Received: 25 April 2023 | Revised: 14 June 2023 | Accepted: 19 June 2023 | Published online: 19 June 2023

RESEARCH ARTICLE

Journal of Data Science and Intelligent Systems
2024, Vol. 2(2) 113–121
DOI: 10.47852/bonviewJDSIS3202998

BON VIEW PUBLISHING

Federated-Based Deep Reinforcement Learning (Fed-DRL) for Energy Management in a Distributive Wireless Network

Victor Kwaku Agbesi[1,*], Noble Arden Kuadey[2], Collinson Colin M. Agbesi[3] and Gerald Tietaa Maale[1]

[1]*School of Computer Science and Engineering, University of Electronic Science and Technology of China, China*

[2]*Department of Computer Science, Ho Technical University, Ghana*

[3]*Computer Science Department, Koforidua Technical University, Ghana*

Abstract: Studies on developing future generation wireless systems are expected to support increased infrastructure development and device subscriptions with densely deployed base stations (BSs). Economically, decreasing BS energy consumption levels and achieving "greenness" remain key factors for the giant industry. Some research works have proposed deep reinforcement techniques to solve energy management (EM) issues in cellular networks. However, these techniques are inefficient in a distributive network environment and expose the devices to privacy issues. Federated learning (FL) is proven to enforce device privacy and train models distributively. Thus, this work proposes an autonomous switching mode framework for BSs based on federated-deep reinforcement learning (Fed-DRL) to address the aforementioned challenges encountered by prior studies. Specifically, we deploy multiple DRL agents to influence the decision of the BS for EM. On the other hand, to make DRL-based decisions feasible and satisfy device quality-of-service, we train the DRL agents distributively by employing the FL concept. The results show the effectiveness of our proposed framework under distributed network scenarios compared with other benchmark algorithms.

Keywords: deep reinforcement learning, device satisfaction, energy management, federated learning

1. Introduction

Energy management (EM) is a key objective in next-generation networks. Studies reveal that base stations (BSs) consume 70–80% of operational energy, which account for 3% of the total energy produced globally and 2–4% of carbon dioxide (CO2) emissions, doubling the energy consumption (EC) rate of 15–20% annually (Githiru et al., 2011). According to Habibi et al. (2019), data traffic density foretells a 1000-fold increase in the next decade. This has motivated researchers to concentrate on EM in wireless networks. As the telecommunications industries sought to promote "greenness," innovative ideas (such as EARTH and Green Touch projects) were initiated (Ahmed et al., 2017).

Researchers came up with an alternative relay-station-based energy-efficient (EE) switching algorithm that turns off BSs during low-traffic intervals. Device association and resource allocation problems were considered in Zhuang et al. (2016) and Mesodiakaki et al. (2014). The model-based technique is set to achieve an EE solution by augmenting a particular independent task for the real-time slot. In Ashraf et al. (2010), sleep strategies were proposed to regulate devices and core networks to optimize power consumption in cellular networks. With the ability to utilize bandwidth more effectively than current cellular communication technologies, cognitive radio (CR) has recently emerged as the most promising next-generation communication technology. The precision of the sensing findings has a significant impact on the CR system's performance. But sensing ambiguities like false alarms and missed detections result in underuse of the spectrum and significant interference to the main user, respectively. In order to solve the problem of sensing ambiguities, the typical frame structure was modified in this research by adding two sensor slots as well as a gearbox slot. Sensing results are recorded in a flag bit up to that period, and the initial sensing slot is kept small and fixed for the specified likelihood of detection (Bala & Ahuja, 2023). When the flag bit status in the current frame differs from the previous frame, the second sensing slot (optional) is utilized. The second sensing slot was optimized to increase the effective throughput and energy efficiency of the secondary communication system while taking the trade-off between sensing throughput and energy efficiency into account. The simulation results are shown to demonstrate the viability of the suggested system, which uses a redesigned frame structure to outperform existing schemes in terms of throughput and energy efficiency.

*__Corresponding author:__ Victor Kwaku Agbesi, School of Computer Science and Engineering, University of Electronic Science and Technology of China, China. Email: vkagbesi@std.uestc.edu.cn

Recently, reinforcement learning (RL) schemes became one of the techniques researchers adopted to solve EM problems in cellular networks (Cheng et al., 2017). Most of these optimization policies focused on deep reinforcement learning (DRL) and Q-learning, taking into account the load on BSs. Liu et al. (2018) use a Deep Q-Network (DQN)-based ON/OFF policy for BSs to control how much energy they use. However, these techniques are inefficient in a distributive network environment. This is because of the unpredictable nature of devices on the network which can be problematic because device mobility impacts the network signal at each period.

Federated learning (FL) (Jiang et al., 2020), as introduced, collaboratively trains a model under the moderation of a main server while protecting the decentralized training set (or distributed) without necessarily exposing its private attributes. It serves as a wrapper over the well-known traditional machine learning (ML) techniques. Its mechanism employs centralized datasets to train similar models that have the ability to train models distributively, record low power consumption, and ensure device privacy. For this reason, our approach solves the EM problem and considers the device quality-of-service (D-QoS) requirement.

1.1. Contribution

In this work, we propose federated-DRL (Fed-DRL), an autonomous switching mode framework for BSs that employs FL to train DRL agents in a distributive manner. FL improves learning performance, ensures device-data privacy, and solves the EM problem. We formulate the switching mode problem as a Markov decision process (MDP), where we define states, actions, rewards, and the next (future) states. Finally, we optimize EC and meet the D-QoS satisfaction requirement with the smallest quantity of active BSs. Extensive simulations and comprehensive analysis are presented in terms of convergence rate, D-QoS satisfaction, and EC in a distributed fashion. The effects of Fed-DRL and the traditional DRL agent are both examined.

The work is structured as follows: Section 2 provides knowledge work, and Section 3 describes the system model. Section 4 deals with problem formulation to solve the BS's EM and provide D-QoS satisfaction. Furthermore, we describe our Fed-DRL framework in Section 5. We provide simulation results and analysis in Section 6 and conclude with Section 7.

2. Knowledge Work

Decreasing operational expenditure (OPEX) has been a key objective in telecommunications, since BS EC increases daily. Wang and Zheng (2015) proposed an EM procedure to predict and adjust the traffic load of BSs according to the mobility of the device. Although EM procedures are shown, there is a high level of EC due to the highly complex and stochastic existence of distributive training.

Recently, RL models have made a lot of progress and have become an interesting area for reducing energy use. In Chen et al. (2022), Hoffmann et al. (2021), Hsieh et al. (2021), Kim et al. (2022), Lee et al. (2020), and Sun et al. (2020), deep learning-based aiding algorithms were introduced to control features in an end-to-end fashion. Data-driven schemes were applied for various optimization sections in wireless communication problems (Li et al., 2018; Sheng et al., 2021; Sun et al., 2020; Xiong et al., 2020). These researchers introduced several learning systems to monitor data consumption rates and determine the EE of devices on the network. These developments inspire many researchers to use DRL-based resources to learn successful BS sleeping strategies. Liu et al. (2018) suggested improving the typical DQN model with action-wise replay capability and flexible compensation balancing to solve non-stationary traffic challenges. However, the distributive network scenario's high-dimensional state and action space can impact the efficacy of the aforementioned methods. To minimize the network's high-dimensional state space and activity, Li et al. (2014) suggested a method based on the actor-critic approach to derive the ON/OFF technique of the BS for the EC issue in the network, and they also included transfer learning (TL) in the actor-critic algorithm to utilize information gained over time. Sharma et al. (2017) also use a TL method for the ON/OFF toggling of BSs in diverse cellular networks. In Nishio and Yonetani (2019), FL was used to randomly choose clients with resource restrictions, allowing the server to combine as many clients' updates as possible and expedite the performance improvement of ML models. To lower OPEX in EM, they analyzed the issue of deploying FL in a cellular network utilized by heterogeneous devices with diverse data resources, computing capabilities, and wireless channel conditions.

In actual use, battery-powered IoT devices complete local training and communicate wirelessly with the main server. However, the constant communication between IoT devices and the main server would require many resources. The authors propose using the intelligent reflecting surface (IRS), a newly developed technology, to re-organize the wireless propagation environment to use the most available resources. In particular, we focus on the crucial problem of energy efficiency in the reconfigurable wireless communication network. Zhang and Mao (2022) propose an energy minimization issue in an IRS-assisted FL system subject to the complete training time restriction. The parameters are jointly configured using an iterative resource allocation technique with quick convergence.

The authors also adopt the FL framework for computation offloading optimization and prove their EM problem (Han et al., 2019; Jiao et al., 2021; Ye et al., 2020). Due to the lack of terrestrial connectivity and the limited battery life of FL users, certain FL tasks may not be possible. Pham et al. (2022) use unmanned aerial vehicles (UAVs) and wireless powered communications (WPCs) for FL networks to solve these issues. The UAV with edge computing and WPC capabilities is deployed as an aerial energy source and as an aerial server to carry out FL activities in order to provide sustainable FL solutions. They put forth the energy-efficient FL (E2FL) algorithm, a combined algorithm of UAV placement, power control, transmission time, model accuracy, bandwidth allocation, and computing resources, with the goal of reducing the overall EC of the aerial server and users after solving the original no convex problem effectively.

Although BS scheduling and EM problems in wireless networks using these techniques have been investigated, only some considered training DRL agents in a distributed fashion. Our work utilizes a federated-DRL-based framework that adjusts preferences for BS EC while still meeting the D-QoS satisfaction requirement.

3. System Model

3.1. Network model

With a collection of BSs as: Every BS $k \in \mathrm{K}$, is linked using devices $i \in \mathrm{I} = \{1, 2, 3, \ldots, I\}$. Assume each device is made up of a local set data element. Every $\beta_i = \{x_{il}, y_{il}\} \times l = 1, x_{il} \in \mathbb{R}$

represents the input vector of the device i while the output is y_{il} respectively. Each device trains a local-FL (L-FL) model on its own dataset. The input of all L-FL models is used to create a global-FL (G-FL). For each BS, B denotes framework bandwidth, and P_k^t denotes total energy utilization for transmission in watts. The path loss in this scenario is computed as:

$$PL(d_{ik}) = 20 \cdot [32.4 + (\log F) \cdot 20 + (\log d_{ik})], \tag{1}$$

where F signifies frequency bandwidth while d_{ik} means the interval between devices i and BS k. We adopt a channel model g_{ik} as (Buzzi et al., 2016):

$$g_{ik} = \left[10^{\frac{-PL(d_{ik})}{20}} \cdot \{b_{ik}\varphi\}^{0.5}\right], \tag{2}$$

where b_{ik} depicts its channel gain, and φ denotes the enormous scope shadow blurring. As the standard deviation and the Gaussian arbitrary variable are signified by σ, PL is indicated as $PL(d_{ik})$. One important use case in our network is the ability to manage transmission between BSs and devices. A device can only be connected to one BS. However, a beam forming system can link a device to multiple BS in edge computing. In this study, we implement a maximum received signal power (MRSP)-based device combination model (Tabassum et al., 2014) to help associate devices on the network with a BS. MRSP is the conventional device affiliation system in which the device decides on the BS from which the full instantaneous signal power is received. We enable a down-link broadcasting rate for devices beyond loss of consensus as in Tian and Jiang (2021). The signal-to-interference-plus-noise ratio (SINR) of a device i aligned to a BS k is denoted as:

$$SINR_{ik} = \frac{\left|g_{ik}^H \times \omega_{ik}\right|^2}{\sum_{u \neq i} \sigma^2 + \left|h_{ik}^H\right|}, i \in \mathrm{I}, k \in \mathrm{K}, \tag{3}$$

where ω_{ik} denotes the beam-formed weights from the BS k to the device i and the spectral density, σ^2, of an added substance white Gaussian variable. The successful data rate can be resolved using a characterized channel transfer bandwidth, B, and SINR as in Tian and Jiang (2021). Adopting the Shannon capacity formula (SCF), the transmit rate r_{ik} of the device i linked with the BS k as:

$$r_{ik} = \{x_{ik} \cdot B \cdot log_2(1 + SINR_{ik})\}, \tag{4}$$

where x_{ik} is a portion of the BS bandwidth B allotted to the device i. For each BS k, an M/G/1 processor sharing system was used, with packets arriving in the Poisson process (PP) with parameter λ_{ik}. The resource time for device i and BS k is indicated with boundary $h_{ik} = \frac{\lambda}{r_{ik}} = \frac{1}{r_{ik}^*}$, and the standardized feasible rate is r_{ik}^*. The average packet size is denoted by λ and is expressed as the mean. The entire time a device demands a resource while in procession for BS k is represented by the average delay τ_{ik}. The delay's average encountered by the device i on BS k is indicated by the property of M/G/1 PS queue as follows:

$$\tau_{ik} = \left\{r_{ik}^* - \lambda_{ik}\right\}^{-1}. \tag{5}$$

Historically, the traffic load has been determined by the arrival rate of systems and the BS sequence. The computational traffic model is unfeasible due to the unpredictable traffic of devices. The notations used in the system architecture are summarized in Table 1.

Table 1
Major notation

Notation	Description
$k/\|\mathrm{K}\|$	Number of BSs
B	System bandwidth
I	Maximum number of devices
λ	Packet arriving rate
P_{max}^t	Maximum transmit per BS
σ	Noise power spectrum density
F	Carrier frequency band
d_{ik}	Distance between BS k and device i
δ	Shadowing effects
η	Steepness coefficient in satisfaction function
r^{min}/τ^{max}	Device demand
P_k^t	Power consumption (active mode)
P_k^s	Power consumption (sleep mode)
L	Mean packet size
x_{ik}	Fraction of the bandwidth of BS k allocated to device i
ρ_k^*	The normalized traffic load on BS k
r_{ik}	Transmit rate of the device i attached to the BS k
τ_{ik}	The average delay e experienced by the device i on BS k
$\mathcal{T}$	Traffic intensity coefficient in the traffic model

3.2. Traffic model

We base this model on the network setting evaluations indicating the difference in BS traffic arrival. The device structure shift is used to model device traffic variation based on the regular trapezoidal traffic pattern. Because cellular traffic is highly dynamic in time and space, we defined it as a time-homogeneous PP with a traffic circulation intensity parameter $\mathcal{T}$ and a stabilized rate $f(t)$ (Alam & Dooley, 2015). This changes the trapezoidal traffic design within a period, differing its probabilistic traffic system $\chi(t)$ in the network. With this, $\chi(t)$ is interpreted as:

$$\chi(t) = (f(t) \cdot \psi), \tag{6}$$

where the function $\psi \sim$ Poi $(\{\mathcal{T}\})$ represents its random variable with parameter $\mathcal{T}$. As a result, if each device arrived at a BS k and sustained a service time h_{ik} per second, standardized traffic load at each BS k at time t is regarded as:

$$\rho_k(t) = \sum_i c(\lambda_{ik}) y_{ik} z_k f(t), \forall = 1, \ldots |\mathrm{K}|, \tag{7}$$

where a portion of B is z_k at BS k allocated to devices.

3.3. Energy consumption

In this part, BSs are made up of continuous load-dependent and non-load-dependent energy utilization that corresponds to their traffic volume (Abdulmula et al., 2019). When the BS is loaded, load-dependent energy utilization is provided by the energy amplifier and transceiver. The load-dependent energy utilization has been mounted with the standardized traffic-load $\rho_k^*(t)$ to determine the energy usage at a stage. Therefore, the overall energy utilization at time t is

$$P_j^t = \left[\{P_k^l \cdot P_k^t\} + P_k^c\right], \tag{8}$$

where P_k^l denotes load dependence, P_k^c denotes constant energy flow, and BS k traffic load is P_k^t at the time slot t. Given the difference in traffic demands, our proposed framework is to regulate the complex network's energy utilization by switching idle BS OFF to efficiently manage energy.

3.4. Utility model

The prerequisite for managing the EC of BSs is to meet D-QoS requirements and ensure device network scalability. To guarantee this, the delay and requisite communication rate must be ensured. With the complexities associated with device i behavior, traffic request per connection, and service permeability, we model this part using a sigmoid function. To meet device satisfaction, we adopt $\xi(.)$ as a function that sets dissimilar optimized goals for both rate constraint and delay as identified in Delaram et al. (2021). In our work, we define our QoS utility as the user's satisfaction with either data rate or delay depending on the application type in our study. The device satisfaction on the rate is defined as:

$$\xi(r_\tau) = \left\{e^{-n(r_\tau - r_\tau^{\min})} + 1\right\}^{-1}, \tag{9}$$

where $r_\tau^{\min}$ is the least rate demand of device i and the sustained n responsible for the device satisfaction curve. We can validate that (a)$\xi(r_\tau)$ is a monotonous function in relation to rI since individual devices would be satisfied if a higher output is achieved above its least requirement, otherwise (b) $\xi(d_\tau)$ of each device is mounted between 0 and 1, $\xi(r_\tau) \in [0, 1]$. Delay on device satisfaction is set as:

$$\xi(d_\tau) = \left\{e^{-n(d_\tau^{\max} - d_\tau)} + 1\right\}^{-1}, \tag{10}$$

where $\tau^{\max}$ denotes the optimum tolerant delay necessary to meet the upper bound delay for the device i. In analyzing device network scalability in our distributed network environment, we set an assumption on the following:

1. Time stationarity = $l_{ik}(t) = l_{ik}$
2. Device Independence, $l_{ik} = f(s_i, s_k)$
3. Switch mode (ON/OFF) = $l_{ik} \in [0, 1]$

Note that the transmission link between device i and BS k is l_{ik}. The geometric distance (d) between the device i and BS $k$is denoted as:

$$d_{ik} = \left[\sqrt{(s_i)^2 + (s_k)^2}\right]. \tag{11}$$

4. Problem Formulation

A tuple $< S, A, \mathcal{P}, R, \gamma >$ is defined as MDP, where A and S represent the finite set of all legitimate states, as well as the finite set of all legitimate actions. The state $\mathcal{P} : S \times A \rightarrow P(S)$ represents a transition probability, where $P\left(s_{(t+1)}|s_t, a_t\right)$ is the transitioning probabilities of time $t+1$ into states, and s_{t+1} is an agent begins action a_t execution at time t. Our reward function $R : S \times A \times S \rightarrow R$ and γ, where $R_t = (s_t, a_t, s_{t+1})$. With the action (a), the switch dynamically situates the BS k into sleep/off mode, i.e., sets $a_k = 0$, else sets $a_k = 1$ (Büttner et al., 2021; Sun et al., 2022).

At any stage t with a traffic-load state $s^{(t)}$, the objective is to discover the ideal policy π^* which corresponds to $s^{(t)}$ of an action $a^{(t)}$ that exploits the action-value task. Let $U\left[|\, s^{(0)}, s^{(1)}, \ldots, s^{(t)}\,|, a^{(t)}\right]$ signify a Markov chain utility. The continuing cumulative reduced reward of $s^{(t)}$ at stage t is assumed:

$$R(s^{(t)}) + \gamma^1 R(s^{(t+1)}) + \gamma^2 R(s^{(t+2)}) + \ldots . \gamma^n R(s^{(t+n)}), \tag{12}$$

where the discount factor $\gamma \in [0, 1]$. The state-value task of a random policy at the stage t is denoted as:

$$V_\pi(s^{(t)}) = E\{\sum_{t=0}^{\infty} \gamma^t R(s^{(t)})\}. \tag{13}$$

The goal of MDP is to find an optimum strategy to exploit the upcoming reward of the resulting agent. From Markov's theory, the policy π can be defined as:

$$V_\pi(s^{(t)}) = E\{R(s^{(t)}, a^{(t)}) + \gamma \sum_{s'} P(s'|s^{(t)}, a^{(t)}) V_\pi(s')\}, \tag{14}$$

where $V_\pi(s^{(t)})$ is the expected utility given the optimum strategy π. With Bellman's mathematic theory, the state-value task for the best policy π is expressed as:

$$\begin{aligned} lV_{\pi_*}(s^{(t)}) = arg\ \max_{a^{(t)} \in \mathbb{A}} \{R(s^{(t)}, a^{(t)}) \\ + \gamma \sum_{s'} P(s'|s^{(t)}, a^{(t)}) V_\pi(s')\}, \end{aligned} \tag{15}$$

where the present reward is $R(s^{(t)}, a^{(t)})$, and the discount factor is γ while the current utility is $V_{\pi_*}(s^{(t)})$ and its future utility is $V_\pi(s')$, respectively. Our objective is to find the optimum strategy $\pi_* = arg\{\max_\pi V_\pi(s)$ that affects the on/off switching results for each BS, which reduces the overall EC in this scenario. With our cell activation, the state space, action space, and reward task are as follows:

a. *State Space:* For $(\forall t = 0, 1, 2, 3, \ldots, 23)$, the state space of the devices and BSs is represented as:

$$S_{(i,k)} = (t, pt, T_{t-1}^{active}, T_{t-1}^{sleep}). \tag{16}$$

In a day, T_{t-1}^{active} and T_{t-1}^{sleep} denote the predicted active and sleep states at a time t. The traffic arriving rate is denoted as pt.

b. *Action Space:* The learning agent must set the ON and OFF strategy at each time t for the best reward. We establish the action as $a_k = 0$, otherwise set $a_k = 1$. Since $a_k \in [0, 1]$ as formulated in Boltzmann distribution probability (Shingu et al., 2021).

c. Reward: Our primary aim is to manage BS EC level and, on the other hand, meet D-QoS requirements. For managing EC, $E(s, a)$, the agent obtains a reward that proves an enhanced system EC built on the switching procedure of the BS. With the D-QoS satisfaction $\delta(s, a)$, the delay optimal metrics and rates are used to assess each device performance. On the basis of the device satisfaction, the agent sets an r^{min} threshold value. With this, we define our reward, $R(s, a)$ as:

$$R(s, a) = (\alpha \cdot E(s, a) + \beta \cdot \delta(s, a)). \tag{17}$$

Note that α and β denote the coefficients that show the significance of device satisfaction and EC. A secularization function is often used to describe a multi-objective reward that reduces multiple goals to a sin-

gle scalar that can be optimized. In this work, we define the given transition, $D_{a'} = (s_{a'}, a_{a'}, s', r_{a'})$, in the FL setting as collected by agent a', and pairs of actions and states $D_{b'} = (s_{b'}, a_{b'})$ as collected by agents b'. The objective is to distributively build policies $\pi^*_{a'}$ and $\pi^*_{b'}$ for agents a' and b', respectively. For simplicity, we considered two federated participating devices. However, the same mechanism could be extended between the two or more class agents. We illustrate state-action policies and Q-functions with respect to a' and b', as: $\left[s_{a'} \in S_{a'}, a_{a'} \in \mathrm{A}_{a'}, \pi^*_{a'}, Q_{a'} \in S_{a'}\right]$ and, $\left[s_{b'} \in S_{b'}, a_{b'} \in \mathrm{A}_{b'}, \pi^*_{b'}, Q_{b'} \in S_{b'}\right]$, respectively. Based on these assumptions, we aim to learn policies $\pi^*_{a'}$ and $\pi^*_{b'}$ of high quality for agents a' and b' by performing efficient switching and meeting D-QoS status.

5. Methodology

We briefly explain the concept of FL and describe our Fed-DRL framework.

5.1. FL algorithm

Algorithm 1 describes the adopted FL procedure. Specifically, the local device and the global server communicate iteratively to train a model. In this setup, the global server computes a weighted average of the resulting models after each device performs one step of gradient descent on an intermediate model using its own local data. There are five parameters: the local mini-batch size B, the number of local epochs E, the fraction of devices i_c to choose for training, a learning rate decay λ, and a learning rate η. Mostly, for stochastic gradient decent (SGD), B, E, λ, η are used. Before the server is updated, the number of iterations needed via the local device is E. The global model w_0 is randomly initialized. A single round of communication involves the following: A subset of devices $S_t, |S_t| = i_c \cdot I1$, shares current global model w_t to all devices in S_t. After they have revised their local models w^i_t to the distributed model, $w^i_t \leftarrow w_t$, each device creates its own local data into batches B to perform E epochs of SGD. Lastly, the local devices upload their trained model w^i_{t+1} to the server to generate new global models by calculating the weight of all local device models.

Algorithm 1: Fed-DRL algorithm

1: Global server:
2: initialize ω_0
3: **for** each round $t = 0,1,2,3, \ldots \ldots$, **do**
4: S_t = (random set of max $(i_c \cdot I,1)$ devices)
5: for each device $I \in St$ in parallel **do**
6: $\omega^i_{t+1} \leftarrow$ DeviceUpdate (i, ω_t)
7: Compute: $\omega_{t+1} \leftarrow \sum_{t=1}^{I} \frac{n_i}{n}(i, \omega_t)$
8: **end for**
9: **end for**
10: DeviceUpdate:
11: **for** each round local epoch, i from 1 E **do**
12: batches $\leftarrow$(*data* P_i *split into batches* B)
13: **for** batch b in batches **do**
14: $\omega \leftarrow \omega - \eta \nabla \pi(\omega; b)$
15: **end for**
16: **end for**
17: return ω to the server

5.2. The proposed Fed-DRL for EM

We describe and demonstrate the effectiveness of our proposed Fed-DRL framework that manages BSs EC in a distributed network architecture. In this process, the local device and the global server communicate with each other iteratively and effectively to train the model. As stated, we consider multiple DRL agents to train our model.

In our distributive scenario, the corresponding agents in a continuous action space schedule the EC of the BS. All agents are considered to initiate their learning process synchronously and select their actions through an initial distribution. Furthermore, the agents add a neural network to receive the initial state function $Q(s_t, a_t)$ and measure $A(s_t, a_t)$ in order to improve model efficiency. Note that the action A is controlled by the switching (ON/OFF) method. Thus, we fixed the action as $a_k \in \{0, 1\}$. If BS $k$is in OFF mode, it indicates $a_k = 0$, otherwise $a_k = 1$. The function $R(s, a) = (\beta \cdot \xi(s, a) + P(s, a) \cdot \alpha)$ defines the reward. Hence, a and β represent the constants objective of EC and device QoS satisfaction. The reward model will independently learn and adapt to the scale of its value to fit the updated scenario according to the score-based merging mechanism. We denote the constant α as:

$$\alpha = \sigma\left(\frac{1}{E} - \xi(\cdot)\right), \tag{18}$$

where $\sigma(\cdot)$ is illustrated as:

$$\sigma(x) = \frac{1}{e^{-x}}. \tag{19}$$

After completing the local training process, each agent sends its trained model to the global server. Finally, all device agents receive globally produced models at the same time. The agents synchronously resume the learning process by using the given global model.

6. Performance Evaluation

This part evaluates our proposed framework with other algorithms. Simulation and experimental findings reveal that our proposed Fed-DRL significantly improves EC and meets D-QoS satisfaction in a distributed network environment. For comparison, we compare our framework against Q-Learning, deep DQN, and dueling DQN, and we adopted performance metrics in terms of energy, D-QoS, and convergence analysis with different mobility scenarios to test the robustness of our proposed framework. The BSs have been arbitrarily distributed within the range of coverage of the distributed network. We presume that our distributed network system initially handles a maximum of 10 BSs. The device bandwidth is set at 20 MHz, and the device sensitivity level for edge devices is set at -120 dBm. The population of participating devices varies according to dynamic mobility to reflect the profile of the traffic model.

Furthermore, we carried out all simulations in a Python 3.8 environment. We experiment on an Ubuntu 20.04 operating system with 16GB of RAM and a RTX 2060 12-core GPU.

Fed-DRL is implemented with both the Tensor Flow and Keras Python library. Table 2 shows other simulation parameters of our experiment.

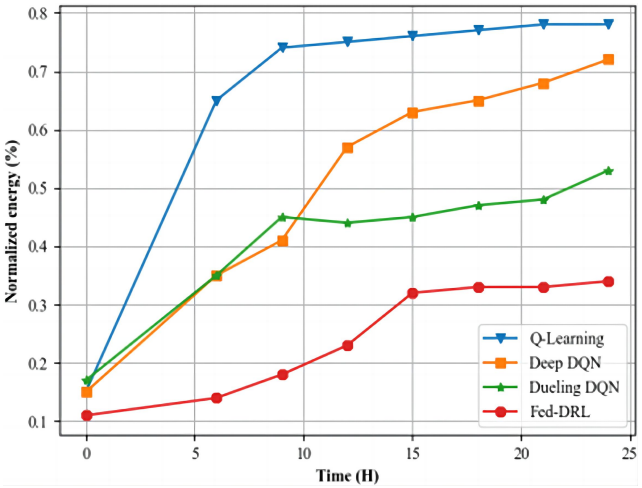
6.1. EC and device QoS satisfaction

In our simulation, we consider 1 h as a decision cycle and observe the performance within 24 h as an episode, as shown in Figure 1.

Table 2
Simulation parameters

Parameters	Values
Number of BSs, k	10−18
The radius of the coverage area for BS	150 m
System bandwidth	20 MHz
Maximum transmit power per BS	1 W
Mean package size	4000 bit
Noise power spectrum density	−174 dBm/Hz
Carrier frequency band	2.4 GHz
Active mode of energy consumption	Equation (8)
Sleep mode of energy consumption	4.3 W
Shadowing effects	(0–8) dB
Packet arriving rate	160 (packet/s)
Discount factor	0.5
Device demand	0.5 Mbps
Steepness coefficient in device satisfaction	1e–5
Bach size	64
Size of replay memory	6000
Learning rate	0.01 s
Decision epoch time, t	10, 20, 40, and 60 min

Figure 1
Stabilized energy consumption

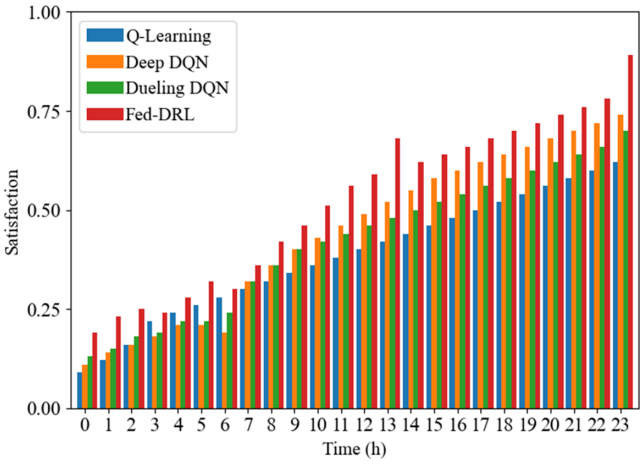


We see that increased traffic load is linked to a high amount of stabilized energy use by BSs. This is because more BSs are needed to meet high device demands.

To control the total amount of energy BSs use, our proposed Fed-DRL algorithm performs better than the benchmarked algorithms in light-load and heavy-load situations within an episode. Based on the results of the experiments, the training performance that was set helps to use the least amount of energy.

Relatively greater EC is seen in Q-learning, which also suffers from the curse of dimensionality as the number of participating devices and BSs increases. We measure our D-QoS satisfaction in Figure 2 to determine the superiority of the switching procedure

Figure 2
Algorithm comparison on device QoS satisfaction



in the distributive network. The assessment value for an appropriate distribution of resources to devices is QoS satisfaction. Comparing D-QoS satisfaction to the benchmark algorithms, we observe that Fed-DRL recorded the highest device level of satisfaction.

We can clearly state that using Q-learning increases OPEX in terms of energy utilization. The participating device's level of QoS satisfaction is reduced by 33% and 25%, respectively, as the episode increases. For participating devices, the deep DQN and dueling DQN maintain an average QoS satisfaction level of about 52%. The proposed Fed-DRL recorded an appraisal device QoS satisfaction of about 67%.

6.2. Effects of decision epoch

We illustrate the EC variance and device satisfaction variance in Figures 3 and 4 under diverse mechanism schemes.

The mobility behavior of devices under diverse epochs permits us to describe statistically how individual BSs react to their respective mobility trails in the distributive network.

Figure 3
Effects of decision epochs for Fed-DRL in BS energy consumption

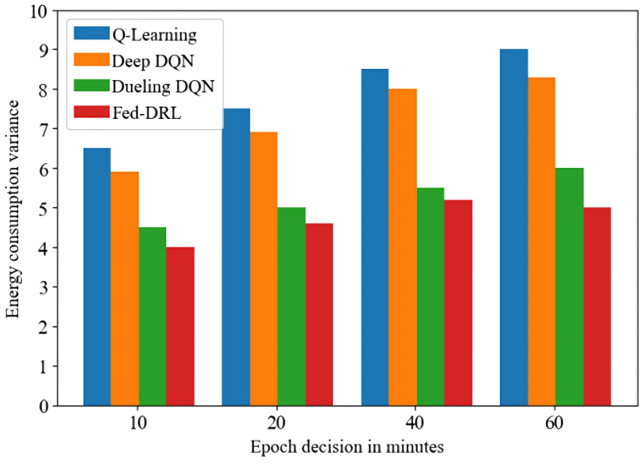


Figure 4
Effects of decision epochs for Fed-DRL in device satisfaction

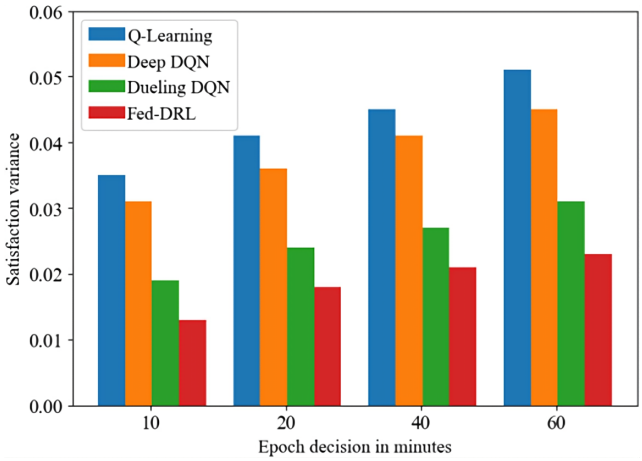
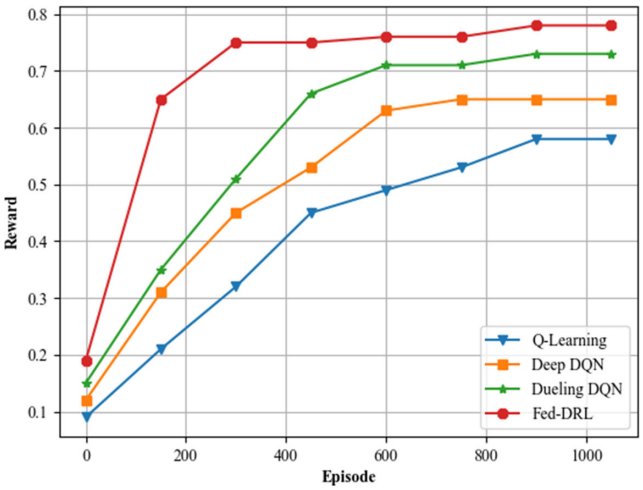


Figure 5
Convergence analysis



In this case, we look at our differences to see what happens at decision epochs 10, 20, 40, and 60. We experienced that with a long interval, the decision epoch upturned the variance between participants' satisfaction and BS's EC. This indication simply points out that the algorithm rises with the decision epoch and takes a long time to converge. However, as illustrated, our proposed Fed-DRL outperforms the benchmark algorithms for D-QoS satisfaction and BS EC. With BS's EC, Figure 3 indicates that Fed-DRL recorded the least variance, as there was an increase in the decision epoch. Compared to the dueling DQN, the EC variance is reduced when the decision epoch is set from 10 to 60.

Similarly, in deep DQN recorded an approximate reduction when observed between 10 and 60 min. In the proposed framework, we realize that setting the decision epoch to 10 min yields the paramount EC of BS. Deep DQN and dueling DQN recorded extraordinary variance owing to their poor convergence output.

Figure 4 shows a comparison of the changes in D-QoS satisfaction in the same scenario. Our research shows that the device requirement for D-QoS satisfaction is met as decision-making periods increase. We also found that the system is very unstable when the decision epoch is high because of the time between decision epoch 40 and decision epoch 60. Our proposed Fed-DRL has a significance value of about 0.008, which is lower than Q-learning, deep DQN, and dueling DQN, which have values of 0.013, 0.033, and 0.028, respectively. The proposed framework has a fairly low average value, which shows that it is strong and can converge for different decision epochs.

6.3. Convergence analysis

In this part, we only concentrated on the convergence analysis of the proposed Fed-DRL framework in terms of cumulative reward, BS EC, and D-QoS satisfaction. This is because the results in these preceding works indicate a slow convergence rate in Q-learning, deep DQN, and dueling DQN algorithms. For our convergence performance, simulation was conducted under countless distributed traffic loads. As shown in Figure 5, the Fed-DRL framework converges speedily at about Episode 260 in terms of cumulative reward. The framework displays consistently lower average EC levels than other results for Q-leaning, deep DQN, and dueling DQN, respectively.

Similarly, the framework recorded a 0.5 satisfaction threshold, while most participating devices in deep DQN and dueling DQN do not converge to the desired satisfaction. The Fed-DRL framework achieves the required satisfaction with lower EC, as it can attain stable convergence. Since our framework trains distributively, it is observed that EC and device satisfaction increase significantly before convergence while minimizing their variance until they attain convergence.

With the convergence performance, we concluded that the proposed framework is robust in a distributed scenario. Our proposed framework seems to outperform deep DQN and dueling DQN because it converges faster at about episode 170. This indicates significant regularity in terms of network scalability. Also, we show that our proposed framework can generate switching configurations that balance all devices and keep track of how much energy they use in a distributed network.

While FL offers several advantages, it has limitations and potential drawbacks, particularly in the above scenario. Communication overhead is one of the drawbacks of this approach. Communication is typically slower and less reliable in a wireless network than in wired networks. Transmitting model updates and gradients over wireless connections can introduce delays and increase communication overhead, impacting the training process. The limited bandwidth and potential packet loss in wireless networks can hinder the efficiency of our proposed approach. In addition, with privacy concerns, keeping the data decentralized and performing model updates locally. However, wireless networks, especially public or unsecured networks, pose additional privacy and security risks. Malicious actors could attempt to intercept or manipulate the communication during the FL process, potentially compromising the integrity and privacy of the data or model.

7. Conclusion

In this work, we propose a switching framework for EM using Fed-DRL. Specifically, we employ FL to help train the DRL agents efficiently because of its ability to improve learning performance, solve EM problems, and finally outperform the traditional DQN in training.

Finally, the framework solves the EM problem by balancing the levels of EC of the BS and achieving device satisfaction with a

minimum number of active BS deployed. Results from the experiment reveal that our proposed Fed-DRL framework outperformed other benchmark algorithms.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

Data available on request from the corresponding author upon reasonable request.

References

Abdulmula, A., Sopian, K., & Haw, L. C. (2019). Power consumption modeling based on real-time data traffic for balancing power supply and energy demand to develop green telecommunication tower: A case study. *Engineering, Technology & Applied Science Research*, *9*(3), 4159–4164. https://doi.org/10.48084/etasr.2742.

Ahmed, F., Naeem, M., & Iqbal, M. (2017). ICT and renewable energy: A way forward to the next generation telecom base stations. *Telecommunication Systems*, *64*, 43–56. https://doi.org/10.1007/s11235-016-0156-4.

Alam, A. S., & Dooley, L. S. (2015). A scalable multimode base station switching model for green cellular networks. In *2015 IEEE Wireless Communications and Networking Conference*, 878–883. https://doi.org/10.1109/WCNC.2015.7127585.

Ashraf, I., Boccardi, F., & Ho, L. (2010). Power savings in small cell deployments via sleep mode techniques. In *IEEE 21st International Symposium on Personal, Indoor and Mobile Radio Communications Workshops*, 307–311. https://doi.org/10.1109/PIMRCW.2010.5670384.

Bala, I., & Ahuja, K. (2023). Energy-efficient framework for throughput enhancement of cognitive radio network by exploiting transmission mode diversity. *Journal of Ambient Intelligence and Humanized Computing*, *14*(3), 2167–2184. https://doi.org/10.1007/s12652-021-03428-x.

Büttner, L., Posch, H., Auer, T. A., Jonczyk, M., Fehrenbach, U., Hamm, B., …, & Böning, G. (2021). Switching off for future — Cost estimate and a simple approach to improving the ecological footprint of radiological departments. *European Journal of Radiology Open*, *8*, 100320. https://doi.org/10.1016/j.ejro.2020.100320.

Buzzi, S., Chih-Lin, I., Klein, T. E., Poor, H. V., Yang, C., & Zappone, A. (2016). A survey of energy-efficient techniques for 5G networks and challenges ahead. *IEEE Journal on Selected Areas in Communications*, *34*(4), 697–709. https://doi.org/10.1109/JSAC.2016.2550338.

Chen, Z., Hu, J., Min, G., Luo, C., & El-Ghazawi, T. (2022). Adaptive and efficient resource allocation in cloud datacenters using actor-critic deep reinforcement learning. *IEEE Transactions on Parallel and Distributed Systems*, *33*(8), 1911–1923. https://doi.org/10.1109/TPDS.2021.3132422.

Cheng, X., Fang, L., Yang, L., & Cui, S. (2017). Mobile big data: The fuel for data-driven wireless. *IEEE Internet of Things Journal*, *4*(5), 1489–1516. https://doi.org/10.1109/JIOT.2017.2714189.

Delaram, J., Houshamand, M., Ashtiani, F., & Valilai, O. F. (2021). A utility-based matching mechanism for stable and optimal resource allocation in cloud manufacturing platforms using deferred acceptance algorithm. *Journal of Manufacturing Systems*, *60*, 569–584. https://doi.org/10.1016/j.jmsy.2021.07.012.

Githiru, M., Lens, L., Adriaensen, F., Mwang'ombe, J., & Matthysen, E. (2011). Using science to guide conservation: From landscape modelling to increased connectivity in the Taita Hills, SE Kenya. *Journal for Nature Conservation*, *19*(5), 263–268. https://doi.org/10.1016/j.jnc.2011.03.002.

Habibi, M. A., Nasimi, M., Han, B., & Schotten, H. D. (2019). A comprehensive survey of RAN architectures toward 5G mobile communication system. *IEEE Access*, *7*, 70371–70421. https://doi.org/10.1109/ACCESS.2019.2919657.

Han, Y., Li, D., Qi, H., Ren, J., & Wang, X. (2019). Federated learning-based computation offloading optimization in edge computing-supported internet of things. In *Proceedings of the ACM Turing Celebration Conference*, 25. https://doi.org/10.1145/3321408.3321586.

Hoffmann, M., Kryszkiewicz, P., & Kliks, A. (2021). Increasing energy efficiency of massive-MIMO network via base stations switching using reinforcement learning and radio environment maps. *Computer Communications*, *169*, 232–242. https://doi.org/10.1016/j.comcom.2021.01.012.

Hsieh, C. K., Chan, K. L., & Chien, F. T. (2021). Energy-efficient power allocation and user association in heterogeneous networks with deep reinforcement learning. *Applied Sciences*, *11*(9), 4135. https://doi.org/10.3390/app11094135.

Jiang, J. C., Kantarci, B., Oktug, S., & Soyata, T. (2020). Federated learning in smart city sensing: Challenges and opportunities. *Sensors*, *20*(21), 6230. https://doi.org/10.3390/s20216230.

Jiao, Y., Wang, P., Niyato, D., Lin, B., & Kim, D. I. (2021). Toward an automated auction framework for wireless federated learning services market. *IEEE Transactions on Mobile Computing*, *20*(10), 3034–3048. https://doi.org/10.1109/TMC.2020.2994639.

Kim, E., Jung, B. C., Park, C. Y., & Lee, H. (2022). Joint optimization of energy efficiency and user outage using multi-agent reinforcement learning in ultra-dense small cell networks. *Electronics*, *11*(4), 599. https://doi.org/10.3390/electronics11040599.

Lee, H., Song, C., Kim, N., & Cha, S. W. (2020). Comparative analysis of energy management strategies for HEV: Dynamic programming and reinforcement learning. *IEEE Access*, *8*, 67112–67123. https://doi.org/10.1109/ACCESS.2020.2986373.

Li, H., Ota, K., & Dong, M. (2018). Learning IoT in edge: Deep learning for the internet of things with edge computing. *IEEE Network*, *32*(1), 96–101. https://doi.org/10.1109/MNET.2018.1700202.

Li, R., Zhao, Z., Chen, X., Palicot, J., & Zhang, H. (2014). TACT: A transfer actor-critic learning framework for energy saving in cellular radio access networks. *IEEE Transactions on Wireless Communications*, *13*(4), 2000–2011. https://doi.org/10.1109/TWC.2014.022014.130840.

Liu, J., Krishnamachari, B., Zhou, S., & Niu, Z. (2018). DeepNap: Data-driven base station sleeping operations through deep

reinforcement learning. *IEEE Internet of Things Journal*, *5*(6), 4273–4282. https://doi.org/10.1109/JIOT.2018.2846694.

Mesodiakaki, A., Adelantado, F., Alonso, L., & Verikoukis, C. (2014). Energy-efficient context-aware user association for outdoor small cell heterogeneous networks. In *2014 IEEE International Conference on Communications*, 1614–1619. https://doi.org/10.1109/ICC.2014.6883553.

Nishio, T., & Yonetani, R. (2019). Client selection for federated learning with heterogeneous resources in mobile edge. In *2019 IEEE International Conference On Communications*, 1–7. https://doi.org/10.1109/ICC.2019.8761315

Pham, Q. V., Le, M., Huynh-The, T., Han, Z., & Hwang, W. J. (2022). Energy-efficient federated learning over UAV-enabled wireless powered communications. *IEEE Transactions on Vehicular Technology*, *71*(5), 4977–4990. https://doi.org/10.1109/TVT.2022.3150004.

Sharma, S., Darak, S. J., & Srivastava, A. (2017). Energy saving in heterogeneous cellular network via transfer reinforcement learning based policy. In *2017 9th International Conference on Communication Systems and Networks*, 397–398. https://doi.org/10.1109/COMSNETS.2017.7945411.

Sheng, S., Chen, P., Chen, Z., Wu, L., & Yao, Y. (2021). Deep reinforcement learning-based task scheduling in IoT edge computing. *Sensors*, *21*(5), 1666. https://doi.org/10.3390/s21051666.

Shingu, Y., Seki, Y., Watabe, S., Endo, S., Matsuzaki, Y., Kawabata, S., . . . , & Hakoshima, H. (2021). Boltzmann machine learning with a variational quantum algorithm. *Physical Review A*, *104*(3), 032413. https://doi.org/10.1103/PhysRevA.104.032413.

Sun, H., Fu, Z., Tao, F., Zhu, L., & Si, P. (2020). Data-driven reinforcement-learning-based hierarchical energy management strategy for fuel cell/battery/ultracapacitor hybrid electric vehicles. *Journal of Power Sources*, *455*, 227964. https://doi.org/10.1016/j.jpowsour.2020.227964.

Sun, X., Qiu, J., Tao, Y., Ma, Y., & Zhao, J. (2022). A multi-mode data-driven volt/var control strategy with conservation voltage reduction in active distribution networks. *IEEE Transactions on Sustainable Energy*, *13*(2), 1073–1085. https://doi.org/10.1109/TSTE.2022.3149267.

Tabassum, H., Siddique, U., Hossain, E., & Hossain, M. J. (2014). Downlink performance of cellular systems with base station sleeping, user association, and scheduling. *IEEE Transactions on Wireless Communications*, *13*(10), 5752–5767. https://doi.org/10.1109/TWC.2014.2336249.

Tian, X. J., & Jiang, Q. L. (2021) Power allocation scheme for maximizing energy efficiency in downlink NOMA Systems. *Journal of Beijing University of Posts and Telecommunications*, *44*(1), 38–44. https://doi.org/10.13190/j.jbupt.2020-031.

Wang, Q., & Zheng, J. (2015). A distributed base station on/off control mechanism for energy efficiency of small cell networks. In *2015 IEEE International Conference on Communications*, 3317–3322. https://doi.org/10.1109/ICC.2015.7248836.

Xiong, X., Zheng, K., Lei, L., & Hou, L. (2020). Resource allocation based on deep reinforcement learning in IoT edge computing. *IEEE Journal on Selected Areas in Communications*, *38*(6), 1133–1146. https://doi.org/10.1109/JSAC.2020.2986615.

Ye, Y., Li, S., Liu, F., Tang, Y., & Hu, W. (2020). EdgeFed: Optimized federated learning based on edge computing. *IEEE Access*, *8*, 209191–209198. https://doi.org/10.1109/ACCESS.2020.3038287.

Zhang, T., & Mao, S. (2022). Energy-efficient federated learning with intelligent reflecting surface. *IEEE Transactions on Green Communications and Networking*, *6*(2), 845–858. https://doi.org/10.1109/TGCN.2021.3126795.

Zhuang, B., Guo, D., & Honig, M. L. (2016). Energy-efficient cell activation, user association, and spectrum allocation in heterogeneous networks. *IEEE Journal on Selected Areas in Communications*, *34*(4), 823–831. https://doi.org/10.1109/JSAC.2016.2544478.

How to Cite: Agbesi, V. K., Kuadey, N. A., Agbesi, C. C. M., & Maale, G. T. (2024). Federated-Based Deep Reinforcement Learning (Fed-DRL) for Energy Management in a Distributive Wireless Network. *Journal of Data Science and Intelligent Systems*, *2*(2), 113–121. https://doi.org/10.47852/bonviewJDSIS3202998