Article

Joint Optimization Algorithm for Small Base Station States Control and User Association in Wireless Caching Networks

Weipeng Wang [1,*] , Jihong Zhao [1,2], Hua Qu [1] and Huijun Dai [1]

1 Department of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China
2 School of Telecommunication and Information Engineering, Xi'an University of Posts and Telecommunications, Xi'an 710061, China
* Correspondence: wangweipeng88@stu.xjtu.edu.cn

Abstract: The energy consumption management of small base stations (SBSs) in wireless caching networks with dense deployment of SBSs is an urgent problem to be solved. This paper jointly optimizes the SBS state control and user association problems in caching networks to reduce network energy consumption while taking into account the average service latency of the network to ensure user experience. First, a new definition of three-state SBSs in caching networks is proposed based on their ability to keep content cache updated. Then, a relaxed threshold setting method is designed and the SBS traffic prediction is used to obtain the initial state information of SBSs in the next period. In order to eliminate the impact on the accessing users when the switched-off base station (BS) wakes up, a SBS state asynchronous switching mechanism is proposed to ensure that the users who switch to the waking SBS can carry out communication services normally, and a user association strategy is constructed with the SBS load as the optimization target. Finally, a joint optimization model of user association and SBS state control (SSC-UA) is constructed to admit and correct the initial state of SBSs to maximize the system gain and obtain the final state strategy for each SBS in the next period. The simulation results show that the proposed SSC-UA algorithm can effectively improve the energy efficiency and reduce the network service delay at the same time.

1. Introduction

With the development of 5G networks, large-scale dense deployment of SBSs is a major feature of future networks, and these SBSs can provide higher service rates and lower latency for users to meet people's network demands [1,2]. At the same time, the dense deployment of SBSs also brings the problem of energy wastage [3,4], and proper energy efficiency management of SBSs is needed to reduce network energy consumption.

In cellular networks, SBSs are deployed to meet the regional peak traffic requirements [5], but the traffic load of individual SBSs can fluctuate widely in space and time based on time, location, and population density distribution [6]. When the SBS coverage area is at a low load because fewer users need to be served, sleep control of SBSs can effectively reduce the energy consumption of cellular networks and improve network energy efficiency. At the same time, when sleep control is applied to the SBS, the users originally connected to the SBS need to switch to other BSs. The traditional user switching scheme is to switch users to the adjacent SBS or macro base station (MBS) with the highest received power, which causes most users to be switched to the MBS because the transmit power of MBSs is much higher than that of SBSs, resulting in excessive load pressure on the MBS and even network blockage in serious cases. Therefore, user association needs to be considered in the energy efficiency management of SBSs to offload users to other turned-on SBSs as much as possible to reduce the load on MBSs. In fact, user association and SBS sleep are highly coupled, as the user association policy determines which SBSs can be put

into sleep, and the SBS sleep control determines which users can be switched to the set of turned-on SBSs. In the user association problem, in order to avoid MBS overload caused by traditional schemes, Corroy et al. [7] proposed a user association scheme that maximizes the total network rate. A user association rule that promotes maximum–minimum fairness among users is proposed in [8]. Ye et al. [9] studied a framework for maximizing fair utility and demonstrated that the framework can achieve different tiers of SBSs between load balancing. However, none of the aforementioned studies consider the SBS energy efficiency. On the other hand, there have been a large number of related studies on SBS state control [10–14]. Baiocchi et al. [10] fixed some SBSs to be turned off during a certain time period according to the periodic fluctuation pattern of traffic. Samulevicius et al. [11] decided whether to shut down a SBS by comparing the SBS load with a set threshold value. Liu et al. [12] proposed a stochastic SBS sleep scheme and a strategic SBS sleep scheme. Zhu et al. [13] investigated the SBS sleep algorithm based on the traffic prediction results and determined the sleep strategy by controlling the SBS transmit power and bandwidth allocation. Lin et al. [14] investigated the joint user association and SBS state set optimization problem by an approximate algorithm. In addition, some works have used machine learning methods for SBS state control. Amine et al. [15] proposed a reinforcement learning algorithm for small cells that adapts their activities subject to service delay constraint. The algorithm intelligently learns from the environment based on the co-channel interference, the cell buffer size, and the expected cell throughput in order to decide the best SBS state control policy. Wu et al. [16] developed a traffic-aware dynamic BS sleep control framework, which presented a novel data-driven learning approach to determine the BS active/sleep modes while meeting lower energy consumption and satisfactory Quality of Service (QoS) requirements. Ju et al. [17] proposed a deep reinforcement learning based approach to control SBS state. The key ingredient of the proposed scheme was to use a decision selection network to reduce the size of action space.

Unlike traditional wireless networks, wireless caching networks need to consider the synchronization of content caches in addition to the traffic distribution changes in the region when controlling the state of SBSs. Since SBSs in caching networks are equipped with cache units, some contents are periodically placed or replaced in the cache units and the content caching function must be taken into account when performing SBS state control. In this paper, we study the joint optimization problem of SBS state control and user association based on energy consumption and latency. First, three states of SBSs in caching networks are defined based on whether they can keep the content cache updated or not. Then, to address the drawback of relying on threshold settings when determining the state of SBSs through traffic load, this paper designs a relaxed threshold determination method, where the threshold settings only determine the initial state of SBSs and provide initial options for subsequent optimization algorithms, thus reducing the complexity. At the same time, in order to eliminate the impact on the access users when the switched-off SBS wakes up, a SBS state asynchronous switching mechanism is proposed to ensure that the users who cut to the waking SBS can carry out communication services normally. Finally, a joint optimization model of SBS state control and user association based on energy consumption and time delay is constructed to admit and correct the initial state of SBSs to maximize the system gain and obtain the final state strategy of each SBS in the next period.

The rest of the paper is organized as follows: Section 2 introduces the system model, including the energy consumption model and the cost of SBS state switching; Section 3 designs a relaxed threshold setting method, and the initial state policy of the SBS is determined by long short-term memory (LSTM) traffic prediction; Section 4 uses the constructed objective function to make a final SBS state decision, which is admitted and modified the initial SBS state by the constructed objective function; Section 5 verifies the performance of the joint optimization algorithm by experimental simulation; and Section 6 summarizes the work of this paper.

2. System Model

A downlink heterogeneous network model consisting of an MBS and multiple SBSs was considered. The MBS is at the center of region $\mathcal{L} \subseteq \mathbb{R}^2$, and a large number of SBSs are deployed around it to meet the peak traffic demand. $B_M = \{1\}$ and $B_S = \{2, 3, \cdots, M\}$ denote the index sets of the MBS and SBSs, respectively, and $B = B_M \bigcup B_S$ denotes the index set of all BSs. SBSs are distributed independently and randomly around MBS, and there are overlapping coverage areas among SBSs. In addition, since SBSs generally use digital subscriber line (DSL) or cable modem (CM) as backhaul links, their capacity is much smaller than the backhaul links of the MBS; therefore, cache units are configured for all SBSs to cache certain contents at SBSs. In contrast, MBSs do not, thus all contents are obtained from the core network through the backhaul link for the MBS. We assume that the set of all possible user requests for content in the region at a given time is $\mathcal{F} = \{1, 2, \cdots, F\}$ and S_f denotes the size of content f. For simplicity, it is assumed that all the contents are of the same size, $S_f = S$, because even if the content is of different sizes, it can be divided into blocks of the same size or encoded into blocks of the same size by encoding techniques. Therefore, the results of this paper can be applied to the scenarios with different content sizes as well. Numerous studies have shown that only a small fraction of content is frequently requested by users and the probability of users requesting content follows Zipf distribution. Assuming that the contents of $\mathcal{F}$ are ordered in descending order of requested probability, the request probability of the fth content is expressed as

$$q_f = \frac{f^{-z}}{\sum\limits_{l=1}^{F} l^{-z}} \tag{1}$$

where z is the skew factor, and a larger z indicates that the content requested by users is more concentrated and a large number of users request only a small part of the content located in the header $\mathcal{F}$.

We assume that the requests for content f from users located at x in the region in period t obey a Poisson process with density $\lambda_f^t(x)$, and that the requests for different contents are Poisson processes which are independent of each other. According to the superposition property of the Poisson process, the user's request for any content at x in period t obeys the Poisson process with density $\lambda^t(x) = \sum\limits_{f=1}^{F} \lambda_f^t(x)$. The traffic load density at x in region $\mathcal{L}$ in period t is defined [14] as $\gamma^t(x) \doteq \sum\limits_{f=1}^{F} \lambda_f^t(x) S_f$, which represents the average content traffic per unit area per unit time, and the unit time in this paper refers to the duration of a period t. In addition, the probability that a user located at x requests content f can be expressed as

$$q_f = \frac{\int_{\mathcal{L}} \lambda_f(x) dx}{\int_{\mathcal{L}} \lambda(x) dx} \tag{2}$$

We assume that all SBSs have the same caching policy, which is to cache some of the content in order of content popularity in the region from high to low, until the cache capacity is full. Let the capacity of the cache unit of the SBS be $F_c S$, i.e., the SBS caches F_c contents and the set of cached contents is $\mathcal{F}_c = \{1, 2, \cdots, F_c\}$.

Since all the SBSs use the same frequency band, when the user at x selects a SBS to access, other turned-on SBSs will cause co-channel interference to the user, so the user's signal to interference plus noise ratio (SINR) can be expressed as

$$\text{SINR}_i^t(x) = \frac{P_i^t H_i(x)}{\sum_{k \in B_{on}^t \setminus \{i\}} P_k^t(x) H_k(x) + \sigma_0^2} \tag{3}$$

where P_i^t denotes the transmit power of SBS, $i \in B_{on}^t$, $H_i(x)$ denotes the channel power gain between the user at x and access SBS, $i \in B_{on}^t$, $\sum_{k \in B_{on}^t \setminus \{i\}} P_k^t(x) H_k(x)$ denotes the sum of interference from other SBSs except access SBS i, and σ_0^2 is the noise power.

The rate available to the user at x from SBS $i \in B_{on}^t$ by Shannon's formula is

$$r_i^t(x) = \mathcal{W} \log_2 \left(1 + \text{SINR}_i^t(x) \right) \tag{4}$$

where $\mathcal{W}$ indicates the bandwidth allocated by the SBS for the user communication link.

We define the SBS load density [18] as $\phi_i^t(x) = \gamma^t(x) / r_i^t(x)$, which represents the proportion of time required to transmit the traffic load, $\gamma^t(x)$, from SBS $i \in B_{on}^t$ in period t. Denoted by $a_i^t(x)$ is the probability that a user located at x in period t accesses SBS $i \in B_{on}^t$. Alternatively, it can be interpreted as the proportion of time that SBS $i \in B_{on}^t$ takes to serve content requests from users at x.

In summary, we can express the feasible domain of the SBS load as

$$\Omega^t(B_{on}^t) = \left\{ \boldsymbol{\rho}^t \left| \begin{array}{c} \rho_i^t = \int_{\mathcal{L}} \phi_i^t(x) a_i^t(x) dx, \\ 0 \leqslant \rho_i^t \leqslant 1 - \vartheta, \\ \sum_{i \in B_{on}^t} a_i^t(x) = 1, \\ 0 \leqslant a_i^t(x) \leqslant 1 \end{array} \right. \right\} \tag{5}$$

where ϑ is a very small positive number. According to queuing theory, full load indicates that the queuing system is unstable, and in a cellular network it means that the requests to access the SBS exceed its processing capacity. Therefore, in order to ensure the normal service of the SBS, $\rho_i^t \leqslant 1 - \vartheta$ is required.

Theorem 1. *The feasible domain $\Omega^t(B_{on}^t)$ of the SBS load is a convex set.*

Proof. Let $\boldsymbol{\rho}^{t1} \in \Omega^t(B_{on}^t)$ and $\boldsymbol{\rho}^{t2} \in \Omega^t(B_{on}^t)$ be two load vectors in the feasible domain, respectively, and $\boldsymbol{\rho}^{t1} \neq \boldsymbol{\rho}^{t2}$. Then there exist user association vectors $\boldsymbol{a}^{t1}(x) = \left(a_1^{t1}(x), a_2^{t1}(x), \cdots, a_b^{t1}(x) \right)$ and $\boldsymbol{a}^{t2}(x) = \left(a_1^{t2}(x), a_2^{t2}(x), \cdots, a_b^{t2}(x) \right)$ satisfying $\rho_i^{t1} = \int_{\mathcal{L}} \phi_i^t(x) a_i^{t1}(x) dx$ and $\rho_i^{t2} = \int_{\mathcal{L}} \phi_i^t(x) a_i^{t2}(x) dx$. For $\forall \theta \in [0,1]$, we have

$$\begin{aligned} \rho_i^t &= \theta \rho_i^{t1} + (1-\theta) \rho_i^{t2} \\ &= \theta \int_{\mathcal{L}} \phi_i^t(x) a_i^{t1}(x) dx + (1-\theta) \int_{\mathcal{L}} \phi_i^t(x) a_i^{t2}(x) dx \\ &= \int_{\mathcal{L}} \phi_i^t(x) \left[\theta a_i^{t1}(x) + (1-\theta) a_i^{t2}(x) \right] dx \\ &= \int_{\mathcal{L}} \phi_i^t(x) a_i^t(x) dx \end{aligned} \tag{6}$$

It results in $a_i^t(x) = \theta a_i^{t1}(x) + (1-\theta) a_i^{t2}(x)$, so both ρ_i^t and $a_i^t(x)$ also satisfy the conditions in $\Omega^t(B_{on}^t)$, so $\Omega^t(B_{on}^t)$ is a convex set. □

2.1. Energy Consumption Model

In order to be able to reduce the power consumption of a SBS by turning off the sleep operation of certain hardware modules within the SBS under low traffic conditions, a clear understanding of the hardware composition of a SBS is required. The hardware composition of a typical SBS is described in [19], which in general consists of three parts, as shown in Figure 1. The first part contains a microprocessor responsible for implementing and managing the standardized radio protocol stack and the associated baseband processing, as well as managing the backhaul link between the SBS and the core network. In addition to the on-chip memory, it has one or more memory components connected to the microprocessor, which are required for various data processing functions and system startup, as well as to enable caching of the contents of the network. The second part contains a field programmable gate array (FPGA) and several other integrated circuits

to implement functions such as data encryption, hardware authentication, network time protocol (NTP), etc. The third part consists of radio frequency (RF) components and power amplifiers for transmitting and receiving data and transmitting high power signals to the transmitting antenna.
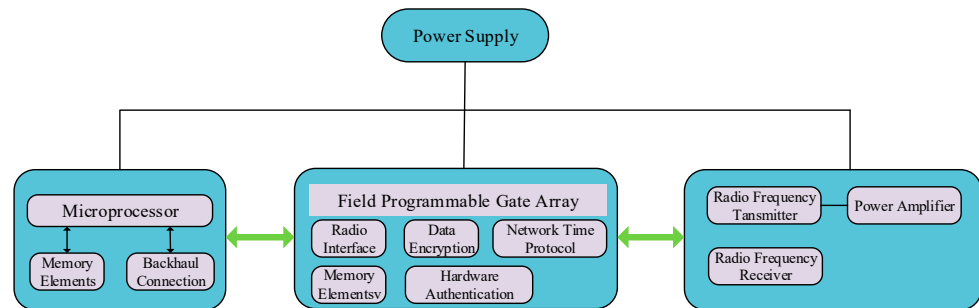

Figure 1. A hardware composition model of a typical SBS.

Reference [12] analyzed the hardware power consumption model of SBSs, finding that the main power consuming parts of SBSs are mainly composed of a RF front-end and a temperature compensated crystal oscillator heating device, where the power consumption of RF front-end accounts for about 45% of the total power consumption and the power consumption of the temperature compensated crystal oscillator heating device accounts for about 7% of the total power consumption. Therefore, shutting down some of the components of the SBSs in certain areas at appropriate times to make the SBSs perform a sleep state can effectively reduce the power consumption of the network and improve the energy efficiency of the network. However, since the state of the SBSs needs to be dynamically adjusted according to the traffic changes in the region, the SBSs that are in sleep state at any one time may need to be turned back on at the next time, and it takes some time to turn on the turned-off components in the SBSs again. Considering the characteristics of caching networks, we define a three-state SBS scheme in caching networks:

- On: the SBS is in a normal service state, and all functional modules are turned on to keep the cache content updated and provide real-time service to users.
- Standby: The SBS is in a light sleep state and only some functional modules of the SBS are turned off, thus ensuring that it can be quickly woken up when necessary. Meanwhile, the SBS can receive content request information from all users in the coverage area during the standby period and update the local cache content regularly according to the preset caching algorithm, so that it can provide content services to users immediately after it is turned on.
- Sleep: A SBS in deep sleep state turns off most of its functional modules and takes some time to be woken up. Moreover, its cache module does not participate in updating popular contents in the network during sleep but can receive content request information from all users in the coverage area, and immediately updates the local cache contents according to the preset caching algorithm after being woken up, and then finally provides services to users.

We denote the state indicator function of the SBS as

$$b_i^t = \begin{cases} 1, & \text{On} \\ 0, & \text{Standby} \\ -1, & \text{Sleep} \end{cases} \tag{7}$$

where Standby and Sleep are both two sleep modes of the SBS, only the degree differs.

According to the type of power consumption of the SBS, the power consumption of the SBS can be divided into two parts: fixed power consumption and variable power consumption, where fixed power consumption mainly consists of circuit power consumption and cooling system power consumption of the SBS, and variable power consumption is mainly

related to the power amplifier of the SBS, which is determined by the traffic load of the SBS. According to the characteristics of SBS power consumption and SBS state classification, the SBS power consumption can be expressed as

$$
P_i^t = \begin{cases} (1 - q_i)\rho_i^t \mathcal{P}_i + q_i \mathcal{P}_i, & b_i^t = 1 \\ P_{standby}, & b_i^t = 0 \\ P_{sleep}, & b_i^t = -1 \end{cases} \tag{8}
$$

where $q_i \in [0, 1]$ indicates the proportion of fixed energy consumption of the SBS to the total power consumption, and $\mathcal{P}_i$ indicates the maximum power consumption of the SBS when the SBS is fully loaded. The percentage of fixed power consumption when the SBS is turned on indicates the power consumption of the SBS in different states, and the wake-up time and power consumption of the SBS in different modes are shown in Table 1.

Table 1. Wake-up time and power consumption of SBS in different states [20].

Mode	Wake-Up Time (s)	Power Consumption
On	N/A	100%
Standby	0.5	50%
Sleep	10	15%

2.2. SBS Service Cost

2.2.1. SBS Service Delay

When a user sends an access request to the SBS, the user waits for the SBS to transmit content for the user according to the queuing theory, so the higher the traffic load of the SBS, the higher the resulting service delay, which is defined as [21]:

$$
d_{i,s}^t(\rho_i^t) = \frac{1}{1 - \rho_i^t} \tag{9}
$$

2.2.2. SBS State Switching Cost

The switching cost for the SBS from on mode to sleep mode or from sleep mode to on mode during two adjacent time periods is mainly referred to as the energy loss and time delay. Assuming that the energy loss of switching the SBS from on mode to sleep mode is 0, and only the energy lost in waking up the sleep state SBS is considered, the energy switching cost of the SBS can be expressed as

$$
P_{i,h}^t = \begin{cases} \beta\left(b_i^t - b_i^{t-1}\right)^- + \gamma, & i \in B_{on}^t, b_i^{t-1} \in (0, -1) \text{ and } b_i^{t-1} \neq b_i^t \\ 0, & i \in B_{on}^t, b_j^{t-1} = 1 \text{ or } b_i^{t-1} = b_i^t \end{cases} \tag{10}
$$

where β denotes the energy consumption required to switch between two adjacent states of SBS, γ is the energy consumption bias, which is to make the SBS state switching cost in line with the circuit energy consumption characteristics. That is, the direct switching between two SBS states is lower than the energy consumption of an additional intermediate state, i.e., the energy consumption of the SBS switching from the Sleep state to the On state is lower than the energy consumption of the SBS first switching from the Sleep state to the Standby state and then switching from the Standby state to the On state. Function $(x)^-$ is defined as

$$
(x)^- = \begin{cases} x, x > 0 \\ 0, x \leqslant 0 \end{cases} \tag{11}
$$

Since the wake-up time required for sleep state SBSs of different depths varies greatly, the user service of a SBS that needs to be woken up is affected by a large delay during the

wake-up period of the SBS. In this paper, we focus on the impact on users within a period, so we define the average wake-up delay of a SBS that needs to be woken up in a period as

$$d_{i,w}^t\left(\rho_i^t\right) = \frac{T_w}{e^{2\rho_i^t}} \tag{12}$$

This indicates that the wake-up delay of the SBS is negatively related to the load of the SBS, where T_w indicates the wake-up time. A higher load indicates that more users access the awake SBS in the t period. For the SBS, the wake-up cost can be spread among more users, so that the average wake-up delay can be lower and the SBS can be woken up more easily. On the contrary, if the relevant SBS has a low load in the t period, its wake-up delay will be high and the difficulty of wake-up will increase, in line with the actual application scenario. Combined with the state of the SBS in the previous period, Equation (12) can be expressed more specifically as

$$d_{i,w}^t\left(\rho_i^t\right) = \begin{cases} \frac{0.1}{e^{2\rho_i^t}}, & b_i^{t-1} = 0 \\ 0, & b_i^{t-1} = 1 \\ \frac{10}{e^{2\rho_i^t}}, & b_i^{t-1} = -1 \end{cases} \tag{13}$$

Thus, the total time delay of the system in period t can be expressed as

$$D^t\left(\boldsymbol{\rho}^t\right) = \sum_{i \in B_{on}^t} d_{i,s}^t\left(\rho_i^t\right) + d_{i,w}^t\left(\rho_i^t\right) \tag{14}$$

3. Initial State of SBSs Based on Relaxed Threshold for LSTM Traffic Prediction

When using the SBS traffic prediction algorithm to study the switching state of the SBS, one generally sets a traffic threshold, where the predicted traffic load of the SBS in the next period is lower than the threshold and is turned off, and vice versa. In addition to the accuracy of the prediction algorithm, the setting of the threshold value is also important. A fixed threshold value set artificially too high will turn off the SBS too much, reducing the quality of service for users or even making them unserved, and a threshold value set too low will result in poor energy saving. This paper corrects the results of SBS switching by subsequent SBS state switching cost and user association algorithms to ensure that unreasonable turn-on or turn-off does not occur, so we can set a higher threshold for SBS turn-on traffic and a lower threshold for SBS sleep traffic, which is then supplemented by subsequent algorithms to make it possible to turn off the SBS to the maximum extent without affecting user service.

3.1. SBSs State Transfer Design

Since there are two levels of sleep in the SBS, Standby and Sleep, we need to set two traffic thresholds, φ_b and φ_s, for Standby and Sleep, respectively. The traffic volume of the SBS in time period t is denoted by T_i^t, and we have

$$b_i^t = \begin{cases} 1, & T_i^t \geq \varphi_b, i \in B_S \\ 0, & \varphi_s \leq T_i^t < \varphi_b, b_i^{t-1} \in \{0,1\}, i \in B_S \\ -1, & 0 \leq T_i^t < \varphi_s \text{ or } \left\{\varphi_s \leq T_i^t < \varphi_b, b_i^{t-1} = -1\right\}, i \in B_S \end{cases} \tag{15}$$

where $b_j^{t-1} \in \{0,1\}$ indicates that only in the $t-1$ period the SBS is in On or Standby state, and in the t period the SBS may be in Standby. The set of SBSs in different states can be denoted as $B_{on}^t = \{i \mid b_i^t = 1, \forall i \in B\}$, $B_{stand}^t = \{i \mid b_i^t = 0, \forall i \in B\}$ and $B_{sleep}^t = \{i \mid b_i^t = -1, \forall i \in B\}$, respectively. The state transfer of the SBS is shown in Figure 2. In order to reduce the loss of the SBS hardware due to frequent switching between different states, it is assumed that the state of the SBS cannot be changed from the Sleep state to the Standby state, and we will now explain the rationality of this assumption.
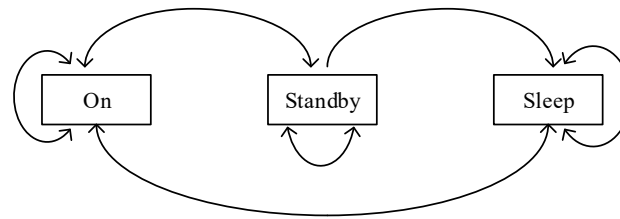
Figure 2. Base station status transfer map.

If only based on the determination result of the traffic threshold, the SBS needs to be switched from Sleep to Standby in period t. There is energy loss in this process, and the SBS in the Standby state does not provide services to users, but only resumes the function of synchronizing the network cache contents and can be quickly woken up in the next period. For this SBS, three states are possible in period $t + 1$; (1) Switching back to the Sleep state makes switching the SBS from Sleep to Standby in period t a negative operation, with no gain except more energy consumption for the SBS; (2) Keeping the Standby state, its SBS state transfer process is Sleep $\rightarrow$ Standby $\rightarrow$ Standby, compared with the state transfer process of keeping the Sleep state unchanged in period t: Sleep $\rightarrow$ Sleep $\rightarrow$ Standby, the former obviously consumes more energy and the final state is also Standby, so it does not reduce the subsequent SBS state switching delay; (3) By switching to the On state, for the SBS from the period $t - 1$ of the Sleep state to the period t of the Standby state and then to the period $t - 1$ of the On state, the SBS experiences two states switching. Each switching state will have an energy loss, and the SBS switching cost formula shows that, compared to the one-time energy loss from Sleep to On, the total energy consumption of going through the two state stages is more, and during the duration of the Standby state, the energy consumed is also much larger than that of the Sleep state. On the other hand, the SBS switching delay from Standby to On is much lower than that from Sleep to On. Next, we analyze the combined cost of these two different state transfer methods.

Assuming that the states of all SBSs except the target SBS under study are kept constant, and letting the duration of each period be τ, the additional costs incurred through the different states are:

Sleep $\rightarrow$ Standby $\rightarrow$ On

$$C1 = \alpha \left(P_{sleep}\tau + P_{stand}\tau + 2(\beta + \gamma) \right) + \frac{0.1}{e^{2\rho_i^t}} \tag{16}$$

Sleep $\rightarrow$ Sleep $\rightarrow$ On

$$C2 = \alpha \left(2P_{sleep}\tau + 2\beta + \gamma \right) + \frac{10}{e^{2\rho_i^t}} \tag{17}$$

Therefore, we have

$$\Delta C = C1 - C2 = \alpha \left[\left(P_{stand} - P_{sleep} \right)\tau + \gamma \right] - \frac{9.9}{e^{2\rho_i^t}} \tag{18}$$

where P_{sleep} and P_{stand} are fixed values, and the energy consumption bias, γ, is generally a small positive number, so the value of ΔC depends on the weighting factor, α, and the period duration, τ. These two parameters are closely related to the network strategy and scheme adopted by the operator. In this paper, the scheme we have chosen focuses on reducing energy consumption and the parameters chosen satisfy $\alpha \left[\left(P_{stand} - P_{sleep} \right)\tau + \gamma \right] \geqslant 9.9$, so $\Delta C \geqslant 0$. Therefore, our assumptions are consistent with the application scenario of this paper.

3.2. Initial State Strategy for SBSs Based on LSTM Traffic Prediction

To initially obtain the state of the SBS in the next period, we use LSTM models to predict the traffic of each SBS in the next period. LSTM is a recurrent neural network (RNN) architecture for making predictions on time-series data, and it is capable of learning long-term dependencies. Compared to traditional RNN models, LSTM models can eliminate the problem of gradient disappearance or gradient explosion. LSTM contains four components: cell states, input gates, output gates, and forgetting gates, which are specially developed to solve the gradient disappearance and gradient explosion problems in traditional RNN networks. Figure 3 shows the internal cell states of an LSTM model cell [22].
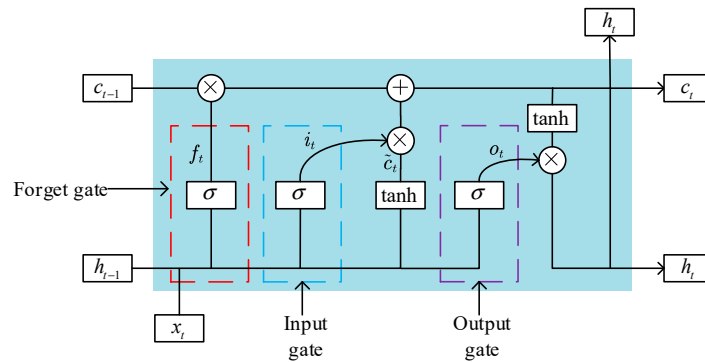

Figure 3. Internal cell states of LSTM model.

The cell state can be thought of as a long continuous chain through which a number of small interactions with cells are generated. It is a line consisting of c_{t-1} and c_t in Figure 3. The first step of the LSTM network is to identify the information that needs to be removed or forgotten from the cell. This process can be handled with a sigmoid layer/forgetting gate layer. The forgetting gate layer can be represented as follows

$$f_t = \sigma\left(w_f[h_{t-1}, x_t] + b_f\right) \tag{19}$$

where x_t, w_f, h_{t-1}, and b_f denote the new input, the weight function, the output of the previous period, and the bias value, respectively. $[h_{t-1}, x_t]$ denotes the sigmoid function, which represents the stitching of two matrices together. The output value of Equation (19) is either 0 or 1. When the output is 0 it means that the value is dropped completely, and when the output is 1 the value is preserved. In the next step, the data to be included in the cell state will be decided. The sigmoid layer, also known as the input gate layer, determines the value to be recovered. The 'tanh' layer produces a new vector of candidate values that can be combined with the current state.

$$i_t = \sigma(w_i[h_{t-1}, x_t] + b_i) \tag{20}$$

$$\widetilde{c}_t = \tanh(w_c[h_{t-1}, x_t] + b_c) \tag{21}$$

where i_t is the input layer and $\widetilde{c}_t$ is a new candidate vector. The new candidate vector value for the next state can be obtained from Equations (20) and (21), i.e.,

$$c_t = f_t * c_{t-1} + i_t * \widetilde{c}_t \tag{22}$$

The last step is to get the output. A sigmoid layer is executed. This selects which part of the cell state is the output, then inserts the cell state and multiplies the output generated by the cell state and the sigmoid gate through the 'tanh' layer (placing the value between -1 and 1). This is done in order to get the output of the decision part.

$$o_t = \sigma(w_o[h_{t-1}, x_t] + b_o) \tag{23}$$

$$h_t = o_t * \tanh(c_t) \tag{24}$$

where o_t is the output of the sigmoid layer and h_t is the output of the decision part. This new output will be input to another layer and then the cycle continues. The flow of predicting the traffic load of each SBS for the next period using the LSTM model is shown in Figure 4.
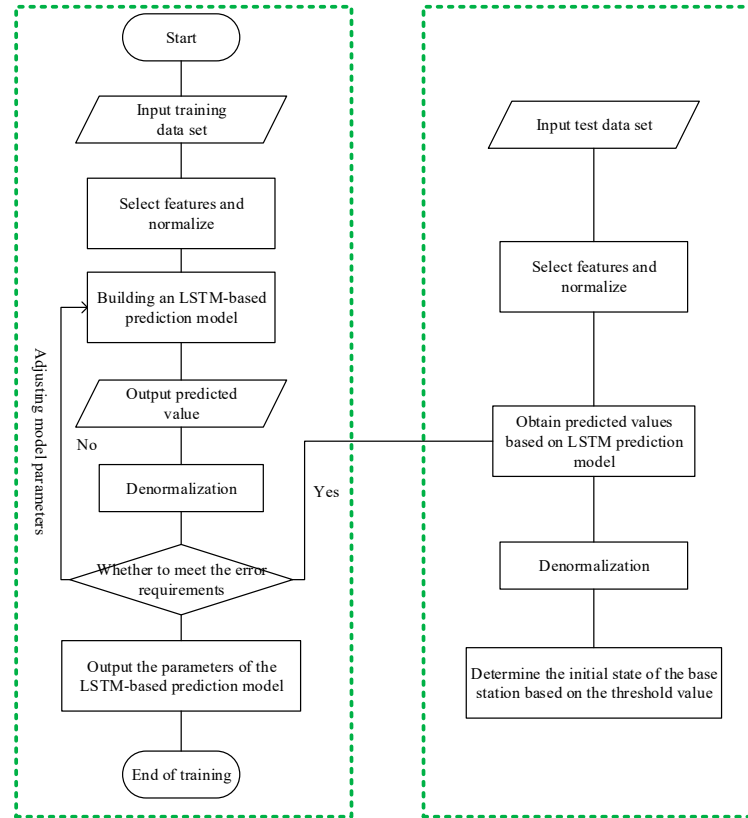

Figure 4. Training and testing flow chart of the LSTM network prediction model.

4. SBS Energy Efficiency Management Decisions

4.1. Problem Formulation

In Section 3, we used the traffic prediction method to obtain a preliminary set of SBS states by setting traffic thresholds and comparing the SBS with the thresholds based on the magnitude of the predicted traffic. Since these state sets are closely related to the threshold settings and there is no scientific mathematical model to strictly control setting the thresholds, this will lead to thresholds set being based on experience which are not universal in different scenarios; therefore, the SBS switching states obtained simply by traffic prediction and threshold settings do not maximize the performance of the network system. In addition, the focus on network performance indicators varies among operators when setting network service strategies, with some focusing on user experience and requiring lower network latency, while others are more sensitive to network energy consumption. All of this has an impact on the switching state of the SBS, so we have constructed a performance function optimization objective for the network:

$$\begin{aligned}
\min_{\boldsymbol{\rho}^t, \mathbf{b}^t} \quad & Q\left(\boldsymbol{\rho}^t, \mathbf{b}^t\right) = D^t\left(\boldsymbol{\rho}^t\right) + \alpha \sum_{i \in B_s} \left(P_i^t + P_{i,h}^t\right) \\
\text{s.t.} \quad & b_i^t = \{-1, 0, 1\}, \quad \forall i \in B \\
& \boldsymbol{\rho}^t \in \boldsymbol{\Omega}^t\left(B_{on}^t\right)
\end{aligned} \tag{25}$$

where α is a weighting factor that represents the operator's trade-off strategy between latency and energy consumption.

Assuming accurate results for traffic prediction based on the LSTM model, a preliminary SBS switching strategy is determined based on the threshold value. For when the SBS state is determined, we have constructed the user associated optimal SBS load optimization objective:

$$\min_{\boldsymbol{\rho}^t} \quad U\big(B_{on}^t\big) = D^t\big(\boldsymbol{\rho}^t\big) + \alpha \sum_{i \in B_{on}^s} (1 - q_i)\rho_i^t \mathcal{P}_i$$
$$\text{s.t.} \quad \boldsymbol{\rho}^t \in \boldsymbol{\Omega}^t\big(B_{on}^t\big) \tag{26}$$

The fixed energy consumption component of the SBS is ignored in (26) because after determining the set of turned-on SBSs, the fixed energy consumption of the SBS is a constant value, independent of the user association policy. Since the average wake-up delay in $D^t(\rho^t)$ is related to the state of the SBS in the previous moment, it is a segmentation function, which complicates the problem analysis. Therefore, in order to eliminate the effect of wake-up delay on user association, we propose a SBS state asynchronous switching mechanism. The mechanism can be described as follows: the MBS makes a SBS switching policy in the region of this period at the 0th second of period t, i.e., it determines B_{on}^t, B_{stand}^t, and B_{sleep}^t, and broadcasts these decisions to all SBSs. Assuming that the process of broadcasting has no time delay, i.e., the SBSs receive the broadcast information at 0 s, after receiving the decision information, the SBSs select different operation modes by combining their current states. For example, if SBS i is in B_{on}^t, if its current state is Sleep, then SBS i starts to enter the wake-up state immediately after receiving the broadcast signal. If its current state is Standby, it starts to enter the wake-up state at 9.9 s after receiving the broadcast signal. If its current state is On, SBS i continues to stay on according to the decision information, as shown in Figure 5.
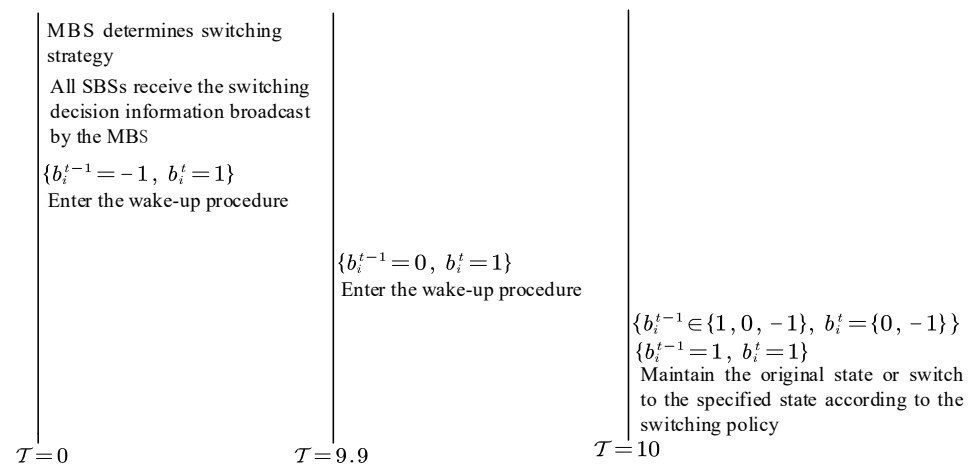

Figure 5. Asynchronous switching mechanism diagram.

According to the asynchronous switching mechanism, all the SBSs will enter the stable state at $\mathcal{T} = 10$ s, and all of the SBSs that need to be turned on at this time are also in a state that can serve the users immediately, so the users are not affected by the SBS state switching delay in choosing the optimal SBS access at this time. For a SBS switching from the Sleep and Standby state to the On state, it enters the switching procedure earlier than other SBSs, which causes additional energy loss and will have an impact on the SBS switching strategy. In this paper, we equate the SBS state switching delay as the energy cost caused by entering the wake-up procedure earlier, so that Equation (25) remains unchanged and Equation (26) ignores the SBS state switching delay, denoted as

$$\min_{\boldsymbol{\rho}^t} \quad U\big(B_{on}^t\big) = \sum_{i \in B_{on}^t} d_{i,s}^t\big(\rho_i^t\big) + \alpha \sum_{i \in B_{on}^t} (1 - q_i)\rho_i^t \mathcal{P}_i$$
$$\text{s.t.} \quad \boldsymbol{\rho}^t \in \boldsymbol{\Omega}^t\big(B_{on}^t\big) \tag{27}$$

Theorem 2. *In Equation (27), the optimal SBS for the user located at x to access in period t is*

$$i^{t^*}(x) = \underset{j \in B_{on}^t}{\arg\max} \frac{c_j^t(x)}{\left(1 - \rho_j^{t^*}\right)^{-2} + \alpha\left(1 - q_j\right)\mathcal{P}_j} \tag{28}$$

Proof. Let $\rho_j^{t^*}$ be the optimal SBS load for Equation (27) and assume that the user can be informed of the current $\rho_j^{t^*}$ through messages broadcast from the SBS. Since the objective function of (27) is a convex function and its feasible domain is also a convex set, (27) is a convex optimization problem. Therefore, we only need to check whether the following inequality holds for the optimal SBS load $\rho_j^{t^*}$:

$$\left\langle \nabla \left(\sum_{i \in B_{on}^t} d_{i,s}^t(\rho_i^t) + \alpha \sum_{i \in B_{on}^t} (1 - q_i)\rho_i^t \mathcal{P}_i \right), \boldsymbol{\rho}^t - \boldsymbol{\rho}^{t^*} \right\rangle \geqslant 0 \tag{29}$$

Let $\boldsymbol{a}^t(x)$ and $\boldsymbol{a}^{t^*}(x)$ be the user association probability vectors when the SBS load is $\boldsymbol{\rho}^t$ and $\boldsymbol{\rho}^{t^*}$, respectively, and the coverage area of the SBS can be obtained according to Equation (28) as

$$\mathcal{L}_i^t = \left\{ x \in \mathcal{L} \, \middle| \, i^t = \underset{j \in B_{on}^t}{\arg\max} \frac{c_j^t(x)}{(1 - \rho^{t*})^{-2} + \alpha\left(1 - q_j\right)\mathcal{P}_j} \right\} \tag{30}$$

Thus, the user association rule can be expressed as

$$a_i^{t*}(x) = 1 \left\{ i = \underset{i \in B_{on}^t}{\arg\max} \frac{c_j^t(x)}{(1 - \rho^{t*})^{-2} + \alpha\left(1 - q_j\right)\mathcal{P}_j} \right\} \tag{31}$$

Inequality (29) can be calculated as follows:

$$\begin{aligned}
&\left\langle \nabla \left(\sum_{i \in B_{on}^t} d_{i,s}^t(\rho_i^t) + \alpha \sum_{i \in B_{on}^t} (1 - q_i)\rho_i^t \mathcal{P}_i \right), \boldsymbol{\rho}^t - \boldsymbol{\rho}^{t*} \right\rangle \\
&= \sum_{i \in B_{on}^t} \left(\left(1 - \rho_i^{t*}\right)^{-2} + \alpha(1 - q_i)\mathcal{P}_i \right)\left(\rho_i^t - \rho_i^{t*} \right) \\
&= \sum_{i \in B_{on}^t} \left(\left(1 - \rho_i^{t*}\right)^{-2} + \alpha(1 - q_i)\mathcal{P}_i \right) \int_{\mathcal{L}} \phi_i^t(x)\left(a_i^t(x) - a_i^{t*}(x) \right)dx \\
&= \int_{\mathcal{L}} \phi_i^t(x) \sum_{i \in B_{on}^t} \frac{\left(1 - \rho_j^{t*}\right)^{-2} + \alpha(1 - q_j)\mathcal{P}_j}{c_j^t(x)} \left(a_i^t(x) - a_i^{t*}(x) \right)dx
\end{aligned} \tag{32}$$

Since $a_i^{t*}(x)$ is an indicator value determined by the maximum value of $\frac{c_j^t(x)}{\left(1 - \rho_j^{t*}\right)^{-2} + \alpha\left(1 - q_j\right)\mathcal{P}_j}$ in Equation (30), i.e., the value of $a_i^{t*}(x)$ also corresponds to the minimum value taken by $\frac{\left(1 - \rho_j^{t*}\right)^{-2} + \alpha\left(1 - q_j\right)\mathcal{P}_j}{c_j^t(x)}$. Therefore, we have $\sum_{i \in B_{on}^t} \frac{\left(1 - \rho_j^{t*}\right)^{-2} + \alpha\left(1 - q_j\right)\mathcal{P}_j}{c_j^t(x)} a_i^t(x) \geqslant \sum_{i \in B_{on}^t} \frac{\left(1 - \rho_j^{t*}\right)^{-2} + \alpha\left(1 - q_j\right)\mathcal{P}_j}{c_j^t(x)} a_i^{t*}(x)$, and the Inequality (29) holds. The theorem is proved. □

With the set of turned-on SBSs determined, Algorithm 1 gives the pseudocode of the optimal user association.

Algorithm 1. Enabling the optimal user association algorithm for the case of base station set determination

---

Initialization: step-length $\nu \in [0,1)$, small positive number ϑ, ε, π, and $\pi > \varepsilon$, $\boldsymbol{\rho}^{t^{(0)}} \in (0, 1 - \vartheta)^M$, $k = 0$, importing the $\widetilde{B}_{on}^t$ obtained by LSTM traffic prediction, $B_{on}^t = \widetilde{B}_{on}^t$;

while $\pi > \varepsilon$

    1. Broadcast load of each SBS $\boldsymbol{\rho}^{t^{(k)}}$;

    2. Users at x access SBS $i^{t^{(k)}}(x) = \underset{j \in B_{on}^t}{arg\max} \dfrac{c_j^{t^{(k)}}(x)}{\left(1 - \rho_j^{t^{(k)}}\right)^{-2} + \alpha\left(1 - q_j\right)\mathcal{P}_j}$, the coverage of SBS i is

$$\mathcal{L}_i^{t^{(k)}} = \left\{ x \in \mathcal{L} \Big| i^{t^{(k)}} = \underset{j \in \widetilde{B}_{on}^t}{arg\max} \dfrac{c_j^{t^{(k)}}(x)}{\left(1 - \rho_j^{t^{(k)}}\right)^{-2} + \alpha\left(1 - q_j\right)\mathcal{P}_j} \right\};$$

    3. The new load of SBS i is $C_i^t\left(\boldsymbol{\rho}^{t^{(k)}}\right) = \min\left\{ \int_{\mathcal{L}_i^{t^{(k)}}} \phi_i^t(x)dx, 1 - \vartheta \right\}$;

    4. $\boldsymbol{\rho}^{t^{(k+1)}} = \nu\boldsymbol{\rho}^{t^{(k)}} + (1 - \nu)C^t\left(\boldsymbol{\rho}^{t^{(k)}}\right)$;

    5. $\pi = \left\| \boldsymbol{\rho}^{t^{(k+1)}} - \boldsymbol{\rho}^{t^{(k)}} \right\|_2$, $k := k + 1$;

end while

Output: $\boldsymbol{\rho}^{t*} = \boldsymbol{\rho}^{t^{(k)}}$, $\mathcal{L}_i^{t*} = \mathcal{L}_i^{t^{(k)}}$

4.2. SBS State Admission and Correction Algorithm

In Section 3, we used a method based on traffic thresholds and traffic forecasts to initially determine the state of the SBS in the period t. However, such decisions are extremely dependent on the setting of traffic load thresholds, which makes the decision results have a large error. Therefore, we need to verify these decision results, accept the correct decisions and correct the incorrect ones, so as to obtain the optimal SBS state. In this paper, we use a heuristic algorithm to confirm the SBS states one by one, and its algorithm complexity is mainly related to the number of SBSs, so we can reduce the complexity by reducing the number of SBSs that need to be verified. In the initial state decision of the SBS, it is assumed that the traffic prediction results are accurate. We set the threshold, φ_b, to be high relative to the general empirical threshold, and therefore the SBS $\widetilde{B}_{on}^t$ which is determined to be turned on in period t by the traffic prediction will be directly accepted without further verification. Similarly, we set φ_s to be low, so that SBS $\widetilde{B}_{sleep}^t$, which is determined to be sleep at time t by traffic prediction, will be accepted directly. Therefore, the main issue to be considered is that for the SBSs in $\widetilde{B}_{stand}^t$, some of them need to be On and some need to be in Sleep mode, so we use the system performance objective function to determine the status of the SBSs in $\widetilde{B}_{stand}^t$ in turn. To facilitate reading, the system performance optimization objective function is rewritten as

$$
\begin{aligned}
&\underset{\boldsymbol{\rho}^t, \boldsymbol{b}^t}{\min} \quad Q\left(\boldsymbol{\rho}^t, \boldsymbol{b}^t\right) = D^t\left(\boldsymbol{\rho}^t\right) + \alpha \sum_{i \in B_s} \left(P_i^t + P_{i,h}^t\right) \\
&\text{s.t.} \quad b_i^t = \{-1, 0, 1\}, \quad \forall i \in B \\
&\qquad\quad \boldsymbol{\rho}^t \in \boldsymbol{\Omega}^t\left(B_{on}^t\right)
\end{aligned}
\tag{33}
$$

Algorithm 2 gives the pseudocode of the joint optimization algorithm for SBS state control and user association to solve (34).

Algorithm 2. Joint optimization algorithm for SBS state control and user association (SSC-UA)

Initialization: The initial $\widetilde{B}^t_{on}$, $\widetilde{B}^t_{stand}$, $\widetilde{B}^t_{sleep}$, and $\widetilde{\boldsymbol{b}}^t$ obtained by LSTM traffic prediction.

while $\widetilde{B}^t_{stand} \neq \varnothing$

　　The optimal load $\widetilde{\boldsymbol{\rho}}^{t*}$ under $\widetilde{B}^t_{on}$ is calculated from Algorithm 1, and the system performance objective function value $Q\left(\widetilde{\boldsymbol{\rho}}^{t*}, \widetilde{\boldsymbol{b}}^t\right)$ is calculated by combining the set $\boldsymbol{b}^{t-1}$ of base station states in period $t-1$ and the set $\widetilde{\boldsymbol{b}}^t$ of predicted states in period t;

　　Select SBS $i = \underset{j \in \widetilde{B}^t_{stand}}{arg\max} T^t_j$, $\widetilde{B}^t_{on} \leftarrow \widetilde{B}^t_{on} \cup \{i\}$, and at the same time adjust b^t_i in $\widetilde{\boldsymbol{b}}^t$ from $b^t_i = 0$ to $b^t_i = 1$ to get the new $\widetilde{\boldsymbol{b}}^t_{new}$, $\widetilde{B}^t_{stand} \leftarrow \widetilde{B}^t_{stand} - \{i\}$;

　　Run Algorithm 1 to compute the optimal load $\widetilde{\boldsymbol{\rho}}^{t*}_{new}$ under the current $\widetilde{B}^t_{on}$. Combine $\boldsymbol{b}^{t-1}$ and $\widetilde{\boldsymbol{b}}^t_{new}$ to compute the new system performance objective function value $Q\left(\widetilde{\boldsymbol{\rho}}^{t*}_{new}, \widetilde{\boldsymbol{b}}^t_{new}\right)$;

　　　if　$Q\left(\widetilde{\boldsymbol{\rho}}^{t*}_{new}, \widetilde{\boldsymbol{b}}^t_{new}\right) \leqslant Q\left(\widetilde{\boldsymbol{\rho}}^{t*}, \widetilde{\boldsymbol{b}}^t\right)$

　　　　$B^t_{on} = \widetilde{B}^t_{on}$, $B^t_{stand} = \widetilde{B}^t_{stand}$, $\boldsymbol{b}^t = \widetilde{\boldsymbol{b}}^t_{new}$;

　　　else

　　　　$B^t_{on} \leftarrow \widetilde{B}^t_{on} - \{i\}$, $B^t_{stand} \leftarrow \widetilde{B}^t_{stand} \cup \{i\}$;

　　　　break;

　　　end if

end while

while　$B^t_{stand} \neq \varnothing$

　　The optimal load $\boldsymbol{\rho}^{t*}$ under B^t_{on} is calculated from Algorithm 1, and the system performance objective function value $Q(\boldsymbol{\rho}^{t*}, \boldsymbol{b}^t)$ is calculated by combining the set $\boldsymbol{b}^{t-1}$ of base station states in period $t-1$ and the set $\boldsymbol{b}^t$ of predicted states in period t;

　　Select SBS $i = \underset{j \in B^t_{stand}}{arg\min} T^t_j$, $\widetilde{B}^t_{sleep} \leftarrow \widetilde{B}^t_{sleep} \cup \{i\}$, and at the same time adjust b^t_i in $\boldsymbol{b}^t$ from $b^t_i = 0$ to $b^t_i = -1$ to get the new $\boldsymbol{b}^t_{new}$, $B^t_{stand} \leftarrow B^t_{stand} - \{i\}$;

　　Combine $\boldsymbol{b}^{t-1}$ and $\boldsymbol{b}^t_{new}$ to compute the new system performance objective function value $Q(\boldsymbol{\rho}^{t*}, \boldsymbol{b}^t_{new})$;

　　　if $Q(\boldsymbol{\rho}^{t*}, \boldsymbol{b}^t_{new}) \leqslant Q(\boldsymbol{\rho}^{t*}, \boldsymbol{b}^t)$

　　　　$B^t_{sleep} = \widetilde{B}^t_{sleep}$, $\boldsymbol{b}^t = \boldsymbol{b}^t_{new}$;

　　　else

　　　　$B^t_{sleep} \leftarrow \widetilde{B}^t_{sleep} - \{i\}$, $B^t_{stand} \leftarrow B^t_{stand} \cup \{i\}$;

　　　　break;

　　　end if

end while

Output: B^t_{on}, B^t_{stand}, B^t_{sleep},

---

5. Simulation Results

5.1. LSTM Model Traffic Prediction

　　The dataset used in this paper is the data traffic of 4G LTE network SBSs on Kaggle (https://www.kaggle.com/datasets/habibmrad1983/lte-data-traffic?select=train.csv accessed on 15 November 2022). The dataset contains data traffic information for 57 SBSs for one year (from 23 October 2017 to 22 October 2018), where the data traffic is the total data capacity served by the SBS to all users in one hour. Therefore, each SBS contains 8733 data, and we used a total of 168 data from the most recent week (16 October 2018 to 22 October 2018) as the test set and 8565 data from the rest of the time (from 23 October 2017 to 15 October 2018) as the training set to train the model and compare the autoregressive integrated moving average model (ARIMA).

　　In predicting SBSs traffic using the LSTM network, we used the SBS traffic of the first five consecutive time series in time period t to predict the SBS traffic. $\boldsymbol{u}_m(t)$ denotes the input to the LSTM network, it can be expressed as

$$\boldsymbol{u}_m(t) = [h_m(t-5), h_m(t-4), \cdots, h_m(t-1)] \tag{34}$$

where $h_m(t)$ is the real traffic of SBS m in time period t. There are two hidden layers with nodes of 64 and 8, and an output layer that performs single-value prediction to obtain the SBS predicted traffic, $\widetilde{h}_m(t)$. The training prediction error is defined as follows

$$\theta = \frac{\sum_t \left| h_m(t) - \widetilde{h}_m(t) \right|^2}{\sum_t |h_m(t)|^2} \tag{35}$$

Figures 6 and 7 compare the results of the LSTM model and ARIMA model for the traffic prediction of Cell_000111 SBS and Cell_000113 SBS, respectively. The traffic prediction values of both prediction models have a high degree of fit with the actual values, and the overall trend remains basically consistent with the actual values. Figures 8 and 9 compare the prediction errors of the LSTM model and ARIMA model for Cell_000111 SBS and Cell_000113 SBS, respectively, and here we use the absolute error to represent the prediction error. Compared with the ARIMA model, the prediction error of the LSTM model is much lower, mainly because of the larger time-varying variability of the SBS traffic. In general, the ARIMA model has better prediction for data with little variation trend and that is more stable overall, such as the exchange rate or the price of a certain stock with more stable price, while the LSTM model has better prediction for more volatile data. Table 2 compares the RMSE and MAE of the two models, and also shows that the LSTM model has better performance in SBS traffic prediction.
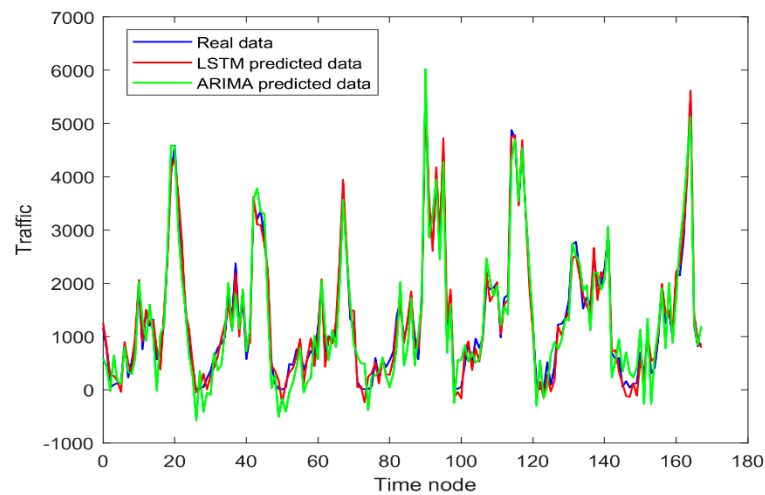
Figure 6. Traffic forecasting for BS Cell_000111.
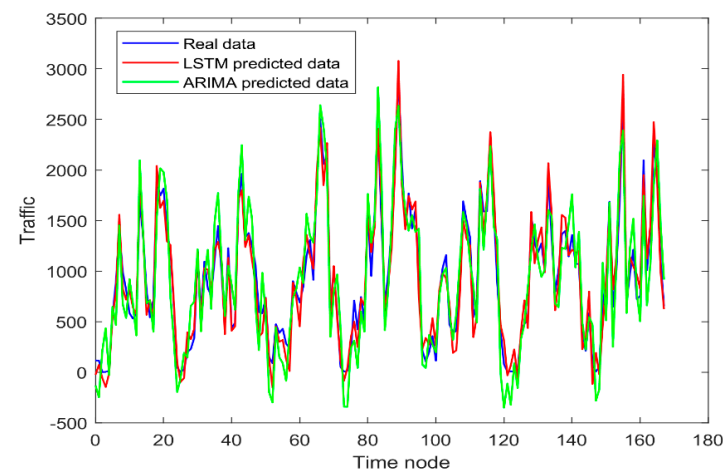
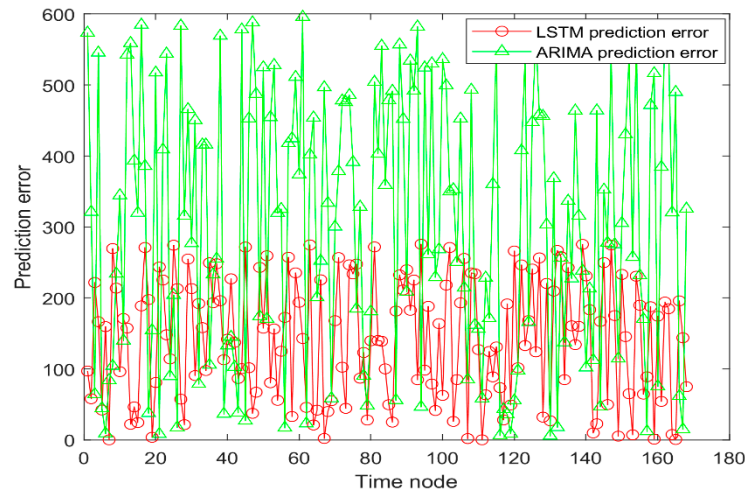
Figure 7. Traffic forecasting for BS Cell_000113.

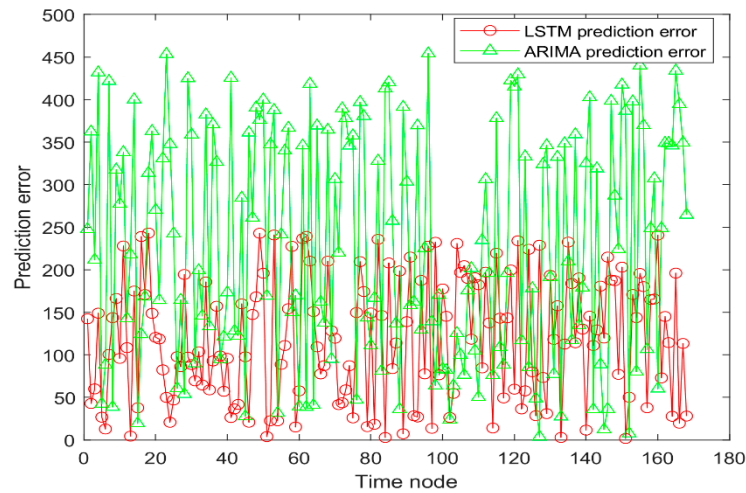**Figure 8.** Traffic prediction error of base station Cell_000111.

Figure 9. Traffic prediction error of base station Cell_000113.

Table 2. Comparing the prediction effect of two models.

BS Number	RMSE		MAE	
	LSTM	**ARIMA**	**LSTM**	**ARIMA**
Cell_000111	135.48	410.9	113.45	354.88
Cell_000113	61.81	474.92	54.79	65.72

5.2. Network System Performance

This section compares the proposed SSC-UA scheme with: (1) The joint optimization of SBS off and user association (BS-off-UA) algorithm, where the SBS states in BS-off-UA are Off and Sleep. (2) The optimization of user association without considering the SBS off (BS-on-UA) algorithm.

Figure 10 compares the power consumption and average delay of the SBSs for each scenario with different α. The average delay of the SBSs during the time period is calculated by the following equation.

$$\overline{D}^t(\boldsymbol{\rho}^t) = \frac{1}{n} \sum_{i \in B_{on}^t} d_{i,s}^t(\rho_i^t) + d_{i,w}^t(\rho_i^t) \tag{36}$$

where n denotes the number of SBSs in the On state in period t. From Figure 10, it can be observed that as the value of α increases, the energy consumption of all three schemes

decreases and the delay increases. This is because the larger α indicates that the proportion of energy consumption in the objective function increases, and the objective function optimization is more sensitive to energy consumption. The final result is to reduce energy consumption by increasing the delay, so as to reduce the value of the overall objective function. Comparing the SSC-UA and BS-off-UA schemes, we can find that the delay of the SSC-UA scheme is always lower than that of the BS-off-UA scheme, but the energy consumption is slightly higher than that of the BS-off-UA scheme. The reason for this result is that there is a Standby state for the SSC-UA scheme, and the energy consumption of these SBSs is larger than that of the Sleep state SBSs. On the other hand, the Standby state SBSs can be woken up faster than the Sleep state SBSs, and the system wake-up delay is much lower than that of the Sleep state SBSs, so the delay is reduced. Although the SSC-UA scheme reduces the latency at the expense of energy consumption, the algorithm complexity of the SSC-UA scheme is much lower than that of the BS-off-UA scheme. This is mainly due to the fact that in the SSC-UA scheme we predict the traffic of the SBSs through the LSTM network and make the initial decision on the state of the SBSs, which greatly reduces the number of SBSs to be considered in the heuristic algorithm. The BS-off-UA scheme needs to take all of the SBSs into account, and the algorithm complexity is much higher than that of the SSC-UA scheme. In addition, because all the SBSs are on in the BS-on-UA scheme, their energy consumption is kept at a high level, and there is no wake-up delay in the system, and the load of each SBS is in a more balanced state, so the overall system delay is low.
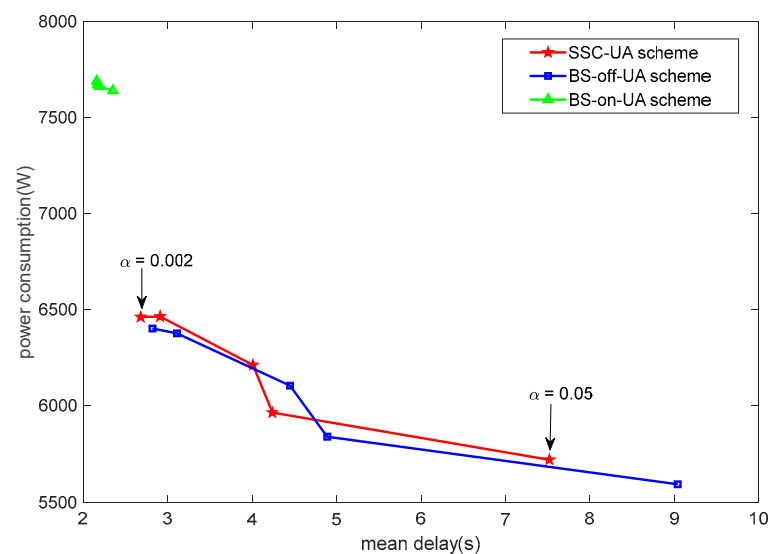
Figure 10. Comparison of average latency and total energy consumption of the three schemes.

Figure 11 compares the cumulative distribution function (CDF) of the delay of the SSC-UA scheme and the BS-on-UA scheme. Compared with the BS-on-UA scheme, the CDF of the delay of the SSC-UA scheme is on the left, which indicates that the network delay of the SSC-UA scheme is smaller than that of the BS-on-UA scheme in general.
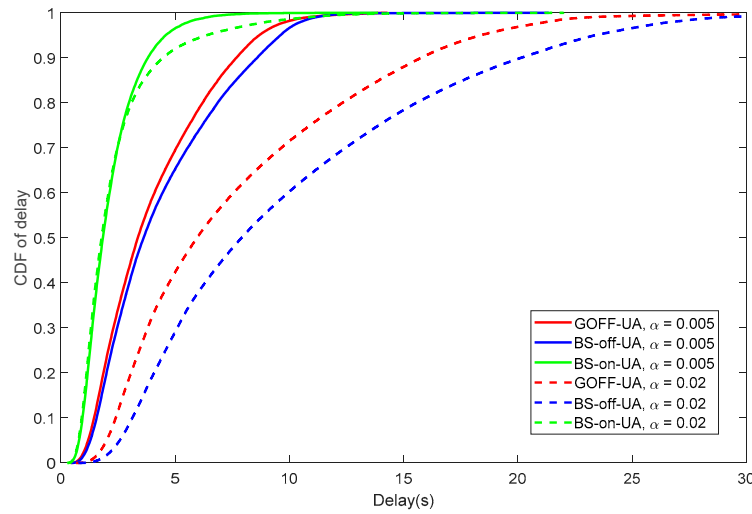
Figure 11. Comparison of time delay CDF of three schemes.

Figures 12 and 13 show the continuous 48-h network conditions of the SSC-UA and BS-on-UA schemes, including the traffic demand, the number of SBSs to be turned on, and the average delay and the energy consumption of the network for each hour, where the left vertical coordinate indicates the number of SBSs, and the right vertical coordinate indicates the normalized traffic demand, average delay and energy consumption. The number of SBSs to be turned on changes with the demand of network traffic in different time periods, and the number of SBSs to be turned on and the traffic demand curves of both schemes keep the same trend. The number of SBSs to be turned on in the SSC-UA scheme and BS-on-UA scheme is basically the same, and the number of SBSs to be turned on in the BS-on-UA scheme is slightly larger than that in the SSC-UA scheme in some cases. Comparing the average delay and energy consumption of the SSC-UA scheme and BS-on-UA scheme, the energy consumption of the SSC-UA scheme is higher than that of the BS-on-UA scheme in all time periods, while the average delay is lower for the SSC-UA scheme than the BS-on-UA scheme, which is consistent with the analysis results in Figure 10.
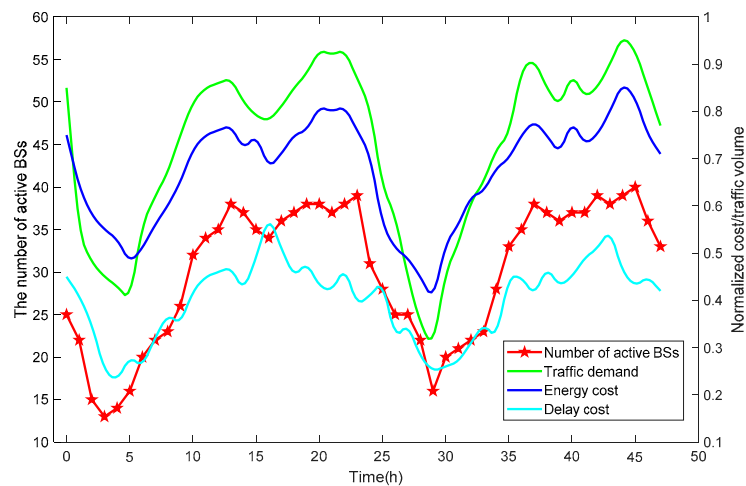

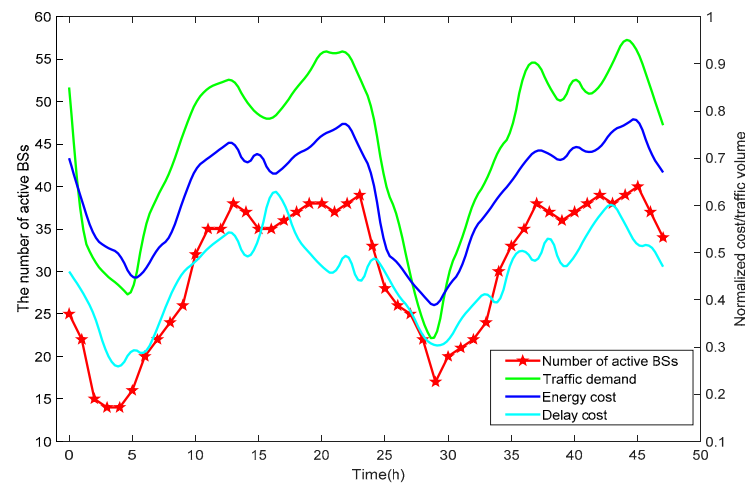**Figure 12.** Network Performance of SSC-UA Scheme.

Figure 13. Network Performance of BS-off-UA scheme.

Figure 14 shows the number of SBSs that are in On, Standby, and Sleep states at different times of the day in the SSC-UA scheme. From Figure 14, it can be seen that when the number of SBSs turned on is relatively small, it means that the current traffic demand of the network is relatively low, and the ratio of the number of SBS in the Sleep state to the number of SBS in the Standby state is larger accordingly. When the number of On SBS increases, the ratio of the number of Sleep state SBSs to the number of Standby state SBSs decreases. This is because when the traffic demand is small, there will be many regional SBSs with low traffic load, and more SBSs can be turned off through the objective function, and these SBSs will not be turned on in a short period of time, which in turn will not have a large impact on the wake-up delay. While when the traffic demand is large, the system will put more switched-off SBSs in the Standby state by weighing the impact of energy consumption and delay, so that they can be woken up quickly in the next period which reduces the average wake-up delay of the network.
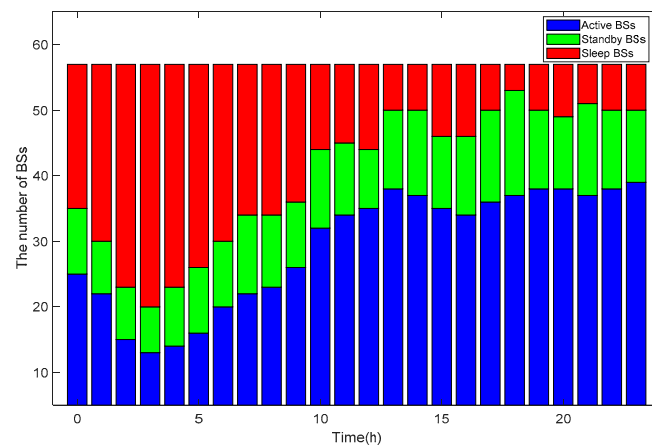

Figure 14. Number of base stations in different states of SSC-UA scheme.

6. Conclusions

This paper first analyzes the energy consumption characteristics of SBSs based on their hardware composition, then adopts a three-state SBS scheme, and finally constructs energy consumption models and delay models for different states of SBSs. Then, the LSTM network is used to predict the traffic load of each SBS in the next period, and the threshold determination method is applied to make a preliminary decision on the state that needs to be switched in the next period of the SBS. Finally, a joint optimization algorithm of SBS state control and user association is proposed to determine whether the preliminary decision

strategy is to be accepted or modified. Simulation results show that the proposed SSC-UA algorithm can effectively improve the energy efficiency and reduce the network service delay at the same time.

References

1. You, X.; Zhang, C.; Tan, X.; Jin, S.; Wu, H. AI for 5G: Research directions and paradigms. *Sci. China Inf. Sci.* **2018**, *62*, 21301. [CrossRef]
2. Hsieh, C.K.; Chan, K.L.; Chien, F.T. Energy-efficient power allocation and user association in heterogeneous networks with deep reinforcement learning. *Appl. Sci.* **2021**, *11*, 4135. [CrossRef]
3. Buzzi, S.; Chih-Lin, I.; Klein, T.E.; Poor, H.V.; Yang, C.; Zappone, A. A survey of energy-efficient techniques for 5G networks and challenges ahead. *IEEE J. Sel. Areas Commun.* **2016**, *34*, 697–709. [CrossRef]
4. Shen, Q.; Ma, Z.; Wang, S. Deploying C-RAN in Cellular Radio Networks: An Efficient Way to Meet Future Traffic Demands. *IEEE Trans. Veh. Technol.* **2018**, *67*, 7887–7891. [CrossRef]
5. Soh, Y.S.; Quek, T.Q.; Kountouris, M.; Shin, H. Energy efficient heterogeneous cellular networks. *IEEE J. Sel. Areas Commun.* **2013**, *31*, 840–850. [CrossRef]
6. Xu, F.; Li, Y.; Wang, H.; Zhang, P.; Jin, D. Understanding Mobile Traffic Patterns of Large Scale Cellular Towers in Urban Environment. *IEEE/ACM Trans. Netw.* **2016**, *25*, 1147–1161. [CrossRef]
7. Corroy, S.; Falconetti, L.; Mathar, R. Dynamic cell association for downlink sum rate maximization in multi-cell heterogeneous networks. In Proceedings of the 2012 IEEE International Conference on Communications (ICC), Ottawa, ON, Canada, 10–15 June 2012; pp. 2457–2461. [CrossRef]
8. Sun, R.; Hong, M.; Luo, Z.Q. Joint downlink SBS association and power control for max-min fairness: Computation and com-plexity. *IEEE J. Sel. Areas Commun.* **2015**, *33*, 1040–1054. [CrossRef]
9. Ye, Q.; Rong, B.; Chen, Y.; Al-Shalash, M.; Caramanis, C.; Andrews, J.G. User Association for Load Balancing in Heterogeneous Cellular Networks. *IEEE Trans. Wirel. Commun.* **2013**, *12*, 2706–2716. [CrossRef]
10. Baiocchi, A.; Chiaraviglio, L.; Cuomo, F.; Salvatore, V. Joint management of energy consumption, maintenance costs, and user revenues in cellular networks with sleep modes. *IEEE Trans. Green Commun. Netw.* **2017**, *1*, 167–181. [CrossRef]
11. Samulevicius, S.; Pedersen, T.B.; Sorensen, T.B.; Micallef, G. Energy savings in mobile broadband network based on load predictions: Opportunities and potentials. In Proceedings of the 2012 IEEE 75th Vehicular Technology Conference (VTC Spring), Yokohama, Japan, 6–9 May 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 1–5.
12. Liu, C.; Natarajan, B.; Xia, H. Small cell SBS sleep strategies for energy efficiency. *IEEE Trans. Veh. Technol.* **2015**, *65*, 1652–1661. [CrossRef]
13. Zhu, Y.; Wang, S. Joint traffic prediction and SBS sleeping for energy saving in cellular networks. In Proceedings of the ICC 2021-IEEE International Conference on Communications, Virtual, 14–23 June 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 1–6.
14. Lin, X.; Wang, S. Joint user association and SBS switching on/off for green heterogeneous cellular networks. In Proceedings of the 2017 IEEE Interna-tional Conference on Communications (ICC), Paris, France, 21–25 May 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1–6.
15. El Amine, A.; Chaiban, J.P.; Hassan, H.A.H.; Dini, P.; Nuaymi, L.; Achkar, R. Energy Optimization with Multi-Sleeping Control in 5G Heterogeneous Net-works using Reinforcement Learning. *IEEE Trans. Netw. Serv. Manag.* **2022**. [CrossRef]
16. 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]
17. Ju, H.; Kim, S.; Kim, Y.; Shim, B. Energy-Efficient Ultra-Dense Network With Deep Reinforcement Learning. *IEEE Trans. Wirel. Commun.* **2022**, *21*, 6539–6552. [CrossRef]

18. Son, K.; Kim, H.; Yi, Y.; Krishnamachari, B. SBS operation and user association mechanisms for energy-delay tradeoffs in green cellular networks. *IEEE J. Sel. Areas Commun.* **2011**, *29*, 1525–1536. [CrossRef]

19. Ashraf, I.; Boccardi, F.; Ho, L. Power savings in small cell deployments via sleep mode techniques. In Proceedings of the 2010 IEEE 21st International Symposium on Personal, Indoor and Mobile Radio Communications Workshops, Istanbul, Turkey, 26–30 September 2010; pp. 307–311. [CrossRef]

20. Vereecken, W.; Haratcherev, I.; Deruyck, M.; Joseph, W.; Pickavet, M.; Martens, L.; Demeester, P. The effect of variable wake up time on the utilization of sleep modes in femtocell mobile access networks. In Proceedings of the 2012 9th Annual Conference on Wireless On-Demand Network Systems and Services (WONS), Courmayeur, Italy, 9–11 January 2012; pp. 63–66. [CrossRef]

21. Newell, C. *Applications of Queueing Theory*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2013.

22. Kurri, V.; Raja, V.; Prakasam, P. Cellular traffic prediction on blockchain-based mobile networks using LSTM model in 4G LTE network. *Peer-to-Peer Netw. Appl.* **2021**, *14*, 1088–1105. [CrossRef]