

Article

Adaptive Capacity Optimization Algorithm Leveraging Joint PHY-MAC Layer Modeling for Dual-Mode Communication Systems

Yuerong Zhao ^{1,*}, Bo Jiang ¹ and Zhixiong Chen ²¹ Yuxi Power Supply Bureau, Yunnan Power Grid Co., Ltd., Yuxi 653100, China; jiangbo20@163.com² Xi'an Actionpower Electric Co., Ltd., Xian 710117, China; chenzx1983@sohu.com

* Correspondence: 18313409984@163.com

Abstract

The extensive deployment of the Power Internet of Things (PIoT) relies on dual-mode communication (HPLC + HRF) for robust data acquisition. However, under massive bursty traffic, conventional static MAC superframe scheduling struggles to reconcile high throughput with stringent reliability constraints. To mitigate this, we propose a dynamic adaptive scheduling scheme. Initially, a joint PHY-MAC layer dual-mode system architecture is proposed. At the MAC layer, a dual-link parallel multiplexing contention access mechanism is applied; at the physical layer, a capacity bottleneck determination model is established, incorporating log-normal–Bernoulli–Gaussian mixed noise and multipath fading. Subsequently, an extended two-dimensional Markov chain analytically derives key performance indicators, including equivalent collision probability, joint outage probability, access delay, and network throughput. Building upon this, a Q-learning-based algorithm is proposed. By constructing an asymmetric penalty–reward function, the central coordinator (CCO) autonomously optimizes the Contention Access Period (CAP) to Contention-Free Period (CFP) ratio under dynamic node scales. Simulations demonstrate this methodology effectively averts channel congestion during extreme concurrent traffic surges. Ultimately, it strictly preserves service reliability while substantially augmenting the concurrent carrying capacity and resource utilization of the dual-mode network.

Keywords: dual-mode communication; concurrent traffic model; parallel multiplexing; superframe structure; Q-learning



Academic Editor: Andreas Mauthe

Received: 17 June 2026

Revised: 18 July 2026

Accepted: 28 July 2026

Published: 5 August 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

With the accelerated advancement of energy digital transformation, the low-voltage distribution station area serves as the “last mile” of the Power Internet of Things (PIoT), where the real-time performance and reliability of data acquisition directly determine the implementation effectiveness of core services including smart meter metering, 15 min load curve acquisition, active fault early warning, and demand side response (DSR) [1,2]. Traditional local communication in low-voltage station areas mostly adopts a single Power Line Communication (PLC) or Radio Frequency (RF) technology, both of which have inherent limitations. To compensate for the shortcomings of single technologies, dual-mode communication integrating High-speed Power Line Communication (HPLC) and High-speed Radio Frequency (HRF) has become the mainstream development direction of local communication in station areas, which can significantly improve communication reliability

and coverage capability through dynamic channel selection and resource multiplexing [3]. However, the time-varying nature of channels in complex station areas, the traffic burstiness caused by multi-service concurrency, and the strict hard constraint of “no less than 95% primary acquisition success rate” in power systems still pose severe challenges to the performance analysis and optimal design of dual-mode communication networks.

In terms of channel characteristics and outage performance analysis, the reliability of the underlying physical link is the fundamental prerequisite that determines the success rate of upper-layer data acquisition. Reference [4] discussed the system architecture and Internet of Things (IoT) applications of hybrid PLC and wireless communication in smart grids, and pointed out that dual-mode physical layer fusion can effectively overcome the fading bottleneck of a single physical medium. Reference [5] analyzed the diversity reception performance of hybrid narrowband PLC and wireless communication, and proved that joint signal processing under composite fading channels can significantly reduce the system outage probability. To further improve spectrum efficiency, Reference [6] derived the closed-form solution for the outage performance of a Non-Orthogonal Multiple Access (NOMA)-based dual-hop hybrid PLC-wireless network under log-normal and Rayleigh hybrid fading channels, which provides theoretical support for the reliability evaluation of the dual-medium physical layer. However, most of these studies focus on the single-layer theoretical analysis of physical layer link connectivity, and rarely conduct joint PHY-MAC layer dimensional analysis that links the underlying physical outage probability to the extremely strict data acquisition success rate (i.e., the 95% success rate red line) in practical services.

In terms of the Medium Access Control (MAC) layer access mechanism and time resource scheduling, designing efficient multiple access protocols and superframe structures is the key to addressing the concurrent burst of massive nodes. Reference [7] proposed a classic two-dimensional Discrete-Time Markov Chain (DTMC) model, which laid the theoretical foundation for accurately evaluating the collision probability and saturated throughput of contention-based access for distributed nodes. For complex heterogeneous dual-mode scenarios, Reference [8] constructed an analytical model for end-to-end access delay of a hybrid PLC and RF network using martingale theory, and quantified the time constraint boundary of dual-mode concurrent access. Reference [9] designed a MAC protocol for dual-mode communication based on adaptive channel selection, which reduces data collisions through real-time channel state sensing. To balance the random access of burst traffic and the collision-free transmission of deterministic long messages, superframe structures combining the Contention Access Period (CAP) and Contention-Free Period (CFP) have been widely adopted. For the time resource allocation within the superframe, Reference [10] designed a superframe control and data buffer optimization scheme for energy harvesting sensor networks to maximize the system throughput, while Reference [11] proposed a priority-based dynamic superframe dedicated time slot allocation strategy. Although the introduction of “multiplexing access” and “dynamic superframe” effectively expands the concurrent capacity of the network, when facing the extremely complex and tidal-fluctuating concurrent traffic in the station area power grid, traditional heuristic or fixed-rule-based superframe ratio adjustment can hardly balance the optimization accuracy and millisecond-level timeliness. This will easily lead to severe CAP congestion under heavy load or massive idle CFP time slots, thus triggering the “barrel effect” of system performance.

To break through the decision-making bottleneck of the above traditional heuristic algorithms in complex time-varying environments, intelligent optimization technologies represented by Reinforcement Learning (RL) have been gradually introduced into the resource scheduling of the MAC layer. Reference [12] provided a comprehensive review

of the applications of deep reinforcement learning in communication networks, and highlighted the advantages of agents in sequential decision-making in unknown and highly dynamic environments. In terms of the intelligentization of specific MAC protocols, researchers have carried out a large number of fruitful attempts. For traditional Carrier Sense Multiple Access (CSMA) protocols, Reference [13] proposed a Q-learning-based parameter tuning method, which significantly reduces the channel collision rate by dynamically and adaptively adjusting the contention probability of nodes. In the dual-mode communication scenario of low-voltage distribution station areas, pure contention-based protocols are prone to drastic performance collapse induced by severe channel collisions under high-concurrency conditions with hundreds of nodes. In contrast, pure scheduling-based protocols are constrained by their inherent access latency, which makes them unable to meet the stringent millisecond-level real-time requirements of burst alarm services. Accordingly, an integrated design of the two mechanisms is imperative to strike a balance between the system's high-concurrency carrying capacity and low-latency access performance. To address the limitation that traditional MAC protocols struggle to adapt to dynamic traffic fluctuations, Reference [14] proposed a Q-learning-based parameter adaptive adjustment scheme. This study extracted features from traffic patterns using reinforcement learning, and enabled the agent to autonomously optimize the length of the Contention Access Period (CAP), completely eliminating the dependence on fixed parameters and significantly improving network robustness under high-contention environments (corresponding to the scenario of dynamic adaptive contention period in the MAC layer). Aiming at the performance bottleneck of conventional Time Slotted Channel Hopping (TSCH) in resource-constrained environments, Reference [15] presented an intelligent scheduling model based on Multi-Agent Reinforcement Learning (MARL) and Double Deep Q-Network (DDQN). By enabling agent collaboration and introducing packet priority awareness, this research realized autonomous joint decision-making for time slot allocation, overcame the shortcomings of heuristic rules, and demonstrated superior adaptability in balancing end-to-end delay and power consumption (corresponding to the scenario of multi-agent collaborative time slot scheduling). To solve the problem of resource mismatch prone to occur in Industrial Internet of Things (IIoT) under high-concurrency scenarios, Reference [16] developed a dynamic resource allocation mechanism based on Deep Reinforcement Learning (DRL). This work leveraged deep neural networks to extract high-dimensional service features and performed autonomous joint scheduling of communication resources, breaking the shackles of static manual models and achieving optimal decision-making performance in complex high-concurrency environments (corresponding to the scenario of intelligent scheduling under resource-constrained high-concurrency conditions). The main contributions of this paper are summarized as follows:

First, targeting the characteristics of harsh communication channels and multi-service concurrent bursts in complex station areas, a joint PHY-MAC layer analysis model of the CSMA/CA-TDMA hybrid superframe for dual-interface parallel multiplexing is established, taking into account the heterogeneous physical characteristics including log-normal fading with hybrid impulse noise for power lines and Rayleigh fading for wireless links.

Second, to accurately evaluate the reliability and real-time performance of the dual-mode communication network, the closed-form expressions of the dual-mode joint outage probability, steady-state collision probability, and limit throughput are derived based on the proposed joint PHY-MAC layer model, using the law of total probability and the extended two-dimensional DTMC theory, with outage probability, system throughput, and single-node access delay adopted as the evaluation metrics. Meanwhile, the impacts of the concurrent node scale and initial contention window parameters on the network limit capacity and access delay are comparatively analyzed.

Third, an adaptive scheduling algorithm for the superframe time slot ratio based on reinforcement learning is proposed. An asymmetric penalty–reward function is defined, which takes into account both the hard constraint of “no less than 95% success rate for 15 min acquisition” and the throughput maximization objective. The Q-learning algorithm is adopted to realize the joint optimization of the time resource parameters of CAP and CFP under dynamic burst traffic. In addition, a closed-loop simulation platform is developed, and the results verify the effectiveness of the proposed algorithm in avoiding concurrent congestion and guaranteeing high-reliability data acquisition.

2. System Model

To address the topological uncertainty in complex station areas and the demand for high-frequency access of massive smart meters, this study constructs a star-structured single-hop dual-mode communication system architecture model tailored for complex station areas. As shown in Figure 1, with the station area Central Coordinator (CCO) as the core node, the model is directly connected to a number of downstream dual-mode smart meter stations (STAs), where all nodes are equipped with both PLC and RF communication interfaces. The system operation complies with the timing logic of the single-hop superframe, and realizes service decoupling through phase division in the time domain.

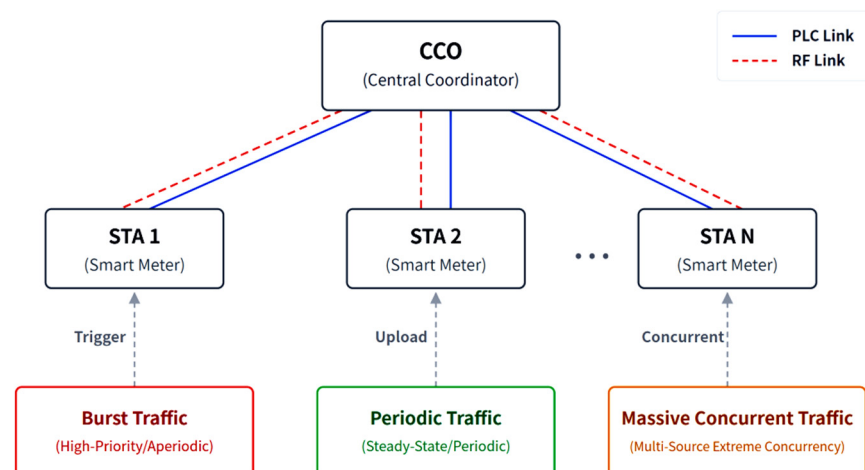


Figure 1. System model.

2.1. Physical Layer Model

In the dual-mode communication system for complex station areas, the fading and noise characteristics of the two links are significantly distinct. The amplitude gain of the PLC channel follows a log-normal distribution. Its noise model incorporates not only background thermal noise, but also the impact of high-amplitude, short-duration impulse noise, and the total noise distribution conforms to the Bernoulli–Gaussian (BG) model based on mixed Gaussian distribution [17]. Let h_{plc} denote the amplitude gain of the PLC channel. According to the definition of the log-normal distribution, its logarithmic form follows a normal distribution with mean μ_{plc} and variance σ_{plc}^2 . Its Probability Density Function (PDF) is expressed as:

$$f_{h_{plc}}(x) = \frac{1}{\sqrt{2\pi}\sigma_{plc}x} \exp\left(-\frac{(\ln x - \mu_{plc})^2}{2\sigma_{plc}^2}\right), x > 0 \quad (1)$$

Noise is the primary bottleneck limiting the performance of PLC systems. In addition to thermal noise (background noise), impulse noise induced by switching operations, motor startup and other grid events is characterized by high amplitude and short duration. The

BG model governs the occurrence of Gaussian pulses via a Bernoulli process. Let n_k denote the noise sample at the k -th time instant:

$$n_k = \omega_k + b_k i_k \quad (2)$$

where $\omega_k \sim N(0, N_G)$ is the persistently existing background Gaussian noise, $b_k \sim \text{Bernoulli}(p)$ is the pulse occurrence indicator, p denotes the probability of pulse occurrence (sparsity), $i_k \sim N(0, N_I')$ is the impulse noise component with power N_I' , and the total noise power in the pulse presence state is $N_I = N_G + N_I'$. Generally, the impulse noise power ratio is defined as $R = N_I/N_G$.

The probability density function (PDF) of the amplitude of the total noise n follows a mixed Gaussian distribution:

$$f_n(n) = \frac{1-p}{\sqrt{2\pi N_G}} e^{-\frac{n^2}{2N_G}} + \frac{p}{\sqrt{2\pi N_I}} e^{-\frac{n^2}{2N_I}} \quad (3)$$

For the wireless channel in the station area environment, smart meters are usually installed in metal meter boxes or on building walls, which brings severe Non-Line-of-Sight (NLOS) propagation and multipath scattering for wireless signals. Let h_{rf} denote the amplitude envelope of the wireless channel, which follows a Rayleigh distribution with parameter Ω . The instantaneous Signal-to-Noise Ratio (SNR) of the wireless channel γ_{rf} follows an exponential distribution, and its probability density function (PDF) is given by:

$$f_{\gamma_{rf}}(\gamma) = \frac{1}{\bar{\gamma}_{rf}} \exp(-\frac{\gamma}{\bar{\gamma}_{rf}}), \gamma \geq 0 \quad (4)$$

where $\bar{\gamma}_{rf}$ is the average SNR of the wireless channel, which is mainly determined by transmit power, antenna gain and path loss. Unlike the PLC channel, interference in the micro-power wireless communication frequency band is mainly manifested as thermal noise and adjacent channel interference, which is typically modeled as Additive White Gaussian Noise (AWGN) with a mean of 0 and variance σ_{rf}^2 .

2.2. MAC Layer Model

To meet the requirements for high-reliability acquisition of 15 min load data across the entire station area and the access of burst alarm services, the MAC layer adopts a single-hop CSMA/CA-TDMA hybrid superframe structure [18]. As shown in Figure 2, based on requirements such as traffic priority, devices are categorized into high-priority and low-priority groups. Furthermore, these devices can concurrently execute the backoff procedure across both channels. The superframe period is strictly divided into three main phases:

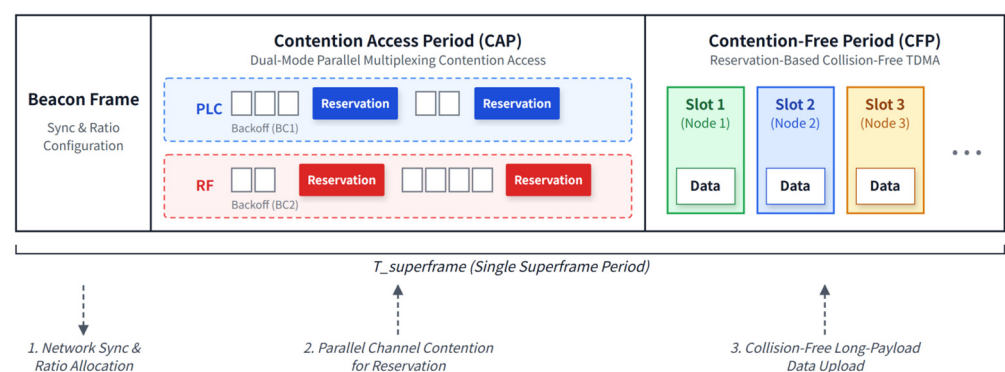


Figure 2. Schematic Diagram of the Superframe Structure.

- (1) **Beacon Phase:** The CCO broadcasts beacon frames, which are used for time synchronization of all nodes in the whole network as well as the division and configuration of the superframe structure, i.e., defining the duration of the CAP and CFP periods.
- (2) **Contention Access Period (CAP):** A dual-interface-based multiplexed CSMA/CA protocol is adopted in this phase. When a station (STA) has data packets to transmit, it initiates backoff countdown simultaneously on both the PLC channel and the Radio Frequency (RF) channel. During the Contention Access Period (CAP) of the superframe, the device repeatedly performs channel sensing to detect an idle state. Upon identifying an idle channel, it transmits a TDMA time slot allocation request, which encapsulates information such as its traffic priority and service requirements. If the backoff counter for one of the two channels decrements to zero, the device attempts to transmit the request, while the remaining devices continue to contend for the alternate channel. In the event that both backoff counters reach zero simultaneously, the device randomly selects one channel for transmission, leaving the other channel available for competing nodes. Should a collision occur, the device advances to the next backoff stage and re-initiates the channel sensing process.

Multiplexed Access Mechanism: The STA does not need to wait for the counters of both channels to decrement to zero. Once the backoff counter (BC1 or BC2) of either channel counts down to zero, the STA immediately transmits a reservation request or short data packets on that corresponding channel.

- (3) **Contention-Free Period (CFP):** For nodes that have completed successful reservation in the CAP phase, the CCO allocates dedicated TDMA time slots to them in the CFP phase. In this phase, nodes exclusively occupy channel resources for data transmission and are free from MAC layer collisions. Thus, data transmission failures are only affected by channel fading from the physical layer.

3. System Performance Analysis

3.1. Outage Probability Modeling

Outage probability is the core metric for quantifying the physical layer transmission reliability of the dual-mode communication system in complex station areas, which is defined as the probability that the instantaneous Signal-to-Noise Ratio (SNR) falls below the preset threshold γ_{th} [19]. To address the inherent fading limitations of single PLC and RF links in the station area, this study adopts a selection combining (SC) reception mechanism to exploit the complementary characteristics of the dual-medium channels. Let γ_{plc} and γ_{rf} denote the instantaneous SNR of the PLC link and the wireless link, respectively. It is assumed that the fading processes of the two links are mutually independent.

For the PLC link, incorporating the characteristics of log-normal fading and BG impulse noise, the closed-form expression of its outage probability under the SNR threshold can be derived using the law of total probability as follows:

$$P_{out}^{plc} = (1 - p)Q\left(\frac{\mu_{PLC} - \ln(h_{th,G})}{\sigma_{PLC}}\right) + pQ\left(\frac{\mu_{PLC} - \ln(h_{th,I})}{\sigma_{PLC}}\right) \quad (5)$$

Since $|h|$ follows a log-normal distribution, $\ln|h| \sim N(\mu_{PLC}, \sigma_{PLC}^2)$. The variable $Z = \frac{\ln|h| - \mu_{PLC}}{\sigma_{PLC}}$ follows a standard normal distribution, which can be characterized by the Gaussian Q-function: $Q(z) = \int_z^\infty \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$ ($P(Z < z) = 1 - Q(z)$). Where $h_{th,G}$ and $h_{th,I}$ denote the critical fading thresholds under background noise and impulse noise, respectively, and p is the pulse occurrence probability.

For the Rayleigh fading channel, the Signal-to-Noise Ratio (SNR) follows an exponential distribution. Its Cumulative Distribution Function (CDF) is exactly the outage probability:

$$P_{out}^{rf} = \int_0^{\gamma_{th}} \frac{1}{\bar{\gamma}_{rf}} e^{-\frac{\gamma}{\bar{\gamma}_{rf}}} d\gamma = 1 - \exp\left(-\frac{\gamma_{th}}{\bar{\gamma}_{rf}}\right) \quad (6)$$

Under the selection combining (SC) mechanism, a communication failure of the system is declared if and only if outages occur on both the PLC and RF channels simultaneously. Accordingly, the outage probability of the dual-mode channel is given by:

$$P_{out}^{Dual} = \{(1-p) \cdot Q\left(\frac{\mu_{PLC} - \ln h_{th,G}}{\sigma_{PLC}}\right) + p \cdot Q\left(\frac{\mu_{PLC} - \ln h_{th,I}}{\sigma_{PLC}}\right)\} \times \{1 - \exp(-\frac{\gamma_{th}}{\bar{\gamma}_{RF}})\} \quad (7)$$

To avoid simultaneously crediting diversity and multiplexing gains, the system operates in a priority-based hybrid framework. For high-priority control traffic operating in Selection Combining (SC) Mode, the physical layer success rate is defined by the joint non-outage probability:

$$P_{success}^{SC} = 1 - P_{out}^{Dual} \quad (8)$$

Conversely, for massive periodic metering traffic operating in Parallel Multiplexing Mode to achieve high throughput, independent streams are transmitted. The equivalent physical layer success rate is therefore bounded by the average individual outage probabilities of the channels. Accounting for the maximum MAC-layer retry limit (), the success rate in multiplexing mode is consistently modeled as:

$$P_{success}^{MUX} = 1 - \left(\frac{1}{2}P_{out}^{plc} + \frac{1}{2}P_{out}^{rf}\right)^{k_{retry}+1} \quad (9)$$

3.2. Throughput Modeling

The core of throughput calculation lies in solving the transmission probability and collision probability of nodes in dual channels. This section extends the classic two-dimensional Markov chain model to accommodate the characteristics of unsaturated Poisson traffic and dual interfaces. It is assumed that the node traffic follows a Poisson process with an arrival rate of λ . Within a unit time interval, the expected number of active nodes actually participating in channel contention is $N_\lambda = N(1 - e^{-\lambda})$. It should be explicitly noted that substituting the effective contention node count N_λ into the saturated-throughput formulation represents a mathematical approximation. This widely adopted simplification assumes that station activity follows a Bernoulli process derived from the Poisson arrival rate, which ensures the mathematical tractability of the coupled Markov model under unsaturated traffic, although it abstracts the transient discrete queueing dynamics.

For the RF wireless channel, the standard Binary Exponential Backoff (BEB) mechanism defined in the IEEE 802.11 [20] protocol is adopted. Let p_{rf} denote the collision probability and τ_{rf} denote the transmission probability. Based on the steady-state solution of the Markov chain, the relationship between τ_{rf} and p_{rf} satisfies [21]:

$$\begin{cases} \tau_{rf} = \frac{2(1-2p_{rf})}{(1-2p_{rf})(W_{rf}+1) + p_{rf}W_{rf}(1-(2p_{rf})^m)} \\ p_{rf} = 1 - (1 - \tau_{rf})^{N_\lambda - 1} \end{cases} \quad (10)$$

where W_{rf} is the initial contention window (CW_{min}), and m is the maximum backoff stage. According to the principle of channel contention, the collision probability p_{rf} is defined as the probability that at least one other active node is transmitting simultaneously when the target node initiates a transmission.

For the PLC channel, let p_{plc} denote the collision probability and τ_{plc} denote the transmission probability in the PLC channel. A system of joint equations is established via the Markov model as follows:

$$\begin{cases} \tau_{plc} = \frac{\sum_{i=0}^{m_1^k-2} t_i^k \prod_{j=0}^{i-1} (1-s_j^k) + \prod_{i=0}^{m_1^k-2} (1-s_j^k) \frac{t_{m_1^k-1}^k}{s_{m_1^k-1}^k}}{\sum_{i=0}^{m_1^k-2} bc_i^k \prod_{j=0}^{i-1} (1-s_j^k) + \prod_{i=0}^{m_1^k-2} (1-s_j^k) \frac{bc_{m_1^k-1}^k}{s_{m_1^k-1}^k}} \\ p_{plc} = 1 - (1 - \tau_{plc})^{N_k-1} \end{cases} \quad (11)$$

where x_l^k denotes the probability that the device transitions to the next backoff stage within l or fewer time slots at backoff stage i , resulting from sensing the PLC channel busy for $dc^k(i) + 1$ times. t_i^k denotes the probability that the device initiates a transmission attempt at backoff stage i of the PLC link. s_i^k denotes the probability that the device achieves successful transmission at backoff stage i of the PLC link. bc_i^k denotes the number of time slots consumed by the device at backoff stage i of the PLC link. For a comprehensive mathematical derivation of these intermediate variables and the step-by-step formulation of the extended Markov model, please refer to Appendix A

$$\begin{cases} x_l^k = \sum_{j=dc^k(i)+1}^l \binom{l}{j} (p_{plc})^j (1-p_{plc})^{l-j} \\ t_i^k = \frac{dc^k(i)+1}{CW_{1i}^k} + \sum_{l=dc^k(i)+1}^{CW_{1i}^k-1} \frac{1}{CW_{1i}^k} (1-x_l^k) \\ s_i^k = (1-p_{plc}) t_i^k \end{cases} \quad (12)$$

The CSMA/CA algorithm is a contention-based protocol. As the number of stations (STAs) increases, the collision probability of the stations rises significantly, which fails to guarantee the bandwidth and delay performance of each individual station. To improve the throughput and channel utilization of the system, this paper extends this algorithm to the CSMA/CA-TDMA hybrid protocol. In this protocol, the channel time is divided into a superframe structure, as shown in the figure. The superframe consists of three parts: the Beacon frame, the CSMA/CA-based CAP, and the TDMA-based CFP. Among them, the Beacon frame is used for network-wide time synchronization. Stations adopt the CSMA/CA protocol for contention-based access in the CAP, and transmit data packets in a time-division manner using the pre-allocated TDMA time slots in the CFP. The total length of the superframe is given by $L_{SF} = T_{Beacon} + T_{CAP} + T_{CFP}$.

During the CAP, stations contend for and reserve time slots via the multiplexed access algorithm. To ensure the mathematical tractability of the coupled dual-mode Markov model and derive closed-form transition probabilities, it is assumed for analytical simplicity that the average successful transmission duration T_s and collision duration T_c are identical for both the PLC and RF channels. While this symmetric assumption facilitates the derivation of the MAC-layer scheduling efficiency, it inherently idealizes the physical heterogeneity of the HPLC and HRF links. The average duration t_{mul} for a successful transmission on a single channel (i.e., one successful reservation) is given by:

$$t_{mul} = T_s + \frac{1-P_{tr}}{P_s P_{tr}} \sigma + T_c \left(\frac{1}{P_s} - 1 \right) \quad (13)$$

where σ denotes the backoff slot, P_{tr} denotes the probability that at least one station is in the transmission state, and P_s denotes the probability of successful transmission. Since the multiplexed access mechanism enables independent and parallel reservation completion on

both the PLC and RF channels, the average number of reservations that can be completed by a single channel within the duration T_{CAP} is $\frac{T_{CAP}}{t_{mul}}$. Under this idealized symmetric assumption, dual-channel multiplexing allows the total number of successfully completed reservation requests (N_{req}) in the CAP phase to reach its theoretical maximum upper bound, which effectively doubles the single-channel capacity:

$$N_{req} = 2 \times \frac{T_{CAP}}{t_{mul}} \quad (14)$$

After entering the CFP, the CCO allocates TDMA time slots according to the reservation status in the CAP phase. It is assumed that the CFP is temporally divided into M TDMA time slots. In the multiplexing mode, the CCO can allocate the PLC channel of the same time slice to station A and the RF channel to station B (or allocate both channels to station A for transmitting different data). Accordingly, the total number of available channel time slots N_{CFP} provided by the system within the CFP is given by:

$$N_{CFP} = 2M \quad (15)$$

Ultimately, within one superframe, the amount of data that can be actually successfully transmitted is determined by the minimum value between the number of successful reservations and the maximum time slot capacity that the system can provide, i.e., the barrel effect.

Let k denote the number of consecutive data packets with a size of $E[P]$ that can be transmitted in each TDMA time slot. Then the throughput S_{mux} of each multiplexed superframe is given by:

$$S_{mux} = \frac{\min(2M, \frac{2T_{CAP}}{t_{mul}}) \cdot k \cdot E[P]}{T_{Beacon} + T_{CAP} + T_{CFP}} \quad (16)$$

3.3. Access Delay Modeling

When evaluating the performance of the proposed network, severe limitations arise from the isolated derivation of network throughput, data acquisition success rate, or access delay alone. Throughput reflects the macroscopic data transmission capability, the data acquisition success rate characterizes the anti-interference performance, while the access delay determines the microscopic real-time performance. The total duration of a superframe is fixed, and there exists a strict zero-sum game relationship between the CAP and CFP. Therefore, it is essential to incorporate access delay into the system modeling. In the proposed model, the access delay is defined as the mathematical expectation of the backoff delay and collision delay incurred by data packets during the CSMA/CA contention process in the CAP.

In practical implementation, the parallel backoff mechanism must address the heterogeneous physical characteristics of the PLC and RF media. To resolve the inherent differences in native slot lengths, a unified logical slot time is enforced at the system level via the CCO's beacon, ensuring that the discrete countdown boundaries of both interfaces remain strictly aligned. Furthermore, a hardware-level tie-breaking arbitration rule is employed to handle timing conflicts, such as the 'dual success' scenario where both backoff counters decrement to zero simultaneously. In this case, the STA randomly selects only one channel to transmit the reservation request and defers the other, which fundamentally eliminates self-interference and guarantees the mathematical stability of the coupled Markov

contention [21]. Let τ_{dual} denote the transmission probability of successfully contending for both channels simultaneously (i.e., simultaneously accessing the dual-mode channels):

$$\tau_{dual} = \sum_{i=0}^{m_1^k} \sum_{j=0}^{m_2^k} p_{ij} \tau_{ij} \quad (17)$$

where $p_{ij,k}$ denotes the probability that a device with priority level k is at backoff stage i on the power line channel and backoff stage j on the wireless channel, respectively. $\tau_{ij,k}$ denotes the probability that a device with priority level k successfully contends for both the wireless and power line channels simultaneously at different backoff stages.

$\tau_{dual,k}$ is used to update and calculate the transmission probabilities of the two channels in the dual-mode system. Since a station that successfully gains simultaneous access to both channels will randomly select one channel for transmission, the parameter $\tau_{dual,k}$ is averaged in this formulation [21]. The average transmission probability τ'_k of the device in the dual-mode channels can be calculated from $\tau'_{1,k}$ and $\tau'_{2,k}$:

$$\tau'_{1,k} = \tau_{1,k} + \tau_{dual,k}/2 \quad (18)$$

$$\tau'_{2,k} = \tau_{2,k} + \tau_{dual,k}/2 \quad (19)$$

$$\tau'_k = (\tau'_{1,k} + \tau'_{2,k})/2 \quad (20)$$

Using the modified transmission probability τ'_k , we can derive the probability P_{tr} that at least one node transmits in an arbitrary time slot of any channel (i.e., the channel is busy), as well as the conditional probability P_s that only one node transmits (i.e., successful reservation):

$$P_{tr} = 1 - (1 - \tau'_k)^N \quad (21)$$

$$P_s = \frac{N \tau'_k (1 - \tau'_k)^{N-1}}{P_{tr}} \quad (22)$$

Channel access is an event dependent on an independent Bernoulli process. The success probability of each access attempt of a node in the CAP phase is $P_{success} = P_{tr} P_s$. Accordingly, the number of time slots N_s that a node needs to experience to achieve one successful access follows a geometric distribution with parameter $P_{success}$.

$$P\{N_s = k\} = P_{succ} (1 - P_{succ})^{k-1} \quad (23)$$

According to the mathematical expectation property of the geometric distribution, the expected number of time slots required by the system on average is given by:

$$E[N_s] = \frac{1}{P_{succ}} = \frac{1}{P_{tr} P_s} \quad (24)$$

Under the CSMA/CA mechanism, not all of these N_s time slots are idle slots with a fixed duration of σ . The actual expected time duration E of a generic time slot is formed by the weighted sum of the time consumption of three mutually exclusive events: idle, successful transmission, and collision:

$$E = (1 - P_{tr})\sigma + P_{tr} P_s T_s + P_{tr} (1 - P_s) T_c \quad (25)$$

The total access delay in the CAP phase must account for both backoff durations and collision penalties. Since channel access under CSMA/CA follows an independent Bernoulli process, the number of access attempts N_s follows a geometric distribution with a success probability of $P_{success} = P_{tr} P_s$. Consequently, the expected number of attempts

is $E[N_s] = 1/P_{tr}P_s$. Furthermore, not all time slots are idle; the expected duration of a generic time slot E is formed by the weighted sum of three mutually exclusive events: an idle slot duration (σ), a successful transmission duration (T_s), and a collision slot duration (T_c). Thus, the overall average access delay E is rigorously derived as the product of the expected number of attempts and the generic time slot duration, as expressed in Equation (24). In summary, the average access delay E of the single-hop dual-mode network in the CAP phase can be rigorously expressed as the product of the expected number of attempts and the duration of the generic time slot:

$$E = E[N_s] \times E = \frac{(1 - P_{tr})\sigma + P_{tr}P_sT_s + P_{tr}(1 - P_s)T_c}{P_{tr}P_s} \quad (26)$$

4. Design of the Core Algorithm

Based on the above joint PHY-MAC layer system analysis model, it is found that there exists a nonlinear coupling correlation between the resource ratio allocation of the CAP and CFP in the superframe structure, and the system node scale, concurrent collision probability, as well as the ultimate throughput. If the CAP is allocated with an excessively small duration, massive nodes will suffer from severe collisions and packet loss due to the failure to complete reservation requests. Conversely, if the CAP is allocated with an excessively large duration, although the reservation success rate is improved, the CFP time slot resources reserved for actual data packet transmission will be severely squeezed, resulting in a significant waste of channel bandwidth. To break through this resource allocation bottleneck under a dynamically varying node scale, this study designs a Q-learning reinforcement learning-based adaptive optimization algorithm for the CAP/CFP ratio.

In this algorithm, as shown in Figure 3, the CCO is taken as the Agent, which autonomously approaches the globally optimal time slot division strategy through periodic state interaction and feedback learning with the network Environment it manages. The core mathematical abstraction and system modeling steps of the algorithm are described as follows:

(1) Definition of State Space and Action Space:

The State Space (S) is defined as the number N of active sensing nodes in the network within the current superframe period $S = \{N | N \in \{20, 30, \dots, 200\}\}$. With the variation in power grid service traffic, the number of nodes presents a dynamic distribution, which constitutes the external environmental state features for the agent's decision-making. A critical property of the proposed model is that the environmental state transition is strongly coupled with the agent's action, satisfying the Markov Decision Process (MDP) paradigm rather than a degenerate contextual bandit. The state at the next superframe $t + 1$, denoted as N_{t+1} , is not merely a product of exogenous grid traffic, but is recursively defined by:

$$N_{t+1} = \lambda_{t+1} + N_{retry}(N_t, a_t) \quad (27)$$

where λ_{t+1} denotes the newly generated exogenous data requests from the power grid, and $S_t(N_t, a_t)$ denotes the number of nodes successfully scheduled and served in the current superframe. Since S_t is strictly determined by the chosen CAP/CFP ratio (a_t), any suboptimal action—such as an inadequate CAP causing contention failure or an insufficient CFP leading to dropped reservations—will diminish S_t . The unserved node ($N_t - S_t(N_t, a_t)$) will remain active and trigger MAC-layer retransmissions in the next superframe, severely inflating N_{t+1} . In the simulation environment, the active node count N at each superframe is dynamically generated by accumulating the newly arrived data requests (modeled as a

Poisson process) and the backlogged nodes that failed to transmit in previous superframes due to contention collisions or capacity limits.

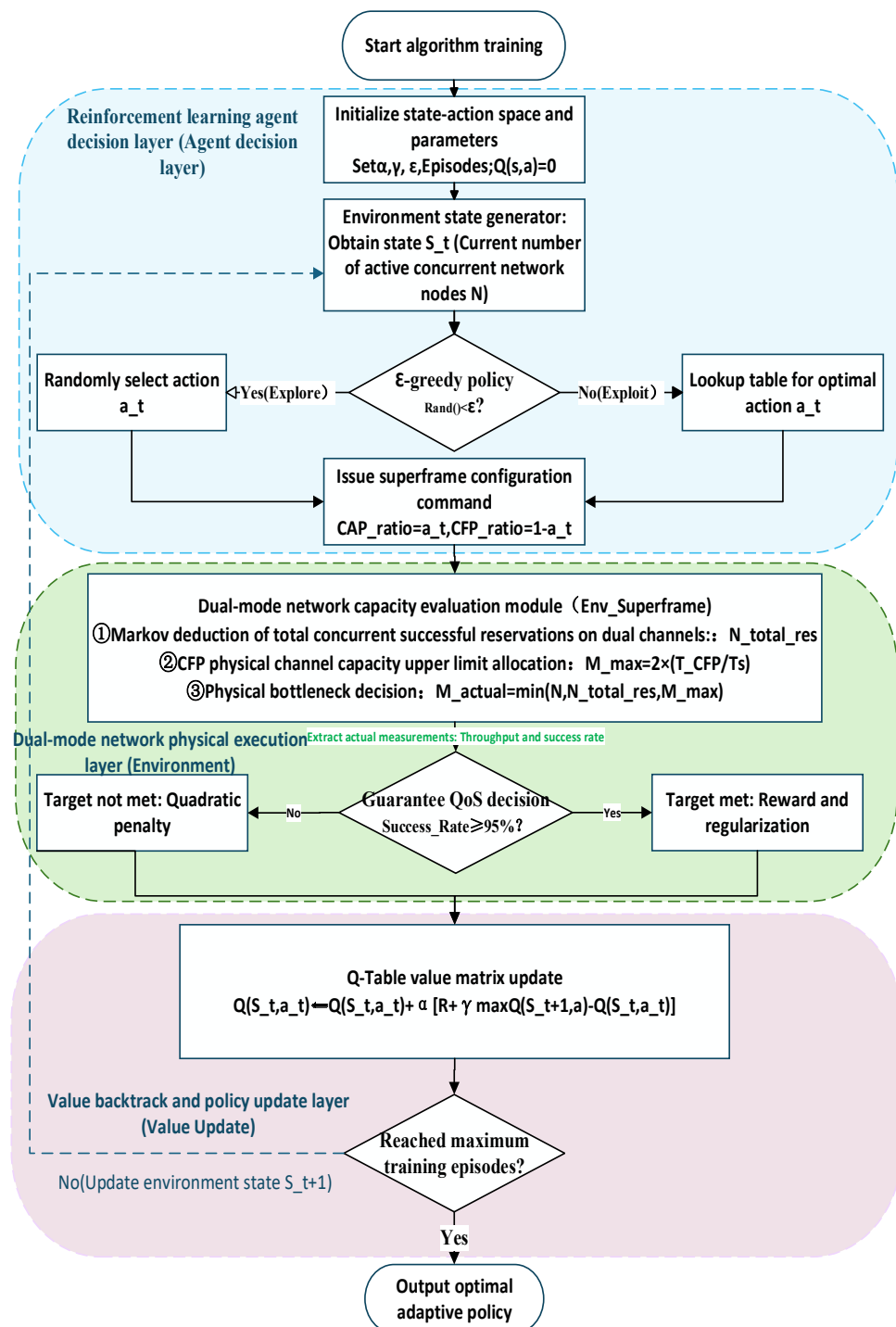


Figure 3. Flowchart of the algorithm.

By abstracting the state space solely to the active node scale N , the algorithm inherently encapsulates macroscopic traffic loads and collision probabilities, while complex queue dynamics and fast-fading channel qualities are efficiently absorbed by the capacity bottleneck model and the asymmetric reward function, thereby preventing the curse of dimensionality. The Action Space (A) is defined as the discretized selection set of the ratio η_{CAP} (CAP ratio) allocated to the CAP in the next superframe by the agent

$A = \{\eta_{CAP} | \eta_{CAP} \in \{0.1, 0.15, \dots, 0.9\}\}$. Correspondingly, the ratio of the CFP is naturally determined as $\eta_{CFP} = 1 - \eta_{CAP}$.

- (2) Environmental Dynamics and System Bottleneck Decision Model: For a given state N and action η_{CAP} , the environment shall accurately evaluate the actual physical capacity of the system. Based on the dual-mode Markov chain model, the time consumption of a single successful reservation t_{succ} is calculated using the analytically derived transmission probability and collision probability. Subsequently, within the allocated T_{CAP} duration, the theoretical numbers of successful reservation requests that can be supported by the network on the PLC and RF dual physical channels are N_{res_p} and N_{res_w} , respectively. The total carrying capacity of concurrent requests is given by: $N_{req} = N_{res_p} + N_{res_w}$.

Meanwhile, in the CFP, the total maximum number of contention-free data time slots that the system can provide is $M_{max} = 2 \times [T_{CFP}/T_{s_data}]$. The “barrel effect” of the system determines the number of nodes M_{actual} that can be successfully scheduled and complete data delivery within the current superframe. It is constrained by the minimum short board among the total number of physical nodes in the network, the number of successful reservations in the CAP, and the time slot capacity of the CFP. Its mathematical criterion is given by:

$$M_{actual} = \min(N, N_{req}, M_{max}) \quad (28)$$

Because the primary objective of the Q-learning agent during massive concurrent bursts is to optimize the capacity of periodic metering traffic, the comprehensive data acquisition success rate is consistently evaluated under the Multiplexing Mode. It is rigorously defined as the product of the MAC layer scheduling efficiency and the per-link multiplexing reliability term:

$$P_{success} = (M_{actual}/N) \times P_{success}^{MUX} \quad (29)$$

- (3) Dual-Constraint Comprehensive Reward Function Formulation The primary task of meter data acquisition is to guarantee the reliability of metering data, followed by the pursuit of throughput maximization. Accordingly, the comprehensive data acquisition success rate of the system $P_{success}$ is defined as the product of the MAC layer scheduling efficiency (M_{actual}/N) and the non-outage probability of the underlying physical layer ($P_{success_PHY}$). To strictly guarantee the Quality of Service (QoS) for high-reliability services, this algorithm constructs an asymmetric penalty–reward function γ_t with a hard constraint threshold:

$$\gamma_t = \begin{cases} -(95.0 - P_{success})^2, & P_{success} < 95.0\% \\ 20 + 10 \times S_{mux} - 0.4 \times \eta_{CAP}, & P_{success} \geq 95.0\% \end{cases} \quad (30)$$

where S_{mux} denotes the effective throughput of the system (in Mbps). The mechanism of this reward function is delicately designed: when the system success rate falls below the 95% threshold, the algorithm imposes a severe penalty amplified in a quadratic form, forcing the agent to rapidly escape from the adverse action interval with high collision or insufficient capacity. The threshold of 95% in Equation (26) is determined by the stringent engineering constraints of low-voltage distribution networks, specifically targeting a 15 min load curve collection success rate of no less than 95%. It serves as a hard QoS boundary. Notably, this threshold is highly flexible and can be dynamically adjusted (e.g., to 98% or 99%) according to specific service requirements, highlighting the adaptability of the proposed Q-learning framework. In contrast, within the safe region where the success rate meets the requirement, the algorithm not only provides positive incentive for the improvement of

throughput S_{mux} , but also introduces a tiny negative regularization term $-0.4 \times \eta_{CAP}$. This term aims to drive the agent to minimize the unnecessary CAP contention overhead under the premise of guaranteeing the same throughput, thus reallocating more time resources to the CFP to reduce the end-to-end delay.

- (4) Q-Table Update Rule: The CCO adopts the ϵ – *greedy* strategy to strike a balance between exploration of the unknown and exploitation of existing experience, and iteratively updates the state-action value matrix $Q(s_t, a_t)$ based on the Bellman equation:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[\gamma_t + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t)] \quad (31)$$

After long-term offline initialization and online adaptation paradigm, the proposed algorithm enables the network to seamlessly and adaptively calculate and deliver the optimal superframe structure ratio without manual intervention when facing burst dense node access, which resolves the medium access crisis in the dual-mode high-concurrency environment from the macro protocol level.

5. Results

To thoroughly verify and analyze the dynamic decision-making performance and convergence characteristics of the proposed Q-learning-based adaptive optimization algorithm for the CAP/CFP ratio in the superframe under different network densities, this study develops a simulation platform integrating a reinforcement learning agent and a discrete-event dynamic evolution environment for the MAC layer. The simulation platform adopted in this work is built upon, but extends, the analytical model and simulation framework proposed in Reference [21]. To ensure the statistical significance and robustness of our findings, the performance evaluation is conducted using a Monte Carlo simulation framework. All reported simulation results represent the statistical means obtained by averaging over 100 independent runs with varying random seeds. To explicitly demonstrate the algorithmic stability against inherent network stochasticity (e.g., random backoffs and Poisson traffic arrivals), the 95% confidence intervals are calculated and depicted as shaded regions (or error bars) alongside the mean curves in the following performance figures.

5.1. Simulation Parameter Configuration

The hyperparameter configuration of the proposed Q-learning adaptive scheduling algorithm in this paper is detailed in Table 1. To achieve smooth convergence in the unknown and highly dynamic station area environment, the algorithm adopts the ϵ – *greedy* strategy. The initial exploration rate ϵ_{init} is set to 1.0 to ensure that the agent fully explores the action space in the initial training phase, and smoothly transitions to the exploitation of existing experience with a decay rate of 0.98. The learning rate $\alpha = 0.1$ effectively avoids oscillations in value updates caused by transient network fluctuations; the discount factor $\gamma = 0.8$ guides the CCO to maximize the throughput of the current superframe while taking into account the long-term stability and congestion avoidance of the network.

Table 1. Core Parameters of the Q-learning Superframe Scheduling Algorithm.

Parameter Category	Parameter	Symbol/Value
Space Definition	State Space	$S \in \{20, 30, \dots, 200\}$
	Action Space	$A \in \{0.10, 0.11, \dots, 0.90\}$
Learning Parameters	Learning Rate	$\alpha = 0.1$
	Discount Factor	$\gamma = 0.8$

Table 1. Cont.

Parameter Category	Parameter	Symbol/Value
Exploration Strategy	Initial Epsilon	$\varepsilon_{\text{init}} = 1.0$
	Minimum Epsilon	$\varepsilon_{\text{min}} = 0.01$
	Decay Rate	$\varepsilon \leftarrow \varepsilon \times 0.98$
Training and Constraints	Episodes	$E_{\text{max}} = 1000$
	QoS Threshold	$SR_{th} \geq 95\%$

5.2. Simulation Result Verification

Figure 4 compares the performance differences in normalized throughput with the increase in node scale between the proposed dual-mode multiplexed access mechanism and the single PLC channel, as well as the single RF channel under a fixed superframe ratio. With the increase in the number of concurrent nodes in the station area, constrained by the contention nature of the CSMA/CA protocol, the channel collision probability increases significantly. A large number of time slot resources are excessively consumed in backoff waiting and collision penalties, resulting in a monotonically decreasing trend in the throughput of all three mechanisms. When the number of concurrent nodes reaches 120, the normalized throughput of the proposed dual-mode multiplexing mechanism (PLC/RF Mux) is approximately 0.70. Quantitatively, the proposed dual-mode multiplexing mechanism achieves a significant throughput improvement of approximately 45.8% compared to the pure RF mode, and an impressive 180.0% improvement compared to the pure PLC mode. This explicitly demonstrates the substantial capacity gain achieved by our parallel CSMA/CA scheduling over conventional single-channel approaches.

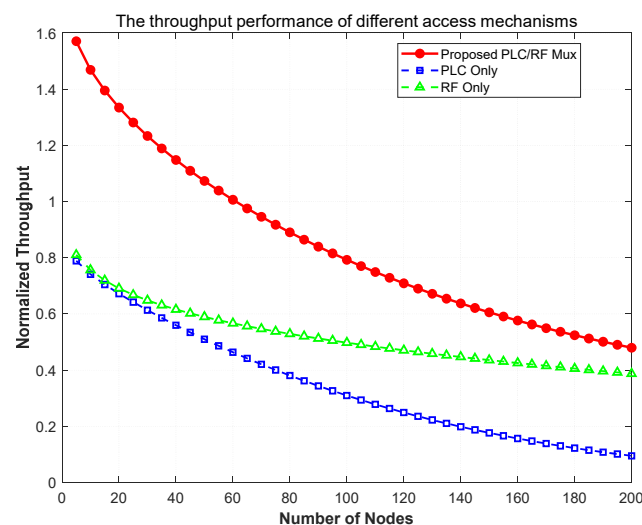


Figure 4. The throughput performance of different access mechanisms.

Figure 5 presents the variation curves of the average node access delay in the CAP with respect to the number of concurrent nodes and the initial contention window. During the light-load phase of the network, the delay difference under different contention window configurations is not significant. However, when the network transitions to the heavy-load state, the system delay exhibits a nonlinear surge. In high-density concurrent scenarios, a narrow initial backoff space can easily cause a large number of nodes to finish backoff at the same time slot, leading to severe collision conflicts, which in turn forces nodes to frequently trigger time-consuming high-order backoff cycles. Appropriately increasing the initial window can effectively disperse the transmission probabilities of each node in the time domain and mitigate transient collisions.

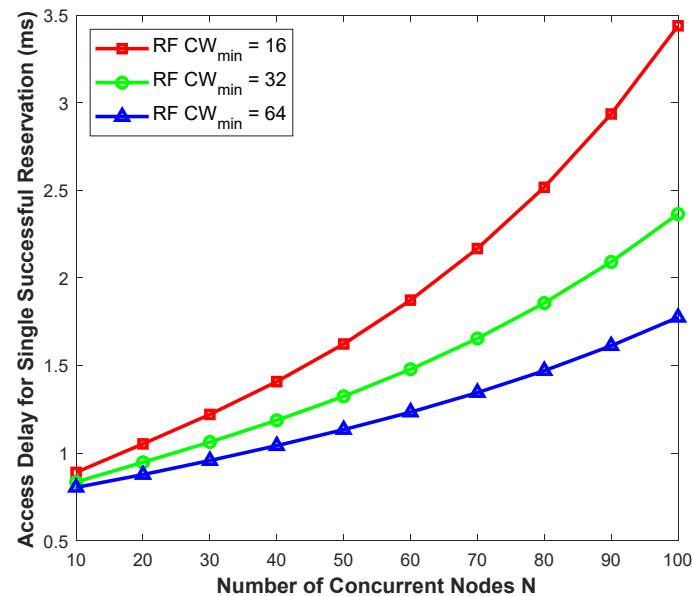


Figure 5. Impact of Different Wireless Contention Windows on Access Delay.

Figure 6 illustrates the variation trend of the steady-state collision probability of the PLC channel with the number of active concurrent nodes N in the station area under different MAC layer backoff parameter configurations. Simulation results show that the collision probability under all three configurations exhibits a monotonically increasing trend with the increase in N . These results objectively demonstrate that merely reducing the contention window or increasing the deferral counter will exacerbate channel collisions in concurrent scenarios. Fixed parameter configurations are difficult to adapt to the dynamic variations in network load in the station area, which provides a theoretical basis for introducing adaptive algorithms to perform joint optimization of joint PHY-MAC layer parameters.

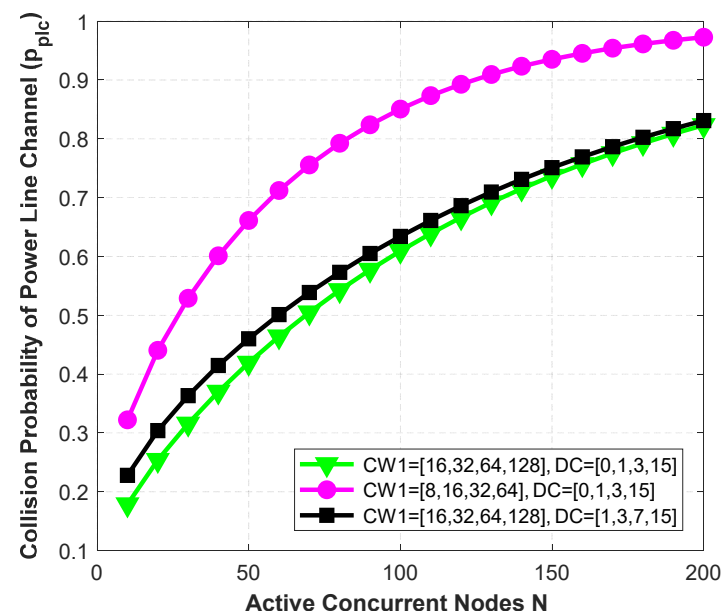


Figure 6. Impact of Different Backoff Parameters on Collision Probability in Power Line Channels.

Figure 7 presents the evolution characteristics of the effective system throughput with respect to the node traffic arrival rate λ under different station area user scales (N). Simulation results show that with the increase in the traffic arrival rate, the effective system throughput increases steadily under all different station area user scales.

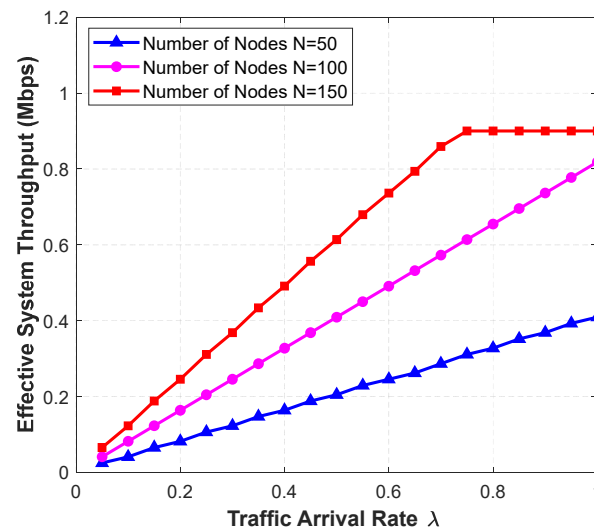


Figure 7. Variation Characteristics of Throughput with Bursty Traffic Load under Different Numbers of Users.

Figure 8 depicts the variation in system throughput with the number of concurrent nodes under three parameter configurations. For different traffic loads, higher arrival rates yield faster throughput growth, while light load results in low channel utilization and limited throughput improvement. As for channel conditions, rising outage probability noticeably degrades throughput and causes earlier saturation, revealing that severe noise and fading become the dominant capacity bottleneck. For fixed CAP ratios, a larger CAP proportion delivers better throughput under high node density, whereas smaller CAP ratios trigger premature throughput saturation due to intensified reservation collisions. These results demonstrate that static superframe settings cannot accommodate dynamic network loads, which validates the necessity of the proposed adaptive scheduling mechanism.

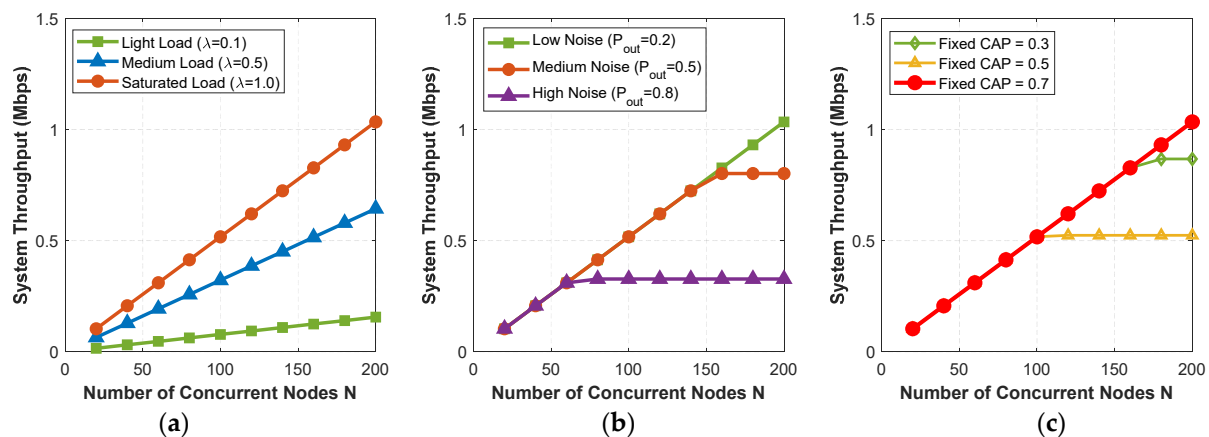


Figure 8. Comprehensive Performance Evaluation Under Heterogeneous Environmental Conditions: (a) Impact of Traffic Scales (λ); (b) Impact of Background Noise; (c) Various CAP/CFP Configurations.

To explicitly verify the core design objective of our proposed dual-mode MAC protocol, Figure 9 presents the data acquisition success rate versus the number of concurrent nodes (N). As quantitatively demonstrated in the figure, under light to medium traffic loads ($N > 100$), both the proposed Q-learning adaptive algorithm and the conventional fixed-ratio baseline (Fixed CAP = 0.5) easily maintain a success rate near 100%. However, as the network transitions into a heavy-load state, the baseline strategy suffers a sharp performance degradation due to severe channel collisions during the fixed CAP and subop-

timal resource utilization. In stark contrast, the proposed Q-learning adaptive algorithm robustly guarantees communication reliability. By intelligently expanding the superframe CAP resources to alleviate real-time collision dynamics, our algorithm maintains a success rate of over 96.5% across the entire tested node spectrum up to $N = 180$.

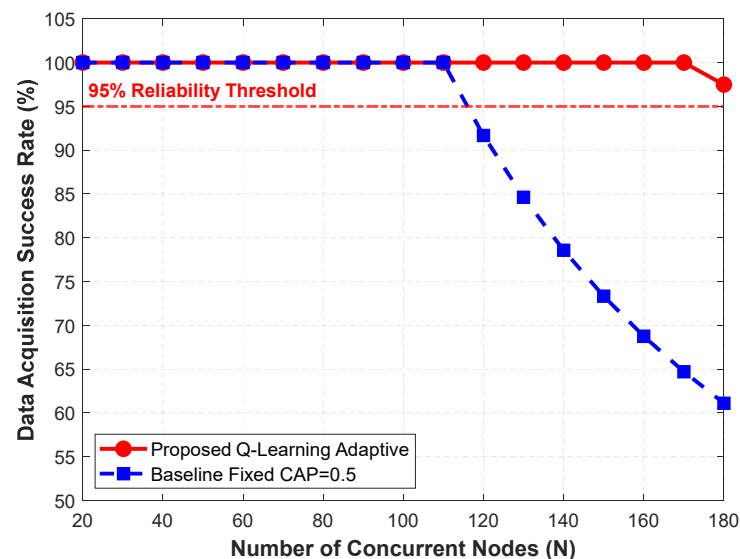


Figure 9. Data Acquisition Success Rate versus the Number of Concurrent Nodes under Different Scheduling Strategies.

To strictly quantify the accuracy of the aforementioned effective node count approximation, we conducted a dedicated packet-level Discrete-Event Simulation (DES). Unlike the analytical formulation, the DES does not utilize the N_λ approximation; instead, it explicitly models the true Poisson packet arrival processes and frame-by-frame queueing contention for all stations. As depicted in Figure 10, the analytical model captures the system throughput with high fidelity. The analytical curve aligns closely with the DES results across various traffic intensities (in $[0.1, 1.5]$), maintaining a maximum relative error of less than 5%. This demonstrates that the approximation is sufficiently accurate and robust for the purpose of dual-mode capacity evaluation and resource allocation.

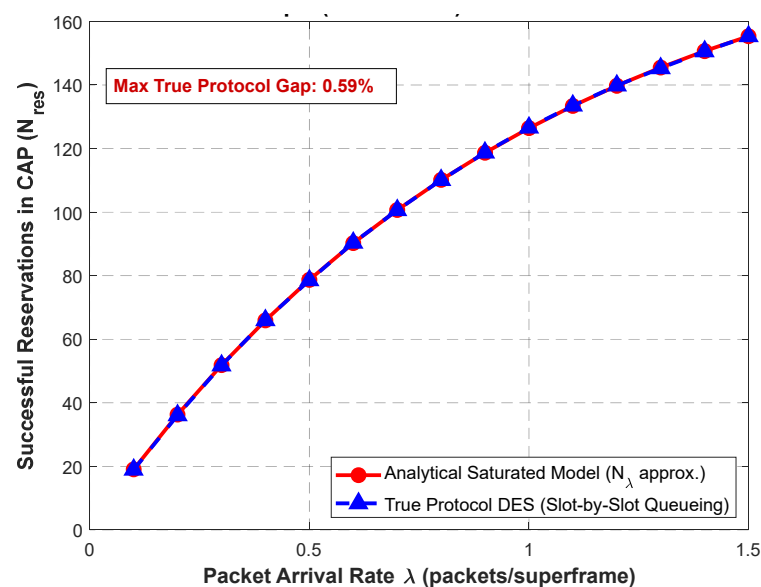


Figure 10. Validation of Equation (8) (RF Channel) via Protocol-Level DES.

Figure 11 illustrates the effective system throughput versus the number of concurrent nodes for the proposed Q-learning adaptive scheduling algorithm and the conventional fixed-ratio (1:1) strategy. Under light network loads, the node collision rate is extremely low; both strategies meet service demands, exhibiting linear throughput growth with highly coincident curves. As the number of concurrent nodes increases, the network transitions to a heavy-load state. Lacking a dynamic adjustment mechanism, the conventional fixed strategy exacerbates ineffective collisions during the CAP and strictly limits the maximum data transmission in the CFP, causing throughput to saturate prematurely at approximately 0.90 Mbps. In contrast, the proposed Q-learning algorithm achieves a significantly higher maximum throughput of 1.30 Mbps in the heavy-load region, delivering a quantitative capacity improvement of approximately 44.4% over the baseline. By continuously sensing network congestion, the algorithm adaptively optimizes the time slot allocation ratio between the CAP and CFP. While maintaining the node reservation success rate, it compresses ineffective contention duration and reallocates available resources toward the data transmission phase. This approach effectively mitigates channel congestion caused by high concurrency and enhances the ultimate transmission capacity of the dual-mode system.

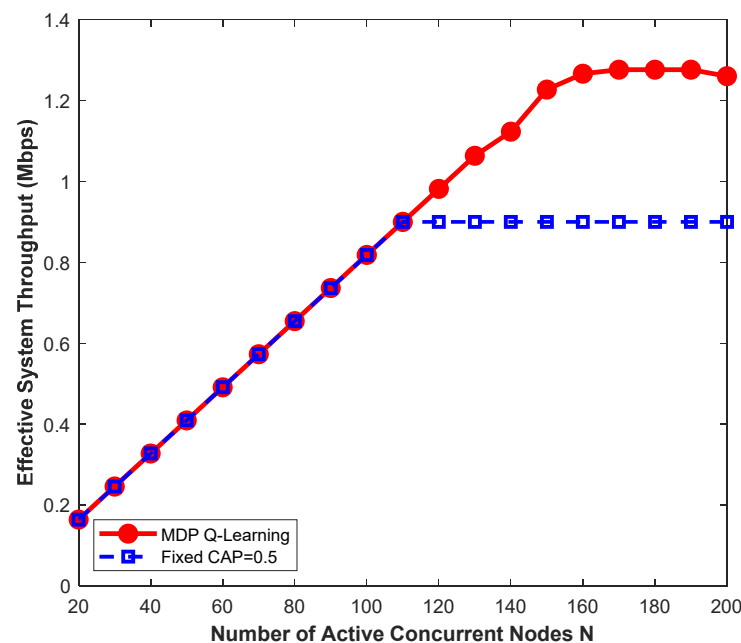


Figure 11. Throughput Comparison: ML-Enhanced vs. Conventional Fixed-Proportion.

Figure 12 depicts the convergence curve of the proposed Q-learning adaptive scheduling algorithm over 2000 training episodes. The light purple line shows the raw accumulative reward per episode, which presents significant stochastic fluctuations caused by random traffic arrival, channel fading, and the ϵ -greedy exploration mechanism. The dark blue smoothed moving average curve clearly demonstrates the learning trend: in the early training stage, the average reward climbs rapidly from negative values, as the agent escapes from low-reward states where the success rate fails the 95% reliability constraint. After approximately 800 episodes, the smoothed reward converges to a stable level around 20 and maintains slight fluctuations thereafter. This convergence characteristic validates that the proposed algorithm can effectively learn the optimal superframe time slot allocation strategy, and the trained policy achieves stable performance under dynamic network conditions.

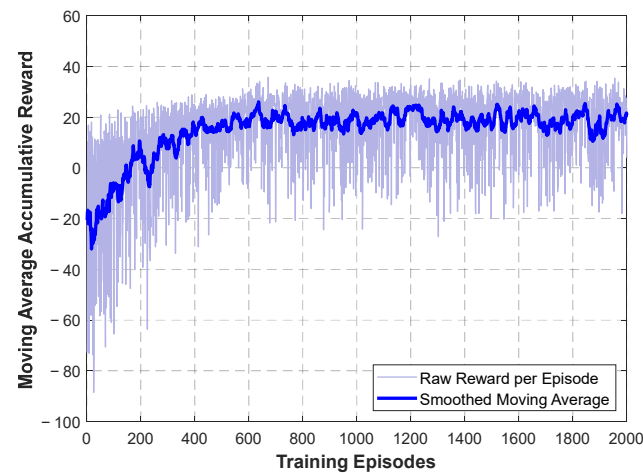


Figure 12. Q-learning policy convergence curve.

Figure 13 verifies the global optimization advantage of the proposed Q-learning superframe scheduling algorithm through horizontal comparison. Facing high-concurrency requests, the conventional fixed-ratio strategy falls into severe congestion due to the lack of environmental perception, resulting in throughput degradation. Compared with other baseline schemes, existing heuristic algorithms (e.g., SA) are limited by the millisecond-level delay constraints of the MAC layer, which restricts their exploration iterations within a single superframe and makes them prone to converge to local suboptimal solutions; while typical online learning algorithms (e.g., UCB) suffer from significant oscillations in system throughput and performance degradation due to the high online trial-and-error cost (Regret). The UCB multi-armed-bandit algorithm is purposefully included as a baseline to highlight the fundamental limitations of modeling this environment without sequential state transitions. To address the above limitations, the proposed algorithm adopts an offline initialization and online adaptation paradigm. The offline phase leverages the closed-form analytical environment to rapidly train a robust initial Q-table. During online deployment, rather than acting as a static lookup table, the Q-learning agent continuously fine-tunes the Q-values based on real-world environmental feedback, effectively compensating for any unpredictable deviations between the analytical theoretical model and actual physical channel fluctuations.

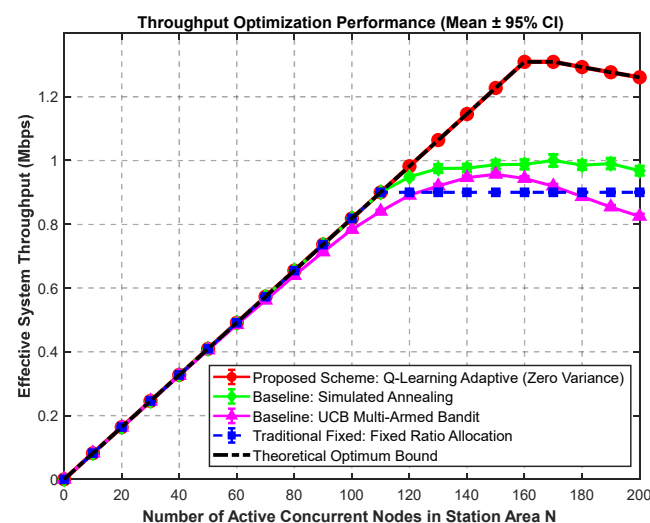


Figure 13. Comparison of Throughput Optimization Performance of Different Time Slot Scheduling Algorithms.

5.3. Discussion and Limitations of the Model

While the proposed Q-learning based dual-mode MAC protocol demonstrates significant performance gains, it is necessary to acknowledge several theoretical approximations and limitations in our current modeling, which point to directions for future research:

- (1) **Approximation for Unsaturated Traffic:** We adapted Bianchi's saturated model using an effective-contending-node count ($N_{\lambda} = N(1 - e^{-\lambda})$) for unsaturated Poisson traffic. While our independent discrete-event simulation validates its steady-state accuracy, this mean-value approximation may exhibit minor deviations during extreme transient bursts.
- (2) **Idealized Dual-Channel Synchronization in CFP:** In the theoretical capacity calculation, the maximum transmission capacity during the Contention-Free Period (CFP) is estimated by aggregating the capacity of both channels (e.g., effectively multiplying the single-channel slot capacity by a factor of 2, minus overhead). This represents a theoretical upper bound, assuming perfect time-domain synchronization and zero cross-channel interference between the PLC and RF interfaces on the same device. In practical physical layer (PHY) implementations, guard bands and synchronization delays may slightly reduce this theoretical maximum throughput.
- (3) **Absence of Hardware Testbed:** All performance metrics rely on high-fidelity discrete-event simulations. Developing a complex FPGA-based dual-interface testbed for large-scale networks ($N = 200$) is reserved for future work.

6. Conclusions

To address the communication bottlenecks caused by massive burst traffic and harsh channel environments in low-voltage distribution networks, this paper proposes a joint PHY-MAC layer analysis and dynamic resource adaptive scheduling framework for dual-mode communication oriented towards highly concurrent scenarios. The study first constructs a joint PHY-MAC layer model for dual-mode networks based on parallel multiplexing, integrating the outage characteristics of physical links with the contention mechanisms of the MAC layer, and derives key performance indicators using an extended two-dimensional discrete-time Markov chain. On this basis, to overcome the resource mismatch limitations inherent in traditional fixed superframes and strictly satisfy the operational hard constraint of a minimum 95% data collection success rate, a Q-learning-based adaptive superframe time slot scheduling algorithm is designed. Simulation results demonstrate that the proposed algorithm can autonomously determine the optimal ratio of the CAP to the CFP according to the dynamic network load, effectively mitigating channel congestion under extreme concurrent surges. While strictly guaranteeing highly reliable data collection, it maximizes the aggregate throughput of the network. Although this study strictly evaluates the proposed scheduling algorithm through a high-fidelity discrete-event simulation platform, we acknowledge the absence of a physical hardware testbed. Developing a large-scale dual-mode hardware testbed with hundreds of nodes presents significant engineering challenges. In our future work, we plan to deploy a small-scale physical prototype to further validate the algorithm's performance under real-world hardware constraints. Future research will also consider introducing a multi-hop dual-mode relay mechanism based on the current single-hop topology to explore deeper-layer resource coordination within larger-scale and more complex distribution networks.

Author Contributions: Methodology, Y.Z.; Investigation, Y.Z. and Z.C.; Writing—original draft, Z.C. and B.J.; Writing—review & editing, Y.Z., Z.C. and B.J. All authors have read and agreed to the published version of the manuscript.

Funding: This paper was funded by the Science and Technology Program of China Southern Power Grid Co., Ltd. (grant number: YNKJXM20240477).

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare that this study received funding from the Science and Technology Program of China Southern Power Grid Co., Ltd. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication. Authors Yuerong Zhao and Bo Jiang were employed by the company Measurement Center of Yunnan Power Grid Co., Ltd. Author Zhixiong Chen was employed by the company Xi'an Actionpower Electric Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

Appendix A. Derivation of the Extended Markov Model for the PLC Channel

Unlike the standard IEEE 802.11 CSMA/CA [20] protocol which primarily relies on the Backoff Phase Counter (BPC) and Backoff Counter (BC), the IEEE 1901 standard [22] introduces a Deferral Counter (DC) to handle harsh power line environments. The DC decreases only when the power line channel is busy. If the DC reaches zero and the channel remains busy, the station immediately jumps to the next backoff phase. Consequently, the transmission probability τ_{plc} cannot be calculated using a simple closed-form Markov equation; rather, it must be derived through an updating iterative function based on renewal reward theory. The analytical expression for τ_{plc} is structured as a fractional equation. The numerator represents the expected number of transmission attempts made before a station successfully transmits a packet, while the denominator denotes the expected total number of time slots consumed during this entire process.

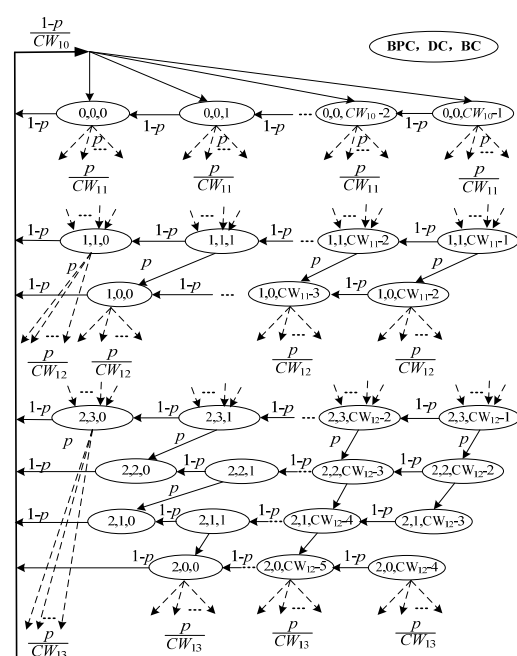


Figure A1. Markov Model for the PLC Channel.

(1) Derivation of the Intermediate Micro-State Variables

To solve the updating iterative function, the microscopic state transitions within a specific backoff stage i must first be quantified. Let CW_{1i}^k denote the contention window

for priority k at stage i , and $dc^k(i)$ denote the corresponding initial DC value. The variable x_l^k represents the transition probability that the station prematurely jumps to the next backoff phase $i + 1$ in l or fewer time slots. This occurs when the station senses the medium busy $dc^k(i) + 1$ times before its BC reaches zero. Assuming channel sensing events are independent with a busy probability p_{plc} , x_l^k follows a cumulative binomial distribution:

$$x_l^k = \sum_{j=dc^k(i)+1}^l \binom{l}{j} (p_{plc})^j (1 - p_{plc})^{l-j} \quad (A1)$$

The variable t_i^k denotes the probability that the station successfully attempts to transmit during phase i (i.e., its BC decrements to zero before the DC expires). Assuming the initial BC value is uniformly distributed across the contention window $[0, CW_{1i}^k - 1]$, t_i^k is derived by accumulating the non-blocking probabilities across all possible slots:

$$t_i^k = \frac{dc^k(i) + 1}{CW_{1i}^k} + \sum_{l=dc^k(i)+1}^{CW_{1i}^k-1} \frac{1}{CW_{1i}^k} (1 - x_l^k) \quad (A2)$$

Consequently, the conditional probability of a successful transmission within stage i , denoted as s_i^k , requires that an attempt is made and no collision occurs on the channel:

$$s_i^k = (1 - p_{plc}) t_i^k \quad (A3)$$

The expected number of time slots consumed at stage i , denoted as bc_i^k , is derived by calculating the mathematical expectation of time elapsed before either a transmission attempt or a forced phase transition occurs:

$$bc_i^k = \frac{1}{CW_{1i}^k} \sum_{l_1=dc^k(i)+1}^{CW_{1i}^k-1} [(l_1 + 1)(1 - x_{l_1}^k) + \sum_{j=dc^k(i)+1}^{l_1-1} j(x_j^k - x_{j-1}^k)] + \frac{(dc^k(i) + 1)(dc^k(i) + 2)}{2CW_{1i}^k} \quad (A4)$$

(2) Derivation of the Macroscopic Transmission Probability (τ_{plc})

By integrating the phase-specific parameters derived above, the global steady-state transmission probability is formulated. The probability that a station reaches stage i without any prior successful transmission is given by the cumulative product of failures from all preceding stages, expressed as $\prod_{j=0}^{i-1} (1 - s_j^k)$.

According to the absorption rules of the IEEE 1901 protocol, when a device reaches the maximum backoff stage (denoted as stage M), it will re-enter this maximum stage continuously upon failure until successful transmission or packet drop. This absorbing behavior generates a geometric sum for the final stage. By applying the renewal reward theorem to aggregate the expected transmission attempts (numerator) and the total expected time slots consumed (denominator) across all stages within a complete packet lifecycle, we obtain the final fractional expression:

$$\tau_{plc} = \frac{\sum_{i=0}^{m_1^k-2} t_i^k \prod_{j=0}^{i-1} (1 - s_j^k) + \prod_{i=0}^{m_1^k-2} (1 - s_j^k) \frac{t_{m_1^k-1}^k}{s_{m_1^k-1}^k}}{\sum_{i=0}^{m_1^k-2} bc_i^k \prod_{j=0}^{i-1} (1 - s_j^k) + \prod_{i=0}^{m_1^k-2} (1 - s_j^k) \frac{bc_{m_1^k-1}^k}{s_{m_1^k-1}^k}} \quad (A5)$$

This explicitly defines the logical origin of Equation (11) and maps the discrete backoff counter dynamics to the macroscopic channel performance metrics.

References

1. Zhou, X.; Zhao, Q.; Zhang, Y.; Yang, H. Analysis on Development Trend of China's Energy and Power System under the "Dual Carbon" Target: Green Power Substitution and Green Hydrogen Substitution. *Proc. CSEE* **2024**, *44*, 6707–6721.
2. Li, Z.; Li, W.; Zhang, Z.; Chang, G.; Lyu, J.; Yue, G.; Li, L.; Ni, W. Research on Development Strategy of Flexibility Resources of China's Power System under the "Dual Carbon" Goal. *Strateg. Study CAE* **2024**, *26*, 108–120. [\[CrossRef\]](#)
3. Wang, Q.; Zheng, Y.; Wu, J. Adaptive Load Balancing for Dual-Mode Communication Networks in the Power Internet of Things. *Electronics* **2023**, *12*, 4366. [\[CrossRef\]](#)
4. Dib, L.d.M.B.A.; Fernandes, V.; Filomeno, M.d.L.; Ribeiro, M.V. Hybrid PLC/Wireless Communication for Smart Grids and Internet of Things Applications. *IEEE Internet Things J.* **2018**, *5*, 655–667. [\[CrossRef\]](#)
5. Sayed, M.; Tsiftsis, T.A.; Al-Dhahir, N. On the Diversity of Hybrid Narrowband-PLC/Wireless Communications for Smart Grids. *IEEE Trans. Wirel. Commun.* **2017**, *16*, 4344–4360. [\[CrossRef\]](#)
6. Samir, A.; Elsayed, M.; El-Banna, A.A.A.; Wu, K.; ElHalawany, B.M. Performance of NOMA-Based Dual-Hop Hybrid Powerline-Wireless Communication Systems. *IEEE Trans. Veh. Technol.* **2022**, *71*, 6548–6558. [\[CrossRef\]](#)
7. Bianchi, G. Performance analysis of the IEEE 802.11 distributed coordination function. *IEEE J. Sel. Areas Commun.* **2000**, *18*, 535–547. [\[CrossRef\]](#)
8. Wu, Y.; Feng, M.; Zhu, X. An End-to-End Delay Analysis Method for Dual-Mode Communication Networks Combining PLC and RF Based on Martingale Theory. In Proceedings of the 2023 International Conference on Networks, Communications and Intelligent Computing (NCIC), Suzhou, China, 17–19 November 2023.
9. Dou, Z.; Bi, Y.; Yang, Z.; Du, J.; Yu, Y.; Wang, P. Research on Wireless and Power Line Dual-mode Communication Based on Adaptive Channel Selection. In Proceedings of the 2025 5th International Conference on Neural Networks, Information and Communication Engineering (NNICE), Guangzhou, China, 10–12 January 2025.
10. Zhu, Y.-H.; Gong, S.; Chi, K.; Li, Y.; Fang, Y. Optimizing Superframe and Data Buffer to Achieve Maximum Throughput for 802.15.4-Based Energy Harvesting Wireless Sensor Networks. *IEEE Internet Things J.* **2021**, *8*, 3689–3704. [\[CrossRef\]](#)
11. Das, K.; Moulik, S.; Chang, C.-Y. Priority-Based Dedicated Slot Allocation With Dynamic Superframe Structure in IEEE 802.15.6-Based Wireless Body Area Networks. *IEEE Internet Things J.* **2022**, *9*, 4497–4510. [\[CrossRef\]](#)
12. Luong, N.C.; Hoang, D.T.; Gong, S.; Niyato, D.; Wang, P.; Liang, Y.-C.; Kim, D.I. Applications of Deep Reinforcement Learning in Communications and Networking: A Survey. *IEEE Commun. Surv. Tutor.* **2019**, *21*, 3133–3174. [\[CrossRef\]](#)
13. De Rango, F.; Cordeschi, N.; Ritacco, F. Applying Q-learning approach to CSMA Scheme to dynamically tune the contention probability. In *Proceedings of the 2021 IEEE 18th Annual Consumer Communications & Networking Conference (CCNC), Las Vegas, NV, USA, 9–12 January 2021*; IEEE: New York, NY, USA, 2021.
14. Wu, C.-M.; Kao, Y.-C.; Chang, K.-F.; Tsai, C.-T.; Hou, C.-C. A Q-Learning-Based Adaptive MAC Protocol for Internet of Things Networks. *IEEE Access* **2021**, *9*, 128905–128918. [\[CrossRef\]](#)
15. Panda, N.; Supriya, M.; Elsts, A. Prioritized and Multi-Agent Reinforcement Learning-Based TSCH Schedulers. *IEEE Open J. Comput. Soc.* **2025**, *6*, 1763–1774. [\[CrossRef\]](#)
16. Yu, D.; Liu, X.; Ning, J.; Wang, S.; Zhu, C.; Zhao, W. Deep Reinforcement Learning-Based AI Task Offloading in Resource-Constrained IIoT Computing Environments. *IEEE Internet Things J.* **2025**, *12*, 54256–54273. [\[CrossRef\]](#)
17. Stiri, S.; Chaoub, A.; Grilo, A.; Bennani, R.; Lakssir, B.; Tamtaoui, A. Hybrid PLC and LoRaWAN Smart Metering Networks: Modeling and Optimization. *IEEE Trans. Ind. Inform.* **2022**, *18*, 1572–1582. [\[CrossRef\]](#)
18. Chen, Z.; Yan, Y.; Zhou, Z. Deterministic Delay Performance Optimization of Dual-Mode Communication Metering Network Based on Peak Age of Information Violation Probability. *Acta Electron. Sin.* **2025**, *53*, 4394–4407.
19. Cao, T. Cross-Layer Performance Analysis and MAC Layer Algorithm Design of Power Line Communication Relay System. Master's Thesis, North China Electric Power University, Beijing, China, 2024.
20. *IEEE Std 802.11-2020 (Revision of IEEE Std 802.11-2016)*; IEEE Standard for Information Technology—Telecommunications and Information Exchange between Systems—Local and Metropolitan Area Networks—Specific Requirements—Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. IEEE: New York, NY, USA, 2021. [\[CrossRef\]](#)
21. Chen, Z.; Zhi, L.; Chen, P.; Qi, Y. An MAC layer algorithm based on power line-wireless dual media channels and multiplexing. *China Commun.* **2022**, *19*, 78–91. [\[CrossRef\]](#)
22. *IEEE Std 1901-2020 (Revision of IEEE Std 1901-2010)—Redline*; IEEE Standard for Broadband over Power Line Networks: Medium Access Control and Physical Layer Specifications—Redline. IEEE: New York, NY, USA, 19 January 2021.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.