Article

Dynamic Multi-Sleeping Control with Diverse Quality-of-Service Requirements in Sixth-Generation Networks Using Federated Learning

Tianzhu Pan [ID], Xuanli Wu * and Xuesong Li [ID]

School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin 150001, China; 20B905048@stu.hit.edu.cn (T.P.); xuesong_l@stu.hit.edu.cn (X.L.)
* Correspondence: xlwu2002@hit.edu.cn; Tel.: +86-130-3004-5392

Abstract: The intensive deployment of sixth-generation (6G) base stations is expected to greatly enhance network service capabilities, offering significantly higher throughput and lower latency compared to previous generations. However, this advancement is accompanied by a notable increase in the number of network elements, leading to increased power consumption. This not only worsens carbon emissions but also significantly raises operational costs for network operators. To address the challenges arising from this surge in network energy consumption, there is a growing focus on innovative energy-saving technologies designed for 6G networks. These technologies involve strategies for dynamically adjusting the operational status of base stations, such as activating sleep modes during periods of low demand, to optimize energy use while maintaining network performance and efficiency. Furthermore, integrating artificial intelligence into the network's operational framework is being explored to establish a more energy-efficient, sustainable, and cost-effective 6G network. In this paper, we propose a small base station sleeping control scheme in heterogeneous dense small cell networks based on federated reinforcement learning, which enables the small base stations to dynamically enter appropriate sleep modes, to reduce power consumption while ensuring users' quality-of-service (QoS) requirements. In our scheme, double deep Q-learning is used to solve the complex non-convex base station sleeping control problem. To tackle the dynamic changes in QoS requirements caused by user mobility, small base stations share local models with the macro base station, which acts as the central control unit, via the X2 interface. The macro base station aggregates local models into a global model and then distributes the global model to each base station for the next round of training. By alternately performing model training, aggregation, and updating, each base station in the network can dynamically adapt to changes in QoS requirements brought about by user mobility. Simulations show that compared with methods based on distributed deep Q-learning, our proposed scheme effectively reduces the performance fluctuations caused by user handover and achieves lower network energy consumption while guaranteeing users' QoS requirements.

Keywords: 6G; network energy saving; base station sleeping; federated learning

1. Introduction

With the emergence and rise of various new smart devices and applications, the traffic in future mobile communication networks is expected to increase significantly. According to a report by the International Telecommunication Union (ITU), user data traffic per month is projected to soar to 5016 exabytes (EB) by 2030, which is about 20 times that of 2023 [1]. To meet such data service demands, the coverage of future sixth-generation (6G) mobile networks will need to be more extensive, the network scale more massive, and the density of base stations further increased. Concurrently, to align with the vision of building an environmentally friendly society and to meet the strategic goals of reaching "carbon peak" and attaining "carbon neutrality", 6G networks are expected to embody green, low-carbon,

cost-effective, and sustainable characteristics [2]. To fulfill the requirements for the 2030 carbon peak objectives, the deployment of 6G networks will have to be managed with minimal increases in network energy consumption compared to existing networks. With the future data traffic projected to grow by several tens of times and under the premise of unchanged total energy consumption, the energy efficiency of 6G networks will need to be improved by several tens of times at least, compared to the current fifth-generation (5G) mobile networks [3].

In actual network scenarios, the power consumption of base stations does not change linearly with the service load. Instead, when a base station is idle (with no load), its power consumption still accounts for 50–60% of the power consumption under full load conditions [4]. This implies that the energy efficiency of base stations is quite low under light-load conditions. Therefore, researchers have proposed that during periods when base stations are not fully loaded, some of the base station's functionalities and hardware components should be shut down to significantly reduce power consumption while still meeting transmission requirements. Ideally, the concept of "0 bit, 0 watt" should be realized, which means that the base station would consume almost zero power when there is no data transmission.

In this paper, we investigate the sleep control problem of base stations, and the remaining of this paper is organized as follows. The related works on base station sleep control are described in Section 2. The system model of the heterogeneous ultra-dense networks and the energy delay trade-off problem is described in Section 3. Federated reinforcement learning for small base station sleep control is introduced in Section 4. In Section 5, the performance of the proposed scheme is evaluated. Finally, Section 6 provides the discussion of this study. The key mathematical notations used in this paper are defined in Table 1.

Table 1. Summary of key notations.

Notation	Description
$\mathcal{M}$	Set of small base stations
$\mathcal{K}$	Set of users
m	The mth small base station
k	The kth user
$r_{m,k}$	Data rate of user k served by small base station m
B_m	Total bandwidth of small base station m
K_m	Number of users served by small base station m
p_m	Transmission power of small base station m
$h_{m,k}$	Channel gain between small base station m and user k
N_0	Thermal noise power spectral density
r_m	Total data rate of small base station m
t_i	The ith time slot
$DP_{m,k}^i$	Data packet of user k served by small base station m arriving at time slot t_i
$v_{m,k}^i$	Remaining data volume of data packet $DP_{m,k}^i$
β_m	Working state indicator of small base station m
L_k	Lifetime of data packets of user k
$d_{m,k}^i$	Remaining lifetime of data packet $DP_{m,k}^i$
t	Current time slot
$\mathcal{B}_m$	Buffer status of small base station m
v_m	Total data volume in the buffer of small base station m
$l_{m,t}$	Packet loss rate of small base station m at time slot t
$p_{m,t}$	Power consumption of small base station m at time slot t
w	Trade-off factor
$\mathcal{S}_m$	State space of small base station m
$\mathcal{A}_m$	Action space of small base station m
$\mathcal{P}_m$	Transition probabilities of small base station m
$\mathcal{R}_m$	Reward function of small base station m
I_m	Interference level of small base station m
θ_m^t	Network parameter of the evaluated Q-network of small base station m
$\hat{\theta}_m^t$	Network parameter of the target Q-network of small base station m
Q	Q-value
γ	Discount factor
s_m^{t+1}	State of small base station m at time slot $t+1$
a_m^t	Action of small base station m at time slot t
y_m^t	Target value
α	Learning rate
g_r	Network parameters of the global model
λ	Step size of local model update
L	Loss function
$\mathcal{D}_m$	Experience replay buffer of small base station m

2. Related Works

In recent years, base station sleeping technology has attracted attention and is regarded as one of the key technologies for enhancing the energy efficiency of future networks. Based on the number of sleep levels, base station sleep mechanisms can be categorized into binary sleep mechanisms and multi-level sleep mechanisms [5]. In the binary sleep mechanism, a base station typically has an active mode and one sleep mode. The active mode is where the base station is fully operational and can serve users with full capabilities. The sleep mode is a low-power state where the base station consumes minimal energy but cannot serve users or serve them with limited capabilities. The base station switches between these two states based on network traffic demands.

In early research, scholars hoped to save energy by controlling the sleep of base stations. For instance, the authors of [6] outlined methods driven by small cells, core networks, and user equipment, noting energy savings between 10% and 60%. The research in [7] explored the trade-off between energy consumption and delay using queuing theory, suggesting a sleep control strategy that optimizes both. Despite these advancements, the widespread implementation of such strategies in 4G networks has been limited due to coverage continuity concerns [8]. With the advent of 5G and ultra-dense networks, there is renewed interest in base station sleep for energy efficiency. Innovative sleep models and artificial intelligence have paved the way for dynamic optimization solutions that adapt to network changes, as highlighted in [3]. Traffic prediction models have become crucial for energy-saving strategies, with various algorithms being compared in [9], and AI-driven approaches like deep neural networks and recurrent neural networks being utilized for predictive analytics in [10]. However, base station sleeping impacts not only the individual base station but also the surrounding network infrastructure, as discussed in [11]. Complexities arise when designing metrics and thresholds to balance the local and network-wide effects of sleeping strategies, especially as networks grow in size and complexity. Advanced solutions now involve game-theoretic approaches to minimize system power consumption while considering base station load [12], as well as dynamic control mechanisms for enhanced energy efficiency in millimeter-wave networks [13]. The intricacies of business load-based sleep in ultra-dense networks (UDNs) were examined in [14], and various sleeping strategies and their trade-offs were analyzed in [15]. AI and machine learning have revolutionized base station sleeping decisions. Neural networks can now predict non-critical stations for sleep initiation, offering a comprehensive solution for optimizing network energy efficiency [16]. Frameworks like MGCN-LSTM have improved traffic predictions [17], and Q-learning has been employed to balance load and energy efficiency [18]. Furthermore, a new data-driven framework has been proposed to manage base station sleep modes, satisfying both energy and quality-of-service (QoS) requirements, with simulations showing significant performance advantages [19]. Multi-agent reinforcement learning methods are being used to optimize access control and base station sleeping in scenarios where human communication and machine communication co-exist [20]. However, the binary sleeping mechanism only allows base stations to remain active or be completely asleep, which often requires a longer transition time and lacks the flexibility to adjust the transmission capacity of the base station dynamically. This rigidity can easily lead to a significant decline in users' QoS.

In a multi-level sleep mechanism, unlike hierarchical sleep mechanisms, there are several levels of sleep depth, each with varying degrees of energy saving and operational readiness. The deeper the sleep level, the more energy is conserved, but the longer it may take for the base station to become fully operational again. This mechanism allows for more granular control of energy consumption, balancing the trade-off between energy savings and QoS [21]. The base station can dynamically adjust its sleep depth based on real-time traffic conditions and predefined policies to optimize both energy consumption and user experience. To enable base stations to enter the appropriate sleep mode, a Q-learning-based algorithm was proposed in [22] to determine the optimal duration that BS can consume at each sleep level to minimize the network energy consumption

under BS sleep/activation delay and user service demand constraints. Ali Ei Amine and colleagues studied multi-level base station sleep mechanisms and proposed a distributed Q-learning approach to control the depth of sleep to achieve a trade-off between power saving, average user delay, and packet loss rate [23,24]. Furthermore, Ali Ei Amine et al. further researched the sleep problem of small base stations in dual-layer heterogeneous networks, considering co-channel interference and user offloading, and achieved the joint optimization of the weighted sum of small base station power consumption and data loss [5]. Meysam Masoudi et al. proposed an online reinforcement learning algorithm to decide in real time the optimal BS during inactive tuning [25], improving the practicality of the algorithm. Sheng Lin et al. performed a rather precise modeling of the user experience delay during base station sleep, including waiting delay and transmission delay, and based on a real 5G dataset, proposed a sequential DRL mechanism for user association and base station state selection. Their simulation results show that the sequential DRL algorithm has a better energy-saving effect than the single DRL algorithm, saving more than 50% of energy compared to greedy, random, and other algorithms [26]. Considering that the AI-based base station state selection mechanism may fail due to insufficient training or changes in network status, leading to an unacceptable severe decline in user QoS, Meysam Masoudi et al. used a Markov process to model the dynamic changes in the system state (including base station status and user status) and proposed a risk-aware base station mode management method assisted by a digital twin. This method ensures that the AI-based sleep decision can only be used when the risk caused by entering base station sleep is less than the risk threshold set by the operator. If the risk exceeds the threshold, it indicates that the model has failed, the sleep decision should be suspended, and retraining should be carried out to ensure the reliability of the base station sleep mechanism [27]. The above studies focus only on the selection of base station sleep modes and do not consider the power control of the base stations, assuming that base stations use all their transmission power, without considering the impact of base station power control on coverage. In fact, studies have shown that base station power control should be jointly optimized with sleep modes [28]. Chang et al. proposed a multi-base station joint downlink power optimization algorithm, where base stations, after determining sleep decisions, choose the appropriate power level based on the user's transmission traffic volume and signal quality using a DQL approach to optimize the system's long-term energy efficiency [29]. However, most of the aforementioned studies are based on distributed learning methods, as mentioned in [5,24]. In these methods, each base station independently trains and makes decisions based on its own observations, and it is assumed that users are either stationary or semi-stationary (moving within the coverage area of a single base station without switching). Once the base station's sleeping decision-maker is sufficiently trained, it can accurately select the optimal sleeping decision. In real-life dense network scenarios, users continuously move and switch, and the quality-of-service requirements of the users served by each base station are in a state of dynamic change. The model previously learned may no longer perform optimally when users leave or join, leading to a decline in performance. This decline in performance due to the different distributions of the training and testing datasets is referred to as concept drift. An overview of the multi-level sleep strategies is shown in Table 2.

In this paper, we propose a federated reinforcement learning-based (FRL-based) base station sleeping control strategy, in which the small base stations can turn into sleep modes when the traffic demand is low to save energy. The main contributions of this paper can be summarized as follows.

In order to more accurately evaluate the impact of base station sleeping on users' quality-of-service, particularly regarding latency, we advanced beyond the approach in [27], which only uses the numbers of users to describe base stations' load. We modeled the transmission requirements and channel capacity for each user individually. Unlike [5,26], which focuses solely on delay-tolerant services, we considered both delay-tolerant and delay-sensitive services. Contrasting the method in [5], which only looks at total throughput

and the buffer state of small base stations, our user-specific model allows more accurate measurements of performance indicators like packet loss rate and transmission delay.

Table 2. Overview of multi-level sleep strategies.

Ref.	Scenario	Method	Contributions	Limitations
[22]	Single BS	QL	Power and delay trade-off	Only a single BS; the impact of inter-cell interference is ignored
[23,24]	MBS, SBSs, and fixed users	QL	Power and delay trade-off	Ignored the impact brought by user movement and handover
[25]	Single BS	QL	Power and delay trade-off	The load of the BS is roughly quantified into only two levels: high and low. It only considers a single BS and the inter-cell interference is ignored
[26]	SBSs and moving users	DQL	Jointly optimizes the user association and SM control to minimize the energy-delay reward	The decision cycle is long; unable to support service types with a latency requirement of less than 100 ms
[27]	Two BSs and fixed users	DQL	Assessed the risk of model performance degradation with environmental changes and provided a retraining mechanism	User mobility is not considered; the BS's load is described solely based on the number of users
[29]	SBSs and fixed users	DQL	Power and SM control to maximize energy efficiency	Only optimizes the system's energy efficiency while ignoring the increased delay caused by BS sleeping
[5]	MBS, SBSs, and moving users	DQL	Determined the optimal sleep mode to minimize power consumption and packet loss rate, providing a user offloading scheme in heterogeneous networks	Users are restricted to moving within the cell; the impact of inter-cell handover is not considered, and only delay-tolerant services are taken into account

We considered a more practical scenario where users are continuously moving, with movement trajectories sourced from real datasets. Compared to the assumptions in [23,24,27,29] where users are stationary and the assumption in [5] where users only move within a cell and do not handover. Our approach evaluates the impact of user movement and handover on system performance. To address the concept drift caused by user mobility and handover, we integrated a federated learning framework, offering better performance and convergence compared to the distributed reinforcement learning method proposed in [5,24].

Through numerical simulations, we analyzed the performance of the proposed algorithm under various network conditions. The simulations demonstrate that our federated reinforcement learning-based approach has significant advantages in mitigating concept drift and accelerating model convergence. In particular, compared to the independent deep Q-learning scheme proposed in [24,26], when there are delay-sensitive service users in the system, our algorithm enables the system to achieve lower power consumption and packet loss rate.

3. System Models and Problem Formulation

3.1. Network Scenario

We consider a two-tier heterogeneous ultra-dense network composed of one macro base station, M small base stations, and K users, as shown in Figure 1. The small base stations and users are randomly distributed within the coverage area of a macro base station, following two independent Poisson Point Processes. The small base stations are connected to the macro base station through the X2 interface to facilitate data backhaul and signaling interaction. It is assumed that all user data are transmitted by small base stations, while the macro base station exclusively handles the transmission of control signals. Furthermore, all small base stations are capable of sharing the whole system bandwidth B. The set of SBSs and users are denoted by $\mathcal{M}$ and $\mathcal{K}$, respectively.

To reduce network energy consumption, small base stations can enter different sleep modes during off-peak periods. These stations may deactivate certain network functions and hardware components to lower power usage. In order to save energy without significantly compromising user QoS, small base stations need to select appropriate sleep modes. Optimizing the sleep control problem for small base stations is a highly complex integer programming problem, requiring extensive searching to find the optimal combination of sleep modes. Therefore, to solve this problem in real time, methods based on reinforcement learning are often used. In this approach, small base stations are treated as agents that continuously interact with the environment to learn near-optimal action-state mappings, thus achieving near-optimal sleep control. This methodology is advantageous

for its adaptability and real-time response, allowing sleep strategies to adjust according to the real-time changes in the network environment. However, in the mechanism of distributed reinforcement learning, each small base station learns the action–state mapping that best matches the characteristics of its users over a previous time frame. As users move and switch, the data distribution during training might not match the distribution during deployment. Consequently, the experience acquired during training may become outdated, leading to what is known as concept drift. To mitigate the issue of concept drift, we have introduced a federated learning mechanism. In this system, small base stations act as agents that conduct reinforcement learning locally to develop a local model. These local models are then periodically uploaded to the macro base station, serving as the central controller. The macro base station aggregates these local models into a global model and then disseminates it back to the small base stations for model updating. This approach helps in synchronizing the learning process across the network and maintaining the relevance of the models in a dynamically changing environment.
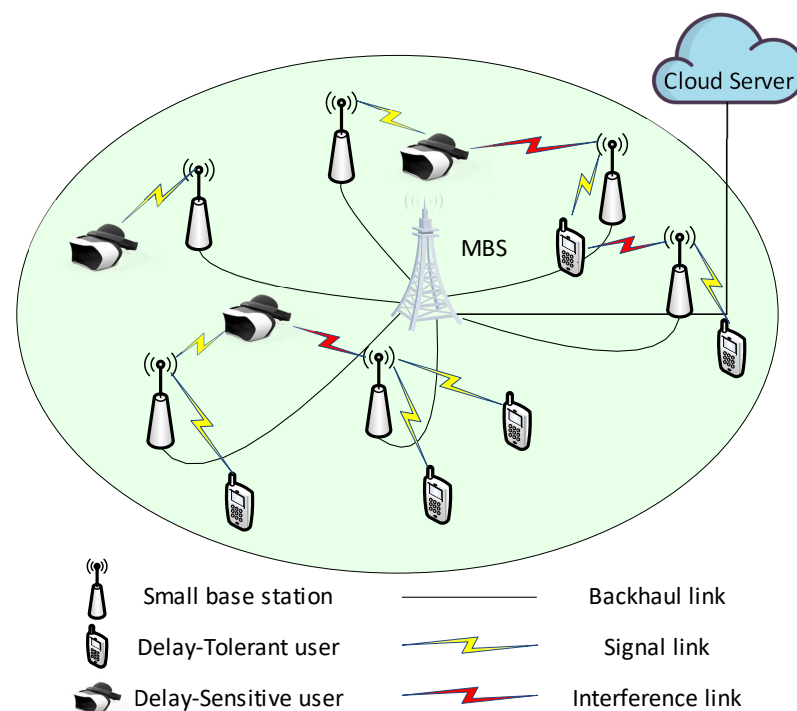

Figure 1. Illustration of the network architecture of a two-tier heterogeneous ultra-dense network.

3.2. Multi-Level Sleep Modes

To model the sleep modes of base stations, we utilize the sleep strategy proposed by GreenTouch in [21], wherein the following four sleep modes are defined:

- SM1: This mode shuts down the shortest time unit, which is an OFDM symbol (71 μs), including the time for shutdown and reactivation. In this mode, only the power amplifier and some processing components are disabled.
- SM2: This mode corresponds to the subframe shutdown, with a shutdown duration of a time slot interval TTI (1 ms). In this mode, the components including the RF channels can be turned off to address more energy.
- SM3: This mode corresponds to the frame shutdown, with a duration of 10 ms. In this mode, deeper shutdown, including carrier shutdown, can be achieved.
- SM4: The deepest sleep level, in which almost all base station components are turned off and only the necessary components for reactivating the base station are retained.

This includes active modes and sleep modes (symbol shutdown, subframe shutdown, frame shutdown, and deep sleep, which can cause significant delay increases and need to

be carefully designed). Higher sleep mode levels can shut down additional base station functions and components, thereby achieving lower power consumption. However, they also necessitate a longer time to reactivate (or deactivate) the base station, potentially causing increased user latency and elevating the risk of compromising user QoS requirements. To support the design of multi-level sleep technology, Ref. [21] provides a modeling method for key indicators such as the power consumption and transition duration of each sleep level. The transition duration conforms to the time-domain frame structure defined by 3GPP, while the power consumption of each sleep level can be determined based on parameters like the type of base station, the number of transmit and receive antennas, and bandwidth. In practical use, these parameters need to be incorporated into complex formulas to obtain the corresponding performance indicators. Fortunately, IMEC Green-Touch also offers an online tool that simplifies the modeling work significantly. This tool allows users to obtain performance indicators for each sleep level based on key parameters configured for the base station. The relevant parameters in this paper are sourced from this online tool [30]. Specifically, we have outlined the characteristics of the sleep levels for 10 MHz 2 × 2 MIMO small base stations in Table 3. This includes details such as the power consumption in both active and sleep modes, and the duration required for activation or deactivation. However, according to the current version of the 3GPP protocol, 5G base stations are required to periodically transmit the Synchronization Signal/Physical Broadcast Channel. The transmission period can be configured to [5, 10, 20, 40, 80, 160] ms. Therefore, in this version, SM4 (Sleep Mode 4) cannot be utilized.

Table 3. Sleep modes.

Mode	Active		SM1	SM2	SM3	SM4
	Full	Idle				
Power	22.7 W	11.7 W	7 W	1.95 W	1.1 W	0.75 W
(De)activation duration	0	0	35.5 µs	0.5 ms	5 ms	0.5 s
Minimum sleep duration	0	0	71 µs	1 ms	10 ms	1 s

The base stations should select appropriate working modes based on the traffic demands and QoS requirements of the users, channel conditions, and interference from neighboring areas. This approach can enhance the system's energy efficiency without significantly reducing the QoS experience for users. This balance is crucial for maintaining network performance that is both energy-efficient and meets user demands.

3.3. Transmisstion and Traffic Model

Given a cell bandwidth B_m and the noise spectral density N_0, the data rate of user k served by SBS m, the cell average data rate, and the number of bits served by SBS m can be calculated as follows:

$$r_{m,k} = \frac{B_m}{K_m} \log_2 \left(1 + \frac{p_m \cdot h_{m,k}}{N_0 \cdot B_m + \sum'_m p_{m'} \cdot h_{m',k}} \right), \tag{1}$$

where p_m is the transmission power of the small base station m, $h_{m,k}$ is the channel gain between the small base station m and user k, and K_m is the number of users served by the small base station m. It is clear that the user data rate is associated with the wireless channel gain, the bandwidth of the small base station, the number of serviced users, and the interference from other base stations. Let r_m denote the total data rate of the small base station m, which can be calculated by

$$r_m = \sum_{k \in \mathcal{K}_m} r_{m,k}. \tag{2}$$

In this paper, we consider two types of user services: delay-tolerant and delay-sensitive services. These categories are significant in the design and optimization of network archi-

tectures, especially in advanced networks like 5G, which promise to support a wide variety of applications with different performance requirements:

- Delay-tolerant services are those where transactions and communications can experience latency without significantly impacting the user's experience. These services are less affected by network congestion and can be scheduled to use network resources during off-peak times to improve overall network efficiency. Typically, delay-tolerant services can accommodate latencies of tens of milliseconds or even beyond 100 ms.
- On the other hand, delay-sensitive services require a fast and often real-time transmission of data to function correctly. High latency can severely impact the quality of these services, leading to poor user experience. For delay-sensitive services, the acceptable latency is often below 10 ms. Examples of delay-sensitive services include real-time voice and video communication, online gaming, and certain forms of interactive applications.

Users' traffic is modeled using an on–off model, where user activity (on state) alternates with periods of inactivity (off state). The on state represents the time during which the user is actively sending or receiving data, while the off state corresponds to the time when there is no data transmission. At the beginning of each time slot, new data packets enter the small base station's buffer. Therefore, the arrival and lifetime of data packets can be illustrated, as shown in Figure 2. The small base station then transmits the data packets to the users it serves over the wireless channel. Let $DP^i_{m,k}$ denote the data packet of the user k served by the small base station m arriving at time slot t_i. Each data packet has a lifetime L_k, which depends on the service type of the user k. It is assumed that the data volume of a $DP^i_{m,k}$, V is fixed. If there is a data packet $DP^i_{m,k}$ in the buffer of the small base station m, then at a later time t, its remaining data volume $v^i_{m,k}$ and lifetime $d^i_{m,k}$ would be

$$v^i_{m,k} = V - \sum_{t_i}^{t} \beta_m r_{m,k} \tag{3}$$

and

$$d^i_{m,k} = L_k - (t - t_i), \tag{4}$$

where β_m denotes the working state indicator of the small base station m, $\beta = 1$ signifies that the small base station m is in active mode, and $\beta = 0$ indicates that the base station is in one of the sleep modes. If the packet fails to complete transmission after exceeding the lifetime upon entering the buffer zone ($d^i_{m,k} < 0$), it will become invalid, leading to packet loss. The base station is assumed to implement a delay-aware scheduling mechanism that prioritizes the transmission of data packets close to their timeout threshold.
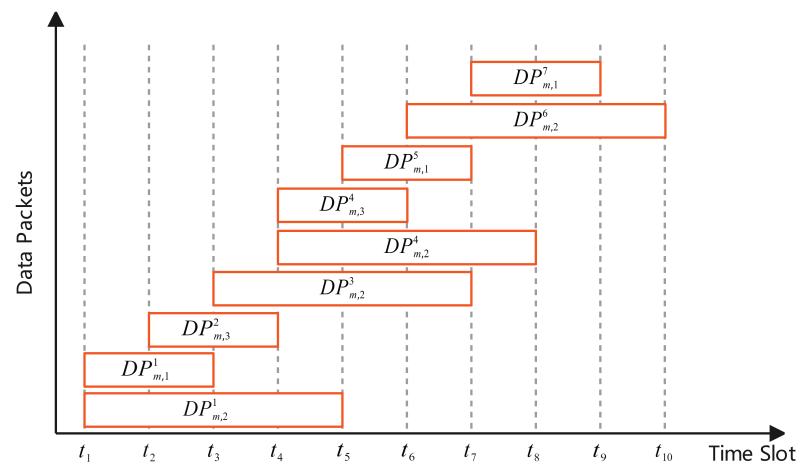

Figure 2. Illustration of the arrival and lifetime of data packets.

The buffer status of a base station can be modeled as the set of all valid data packets in the buffer that have not yet completed transmission:

$$\mathcal{B}_m = \{DP^i_{m,k} | d^i_{m,k} > 0, \forall k \in \mathcal{K}_m\}. \tag{5}$$

The total data volume v_m in the buffer of the small base station m can be calculated as

$$v_m = \sum_{DP \in B_m} v^i_{m,k}. \tag{6}$$

When the small base station is in sleep mode, data transmission is paused, but new data packets can still arrive. Figure 3 provides an illustrative representation of the dynamic changes in the base station's buffer state.
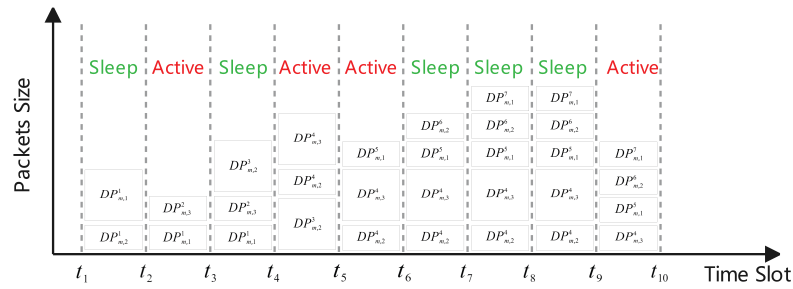

Figure 3. Illustration of the arrival and transmission of data packets in the buffer.

3.4. User Moving and Association

In this paper, we use the best RSRP user association strategy. As users move and the wireless environment changes, when a base station has a higher RSRP than a user's current service base station, inter-cell handover occurs. To investigate the impact of user mobility on base station sleep, we used real user trajectories in our research. Our user trajectory data come from the open-source dataset, GeoLife GPS Trajectories, released by Microsoft, which collects real user movement trajectories from the real world.

3.5. Power Model and Problem Formulation

The power consumption of small base stations under sleep modes has been provided in Table 2. For base stations in the active state, we model their power consumption with the following linear model:

$$p^{total}_m = \rho_m(p_{Full} - p_{Idle}) + p_{Idle}, \tag{7}$$

where p_{Full} and p_{Idle} represent the power consumption of the base station under full load and idle conditions, respectively. ρ_m is the load ratio of the small base station m.

Based on the system model, the SBS sleep control problem can be formulated as follows:

$$\min_{a \in \{Active, SM1, SM2, SM3\}} \sum_{t=1}^{+\infty} ELR = \sum_{t=1}^{+\infty} \sum_{m=1}^{M} \omega \cdot l_{m,t} + (1-\omega) \cdot p_{m,t}, \tag{8}$$

where $l_{m,t}$ is the packet loss rate of the small base station m at time slot t. Packet loss means that the data packet has not been transmitted within its lifetime, ω is the trade-off factor used to quantify the priority of energy conservation and ensure the packet loss rate. The goal is to jointly minimize the long-term energy-loss reward (ELR), which is defined as the sum of the total packet loss rate and the total power consumption of the small base stations. To reach this goal, the small base stations need to choose the appropriate operating mode at the beginning of each time slot based on their own buffer status, channel conditions, and other factors.

4. Federated Reinforcement Learning for Small Base Station Sleep Control

To reduce power consumption while avoiding packet loss, a multi-level base station sleep control scheme based on reinforcement learning was proposed in [5], where small base stations independently learn the optimal active control strategy to minimize the weighted sum of power consumption and packet loss, thereby achieving a trade-off between energy saving and user QoS. However, in [5], users are assumed to move only within their serving cell, that is, there is no handover between the small cells. Moreover, the packet reception rate of each base station depends solely on the number of users it serves, and under this assumption, the arrival and lifetime distribution of the packets to the small base stations remain constant. Once the deep Q-network is trained, the small base stations can obtain an optimal sleep control strategy adapted to their service demands.

However, such assumptions are challenging to guarantee in real-world scenarios. For instance, the service demands of users do not arrive at a constant rate; each user has different data rate and latency requirements, leading to variable service demands for small base stations. Furthermore, as users move, there are continuous inter-cell handovers between small base stations, causing the user set served by each small base station to dynamically change. This change affects the arrival rate and lifetime of packets at the small base stations. The dynamic variation in the user set served by small base stations, changes in wireless channels, and user QoS requirements may cause a data distribution shift between the samples collected during the training of the reinforcement learning for optimal sleep control strategy and the samples processed after base station deployment. This discrepancy between training and test data can lead to concept drift, causing the trained base station sleep control strategy to deviate from the optimal strategy. This deviation can lead to inaccurate or erroneous decisions, reducing the energy-saving effectiveness or service quality of the base stations and even causing severe packet loss.

In this paper, we introduce a federated learning approach to mitigate concept drift. Federated learning is a distributed machine learning framework that enables multiple nodes (for example, mobile devices or small base stations) to collaboratively train a shared model while keeping the data localized. In addressing concept drift caused by dynamic network changes, federated learning offers several advantages:

- Data diversity: Federated learning aggregates learned information across different devices and users, which increases the model's generalization capability by considering the variability in environments and user behaviors.
- Real-time updates: Base stations can update their local models and exchange parameters with a central server or other base stations periodically, keeping the global model current with the latest network conditions and user behaviors.
- Continuous learning: The federated network engages in ongoing learning to better adapt to dynamic changes in channels and user demands, thereby minimizing the impact of concept drift.

Markov Decision Process Model of Multi-Level Sleeping Control

In this paper, the multi-level sleep control for each small base station is modeled as a four-tuple Markov Decision Process (MDP) $(\mathcal{S}_m, \mathcal{A}_m, \mathcal{P}_m, \mathcal{R}_m)$, where $\mathcal{S}_m$ is the state space of the small base station m, $\mathcal{A}_m$ is the action space of the small base station m, $\mathcal{P}_m$ is the transition probability from the current state S_m^t to the next state S_m^{t+1}, and R_m is the reward function. In the base station sleep control strategy studied in this paper, a small base station selects an action a based on the current state S_t, according to a certain policy. Subsequently, the base station's state transitions to S_m^{t+1}, and it receives a reward. The small base station continuously updates its policy based on the received reward using reinforcement learning methods until it converges to an optimal or near-optimal sleep control policy. Accordingly, aligned with the features of the problem, the state space $\mathcal{S}_m$, action space $\mathcal{A}_m$, transition probability $\mathcal{P}_m$, and reward function R_m are defined as follows:

- The state space $\mathcal{S}_m = \{r_m, v_m, I_m\}$ represents a combination of the current achievable data rate, buffer data amount, and the level of interference for the small base station m.

Here, the current achievable data rate can be calculated using Equations (1) and (2), the buffer data amount can be calculated using Equation (6), and the level of interference represents the proportion of throughput loss due to interference and is expressed by the following equation:

$$I_m = \frac{\log_2(1 + \frac{p_m \cdot h_{m,k}}{N_0 \cdot B_m + \sum'_m p_{m'} \cdot h_{m',k}})}{\log_2(1 + \frac{p_m \cdot h_{m,k}}{N_0 \cdot B_m})}. \tag{9}$$

- The action space $\mathcal{A}_m = \{Active, SM1, SM2, SM3\}$ consists of possible operational states for the small base station m, such as the active mode and three sleep modes. The small base station m selects an action from $\mathcal{A}_m$ based on the ϵ-greedy policy at the beginning of each time slot.
- The transition probabilities $\mathcal{P}_m(s'|s,a)$ represent the likelihood of moving from state s to state s' when action a is taken.
- The reward function $R_m(s,a)$ quantifies the immediate benefit of taking action a in state s, typically balancing energy consumption against service quality. Considering the characteristics of problem (8), the reward function is defined as

$$R_m = -ELR_m = -(\omega \cdot l_{m,t} + (1 - \omega) \cdot p_{m,t}). \tag{10}$$

The base station can calculate rewards based on power consumption and the packet loss rate in time slot t.

To address the issue of "concept drift" caused by user mobility, this paper proposes an FRL-based small base station sleep control scheme based on the federated learning framework proposed in [31,32]. The process flow of the small base station sleep control scheme based on federated learning is shown in Figure 4, and the execution of this scheme can be summarized as follows: the small base stations first train their local models independently based on their environment; then, they periodically upload their local models to a central data unit for model aggregation to obtain a global model; once the aggregation is performed, the central data unit broadcasts the global model to all small base stations and; upon receiving the global model, the small base stations update their local models and continue with the next round of local model training. The details of this scheme are as follows.
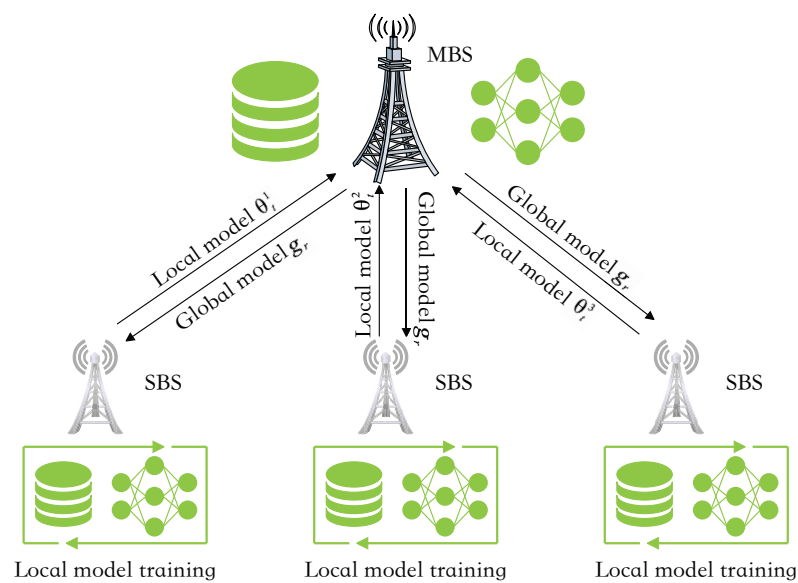

Figure 4. Flowchart of federated learning-based small base station sleep control scheme.

1. Local Model Training

    The small base stations initially engage with their environment to refine their local models, targeting improved effectiveness in present conditions. This research utilizes the discrete action deep reinforcement learning algorithm, double deep Q-network (DDQN), for local model training, considering the expansive state space and compact action space in base station sleeping scenarios. DDQN, with its replay buffer and the separation of action selection from its evaluation, tackles the challenges posed by the vast state space by diminishing data correlation. The update mechanism of DDQN is outlined as follows:

    $$y_m^t = R_{t+1} + \gamma Q\left(s_m^{t+1}, \arg\max_{a_t^m} Q\left(s_m^{t+1}, a_m^t; \theta_m^t\right); \hat{\theta}_m^t\right), t = 1, 2, 3, \cdots, \quad (11)$$

    where

    $$a_m^t = \begin{cases} \arg\max_{a_m^t \in \mathcal{A}} Q\left(s_m^{t+1}, a_m^t; \theta_m^t\right), & \text{if } \sigma > \epsilon, \\ rand(\mathcal{A}), & otherwise \end{cases}, \forall m \in \mathcal{M}, t = 1, 2, 3, \cdots \quad (12)$$

    is the ϵ-greedy policy used for base station sleeping control, where $\sigma \sim U(0, 1)$ is a random number that follows a uniform distribution and ϵ is a preset probability threshold. θ_t^m and $\hat{\theta}_t^m$ are the network parameters for the Q-network and the target network, respectively, and $\gamma \in [0, 1]$ is the discount factor. For a base station m, if it is in state s_t^m and takes action a_t^m at time slot t, its state–action value is given by

    $$Q\left(s_m^t, a_m^t\right) = \mathbb{E}\left[\sum_{n=t}^{T} \gamma^n r_m^t | s_m^t, a_m^t\right], t = 1, 2, 3, \cdots. \quad (13)$$

    The objective of DDQL is to minimize the gap between Q and $\hat{\theta}$, i.e., the loss function. Here, the loss function of the DDQN, *L*, is defined as

    $$L\left(\theta_m^t\right) = \mathbb{E}\left[\left(y_m^t - Q\left(s_m^t, a_m^t; \theta_m^t\right)\right)^2\right], t = 1, 2, 3, \cdots. \quad (14)$$

    Furthermore, gradient descent is employed to update the network parameters θ_t^m in DDQL. The update rule in DDQN is given by

    $$\theta_m^{t+1} = \theta_m^t + \alpha\left[y_m^t - Q\left(s_m^t, a_m^t; \theta_m^t\right)\right]\nabla Q\left(s_m^t, a_m^t; \theta_m^t\right), t = 1, 2, 3, \cdots, \quad (15)$$

    where α is the learning rate and τ is the period of model aggregation. After training for τ time slots locally, the small base stations send the model parameters θ_t^m to the central data unit for global model aggregation.

2. Global Model Aggregation

    The global model aggregation occurs at the central data unit. Once the local models θ_t^m from all small base stations are received, the global model is calculated using:

    $$g_r = \frac{\sum_{m \in \mathcal{M}} K_m \theta_m^t}{K}, \quad \mod (t, \tau) = 0, \quad (16)$$

    where $K = \sum_{m \in \mathcal{M}} K_m$ represents the total number of training samples, K_m stands for the total number of training samples at the base station m, and g_r is the global model in the rth aggregation round. After the global model aggregation, the central data unit broadcasts the obtained global model g_r to all small base stations for updating their local models.

3. Local Model Update

The local model update process occurs at the small base station. After receiving the global model from the central data unit, the small base station updates its local model based on the global model, which serves as the foundation for subsequent local model training. Let us assume each communication period consists of τ time slots. We denote θ_t^m as the dormant model of base station m at time slot t and g_r as the global model in the rth round of communication. At the beginning of the rth communication cycle, the small base station m receives the global model g_r broadcasted by the central node. The local model is then updated as

$$\theta_m^{t+1} = g_r - \frac{\lambda}{K_m} \sum_{m=1}^{M} \nabla L(\theta_m^t), \quad \mod (t, \tau) = 0, \tag{17}$$

where λ is the step size. After the model update, the small base station interacts with the environment and performs the next round of local model training.

The workflow of the proposed federated reinforcement learning-based base station sleeping control algorithm can be summarized as follows. Each small base station trains a local Q-network to guide base station sleeping decisions based on the reinforcement learning framework. When the model aggregation process starts, each small base station uploads its local Q-network parameters (including network weights, biases, etc.) to the macro base station. The macro base station aggregates the collected local models based on the aggregation rule shown in Equation (16) to obtain a global model. This global model is then shared with each small base station, allowing them to update their models and perform the next round of local training on the basis of the global model. This helps the small base stations to obtain Q-networks that are more suitable for the characteristics of all users in the network. The details of the proposed algorithm are given in Algorithm 1.

As analyzed in references [31,32], since each communication round involves model aggregation, model updates, and τ rounds of local training, the computational complexity of the proposed algorithm is $\mathcal{O}(J(M + \tau))$, where J represents the number of communication rounds. The local model training process can be considered similar to the independent reinforcement learning described in references [5,26], where each small base station maintains its own experience replay buffer and Q-network, with no information exchange between the small base stations or between small and macro base stations. The model aggregation process can be seen as a process of "experience sharing" among the small base stations through the global model aggregated by the macro base station. In this process, the small base stations transmit local model parameters to the macro base station, and then the macro base station transmits the global model parameters back to the small base stations. This will incur certain signaling overheads, but in the network architecture of this paper, this interaction process is carried out through the X2 interface connected by an optical fiber between the base stations, making the communication cost acceptable. The framework of base station sleeping decisions based on federated learning proposed in this paper, compared to a completely independent reinforcement learning approach, can bring significant performance advantages.

As local model training, global model aggregation, and global model updates alternate, the small base stations continually update their deep Q-network, allowing the sleeping control strategy to dynamically adapt to changes in the network environment. The details of the proposed federated reinforcement learning-based base station sleeping control algorithm are given in Algorithm 1.

Algorithm 1 Federated reinforcement learning-based base station sleeping control.

1: **while** Online **do**
2:     **procedure** Local Model Training
3:     **for** $m \in \mathcal{M}$ **do**
4:         Visit state s_m^t.
5:         Select an action a_m using ϵ-greedy policy (12).
6:         Obtain reward r_m and next state s_m^{t+1}.
7:         Store $(s_m^t, a_m^t, r_m^t, s_m^{t+1})$ into $\mathcal{D}_m$.
8:         Randomly select samples D_m from $\mathcal{D}_m$.
9:         Train Q-network using gradient descent method (15) based on the loss function (14) and samples D_m.
10:        Update $\hat{Q} = Q$.
11:    **end for**
12:    **procedure** Global Model Aggregation
13:    $g_r = \frac{\sum_{m \in \mathcal{M}} K_m \theta_m^t}{K}$.
14:    **procedure** Local Model Update
15:    $\theta_m^{t+1} = g_r - \frac{\lambda}{K_m} \sum_{m=1}^{M} \nabla L(\theta_m^t)$.
16: **end while**

---

5. Results

In this section, we evaluate the performance of the algorithm proposed in this paper through simulation. We consider a dense urban environment, where there are 30 users served by four SBSs distributed in the coverage of an MBS. Users move according to trajectories sampled from reality, with the user trajectory data sourced from the dataset GeoLife GPS Trajectories. The small base stations will use the FRL-based base station sleeping control method proposed in this text to select the appropriate working mode based on the status of their buffer area, channel, interference, etc., in order to minimize the *ELR*. The detailed parameters of the simulation are summarized in Table 4.

Table 4. Simulation parameters

Parameters	Values
The total bandwidth of the small base stations B_m	10 MHz
Carrier frequency f_c	5 GHz
Number of time slots	7×10^4
Real-world duration of simulation	70 s
Duration of a time slot	1 ms
The lifetime of a delay-tolerant data packet	100 ms
The lifetime of a delay-sensitive data packet	20 ms
Data volume of a data packet	10 kbit
Spectral density of thermal noise N_0	-174 dBm/Hz
Number of layers in neural network	20
Target network update interval step	4
Discount factor	0.9
Learning rate	0.01
Replay memory size	8000
Minibatch size	400
Cycle of model aggregation	50

Figure 5 shows the evolution of the ELR with the number of epochs. We simulated 70,000 epochs, corresponding to 70 s in the real world. From the simulation, it is observed that the DRL model converges. In the simulation of this paper, the first 10,000 epochs (10 s) are dedicated to free exploration without training the Q-network. Between 10,000 epochs and 20,000 (10 s and 20 s), deep reinforcement learning begins to train the Q-network. It is noted that as the training progresses, the DRL model can converge within 1000 ms and significantly reduce the ELR, indicating that the base station has lower weighted power

consumption and packet loss. After 20,000 epochs (20 s), the FRL-based algorithm starts to perform periodic model aggregation, with the small base stations sharing experiences within the model. It is observed that after 20,000 epochs (20 s), the ELR of the DQL-based algorithm remains relatively stable with fluctuations, while the ELR of the FRL-based algorithm further decreases. This is because, in the DQL-based algorithm, the small base stations independently interact with the environment to train the sleeping control model. However, as users move, the channel conditions and service types of the small base stations' served users dynamically change. This breaks the assumption of training and testing datasets being from the same distribution, leading to concept drift. The small base stations struggle to adapt quickly and in real time to the dynamic changes in the environment, causing performance fluctuations. On the other hand, in the FRL algorithm, the small base stations periodically interact with macro base stations for model aggregation and updating. This allows for a certain degree of experience sharing among the small base stations, enhancing the adaptability of the model. Even with users switching cells, the new small base station gains some experience with these users. The FRL-based algorithm adapts more quickly to environmental changes, thus showing better performance.

Figure 6 shows the proportion of various operating states used by base stations when employing the FRL-based small base station sleeping control algorithm at different values of ω. It can be seen that small base stations choose their operating modes according to the ε-greedy strategy, including Active, SM1, SM2, and SM3, to save power.

As ω increases, the small base stations tend to choose higher levels of sleeping to save more power. Notably, when $\omega = 1$, the proportion of the small base stations choosing the SM3 mode exceeds 90%, with other modes being chosen almost exclusively during the exploration phase of the ε-greedy strategy. This allows the system to achieve the lowest power consumption but at a high risk of packet loss. Conversely, when $\omega = 0$, the small base stations almost always choose the active mode, ensuring continuous data transmission but potentially resulting in high network power consumption.
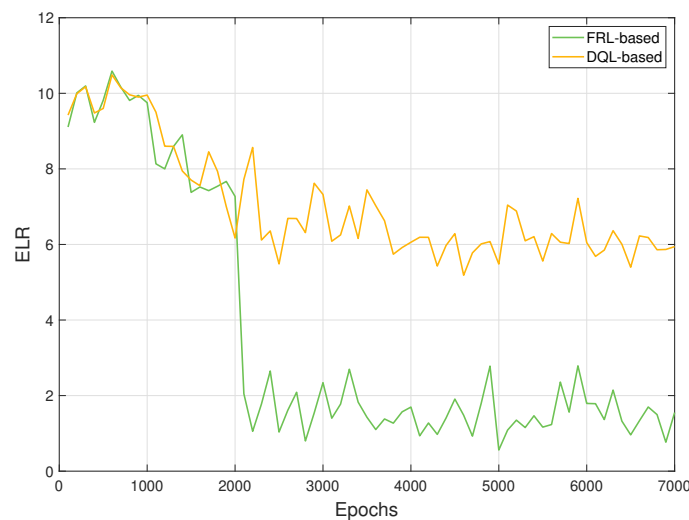

Figure 5. Evolution of the ELR with the number of epochs.

In the evaluation section focusing on the power consumption and latency performance of our algorithm, we adopt a comparative analysis approach. We provide performance evaluation results of both the algorithm proposed in this paper and those presented in existing related research. This methodology is employed to clearly discuss the relationship and differences between our study and the most recent mainstream solutions in the field. The comparison algorithms include DQL-based (similar to [5,26]), all-active (small base stations always in active mode), and random (small base stations randomly choose their work mode). Figure 7 and Figure 8, respectively, showcase the average power consumption and delay of the networks employing the federated reinforcement learning (FRL)-based

small base station sleeping control algorithm and other algorithms. Among these four algorithms, the random scheme displays higher levels of power consumption and delay because the small base stations randomly select their operating states without considering user business, channel conditions, or other network states.
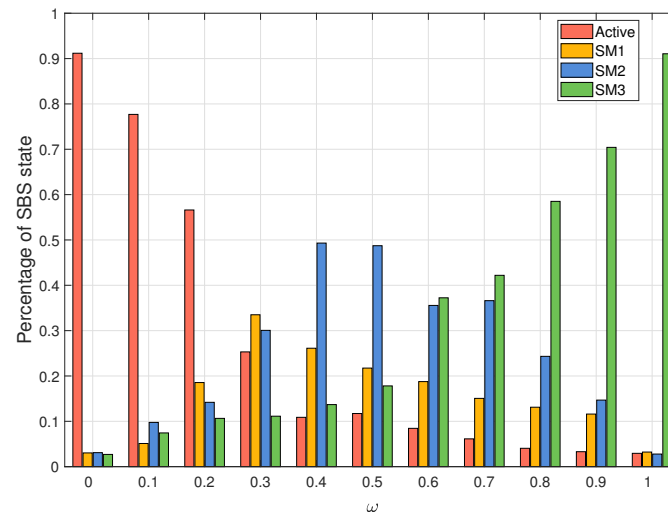

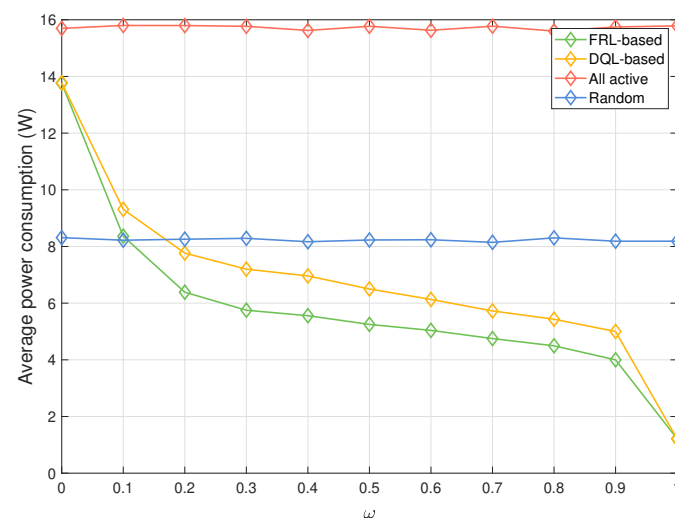**Figure 6.** The proportion of various operating states used by base stations with different ω.

Figure 7. Average power consumption with different ω.

In contrast, the all-active scheme exhibits the lowest delay levels but the highest power consumption. This is because, with the small base stations not entering any sleep mode and remaining active, the network does not experience additional delays due to base station sleeping. However, this leads to significant power wastage in scenarios where the system is not fully loaded.

The two intelligent sleeping control algorithms, FRL-based and DQL-based, achieve better energy-saving effects while maintaining relatively low delays. As the weight factor ω increases, the network's average power consumption gradually decreases, while the average delay increases, aligning with the characteristics of the designed weight factor. When ω is 0 or 0.1, to ensure very low packet loss performance in the system, the active mode dominates the small base station's choices. The power consumption levels of both algorithms are higher than the random scheme, but their delay levels are close to the method without base station sleeping. This is because $\omega = 0$ and $\omega = 1$ represent two of the simplest sleeping objectives: base stations choosing active mode as much as possible

to ensure transmission performance and choosing the deepest sleep mode to achieve the lowest power consumption, respectively. Both learning-based methods can learn the optimal strategy, hence displaying similar performance.
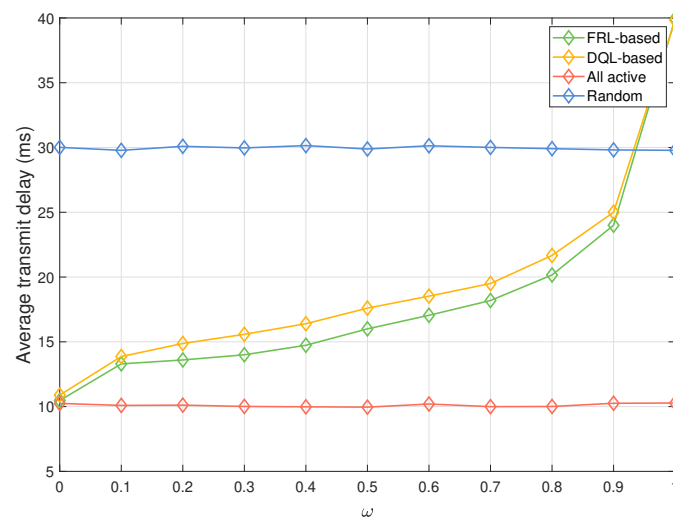

Figure 8. Average transmit delay with different ω.

However, when ω takes other values, the small base stations must consider the trade-off between power consumption and transmission performance when selecting their operating modes. The FRL-based algorithm shows stronger performance than the DQL-based algorithm because it mitigates the issue of concept drift through model aggregation, making the model more adaptable to user mobility. On average, the FRL-based algorithm achieves a 13.9% reduction of power consumption and a 5.8% reduction of delay compared to the DQL-based algorithm. Compared to the scenario without base station sleeping, it achieves a 62.7% reduction of power consumption.

6. Discussion

In this paper, we investigated the issue of network energy saving through base station sleeping in dense network environments. To reduce network power consumption while preventing a significant decrease in users' QoS, we introduced a federated reinforcement learning-based multi-level sleeping control method for small base stations, aimed at minimizing the network's overall power consumption and packet loss. The small base stations choose between active or dormant modes at the start of each time slot depending on their buffer load, channel conditions, and neighboring interference. They interact with the environment using double deep reinforcement learning to obtain a near-optimal sleeping control policy. Furthermore, recognizing the concept drift problem due to user mobility in actual networks, we adopted a federated learning mechanism. This involves periodic model aggregation and updates, allowing small base stations to share experiences, thereby enhancing the performance of the sleeping control algorithm. The simulation shows that, compared to scenarios without base station sleeping, our proposed method can save over 62.7% in power consumption. Additionally, relative to the sleeping control method based solely on DQL, our approach can reduce power consumption by 13.9% and average latency by 5.8%.

The federated learning-based base station sleeping control technique proposed in this paper is conducive to achieving a more energy-efficient 6G network. Employing a decentralized learning and decision-making process enables small base stations to autonomously adjust their sleep states based on real-time data and environmental changes. This approach not only enhances overall energy efficiency and the packet loss of the network compared with completely independent learning schemes but also reduces dependence on centralized control, thereby increasing network reliability and responsiveness. Furthermore, the appli-

cation of this technology may also lead to further improvements in user experience, service quality, and environmental adaptability in 6G networks, particularly in dynamic and constantly changing network environments. In future work, it would be valuable to explore different federated learning architectures and algorithms in the base station sleep control, especially those that can efficiently handle non-independently and identically distributed data, which is a common challenge in real-world networks. Advanced algorithms that can better manage the skewed and unbalanced data distributions often found in dense network environments would likely yield improvements in both energy savings and QoS. Additionally, our study has certain limitations regarding network architecture. The scenario we investigated assumes that small base stations have ideal backhaul links to support the signaling required for transmitting neural network parameters during model aggregation and model update processes. However, considering that future 6G networks are likely to support various networking methods, including wireless backhaul for small base stations, it would be interesting to explore the impact of the signaling overhead required for federated learning on performance. Developing an optimal base station sleeping control technology that accounts for signaling overhead could be a fascinating research direction. This would potentially enhance the practical applicability of the algorithm.

Author Contributions: Methodology, T.P. and X.W.; Formal analysis, T.P.; Resources, X.W.; Data curation, T.P.; Writing—original draft, T.P.; Writing—review & editing, X.W. and X.L. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The data presented in this study are available in this article.

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

References

1. ITU-R. IMT Traffic Estimates for the Years 2020 to 2030. M. 2370-0. 2015. Available online: https://www.itu.int/pub/R-REP-M.2370-2015 (accessed on 17 March 2023).
2. López-Pérez, D.; De Domenico, A.; Piovesan, N.; Xinli, G.; Bao, H.; Qitao, S.; Debbah, M. A Survey on 5G Radio Access Network Energy Efficiency: Massive MIMO, Lean Carrier Design, Sleep Modes, and Machine Learning. *IEEE Commun. Surv. Tutor.* **2022**, *24*, 653–697. [CrossRef]
3. Mao, B.; Tang, F.; Kawamoto, Y.; Kato, N. AI Models for Green Communications Towards 6G. *IEEE Commun. Surv. Tutor.* **2022**, *24*, 210–247. [CrossRef]
4. Guan, L.; Ding, Y.; Li, R.; Su, T.; Hu, L.; Wang, T.; Hu, N. Network energy saving technologies for green 5G. *Telecommun. Sci.* **2022**, *38*, 167–174. [CrossRef]
5. Amine, A.E.; Chaiban, J.P.; Hassan, H.A.H.; Dini, P.; Nuaymi, L.; Achkar, R. Energy Optimization with Multi-Sleeping Control in 5G Heterogeneous Networks Using Reinforcement Learning. *IEEE Trans. Netw. Serv. Manag.* **2022**, *19*, 4310–4322. [CrossRef]
6. Ashraf, I.; Boccardi, F.; Ho, L. SLEEP mode techniques for small cell deployments. *IEEE Commun. Mag.* **2011**, *49*, 72–79. [CrossRef]
7. Niu, Z.; Zhang, J.; Guo, X.; Zhou, S. On energy-delay tradeoff in base station sleep mode operation. In Proceedings of the 2012 IEEE International Conference on Communication Systems (ICCS), Singapore, 21–23 November 2012; pp. 235–239. [CrossRef]
8. Ju, H.; Kim, S.; Kim, Y.; Lee, H.; Shim, B. Energy-Efficient Ultra-Dense Network using Deep Reinforcement Learning. In Proceedings of the 2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Atlanta, GA, USA, 26–29 May 2020; pp. 1–5. [CrossRef]
9. Gao, Y.; Chen, J.; Liu, Z.; Zhang, B.; Ke, Y.; Liu, R. Machine Learning based Energy Saving Scheme in Wireless Access Networks. In Proceedings of the 2020 International Wireless Communications and Mobile Computing (IWCMC), Limassol, Cyprus, 15–19 June 2020; pp. 1573–1578. [CrossRef]
10. Donevski, I.; Vallero, G.; Marsan, M.A. Neural Networks for Cellular Base Station Switching. In Proceedings of the IEEE INFOCOM 2019—IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), Paris, France, 29 April–2 May 2019; pp. 738–743. [CrossRef]
11. Chang, P.; Miao, G. Optimal Operation of Base Stations with Deep Sleep and Discontinuous Transmission. *IEEE Trans. Veh. Technol.* **2018**, *67*, 11113–11126. [CrossRef]
12. Zheng, J.; Cai, Y.; Chen, X.; Li, R.; Zhang, H. Optimal Base Station Sleeping in Green Cellular Networks: A Distributed Cooperative Framework Based on Game Theory. *IEEE Trans. Wirel. Commun.* **2015**, *14*, 4391–4406. [CrossRef]

13. Feng, M.; Mao, S.; Jiang, T. Dynamic Base Station Sleep Control and RF Chain Activation for Energy-Efficient Millimeter-Wave Cellular Systems. *IEEE Trans. Veh. Technol.* **2018**, *67*, 9911–9921. [CrossRef]

14. Zhu, Y.; Wang, S. Joint Traffic Prediction and Base Station Sleeping for Energy Saving in Cellular Networks. In Proceedings of the ICC 2021—IEEE International Conference on Communications, Montreal, QC, Canada, 14–23 June 2021; pp. 1–6. [CrossRef]

15. Wu, J.; Wong, E.W.M.; Chan, Y.C.; Zukerman, M. Power Consumption and GoS Tradeoff in Cellular Mobile Networks with Base Station Sleeping and Related Performance Studies. *IEEE Trans. Green Commun. Netw.* **2020**, *4*, 1024–1036. [CrossRef]

16. Guerra, I.; Yin, B.; Zhang, S.; Cheng, Y. Optimization of Base Station ON-Off Switching with a Machine Learning Approach. In Proceedings of the ICC 2021—IEEE International Conference on Communications, Montreal, QC, Canada, 14–23 June 2021; pp. 1–6. [CrossRef]

17. Lin, J.; Chen, Y.; Zheng, H.; Ding, M.; Cheng, P.; Hanzo, L. A Data-driven Base Station Sleeping Strategy Based on Traffic Prediction. *IEEE Trans. Netw. Sci. Eng.* **2021**, 1–16. [CrossRef]

18. Zhang, Q.; Xu, X.; Zhang, J.; Tao, X.; Liu, C. Dynamic Load Adjustments for Small Cells in Heterogeneous Ultra-dense Networks. In Proceedings of the 2020 IEEE Wireless Communications and Networking Conference (WCNC), Seoul, Republic of Korea, 25–28 May 2020; pp. 1–6. [CrossRef]

19. Wu, Q.; Chen, X.; Zhou, Z.; Chen, L.; Zhang, J. Deep Reinforcement Learning with Spatio-Temporal Traffic Forecasting for Data-Driven Base Station Sleep Control. *IEEE/ACM Trans. Netw.* **2021**, *29*, 935–948. [CrossRef]

20. Shi, Z.; Liu, J.; Zhang, S.; Kato, N. Multi-Agent Deep Reinforcement Learning for Massive Access in 5G and Beyond Ultra-Dense NOMA System. *IEEE Trans. Wirel. Commun.* **2022**, *21*, 3057–3070. [CrossRef]

21. Debaillie, B.; Desset, C.; Louagie, F. A Flexible and Future-Proof Power Model for Cellular Base Stations. In Proceedings of the 2015 IEEE 81st Vehicular Technology Conference (VTC Spring), Glasgow, UK, 11–14 May 2015; pp. 1–7. [CrossRef]

22. Salem, F.E.; Altman, Z.; Gati, A.; Chahed, T.; Altman, E. Reinforcement Learning Approach for Advanced Sleep Modes Management in 5G Networks. In Proceedings of the 2018 IEEE 88th Vehicular Technology Conference (VTC-Fall), Chicago, IL, USA, 27–30 August 2018; pp. 1–5. [CrossRef]

23. Amine, A.E.; Dini, P.; Nuaymi, L. Reinforcement Learning for Delay-Constrained Energy-Aware Small Cells with Multi-Sleeping Control. In Proceedings of the 2020 IEEE International Conference on Communications Workshops (ICC Workshops), Dublin, Ireland, 7–11 June 2020; pp. 1–6. [CrossRef]

24. El-Amine, A.; Iturralde, M.; Haj Hassan, H.A.; Nuaymi, L. A Distributed Q-Learning Approach for Adaptive Sleep Modes in 5G Networks. In Proceedings of the 2019 IEEE Wireless Communications and Networking Conference (WCNC), Marrakesh, Morocco, 15–18 April 2019; pp. 1–6. [CrossRef]

25. Masoudi, M.; Khafagy, M.G.; Soroush, E.; Giacomelli, D.; Morosi, S.; Cavdar, C. Reinforcement Learning for Traffic-Adaptive Sleep Mode Management in 5G Networks. In Proceedings of the 2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications, London, UK, 31 August–3 September 2020; pp. 1–6. [CrossRef]

26. Lin, S.; Qiu, C.; Tan, J.; Wang, X.; Yang, Y.; He, Y.; Jiang, J. DADEs: 5G Dual-Adaptive Delay-aware and Energy-saving System with Tandem Learning. In Proceedings of the GLOBECOM 2022—2022 IEEE Global Communications Conference, Rio de Janeiro, Brazil, 4–8 December 2022; pp. 1–6. [CrossRef]

27. Masoudi, M.; Soroush, E.; Zander, J.; Cavdar, C. Digital Twin Assisted Risk-Aware Sleep Mode Management Using Deep Q-Networks. *IEEE Trans. Veh. Technol.* **2023**, *72*, 1224–1239. [CrossRef]

28. Xu, X.; Yuan, C.; Chen, W.; Tao, X.; Sun, Y. Adaptive Cell Zooming and Sleeping for Green Heterogeneous Ultradense Networks. *IEEE Trans. Veh. Technol.* **2018**, *67*, 1612–1621. [CrossRef]

29. Chang, Y.; Chen, W.; Li, J.; Liu, J.; Wei, H.; Wang, Z.; Al-Dhahir, N. Collaborative Multi-BS Power Management for Dense Radio Access Network using Deep Reinforcement Learning. *IEEE Trans. Green Commun. Netw.* **2023**, *7*, 2104–2116. [CrossRef]

30. GreenTouch. Power Model for Today's and Future Base Stations. Available online: https://www.powermodel.be/live/ImecPowerModel.shtml (accessed on 17 March 2023).

31. Liu, Y.J.; Feng, G.; Sun, Y.; Qin, S.; Liang, Y.C. Device Association for RAN Slicing Based on Hybrid Federated Deep Reinforcement Learning. *IEEE Trans. Veh. Technol.* **2020**, *69*, 15731–15745. [CrossRef]

32. Xue, Q.; Liu, Y.J.; Sun, Y.; Wang, J.; Yan, L.; Feng, G.; Ma, S. Beam Management in Ultra-Dense mmWave Network via Federated Reinforcement Learning: An Intelligent and Secure Approach. *IEEE Trans. Cogn. Commun. Netw.* **2023**, *9*, 185–197. [CrossRef]