obvious differences, and, for the same reason, the CSI average values corresponding to different antennas in different groups also showed differences. It is worth adding that changing the position of the AP actually corresponded to changing the position of each receiving antenna of the network card, because the wireless channel had interchangeability, which was also the main reason for the large difference in the average CSI value of each group.

It can be observed from Figure 6 that the variance of different data groups showed obvious differences. The variance of groups 2, 3, and 10 was generally larger, thus indicating that the CSI data of these groups fluctuated greatly in the time domain; that is, the correlation in the time domain was poor. In comparison, the overall variance of the remaining data groups was relatively small, but, for each group of data, there was CSI

variance intensity among different antennas alternates. The fluctuation characteristics of the CSI data curve can also be observed in combination with Figures 2–4.

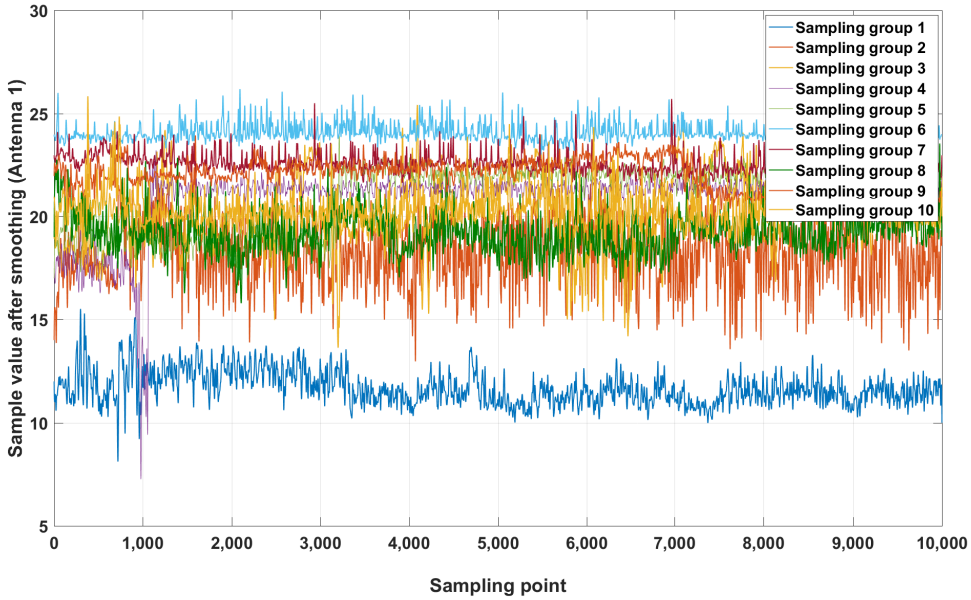

Figure 2. The smoothed CSI value received by antenna 1.

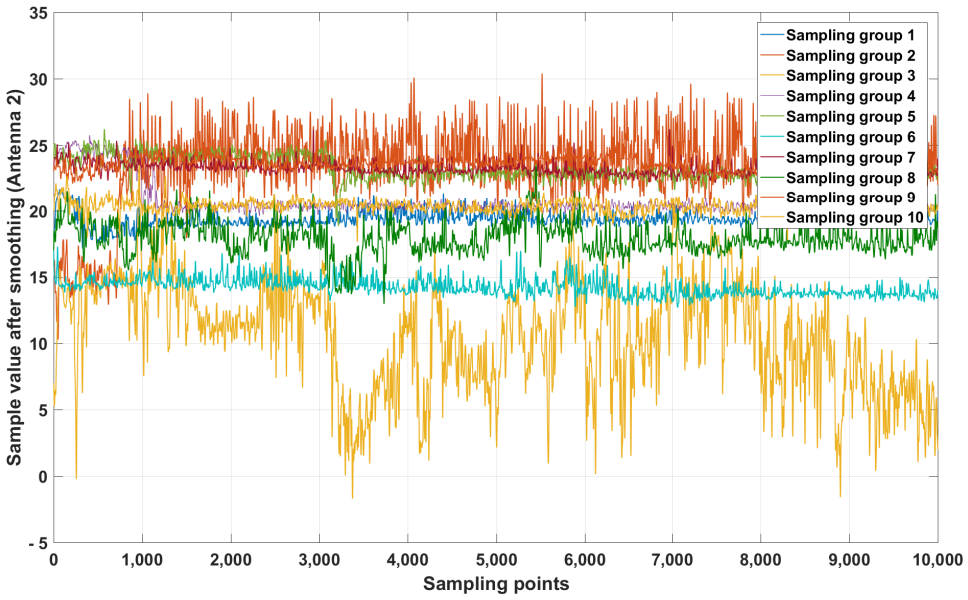

Figure 3. The smoothed CSI value received by antenna 2.

Next, we calculated the correlation between the data of different antennas corresponding to the 10 groups of collected CSI data in order to show the signal reception coherence under the condition of tiny spatial differences between different antennas. It can be seen from Figure 7 that there was a certain correlation between the data of the different antennas. Except for the relatively small correlation between the data of groups 3 and 10, the correlation between the data of other groups of antennas was above 0.1, and the highest correlation of each group of data did not exceed 0.9. Except that the correlation between the three pairs of antennas in groups 6, 7, 8, and 9 was higher than 0.6, the correlation between the antenna data of other groups was not high. This indicates two points: (1) There were obvious differences in the correlation between antennas of each data group; (2) Although

the correlation between data of the same group of antennas was significant, it did not reach a completely consistent level. This shows that it is necessary to collect CSI data from multiple antennas, and the differences of CSI data in space and time can be better presented by multiple antennas.

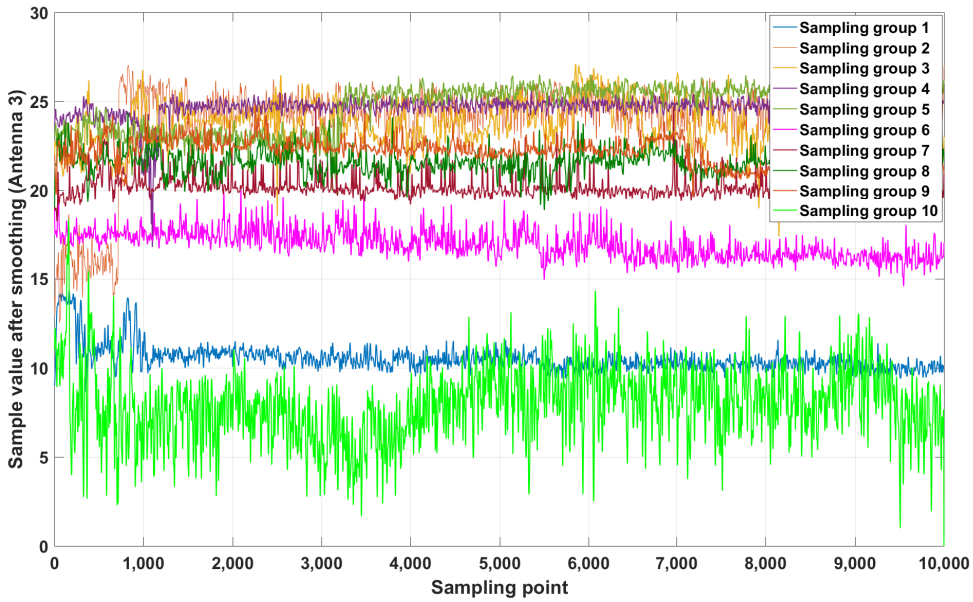

Figure 4. The smoothed CSI value received by antenna 3.

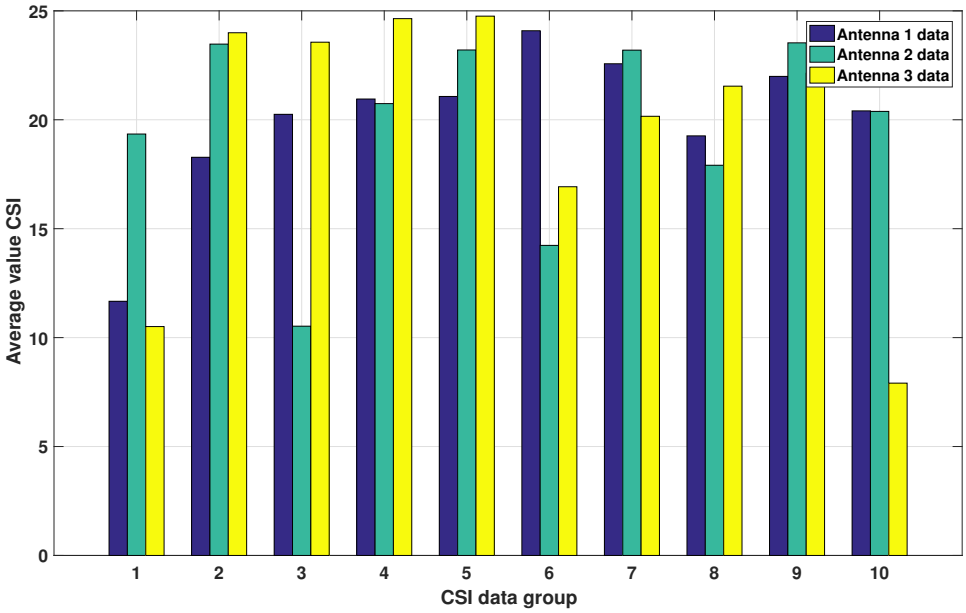

Figure 5. The average of obtained CSI data.

In order to further reflect the differences between the collected CSI data in different groups, we calculated the correlation between data in different groups. The results are shown in Figure 8. For convenience of display, we selected the first group of data as the reference data to calculate the correlation with the other nine groups of CSI data. It can be clearly seen from the figure that, unlike Figure 7, the correlation between groups was significantly lower, and the highest value (absolute correlation) did not exceed 0.25. For different antennas, the correlation between data groups was also different, which again reflects the differences in signal attenuation in space.

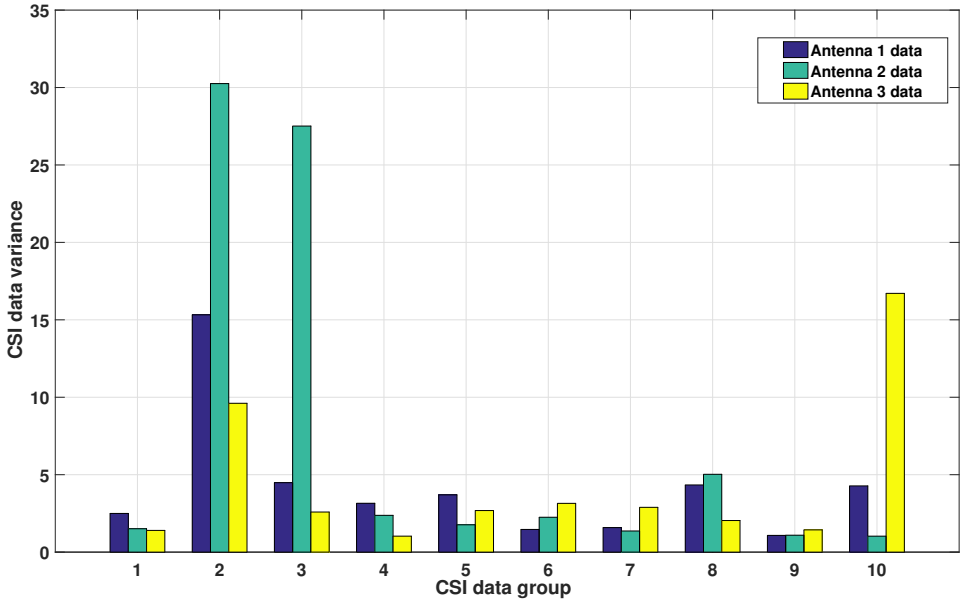

Figure 6. The variance of obtained CSI data.

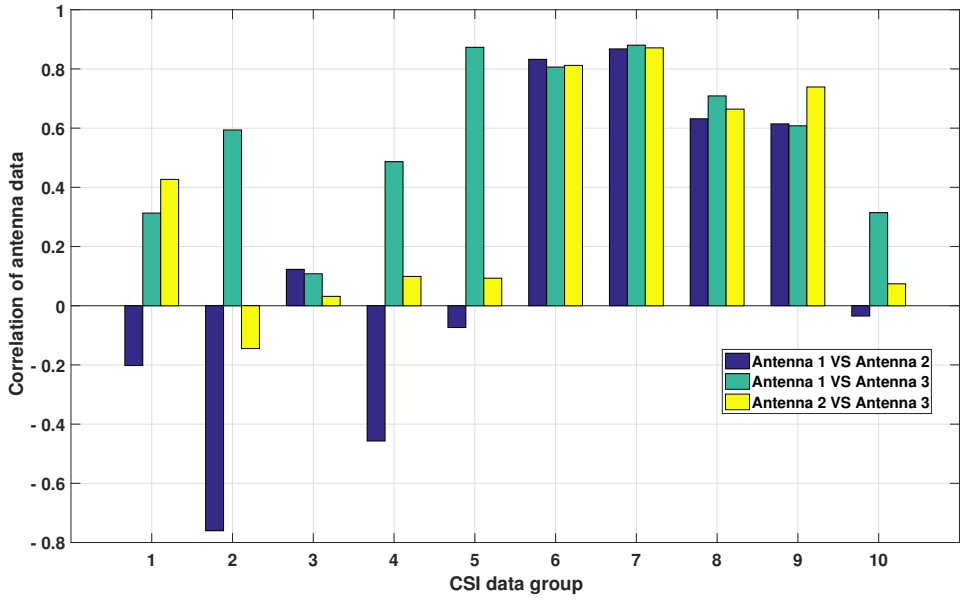

Figure 7. Correlation of different antenna data within the CSI data set.

Finally, we calculated the correlation between the CSI data in the same group under different time shifts in order to show the time domain coherence of the CSI data. The results are shown in Figure 9. Without a loss of generality, the data of antenna 1 was selected for calculation. It can be observed from the figure that, except for the data of groups 2, 6, 7, and 10, the data of other groups showed obvious correlation in the time domain. As the time domain offset increased, the correlation of the data of groups 3 and 8 showed a decreasing trend, while the data of groups 1, 4, and 7 showed a relatively regular oscillation characteristic. Overall, the correlation of each group of data in the time domain was less than 0.8, thus reflecting the differences of each group of data in the time domain and indicating that a 100% accuracy of the CSI data (identification) cannot be obtained through linear prediction or deduction.

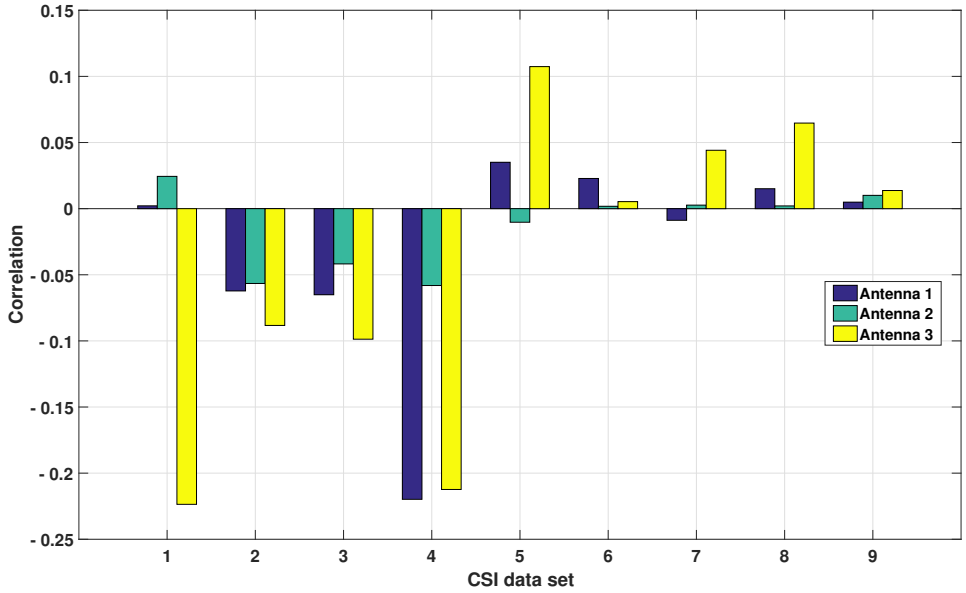

Figure 8. Correlation between different CSI data groups.

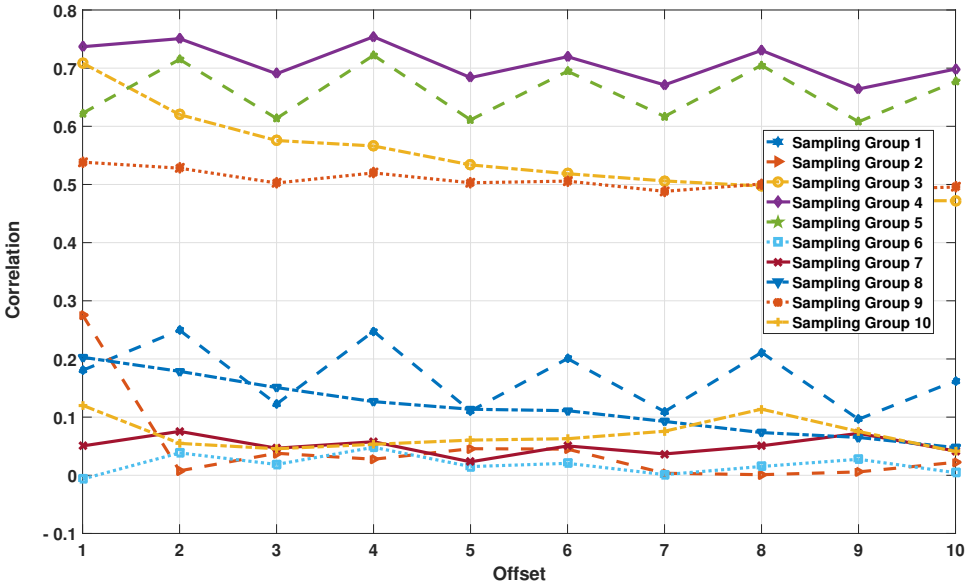

Figure 9. Correlation of CSI data from the same group and antenna at different offsets.

4. Proposed Intelligent Identification Method

Reinforcement learning techniques have the advantage of high efficiency in intelligent decision making. This is because they can reflect the decision environment, decision actions, and feedback after taking certain decision actions in the environment, and the techniques achieve this relatively conveniently and intuitively based on the setting of states, actions, and rewards. On the basis of inheriting the excellent characteristics of reinforcement learning techniques, deep reinforcement learning techniques effectively solve the dimensionality disaster faced by reinforcement learning techniques. By using deep neural networks, they can quickly fit the mapping relationship between elements in huge state space and action space. The above characteristics of deep reinforcement learning techniques enable them to be well suited for a wireless device authentication scenario based on the physical layer concerned in this paper. On the one hand, the state elements reflecting the characteristics of the physical layer channel can be obtained through simple data preprocessing. On the other hand, the authentication results and effects correspond to

the action elements and reward details, respectively. Through a certain amount of sample training, the constructed model can closely matched with the considered scenario. We summarize the DRL-based intelligent authentication framework relying on CSI information in Figure 10. We now introduce the proposed method in detail.

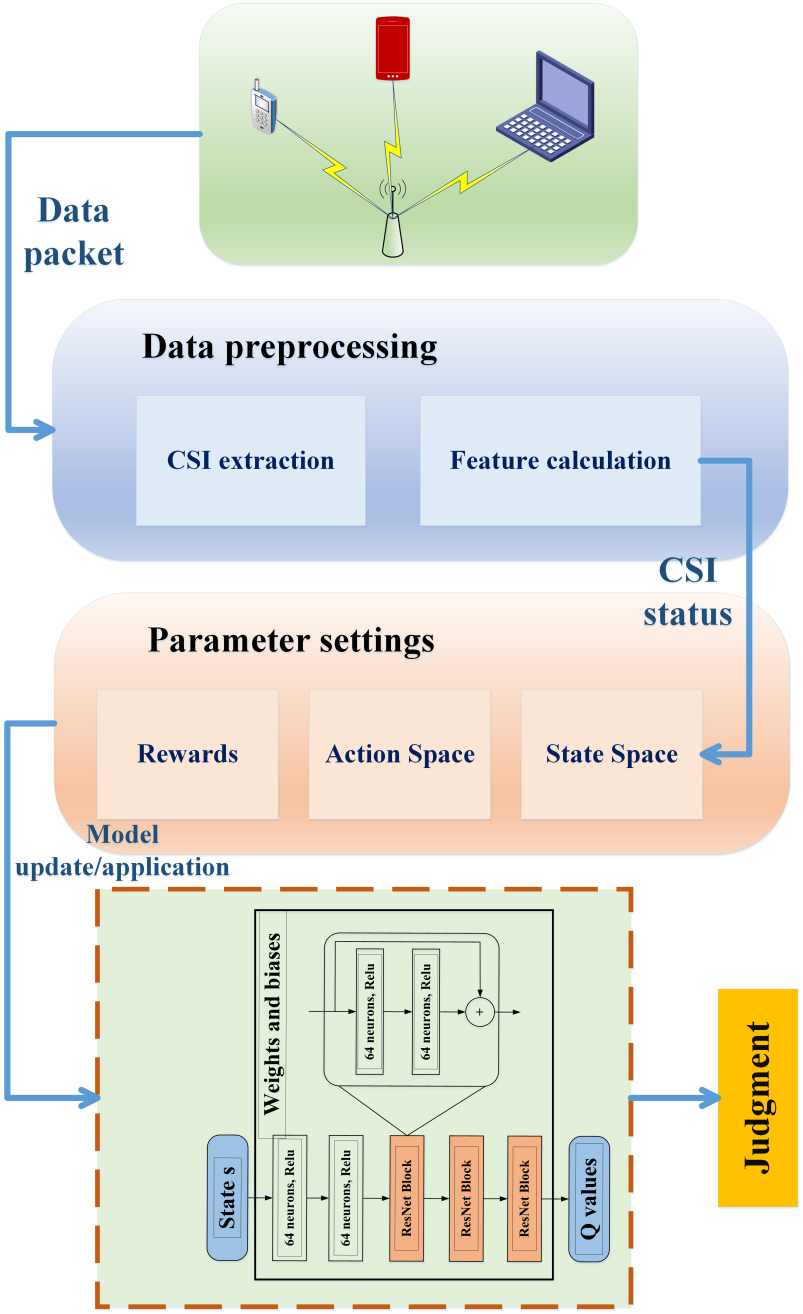

Figure 10. The framework for intelligent identity authentication based on CSI information.

4.1. State Space

For the state setting of the deep reinforcement learning method, it is necessary to comprehensively and objectively reflect the situation of the environment in which the agent is located. The data used in this paper is the CSI information of the wireless channel. This information can characterize the spatial and temporal attenuation process of the transmitter signal in the frequency domain at a fine granularity. Considering that the number of elements in the state space and action space of the DQN method can significantly affect the training and application complexity of the method, thus, in order to compress the size of the state space, we performed feature extraction on the data in the previous chapter. Under

the premise of not affecting the information carried by the collected data itself, we used the principal component analysis (PCA) method [37] to reduce the dimensionality of the data. This process can be mathematically expressed as follows:

$$\chi = f_{PCA}(\Psi_{CSI}), \tag{12}$$

where Ψ_{CSI} represents the raw CSI data information received by the gateway, and f_{PCA} denotes the PCA processing operation. The dimension of Ψ_{CSI} is 3×30, which corresponds to the CSI sampling data in Section 2.2. Through this processing operation, the size of the method state space is significantly reduced.

Considering the extreme variability of the CSI value data after dimensionality reduction, we normalized the data to a range of 0 to 1 based on their original value range. Moreover, considering that the continuous real number of state data means that the deep reinforcement learning state space is extremely large, we performed equi-distant discretization processing on the data. Meanwhile, in order to provide the deep reinforcement learning model with an adequate amount of information so that it could determine the provenance of CSI data, we also encompassed some data features in the state. Mathematically, the state can be expressed as $s = \left[\widehat{\chi}, \widehat{\eta}, \widehat{\mu}\right]_{1\times(M+6)}, s \in S$, where $\widehat{\chi}$ is the CSI vector data received by the gateway node after PCA processing at a certain time, $\widehat{\chi} = \left[\widehat{\chi}_1, \widehat{\chi}_2, \ldots \widehat{\chi}_M\right], 0 \leq \chi_m \leq 1$ holds, M denotes the number of principal components, and $\widehat{\eta}$ and $\widehat{\mu}$ respectively denote the mean and variance of the CSI data after discretization.

4.2. Action Space

Corresponding to the previous analysis, we adopted the terminal identification approach for authentication. Therefore, the output of the deep reinforcement learning model was the terminal identifier. In other words, the action space of the method corresponded to the number of terminals. The action setting of the method was the terminal identifier to achieve the purpose of identity classification determination.

4.3. Action Reward

For the authentication process of the wireless channel, the action of deep reinforcement learning model is the identity identifier of the terminal. There was no pre-level distinction between the terminals considered in this paper. Then, for the authentication method, its successful authentication or not corresponded to whether the terminal identification was correct or not. Therefore, the reward setting of the proposed method was measured in positive or negative values. In the setting of this paper, when the gateway (agent) positively authenticates the terminal, the reward r takes the value of 10, otherwise, r takes the value of -100, in order to guide the training process of the model to drive towards high reward values.

4.4. Other Settings

To address the degradation problem in the existing deep neural network learning process, we adopted the residual network (ResNet) [36] as the network structure of deep reinforcement learning. The residual network realizes skip connection through internal residual blocks, which significantly alleviates the gradient disappearance problem caused by increasing the depth of neural networks. In order to improve the learning rate of the DQN network and avoid the problem of training oscillation in the deep reinforcement learning method, a double deep neural network (i.e., the main network and the target network) was adopted to map the relationship between state and action. By synchronizing the parameters of the target network with a certain delay, it can better guide the training amplitude and training direction of the main network, thereby avoiding the unstable situation of multi-dimensional fluctuation. In addition, a ResNet with eight hidden layer structures was utilized to estimate the Q-value of the reinforcement learning. The optimizer

chose Adam, and the activation function chose ReLU [38,39]. The input and output of the network corresponded to the dimensions of the state and action, respectively.

4.5. Summary of the Method

This subsection summarizes the proposed physical layer identity authentication (DQN-PIA) method based on the deep Q-network in Algorithm 1. Among them, $\varepsilon, \varepsilon_{decay}$, and $\varepsilon_{\min}$ represent the initial value, attenuation value, and minimum value of the ε-greedy algorithm, respectively. The Equation (13) in Step 12 of Algorithm 1 is similar to Equation (8), and it can be represented as follows:

$$L(\boldsymbol{\theta}, \boldsymbol{\theta}') = \mathrm{E}\left[\left(r(s,a) + \vartheta \max_{a+1} Q(s+1, a+1 \mid \boldsymbol{\theta}') - Q(s,a \mid \boldsymbol{\theta})\right)^2\right], \tag{13}$$

where $\boldsymbol{\theta}$ is the weight of the target network. It is worth emphasizing that the method stores the learning experience of the network by using the history playback library $\boldsymbol{\Gamma}$, and it obtains small batches of data by referring to a certain priority for network training so as to avoid the dilemma of overfitting in the training process. For the training of the network, the method also sets an activation threshold: that is, when the threshold for the number of actions is reached, it can avoid frequent learning operations and allow the intelligent agent to obtain a sufficient exploration experience. Meanwhile, the method sets the frequency threshold ϕ for the target network to update the weights of the main network, and it does this mainly to prevent the main network from entering the overfitting state during the training process, thus ensuring the performance of the method and increasing the convergence speed of the training process.

Algorithm 1: The proposed DQN-PIA method

1: **Input**: State space $\boldsymbol{S}$, action space $\boldsymbol{A}$, discount rate ϑ, learning rate ξ.
2: **Output**: Weight matrix of the main DNN $\boldsymbol{\theta}$, terminal category determination (i.e., model actions).
3: **Initialize**: $\boldsymbol{\theta} = \boldsymbol{\theta'}$, replay memory library $\boldsymbol{\Gamma}$, $\varepsilon, \varepsilon_{\min}, \varepsilon_{\text{decay}}$, threshold for loop execution I, threshold for model training L_t, update threshold of the target DNN F, and $w = 0$.
4: **While** $\varepsilon > \varepsilon_{\min}$ **Do**
5: $\quad \varepsilon \leftarrow \varepsilon \times \varepsilon_{\text{decay}}$
6: $\quad$ **For** $i \leftarrow 1, 2, \ldots, I$ **Do**
7: $\quad\quad$ Generate a random value among 0 to 1, and perform Equation (10);
8: $\quad\quad$ Choose action a, obtain the reward, and store quad (s, a, r_w, s') into $\boldsymbol{\Gamma}$;
9: $\quad\quad$ **If** $i > L_t$ **Do**
10: $\quad\quad\quad w = w + 1$;
11: $\quad\quad\quad$ Randomly select a batch of experience samples from $\boldsymbol{\Gamma}$;
12: $\quad\quad\quad$ Obtain $L(\boldsymbol{\theta}, \boldsymbol{\theta'})$ through calculating Equation (13);
13: $\quad\quad$ **End If**
14: $\quad\quad$ **If** w **mod** F **= 0 Do**
15: $\quad\quad\quad w = w + 1$;
16: $\quad\quad\quad s \leftarrow s', \boldsymbol{\theta'} \leftarrow \boldsymbol{\theta}$;
17: $\quad\quad$ **End If**
18: $\quad$ **End For**
19: **End While**

The proposed DQN-PIA method in this paper is consistent with typical deep learning methods, and both are offline decision methods that require pre-model training. The characteristic of this type of method is that it relies on a large amount of sample data for model

parameter calculation during the training process, while, in testing or application, it only involves simple forward linear and nonlinear mapping processing. It can be considered that the complexity of the model after deployment is relatively low. It should be added that the method proposed in this paper significantly reduced the volume of space and reduced the complexity of model training and deployment by over 80% in the process of state setting. In order to obtain the decision model of the method proposed in this paper, the gateway device can generally be required to train or update the parameters of the model through offline or online parallel methods, which is combined with its perceptual recorded historical channel state sample data, without affecting the normal working process.

5. Simulation and Performance Analysis

5.1. Simulation Settings

In order to verify the performance of the DQN-PIA method, we used Keras [40,41] as the simulation platform for deep reinforcement learning. Without a loss of generality, we selected the first six groups of CSI data for the performance verification of the proposed method. Due to the spatial differences in CSI information of the wireless devices, these six groups of CSI data could be identified as originating from six different terminals (We believe that the difference in CSI information of the same terminal at different spatio-temporal dimensions is not higher than that of different terminals at different spatio-temporal dimensions. This means that if the proposed method has good discrimination ability for the same terminal at different spatio-temporal dimensions, it also has good discrimination ability for different terminals at different spatio-temporal dimensions). We selected the number of principal components M as six by referring to the threshold of not less than 99% for the sum of the explained variance ratio of the component. The relevant CSI data were trained and learned after the aforementioned preprocessing operations. In this paper, temporal continuous perception data was segmented, with each data size containing 500 samples and a total of 10,000. A total of 80% of the data was used as the training set, while the remaining 20% was set as the test set. The discount factor for Algorithm 1 during operation was set to 0.9, and the exploration factor for the greedy algorithm was set to $\varepsilon \in [0.9, 0.005]$, and $\varepsilon_{\min} = 0.005$ with a decay rate of 0.995. The learning rate was set to 0.01, the size of the experience playback library was set to $\mathbf{\Gamma}$ = 500, the size of the small batch capacity was set to 32, and the update frequency of the target network was set to ϕ = 500. The settings of the simulation process's related parameters are summarized in Table 1.

Table 1. Parameters in the simulation.

Hyper-Parameter	Value
Total number of CSI data entries	10,000
Number of training set entries	8000
Number of test set entries	2000
Number of sampling points per entry	500
Decay rate of ε, ε_{decay}	0.995
The minimal value of ε, $\varepsilon_{\min}$	0.005
Experience–Replay memory capacity $\mathbf{\Gamma}$	500
Number of Wi-Fi terminals	6
Update threshold of the target network F	500
Experience–Replay minibatch size	32
Discount factor ϑ	0.9
Learning rate	0.01
Number of principal components M	6
Deep reinforcement learning platform	Keras

In order to fully demonstrate the advantages of the DQN-PIA proposed in this paper in terms of terminal authentication performance, the following seven additional types of methods were used for performance comparison:

- SVM, the support vector machine method [42], which creates a predictive classification model based on the given features of the dataset elements and uses radial basis function kernels in simulation.
- MLP, the multi-layer perceptron method [24], which is a feedforward supervised neural network trained with a standard backward transmission algorithm. In the simulation, the network structure is set as two hidden layers, and the number of units is 15 and 10, respectively.
- KNN, the K-nearest neighbor method [20], which determines the type of sample by finding the K-closest samples in the feature space and observing the attribute labels of the latter. We set K = 20.
- RF, the random forest method [43], which is a classifier that includes multiple decision trees. The mode of the category output by individual classification trees determines the output category of the method. In this paper, the maximum number of features was set as 15 in the simulation.
- LR, the logistic regression method [44], involves using linear or non-linear models to associate inputs and outputs. The LR model is a classic statistical method that has advantages in practicality. In this paper, a penalty term of two norm was set for the method in simulation.
- CNN, the method proposed in [24], followed the process of this method (i.e., Figure 2 and Algorithm 1 in [24]) to process the CSI data and to authenticate the terminal.
- LSTM, the long short-term memory method [45], is an improved method of recurrent neural networks that utilizes a three-gate structure for the proper handling of historical information through reasonable planning. In the simulation, three hidden layers were set, including 15, 25, and 20 blocks, respectively.

5.2. Performance and Analysis

Firstly, we verified the convergence of the proposed method on the training set. With the same number of rounds but a different number of trainings in each round as the reference factor, the convergence of the method was compared, as shown in Figure 11. In the training and testing process, we set the rewards for correct and incorrect identity authentication to 10 and −100, respectively. Each point in the figure represents the average reward value corresponding to the number of trainings. It can be seen from the figure that, based on the training set data, the method showed a gradual convergence trend under different numbers of trainings, and the method basically reached the initial convergence state at 500 rounds in six subgraphs, thus indicating the correctness of the method parameters setting. The set parameters played a positive guiding role in the learning of the model method. In addition, we can also observe from the figure that, as the number of trainings in each round increased, the training convergence curve of the method became gradually smoother, and the convergence results of each subgraph were 9.13, 9.22, 9.35, 9.46, 9.65, and 9.82, respectively (the closer the convergence value is to 10, the more in place the model training is), which also corresponded to the objective reality. The more the model explores and utilizes, the better it can understand the rules of getting along with the environment. However, in the setting of model parameters, more trainings corresponded to a longer training delay.

For the test set of CSI data corresponding to six sensing terminals, we verified the identification performance of various methods. Figure 12 (on the last page of this paper) shows the authentication confusion matrix of the utilized methods. The horizontal axis in each subgraph represents the terminal discrimination of the method, while the vertical axis represents the true identity of the terminal. It can be observed from the eight subgraphs that the DQN-PIA method proposed in this paper achieved an accuracy rate of over 98% in the authentication of six terminals, with an overall authentication accuracy of 98.9%, which was the scheme with the highest authentication accuracy of all methods. In comparison, the long short-term memory (LSTM) method, which is good at processing time series data, showed a performance second only to the method proposed in this paper, with an overall

authentication accuracy of 97.7%. The CNN method was also satisfactory for terminal authentication performance, with an authentication accuracy only slightly lower than DQN-PIA and LSTM; this is because this method can obtain the features of the CSI matrix through convolution, which can be used for terminal identity authentication. Among the eight methods, the logistic regression method assumed that there was a certain direct nonlinear mapping relationship between the CSI data and the terminal identity. Because the assumption made was relatively simple and intuitive compared with other methods, the performance of the method was the worst, with an overall authentication accuracy of 88.5%.

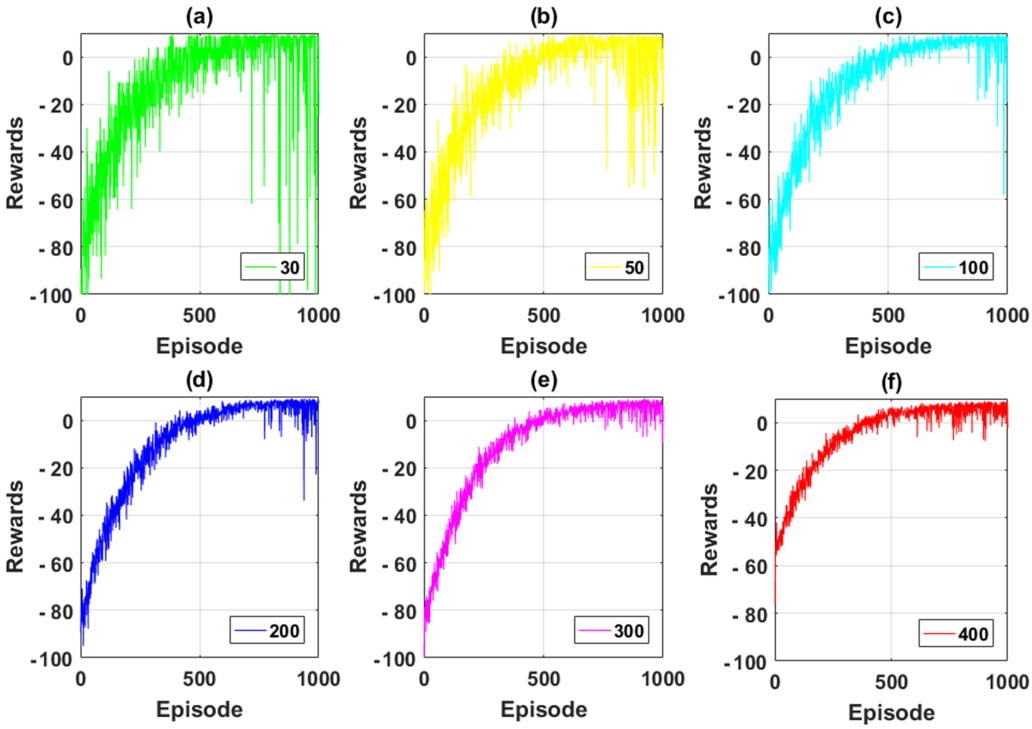

Figure 11. Convergence of the proposed DQN-PIA method under different training rounds. (**a**) Total of 30 trainings per round. (**b**) Total of 50 trainings per round. (**c**) Total of 100 trainings per round. (**d**) Total of 200 trainings per round. (**e**) Total of 300 trainings per round. (**f**) Total of 400 trainings per round.

5.3. Discussion

We can observe from Figure 12 that, except for the LR method, the overall identification accuracy of other methods was over 90%. This shows that there is indeed a clear connection between the channel state information and the identity of the terminal. It is feasible to authenticate the identity of terminals in the wireless communication system based on the physical layer. The authentication method based on deep reinforcement learning proposed in this paper showed superior authentication accuracy performance compared with other methods. Furthermore, considering that the identification model of this method can converge quickly, which indicates that the training complexity of the method is low and it can adapt to wireless channel environments with changing characteristics, we can conclude that method DQN-PIA is suitable for the scenario and has better identification ability for wireless terminals compared to commonly used methods.

For future work, from the perspective of method application, we believe that there are three steps that need to be carried out: Firstly, we need to verify the authentication efficiency and accuracy of the proposed method for a large number of terminals. Secondly, we need to verify the method's ability to authenticate and recognize different types of terminals. Finally, we need to design authentication protocols tailored to the considered

scenario in order to avoid affecting authentication efficiency and network availability when authentication errors occur in the proposed method.

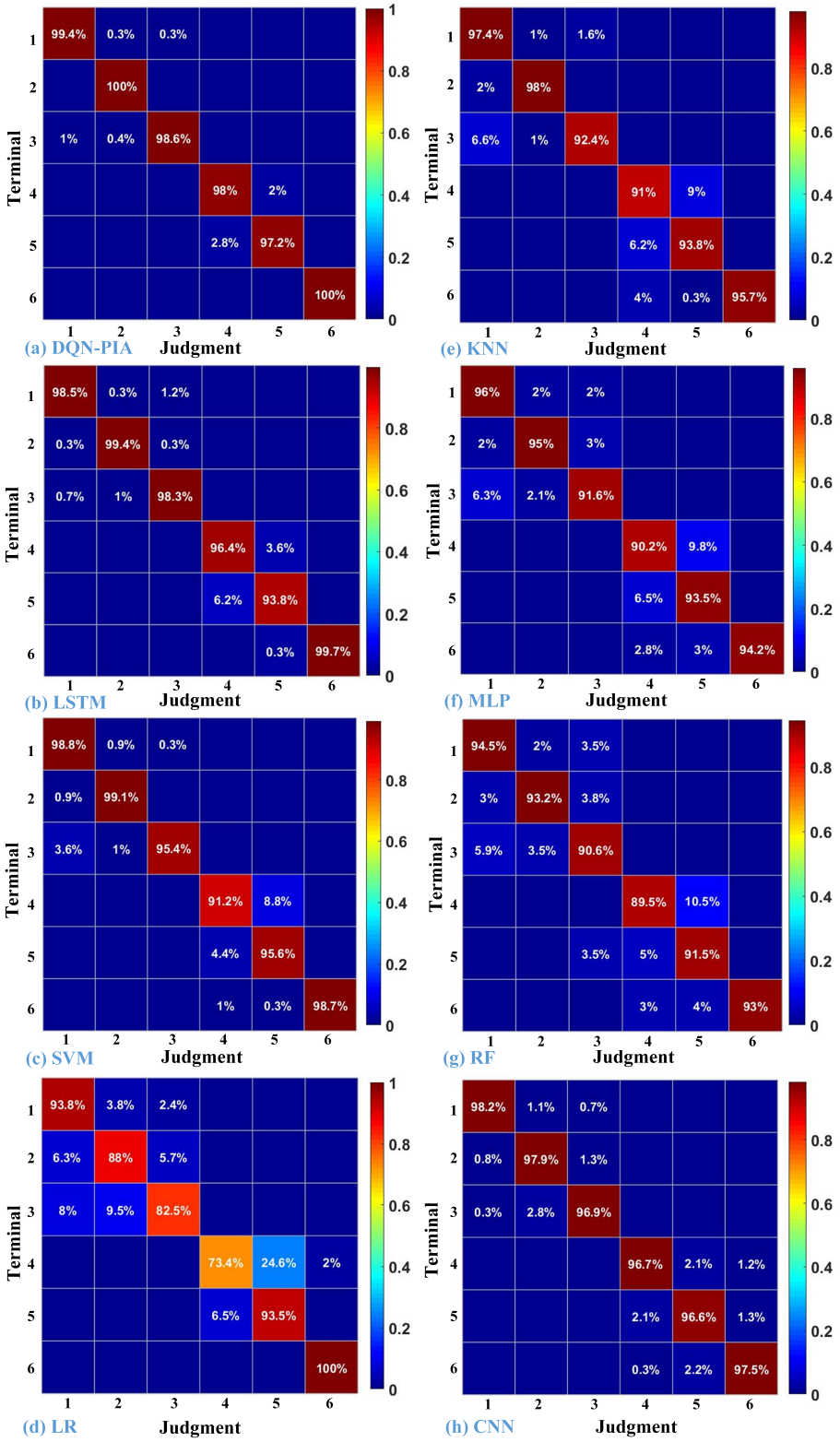

Figure 12. Identity authentication confusion matrix of each comparison method. The authentication accuracy of DQN-PIA, LSTM, SVM, LR, KNN, MLP, RF, CNN is 98.9%, 97.7%, 96.4%, 88.5%, 94.7%, 93.4%, 92.05%, and 97.3%, respectively.

6. Conclusions

In response to the security and stability risks of traditional cryptographic-based certificate identity authentication methods in IoT scenarios, such as low efficiency and large latency, we analyzed the collected CSI data in real environments, which indicated that CSI data can support identity authentication for devices in the scenario. A physical layer identity authentication method based on deep reinforcement learning was explored and proposed; this method was based on the channel state information of the wireless channel to distinguish the identity of the terminal. By designing the deep neural network with reasonable states, actions, rewards, and methods, the proposed scheme had higher authentication accuracy compared to the comparison methods and significantly lower computational requirements compared to traditional authentication methods.

Author Contributions: Conceptualization, Y.L. and P.Z.; Investigation, Y.W. and X.L.; Methodology, P.Z. and H.L.; Project administration, P.Z. and H.J.; Validation, Y.L., Y.W. and H.L.; Writing—original draft, Y.L. and P.Z.; Writing—review and editing, Y.W. and H.J. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Fundamental Research Funds for the Central Universities (no. 328202206 and 3282023034), the Beijing Natural Science Foundation (no. L192002), and in part by the "Advanced and sophisticated" discipline construction project of the universities in Beijing (no. 20210013Z0401), the China National Key R&D Program (no. 2020YF-B1808000).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

IoT	Internet of Things
DRL	deep reinforcement learning
DQN	deep Q-Network
SNR	signal-to-noise ratio
MDP	signal-to-interference-noise ratio
ResNet	residual network
MDP	Markov decision process
DNN	deep neural network
QoS	quality of service
UE	user equipment
CSI	channel state information
RSS	receive signal strength
CIR	channel impulse response
CFO	carrier frequency offset
SVM	support vector machine
NIC	network interface card
AP	access point
PCA	principal component analysis
MLP	multi-layer perceptron
KNN	K-nearest neighbor
RF	random forest
LR	logistic regression
LSTM	long short-term memory
CNN	convolutional neural networks

References

1. Tran-Dang, H.; Krommenacker, N.; Charpentier, P.; Kim, D.-S. Toward the Internet of Things for Physical Internet: Perspectives and Challenges. *IEEE Internet Things J.* **2020**, *7*, 4711–4736. [CrossRef]
2. Nguyen, D.C.; Ding, M.; Pathirana, P.N.; Seneviratne, A.; Li, J.; Niyato, D.; Dobre, O.; Poor, H.V. 6G Internet of Things: A Comprehensive Survey. *IEEE Internet Things J.* **2022**, *9*, 359–383. [CrossRef]
3. You, X.; Wang, C.X.; Huang, J.; Gao, X.; Zhang, Z.; Wang, M.; Huang, Y.; Zhang, C.; Jiang, Y.; Wang, J.; et al. Towards 6G wireless communication networks: Vision, enabling technologies, and new paradigm shifts. *Sci. China Inf. Sci.* **2021**, *64*, 1–74. [CrossRef]
4. Zhang, Z.; Xiao, Y.; Ma, Z.; Xiao, M.; Ding, Z.; Lei, X.; Karagiannidis, G.K.; Fan, P. 6G Wireless Networks: Vision, Requirements, Architecture, and Key Technologies. *IEEE Veh. Technol. Mag.* **2019**, *14*, 28–41. [CrossRef]
5. Alsharif, M.H.; Jahid, A.; Kelechi, A.H.; Kannadasan, R. Green IoT: A Review and Future Research Directions. *Symmetry* **2023**, *15*, 757. [CrossRef]
6. Cui, T.; Yang, R.; Fang, C.; Yu, S. Deep Reinforcement Learning-Based Resource Allocation for Content Distribution in IoT-Edge-Cloud Computing Environments. *Symmetry* **2023**, *15*, 217. [CrossRef]
7. Kanellopoulos, D.; Sharma, V.K. Dynamic Load Balancing Techniques in the IoT: A Review. *Symmetry* **2022**, *14*, 2554. [CrossRef]
8. Abbas, G.; Mehmood, A.; Carsten, M.; Epiphaniou, G.; Lloret, J. Safety, Security and Privacy in Machine Learning Based Internet of Things. *J. Sens. Actuator Netw.* **2022**, *11*, 38. [CrossRef]
9. De Santis, A.; Ferrara, A.L.; Flores, M.; Masucci, B. Continuous Entity Authentication in the Internet of Things Scenario. *Appl. Sci.* **2023**, *13*, 5945. [CrossRef]
10. Falayi, A.; Wang, Q.; Liao, W.; Yu, W. Survey of Distributed and Decentralized IoT Securities: Approaches Using Deep Learning and Blockchain Technology. *Future Internet* **2023**, *15*, 178. [CrossRef]
11. Chen, C.; Guo, H.; Wu, Y.; Shen, B.; Ding, M.; Liu, J. A Lightweight Authentication and Key Agreement Protocol for IoT-Enabled Smart Grid System. *Sensors* **2023**, *23*, 3991. [CrossRef] [PubMed]
12. Thapa, S.; Bello, A.; Maurushat, A.; Farid, F. Security Risks and User Perception towards Adopting Wearable Internet of Medical Things. *Int. J. Environ. Res. Public Health* **2023**, *20*, 5519. [CrossRef] [PubMed]
13. Bhushan, B.; Kumar, A.; Agarwal, A.K.; Kumar, A.; Bhattacharya, P.; Kumar, A. Towards a Secure and Sustainable Internet of Medical Things (IoMT): Requirements, Design Challenges, Security Techniques, and Future Trends. *Sustainability* **2023**, *15*, 6177. [CrossRef]
14. Budati, A.K.; Vulapula, S.R.; Shah, S.B.H.; Al-Tirawi, A.; Carie, A. Secure Multi-Level Privacy-Protection Scheme for Securing Private Data over 5G-Enabled Hybrid Cloud IoT Networks. *Electronics* **2023**, *12*, 1638. [CrossRef]
15. Zhao, J.; Hu, H.; Huang, F.; Guo, Y.; Liao, L. Authentication Technology in Internet of Things and Privacy Security Issues in Typical Application Scenarios. *Electronics* **2023**, *12*, 1812. [CrossRef]
16. Chahoushi, M.; Nabati, M.; Asvadi, R.; Ghorashi, S.A. CSI-Based Human Activity Recognition Using Multi-Input Multi-Output Autoencoder and Fine-Tuning. *Sensors* **2023**, *23*, 3591. [CrossRef]
17. Mesa-Cantillo, C.M.; Sánchez-Rodríguez, D.; Alonso-González, I.; Quintana-Suárez, M.A.; Ley-Bosch, C.; Alonso-Hernández, J.B. A Non Intrusive Human Presence Detection Methodology Based on Channel State Information of Wi-Fi Networks. *Sensors* **2023**, *23*, 500. [CrossRef]
18. Chen, B.; Song, Y.; Zhu, Z.; Gao, S.; Wang, J.; Hu, A. Authenticating Mobile Wireless Device through Per-packet Channel State Information. In Proceedings of the 2021 51st Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops (DSN-W), Taipei, Taiwan, 21–24 June 2021; pp. 78–84. [CrossRef]
19. Wang, S.; Huang, K.; Xu, X.; Zhong, Z.; Zhou, Y. CSI-Based Physical Layer Authentication via Deep Learning. *IEEE Wirel. Commun. Lett.* **2022**, *11*, 1748–1752. [CrossRef]
20. Yang, S.; Wang, Y.; Yu, X.; Gu, Y.; Ren, F. User Authentication leveraging behavioral information using Commodity WiFi devices. In Proceedings of the 2020 IEEE/CIC International Conference on Communications in China (ICCC), Chongqing, China, 9–11 August 2020; pp. 530–535. [CrossRef]
21. Yu, B.; Yang, C.; Ma, J. Continuous Authentication for the Internet of Things Using Channel State Information. In Proceedings of the 2019 IEEE Global Communications Conference (GLOBECOM), Waikoloa, HI, USA, 9–13 December 2019; pp. 1–6. [CrossRef]
22. Kong, H.; Lu, L.; Yu, J.; Chen, Y.; Tang, F. Continuous Authentication Through Finger Gesture Interaction for Smart Homes Using WiFi. *IEEE Trans. Mob. Comput.* **2021**, *20*, 3148–3162. [CrossRef]
23. Liu, H.; Wang, Y.; Liu, J.; Yang, J.; Chen, Y.; Poor, H.V. Authenticating Users Through Fine-Grained Channel Information. *IEEE Trans. Mob. Comput.* **2018**, *17*, 251–264. [CrossRef]
24. Li, X.; Huang, K.; Wang, S.; Xu, X. A physical layer authentication mechanism for IoT devices. *China Commun.* **2022**, *19*, 129–140. [CrossRef]
25. Furrer, S.; Dahlhaus, D. Multiple-Antenna Signaling Over Fading Channels with Estimated Channel State Information: Capacity Analysis. *IEEE Trans. Inf. Theory* **2007**, *53*, 2028–2043. [CrossRef]
26. Sutivong, A.; Chiang, M.; Cover, T.M.; Kim, Y.-H. Channel capacity and state estimation for state-dependent Gaussian channels. *IEEE Trans. Inf. Theory* **2005**, *51*, 1486–1495. [CrossRef]
27. Liu, J.; Elia, N.; Tatikonda, S. Capacity-Achieving Feedback Schemes for Gaussian Finite-State Markov Channels with Channel State Information. *IEEE Trans. Inf. Theory* **2015**, *61*, 3632–3650. [CrossRef]

28. Aubry, A.; Tulino, A.M.; Venkatesan, S. Multiple-access channel capacity region with incomplete channel state information. In Proceedings of the 2010 IEEE International Symposium on Information Theory, Austin, TX, USA, 13–18 June 2010; pp. 2313–2317. [CrossRef]
29. Halperin, D.; Hu, W.; Sheth, A.; Wetherall, D. Tool release: Gathering 802.11n traces with channel state information. *ACM Sigcomm Comput. Commun. Rev.* **2011**, *41*, 53. [CrossRef]
30. Zuo, P.; Sun, G.; Li, Z.; Guo, C.; Li, S.; Wei, Z. Towards Secure Transmission in Fog Internet of Things Using Intelligent Resource Allocation. *IEEE Sens. J.* **2023**, *23*, 12263–12273. [CrossRef]
31. Sutton, R.S.; Barto, A.G. *Reinforcement Learning: An Introduction*; MIT Press: Cambridge, MA, USA, 2018.
32. Zhao, X.; Yang, R.; Zhang, Y.; Yan, M.; Yue, L. Deep Reinforcement Learning for Intelligent Dual-UAV Reconnaissance Mission Planning. *Electronics* **2022**, *11*, 2031. [CrossRef]
33. Ud Din, A.F.; Mir, I.; Gul, F.; Mir, S.; Saeed, N.; Althobaiti, T.; Abbas, S.M.; Abualigah, L. Deep Reinforcement Learning for Integrated Non-Linear Control of Autonomous UAVs. *Processes* **2022**, *10*, 1307. [CrossRef]
34. Zuo, P.; Wang, C.; Wei, Z.; Li, Z.; Zhao, H.; Jiang, H. Deep Reinforcement Learning Based Load Balancing Routing for LEO Satellite Network. In Proceedings of the 2022 IEEE 95th Vehicular Technology Conference: (VTC2022-Spring), Helsinki, Finland, 19–22 June 2022; pp. 1–6.
35. Yu, Y.; Wang, T.; Liew, S.C. Deep-Reinforcement Learning Multiple Access for Heterogeneous Wireless Networks. *IEEE J. Sel. Areas Commun.* **2019**, *37*, 1277–1290. [CrossRef]
36. Lange, S.; Riedmiller, M. Deep auto-encoder neural networks in reinforcement learning. In Proceedings of the 2010 International Joint Conference on Neural Networks (IJCNN), Barcelona, Spain, 18–23 July 2010; pp. 1–8.
37. Yang, J.; Zhang, D.; Frangi, A.F.; Yang, J.Y. Two-dimensional PCA: A new approach to appearance-based face representation and recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2004**, *26*, 131–137. [CrossRef]
38. Dahl, G.E.; Sainath, T.N.; Hinton, G.E. Improving deep neural networks for LVCSR using rectified linear units and dropout. In Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 26–31 May 2013; pp. 8609–8613. [CrossRef]
39. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [CrossRef]
40. Keras: Deep Learning for Humans. Available online: https://keras.io/ (accessed on 21 April 2023).
41. Ziegler, J.L.; Arn, R.T.; Chambers, W. Modulation recognition with GNU radio, keras, and HackRF. In Proceedings of the 2017 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN), Baltimore, MD, USA, 6–9 March 2017; pp. 1–3. [CrossRef]
42. Huang, G.-B.; Zhou, H.; Ding, X.; Zhang, R. Extreme Learning Machine for Regression and Multiclass Classification. *IEEE Trans. Syst. Man Cybern. Part (Cybern.)* **2012**, *42*, 513–529. [CrossRef] [PubMed]
43. Ham, J.; Chen, Y.; Crawford, M.M.; Ghosh, J. Investigation of the random forest framework for classification of hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 492–501. [CrossRef]
44. Krishnapuram, B.; Carin, L.; Figueiredo, M.A.T.; Hartemink, A.J. Sparse multinomial logistic regression: Fast algorithms and generalization bounds. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 957–968. [CrossRef] [PubMed]
45. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]