An Optimal Selection Algorithm Based on Q-Learning For Dual Media Communication System With Full-Duplex Relays

Huifang Wang¹, Shujuan Zhang¹, Zhixiong Chen²*, Meiru Huo¹, Lei Fan¹ and Lixia Zhang¹

¹ Information and Communication Branch, State Grid Shanxi Electric Power Company, Taiyuan, Shanxi, 030021, China
² School of Electrical and Electronic Engineering, North China Electric Power University, Baoding, Hebei, 071003, China
*Corresponding author’s e-mail: zxchen@ncepu.edu.cn

Abstract. Full-duplex communication technology can make full use of spectrum resources through simultaneous co-frequency transmission, but the interference of the signal sent by the device itself degrades performance. The paper proposes a joint antenna and relay selection algorithm based on Q-learning, which selects the best relay with the best transmit-receive configuration for the communication between the source and the target. Specifically, by learning the dual-medium cooperative communication environment to obtain feedback in terms of interruption probability and channel capacity, so as to determine the best relay and antenna. Finally, the paper verifies the effectiveness of the algorithm by simulation.

1. Introduction
Power Line Communication (PLC) and wireless communication technology have broad application prospects in the fields of smart electricity and the Internet of Things. The wireless communication network access is flexible and the networking is simple. However, the wireless communication is easily affected by distance, obstacles (such as walls), etc., and the signal fading is relatively large[1]. Literature [2] investigated a hybrid communication with both power line and wireless communication, and exerts their respective communication advantages to make signals better transmitted in the communication process.

With the development of the Internet of Things, the number of parameters that need to be optimized in the network has increased sharply, and the computational overhead required by traditional optimization algorithms is too large. In response to this development trend, artificial intelligence technology can be applied to the communication field to solve the problems encountered [3-5]. Among them, Q-learning complexity is relatively low, and less channel feedback information is required. Literature [6] studied the power control problem based on game theory in two relay modes of FD and HD. Aiming at energy efficiency and data rate maximization, the Q-learning algorithm is used to solve the best Nash equilibrium point. The above research shows that Q-learning can be effectively used in cooperative communication systems.

In the existing communication system, regardless of the frequency division duplex (FDD) mode and the time division duplex (TDD) mode, the uplink and downlink communication is realized by occupying different resources, avoiding the communication between the uplink and the downlink. Interference, which obviously wastes half of the resources. Therefore, the use of full-duplex
technology and the same channel to achieve simultaneous two-way transmission has become one of the important means of mining spectrum resources in the subsequent evolution of communications[7]. In reference to the problem of relay selection in the design of dual-media systems, the literature [8] proposed and implemented a joint relay and antenna selection scheme, and selected the best relay with the best transmitter-receiver configuration to perform the communication between the source and the target. The communication increases the spatial diversity between two communication nodes and reduces the influence of self-interference in full-duplex communication. However, this paper only uses wireless communication channels and does not consider the large-signal fading. Therefore, the paper studies the problem of relay selection in full-duplex hybrid communication, and considers the use of machine learning to solve this problem.

This paper is organized as follows. In Section 2, we describe the Full-duplex system combining power line and wireless and propose the channel selection problem. In Section 3, we propose a joint antenna and relay selection algorithm based on Q-Learning Numerical results and analyses are presented in Section 4. And in section 5, we conclude the paper.

2. Full-duplex System model

This research describes the design of a practical full-duplex system combining power line and wireless. In this system, the full-duplex communication is realized, requiring communication to be carried out in both directions within a given time frame. In order to improve the spatial diversity and error performance of the full-duplex (FD) relay network, this topic proposes a joint relay and antenna selection method. In this system, information can only be transmitted between the source and the target with the help of the relay R. The relay consists of a power line relay and a wireless relay equipped with two antennas. When transmitting a signal, according to the instantaneous channel state information of each relay, select the power line relay or select one antenna in the wireless relay as the transmitter and the other as the receiver.

![Figure 1. Full-duplex system model.](image)

In order to better represent the system and perform mathematical analysis on the proposed system design, we consider the system model as shown in the figure. Considering that the relay R is selected according to the above method, it is used to forward the signal between the source and the target. When the relay works in full-duplex mode, the relay R receives the source signal and the signal sent by itself at the same time, so the power line and wireless signal received by the relay at each location are shown in (1) and (2):

\[ y_R^P = h_{SR}^P | h_{RR}^P | x_S^P + h_{RR}^P | x_R^P + n_R^P \]  
\[ y_R^W = h_{SR}^W | h_{RR}^W | x_S^W + h_{RR}^W | x_R^W + n_R^W \]  

Where, \( P_S^k \) and \( P_R^k \) represent the transmission power of S and R, \( x_S^P \) and \( x_R^P \) represent transmission signals of S and R in the power line, \( x_S^W \) and \( x_R^W \) represent transmission signals of S and R in the wireless channel. The power line channel fading \( h_{ij}^P \) satisfies the lognormal distribution, and the wireless channel fading \( h_{ij}^W \) satisfies the Nakagami distribution. The
and represent the self-interference parts of power line and wireless relay respectively. In addition, and are noise items in power lines and wireless.

Under the DF protocol, the relay R performs hard decision decoding on the received signal and then sends the signal to the node D. The signals at D are

\[ y_D^P = H_{RD} \sqrt{P_R} x_R^P + n_D^P \]
\[ y_D^W = H_{RD} \sqrt{P_R} x_R^W + n_D^W \]

where and represent the noise items at D.

According to equations (1) and (2), the instantaneous signal-to-noise ratio of the power line and wireless S-R link can be obtained as

\[ \eta_{SR}^P = \frac{P_R |h_{SR}^P|^2}{P_e + N_R} \]
\[ \eta_{SR}^W = \frac{P_e |h_{SR}^W|^2}{P_e + N_R} \]

where and represent the noise powers at the relay.

According to equations (3) and (4), the instantaneous signal-to-noise ratio of the power line and wireless channel R-D link are

\[ \eta_{RD}^P = \frac{P_e |h_{RD}^P|^2}{N_D^P} \]
\[ \eta_{RD}^W = \frac{P_e |h_{RD}^W|^2}{N_D^W} \]

When the system adopts the DF protocol, the amount of mutual information is determined by the link with the smaller amount of mutual information in the two links of S-R and R-D. Therefore, the instantaneous amount of mutual information based on the wireless and the power line channels of the full-duplex relay system are:

\[ I_{FD}^W = \log (1 + \min(\eta_{SR}^W, \eta_{RD}^W)) \]
\[ I_{FD}^P = \log (1 + \min(\eta_{SR}^P, \eta_{RD}^P)) \]

3. Joint Antenna and Relay Selection Algorithm Based on Q-Learning

Assume that each wireless relay node contains two antennas. In order to improve the system performance, the Q-learning algorithm is used to select the relay with the best performance. This optimization goal can be expressed as:

\[ c = \text{argmax} \{ I_{FD}^P, I_{FD,A_1 \rightarrow A_2}^W, I_{FD,A_2 \rightarrow A_1}^W \} \]

where is the amount of mutual information of the system when the power line is used as a relay to forward information, and and represent the two antennas of the wireless relay. is the amount of mutual information of the transmission system when receives a signal from the source node and transmits it to the target node of as a transmitter. Similarly, is the channel capacity when acts as a receiver on the relay node and forwards information to in order to retransmit it to the destination. Only the relay that can maximize the system capacity can communicate between the source and the destination.

According to the proposed system model, the Q-learning algorithm is used to select the best transmitter-receiver structure for end-to-end communication. The system model improves the traditional antenna selection scheme and adds additional degrees of freedom.

Q-learning is an algorithm in reinforcement learning that has nothing to do with the environment model. Each state-action pair in Q-learning corresponds to a Q value Q(s, a), and Q(s, a) is stored in state s The cumulative return value of action a when taken. The Q value is defined as:

\[ Q(s, a) = (1 - \alpha)Q(s, a) + \alpha [R(s, a) + \gamma \max \limits_{a'} Q(s', a')] \]
where $\gamma_{\text{max}} Q(s', a')$ represents the best-estimated value of the next state $s'$; $\alpha \in [0,1]$ is the learning rate, which controls the continuous time. The obtained Q value; $\gamma \in [0,1]$ is a discount factor used to indicate the importance of long-term feedback relative to immediate feedback. In the learning process, the relay selects a behaviour to execute, that is, selects a suitable relay to send and receive signals, and obtains rewards based on the behaviour. The reward is:

$$R(t) = I(a_t) - I(a_{t-1})$$  \hspace{1cm} (13)

where $I(a_t)$ represents the amount of mutual information obtained by the relay performing the current action $a_t$, and $I(a_{t-1})$ represents the amount of mutual information obtained by the relay before performing the action $a_{t-1}$. After receiving the reward $R(a_t)$, update the Q value until the Q table converges.

The Q-learning algorithm process is as follows:
1. Establish a Q value table, initialize each state-action, the corresponding Q value $Q(s, a) = 0$.
2. Judge whether the Q value table converges. If it converges, the algorithm is completed and exit the program; if it does not converge, proceed as follows:
   a) Observe the current state $s_t$;
   b) Select action $a_t$ according to the current strategy;
   c) Perform action $a_t$ and get an immediate return;
   d) Observe the new state $s_{t+1}$ and update the Q value table.

Q-learning accumulates historical experience through the continuous perception of the environment. After continuous trial and error and continuous reinforcement, the learning subject can choose the best action goal. The Q-learning algorithm selects the behaviour with the largest Q value, that is, selects the receiving-transmitting device that maximizes the amount of mutual information to transmit signals.

### 4. Simulation Results and Analysis

Unless otherwise specified, the parameters in the system model are set as follows: (1) The system transmit power of the source node and the relay node is 1; (2) $SNR$ represents average signal-to-noise ratio of the channel; (3) Bernoulli-Gaussian impulse noise parameters: $p=0.1$, $k=0.02$; (4) Power line channel fading parameter $\sigma_4=4$dB, wireless channel fading parameter $m_4 = 1.5$, $m_8 = 2$; (5) Q-learning parameters: learning rate $\alpha=0.5$, discount factor $\gamma=0.9$.

In the case of changing various parameters, for the dual-media communication full-duplex relay system, the performance of the joint relay and antenna selection algorithm based on Q-learning is analysed.
Figure 3 compares and analyses the channel capacity obtained by the communication system using different algorithms to select relays and antennas. It can be seen that with the increase of the signal-to-noise ratio, the performance of the wireless channel exceeds that of the power line channel. In the case of a fixed signal-to-noise ratio, as the fading of the power channel increases, the amount of mutual information of the power line channel is lower than that of the wireless channel. However, using Q-learning joint relay and antenna selection, the channel capacity is always the largest.

Figure 4 compares the relationship between outage probability and $S_{NR}$ of different relay selection algorithms. It can be seen that in terms of outage probability, when using Q-learning algorithm for joint relay and antenna selection, the system performance is always better than using only power line relay or wireless relay, which is the same as the conclusion in the literature, indicating that Q-learning The algorithm does not need to perform complex calculations, and effectively performs relay selection.
In order to verify the effectiveness of the Q-learning algorithm in the research of the relay selection algorithm, Figure 5 compares the channel capacity obtained by selecting the best relay through the Q-learning algorithm under different power allocations for power lines and wireless. T represents the power ratio between the power line and the wireless channel. It can be seen from Figure 5 that there are different optimal power allocation ratios under different signal-to-noise ratio conditions. Subsequent power allocation can be combined for relay selection.

5. Conclusion
For power line and wireless parallel full-duplex communication systems, the paper proposes a joint relay and antenna selection algorithm based on Q-learning, which can select the best relay and transmitter and receiver configurations according to the instantaneous state of the channel to improve the performance from the source to the target. It allows wireless communication and power communication with limited access to play their respective advantages in communication, so that signals can be transmitted better in the communication process. The paper verifies the comparison between the proposed system and the traditional system through simulation, and analyses the results. Monte Carlo simulation is used to verify the effectiveness of the Q-learning algorithm with lower complexity and subsequent power allocation can be combined for relay selection.

Acknowledgments
This paper was funded by the Science and Technology Project of State Grid Shanxi Electric Power Company (contract number: SGSXXT00JFS2100106).

References
[1] Dib,L.,Fernandes,V.,Filomeno,M.,&Ribeiro,M.V. (2017) Hybrid plc/wireless communication for smart grids and internet of things applications.IEEE Internet of Things Journal.,5:655-667
[2] CHEN Zhixiong, JING Yifang, HAN Dongsheng (2019) Performance analysis of dual-media cooperative communication based on wireless and power line under hybrid fading. International Journal of Distributed Sensor Networks.,15:1-11.
[3] HU Zhengwei, HE Dongmei, XIE Zhiyuan (2019) OFDM signal separation method for power line communication based on FastICA algorithm. Electric Power Automation Equipment.,39: 212-217.
[4] Kaur, J.,Khan,M.A.,Iftikhar,M.,Imran, M.and Emad Ul Haq,Q. (2021) Machine Learning Techniques for 5G and Beyond. IEEE Access.,9: 23472-23488.
[5] Tonello,A.M.,Letizia,N.A.,Righini,D.andMarcuzzi, F. (2019) Machine Learning Tips and Tricks for Power Line Communications.IEEE Access.,7: 82434-82452.
[6] LIANG Yingchang, TAN Junjie, Dusit, & Niyato. (2020) An overview of the research of intelligent wireless communication technology. Journal of Communications, 7: 1-17.

[7] Shams, F., Bacci, G., Luise, M. (2015) Energy-Efficient Power Control for Multiple-Relay Cooperative Networks Using Q-learning. IEEE Transactions on Wireless Communications, 14:1567-1580.

[8] Yang, K., Cui, H., Song, L., & Li, Y. (2014) Joint relay and antenna selection for full-duplex AF relay networks. IEEE International Conference on Communications, ICC 2014: (4454–4459)

[9] CHEN Zhixiong, ZENG Honghai, HAN Dongsheng (2021) Full-duplex and Falf-duplex adaptive power line relay and power optimal allocation algorithm. Electric Power, 54:199-207.