https://doi.org/10.1590/jatm.v17.1388
ORIGINAL PAPER

Deep Q Network Based Power Allocation for Uplink 5G Heterogeneous Networks

Madhusudhanan Sampath[1], Amalorpava Mary Rajee Samuel[2,*], Yamuna Devi Manickam Malu[3], Sujatha Chinnathevar[4], Sridevi Cheguri[5]

1. Rajalakshmi Engineering College – Department of Computer Science Engineering – Chennai – India.
2. University of Witwatersrand – Department of Electrical and Information Engineering – Johannesburg – South Africa.
3. Koneru Lakshmaiah Education Foundation – Department of Computer Science Engineering – Vijayawada – India.
4. SSM Institute of Engineering and Technology – Department of Computer Science Engineering – Dindigul – India.
5. St. Martin's Engineering College – Department of Electronics and Communication Engineering – Secundrabad – India.

*Correspondence author: maryrajee@gmail.com

ABSTRACT

The next generation heterogeneous network (HetNet) consists of multiple technologies for the device to improve their quality of service (QoS) parameters for the ubiquitous connectivity. The key technology has the ability to use machine intelligence in the design of an energy-efficient HetNet. That allows internet of things (IoT) devices to choose which base station (BS) to connect with for optimal performance; the proposed QoS-aware deep Q network (Q-DQN) algorithm adapts an energy-efficient reward function to improve the performance of femto BS IoT devices without deviating from macro BS. The main objective is to ascertain the QoS requirement that should exceed the threshold level. The performance of the proposed work is validated through the QoS, system capacity, and energy efficiency. A dynamic power selection strategy in a HetNet is based on the Q-DQN algorithm subject to network QoS parameters. The Q-DQN power allocation using reinforcement learning in uplink HetNet offers a powerful approach to managing the complex trade-offs between QoS requirements and energy efficiency. By dynamically adjusting power levels based on real-time conditions, the comparative results are evident that the proposed algorithm provides improved capacity and energy efficiency in proportion to the escalating throughput enhancement.

Keywords: DQN; HetNet; Uplink heterogeneous network.

INTRODUCTION

The rapid growth of wireless technologies enables the user devices to access the internet at a speed of faster data transfer of about 100 Mbps than previous generations. The key enablers of next-generation networks are connectivity, data rate, capacity, low latency, and higher energy efficiency. To handle the mobile data traffic and continue providing high quality of service (QoS) to the user devices, the heterogeneous network (HetNet) is an attractive solution to offload the internet of things (IoT) devices' traffic. In which co-existence and interworking of multiple radio access technologies and multiple base stations (BSs) with different powers are integrated. In the concept of the 5G network, small cells are the offloading solution in a self-organized network equipped with reinforcement learning (RL). The study of machine learning technique involves the decision-making ability using Markov decision process. In RL-based HetNet, each IoT device equipment is the decision-making agent in each time step, interacting between the agents with its external environment by executing an action. The 5G network is constructed with macrocells surrounded by the small cells to distribute the overload and maintain the connectivity among the IoT devices.

Received: Dec. 30, 2024 | **Accepted:** May 03, 2025
Peer Review History: Single Blind Peer Review.
Section editor: Paulo Renato Silva

LITERATURE REVIEW

Literature review based on energy efficiency aspects

The importance of green communication is even more significant with respect to 5G wireless systems. Energy efficiency is the QoS parameter to define the ability of successful transmission (bits per joule of energy), where the energy computation includes core and radio access, legacy cellular technologies, and cloud data. To drive focus on energy efficiency, it is important to know how an intelligent network is exploited with a higher degree of automation and cooperation. It is rigid to manage a centralized network management approach in dense networks. Power allocation is a main problem in an interference-limited network. This intelligent network environment is dynamic in nature. Power allocation and resource allocation are the major constraints that are dealt with by the application of the RL technique (Wei *et al.* 2022) suggested a Q-learning method of Q-learning Based Distributed Power Allocation (Q-DPA), which is useful for dynamic power allocation, to enable power adaptation using Markov Decision Process (MDP). The optimal power allocation is derived under power constraints for sum value and individual value. The optimal solution is dependent on signal-to-interference-plus-noise ratio (SINR) values. Geometric programming (GP) problems are used to find optimal solutions in the high SINR regime. Multiple GPs are used for solving the problem of power allocation in the low SINR regime (Sami *et al.* 2021). When making decisions in sequential order, offered a traditional formalization in which actions have consequences that extend beyond the immediate rewards. As a result, mobile IoT devices might use future rewards to predict decision-making under unknowable network variables like channel availability and resource allocation.

Zhou *et al.* (2021) proposed cell selection using a Markov decision process framework. The performance of the cell selection strategy depends on several factors such as the quality of the device's link and dynamic channel load. The authors solved cell the allocation problem using value iteration algorithm (VIA) with MDP. The allocation information is shared among the femtocells and the decision is taken via MDP. This technique makes judicious handoff choices in the midst of challenging radio propagation. This will reduce unnecessary handoff and lower the signal overhead, thereby providing good QoS to the IoT device equipment. Dynamic channel load and link quality are the main factors involved in MDP Cooperative Q-learning is effectively involved in the design of transmit power to achieve a better QoS. The IoT devices' data rate and its threshold are maintained to ascertain the connectivity.

To help the power allocation method for industrial IoT (IIoT) HetNet, Kunpeng *et al.* (2021) and Yuan *et al.* (2022) presented a distributed Q-learning approach.]. To guarantee the seamless information coverage, reward functions are proposed to meet the QoS of the IoT device equipment. The learning rates and their cooperation are investigated to improve fairness. Fair power allocation is achieved through the proximity of various BSs to their associated IoT devices. Wang *et al.* (2022) proposed self-organizing network (SON) related objectives such as coverage optimization, interference control, and energy-saving management self- organizing network architecture and their energy management are provided for network elements. Yejun *et al.* (2023) performed a Q-learning optimization using a singular value decomposition, which makes it challenging in real-time implementation. Key technologies discussed in the previous research are used to reduce capital expenditure (CAPEX) and operator expenses (OPEX), operational cost, and total cost of ownership (TCO). The various techniques to facilitate network connectivity in remote areas are dependent on multiple layers of coordination and management among the network components. It becomes to be effective if it is managed by machine intelligence algorithms (Qian *et al.* 2023). The next-generation 5G network depends on a variety of strategies, but new studies have found that resource allocation and power control play the biggest roles in creating a greener network (Weidang *et al.* 2022).

The design of an energy-efficient reward function is the novelty in attaining the high-energy efficiency for small cell devices without compromise. The same energy efficiency is attained for the devices when connecting to macrocell or small cell (Rajee and Merline 2020). The main contributions are listed as follows. This work proposes a QoS-aware algorithm adapting a cooperative Q-learning method. A novel energy-efficient reward function with an efficient energy threshold constraint for the femto user equipment (FUE) devices based on IoT devices. A dynamic power allocation strategy is done through MDP and the decision is based upon the energy-efficient reward function. Hence, all the UE devices are equally treated without any QoS deviation (Nagarajan *et al.* 2023).

Quality of service needs

6G networks have diverse QoS requirements for different devices and applications, such as latency, reliability, throughput, and energy efficiency (Rajee *et al.* 2023). For instance, crucial tasks such as automated cars or long-distance surgery require extremely quick response times and top-notch dependability, whereas some tasks may focus more on conserving energy.

Challenge of power allocation

The power allocation in a manner that satisfies the various qualities of service needs of IoT devices is difficult for maintaining the small power. Efficient distribution of power is essential to ensure that important devices receive the resources they need without energy being wasted (Sampath *et al.* 2024). The novelty of the proposed system deals with the DQN approach by integrating machine intelligence, QoS adaptation, and energy-efficient RL. The feature analysis is given in Table 1.

The system model section proposes the conceptualization of the main study, consequently, the power allocation strategy is explained through algorithm, and the evaluation results are shown with explanation.

Table 1. Comparative feature analysis.

Feature	Existing approach	Proposed DQN approach
Power allocation	Static or heuristic based programming	Dynamic real-time adaptation
Energy efficiency	Limited constraints	Energy-efficient reward function
UE connectivity	Random	Intelligent association
Learning methodology	Traditional RL	DQN

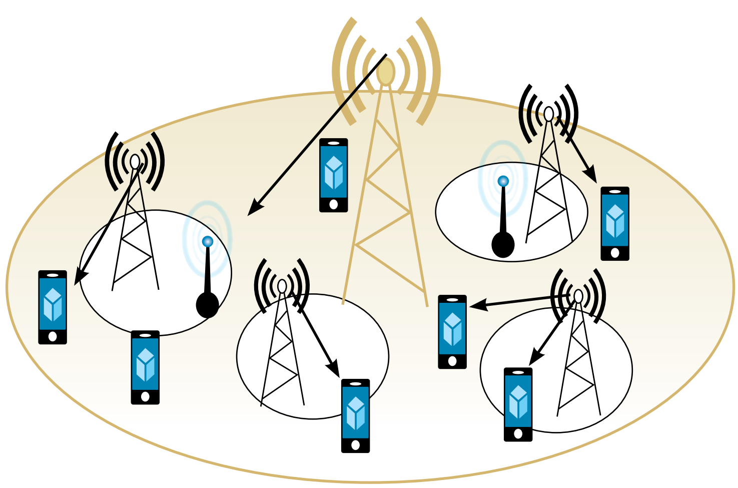
Source: Elaborated by the authors.

SYSTEM MODEL

Next-generation HetNet

A HetNet comprises of different cells. The high transmitted power is contributed by the macrocell, which will cover macro UE (MUE) devices in the primary tier, while FUEs, which are serviced by the femto cells, are available in the surrounding layer.

Figure 1 depicts a two-tier HetNet as an example. The proposed two-tier HetNet is deployed in the Euclidean plane. The primary part of the HetNet is the macrocell and k femtocells are surrounded in the outer layers. There is an IoT device linked to each of the cells. This HetNet simplified model affects the learning algorithm's precision. Considered is a dense HetNet in



Source: Rajee SAM, 2023.

Figure 1. An illustration of HetNet.

which interference between and within cells is a result of the densification (Samuel and Arulraj 2023). Therefore, our proposed study focuses on optimization problem in the uplink and downlink of the femtocell network so as to limit the interference and offer the necessary QoS parameter for the signal quality. The macrocell BS (MBS) and their counterpart UE devices pair is designated as index 0, and femto BS (FBS) and their counterpart UE devices pair couple from the set K are denoted by the index k. The interference in the received side which operates on a subband, is present (Yamuna *et al.* 2022). Consequently, the following formula is used to calculate the SINR over subband s. The index 0,0 denotes the MBS and MUE collaboration, whereas as k,0 represents interference from FBS with MUE:

$$SINR_{MUE} = \gamma_m = \frac{P_0 \left|h_{0,0}\right|^2}{\sum_{k \in K} P_k \left|h_{k,0}\right|^2 + N_0} \tag{1}$$

where P_0 refers to macrocell transmitted power and $h_{0,0}$ is the channel gain between the MBS and MUE. Similarly, the instantaneous SINR of FUE is computed as follows:

$$SINR_{FUE} = \gamma_\mathrm{k} = \frac{P_k \left|h_{k,k}\right|^2}{P_0 \left|h_{0,k}\right|^2 + \sum_{j,k \in K} P_j \left|h_{j,k}\right|^2 + N_k} \tag{2}$$

where $h_{k,k}$ is the channel gain between the desired link of FBS and FUE, P_j is the transmitted power of FBS, $h_{0,k}$ and $h_{j,k}$ are the interference channel gain, and N_k denotes the additive white Gaussian noise (AWGN). The mathematical formulation for calculating the capacity (Namaskaram *et al.* 2024; Yuan *et al.* 2022) of MUE and FUE are given as follows:

$$C_{MUE} = log_2(1 + SINR_{MUE}) \tag{3}$$

$$C_{FUE_k} = log_2\left(1 + SINR_{FUE_k}\right), k \in K \tag{4}$$

The total system capacity is calculated as:

$$C_{system} = BW\left(C_{MUE} + C_{FUE_k}\right), k \in K \tag{5}$$

$$EE_{eff} = \frac{Maximum\ Achievable\ Capacity}{Average\ Network\ Power\ Consumption} \tag{6}$$

$$EE_{eff} = \frac{C_{system}}{P_{MBS} + P_k}, k \in K$$

The total capacity is formulated using the bandwidth available for MBS and FBS, energy efficiency is further calculated using Eq. 6. The main objective is to design the energy-efficient reward function of the proposed learning environment.

POWER ALLOCATION ANALYSIS

Reinforcement learning

The Q-learning issue involves an environment and multiple agents. It is utilized for engaging with the ever-changing environment through trial-and-error experiences. In 6G IoT networks on a big scale, various agents (such as IoT devices, BSs) need to synchronize their activities. Multi-agent RL enables these agents to gain knowledge and adjust their strategies within a common setting, whether through collaboration or competition. In collaborative multi-agent RL, agents collaborate to maximize a common reward, beneficial in situations where IoT devices must work together to enhance performance across the entire network (Shasha 2023; Yaser *et al.* 2024). In competitive multi-agent RL, agents engage in rivalry for scarce resources (e.g., bandwidth, power), beneficial in settings with conflicting goals. In decentralized multi-agent RL, every agent independently learns its policy, enabling scalability in large networks where centralized control is not feasible.

The RL framework tuple is defined as (A: X: P_r: R) with the following definitions. The tuple is a set of actions, states, state transition probability and reward function. The actions of the agents are specified as:

$$A = \{a_1, a_2 \ldots\ldots a_n\} \tag{7}$$

X is the state set of all the agents. An agent is represented in its state when agents move from state X to the new state X* after taking action as:

$$X = \{X_1, X_2, \ldots\ldots X_n\} \tag{8}$$

The state transition probability is the probabilistic transition from current state to next state after observing their actions and related rewards. r_i^{MUE} *and* r_i^{FUE} are the reward function when performing action a in state x:

$$r_i^{MUE} = \begin{Bmatrix} 1 \; if \; \gamma_m \geq \gamma_T \\ -1 \; otherwise \end{Bmatrix} \tag{9}$$

$$r_i^{FUE} = \begin{Bmatrix} 0 \text{ if } \gamma_k/P \leq t_a \\ 1 \text{ if } t_a < \gamma_k/P \leq t_b \\ 2 \text{ if } \gamma_k/P \geq t_b \end{Bmatrix} \tag{10}$$

- Agents: the agents are the MUE and FUE. The QoS improvement is guaranteed with the SINR target.
- Actions: the learning agents are initially allotted with predefined transmission power levels starting from 2 dB to 18 dB.
- States: the number of states in this DQN is 12 states for FUE and six states for MUE mutually migrated from one state to another state after observing the reward. The energy efficiency ratio between the SINR and its threshold is given in the Eq. 11. These three indicators are defined in the following equation:

$$I_{\gamma_m} = \begin{Bmatrix} 1 \text{ if } \gamma_m \geq \gamma_T \\ 0 \text{ otherwise} \end{Bmatrix} \quad I_{\gamma_k} = \begin{Bmatrix} 1 \text{ if } \gamma_k \geq \gamma_T \\ 0 \text{ otherwise} \end{Bmatrix}$$

$$I_{\gamma} = \begin{Bmatrix} 0 \text{ if } \gamma_k/P \leq t_a \\ 1 \text{ if } t_a < \gamma_k/P \leq t_b \\ 2 \text{ if } \gamma_k/P \geq t_b \end{Bmatrix} \tag{11}$$

An optimal Q-value is defined as:

$$Q(x,a) = E\{R(x,a)\} + \beta \sum_{t=0}^{\infty} P_r(x^*|x,a)Q^*(x,a) \tag{12}$$

The Q-value Q(x,a) is updated after each iteration, the discount factor and learning factor is chosen to give the optimal value of Q(x,a).

Co-operative power allocation

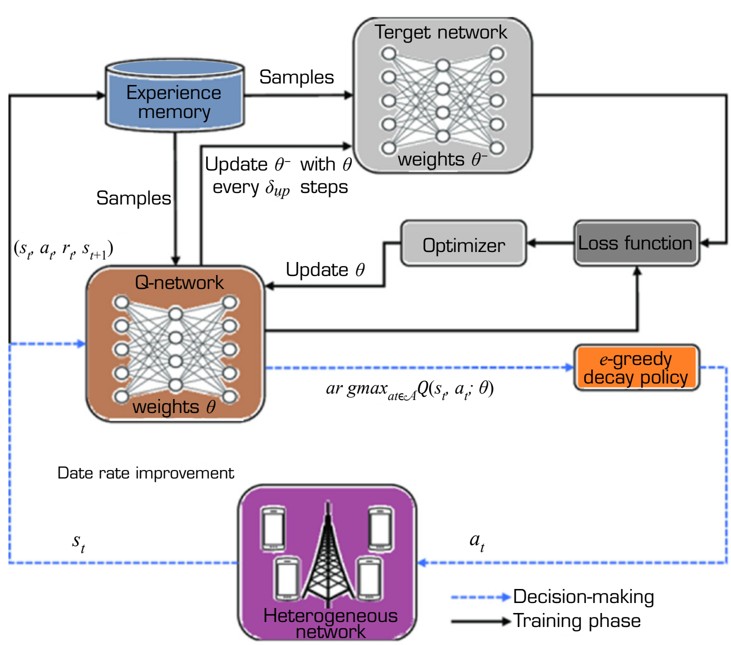
The co-operative learning is adopted in the proposed DQN framework, the objective of the co-operative learning among agent is to maximize the EE after allocating power allocation to the UEs.

Maximize EE_{eff}. Subject to:

$$\begin{aligned} &C.1:\ C_{FUE} \geq \gamma_T^* \\ &C.2:\ C_{MUE} \geq \gamma_T^* \end{aligned} \tag{13}$$

The subjective constraints are guaranteed that the MUE and FUE do not fall below the predefined thresholds defined in Eqs. 1 and 2.

Traditional BS activation control utilizing DQN presupposes centralized management and models the complete system with a singular action value function. Nonetheless, this approach is not suitable for dense HetNet with numerous BSs due to repeated actions. In the suggested approach shown in Fig. 2, every BS acts as an agent with its unique action value function, allowing the DQN-based control of BS activation to be suitable for dense small-cell networks. Various definitions of states are analyzed from



Source: Elaborated by the authors.

Figure 2. DQN framework.

the perspectives of the load necessary for obtaining the state variables and the performance throughput that can be achieved. Computer simulations indicate that the suggested method attains greater throughput than the traditional probabilistic control approach and that it can be adjusted to various system conditions.

QoS-aware DQN algorithm is as:

- Initialize: Qi (x, a)=0 for all x, a
- For: all events t=1,2,… do
- Do: optimal action π*(s,a)
- Allocate: transmit power for each cell
- Observe: reward and maximize EE
- Check if: constraints of Eq. 9 are satisfied, then
- Update: Qi (x, a) end for

REWARD

The reward function that is selected is important since it determines the system's learning objective (QoS). The MUE and FUE's respective instantaneous SINR and FUE's transmission power, as well as two specified thresholds, are used to calculate the reward function that should be used. There are basically positive and negative rewards to maintain the QoS of an IoT device.

QoS-aware deep Q network (Q-DQN) convergence analysis

The proposed QoS-aware algorithm convergence is tested through fairness performance popularly known as Jain's fairness index, and is given in Eq. 14.

$$f(x_1, x_2, ...x_n) = \frac{\left(\sum_{i=1}^{J} x_i\right)^2}{J\sum_{i=1}^{J} x_i^{\,2}} \tag{14}$$

The fairness index is used to check the convergence of the proposed DQN algorithm.

RESULTS AND DISCUSSION

One macrocell with a radius of Rm = 500 m is supposed to exist in an urban model HetNet, whereas Rf = 50 m exists for each femtocell. These femtocells support numerous MUEs and FUEs ranging in number from two to 40. Every time slot has a single FUE and MUE that is in use. Therefore, power allocation that requires dynamic interactions is what online learning is supposed to do. The operating frequency of the mmWave network is 28 GHz and the bandwidth for each IoT devices is W = 20 MHz. For line-of-sight (LOS) and non-LOS (NLOS) links, the route loss exponents are 2 and 4, respectively. In this work, capacity and energy efficiency are taken into account as QoS indicators. Table 2 includes a list of the simulation parameters.

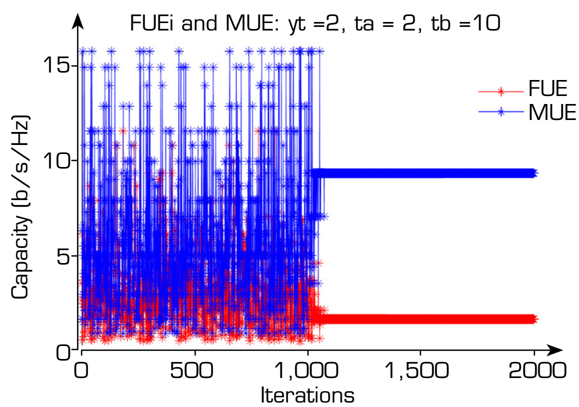
The choice of the reward function is critical as it defines the objective of the system (QoS) and its learning. The rationale behind these reward functions is to maintain high capacity and protection for the MUE and let FUEs utilize the same spectrum as much as possible without interfering with the MUE. The MUE will be rewarded +1 if its SINR is higher than the threshold, and FUE will be rewarded based on the Eq. 11 for the values 0, 1, and 2.

Table 2. Simulation parameters.

Symbol	Quantity
Bandwidth	100 MHz
Learning rate	0.9
Discount rate	0.9
FBS transmission power	17 dBm
MBS transmission power	46 dBm
Threshold	1:2:10
Noise power	-174 dBm/Hz
Carrier frequency	28 GHz

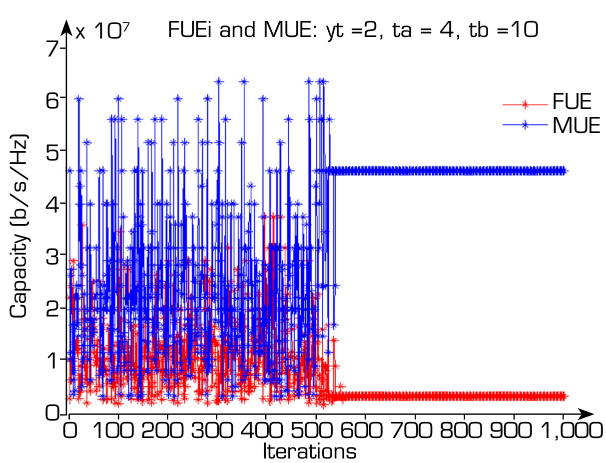
Source: Elaborated by the authors.

This assessment demonstrates how well the online learning system can distribute resources and maintain the level of service provided to UE devices. These specifications match this criterion to ensure that the MUE consistently maintains a QoS level beyond a predetermined threshold. The achieved average energy efficiency of the network is plotted in Fig. 3. The number of iterations is fixed as 2,000. The optimality of the DQN algorithm is shown in the Figs. 4 and 5 to ensure the efficacy of the results.



Source: Elaborated by the authors.

Figure 3. Energy efficiency of a network.

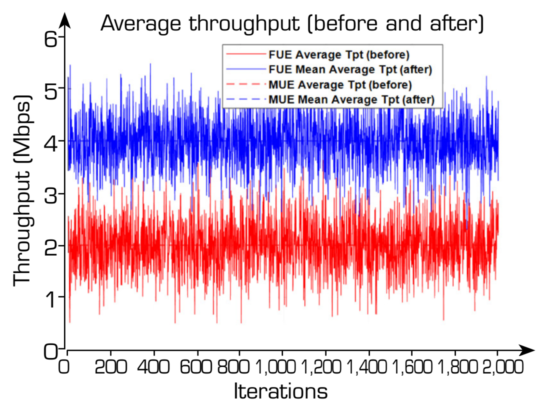


Source: Elaborated by the authors.

Figure 4. Energy efficiency of a network.

The system's performance assessment is determined by its energy efficiency, capacity, and QoS. This assessment shows how well the online learning method can efficiently distribute power and uphold the QoS for users. These parameters have been outlined in the relevant literature. The QoS is ensured by the EE of the network with FUE and MUE capacities.

The sum rate of the system energy usage is seen in Fig. 5. The requirement is a total computing data of 200 Mbps, and as the number of UEs rises, so does the system energy consumption of the four algorithm methods. This result can be attributable to the fact that SDs will be dispersed at random around the region as their quantity grows. Device energy consumption will rise as it must fly farther to cover a larger area and more UE devices.

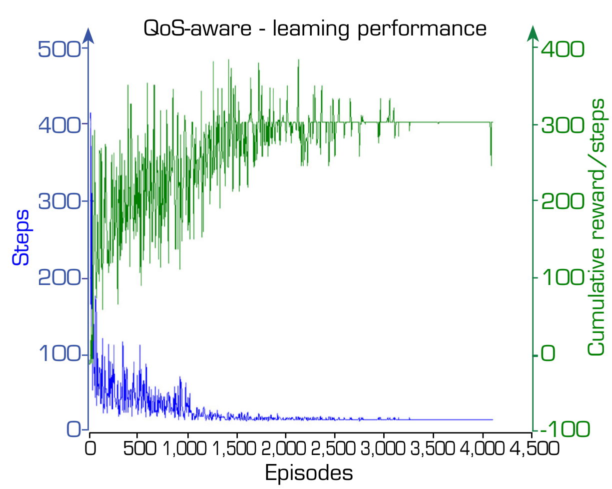


Source: Elaborated by the authors.

Figure 5. Throughput of a network.

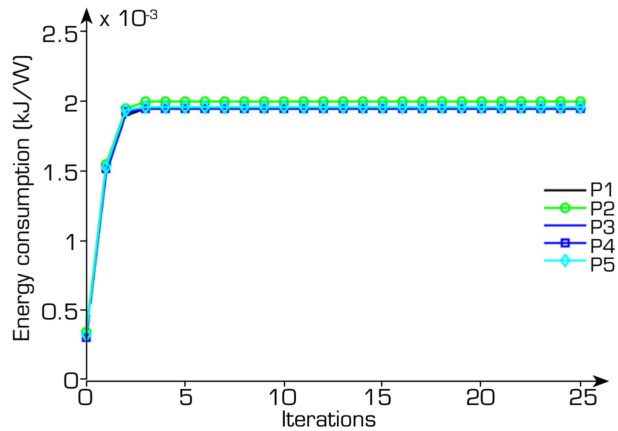
The cumulative reward is shown in Fig. 6. The energy consumption of the HetNet is shown in Fig. 7. The computational efficiency between the UE devices is shown in Fig. 8. It demonstrates that as distance tends to rise, so does the system latency. The comparative study is shown with prevailing algorithms such as DRL, actor-critic, and greedy algorithms.

QRL performed better in the simulation results when it came to the quantity of assigned tasks and overall priority. The job was allocated to devices by the suggested QRL algorithm, and the task performance demonstrated the algorithm's efficacy.



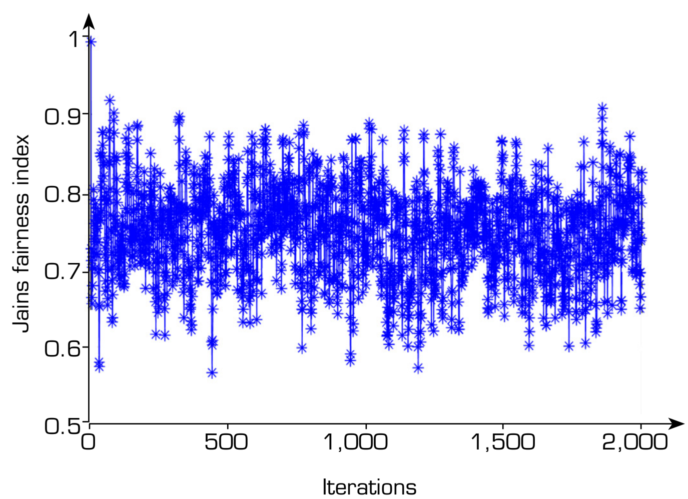
Source: Elaborated by the authors.

Figure 6. Cumulative reward vs. steps.



Source: Elaborated by the authors.

Figure 7. System power consumption.



Source: Elaborated by the authors.

Figure 8. Jain's fairness performance.

In addition to this, Table 3 is an observation of agent sum capacity. When the number of FUEs is increased, the sum capacity is also increased. The learning rate is varied in the testing phase to check the convergence. The sum capacity depends on the learning rate and their convergence. The convergence of the QoS algorithm is good up to 13 FUEs. After testing multiple inputs, it takes a long time to converge due to state space.

The comparative study report is given in Table 4 to ensure that the proposed dynamic power allocation strategy adapts to real-time network conditions.

Table 3. Sum capacity.

No. of devices	Sum capacity (Mb/s/Hz)			
Learning rate	4 Agents	6 Agents	8 Agents	12Agents
α = 0.3	0.481	0.652	0.764	0.898
α = 0.5	0.493	0.694	0.742	0.920
α = 0.7	0.498	0.696	0.796	0.990
α = 0.9	0.499	0.699	0.798	0.991

Source: Elaborated by the authors.

Table 4. Comparative analysis with existing works.

Contributors	Objective	Performance metrics (%)	Limitation
[16]	Power allocation and user association	Data rate (20-30) Power consumption (15)	Computational complexity in dense environment
[17]	Bandwidth allocation in edge computing	Spectral efficiency (25) Energy cost (10)	Limited real-time adaptability
[18]	Network slicing with power allocation	Energy efficiency (35) Throughput (20)	Communication overhead in federated learning
Proposed DQN	QoS-based resource allocation	Energy efficiency (37) Throughput (25) Power consumption (15)	Training phase is needed to assign UE's to BS allocation

Source: Elaborated by the authors.

CONCLUSION

In this proposed QoS-aware power control algorithm, energy efficiency and throughput are the main constraints. It is implemented to guarantee the IoT devices have a good QoS service. The performance is superior compared with the conventional power control technique. The UEs are tested and trained in the Q-learning environment. The training phase of the proposed work is shown in every aspect. The energy-efficient reward function is well designed in the dynamic environment, and the power selection strategy is able to produce maximum energy efficiency without any power compensation to lower energy efficiency. Results reveal that the optimal power control is possible without loss of connectivity using the QRL algorithm. In future, the same work can be carried out with the impact of blockages and penetration losses.

CONFLICT OF INTEREST

Nothing to declare.

AUTHORS' CONTRIBUTION

Conceptualization: Sampath M; **Software:** Sampath M; **Validation:** Samuel AMR; **Formal analysis:** Samuel AMR; **Investigation:** Malu YDM; **Resources:** Chinnathevar S; **Data Curation:** Cheguri S; **Writing - Original Draft:** Samuel AMR; **Writing - Review & Editing:** Samuel AMR and Chinnathevar S.; **Final approval:** Samuel AMR.

DATA AVAILABILITY STATEMENT

The data will be available upon request.

FUNDING

Not applicable.

ACKNOWLEDGMENTS

Not applicable.

REFERENCES

Kunpeng Y, Shan H, Sun T, Hu R, Wu Y, Yu L (2021) Reinforcement learning-based mobile edge computing and transmission scheduling for video surveillance. IEEE Trans Emerg Topics Comput 10(2):1142-1156. https://doi.org/10.1109/TETC.2021.3073744

Li Q, Shi L, Zhang Z, Zheng G (2023) Resource allocation in UAV-enabled wireless-powered MEC networks with hybrid passive and active communications. IEEE Internet Things J 10(3):2574-2588. https://doi.org/10.1109/JIOT.2022.3214539

Nagarajan S, Rani PS, Vinmathi MS, Subbareddy V, Saleth ALM, Subhahan DA (2023) Multi-agent reinforcement learning for resource allocation in container based cloud environment. Expert Syst 42(1):e13362. https://doi.org/10.1111/exsy.13362

Namaskaram K, Rajagopal P, Rathinam BM, Rajee M, Cheguri S, Nisha MF (2024) RIS enabled NOMA for resource allocation in beyond 5G networks. J Eng Sci Technol Rev 17(1):8-15. https://doi.org/10.25103/jestr.171.02

Rajee SAM, Merline A (2020) Machine intelligence technique for blockage effects in next-generation heterogeneous networks. Radioengineering 29(3). https://doi.org/10.13164/re.2020.0555

Rajee SAM, Merline A, Devi MMY (2023) Game theoretic model for power optimization in next-generation heterogeneous network. SIViP 17:3721-3729. https://doi.org/10.1007/s11760-023-02599-8

Sami H, Otrok H, Bentahar J, Mourad A (2021) AI-based resource provisioning of IoE services in 6G: a deep reinforcement learning approach. IEEE Trans Netw Serv Manag 18(3):3527-3540. https://doi.org/10.1109/TNSM.2021.3066625

Sampath M, Duraisamy AK, Samuel AMR, Manickam Malu YD, Nirmala M (2024) Unmanned aerial vehicle path planning using water strider algorithm. Int J Ind Eng Theory Appl Pract 31(3):429-438. https://doi.org/10 .23055/ijietap.2024.31.3.9861

Samuel AMR, Arulraj M (2023) Performance analysis of flexible indoor and outdoor user distribution in urban multi-tier heterogeneous network. Int J Mob Commun 21(1):119-133. https://doi.org/10.1504/IJMC.2023.127376

Shasha Z (2023) Energy efficient resource allocation method for 5G access network based on reinforcement learning algorithm. Sustain Energy Technol Assess 56:103020. https://doi.org/10.1016/j.seta.2023.103020

Wang D, Tian J, Zhang H, Wu D (2022) Task offloading and trajectory scheduling for UAV-enabled MEC networks: an optimal transport theory perspective. IEEE Wireless Commun Lett 11(1):150-154. https://doi.org/10.1109/LWC.2021.3122957

Wei P, Guo K, Li Y, Wang J, Feng W, Jin S (2022) Reinforcement learning-empowered mobile edge computing for 6G edge intelligence. IEEE Access 10:65156-65192. https://doi.org/10.1109/ACCESS.2022.3183647

Weidang Lu, Yu Ding, Yunqi Feng, Guoxing Huang, Nan Zhao, Arumugam N (2022) Dinkelbach-guided deep reinforcement learning for secure communication in UAV-aided MEC networks. Paper presented 2022 IEEE Global Communications Conference. IEEE; Rio de Janeiro, Brazil. https://doi.org/10.1109/GLOBECOM48099.2022.10001574

Yamuna DMM, Ahila A, Samuel AMR, Indira NR (2023) Reconfigurable intelligent surface aided indoor and outdoor user distribution in heterogeneous network. Int J Comput Netw Inf Secur 14(4):89-97. https://doi.org/10.5815/ijcnis.2022.04.07

Yaser A, Saleh Yi, Hashem K, Thomas K (2024) Mobility aware and energy-efficient federated deep reinforcement learning assisted resource allocation for 5G-RAN slicing. Comput Commun 217:166-182. https://doi.org/10.1016/j.comcom.2024.01.028

Yejun H, Youhui G, Haixia C, Mohsen G (2023) Fairness-based 3-D multi-UAV trajectory optimization in multi-UAV-assisted MEC system. IEEE Internet Things J 10(13):11383-11395. https://doi.org/10.1109/JIOT.2023.3241087

Yuan Z, Jie T, Xiaofang D, Jingping Q, Dianjie (2022) Deep reinforcement learning-based resource allocation for D2D communications in heterogeneous cellular networks. Digit Commun Netw 8(5):834-842. https://doi.org/10.1016/j.dcan.2021.09.013

Zhou P, Xie Y, Niu B, Pu L, Xu Z, Jiang H (2021) QoEaware 3D video streaming via deep reinforcement learning in software defined networking enabled mobile edge computing. IEEE Trans Netw Sci Eng 8(1):419-433. https://doi.org/10.1109/TNSE.2020.3038998