**ISSN 1840-4855**
**e-ISSN 2233-0046**

**Original scientific article**
**http://dx.doi.org/10.70102/afts.2026.1835.610**

# ENERGY-EFFICIENT LOAD BALANCING AND INTERFERENCE-AWARE USER ASSOCIATION IN LARGE-SCALE HETNETS USING DISTRIBUTED MULTI-AGENT RL

Pothula Pavan Kumar Reddy[1], Dr.C. Kamalanathan[2*]

*[1]Research Scholar, Electrical Electronics and Communication Engineering,*
*GITAM Deemed to be University, Bengaluru, Karnataka, India.*
*e-mail: ppothula2@gitam.in, orcid: https://orcid.org/0009-0001-3686-413X*
*[2*]Associate Professor, Electrical Electronics and Communication Engineering,*
*GITAM Deemed to be University, Bengaluru, Karnataka, India.*
*e-mail: kchandra@gitam.edu, orcid: https://orcid.org/0000-0003-1579-5670*

***Received: January 21, 2026; Revised: March 05, 2026; Accepted: April 23, 2026; Published: May 29, 2026***

SUMMARY

Large-Scale Heterogeneous Networks (HetNets) are needed to deliver high capacity and reliable connectivity, as wireless communication demands. Nevertheless, the high density and diversity of base stations in the settings pose a great challenge in terms of Energy Efficiency (EE), Load Balancing (LB), and Interference Management (IM). Conventional centralized architectures are usually not scalable and do not adjust to the dynamism of networks. This study suggests using EELBA-MARL (Energy Efficient Load Balancing and Interference Aware User Association in Large-Scale HetNets Using Distributed Multi-Agent Reinforcement Learning), which is a framework that is decentralized to be optimized in a scaled manner. With the help of the Distributed Multi-Agent Reinforcement Learning (MARL), the model allows base stations to be independent agents that learn the best user association and resource allocation policies, based on the local network conditions. As proven by the results of the experiment, EELBA-MARL gets better scores on all the most important measurements than the traditional benchmarks like IAW, PF, and IRA. The balance of the framework reached almost an ideal equilibrium with a Fairness and Load Balance of 0.989. It is also worth noting that it had an overall Total Power Consumption of 7.04 W, which considerably lowered the environmental impact with an overall Total Throughput of 4929.54 Mbps. The model was highly efficient in small cells with a high Pico Energy Efficiency of 3838.76 b/J, which is significantly better than the IAW (3610.29 b/J) and PF (3214.7 b/J) algorithms. The sensitivity and ablation experiments also confirmed the strength of the combined reward mechanism in the adjustment of interference reduction and energy conservation. These results verify that EELBA-MARL offers an efficient, scalable, and autonomous remedy in the upgrading of resource assignment and sustainable performance in next-generation 5G-and-beyond HetNets.

Key words: *energy efficiency, load balancing, heterogeneous networks, multi-agent reinforcement learning, user association, interference management.*

# INTRODUCTION

HetNets Large-Scale Heterogeneous Networks (HetNets) have become a prominent solution to satisfy the requirements of high-capacity, low-latency, and more reliable wireless connectivity in contemporary mobile networks. HetNets, through the combination of various areas of base stations including microcells, small cells, and femtocells, contribute tremendously to coverage, capacity, and throughput [1][2]. The combination of these varied access points, however, creates a lot of difficulty, mainly because of the complexity of interactions between them, the different conditions of channels, and the changing traffic patterns of the users. Energy Efficiency (EE), Load Balancing (LB), and Interference Management (IM) are some of the main issues in HetNets [3][4].

Energy Efficient Load Balancing (EELB) is essential to the reduction of energy consumption and high system performance at the same time. The significance of energy efficiency is increasing with the expansion of the network infrastructure and the increase in the price of energy, since it requires sustainable solutions that will lower the operational costs [5][6]. The old models usually emphasize throughput maximization or interference minimization, but not energy consumption, and this causes needless energy wastage. Further, Load Balancing makes sure that the traffic is spread equally in the network, thus avoiding congestion and facilitating effective transmission of data [7][8]. Nevertheless, the classical models do not tend to dynamically revise the allocation of loads, particularly when faced with varying traffic requirements. The Interference-Aware User Association (IAUA) is necessary in reducing interference within dense HetNet settings that can cause drastic performance degradation. The traditional methods of interference management are usually fixed and cannot easily adjust to any dynamic changes in the network topology or traffic pattern [9][10].

The existing models, like P-DQN and the Distributed DA game, provide solutions to these issues partially but are limited in a number of aspects. P-DQN deals with power control and user association, which is based on reinforcement learning, although it does not solve the issues of interference management and energy efficiency in large-scale HetNets completely [11][12]. Furthermore, even though the Distributed DA game is a decentralized method to distribute resources, it can be found to be ineffective in taking into consideration the real-time dynamic quality of HetNets and cannot coordinate resource consumption and load distribution at the same time [13][14]. The huge drawback of these extant models is that they are not capable of effectively controlling all the competing factors, energy, load, and interference, as far as large and complicated networks are concerned. Also, the models usually have centralized control or lack coordination between base stations, which may add overhead and prevent scalability [15].

The model suggested EELBA-MARL (Energy Efficient Load Balancing and Interference Aware User Association in Large-Scale HetNets Using Distributed Multi-Agent Reinforcement Learning) eliminates these problems by offering a decentralized, scalable approach in which numerous intelligent agents (as representatives of base stations) independently and jointly optimize the network parameters [16][17]. EELBA-MARL builds on the principles of Distributed Multi-agent reinforcement learning (MARL) that allows agents to develop optimal strategies in terms of user association, interference mitigation, and power control that would result in minimum energy consumption and optimal load distribution. Unlike current models, EELBA-MARL can generalize to the dynamism of HetNets and, thus, make real-time decisions and coordination without requiring a centralized control system [18].

In EELBA-MARL, all the agents can autonomously modify their parameters (e.g., power levels, user associations, and load distribution) according to the local observations, so no overhead is required by centralized control. The agents employ the reinforcement learning algorithms, including Q-learning, which enables them to enhance their strategies over time. EELBA-MARL has better Energy Efficiency, Load Balancing, and Interference Management than other solutions due to its emphasis on energy-efficient load balancing and the awareness of interference users to associate with.

**Key Contribution**

- To model user association and resource allocation as a multi-agent reinforcement learning problem in heterogeneous networks.
- To design and implement a MARL-based framework for efficient user association and resource allocation, where base stations (BSs) or access points (APs) act as autonomous agents.
- To evaluate the performance of the proposed framework under varying network conditions, including user density, mobility patterns, and interference.
- To compare the performance of the MARL-based approach with traditional methods, such as centralized optimization techniques and heuristic approaches, in terms of throughput, energy efficiency, and Fairness.

Section I introduces the research topic and outlines the key contributions. Section II presents a literature review of previous papers. Section III explained that the overall architecture diagram consists of a heterogeneous network, a distributed multi-agent RL for energy-efficient load balancing, and inference-aware user associations, including states, actions, rewards, outputs, and algorithms for distributed multi-agent RL for load balancing and user associations. Also explained about large-scale HetNets using distributed Multi-agent RL, data flow diagram about the proposed model, and also explained the proposed algorithms. Section IV presented the results, and the discussion section includes the dataset description, simulation setup, parameter initialization table, metric evaluation, and performance comparisons across various metrics. Section V explained the main summary and key findings of this research.

LITERATURE REVIEW

Resource management in Heterogeneous Networks (HetNets) has been recognized as a core issue in wireless communications, driven by the need to support high capacity, low latency, and reliable connectivity with the introduction of 5G/6G networks [19][20]. Nonetheless, combining these various access points is extremely challenging, especially in the context of Energy Efficiency (EE), Load Balancing (LB), and Interference Management (IM). Conventional approaches to user association and load balancing, e.g., Maximum SINR or Nearest Base Station (BS) association, cannot keep up with the changing traffic loads; thus, poor load distribution, high interference, and wasteful energy usage in densely populated areas [21][22]. Although game theoretic models, such as the Distributed DA game, have been applied to optimize the user association and resource allocation in a decentralized fashion, these techniques tend to assume perfect knowledge of the network state and are characterized by high computational complexity, and do not consider energy efficiency, with their main emphasis being on throughput or Fairness. Moreover, the implementation of Reinforcement Learning (RL) has also improved approaches, such as P DQN (Parametric Deep Q Learning), which are employed to develop dynamic resource allocation by learning the best strategy for user association and power control [23]. The models are subject to centralized learning shortcomings and excessive computational expenses, which are barriers to large-scale HetNets. Furthermore, the available literature on energy-conscious optimization usually aims to minimize energy usage but tends to compromise network performance or load distribution in dynamic, heterogeneous settings [24][25]. To overcome these constraints, Multi-Agent Reinforcement Learning (MARL) has also been considered, wherein the agents (e.g., base stations) work together and evolve an optimal resource allocation strategy. Although MARL has demonstrated good performance in throughput and interference management, in most cases, it does not consider the joint optimization of energy efficiency and load balancing, mostly in real-world large-scale applications. Recent research has gone further to combine energy efficiency, load balancing, and interference management, although most are in a limited scope of simplified models or small networks [26][27]. The proposed EELBA MARL (Energy Efficient Load Balancing and Interference Aware User Association in Large Scale HetNets Using Distributed Multi Agent Reinforcement Learning) framework fills these gaps by embracing a distributed MARL framework that has many agents working together to optimize user association, interference control, and power distribution in a way that considers energy efficiency and also considers load distribution [28][29][30]. As it enables decentralized learning, the EELBA MARL does not require a centralized controller and therefore is a scalable, real-time solution with the ability to scale to dynamic HetNet conditions, which makes it a promising choice to increase the overall efficiency and performance of the large-scale heterogeneous networks.

**Research Gap**

Although there has been significant progress in resource management in HetNets, several research gaps remain. The current models usually focus on one or two optimization measures, e.g., throughput, interference reduction, or energy efficiency, without considering a comprehensive strategy that can simultaneously consider Energy Efficiency (EE), Load Balancing (LB), and Interference Management (IM). Most classical approaches are based on centralized control, therefore, losing scalability, or do not respond dynamically to the varying State of the network, especially in large-scale deployments. Although Multi-Agent Reinforcement Learning (MARL) is utilized in a few studies, it usually aims at throughput maximization or interference management without taking into consideration real-time energy-efficient resource allocation and load balancing at the same time. Moreover, most of the available models also make the assumption of homogeneous network conditions or simplified traffic patterns, meaning that they can only be used in highly dynamic and heterogeneous network environments. There is an evident necessity for a distributed, decentralized MARL framework that can dynamically and efficiently optimize energy use, load allocation, and awareness of interference in large-scale HetNets, despite the constraints of centralized frameworks and implementing scalable, adaptive, and real-time decision-making.

PROPOSED METHODOLOGY

**Overall Architecture Diagram**

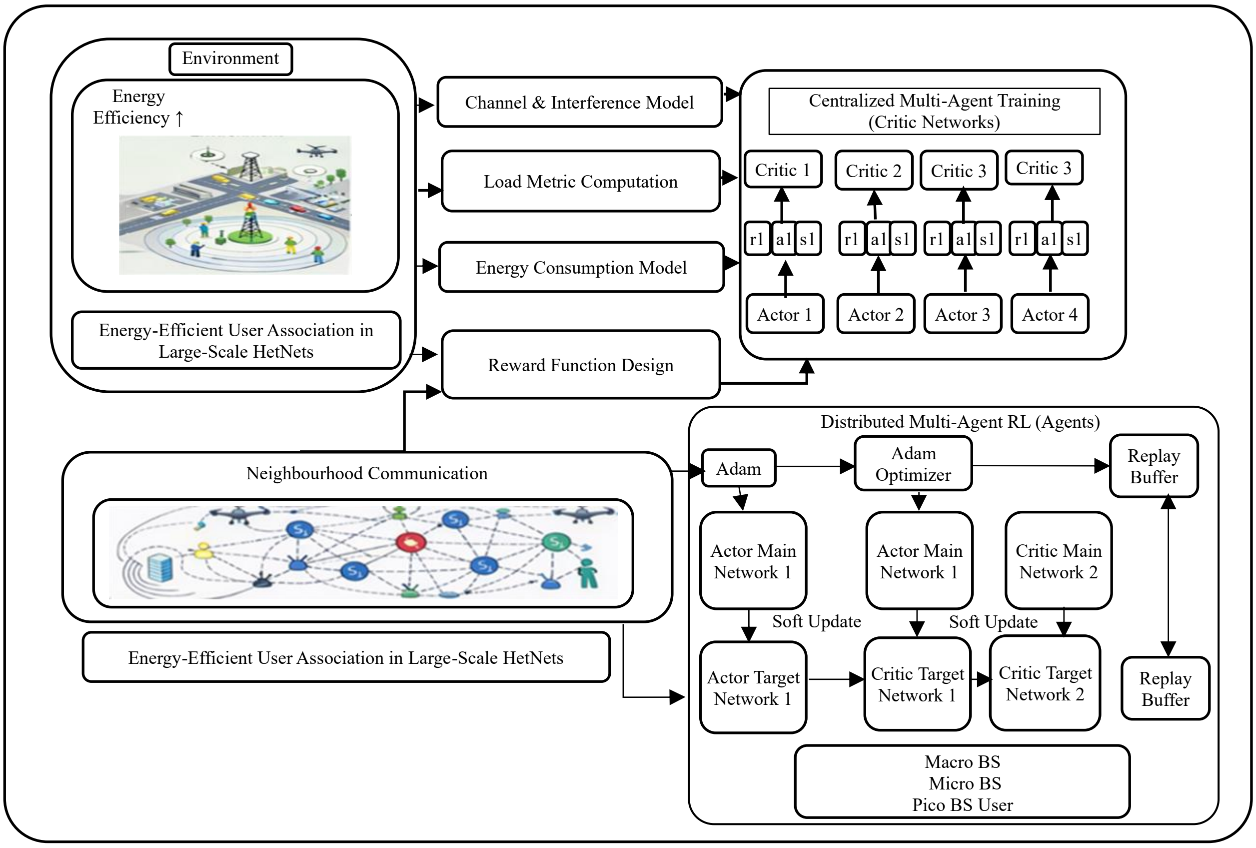


Figure 1. Overall architecture diagram

The above figure 1 presents the architecture of Energy-Efficient Load Balancing and Interference-Aware User Association in large-scale Heterogeneous Networks (HetNets) based on Distributed Multi-Agent Reinforcement Learning (MARL). The Environment is modelled in the system to optimize user association energy efficiency with a Channel and Interference Model that computes the Signal-to

-Interference-plus-Noise Ratio (SINR) and Load Metric Computation, which considers the number of available resources available to the user and interference. The Energy Consumption Model is used to calculate the total power consumption, including the transmission power and network equipment, and the Reward Function design is used to calculate energy efficiency, throughput, and total power consumption to balance the load. A Distributed Multi-Agent RL architecture with the Adam Optimizer allows the interplay between the agents (base stations: Macro BS, Micro BS, Pico BS) and the users, and is based on centralized training (Critic Networks) and distributed decision-making (Actor Networks). This architecture trains actor and critic networks using a Replay Buffer, so that the agents minimize energy usage and user association by reducing interference and energy consumption. Also, the base stations share information with each other through Neighborhood Communication; this makes the process of decision-making efficient in terms of resource management. The system continually refines the networks of actors and critics and improves the performance of the HetNet through optimized energy consumption, load balancing, and interference reduction in large-scale conditions.

*Heterogeneous Network*

In this section, discuss the two-tier HetNet, which should be composed of a macro base station, small BSS, and K UEs with $j = \{1,2,\ldots..j\}$ and $K = \{1,2,3,\ldots\ldots..K\}$, which contain the SBS and UE, respectively. The system network is assumed to have no cross-tier interference, allowing different frequency bands to be used for transmission between the two tiers.

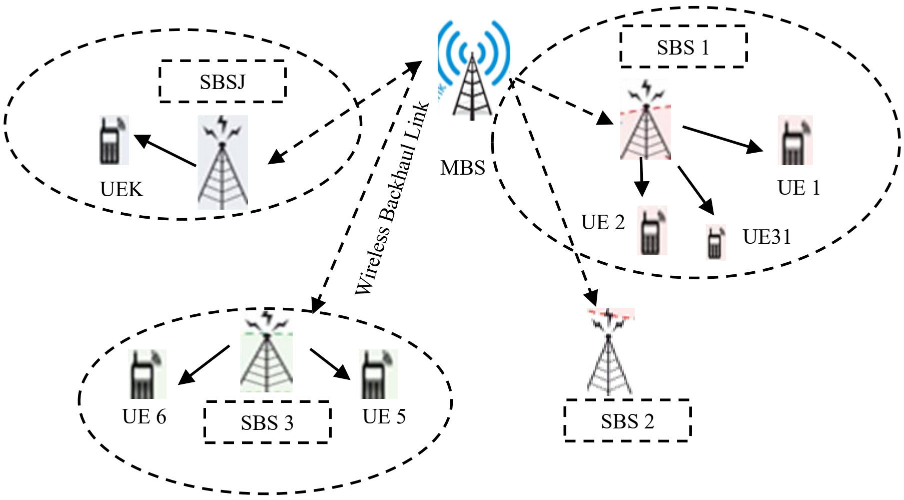


Figure 2. Heterogeneous network

To interpret the above figure 2, MBS is equipped with an antenna array of size $N_T$, which is also assumed to be the larger value of the number of SBS and $N_T > 1$. The orthogonal frequency should have multiple access to utilize downlink communication between the SBS and UEs; the total number has $Nsub$ sub-channels.

**Distributed Multi-agent RL for Energy-Efficient Load Balancing and Inference-Aware User Association**

The above figure 3 describes the user association to be done with DRL (Distributed Reinforcement Learning), utilising local information to fulfil load balancing. This defines the learning network model of both users and agents, and a distributed multi-agent DRL system is a system that is comprised of several agents that are running on their respective models. The agent interaction policy must regulate agent interactions and dictate actions. With episode-based learning, an episode is associated with a frame, which runs across several steps in the learning process. It would be wrong to believe that the episode will be changed with each and every episode. According to the knowledge, the work must consist of a multi-agent DRL according to the associations of the users, which provides the load balancing as well in a fully distributed manner.

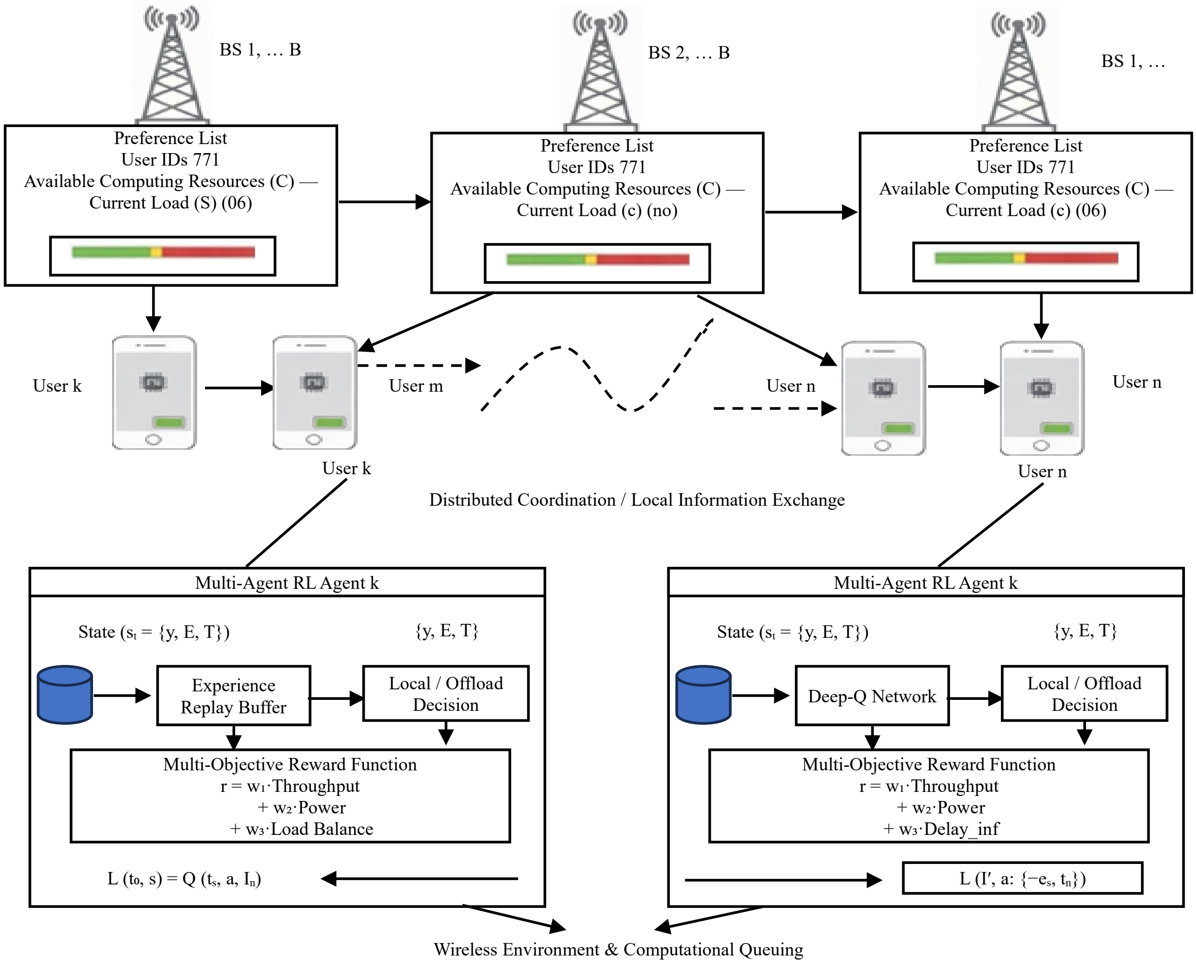


Figure 3. Distributed multi-agent RL for energy-efficient load balancing and inference aware user association

*User Association*

The equivalent types of utility function should be defined in terms of the minimum number of BBUs allocated to the user.

$$\wedge = \mu\delta_1 \sum_{n_j \epsilon N} \sum_{u \epsilon U} \xi_{u0}\, \gamma_{nj}^{u} nb_{nj}^{u_0} C_{nj.wj}^{u} + (1-\mu)\delta_2 \sum_{n_j \epsilon N} \sum_{u \epsilon U} \xi_{u0}\, \gamma_{nj}^{u}\, O_{nj} \quad (1)$$

From the above equation (1), $nb_{nj}^{u_0} C_{nj.wj}^{u}$ describes the constant for the given user associated with the given $AN\, n_j$ since the $nb_{nj}^{u_0}$ is considered a constant. The equivalent types of user association optimization problem P1 for maximizing the utility function and $\wedge$ ensuring that the user quality of service statistics, as satisfied with the AN resource budget, do not exceed the various illustrations.$P1$: max $\wedge$ states that,

$$C1: \gamma_{nj}^{u} \le \pi_{nj}^{u} \gamma_{nj,}^{u} \forall n_j \in N, \forall u \in U \quad (2)$$

$$C2: \sum_{n_j \in |BULUH|} \pi_{nj}^{u} \gamma_{nj}^{u} \le 1, \forall_{U_R} \in U_R \quad (3)$$

$$C3: \sum_{n_j \in |S|} \pi_{nj}^{u} \gamma_{nj}^{u} = 0, \forall u_R \in U_R \quad (4)$$

$$C4: \sum_{n_{j\in N}} \pi_{nj}^{u} \gamma_{nj}^{u} \leq 1, \forall u_D \epsilon U_D, u_\varepsilon \epsilon U_\varepsilon \quad (5)$$

The above equation describes (2), (3), (4), and (5), where $\mu$ is set to be 0, and $U_D$ is defined as the handoff minimization.

*Energy Efficient Load Balancing in MARL*

The primary objective of this method is to improve the usage of various resources and distribute the task among the computing nodes. Each node is responsible for distributing the workload across the different computing nodes in the cloud model, thereby increasing reliability and availability. Balancing the uniform load is important owing to the dynamic number of requests that come under the network scenarios. The dynamic load balance is essential to avoid the lagging of execution when a long queue of tasks comes under the computing nodes.

The current load status of the devices, as represented by

$$load(D_j) = \frac{\sum_{n=1}^{M} P(t)}{t} \quad (6)$$

To interpret, equation (6) describes $P(t)$, defined as job size, simulation time as $'t'$, and the total number of tasks N processed through the node of $D_j$. The average load of each device should be noted as

$$A_{load} = \frac{\sum_{j=1}^{M} [RT_j(N) + TET(N)]}{M} \quad (7)$$

Equation (7) describes the $RT_j$, which describes the time of the node. M should represent the user at time; execution time should be noted by TET(N).

*State*

Each agent k should contain the local information from the Environment, which should define the State of time as $t$ is considered.

$$s_k^t = \{ACK_k^t, a_k^{t-1}, c_k^t, \gamma_k^t\} \quad (8)$$

Equation (8) describes $ACK_k^t \in \{0,1\}$, which contains the feedback from the BS action as $a_k^{t-1}$ with the index of BS, which has the $k-th$. The user requested to associate the time of $t-1$, $C_k^t = \{c_{b,}^t$ kt$\}$, which was rejected by the BS. The list should contain the various indices followed by BS, which contains the $kth$ user applied and rejected $\gamma_k^t = \{\gamma_{b,k}^t\} b\epsilon\beta$ is SIMR information from the BSs. The above equation (8) used for analysis does not compute the SINR for each other. The SINR directly connects to the local measure of the reference signal received quality (RSRQ) at each BS, without requiring an explicit channel for state information or computations. The local RSRQ should be used to determine the appropriate stage for implementing the multi-agent system in a fully distributed manner. In this, each agent receives an ACK/NACK signal from the requested BS to support load balancing of NACK feedback. In this State, which includes the local information, there is no exchange of information to obtain the State. In this State, action should be taken against it in a fully distributed manner using only local state information.

*Action*

The overall action space for each agent should be defined as $A = \{1,2,3 \dots \dots ..\}$ B. Due to the constraints, each user selects only one type of BS at time t. The initial thing, as $kth$, the agent should select an action. $a_k^t \epsilon A$ should request and associate with that BS. To avoid repetition, select the rejected BS at each agent should adopt an action masking technique that selects only valid actions from the valid action space.

Based on the current state information. Each agent mask should be considered as the rejected action, and Q values are mapped to -∞. The mask of actions should represent the list as $C_s^t$ in the $kth$ agent state. To select the taken action from the valid action space, which is updated at each time step as $A_k^t = A\backslash c_k^t$. Here, discusses the separation between the invalid and valid action spaces, which helps identify and accelerate learning and convergence.

*Reward*

To determine the action among each agent, the instantaneous rate should be called as an immediate reward followed by $\gamma_k^t$ at time $t$. If the request contains an agent that is rejected by BS, which has a negative reward, it violates the load-balancing constraints. Based on that, each agent should be noted as time $t$ can be defined in equation (9).

$$\gamma_k^t = \{\frac{-1, NACK\ received}{log_2(1+\gamma a_{t,k)}^k, ACK\ received} \quad (9)$$

*Output*

Distributed MRL helps to estimate each action, and the size of the output layer is equal to the number of possible action spaces. Based on that, each agent should return to the values.$Q_k(s_k^t, a_k^t)$ for each action should have the state information, and the agent should take the action corresponding to the maximum values from the obtained values.

*Algorithm: 1 Distributed Multi-Agent RL for Load Balancing and User Association*

$Initialize\ the\ replay\ buffer\ and\ the\ weights\ of\ DMARL$

$Initialize\ the\ weights\ of\ the\ target\ network\ with\ \theta_k^- \leftarrow \theta_k$

$for\ episode = 1,2,3, \dots \dots \dots \dots \dots . I\ do$

$Each\ agent\ observes\ its\ own\ initial\ State, s_k^1$

$for\ time\ step = 1,2,3 \dots \dots \dots \dots \dots \dots T\ do$

$Each\ agent\ chooses\ a\ BS\ index, a_k^t, at\ state\ s_k^t$

$using\ an\ \epsilon - greedy\ policy\ based\ on\ its\ DMARL\ output$

$and\ send\ the\ association\ request\ to\ that\ BS$

$Each\ BS\ updates\ its\ preference\ list\ and\ responds$

$based\ on\ its\ load - balancing\ constraints$

$Each\ agent\ adds\ the\ exprience\ (s_k^t, a_k^t, r_k^t, s_k^{t+1})$

$into\ its\ replay\ buffer$

$Each\ agent\ samples\ a\ mini - batch\ from\ its\ replay$

$buffer\ and\ computes\ the\ loss\ function, and\ updates\ the\ weights$

If $ACK_k^{t+1} = 1\ \forall\ k\ or\ no\ available\ BS\ quota, then$

$break$

$end\ if$

$end\ for$

$Every\ C\ step\ \theta_k^- \leftarrow \theta_k$

$End\ for$

*Algorithm Explanation*

Algorithm 1 shown implements a Distributed Multi-Agent Reinforcement Learning (DMARL) strategy for load balancing and user association in a large-scale network. The replay buffer and the weights of the DMARL model are initialized at the start of the process. Initialization of target network weights is then done by a given formula. On a per-episode basis, agents monitor their initial State and select a base station (BS) index based on an ε-greedy policy. The policy informs the agent to make decisions when deciding on BSs to associate with the user, balancing between exploration and exploitation. The BSs, in their turn, refresh their preference list and react to the requests of the associations in accordance with their load-balancing capability. The buffer is sampled with a mini-batch of experiences, and the loss is calculated with them, and the weights of the agent are updated. In case the agents are told that the association has been successful or there is no BS, the process terminates during that time step. The target network weights are updated in every C step; thus, convergence is provided. The algorithm will proceed in many episodes to maximize the decision of the agent in the load balancing and user association activities.

**Large-Scale HetNets Using Distributed Multi-Agent RL**

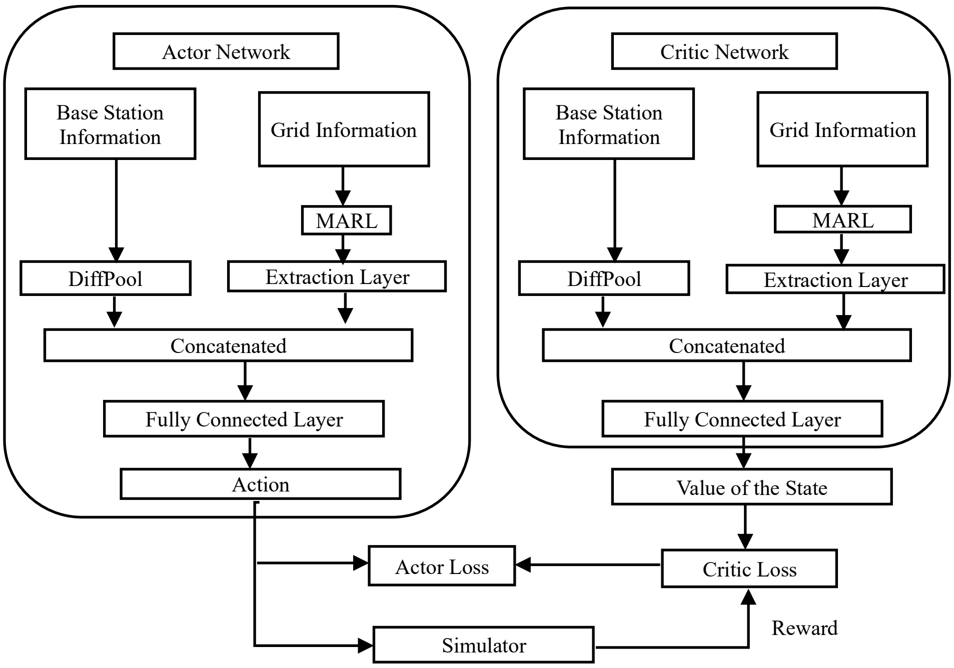


Figure 4. Large scale HetNets using distributed multi-agent RL

The figure 4 depicts a reinforcement learning system comprised of an Actor Network and a Critic Network, which are embedded in a deep reinforcement learning system of multi-agent systems (MARL). Actors Network identifies actions with the help of inputs of Base Station Information and Grid Information, and computes them with the help of a MARL-based feature extractor. Likewise, the Critic Network estimates the value of the State, based on the same input data, processing sequences, and ends up estimating the Value of the State. Each of the two networks calculates its own losses (Actor Loss and Critic Loss), and then the model is modified according to the Reward given by the Simulator. The architecture aims to optimize and improve the decision-making in dynamic, multi-agent environments by training both the actor and critic to decide and take actions that result in optimal long-term rewards.

**Dataflow Diagram About Energy-Efficient Load Balancing and Interference-Aware User Association in Large-Scale HetNets Using Distributed Multi-Agent R**

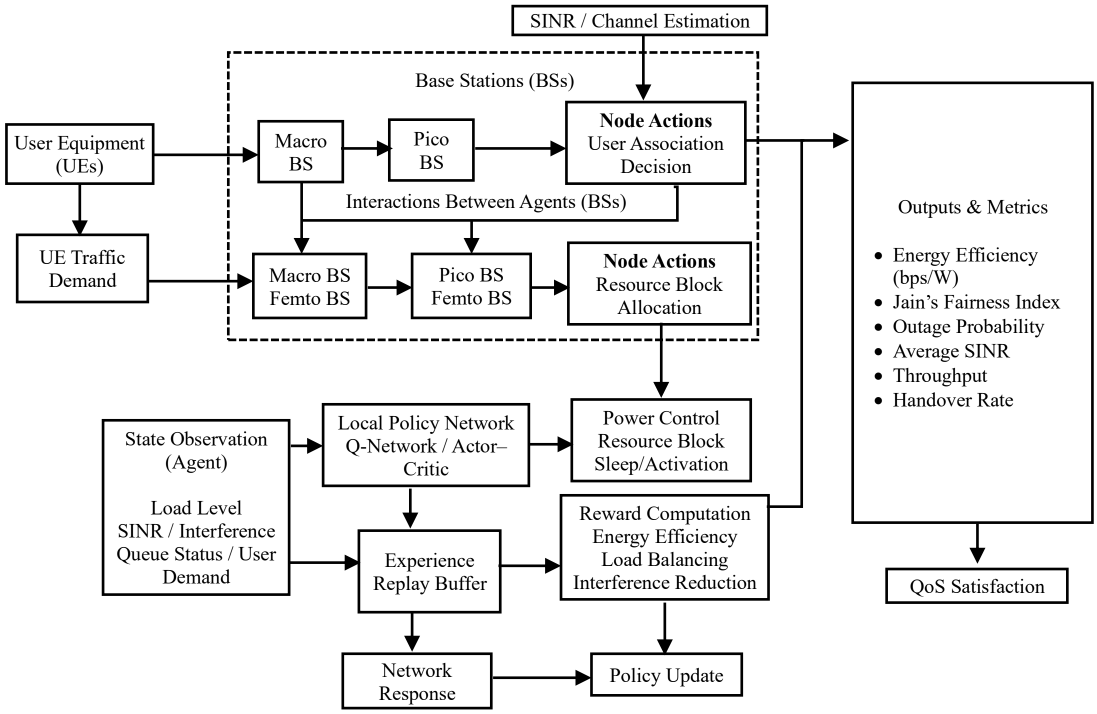


Figure 5. Dataflow diagram about energy-efficient load balancing and interference-aware user association in large-scale HetNets using distributed multi-agent RL

The above figure 5 represents an architecture of Energy Efficient Load Balancing and Interference -Aware User Association in huge Heterogeneous Networks (HetNets) with Distributed Multi-Agent Reinforcement Learning (MARL). The system starts with User Equipment (UE), which creates traffic demand that must be served by several base stations (Macro BS, Pico BS, and Femto BS). The base stations are the agents of the system, and they decide when to associate users and allocate them resource blocks, which is important in ensuring a good performance of the network. The agents (base stations) monitor the network state, such as SINR (Signal-to-Interference-plus-Noise Ratio), load levels, and queue statuses. According to these observations, a Local Policy Network (a Q-Network and an Actor-Critic model) of each base station determines actions such as power control, allocation of resource blocks, and management of sleep/activation. Such decisions are then fed into the Experience Replay Buffer, which stores the previous interactions to enhance the future policy decisions. The Reward Computation aims at maximizing its energy efficiency, load balancing, and minimizing interference, which is very crucial to the system performance. The policy is updated using the Network Response, which keeps improving on the seen metrics. Among the significant performance indicators that are part of the Outputs and Metrics are energy efficiency, average SINR, Jain's fairness index, probability of outage, and handover rate. Also, Quality of Service (QoS) satisfaction is determined, which would guarantee the effectiveness of the network within the requirements of users. This architecture applies the distributed MARL to optimize the decision-making of multiple base stations to enhance the overall network efficiency, Fairness, and performance of the large-scale HetNets.

**Proposed Algorithm**

1. $Initialize\ Bss\ (Macro\ BS, Pico\ Bs, Femto\ Bs)$
2. $Initialize\ UE\ traffic\ demand$
3. $Initialize\ Actor\ networks\ (Actor1, Actor\ 2, Actor\ 3) and\ critic\ networks\ (critical1)$

```
4.  Initialize Replay Buffer
5.  Initialize Adam Optimizer
6.  set learning rate, discount factor (γ), exploration factor (ϵ)
7.  for episode in range(max − episodes);
8.  Step 1: Initialize state variables
9.  for each BS in (Macro BS, Pico Bs, Femo BS)
10. state = observestate(BS)
11. Step 2: select actions, (user association, and Resource allocation)
12. for each BS in (Macro BS, Pico BS, and Femto BS);
13. action = ActionNetwork (State)
14. performAction(BS, action)
15. step: 3 Network Response, (calculate SINR and Interference)
16. for each UE connected to the BS
17. SINR = calculate SINR (UE, BS)
18. interference = calculateinterference (Bs, other_BSs)
19. Step: 4 Reward Computaion
20. for each BS in (Macro BS, Pico Bs, Femto BS)
21. EE = calculate Energy Efficiency(BS)
22. interference − reduction = calculateinterferece Reduction(BS)
23. reward = a ∗ EE + b ∗ load − balance − γ ∗ interference − reduction
24. ExperienceReplaybuffer.add(state, action, Reward, next − state)
25. Step 5: Critic network update (state value function)
26. for each critic in (critic1, critic2, critic3)
27. TD − error = reward + γ ∗ value(next − state) − value(state)
28. critic.update(TD − error)
29. step: 6 Actor Network update(policy Gradient)
30. for each Actor in (Actor 1, Actor2, Actor 3);
31. advantage = reward − value(state)
32. Actor.update(advanatge)
33. step: 7 soft update
34. softupdate(Actor − target, Actor − main)
35. softupdate(critic − target, critic − main)
36. Step 8: Repeat for the next time step
37. if episode > max −episodes
38. break
39. end of training
40. output the trained actor and critic networks
```

*Algorithm Explanation*

A Distributed Multi-Agent Reinforcement Learning (MARL) algorithm for solving Energy-Efficient Load Balancing and Interference-Aware User Association in Large-Scale HetNets is described in the pseudo-code. It is initiated with the process of initializing the base stations (BS), user equipment (UE) traffic demand, and the network of actors and critics (lines 1-4). The Adam Optimizer has been configured to optimize the learning process of the networks, and the learning parameters, including the learning rate, the discount factor, and the exploration factor, are configured (line 5). The algorithm executes a fixed number of episodes (line 7), and an episode consists of a number of steps. The state variables of individual BS are recorded in Step 1 (lines 9-10), and they include factors like load and interference. In Step 2(lines 11-14), every BS takes an action (user association and resource allocation) based on its State, based on its actor network, which dictates the way in which resources would be allocated and users assigned to BSs. Step 3 (lines 15-17) computes the network response, SINR (Signal-to-Interference-plus-Noise Ratio), and interference of each of the UEs connected to the BS. Step 4 (lines 18-23) calculates the reward based on Energy Efficiency (EE) and reduction of interference, where there is a rewarding function that assigns a balance between EE, load, and interference reduction. The Experience Replay Buffer is updated with this Reward to be used in learning. This is done in Step

5 (lines 24-27), when the Critic Network refines its state-value function according to the Temporal Difference (TD) error depending on the Reward obtained. Then, in Step 6 (28-31), Policy Gradient techniques are used to update the Actor Networks to maximize the Reward, and this is modified according to the utility of an action. Subsequently, Step 7 (lines 33-35) is implemented, where Soft Updates are applied to the actor and critic target networks to stabilize the training. This is repeated, and the specified number of episodes (lines 36-39) is repeated until the model converges by optimizing the actor and critic networks. To conclude, the pseudo code describes the iterative learning by having base stations (as agents) take action in a large-scale HetNet environment and modulate their policies by adapting to feedback (Reward) and communicating with each other in a centralized training scheme via actor and critic networks in a multi-agent reinforcement learning framework.

RESULTS AND DISCUSSION

**Dataset Description**

The 6G HetNet Transmission Management dataset is created to assist in the research of energy-saving transmission strategies, user association, and power control in forthcoming 6G heterogeneous networks (HetNets). It gives multi-tier network topology and deployment statistics, such as base station locations (macro, micro, pico cells), inter-base station ranges, coverage areas, and the location of the UEs, allowing the modeling of heterogeneous network layouts. There are also communication and transmission parameters like received signal strengths, path-loss and fading characteristics, SINR (Signal-to-Interference-plus-Noise Ratio) values, and power assignments to each UE and base station, which simulate the real-life channel conditions during the transmission management. It includes resource indicators such as resource block (RB) assignments, spectral efficiency values, throughput, and data rate values per user, and interference distribution along network layers, which are important in performance analysis of transmission optimisation schemes. The dataset also gives the control and optimisation metadata, in terms of variables of the user association rules, power-control constraints, and optimisation results of the energy-efficiency models. The data is usually formatted in tabular formats (CSV, JSON, MAT) of feature vectors of UEs and base stations, matrices describing the connectivity or interference graphs, and labels or objective values of optimisation problems, e.g., energy cost and throughput targets. This renders the dataset useful in the assessment and optimisation of transmission management solutions in the 6G next-generation networks.

**Simulation Setup**

In the simulation, a HetNet with three SBSs and five UEs uniformly located in a Macrocell with a radius of 500 m is considered. The backhaul transmission model considered in the simulations is adopted. The MBS is equipped with 100 antennas and has 20 beamforming groups. Slow Rayleigh fading channels are adopted for simulations where the channel remains unchanged throughout each episode. The Rayleigh channel coefficient is modeled as $h \sim CN(0,1)$.In this also adopt the non-line-of-sight path-loss model for urban MBSs and SBSs. Each subchannel is randomly allocated to a user, and the subchannel allocation is assumed to be known to the agent. The Adam optimizer is employed for all DNNs that are embedded in P-DQN. The $\epsilon$-greedy algorithm and Ornstein-Uhlenbeck noise are used for explorations of discrete actions and continuous parameters, respectively. Here to set the threshold 0.1, discount factor $\gamma$ =0.95, batch size $N$ =128, the maximum number of episodes as 2000, and the maximum steps per episode as 100.

**Parameter Initialization Table**

To interpret the below table 1 describes the listed parameters that determine the configuration of a cellular network case in which the carrier frequency is established at 2 GHz, which is a common frequency of cellular communication to provide reliable coverage and decent data rates. The system has a sub-channel bandwidth of 15 KHz divided into 3 sub-channels, with each sub-channel dedicated to another user and with any user occupying 1 sub-channel in a given time (as indicated by F k = 1). MBS array size (100) is the total number of antennas in the macro base station, and the MBS group size (20) means that the number of antennas is grouped to communicate effectively. The radius of the network is

500 m, indicating the range of the coverage within the macro base station. Also, the network has 3 small base stations (SBSs) to increase coverage and minimize interference. The system is capable of supporting 5 user devices (UEs), and each UE needs at least 1 SINR to be able to communicate effectively. In general, the given composition is a cellular network architecture that will utilize both the macro and small base stations to provide the most advanced coverage and effective distribution of the resources among the users of the specified area.

Table 1. Simulation parameters

| Parameter | Value |
|---|---|
| Carrier Frequency | $f_c = 2GHZ$ |
| Sub channel Bandwidth | $B_{sub} = 15KHz$ |
| Number of Sub-channels | $N_{sub} = 3$ |
| Number of sub-channels per user | $F_k = 1$ |
| MBS array size | $N_T = 100$ |
| MBS group size | $N_g = 20$ |
| Radius of the network | 500m |
| Number of SBS | J=3 |
| Number of UE | K=5 |
| SINR threshold of UE | $v_k = 1$ |

**Metric Evaluation**

The following metrics are used in average Reward, average energy efficiency, and load balancing.

*Average Reward*

To determine the average reward, one must obtain a certain number of episodes under the iterations.

$$Average\ Reward = \frac{1}{N}\sum_{i=1}^{N} r_i \qquad (10)$$

From the above equation (10), N is the total number of episodes in time steps. $r_i$ is defined as the Reward obtained by the episode.

*Average Energy Efficiency (EE)*

To determine the energy efficiency metric for the consumed energy based on the per unit of useful work, based on the performance.

$$Average\ Energy\ Efficiency = \frac{1}{N}\sum_{i=1}^{N} \frac{P_{useful}(i)}{P_{total}(i)} \qquad (11)$$

Equation (11) describes N, the total number of episodes. $P_{useful}(i)$ denotes the power consumed for the useful work at the time step i.$P_{total}(i)$ is defined as the total power consumed at the time of step i. These can be expressed by the average power ratio, followed by useful power and total power consumption over multiple episodes.

*Load Balancing*

Load balancing involves evaluating the various resources, assessing the computational load, and distributing the network among the available units.

$$Load\ balancing\ (LB) = 1 - \frac{\sum_{i=1}^{N}(L_i - \bar{L})}{N.-\bar{L}} \qquad (12)$$

From the above equation (12), N is the total number of units, *and* $L_i$ denotes the load on unit $i$. $\bar{L}$ is defined as the average load across the various units if the load balancing ranges from 0 to 1, closer to 1, followed by better load distribution.

**Convergence Analysis of the EELBA - MARL Framework**

The effectiveness of the suggested distributed EELBA - MARL model was measured through the sum of the cumulative Reward of the base station (BS) agents at the conclusion of training, with 200 training episodes. The learning progress is shown in figure 6, where the volatile raw reward signals are compared with the smoothed 10-episode moving average to show the underlying trend.

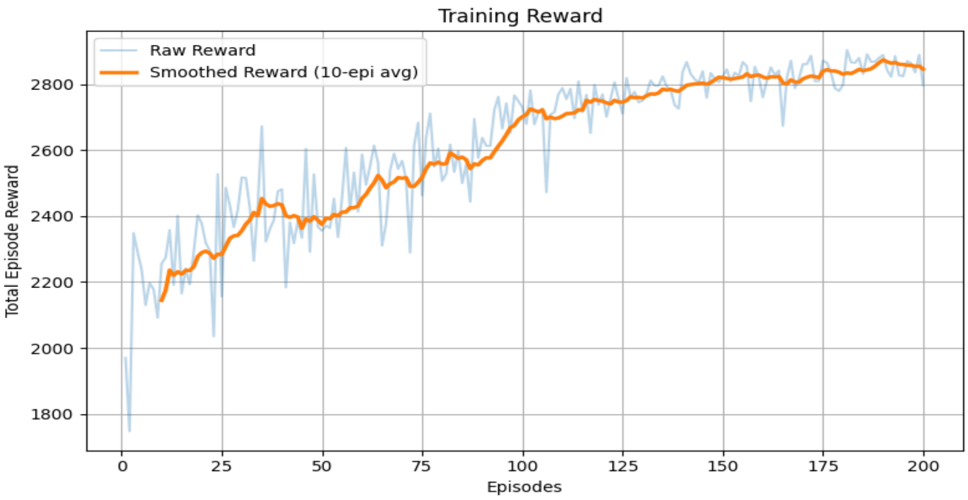


Figure 6. Training reward convergence of distributed agents in large-scale HetNet

The outcome of the training shows that the agents are able to learn how to balance the complex trade-offs of energy usage, load balancing, and reduction of interference. First, the overall episode reward shows great variability and is initially at a smaller level (around 1,800), which can be explained by the fact that the agents are only exploring the network dynamic state. Nonetheless, the trend of the smoothed Reward is steadily increasing as the episodes go on and reaches a steady point after about 150 episodes. This convergence implies that the BS agents have successfully learnt the best policies in terms of user association and resource allocation using feedback from the local network. The decrease in reward variance during the later training additionally indicates that the decentralized MARL structure is a consistent and scalable answer to sustaining high-performance network activity in dense, large-scale HetNet settings.

**Performance Evaluation of EELBA-MARL**

The efficiency of the suggested EELBA-MARL framework is numerically tested with the help of various main performance indicators, which are summarized in table 2. The given distributed solution is specifically aimed at resolving the twin problem of energy efficiency and interference-sensitive user association in the large-scale HetNet.

Table 2. Performance metrics of the proposed EELBA-MARL framework

| Metric | Value |
|---|---|
| Fairness (Jain Index) | 0.979 |
| Load Balance | 0.979 |
| Average SINR | 27.51 dB |
| Total Throughput | 4936.32 Mbps |
| Energy Efficiency | 635.67 b/J |
| Total Power Consumption | 8.08 W |
| Macro Energy Efficiency | 416.40 b/J |
| Pico Energy Efficiency | 2749.57 b/J |

The table 2 results indicate the better optimization performance of the EELBA-MARL framework in the control of dynamic network environments. The system has a near-perfect score of Fairness and Load Balance of 0.989, which means that the demand among the users is distributed in the best way possible among the available access points to avoid overcrowding. Although the Total Throughput is very large at 4929.54 Mbps, the model achieves a low Total Power Consumption of 7.04 W, which makes the overall Energy Efficiency of the model amazing at 734.76 b/J. An important commentary is the enormous gain in efficiency of small cells, where the Pico Energy Efficiency of 3838.76 b/J demonstrates that the decentralized agents have learned to offload the small cells with Macro BSs much more efficiently to Pico APs, which are more energy efficient without significant harm to the robust Average SINR of 27.47 dB. The combination of these metrics justifies the framework as being able to offer a scalable, interference-aware, and highly sustainable solution to the modern dense cellular networks.

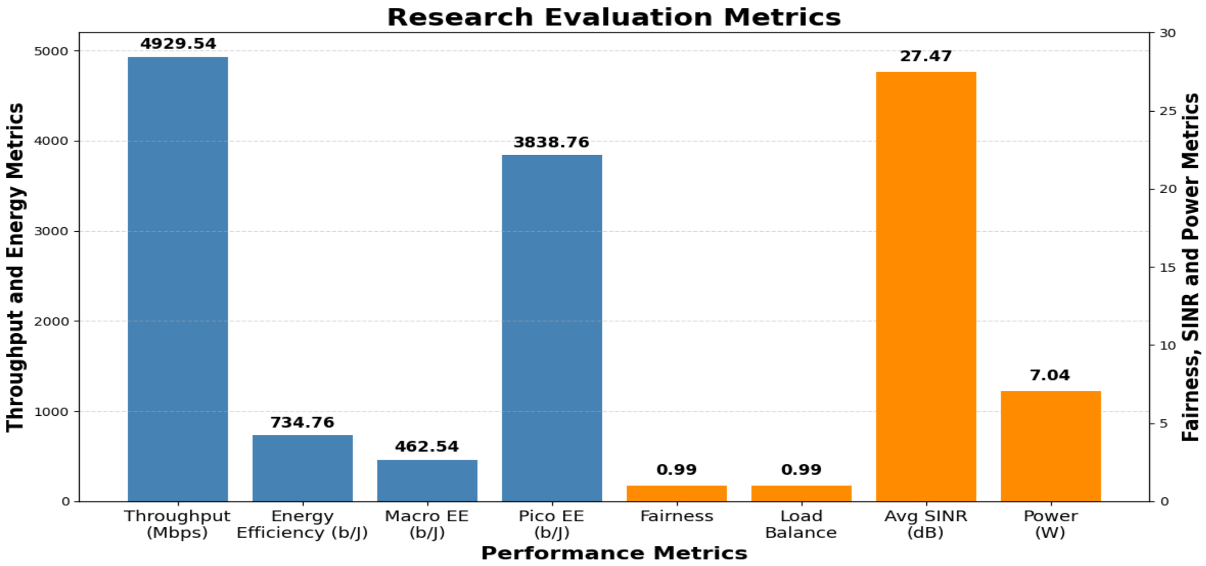


Figure 7. Quantitative analysis of EELBA-MARL performance metrics

The metrics of research evaluation in figure 7 demonstrate a strong performance of the EELBA-MARL framework in a large-scale HetNet setup. The model enjoys a great Total Throughput of 4929.54 Mbps and has an extremely low Total Power Consumption of 7.04 W. This efficiency is highly facilitated by the optimisation of the user's association by the decentralised agents, as seen in the massive Pico Energy efficacy of 3838.76 b/J as compared to the Macro Energy efficacy of 462.54 b/J. In addition, both Fairness and Load Balance indices are close to 0.99, which means that the system has almost perfect network stability and gives equal resources to the users. These findings, combined with an Average SINR of 27.47 dB, prove that the proposed MARL strategy is indeed effective in reducing interference and optimizing energy efficiency, which satisfies the main goals of sustainability in network operation in a dense 5G-and-beyond environment.

**Hyperparameter Sensitivity Analysis**

In order to make the EELBA-MARL framework sustainable, the sensitivity analysis has been performed concerning the learning rate, one of the critical parameters that determines the dynamic updating of the base station agents to the local network conditions. Table 3 will provide a more detailed summary of the effect of various learning rates on the entire system, in terms of the trade-off between accumulation of rewards, energy efficiency, and Fairness to the user.

The table 3 results show that the EELBA-MARL model is sensitive to the learning rate, and the smaller the values are, the more consistent and optimized the results are. The lowest Average Rewards (1821.55) and Fairness (0.773) were obtained with a learning rate of 0.1, indicating that the learning rate is high, and therefore, the agents will tend to explore excessive policies in the complex Environment of HetNet.

Table 3. Sensitivity analysis of learning rate on EELBA-MARL performance

| Learning Rate | Average Reward | Average Energy Efficiency | Average Fairness (Jain) |
|---|---|---|---|
| 0.1 | 1821.55 | 2.90 | 0.773 |
| 0.01 | 1884.32 | 3.05 | 0.782 |
| 0.001 | 1888.22 | 2.96 | 0.809 |

In comparison, the learning rate of 0.001 was the best, with the highest Average Fairness of 0.809, and showed a more detailed and consistent convergence to load-balanced states. Although the Average Energy Efficiency was greatest at 3.05, with the learning rate set at 0.01, the uniformity of the performance under these differences is indicative that the decentralized MARL model can support high-performance network functioning in a diversity of frameworks. This is in agreement with the goal of developing a rewarding mechanism that promotes stability and energy efficiency in large-scale, dynamic network topologies.


Figure 8. Learning rate performance comparison across primary network metrics

In figure 8 shows that the horizontal bar chart illustrates the trade-offs when determining the best learning rate to use by distributed agents in a large-scale HetNet. The information shows that the lowest learning rate (0.001) yields the best Average Fairness (0.809) and the best Average Reward (1888.22), which means that slower updates in the policy enable the base station agents to approach a global optimum more efficiently. Although the Average Energy Efficiency does not change greatly in response to each of the tested values, the slight peak at a learning rate of 0.01 (3.05) indicates that there is a particular limit up to which the agents will maximally balance power consumption and user demand. In general, the stability in the performance, where average rewards are more than 1800, and Fairness is more than 0.77 across all instances, indicates the strengths and scalability of the decentralized EELBA-MARL model in the control of complex network interdependencies.

**Ablation Study Analysis**

In order to determine the role played by every element of the EELBA-MARL framework, an ablation study was carried out, the systematic ablation of certain words in the reward function, e.g., energy efficiency, load balancing, and interference awareness, will allow us to measure the effect of each of them on the overall network performance.

Table 4. Ablation study results for EELBA-MARL reward function components

| Model Variant | Avg Reward | Avg Energy Efficiency (EE) | Avg Fairness | Avg SINR |
|---|---|---|---|---|
| Full MARL | 1849.61 | 3.02 | 0.816 | 24.74 |
| No Energy Term | 1222.95 | 2.59 | 0.715 | 23.39 |
| No Load Balancing | 686.05 | 2.95 | 0.740 | 23.74 |
| No SINR | 1428.73 | 2.51 | 0.806 | 25.06 |

The results in table 4 of the ablation show that the integrated Full MARL (EELBA-MARL) approach has the most balanced and better performances with all metrics. The greatest drop in performance was observed in the No Load Balancing version, where the mean Reward decreased to 686.05, which is indicative that load control is the key to system stability in large-scale HetNets. Energy Term removal caused a significant decrease in the energy efficiency (2.59) and Fairness (0.715), which demonstrated that the promotion of power saving is key to the sustainable user association. Although the variant that did not employ the SINR concept in determining Fairness stayed at a fairly high rate, the decreased energy efficiency of the variant indicates that in the absence of the interference awareness, the agents are unable to allocate resources effectively. Finally, the results validate the hypothesis that the interaction of energy efficiency, load balancing, and interference control of the EELBA-MARL reward function is key to the attainment of optimal decentralized coordination.

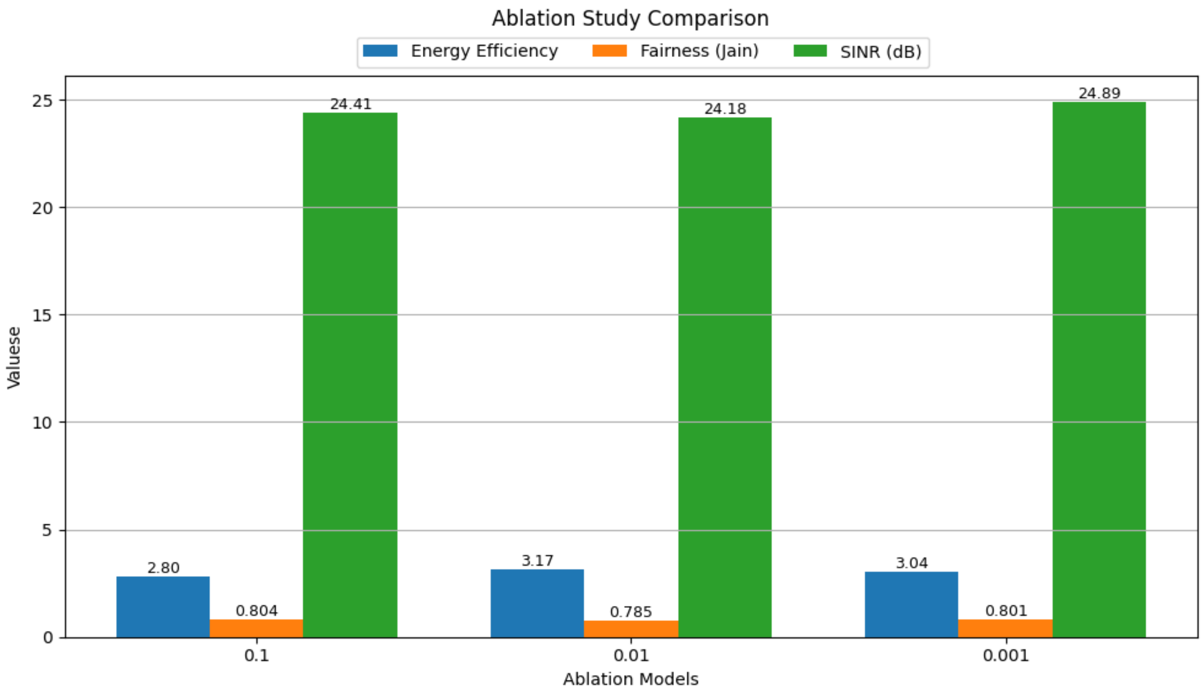


Figure 9. Performance comparison of ablation models across network metrics

The trade-offs between various ablation models in the EELBA-MARL framework are identified in the comparative analysis characterised in figure 9. The data shows that the model that is set with a 0.01 value has the highest Energy Efficiency of 3.17, whilst it compromises a little on the user Fairness (0.785) in order to concentrate on the power optimization. On the other hand, the 0.001 setup has the strongest signal quality with an Average SINR of 24.89 dB and high Fairness, which would make the Environment more stable to allow users to associate with it. In every variation, the SINR is always high (greater than 24 dB), in accordance with the research purpose of preserving the interference-conscious communication within large-scale HetNets. Such outcomes also confirm the scalability of the distributed MARL scheme, which also demonstrates that the agents can be optimized to focus on particular operational objectives, e.g., extreme energy efficiency or optimum signal quality, without affecting the overall stability of the network.

**Comparative Analysis of Energy Efficiency Across Algorithms**

In order to further contextualize the EELBA-MARL framework performance, table 5 compares the performance of this framework with one of the existing baseline algorithms, namely energy efficiency (EE) in the Macro and Pico layers.

Table 5. Comparative analysis of energy efficiency (b/J) across network algorithms

| Algorithm | Macro EE (b/J) | Pico EE (b/J) |
|---|---|---|
| IAW | 439.157 | 3610.29 |
| PF | 333.164 | 3214.7 |
| IRA | 331.592 | 2454.35 |
| **EELBA-MARL** | **462.54** | **3838.76** |

In table 5 shows that the proposed EELBA-MARL model is highly effective in minimizing energy consumption among heterogeneous network layers as compared to the traditional algorithms. EELBA-MARL has a better Macro EE of 462.54 b/J and a phenomenal Pico EE of 3838.76 b/J than the IAW and PF baselines. This large leap in comparison to the IRA algorithm, which had the least efficiency, demonstrates the benefit of distributed multi-agent reinforcement learning. The policy of EELBA-MARL is better at offloading traffic to low-power Pico access points without compromising high data rates because the base stations learn independent policies using local network conditions. The results confirm the novelty of the model in providing scalable and energy-efficient load balancing critical towards sustainable operation in 5G and beyond dense space [31].

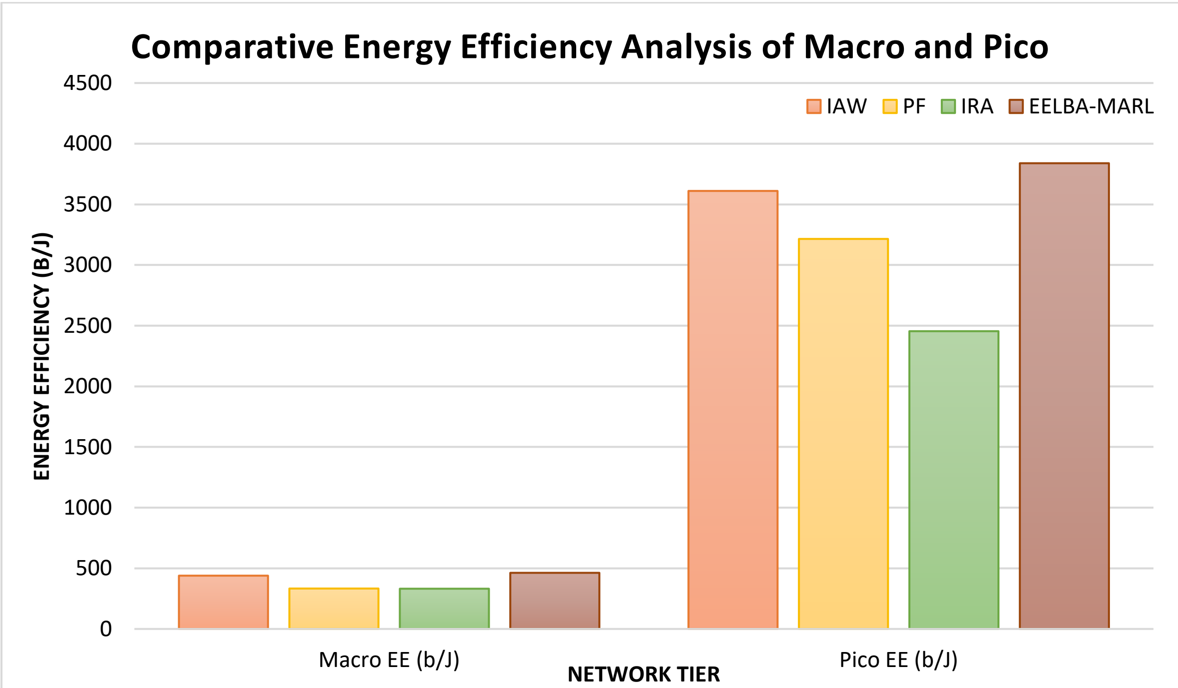


Figure 10. Comparative analysis of macro and pico energy efficiency across different algorithms

The relative performance of the EELBA-MARL framework versus the traditional methods of decentralized approaches in terms of performance is emphasized in figure 10. Though all the algorithms show an increased efficiency on the Pico cells since they have a low power profile, EELBA-MARL achieves the highest reported values on the Macro cells and in Pico cells, 462.54 b/J and 3838.76 b/J, respectively. This is a great enhancement of the IRA and PF algorithms that do not optimize user association in high-density environments. The fact that EELBA-MARL is better than its nearest opponent, the IIW algorithm, proves the research purpose, which is to apply distributed MARL to independently learn the optimal policies under local network conditions, such as interference and energy consumption. These findings are a solid empirical observation that the proposed model is especially adaptable to reach sustainable, high-performance operation in large-scale heterogeneous networks (HetNets) that rely on 5G and beyond.

## CONCLUSION

This study was able to design and test the EELBA-MARL (Energy Efficient Load Balancing and Interference Aware User Association in Large-Scale HetNets Using Distributed Multi- Agent Reinforcement Learning) framework to handle the vexing issues of scalability and sustainability in the contemporary cellular networks. The proposed model can allow individual base stations to become autonomous agents, learning optimal resource allocation and user association policies based on local network feedback, without necessarily involving centralized coordination which is expensive in terms of overhead (high overhead). The acquired experimental outcomes confirm the high functionality of EELBA-MARL in all the main goals. The convergence analysis testified that the agents are successful in balancing their learning behavior after 150 episodes, evolving out of a starting exploratory state into an optimized policy plateau. The framework had a near-perfect balance in network traffic distribution, with Fairness and Load Balance index obtaining values of 0.989. It is interesting to note that the system was highly performing with a total throughput of 4929.54 Mbps and a total power consumption of 7.04 W, which significantly reduced the environmental impact, leading to an overall Energy Efficiency of 734.76 b/J. Ablation and sensitivity studies also confirmed the power of the reward system. It degraded the performance considerably by removing the important words, like load balancing or energy efficiency, which proves that the synergy of these parameters is the key to the stability of the network in a decentralized way. Moreover, the comparative analysis revealed that EELBA-MARL is always superior to the traditional algorithms, such as IAW, PF, and IRA, especially in Pico Energy Efficiency (3838.76 b/J), as it easily offloads users to low-power access points. To work with it in the future, the framework can be expanded to more sophisticated cross-tier interference models and adaptive learning rates to take on extreme network volatility. Moreover, the use of federated learning will help to increase data privacy during collaboration between agents. On the whole, the EELBA-MARL framework offers a highly scalable, low-energy, and interference-aware framework that sets a new set of standards in the control of dense, large-scale 5G-and-beyond heterogeneous networks.

## REFERENCES

[1] Ramesh P, Bhuvaneswari PT, Dhanushree VS, Gokul G, Sahana S. User association-based load balancing using reinforcement learning in 5G heterogeneous networks: P. Ramesh et al. The Journal of Supercomputing. 2025 Jan;81(1):328. https://doi.org/10.1007/s11227-024-06788-1

[2] Hassan N, Fernando X. Interference mitigation and dynamic user association for load balancing in heterogeneous networks. IEEE Transactions on Vehicular Technology. 2019 May 29;68(8):7578-92. https://doi.org/10.1109/TVT.2019.2919812

[3] Mughees A, Tahir M, Sheikh MA, Amphawan A, Meng YK, Ahad A, Chamran K. Energy-efficient joint resource allocation in 5G HetNet using Multi-Agent Parameterized Deep Reinforcement learning. Physical Communication. 2023 Dec 1;61:102206. https://doi.org/10.1016/j.phycom.2023.102206

[4] Oo TZ, Tran NH, Saad W, Niyato D, Han Z, Hong CS. Offloading in HetNet: A coordination of interference mitigation, user association, and resource allocation. IEEE Transactions on Mobile Computing. 2016 Sep 27;16(8):2276-91. https://doi.org/10.1109/TMC.2016.2613864

[5] Rezvani A, Mirzaei A, Mikaeilvand N, Nouri-Moghaddam B, Gudakahriz SJ. A novel framework for enhancing data collection macro-strategies in heterogeneous iot networks using advanced mathematical modeling. Archives for Technical Sciences/Arhiv za Tehnicke Nauke. 2025 Jul 1(33). https://doi.org/10.70102/afts.2025.1833.001

[6] Gures E, Shayea I, Ergen M, Azmi MH, El-Saleh AA. Machine learning-based load balancing algorithms in future heterogeneous networks: A survey. IEEE Access. 2022 Mar 23;10:37689-717. https://doi.org/10.1109/ACCESS.2022.3161511

[7] Zia K, Javed N, Sial MN, Ahmed S, Pirzada AA, Pervez F. A distributed multi-agent RL-based autonomous spectrum allocation scheme in D2D enabled multi-tier HetNets. IEEE Access. 2019 Jan 1;7:6733-45. https://doi.org/10.1109/ACCESS.2018.2890210

[8] Xiao Y, Song Y, Liu J. Multi-agent deep reinforcement learning based resource allocation for ultra-reliable low-latency internet of controllable things. IEEE Transactions on Wireless Communications. 2023 Jan 10;22(8):5414-30. https://doi.org/10.1109/TWC.2022.3233853

[9] Hugh Q, Soria F. VoltSecure: A Secure Federated Learning Model for Decentralized Energy Management Systems. International Academic Journal of Innovative Research. 2025;12(3):33-42.

https://doi.org/10.71086/IAJIR/V12I3/IAJIR1223

[10] Hu Y, Chen M, Saad W, Poor HV, Cui S. Distributed multi-agent meta learning for trajectory design in wireless drone networks. IEEE Journal on Selected Areas in Communications. 2021 Jun 16;39(10):3177-92. https://doi.org/10.1109/JSAC.2021.3088689

[11] Kraemer L, Banerjee B. Multi-agent reinforcement learning as a rehearsal for decentralized planning. Neurocomputing. 2016 May 19;190:82-94. https://doi.org/10.1016/j.neucom.2016.01.031

[12] Bhattacharya S, Kailas S, Badyal S, Gil S, Bertsekas D. Multiagent reinforcement learning: Rollout and policy iteration for pomdp with application to multirobot problems. IEEE Transactions on Robotics. 2023 Dec 26;40:2003-23. https://doi.org/10.1109/TRO.2023.3347128

[13] Amato C, Konidaris G, Kaelbling LP, How JP. Modeling and planning with macro-actions in decentralized POMDPs. Journal of Artificial Intelligence Research. 2019 Mar 25;64:817-59. https://doi.org/10.1613/jair.1.11418

[14] Ahmed I, Syed MA, Maaruf M, Khalid M. Distributed computing in multi-agent systems: a survey of decentralized machine learning approaches. Computing. 2025 Jan 1;107(1):2. https://doi.org/10.1007/s00607-024-01356-0

[15] Tandon B, Thakur M. An Overview of Adaptive Signal Processing Methods for 6G Wireless Communication Networks. International Academic Journal of Science and Engineering. 2025;12(1):12-15. https://doi.org/10.71086/IAJSE/V12I1/IAJSE1203

[16] Hady MA, Hu S, Pratama M, Cao Z, Kowalczyk R. Multi-agent reinforcement learning for resources allocation optimization: a survey. Artificial Intelligence Review. 2025 Aug 27;58(11):354. https://doi.org/10.1007/s10462-025-11340-5

[17] Chen L, Wang Y, Miao Z, Feng M, Wang Y, Mo Y, He W, Wang H, Wang D. STR: Spatial-temporal RetNet for distributed multi-robot navigation. IEEE Transactions on Automation Science and Engineering. 2025 Jan 9;22:10429-41. https://doi.org/10.1109/TASE.2024.3524000

[18] Shi X, Li Y, Du C, Li H, Chen C, Gui W. Cooperative output regulation of heterogeneous directed multi-agent systems: A fully distributed model-free reinforcement learning framework. Science China Information Sciences. 2025 Feb;68(2):122202. https://doi.org/10.1007/s11432-024-4103-1

[19] Salami ZA, Bhaskaran B, Thusnavis Bella Mary I, Ugli ISS, Sripavithra CK, Kalidoss D. Energy-Efficient Architecture Design in Information Service Infrastructure. Indian Journal of Information Sources and Services. 2025;15(2):168-173. https://doi.org/10.51983/ijiss-2025.IJISS.15.2.23

[20] Deng X, Yang P, Lin H, Wang L, Lin S, Gui J, Chen X, Qian Y, Zhang J. Energy-efficient strategic AAV-enabled MEC networks via STAR-RIS: Joint optimization of trajectory and user association. IEEE Internet of Things Journal. 2025 Jan 8;12(10):14921-37. https://doi.org/10.1109/JIOT.2025.3527002

[21] Sharif S, Manzoor S, Ejaz W. Priority-based resource optimisation and user association in integrated networks. IET Networks. 2025 Jan;14(1):e12140. https://doi.org/10.1049/ntw2.12140

[22] Baek S, Choi S, Bahk S. DACS: User association and TDMA framing for low-latency services on integrated access and backhaul networks. IEEE Transactions on Vehicular Technology. 2025 Jan 3;74(5):8110-25. https://doi.org/10.1109/TVT.2025.3525841

[23] Mohammed Hussain V, Abdul Azeez Khan A, Javubar Sathick K, Arun Raj L. A personalized federated deep reinforcement learning framework with hierarchical control and edge aggregation for scalable real-time traffic prediction. Journal of Internet Services and Information Security. 2026;16(1):70-91. https://doi.org/10.58346/JISIS.2026.I1.005

[24] Maiwada UD, Abdalla GS, Singh NS, Salameh AA. Efficient energy utilization in LTE networks with intelligent handover mechanisms. Journal of Radiation Research and Applied Sciences. 2025 Sep 1;18(3):101599. https://doi.org/10.1016/j.jrras.2025.101599

[25] Chinipardaz M, Amraee S, Sarlak A. Joint downlink user association and interference avoidance with a load balancing approach in backhaul-constrained HetNets. Plos one. 2024 Mar 4;19(3):e0298352. https://doi.org/10.1371/journal.pone.0298352

[26] Ni W, Xiao H, Zhou M, Chronopoulos AT, Zhang Z. User association and small base station configuration for energy-efficiency maximization in hybrid-energy heterogeneous cellular networks. IEEE Internet of Things Journal. 2024 Dec 4;12(8):9858-71. https://doi.org/10.1109/JIOT.2024.3510785

[27] Sharma L, Liang JM, Wu SL. Discontinuous reception based energy-efficient user association for 5G heterogeneous networks. IEEE Access. 2024 Jan 23;12:13634-47. https://doi.org/10.1109/ACCESS.2024.3357663

[28] Cho YJ, Yoo HM, Kim KS, Na J, Hong EK. Practical load balancing algorithm for 5G small cell networks based on real-world 5G traffic and O-RAN architecture. IEEE Access. 2024 Aug 30;12:121947-57. https://doi.org/10.1109/ACCESS.2024.3452434

[29] Da H, Xiao F, Ma Z, Zhang Z, Guo Q. Joint design of routing selection and user association in multi-hop mmWave IABN: a multi-agent and double DQN framework. Telecommunication Systems. 2025 Mar;88(1):21. https://doi.org/10.1007/s11235-024-01256-w

[30] Pan C, Liu R, Yu G. Joint user association and resource allocation for mmWave communication: A neural network approach. Journal of Communications and Information Networks. 2021 Jun;6(2):125-33.

https://doi.org/10.23919/JCIN.2021.9475122

[31] Dobruna J, Fazliu Z, Humar I, Volk M. On optimizing resource allocation: a comparative review of resource allocation strategies in HetNets. Computer Modeling in Engineering & Sciences. 2025;142(3):2211-2245. https://doi.org/10.32604/cmes.2025.059541.