Journal of Information Systems Engineering and Management
2025, 10(18s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Machine Learning Approach for Optimization of Quality of Service in Self Organizing Heterogeneous Wireless Network

[1*]Dr. G. U. Patil, [2]Dr. A. D. Vishwakarma, [3]Dr. T. H. Jaware, [4]Dr. P. Subramanium

[1]Assistant Professor, Department of Electronics and Telecommunication Engineering, Hindi Seva Mandal's, Shri Sant Gadge Baba College of Engineering and Technology, Bhusawal, Maharashtra.*

[2]Associate Professor, Department of Artificial Intelligence and Data science Engineering, Godavari Foundation's, Godavari College of Engineering, Jalgaon, Maharashtra.

[3]Associate Professor, Department of Electronics and Telecommunication Engineering, S.E.S's, R. C. Patel Institute of Technology, Shirpur, Maharashtra.

[4]Assistant Professor, Department of Computer Science and Engineering, Hindi Seva Mandal's, Shri Sant Gadge Baba College of Engineering and Technology, Bhusawal, Maharashtra.

[1]Corresponding Author Email: gajanan.bsl@gmail.com*

ARTICLE INFO

Received: 18 Dec 2024

Revised: 30 Jan 2025

Accepted: 08 Feb 2025

ABSTRACT

Autonomous wireless networks (AWNs) face significant challenges due to interference arising from their diverse nature and high density. Employing machine learning as an effective data-driven approach presents a promising avenue for automatic power configuration and related settings. This study outlines an innovative technique for modeling a more densely populated network by incorporating femto or pico cells, simultaneously addressing power optimization challenges through a novel reward function in a distributed network. The methodology integrates particle swarm optimization to ensure optimal parameter selection within the reward function, thereby guaranteeing the fundamental Quality of Service (QoS) for microcell users with minimal power requirements. The proposed distributed power allocation method, rooted in Q-learning, undergoes evaluation against Markov decision states. Results demonstrate that the PSO-optimized solutions outperform the greedy algorithm in achieving superior outcomes.

Keywords: PSO, AWNs, Markov Decision Process, Q-learning, Greedy.

1. INTRODUCTION

Current cellular networks are denser and highly heterogeneous along with small femto and pico cells where Self-organization features become essential [1] - [4]. Self-configuration, self-optimization, and self-correction are the important features of autonomous networks. Majorly it can set up a newly installed base station (BS), managing resources, and managing network outages [5]. Hence the intention of autonomous network seeks to minimize human intervention when using network metrics to optimizing the cost of installation, configuring, and maintaining the network. Self-organizing networks, in fact, plays two main factors: intelligence based on history and autonomous adaptability according to conditions [2], [3]. Hence, machine learning techniques became a popular approach for processing of data and improvement in self-organizing networks [6] [7].

Within the realm of machine learning, Signal-to-Interference-Plus-Noise Ratio (SINR) matrices find application in constructing autonomous networks capable of adaptively enhancing prior performance metrics. Machine learning-based reinforcement learning (RL) emerges as a robust and extensively employed technique for dynamically regulating power in wireless systems. In this context, RL endeavors to augment the transmit power of base stations (BSs) to maximize overall network performance. Distinguishing itself from supervised learning approaches, RL operates without the need for predefined inputs and outputs during the learning phase. RL leverages network cooperation to derive its capabilities effectively [8].

Reinforcement learning has proven its efficacy across a spectrum of wireless communication applications, encompassing resource management [9] [10] [11] [12] [13] [14], energy capture [15], and flexible spectrum access [16] [17]. A comprehensive exploration of RL's application in wireless communication is presented by the authors of [18]. The Q-Learning model's inherent adaptability makes it particularly suitable for scenarios characterized by

continually changing network statistics [19]. Additionally, owing to its low computing complexity, Q-learning can be executed in a distributed manner using a base station [1]. Consequently, Q-learning facilitates enhanced computational efficiency, scalability, and fault tolerance within expansive networks. Nevertheless, formulating an effective reward function, crucial for fostering learning and mitigating the challenges of illusion or weaning, poses a non-trivial task [20].

Addressing this, the present research introduces a novel Q-learning strategy grounded in a reward function to optimize transmit power for individual base stations, ensuring varied allocation to minimize the overall network footprint. Key aspects underscored include:

- The incorporation of an optimized PSO-based reward function to enhance QoS for interactions between macro and femto cells and their respective base stations in a congested network.
- The development of a multi-agent Markov State Decision Process (MDP) and policy evaluation to further refine the strategy.

2. LITERATURE REVIEW

Various endeavors have been put forth to tackle Quality of Service (QoS) concerns in wireless communication, and publications [9]-[14] specifically present diverse reward function propositions for optimizing power allocation among femtocell base stations (FBS). These reward functions are tailored to enhance the performance of the FBS network, taking into consideration factors such as user satisfaction, energy efficiency, and network capacity. The primary objective is the holistic enhancement of QoS for network users.

In [9], the authors introduce an independent Q-learning methodology to regulate transmission power in a secondary base station. Leveraging Q-learning, a widely adopted reinforcement learning algorithm, this approach seeks to fine-tune power allocation among FBSs in the network. Additionally, the methodology incorporates the Signal-to-Interference-Plus-Noise Ratio (SINR) metric to uphold QoS standards for primary users. However, as noted, a noteworthy limitation of the [9] approach is its failure to account for the QoS of secondary users. Consequently, there exists a need for further exploration to encompass the QoS considerations of all users within the network.

Turning attention to [10], the authors propose a collaborative Q-learning strategy designed to maximize the overall communication rate for macrocell users while maintaining a predefined threshold. This approach considers the QoS of macrocell users by using Q-learning to optimize the power distribution among FBSs, with the goal of maintaining a certain threshold for the macrocell users' communication rate. Similarly, in [11], the authors propose a reward function that utilizes the proximity of FBSs and the aggregate FBS throughput. The goal is to achieve an unbiased distribution of power in the network. However, as you mentioned, this approach may have a limitation in that it does not take into account the lower thresholds for FBS rates, which could lead to a lack of support for FBSs due to increased network density and interference. Therefore, this approach may not be suitable for dense networks. It is worth mentioning that considering the trade-offs between different users and different network conditions while designing the reward function is very important, to achieve a balance between different QoS parameters.

The authors of [12] appear to have modeled the problem of interference management between microcells and femtocells as a non-cooperative scenario, which means that the cells are not working together and have competing interests. This approach may be used to understand and address issues of interference in communication networks that involve both microcells and femtocells. The authors of [13] seem to have suggested an approach with a spherical nature to enhance bandwidth for users situated at the periphery of a cell, while concurrently ensuring fairness between macrocell and femtocell users. The adoption of a spherical approach suggests that the authors might be leveraging geometric or spatial network characteristics to formulate their solution. The objective of this method is to augment the available bandwidth for users at the cell's edge, ensuring uniform service levels for all users irrespective of their connection to a specific cell type.

In [14], the authors seem to advocate the utilization of a reward function exhibiting exponential properties to minimize energy consumption in networks incorporating Long-term Evolutionary Femtocells (LTE). This strategy seeks to strike a balance among the femtocells in the network without necessitating a predefined equilibrium. The incorporation of a reward function with exponential properties suggests that the authors could be employing a

mathematical model sensitive to variations in specific variables, adaptable to diverse scenarios. The primary aim of this approach is to reduce energy consumption in the network, thereby enhancing overall efficiency and performance. Unlike previous studies in dense networks, which lacked the application of a reward function, this paper emphasizes the importance of establishing a lower limit for achievable femtocell levels. The authors also note that earlier studies only employed the reward function to constrain macrocell utilization to the required Quality of Service (QoS), neglecting other considerations. Such an approach may lead to interference issues as femtocell base stations (FBS) might employ additional power to enhance performance, negatively impacting neighboring femtocells and diminishing overall network efficiency.

This paper proposes a shift in focus towards congested networks, employing optimization algorithms to devise universal solutions and a comprehensive structure for portraying heterogeneous networks as a multi-agent Local Area Network (LAN). Additionally, it introduces a method for formulating reward functions aligned with network QoS requirements.

The paper's organization is structured as follows: the third section delineates the proposed methods, outlining the techniques and algorithms developed by the authors to address the identified problems. The fourth section provides an analysis of the results, which includes a detailed examination of the data and findings obtained from applying the proposed methods. Finally, the fifth section contains the conclusive remarks, which summarize the main contributions of the paper and provide an overall evaluation of the proposed methods.

3. PROPOSED METHODOLOGY

3.1 System Model:

The research introduces a novel framework for a cell within a diverse network, comprising a Macro Base Station (MBS) and a Femto Base Station (FBS). Each FBS caters to an individual user, referred to as Femto User Equipment (FUE), while the MBS addresses Macro User Equipment (MUE). Our emphasis lies in the distribution of power in congested and sporadic downlinks, where congestion introduces considerable disruptions. The assumption here is that network users operate within the same narrowband spectrum, and signaling is treated as subcarriers within the wideband multicarrier signal or its equivalent. The comprehensive configuration of the network is visually represented in Figure 1. It's noteworthy that although the presented approach showcases FBS and MBS servers as distinct users, it seamlessly lends itself to scenarios involving multiple served users.

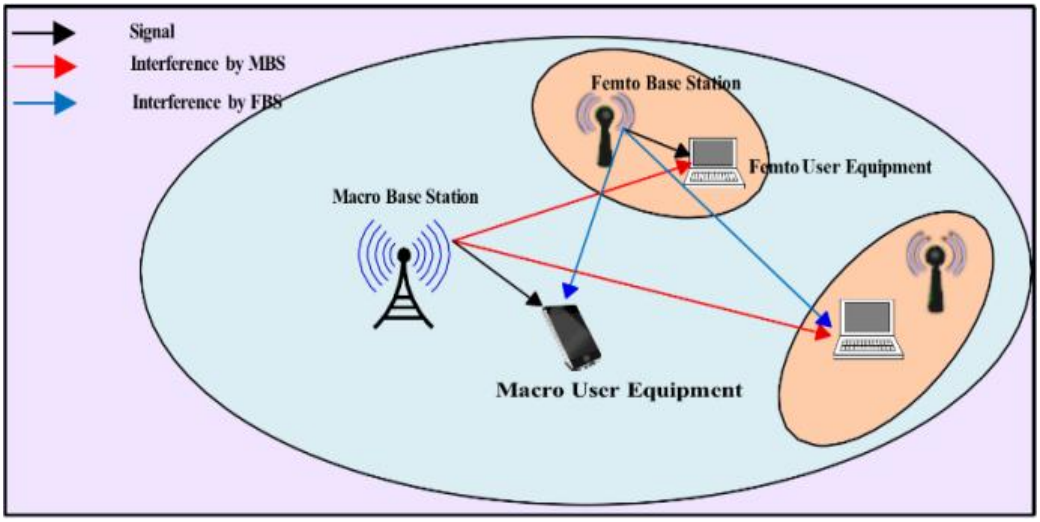

Figure 1: Femtocell Network [1]

Within a diverse mesh cell structure, a downlink configuration is employed, with a solitary macro base station (MBS) catering to multiple users distributed across distinct orthogonal subbands. Each user is assured a minimum average Signal-to-Interference-Plus-Noise Ratio (SINR) on their designated subband while receiving service from the MBS. Expanding beyond the macro base station (MBS), additional femto-cell base stations (FBS) are strategically positioned within the macro cell coverage area. Each FBS autonomously selects one of the subbands and manages communication for the corresponding femto user equipment (FUE). A set of FBSs are assumed to

operate globally across all subbands. Each FBS guarantees a constant mini-mum average SINR for the FUEs it serves. It is assumed that dense network uses a power distribution method in which power is concentrated in the downstream femtocell network to manage interference levels and ensure that users have the minimum required SINR. The proposed solution is focused on sub-bands and is only considering uplink. The overall configuration of the network is shown in Figure 1. In this scenario; the MUE is receiving a signal from an MBS with index 0. The signal also contains interference from an FBS with index, which is part of a set of FBSs, and thermal noise. The MUE is operating in a specific subband of the available spectrum, denoted as. The SINR experienced by the MUE will undergo alterations due to the existence of interference. This factor necessitates consideration when assessing the quality of the signal received by the MUE.

$$SINR_{MUE} = \frac{p_0|h_{0,0}|^2}{\underbrace{\sum_{k\in\mathcal{K}} p_k|h_{k,0}|^2}_{Femtocells' Interference} + N_0} \quad (1)$$

Let p_0 denote the power transmitted by the Macro Base Station (MBS), and $h_{0,0}$ represent the gain of the MBS for the channel to the Macro User Equipment (MUE). Furthermore, the power transmitted by the k^{th} Femto Base Station (FBS) is denoted as p_k, and the channel gain from the k^{th} FBS to the MUE is represented by $h_{k,0}$. Lastly, N_0 stands for the Additive White Gaussian Noise (AWGN) within the system.

3.2 Analyzing Policy and the Markov Decision Process:

The Markov Decision Process (MDP) stands as a mathematical framework employed for conceptualizing decision-making scenarios wherein an agent engages with an environment. An MDP encompasses a collection of states, a repertoire of actions, and a series of probabilistic transition dynamics elucidating how the agent's actions influence the likelihood of transitioning from one state to another. The overarching objective for the agent is to opt for actions that optimize a designated reward function.

In the realm of Markov Decision Processes (MDP), the policy of an agent is a function that establishes mappings from states to actions. This policy may adopt a deterministic form, wherein a specific action is chosen for each state, or a stochastic form, wherein a probability distribution over actions is determined for each state. The assessment of a policy within an MDP revolves around gauging the anticipated long-term reward that the agent is poised to receive while adhering to the specified policy. Techniques such as value iteration or policy iteration are instrumental in undertaking this evaluation. These methodologies entail iteratively estimating the value associated with each state and/or the action-value linked to each state-action pair under the policy under consideration.

Following the evaluation of a policy, comparisons can be made with other policies to discern the most optimal one for deployment in the MDP. This often involves contrasting the anticipated long-term rewards for each policy, with the policy yielding the highest expected reward being deemed the optimal choice for the MDP. A singular MDP agent encompasses the agent itself, an environment, a set of actions, and a set of states. The agent, guided by its policy, selects actions with probabilities contingent upon each state. As the agent engages with the environment, the state of the environment undergoes changes, and the agent accrues a reward for each executed action. The overarching objective for the agent remains the maximization of cumulative rewards over time.

In scenarios where multiple agents interact within the same environment, the setup is designated as a multi-agent MDP. Each agent in this context possesses its unique policy, with a focal aim to maximize individual rewards. Nevertheless, it is essential to note that the actions undertaken by one agent can exert influence on the rewards obtained by other agents. In such instances, effective coordination among agents becomes imperative to achieve an optimal collective outcome. The representation of agents within a multi-agent MDP involves a set of K agents, denoted as $\mathcal{K}$. The structure of the multi-agent MDP is defined by a tuple $(\mathcal{A}, \mathcal{X}, P_r, \mathcal{R})$, where $\mathcal{A}$ denotes the set of actions, $\mathcal{A}$ stands for the set of states, P_r signifies the set of transition dynamics, and $\mathcal{R}$ represents the set of rewards [21]. Agents collaboratively select actions to traverse from one state of the model to the next, with the overarching objective of maximizing the rewards collectively provided by all agents. The evaluation of a policy in this multi-agent MDP context incorporates the utilization of discounted future rewards.

Set $\mathcal{A}$ encompasses all conceivable actions within the purview of agents. Each agent, denoted as k, selects an action from its designated set, $\mathcal{A}_k$. The collective set of shared actions, denoted as $\mathcal{A}$, is derived as the Cartesian product of individual action sets, spanning $\mathcal{A}_1$ to $\mathcal{A}_K$. This implies that an action, denoted as 'a,' attains the status of being

common when it can be articulated as a composite outcome drawing from each unique set of individual actions.

The configuration of the system state hinges on an array of variables denoted as X, with each individual variable denoted as X_i, where the index i spans from 1 to n. The comprehensive set of states is articulated as $\mathcal{X} = \{X_1, X_2, \dots, X_n\}$, and the depiction of a distinct state is encapsulated by $x \in \mathcal{X}$. It's crucial to note that each variable, X_i, is affiliated with a designated network component within the overarching system.

P_r signifies the probability associated with the transition from the present state x to the subsequent state x' through the execution of a specific operation denoted as a. Its influence extends to the inter-agent communication within the system, dictating the probabilities linked to diverse actions resulting in distinct states.

$\mathcal{R}$ denotes a reward function designed to allocate a numerical value to the combination of state x and action a, portraying the degree of desirability associated with occupying that state and executing that particular action. In the realm of reinforcement learning, this reward function assumes a pivotal role in steering the decision-making endeavors of the agent.

In the context of a strategic scenario, denoted as $\pi: \mathcal{X} \to \mathcal{A}$, a function is established for mapping states, x, to corresponding actions, a. This implies that for any given state x, the representation $\pi(x)$ encapsulates the action to be executed. The anticipation of $\pi(x)$'s trajectory is facilitated through the introduction of value functions, namely $V_\pi(x)$ and $Q_\pi(x, a)$. The function $V_\pi(x)$ encapsulates the anticipated cumulative reward originating from state x, presuming the agent adheres to the strategy outlined by π. On the other hand, the action-value function $Q_\pi(x, a)$ encapsulates the anticipated cumulative reward stemming from state x when action a is undertaken while following the strategy π. The governing directive for the scenario where $x' \in \mathcal{X}$ is articulated in [21]:

$$V_\pi(x') = \mathbb{E}_\pi\left\{\sum_{t=0}^{\infty} \beta^t \mathcal{R}^{(t+1)} \middle| x^{(0)} = x'\right\} \quad (2)$$

Incorporating a scale factor denoted as β, which spans the range from 0 to 1, becomes instrumental in harmonizing the relative significance assigned to immediate and future rewards. Here, $\mathcal{R}^{(t+1)}$ signifies the reward at time $(t + 1)$, and $x^{(0)}$ denotes the initial state. The action-value function $Q_\pi(x, a)$ assumes a pivotal role, delineating the expected cumulative reward attributed to the execution of the joint action (a, y) in the ensuing iteration, as elucidated in reference [21].

$$Q_\pi(x, a) = \mathcal{R}(x, a) + \beta \sum_{x' \in \mathcal{X}} P_r(x'|x, a) V_\pi(x') \quad (3)$$

3.3 Factored Markov Decision Process (FMDP):

An instance of a Factored Markov Decision Process (FMDP) is categorized under the broader framework of Markov Decision Processes (MDP). What sets the FMDP apart is its unique portrayal of state and action spaces, which involves a combination of numerous smaller, independent components or factors. This distinctive modeling approach of the FMDP introduces a streamlined representation for the MDP. It achieves this by segmenting the overarching problem into more manageable subproblems, allowing for independent resolution, which can then be amalgamated to derive the comprehensive solution. Additionally, the factorization of the MDP allows for the use of more efficient methods to solve the MDP, such as using value function factorization techniques or exploiting the problem's underlying structure. This makes FMDPs a useful tool in decision-making problems with large state spaces, where traditional MDPs may become intractable. The reward function in an FMDP is also factorized [21]:

$$\mathcal{R}(x, a) = \sum_{k \in \mathcal{K}} \mathcal{R}_k(x_k, a_k) \quad (4)$$

In this context, the notation $\mathcal{R}_k(x_k, a_k)$ is employed to denote the reward function pertinent to the k^{th} agent. The symbol $\mathcal{R}_k(x_k, a_k)$ encapsulates the reward function specific to the k^{th} agent within the system. This function operates with two inputs: x_k, indicative of the state of the k^{th} agent, and a_k, representing the action executed by the k^{th} agent. The resultant output of the function corresponds to the reward or value associated with the particular state-action pairing. Employed as a metric, this function serves the purpose of assessing the performance exhibited by the k^{th} agent within the given environmental context.

3.4 Femtocell Network Markov Decision Process:

A Femtocell Network Markov Decision Process (MDP) serves as a mathematical framework designed to encapsulate decision-making processes within a network of femtocells. Femtocells, characterized as diminutive, low-power

cellular base stations, find application in residential or commercial spaces to enhance wireless coverage. The MDP framework operates by considering the system's distinct state at each time step, with the overarching objective of identifying optimal actions within each state. These actions are strategically chosen to maximize a specific objective, be it network throughput or energy efficiency. The system's state encompasses crucial details pertaining to user locations, channel conditions, and femtocell power levels. On the other hand, the network operator possesses a range of actions at their disposal, including adjusting femtocell power levels, resource allocation to users, and fine-tuning the cell-edge data rate. Leveraging the MDP framework facilitates the determination of an optimal policy for governing the femtocell network, thereby enabling real-time decision-making regarding resource allocation and interference management.

A Femtocell network typically includes the following elements:

• Femtocells: These are small, low-power cellular base stations that can be installed in homes or businesses to improve wireless coverage. They use the same technology as macrocell base stations, but have a smaller coverage area and lower power output.

• Access Points (APs): These are devices that connect the femtocells to the internet and allow users to access the cellular network.

• Subscriber devices: These are the mobile devices, such as smartphones or tablets that connect to the femtocells to access the cellular network.

• Backhaul: This is the connection that links the femtocells to the macrocell network, allowing them to communicate with the core network and the internet.

• Management System: This is the system that manages the femtocells and controls their configuration, operation, and maintenance.

• Security Measures: Femtocell network have security measures to prevent unauthorized access to the network and protect users' privacy.

• Resource Management: Resource management strategies such as power control, handover, and interference management are implemented to optimize the performance of the femtocell network.

• Mobility Management: It enables the smooth handover of the mobile device between different femtocells and macrocells.

• QoS: it ensures that the services provided by the femtocell network are reliable and meet the requirements of different types of applications and users.

• Environment: The environment in a Femtocell network refers to the physical and wireless conditions of the network, such as the location of users, channel conditions, and interference levels. It also includes the actions of other agents in the network.

• Agent: The agent in a Femtocell network is the decision maker, such as the network operator, who controls the configuration and operation of the femtocells.

• Set of Actions($\mathcal{A}_K$): Within a Femtocell network, the array of actions at the disposal of the agent encompasses the ability to modify femtocell power levels, allocate resources to diverse users, and fine-tune the cell-edge data rate.

• State Set($\mathcal{X}_k$): In a Femtocell network, the collection of states encompasses details regarding the present condition of the network. This incorporates information about user locations, channel conditions, and femtocell power levels. The agent utilizes this information as a basis for decision-making, influencing the control of femtocells and resource allocation.

3.5 Q-Learning Approach for Power Allocation:

When considering power allocation within wireless communication systems, Q-learning emerges as a widely adopted strategy for identifying the most effective power allocation strategy. Comparable to its application in other optimization scenarios, Q-learning conventionally relies on a consistent learning rate. Nevertheless, investigations have indicated that introducing a diminishing learning rate over a finite number of iterations can enhance the efficacy of the Q-learning algorithm. A detailed account of this approach is outlined in reference [14], with specific

learning rates referenced in [21]. This approach can be applied to various wireless communication scenarios including cooperative communications, cognitive radio networks, and cellular networks.

$$\alpha^{(t)}(x,a) = \frac{1}{[1+t(x,a)]} \tag{5}$$

In the realm of reinforcement learning, the notation $t(x,a)$ is a commonly employed symbol denoting the frequency with which the agent has encountered the state-action pair (x,a) before the present time step denoted by t. This particular information serves a pivotal role across diverse algorithms, influencing the adjustment of the agent's policy and value estimations.

3.6 Reward Function:

The suggested scheme for the reward function holds paramount significance in delineating the objectives of the evolving FBS. A methodical strategy for formulating the reward function, grounded in the inherent characteristics of the optimization algorithm, proves instrumental in directing agents toward the intended behavior or outcome. Furthermore, incorporating modifications to the agents' conduct throughout the learning trajectory offers an additional avenue to guide them toward the targeted actions or states, as detailed in reference [21].

$$Q(z',a) \leftarrow Q(z',a) + a^{(t)}(x',a)\left[\mathcal{R}(z',a) + \beta \max_{a'} Q(z'',a') - Q(z',a)\right] \leftarrow a^{(t)}(z',a)\left[f(\cdot)\beta \max_{a'} Q(z'',a')\right] + \underbrace{a^{(t)}(z',a)C}_{A} \tag{6}$$

As articulated in equation (6), the depiction of state transition delineates the shift from the current state (x′) to the subsequent state (x′′). The Q value corresponding to state x′ undergoes degradation due to the influence of the term (A). If the magnitude of (A) exceeds 0, it exhibits an inverse relationship with the reduction in state x′. This signifies that an escalation in the value of (A) leads to an intensified decline in the value of state x′.

The reward function governing the k^{th} FBS, denoted as R_k, takes the form of a differentiable and continuous function. This function operates with four inputs, namely $(r_0, r_k, \Gamma_0, \Gamma_k)$, and translates them into a numerical value within the set of real numbers ($\mathbb{R}$). The function is defined as per the reference [21], and the integers k_1 and k_2 are also defined in the same reference [21].

$$R_k(r_0, r_k, \Gamma_0, \Gamma_k) = [r_0 - \log_2(1+\Gamma_0)]^{k_1} + [r_k - \log_2(1+\Gamma_k)]^{k_2} \tag{7}$$

The declaration suggests that when the rate of the FBS or MUE surpasses a specified minimum threshold, opting for an action that elevates the rate results in a decrement in the premium. This characteristic operates counter to the objective of augmenting the collective network bandwidth.

3.7 Fitness Function for Optimization:

The present study employs PSO to refine the parameters of the learning rate (α) and discount factor (β) within the constructed model. A fitness function plays a pivotal role in assessing the efficacy of various pairings of α and β. The utilization of PSO facilitates the identification of optimal values that yield the most favorable outcomes based on the criteria defined by this function.

$$F = Optmize(\alpha, \beta) \tag{8}$$

3.8 Particle Swarm Optimization (PSO):

Particle Swarm Optimization stands as a population-centric optimization algorithm drawing inspiration from the collective behavior observed in bird flocks or fish schools. It employs a cluster of "particles" navigating through the search space, dynamically adjusting their positions influenced by both individual experiences and the collective wisdom gleaned from other particles within the swarm. It has been used in various research fields and subsequent applications. PSO considers the solutions in the search space as particles, each particle moves best past p_{best}position and g_{best} global position with variable speed according to the following equation [22] [23]:

$$v_{id}(t) = v_{id}(t-1) + c_1 r_1\big(p_{best}(t-1) - x_{id}(t-1)\big) + c_2 r_2\big(g_{best}(t-1) - x_{id}(t-1)\big) \tag{9}$$

$$x_{id}(t) = x_{id}(t-1) + v_{id}(t) \tag{10}$$

The speed increases as the particle moves closer to $p_{best}$and p_{best}. PSO has proven its efficiency in dealing with

different problems [23]. Figure 2

Reset swarm parameters and S dimension;
Initialize the velocities and random positions of the particles in each dimension of the search space;
Initialization of random velocity and position of particles in all dimensions of the search space;
$p_{best} = x_{id}$ for each particle;
Determine the $f(x_{id})$ of each particle;
Determine the g_{best}; // the best p_{best}
As long as (the stop condition is not checked) **do**
For (i in the range from 1 to S) **do**
Determine the novel velocity utilizing the equation (9);
Discover the novel position utilizing the equation (10);
Estimate $f(x_{id})$ of each particle;
If ($f(x_{id})$ is superior than $f(p_{best})$) **then**
$p_{bestid} = x_{id}$;
If ($f(p_{best})$is better than $f(g_{best})$) **then**
$g_{best} = p_{best}$;
End For
End as long as
Show the best solution found g_{bestd};
End

Figure 2: PSO algorithm

4. RESULTS AND DISCUSSION

In this segment, we present the outcomes of the Q-learning approach implemented with PSO optimization. To commence, let's delve into the intricacies of modeling and the associated variables. A simulation of a femtocell system unfolded with a MBS, five MUEs, and multiple FBSs, each catering to a singular FUE (refer to Figure 3). The operational assumption entails FBS and MBS functioning with an equivalent channel bandwidth at a frequency of f = 2.4 GHz. The path loss model governing the MUE-to-MBS and FBS-to-FUE connections is delineated by the following expression:

$$\text{pathloss} = \text{constantpathloss} + 10\text{nlog}_{10}\left(\frac{d}{d_0}\right) \quad (11)$$

The defined parameters are as follows: $d_0 = 10m$, $PL_0 = 76.9\ dB$, and $n = 6$.

Path loss refers to the reduction in power density (attenuation) experienced by an electromagnetic wave as it traverses through space. The study employs an internal interaction model—a mathematical representation elucidating the dynamics among various components of the network, namely, the MUE, FBS, and FUE. This internal interaction model captures the nuances of path loss in the connectivity between each MUE and FBS, as well as the interconnection between each FBS and the FUE of another FBS. This modeling approach is instrumental in comprehending the wireless network's performance and behavior, paving the way for the identification of potential enhancements. Table 1.

Table 1: Parameter table

Parameter	Value
Power (Pmin)	-30dBm
Power (Pmax)	35dBm
Step size	2.5 dBm
N_{power}	11

The simulation results are visually presented in Figure 3, delineating the spatial states of the FBSs concerning the MUE. Figure 4 provides a comparative visualization, illustrating the minimum, maximum, and mean transmit power across varying numbers of FBSs. Employing the PSO-Q-Learning algorithm for transmit power optimization unfolds in Figure 5, spotlighting the noteworthy achievement of the proposed PSO-optimized Q-Learning approach. Evidently, this approach surpasses alternative methodologies by attaining the lowest power consumption, showcasing its superior performance. The overarching objective of this study is to refine power consumption dynamics and elevate the operational efficiency of the proposed femtocell network.

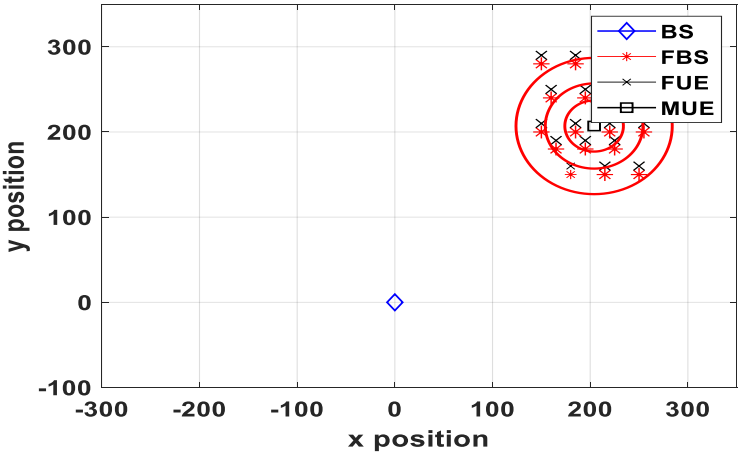

Figure 3: Conceptualization of the proposed system model

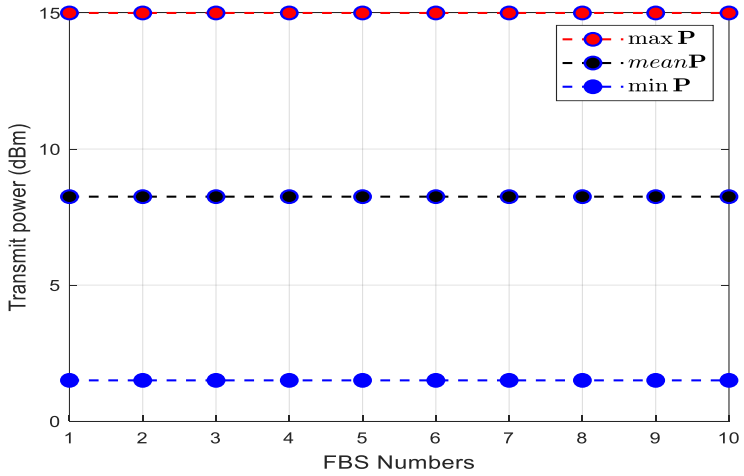

Figure 4: Representation of the minimum, maximum, and mean transmit power for different number of FBSs

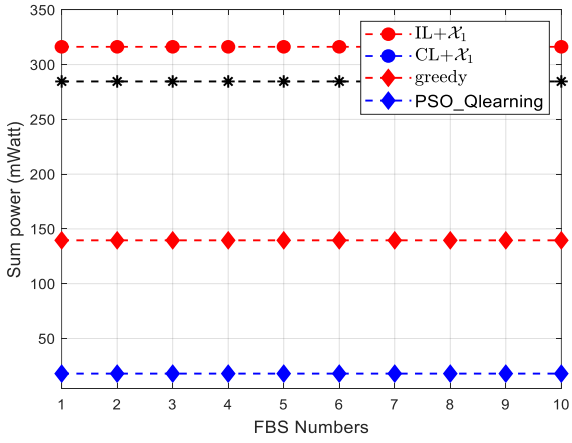

Figure 5: Performance evaluation for the aggregate transmit power of FBSs across diverse techniques

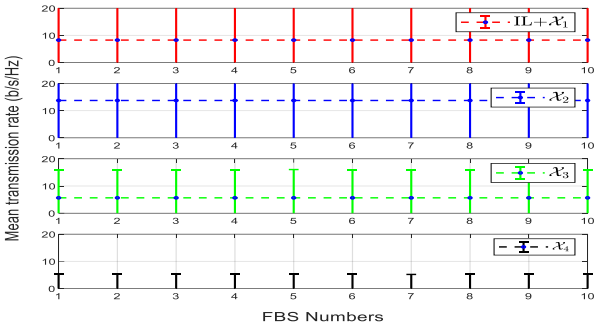

Figure 6: Mean transmission rate performance comparison

The simulation encompasses the examination of two distinct state sets, denoted as X_1 and X_2. The set $X_1 = \{\mathcal{X}_1, \mathcal{X}_3, \mathcal{X}_4\}$ conveys the state information of FUE to the FBS, while the set $X_2 = \{\mathcal{X}_2, \mathcal{X}_3, \mathcal{X}_4\}$ conveys the state information of Macro User Equipment (MUE) to the FBS. These states encapsulate details regarding the relative positioning of the FBS concerning both the MBS and MUE, respectively. The simulation results, shown in Figure 6, indicate that when the FBS is in the X_2 state, it utilizes the maximum power available for transmission. This results in the least troublesome for MUE and the lowest bit rate. The X_2 state set also provides the FBS with information about the MUE's QoS status, which leads to better Interference-Limited (IL) performance than the X_1 state set, as seen in Figure 6a.

The outcomes of the simulation, as depicted in Figures 7 and 8, indicate the algorithm's adeptness in judiciously allocating resources for FUEs, validating its efficacy. Furthermore, the results underscore the enhanced capability and superiority of the proposed PSO-Q-Learning algorithm compared to alternative methodologies.

To substantiate this claim, a comparative analysis was conducted with a prior study conducted by Amiri et al., (2019) [21]. The findings reveal that the proposed PSO-based Q-Learning approach outperforms the aforementioned research in both power consumption efficiency and overall performance metrics.

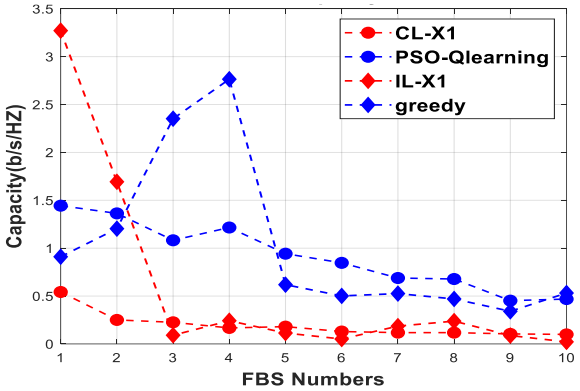

Figure 7: Comparative graph illustrating the capacity assessment for FUEs

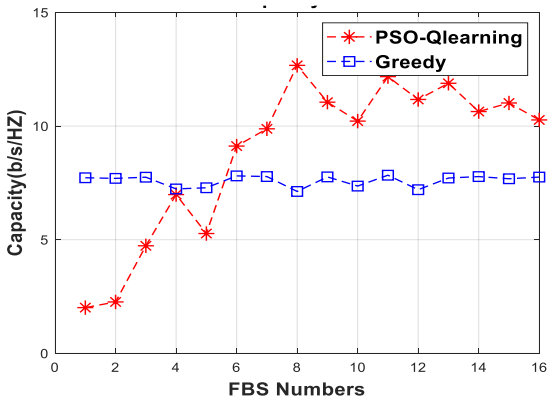

Figure 8: Graph comparing the cumulative capacity of FUEs

4.1 Comparative Analysis with Previous Research Works:

This research work is based on various reference papers which we used as the base or reference papers for research work. The comparative analysis for the results of base papers and developed work is presented in Figures. Moreover, a tabular comparison is also presented for the same for comparative analysis purpose. Figure 8 and Table 2.

Table 2: Results for FUEs Capacity (b/s/Hz)

FBS No	PSO-Q Learning	IL-X1	Greedy	CL-X1
1	**1.5**	3.5	0.9	0.5
2	**1.4**	1.7	1.3	0.3
3	**1.1**	0.1	2.4	0.2
4	**1.3**	0.2	2.7	0.2
5	**0.9**	0.1	0.6	0.1
6	**0.8**	0.1	0.5	0.1
7	**0.7**	0.1	0.5	0.2
8	**0.6**	0.3	0.5	0.2
9	**0.5**	0.1	0.4	0.1
10	**0.5**	0.1	0.5	0.1

From the above comparative results for FUEs, it is clearly shows that the FUEs capacity gets increased by PSO Q-learning technique as compared to individual learning, greedy and cooperative learning technique as we go from femto base station 1 to 10. Hence we can say that the proposed PSO Q-learning technique is much better than the individual learning, greedy and cooperative learning technique.

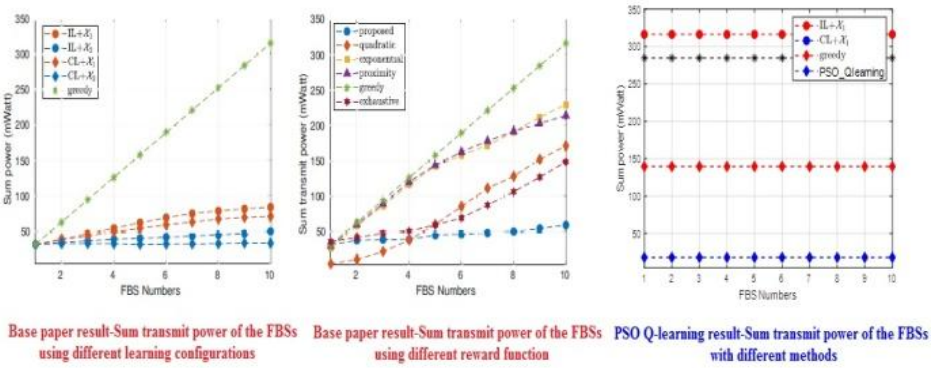

Figure 9: Comparative analysis of sum transmit power of the FBSs

In above Figures 9, the first two graphs represent the result achieved by the method proposed by Amiri et al., [89]. Also, the third graph shows the simulation outcome of proposed PSO based Q-learning method. After analyzing the comparative graphs, it is clear that the proposed method outperforms previous research work on the basis of sum transmit power.

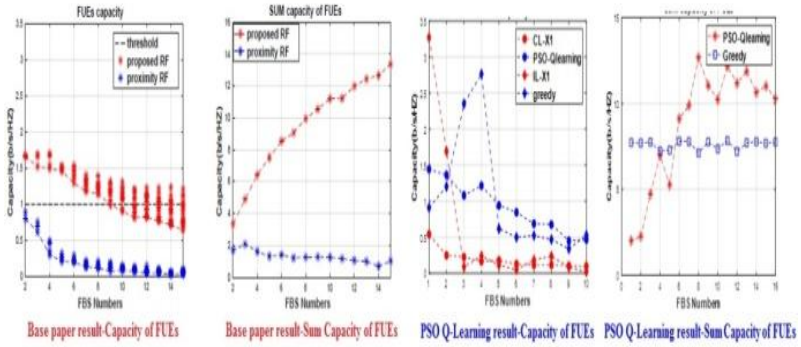

Figure 10: Comparative analysis of capacity and sum capacity of FUEs

In above Figures 10, the first two graphs represent the result achieved by the method proposed by Amiri et al., [2]. Also, the third and fourth graphs show the simulation outcomes of proposed PSO based Q-learning method. After analyzing the comparative graphs, it is clear that the proposed method outperforms previous research work on the basis of capacity and sum capacity of FUEs. Table 3

Table 3: Comparative Analysis with Previous Research Works

Sr. No.	Paper	Parameter	Comparative Analysis
1	Amiri et al., [21]	Sum Transmit Power of the FBSs	The PSO Q-learning approach is more efficient with low power consumption because it uses the lowest power.
2	Amiri et al., [1]	Capacity and Sum Capacity of FUEs	The capacity of FUEs exhibits a noteworthy uniformity across diverse positions, underscoring the algorithm's equitable resource distribution. The cumulative capacity of the network displays a progressive inclination for all configurations of FBSs, consistently surpassing the capacity observed in the foundational research approach.

5. CONCLUSION

This article introduces a two-tier femtocell network featuring a learning framework that employs a decentralized PSO-based strategy. The primary objective is to optimize transmit power, catering to service users while mitigating interference issues arising from multiple cells. The novelty lies in its adaptability, enabling the seamless integration of new femtocells into the network. Simultaneously, existing femtocells undergo a learning process to dynamically adjust transmit power, ensuring optimal support for service users and preventing interference from multiple cells. The versatility of the proposed approach extends its applicability to the facilitation of self-organizing networks (SON) through machine learning for efficient femtocell network management. Moreover, it serves as a versatile testbed for evaluating the efficacy of diverse pedagogical approaches, including reward functions, Markov state models, and learning outcomes. Beyond femtocell networks, this proposed structure can find utility in other limited interference networks, such as cognitive radio frameworks.

DATA AVAILABILITY STATEMENT

All the data is collected from the simulation reports of the software and tools used by the authors. Authors are working on implementing the same using real world data with appropriate permissions.

FUNDING

No fund received for this project

CONFLICTS OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

[1] Amiri, R., Mehrpouyan, H., Fridman, L., Mallik, R. K., Nallanathan, A. and Matolak, D., 2018, May. A machine learning approach for power allocation in HetNets considering QoS. In 2018 IEEE International Conference on Communications (ICC) (pp. 1-7). IEEE.

[2] O. G. Aliu, A. Imran, M. A. Imran, and B. Evans, “A survey of self organization in future cellular networks,” IEEE Commun. Surv. Tutor., vol. 15, no. 1, pp. 336–361, First Quarter 2013.

[3] J. Moysenand L. Giupponi, “From 4G to 5G: Self-organized network management meets machine learning,” CoRR, vol. abs/1707.09300, 2017. [Online]. Available: http://arxiv.org/abs/1707.09300

[4] M. Agiwal, A. Roy, and N. Saxena, “Next generation 5G wireless net-works: A comprehensive survey,” IEEE Commun. Surv. Tutor., vol. 18, no. 3, pp. 1617–1655, Third quarter 2016.

[5] P. V. Klaine, M. A. Imran, O. Onireti, and R. D. Souza, “A survey of machine learning techniques applied to self-organizing cellular networks,” IEEE Commun. Surv. Tutor., vol. 19, no. 4, pp. 2392–2431, Fourth quarter 2017.
[6] A. Imran, A. Zoha, and A. Abu-Dayya, “Challenges in 5G: how to empower SON with big data for enabling 5G,” IEEE Network, vol. 28, no. 6, pp. 27–33, Nov 2014.
[7] R. Li, Z. Zhao, X. Zhou, G. Ding, Y. Chen, Z. Wang, and H. Zhang, “Intelligent 5G: When cellular networks meet artificial intelligence,” IEEE Wirel. Commun., vol. 24, no. 5, pp. 175–183, Oct 2017.
[8] R. S. Sutton and A. G. Barto, Introduction to Reinforcement Learning, 1st ed. Cam-bridge, MA, USA: MIT Press, 1998.
[9] A. Galindo-Serrano and L. Giupponi, “Distributed Q-learning for aggregated interference control in cognitive radio networks,” IEEE Trans. Veh. Technol., vol. 59, no. 4, pp. 1823–1834, May 2010.
[10] H. Saad, A. Mohamed, and T. El Batt, “Distributed cooperative Q-learning for power allocation in cognitive femto cell networks,” in Proc. IEEE Veh. Technol. Conf., pp. 1–5, Sep 2012.
[11] J. R. Tefftand N. J. Kirsch, “A proximity-based Q-learning reward function for femto cell networks,” in Proc. IEEE Veh. Technol. Conf., pp. 1–5, Sep 2013.
[12] M. Bennis, S. M. Perlaza, P. Blasco, Z. Han, and H. V. Poor, “Self-organization in small cell networks: A reinforcement learning approach,” IEEE Trans. Wireless Commun., vol. 12, no. 7, pp. 3202-3212, Jul 2013.
[13] B. Wen, Z. Gao, L. Huang, Y. Tang, and H. Cai, “A Q-learning-based downlink re-source scheduling method for capacity optimization in LTE femto cells,” in Proc. IEEE. Int. Comp. Sci. and Edu., pp. 625-628, Aug 2014.
[14] Z. Gao, B. Wen, L. Huang, C. Chen, and Z. Su, “Q-learning-based power control for LTE enterprise femto cell networks,” IEEE Syst. J., vol. 11, no. 4, pp. 2699-2707, Dec 2017.
[15] M. Miozzo, L. Giupponi, M. Rossi, and P. Dini, “Distributed Q-learning for energy harvesting heterogeneous networks,” in Proc. IEEE. ICCW, pp. 2006–2011, Jun 2015.
[16] B. Hamdaoui, P. Venkatraman, and M. Guizani, “Opportunistic exploitation of band-width resources through reinforcement learning,” in Proc. IEEE GLOBE COM, pp. 1–6, Nov 2009.
[17] G. Alnwaimi, S. Vahid, and K. Moessner, “Dynamic heterogeneous learning games for opportunistic access in LTE-based macro/femto cell deployments,” IEEE Trans. Wireless Commun., vol. 14, no. 4,pp. 2294–2308, Apr 2015.
[18] K. L. A. Yau, P. Komisarczuk, and P. D. Teal, “Reinforcement learning for context awareness and intelligence in wireless networks: Review, new features and open is-sues,” J. Netw. Comput. Appli., vol. 35, no. 1, pp. 253–267, Jan 2012.
[19] C. J. C. H. Watkins and P. Dayan, “Q-learning,” Machine Learning, vol. 8, no. 3, pp. 279–292, May 1992.
[20] L. Matignon, G. J. Laurent, and N. Le Fort Piat, “Reward function and initial values: Better choices for accelerated goal-directed reinforcement learning,” in Proc. ICANN, pp. 840–849, 2006.
[21] Amiri, R., Almasi, M. A., Andrews, J. G. and Mehrpouyan, H., 2019. Reinforcement learning for self-organization and power control of two-tier heterogeneous networks. IEEE Transactions on Wireless Communications, 18(8), pp. 3933-3947.
[22] Kalpana, N., Gai, H. K., Kumar, A. R. and Sathya, V., 2021. Optimal resource allocation based on particle swarm optimization. Advances in Communications, Signal Processing, and VLSI: Select Proceedings of IC2SV 2019, p.199.
[23] Bansal, J.C., 2019. Particle swarm optimization. In Evolutionary and swarm intelligence algorithms (pp. 11-23). Springer, Cham.