Computers, Materials & Continua

Tech Science Press

DOI: 10.32604/cmc.2024.047193

ARTICLE

Deep Learning-Based Secure Transmission Strategy with Sensor-Transmission-Computing Linkage for Power Internet of Things

Bin Li*, Linghui Kong, Xiangyi Zhang, Bochuo Kou, Hui Yu and Bowen Liu

State Grid Beijing Urban District Power Supply Company, Beijing, 100034, China

*Corresponding Author: Bin Li. Email: boylibin@126.com

Received: 28 October 2023 Accepted: 28 December 2023 Published: 26 March 2024

ABSTRACT

The automatic collection of power grid situation information, along with real-time multimedia interaction between the front and back ends during the accident handling process, has generated a massive amount of power grid data. While wireless communication offers a convenient channel for grid terminal access and data transmission, it is important to note that the bandwidth of wireless communication is limited. Additionally, the broadcast nature of wireless transmission raises concerns about the potential for unauthorized eavesdropping during data transmission. To address these challenges and achieve reliable, secure, and real-time transmission of power grid data, an intelligent security transmission strategy with sensor-transmission-computing linkage is proposed in this paper. The primary objective of this strategy is to maximize the confidentiality capacity of the system. To tackle this, an optimization problem is formulated, taking into consideration interruption probability and interception probability as constraints. To efficiently solve this optimization problem, a low-complexity algorithm rooted in deep reinforcement learning is designed, which aims to derive a suboptimal solution for the problem at hand. Ultimately, through simulation results, the validity of the proposed strategy in guaranteed communication security, stability, and timeliness is substantiated. The results confirm that the proposed intelligent security transmission strategy significantly contributes to the safeguarding of communication integrity, system stability, and timely data delivery.

KEYWORDS

Secure transmission; deep learning; power Internet of Things; sensor-transmission-computing

1 Introduction

With the advancement of power systems, power communication plays an increasingly important role in the field of modern energy [1,2]. Power communication is not only used for data transmission and exchange of control commands but also plays a key role in the process of energy production, transmission, and distribution. However, the data transmission involved in power communication covers a large amount of critical information, involving key aspects such as power supply stability, load management, equipment status monitoring, and emergency response [3]. Therefore, the security and reliability of power communication have become the cornerstone of the stable operation of the power system.

In today's digital environment, power communication is facing increasing cyber security threats (Data leaks, malware, and denial-of-service attacks are constantly evolving). New challenges have been posed to the security and stability of power communication [4–6]. Attackers may attempt to steal sensitive data, damage critical equipment, and disrupt communication links, thereby causing severe impacts to power systems. For example, attackers may affect the operation and control of the power system by tampering with data transmission intercepting communication traffic, and even causing equipment damage and power interruption, bringing huge economic and security risks to society.

In response to these threats, ensuring the safe transmission of power communication has become a top priority [7–9]. Traditional communications security measures are no longer sufficient and a more advanced and comprehensive approach is required to address the ever-changing threat environment [10]. Based on sensor-transmission-computing, this paper proposed a secure transmission strategy for power communication. This strategy is based on the Distributed Proximal Policy Optimization (DPPO) algorithm of Deep Reinforcement Learning (DRL) and selects the relay nodes in the signal transmission process of the electric power Internet of Things (IoT), aiming at ensuring data confidentiality and integrity [8]. Based on security and availability, the system's confidentiality capacity is maximized, and the overall security and robustness of the power communication system are improved.

The chapters of this paper are arranged as follows:

Sections 1 and 2 reviewed the importance of power communication and the current security challenges, followed by an introduction to traditional communication security measures and their limitations.

Section 3 elaborated on the proposed secure transmission strategy for power communication, including its core principles, key technologies, and application scenarios.

Section 4 constructed the strategy model and problem model, and transformed them into the Markov Decision Process (MDP) process.

Sections 5 and 6 solved the problem model based on DRL, and compared and analyzed it with traditional methods through simulation experiments.

Section 7 summarized the innovations and advantages of this strategy will be highlighted, and future research directions will be envisioned.

Through the research of this thesis, we provided a new relay strategy selection method for the secure transmission of power communication. This strategy is not only of great significance to the power industry but also has certain reference significance for data transmission and network security in other key areas. It is hoped that the strategy of this paper can bring useful guidance and inspiration for further research and practice in the field of power communication security.

2 Related Works

With the digital and intelligent transformation of the power system, the secure transmission of power communication has attracted widespread attention, and many studies have been devoted to developing various strategies and methods to guarantee the integrity and availability of power communication data. In this section, we review research work in related fields, including traditional power communication security strategies, network isolation techniques, encrypted communication methods, and AI-based secure transmission strategies.

2.1 Traditional Power Communication Security Strategy

The traditional power communication system mainly relies on the closed-loop structure, and the data is transmitted at the physical level, which is relatively closed, so the security risk is relatively low [11]. However, with the application of information technology, the power system is gradually changing to an open network, which introduces new security challenges. Early traditional power communication systems usually adopted security policies based on firewalls and access control lists to protect the system from external attacks. For example, references [12] and [13] presented the implementation of a firewall-wise system intended to protect a target distribution network, which communicates with field devices and remotely controls processes from a computer screen. However, these traditional methods may become insufficient in the face of complex network environments and advanced threats.

2.2 Network Isolation Technology

To reduce the impact of external attacks on the power communication system, some studies have focused on using network isolation technology to isolate the power communication network from other networks. Technologies such as virtual private networks (VPNs) and isolation gateways are widely used to physically isolate power communication data from other data, thereby reducing the risk of external attacks. Reference [14] introduced five solutions to the sparse reward problem, including reward design and learning, experience replay, exploration and exploitation,multi-goal learning, and auxiliary tasks, to solve the power communication problem. Reference [15] proposed a novel personalized federated learning with privacy preservation (PDP-PFL) algorithm based on information fusion. Reference [16] proposed a kind of method for improving the generating unit-tripping strategy using deep reinforcement learning. However, as the complexity of power communication networks increases, a single isolation technique may not provide sufficient security.

2.3 Encrypted Communication Method

Encrypted communication is a common method of ensuring data confidentiality by encrypting transmitted data decrypt. Symmetric encryption and asymmetric encryption are two common encryption methods. Many research works have explored how these encryption methods can be applied to power communication to protect the confidentiality of sensitive data [17]. Reference [18] analyzed the current state of the digital twins' paradigm and classified potential threats related to it while considering its functional layers and operational requirements to achieve a more complete and useful classification. However, in the power communication scenario with high real-time requirements, encrypted communication may introduce a large delay, thereby affecting system performance.

2.4 Artificial Intelligence-Based Security Transmission Strategy

In recent years, artificial intelligence (AI)-based secure transmission strategies have gradually attracted attention. AI technology can analyze vast amounts of network data, detect anomalous behavior, and identify potential threats. For example, deep learning models can be used for intrusion detection and behavioral analysis, helping to discover unknown attack patterns [19]. In addition, reinforcement learning algorithms also have potential in security decision-making, adapting to changing threat environments [20]. However, the application of AI-based methods in the field of power communication is still in its infancy, requiring more research and experimental verification.

In conclusion, research in the field of power communication security transmission covers traditional security strategies, network isolation technology, encrypted communication methods, and strategies based on artificial intelligence. Although these methods have improved the security of power communication to some extent, there are still many challenges and limitations. Therefore, this paper proposes a new secure transmission strategy for power communication to address the deficiencies of existing methods and provide a more efficient and flexible solution. In the next section, we detail the rationale and design ideas of the proposed strategy.

3 System Model

3.1 Network Model

Power Internet of Things is an innovative model that applies IoT technology to power systems, aiming to realize intelligent monitoring, management, and optimization of power systems. It closely combines the IoT technology with the power system to build an intelligent power network, reliability, and sustainability of the power system [21–23]. The network model of the power is based on the core characteristics of the IoT technology, that is, the connection between devices, data exchange, and intelligent decision-making are realized through wireless sensors, communication technology, and data analysis in the power system, this means real-time monitoring of key elements such as power equipment, lines, and transformers through the deployment of sensor devices, smart meters, and data communication technologies [24–26]. With the improvement of hardware technology, the development of power IoT has entered a new stage—the edge collaboration stage of computing vertical distribution enabled by edge computing. As shown in Fig. 1, this diagram illustrates a network architecture model for data sensor, transmission, and collaborative computing. This model, commonly referred to as the sensor-transmission-computing linkage, has gained popularity in power IoT systems in recent years.

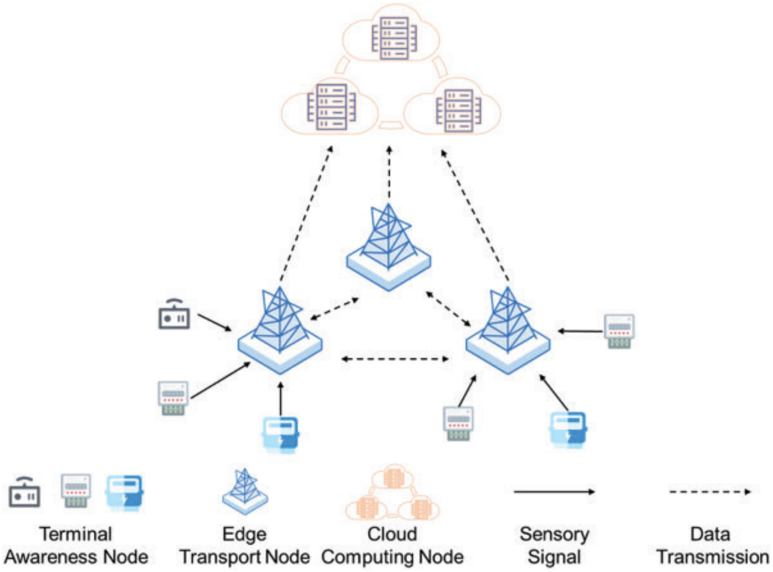

Figure 1: System model

Under this model, a complete computing task would be divided into multiple segments, distributed among end nodes, edge nodes, and cloud processing centers. The functions of various entity nodes are as follows:

1) Terminal sensor node: The terminal perception node of the power Internet of Things refers to the intelligent devices or sensors distributed in all links of the power system. End nodes can be installed

in various links such as power generation, power transmission, and power distribution to collect real-time power data, such as current, voltage, frequency, power, etc. Through data collection and a small amount of calculation, the operating status of the power system can be understood in real-time and fed back to adjacent edge nodes, which helps to quickly find problems and abnormalities.

2) Edge transmission node: The edge transmission node plays the role of data collection, preliminary processing, and real-time communication in the power Internet of Things. By analyzing and responding to the data locally, it provides real-time data support for operators and managers, helping them make faster and more accurate decisions, and improving the real-time efficiency of power systems.

3) Cloud computing node: This is the core of the power IoT, and it is a platform covering functions such as data storage, processing, analysis, and management. At this layer, data is aggregated, stored, processed, and analyzed for intelligent decision-making and control.

The introduction of the power Internet of Things network model will bring new intelligence and automation capabilities to the power system, which will help improve the reliability, availability, and maintainability of the power system. However, data privacy and network security issues are also worth noting to guarantee system stability. The power IoT network model will bring revolutionary changes to the power industry and pave the way for the development of future smart power systems.

3.2 Relay Process Model

Based on the sensor-transmission-computing linkage network architecture, this paper proposes an intelligent security transmission strategy based on DRL. To maximize the system confidentiality capacity, the confidentiality rate and calculation rate of information during wireless transmission are maximized.

As shown in Fig. 2, the power wireless network system model can be composed of a source node S representing sending data, a destination node D representing receiving data, multiple trusted relay nodes $\{node_i | 1 \leq i \leq N\}$, and multiple eavesdropping nodes composed of $\{e_i | 1 \leq i \leq E\}$, trying to steal data from relay nodes.

In a network containing multiple relay nodes, a cooperative transmission scheme can be used to improve the secrecy capacity of the system.

The cooperative transmission scheme is a communication strategy that improves the performance of the communication system by coordinating multiple transmission nodes (such as terminals, relays, servers, etc.). This cooperative transmission can play a role in many aspects, including increasing transmission rate, reducing transmission delay, increasing network capacity, reducing energy consumption, etc. Common cooperative transmission schemes include:

1) Multi-hop transmission: Multiple relay nodes are set up in the network, and data can be transmitted through multiple relay nodes in a hop-like manner, thereby expanding the transmission range and coverage.

2) Network coding: Using network coding technology, the terminal can encode data during transmission so that the receiving node can reconstruct the original data through decoding. This is especially useful in combating channel errors and packet loss.

3) Multi-path transmission: The data is transmitted to multiple paths at the same time, and the receiving end can receive data from different paths to increase transmission reliability and reduce the impact of channel attenuation.

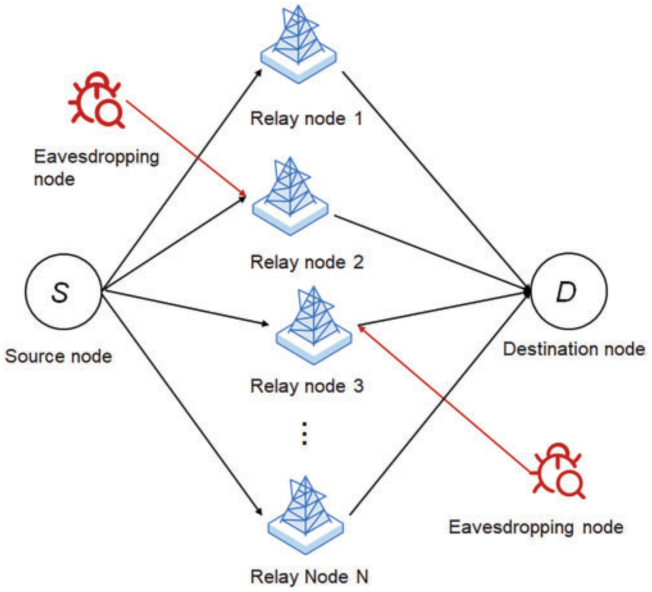

Figure 2: Relay process model

4) Joint transmission: Multiple terminals or nodes can jointly transmit data in time and frequency to improve spectrum efficiency and system capacity.

The collaborative transmission scheme can be customized according to specific application scenarios and communication requirements. By exploiting the collaboration among different nodes, the system efficiency and reliability can be improved. These solutions are usually applied in different communication standards, such as wireless networks, IoT, etc.

The transmission strategy proposed in this paper aims to use the cooperative work of relay nodes to enhance the confidentiality of communication links through reasonable cooperation. The following is a description of the cooperative transmission scheme proposed in this paper.

Assume that there are multiple relay nodes, namely $N > 1$, the information transmission is carried out in $\{T_i | (1 < i \leq N + 1)\}$ time slots. In T_1 time slot, the source node S transmits a signal s, the power of s is P_s. Therefore, the signals received at the destination node D, the i-th relay node $node_i$, and the i-th eavesdropping node e_i are, respectively:

$$y_D(s) = \sqrt{P_s} h_{SD} s + z_D(s) \tag{1}$$

$$y_{node_i}(s) = \sqrt{P_s} h_{Snode_i} s + z_{node_i}(s) \tag{2}$$

$$y_{e_k}(s) = \sqrt{P_s} h_{Se_k} s + z_{e_k}(s) \tag{3}$$

Among them, h_{SD} represents the complex channel gain from S to D, h_{Snode_i} represents complex channel gain from S to $node_i$, h_{Se_k} represents complex channel gain from S to the k-th eavesdropping node e_k. The relay node $node_i$ performs an XOR encoding operation on the received signal and its observation results to obtain the encoded data $y'_{node_i}(s)$. z_D, z_{node_i} and z_{e_k} are additive white Gaussian noises with received variance σ^2.

In the next time slot, the communication between S and D is realized through the correct cooperation of the relay node to encode signal s. Assume that the trusted relay node of $T \leq N$ is selected from the relay nodes of N, and continuously sends data to D in a time slot T. Therefore, in the a-th time slot, the signals received by D, $node_i$, e_k are, respectively:

$$y_D\left(a\right) = \sqrt{P_a} h_{node_a D} s + z_D\left(a\right) \tag{4}$$

$$y_{node_i}\left(a\right) = \sqrt{P_a} h_{node_a node_i} s + z_{node_i}\left(a\right) \tag{5}$$

$$y_{e_k}\left(a\right) = \sqrt{P_a} h_{node_a e_k} s + z_{e_k}\left(a\right) \tag{6}$$

Among them, $a = 1, 2, \ldots, N + 1$, $i = a + 1, a + 2, \ldots, N + 1$, P_a are transmission power of the a-th trusted relay node, $h_{node_a D}$ represents the complex channel gain of the link between the a-th trusted relay node and D, $h_{node_a node_i}$ represents the complex channel gain of the link between a-th trusted relay node and the i-th trusted relay node, $h_{node_a e_k}$ represents complex channel gain of the link between the a-th trusted relay node and the k-th eavesdropping node. Similarly, $node_i$ performs the XOR coding operation on the received signal and its observation results to obtain the coded data $y'_{node_i}\left(a\right)$.

Assume that the transmit power of the relay node is P_{Total}/T, where P_{Total} is the total transmit power of all trusted relay nodes. All relay nodes use maximum ratio combining technology to process the received signals and perform XOR coding operations. According to Shannon's theorem, in a time slot $a + 1$, the signal rates received by the destination node D, $node_i$, e_k are, respectively:

$$\vartheta_D = \frac{1}{T+1} \mathrm{lb}\left(1 + \alpha_{S,D} P_s + \alpha_{node,D}\left(P_{Total}/T\right)\right) \tag{7}$$

$$\vartheta_{node_i} = \begin{cases} \mathrm{lb}\left(1 + \alpha_{S,node_i} P_s\right) \\ \dfrac{1}{i} \mathrm{lb}\left(1 + \alpha_{S,node_i} P_s + \alpha_{node_i}\left(P_{Total}/T\right)\right) \end{cases} \tag{8}$$

$$\vartheta_{e_k} = \frac{1}{T+1} \mathrm{lb}\left(1 + \alpha_{S,e_k} P_s + \alpha_{node,e_k}\left(P_{Total}/T\right)\right) \tag{9}$$

where

$$\alpha_{S,node_i} = \left|h_{S,node_i}\right|^2 / \delta^2 \tag{10}$$

$$\alpha_{node,D} = \sum_{i=1}^{T} \left|h_{node_i,D}\right|^2 / \delta^2 \tag{11}$$

$$\alpha_{node_i} = \sum_{t=1}^{i-1} \left|h_{node_t,node_i}\right|^2 / \delta^2 \tag{12}$$

$$\alpha_{node,e_k} = \sum_{i=1}^{T} \left|h_{node_i,e_k}\right|^2 / \delta^2 \tag{13}$$

Express the SNR ratio of the received signal at the destination node D, $node_i$, e_k. The above formula expresses the maximum transmission rate of the channel, that is, the signal rate. In the secure communication of the power IoT, if encryption technology is introduced to protect the data, the channel capacity can also be used to calculate the confidentiality capacity, that is the maximum number of confidential bits that can be transmitted.

The calculation of confidentiality requires attention to channel limitations, as well as the impact of encryption and decryption. In secure communications, part of the channel capacity may be used to transmit keys, while another part is used to transmit the actual secret information. Therefore, the secrecy capacity is expressed as:

$$C_D = \vartheta_D - w_D \tag{14}$$

$$C_{node_i} = \vartheta_{node_i} - w_{node_i} \tag{15}$$

$$C_{e_k} = \vartheta_{e_k} - w_{e_k} \tag{16}$$

Among them, w_D is the payload channel capacity of the destination node D, w_{node_i} is the payload channel capacity of $node_i$, w_{e_k} represents the payload channel capacity of e_k. Therefore, in the proposed transmission strategy, the system's achievable secrecy capacity is:

$$C = max\left\{C_D - \sum_k C_{e_k}, 0\right\} \tag{17}$$

To ensure the feasibility of the above transmission strategy, it is necessary to ensure that the relay node can correctly encode and decode the signal from the source node, the outage probability constraint, which can be expressed as:

$$C_{node_i} \geq C_{\min} \tag{18}$$

where $C_{\min}$ is the minimum load channel capacity to ensure successful encoding and decoding.

In addition, it is necessary to ensure that the eavesdropping node cannot intercept the relay signal, otherwise, the confidentiality of the signal reaching the destination node D cannot be guaranteed, that is, the interception probability constraint needs to be satisfied, which can be expressed as:

$$C_{e_k} \leq C_{max} \tag{19}$$

Among them, C_{max} is the maximum load channel capacity to ensure that the eavesdropping node cannot intercept the relay signal.

In the case of Nrelay nodes, when $node_i$ relay nodes are activated, $(node_i, N), 1 \leq i \leq N$ can be considered, and the possible relay selection strategy is:

$$\sum_{i=1}^{N} (node_i, N) = \sum_{i=1}^{N} \frac{N!}{(N-i)!} = \left(2^N - 1\right) \tag{20}$$

Therefore, the goal of this paper is to find a relay selection strategy under all relay selection strategies, which can maximize the secrecy capacity under the constraints of outage probability and interception probability.

4 Markov Decision Process Analysis of Relay Strategy Selection

4.1 Action

In the problem scenario described in this paper, the number of relay nodes is N, so the Markov Decision Process (MDP) action in reinforcement learning is defined as:

$$A = \{a_1, a_2, .., a_n\}, a_i \in \{0, 1\} \tag{21}$$

$a_i = 1$ indicates that $node_i$ is selected as the node for relay forwarding.

4.2 State

Since the selection of the relay strategy is only related to the signal status received by the current destination node, relay node, and eavesdropping node, the MDP state can be expressed as:

$$S = \left\{y_D, y_{node_i}, y_{e_k}\right\}, i \in [1, n], k \in [1, K] \tag{22}$$

In the scenario of this paper, the next state after the state transition is not deterministic, because the target node of the next relay is not known. Therefore, the state transition probability is related to the action and the current state, namely:

$$P_{ss}^{a} = P\left[S_{t+1} = s' \mid S_t = s, A_t = a\right] \tag{23}$$

4.3 Reward

The reward represents the environment's evaluation of the last action feedback. In the scenario of this paper, optimization aims to maximize the confidentiality capacity while balancing the constraints of outage probability and interception probability. Therefore, the reward for the model is:

$$R = C = max\left\{C_D - \sum_k C_{e_k}, 0\right\} \tag{24}$$

5 Relay Strategy Selection Algorithm Based on DRL

5.1 DPPO

The proximal policy optimization (PPO) algorithm uses the same neural network structure as the actor-critic. PPO is based on the actor-critic model, which uses the ratio of the updated new strategy to the old strategy to update, limiting the magnitude of the strategy update makes the whole algorithm more stable.

The DPPO algorithm is based on the PPO algorithm, adding multiple threads and increasing the learning rate. The DPPO algorithm uses the advantage function $A_\pi(s, a)$, which represents the advantage of choosing an action a in state s, which has a similar effect to the Q-value of Q-learning.

The difference is that $A_\pi(s, a)$ evaluates the quality of each action taken in a certain state relative to the average return, that is, the advantage of taking this action, while the Q-value evaluates the quality of each action taken in a certain state. $A_\pi(s, a)$ can be expressed as:

$$A_\pi(s, a_t) = R_\pi(s, a) - V_\pi(s) \tag{25}$$

$$R_\pi(s, a) = r(s, a) + \gamma \sum_{s \in S} P_{ss'} V_\pi(s') \tag{26}$$

$$\begin{aligned} V_\pi(s) &= \sum_{a \in A} \pi(a \mid s)\left[r(s, a) + \gamma \sum_{s \in S} P_{ss'} V_\pi(s')\right] \\ &= \sum_{a \in A} \pi(a \mid s) R_\pi(s, a) \end{aligned} \tag{27}$$

$$\eta(\pi) = \eta(\pi) + E_\pi\left[\sum_{t=0}^{\infty} A_\pi(s, a)\right] \tag{28}$$

$$J = max_{\theta} E_{\pi} \left[\sum_{t=0}^{\infty} A_{\pi} (s, a) \right] \tag{29}$$

The above formula represents the expected value of the cumulative reward report of the new strategy relative to the old strategy as the network parameters are continuously updated. The above formula represents an objective function. In practice, to have higher robustness, the method of selecting a clip proxy target instead of the update method of the above formula can be expressed as Table 1.

$$J(\theta) = E_t \left(\min \left\{ r_t(\theta) \hat{A}_t, \text{clip} \left[r_t(\theta), 1 - \varepsilon, 1 + \varepsilon \right] A_t \right\} \right) \tag{30}$$

$$J = max_{\theta} E_{\pi} \left[\sum_{t=0}^{\infty} A_{\pi} (s, a) \right] \tag{31}$$

where $r_t(\theta)$ is the probability ratio of the old and new strategies, $(1 - \varepsilon, 1 + \varepsilon)$ is the boundary range of the clip, and $\hat{A}_t$ is the estimated value of the advantage function for n time steps.

Table 1: DPPO-based relay strategy selection algorithm

```
Algorithm 1 DPPO-Based Relay Strategy Selection Algorithm
Input: Network Initial State
/*Global Agent*/
Repeat $T = 1, 2, \ldots, N$
    Repeat $t = 1, 2, \ldots, N$
        According to strategy to get $s_t, a_t$
        According to formula (24) to get $r_t$
        Approximate calculation, get $R_{\pi}(s_t, a_t)$, $v_{\pi}(s_t)$, $A_{\pi}(s_t, a_t)$
    End repeat
    Input the obtained data to the agent of DPPO.
    Update the old strategy to the existing strategy.
    Optimize agent loss, update strategy and critic network parameters.
End repeat
/*Sub Agent*/
Repeat $T = 1, 2, \ldots, N$
    Repeat $t = 1, 2, \ldots, N$
        According to strategy to get $s_t, a_t$
        According to formula (24) to get $r_t$
        Approximate calculation, get $R_{\pi}(s_t, a_t)$, $v_{\pi}(s_t)$, $A_{\pi}(s_t, a_t)$.
        Input the obtained data to the agent of PPO.
        Update the old strategy to the existing strategy.
    End repeat
    Repeat $m = \{1, 2, \ldots, U\}$
        Calculate $J(\theta) = E_t \left( \min \left\{ r_t(\theta) \hat{A}_t, \text{clip} \left[ r_t(\theta), 1 - \varepsilon, 1 + \varepsilon \right] A_t \right\} \right)$
```

(Continued)

Table 1 (continued)

```
            Pass the consolidated information to the general agent of DPPO.
            Update actor network parameters.
            Update critic network parameters.
            Wait until all threads' agents have finished running.
        End repeat
        The agent of DPPO calculates the gradient.
        Update strategy network parameters for the agent of DPPO.
End repeat
Output: The Selection Strategy of Relay Nodes
```

5.2 DPPO-Based Relay Policy Selection Algorithm

Using the DPPO algorithm to solve the relay strategy selection scheme (RSSA) optimization problem, the algorithm flow is shown in Table 1.

6 Simulation and Analysis

For benchmarking, we choose to compare the 2-hop transmission scheme with RSSA. In addition, to verify the effectiveness of DPPO in solving the problem of relay strategy selection, this paper simulates the simulation results of CNN and Q-learning and compares them.

Assume that the number of source nodes and destination nodes is 1, the number of relay nodes is 5–20, the number of eavesdropping nodes is 1–5, the positions of eavesdropping nodes are random, and all channels have the same Independently distributed Rayleigh fading, set the total transmit power $P_s = 50\,dBm$, noise power $\sigma^2 = -20\,dBm$, and path loss exponent to 3.5.

In this paper, the number of threads in the DPPO algorithm is set to 4, the reward discount factor is 0.95, and the policy update step is 200, that is, when the agent performs 200 trainings, the network parameters are updated once. In the reference CNN model, the learning rate is 0.001, and the parameters such as step size are consistent with DPPO. In the reference Q-learning model, the initial Q-value is 1, and the remaining parameters are consistent with DPPO.

Figs. 3 and 4 are the convergence of different reinforcement learning algorithms with the training process under the same system environment. It can be observed that the loss of the DPPO algorithm gradually decreases as the training progresses, indicating that model performance is gradually improved, the error of the model in prediction is reduced, and the accuracy rate is gradually improved, which means correct classification ratio of the model in the sample is constantly increasing. The model fits the training data better. However, the loss curve of the Q-learning algorithm is relatively high in the early stage of training, the model performance is poor, and more training is needed to reduce the error. The performance of CCN algorithm data is the worst. In the early stage of training, the loss drops rapidly and the accuracy rate increases, but after a certain level, the loss hardly drops, and may even start to rise. At the same time, the accuracy rate tends to stabilize or even decline. Through the above data analysis, we can see that the DPPO algorithm can solve the problem of relay strategy selection, and it is better than other learning algorithms.

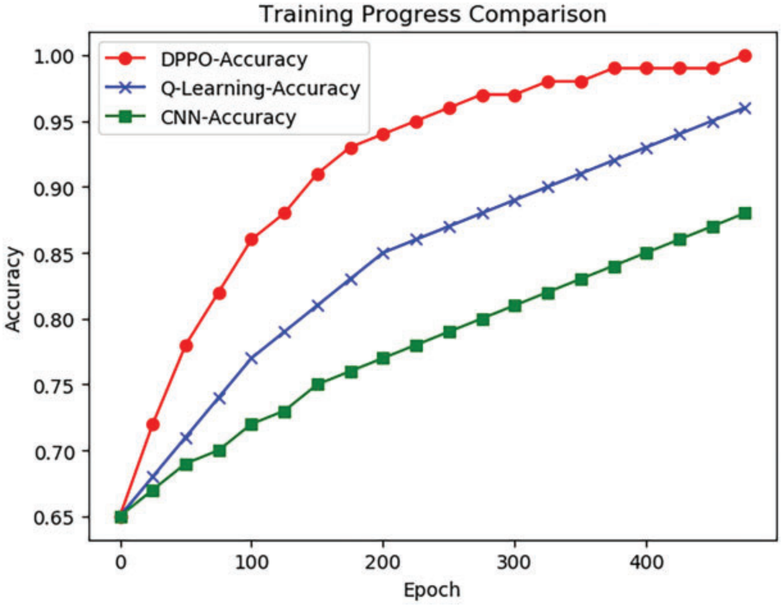

Figure 3: Training process comparison 1

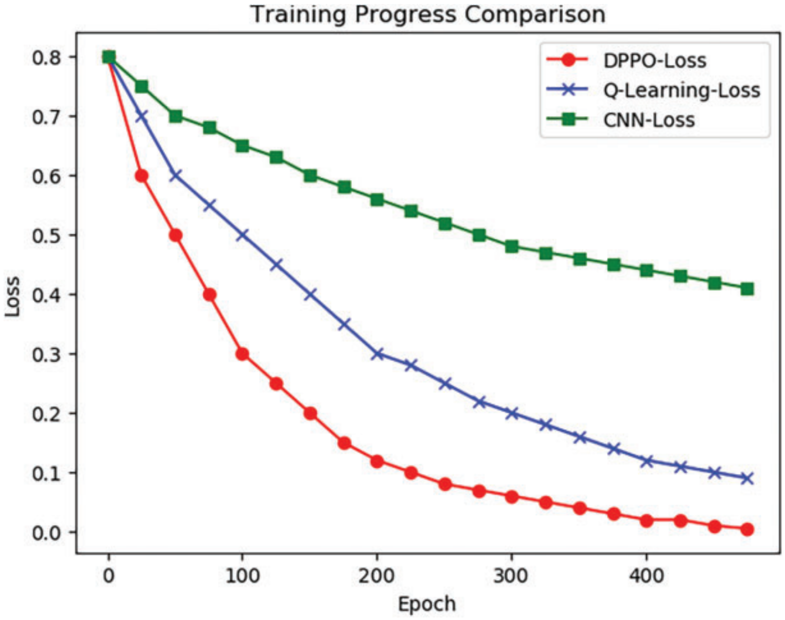

Figure 4: Training process comparison 2

Figs. 5 and 6 show how the secret capacity varies with the number of relay nodes. From the simulation figure, we can see that 2-hop transmission has the lowest system secrecy capacity because as the number of relay nodes increases, each relay node has a potential security risk. This is especially important for data transfers that contain sensitive information, as intermediate nodes could be the target of attacks or data breaches. In the RSSA proposed in this paper, the optimal relay node can be selected to forward the information, and the data is transmitted on multiple paths instead of being concentrated on a single path, which can reduce the attacker's ability to intercept or interfere with the entire network. Possibility of data transfer. The DPPO algorithm is also significantly better than CNN and Q-learning in solving the problem of relay strategy selection. This is because in a loaded network environment, the dimensions of the system's action space and state space are very high, and the spatial dimension is very important for CNN. Compared with Q-learning, it is very fatal. DPPO can still maintain an efficient and accurate solution speed in a high-latitude space, to obtain a more accurate, optimal transmission path to maximize the system's confidentiality capacity.

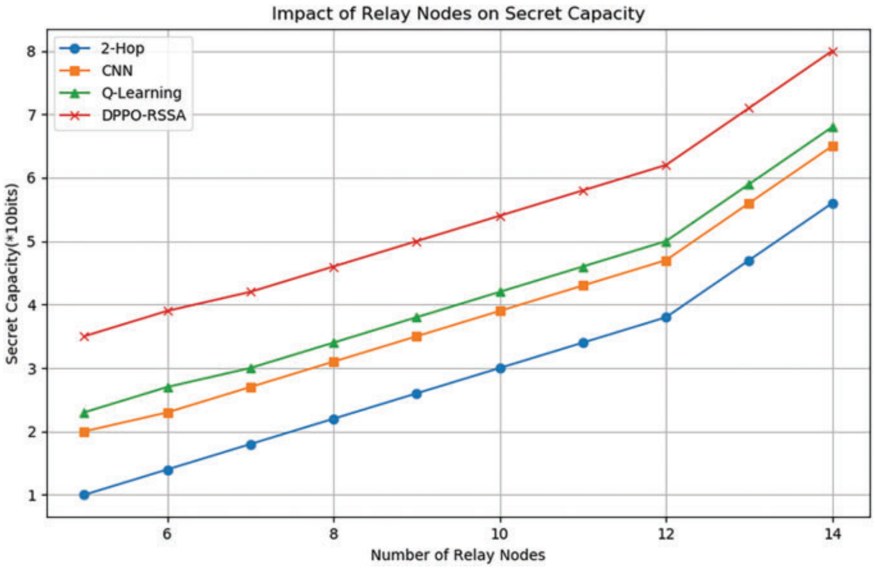

Figure 5: Number of relay nodes-secret capacity

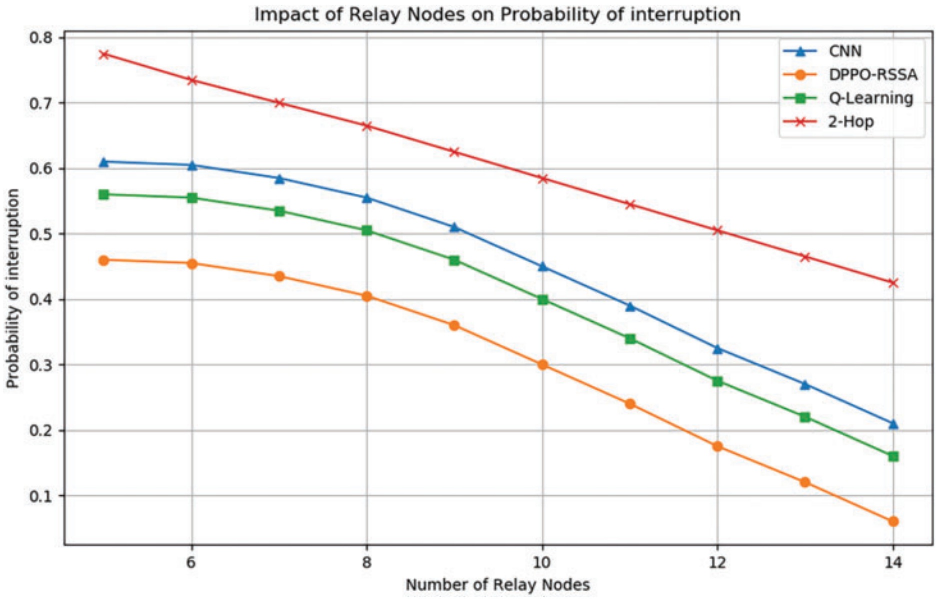

Figure 6: Number of relay nodes-outage rate

Figs. 7 and 8 show how the secret capacity and outage rate vary with the number of eavesdropping nodes. From the simulation figure, we can see that the 2-hop transmission has the lowest system secrecy capacity and the highest interruption rate because as the number of eavesdropping nodes increases, each relay node in the 2-hop transmission mode is A link where if one of the nodes fails or is interrupted, the entire data transmission may fail or be affected. In this case, the RSSA transmission proposed in this paper has a lower transmission failure rate than the 2-hop transmission. Similarly, the advantages of the DPPO algorithm in dealing with high-latitude space problems can also be reflected in the above figure, that is, in the same network environment, applying the DPPO algorithm to solve the problem of relay strategy selection can obtain a better solution.

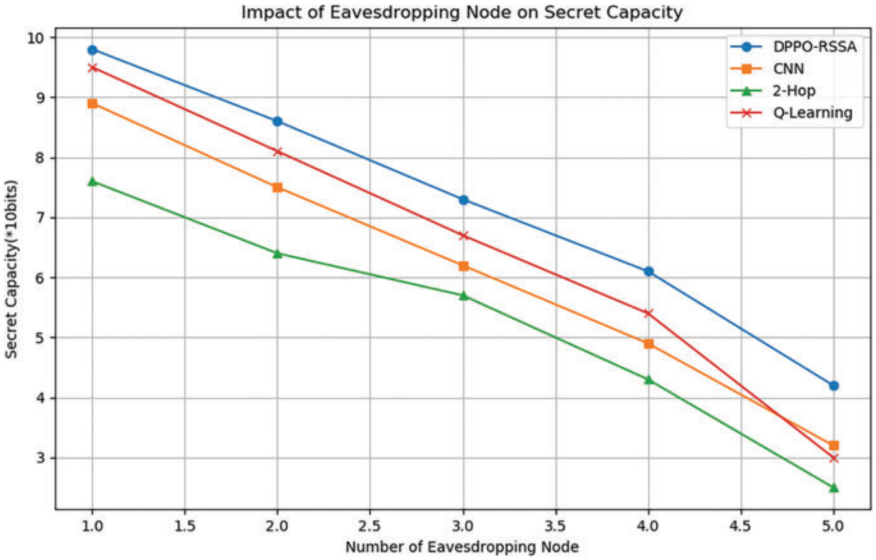

Figure 7: Number of relay nodes-secret capacity

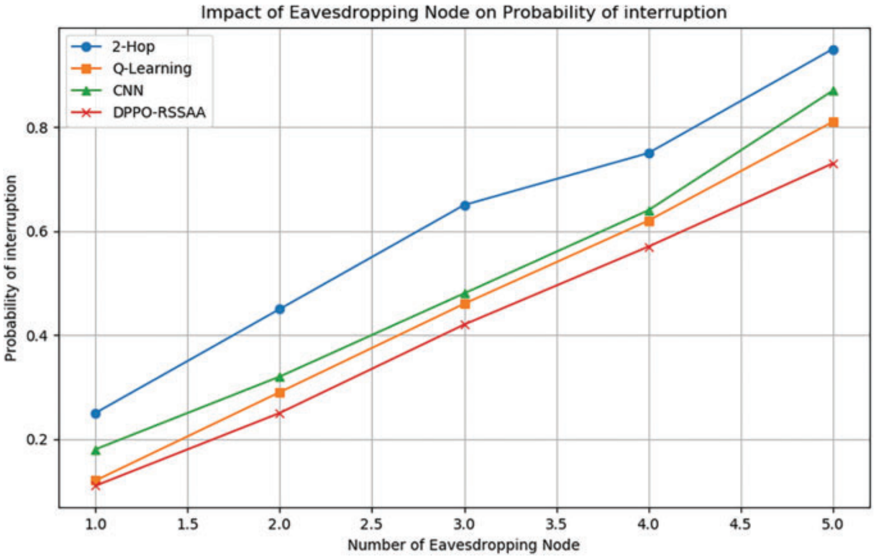

Figure 8: Number of eavesdropping nodes-outage rate

7 Conclusion

The automatic collection of power grid status information, coupled with real-time multimedia interactions between the front and back ends during accident management, has resulted in a substantial volume of power grid data. While broadcast nature of wireless transmission raises apprehensions regarding the potential for unauthorized eavesdropping during data transmission. To address these issues and ensure reliable, secure, and real-time transmission of power grid data, this paper introduces an intelligent security transmission strategy that incorporates a sensor-transmission-computing linkage. The primary goal of this strategy is to maximize the system's confidentiality capacity. To achieve this objective, an optimization problem is formulated, with interruption probability and interception probability as constraints. To efficiently address this optimization problem, a low-complexity algorithm based on DPPO is developed, aiming to provide a suboptimal solution. Through simulation results, the paper validates the effectiveness of the proposed intelligent security transmission strategy in enhancing communication security, system stability, and timely data delivery. These findings underscore the strategy's valuable contributions to maintaining communication integrity, ensuring system stability, and achieving timely data transmission. However, the strategy's performance may also be influenced by evolving cybersecurity threats and wireless communication technologies, which can change over time, these issues need to be further improved in future research work.

Acknowledgement: Thanks to State Grid Beijing Urban District Power Supply Company for supporting this paper.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: Study conception and design: Bin Li, Linghui Kong; data collection: Bin Li; analysis and interpretation of results: Bin Li, Linghui Kong, Xiangyi Zhang, Bochuo Kou; draft manuscript preparation: Hui Yu, Bowen Liu. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: Due to the nature of this research, participants of this study did not agree for their data to be shared publicly, so supporting data is not available.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] P. R. Desai, S. Mini, and D. K. Tosh, "Edge-based optimal routing in SDN-enabled industrial Internet of Things," *IEEE Internet Things J.*, vol. 9, no. 191, pp. 18898–18907, 2022.

[2] Q. Qi, X. M. Chen, C. J. Zhong, and Z. Y. Zhang, "Integrated sensing, computation and communication in B5G cellular Internet of Things," *IEEE Trans. Wirel. Commun.*, vol. 20, no. 1, pp. 332–344, 2021.

[3] X. Li *et al.*, "Processing-while-transmitting: Cost-minimized transmission in SDN-based STINs," *IEEE/ACM Trans. Netw.*, vol. 30, no. 1, pp. 243–256, 2022.

[4] H. S. Xu, J. Wu, J. H. Li, and X. Lin, "Deep-reinforcement-learning-based cybertwin architecture for 6G IIoT: An integrated design of control, communication, and computing," *IEEE Internet Things J.*, vol. 8, no. 22, pp. 16337–16348, 2021.

[5] N. Chen, T. Qiu, L. P. Zhao, X. B. Zhou, and H. S. Ning, "Edge intelligent networking optimization for Internet of things in smart city," *IEEE Wirel. Commun.*, vol. 28, no. 2, pp. 26–31, 2021.

[6] L. Sun, J. Liang, C. Zhang, D. Wu, and Y. Zhang, "Meta-transfer metric learning for time series classification in 6G-supported intelligent transportation systems," *IEEE Trans. Intell. Transp. Syst.*, pp. 1–11, 2023. doi: 10.1109/TITS.2023.3250962.

[7] Y. C. Yi, J. Cai, and Z. Su, "A multi-user mobile computation offloading and transmission scheduling mechanism for delay-sensitive applications," *IEEE Trans. Mob. Comput.*, vol. 19, no. 1, pp. 29–43, 2020.

[8] L. J. He, J. D. Li, M. Sheng, R. Z. Liu, K. Guo and D. Zhou, "Dynamic scheduling of hybrid tasks with time windows in data relay satellite networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 5, pp. 4989–5004, 2019.

[9] L. Sun, J. Wu, Y. Xu, and Y. C. Zhang, "A federated learning and blockchain framework for physiological signal classification based on continual learning," *Inf. Sci.*, vol. 630, pp. 586–598, 2023.

[10] D. Zhou, M. Sheng, J. D. Li, and Z. Han, "Aerospace integrated networks innovation for empowering 6G: A survey and future challenges," *IEEE Commun. Surv. Tutor.*, vol. 25, no. 2, pp. 975–1019, 2023.

[11] Y. Shen and Q. Liu, "Proximal policy optimization based on self-directed action selection," *Comput. Sci.*, vol. 48, no. 12, pp. 297–303, 2021.

[12] H. Eslava, L. A. Rojas, and D. Pineda, "An algorithm for optimal firewall placement in IEC61850 substations," *J. Power Energy Eng.* , vol. 3, no. 4, pp. 16–22, 2015.

[13] K. Spiteri, R. Urgaonkar, and R. K. Sitaraman, "BOLA: Near–optimal bitrate adaptation for online videos," *IEEE Trans. Netw.*, vol. 28, no. 4, pp. 1698–1711, 2020.

[14] W. Y. Yang, C. J. Bai, C. Cai, Y. N. Zhao, and P. Liu, "Survey on sparse reward in deep reinforcement learning," *Comput. Sci.*, vol. 47, no. 3, pp. 182–191, 2020.

[15] Q. Zhiguo, T. Yang, M. Ghulam, and T. Prayag, "Privacy protection in intelligent vehicle networking: A novel federated learning algorithm based on information fusion," *Inf. Fusion.*, vol. 98, no. 101824, pp. 1–12, 2023.

[16] W. Liu, D. Zhang, X. Wang, J. Hou, and L. Liu, "A decision making strategy for generating unit tripping under emergency circumstances based on deep reinforcement learning," in *Proc. CSEE*, vol. 38, no. 1, pp. 109–119, 2018.

[17] P. Mach and Z. Becvar, "Device-to-device relaying: Optimization, performance perspectives, and open challenges towards 6G networks," *IEEE Commun. Surv. Tutor.*, vol. 24, no. 3, pp. 1336–1393, 2022.

[18] Q. Zhiguo, L. Yang, and T. Prayag, "QNMF: A quantum neural network based multimodal fusion system for intelligent diagnosis," *Inf. Fusion.*, vol. 100, no. 101913, pp. 1–13, 2023.

[19] S. Bi and Y. J. Zhang, "Computation rate maximization for wireless powered mobile-edge computing with binary computation offloading," *IEEE Trans. Wirel. Commun.*, vol. 17, no. 6, pp. 7457–7469, 2020.

[20] H. Ke, J. Wang, L. Deng, Y. Ge, and H. Wang, "Deep reinforcement learning-based adaptive computation offloading for MEC in heterogeneous vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 7, pp. 7916–7929, 2020.

[21] Z. G. Qu, X. Z. Liu, and M. Zheng, "Temporal-Spatial quantum graph convolutional neural network based on Schrödinger approach for traffic congestion prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 8, pp. 8677–8686, 2023.

[22] A. E. Mostafa and V. W. S. Wong, "Transmit or backscatter: Communication mode selection for narrowband IoT systems," *IEEE Trans. Veh. Technol.*, vol. 71, no. 5, pp. 5477–5491, 2022.

[23] P. Cruz, N. Achir, and A. C. Viana, "On the edge of the deployment: A survey on multi-access edge computing," *ACM Comput. Surv.*, vol. 55, no. 5, pp. 1–34, 2022.

[24] L. Sun, M. Zhang, B. Wang, and P. Tiwari, "Few-shot class-incremental learning for medical time series classification," *IEEE J. Biomed. Health Inform.*, pp. 1–11, 2023. doi: 10.1109/JBHI.2023.3247861.

[25] C. Chen, R. Guo, W. Zhang, J. Yang, and C. K. Yeo, "Optimal sequential relay-remote selection and computation offloading in mobile edge computing," *J. Supercomput.*, vol. 78, no. 1, pp. 1093–1116, 2022.

[26] S. Bi, L. Huang, H. Wang, and Y. J. A. Zhang, "Lyapunov-guided deep reinforcement learning for stable online computation offloading in mobile-edge computing networks," *IEEE Trans. Wirel. Commun.*, vol. 20, no. 11, pp. 7519–7537, 2021.