I.J. Wireless and Microwave Technologies, 2026, 4, 275-290
Published Online on August 8, 2026 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijwmt.2026.04.15

Federated and Communication-Efficient Decentralized Meta-Reinforcement Learning for Dynamic Spectrum Access in Cognitive Radio–Enabled 5G IoT Networks

Jayesh Kumar Dabi*
Department of Electronics & Communication Engineering, SVCE Indore, RGPV, India
E-mail: jayeshdabi@svceindore.ac.in
ORCID iD: https://orcid.org/0009-0005-9453-5704
*Corresponding Author

Priyadarshi Ashok Dahat
Department of Electronics &Telecommunication Engineering, IET, DAVV Indore, India
E-mail: pdahat@ietdavv.edu.in
ORCID iD: https://orcid.org/0000-0002-0049-2522

Received: April 3, 2026; Revised: May 25, 2026; Accepted: July 14, 2026; Published: August 8, 2026

Abstract: Dynamic spectrum access (DSA) in 5G IoT setups with cognitive radio is characterized by rapid and decentralized decision-making processes in highly non-stationary wireless environments, limited communication needs, and restrictive bounds. In this work, we present F-DMRL, a federated, communication-efficient decentralized meta-reinforcement learning framework for allowing a massive number of IoT devices to meta-learn collectively about spectrum-access strategies in a decentralized way without centralized control and without an extensive amount of inter-agent communication. Our method incorporates lightweight federated meta-parameter aggregation with gradient sparsification and periodic communication, allowing devices to only compress the meta-updates during this process and then adapt locally for task-specificity. We have presented analytical speedup guarantees and upper bounds on communication cost under bounded environmental drift and shown that using the approach proposed here, F-DMRL preserves convergence properties while posing a large reduction in coordination overhead at the same time. Simulations across various 5G IoT spectrum environments showed that F-DMRL performed faster adaptation (up to 45% fewer episodes), higher spectral efficiency, and lower interference probability compared to centralized meta-RL, federated DRL, and traditional decentralized RL baselines. Simulation results averaged across 10 independent runs demonstrate improvements of 45% faster adaptation and 60–80% lower communication overhead relative to baseline methods, while maintaining stable convergence.

Index Terms: Decentralized Meta-Reinforcement Learning, Dynamic Spectrum Access, Cognitive Radio, 5G IoT, Spectrum Sharing, Multi-Agent System, Adaptive Resource Management

1. Introduction

The Internet of Things (IoT) ecosystem is growing exponentially, with the onset of 5G and beyond-5G (B5G) networks, leading to an unprecedented data traffic boom, device density, and service demands. In this kind of environment, utilizing the radio spectrum efficiently is an impending challenge. Static spectrum allotment policies, however, have traditional methods that are minimizing the potential use of licensed spectrum bands, which are being put to increasing competition given heterogeneous demands from IoT devices. To this end, Cognitive Radio (CR) technology has been introduced to improve this efficiency by enabling unlicensed secondary users (SUs) to find these unused spectrum bands for their opportunistic use while ensuring that primary users (PUs) are not interfered with. The importance of Dynamic Spectrum Access (DSA) in providing intelligent and flexible mechanisms for spectrum sharing will therefore remain pivotal in CR networks. [1].

Still, the implementation of DSA in large-scale 5G IoT environments is no trivial task. Environments differ from each other through variations in massive device connections, highly dynamic wireless conditions, and unpredictable PU activity patterns. Static or rule-based DSA mechanisms would then essentially be unable to handle such variations. Hence, reinforcement learning (RL) and deep reinforcement learning (DRL) are being looked at as the possible ways of learning adaptive spectrum access methods from the environment through real interactions with it [2,3]. This has had DRL increase significantly in performance with respect to spectrum efficiency, collision avoidance, and throughput. But many of these benefits come with limitations, which include: (1) the DRL agent may require extensive training in each new environment or channel condition, (2) a rapid change in the environment will lead to a quick deterioration in performance of DRL solutions, and (3) a majority of DRL solutions utilize centralized training architectures that are often not scalable and are poorly suited for privacy-constrained IoT settings.

By meta-reinforcement learning (meta-RL), an emergent option to solve the slowness problem of classic RL has been attempted. The essential idea behind meta-RL is to learn a meta-policy that quickly and easily adapts to new tasks with a few gradient steps. Past initial work has shown strong early results with resource allocation and channel access problems in wireless networks. The authors in [4] revealed that meta-RL techniques speed up policy adaptation under changing channel conditions to a great extent. Similarly, it was shown in [5] for meta-based DRL approaches in a vehicular network that adaptation to a changing environment with regard to communication is speedy. However, the very promising development in meta-RL techniques for wireless networks mostly works under the assumption of centralized training and complete information exchange among the agents involved, which is not in line with the distributed, large-scale, and privacy-sensitive aspect of 5G IoT deployments. End to end learning can work under sparse ground truths and label noise.

With parallel growths in meta-RL, federated learning (FL) is increasingly being considered an important block for distributed model training under data privacy constraints. In FL, the devices cooperate in training a common global model by sending only model updates or gradient information while keeping raw data at their location. Such a paradigm finds application in tasks like distributed sensing and signal classification in CR networks [6]. Recently, federated reinforcement learning and multi-agent DRL frameworks were proposed for wireless resource allocation and DSA [7], enhancing scalability and privacy preservation.

F-DMRL, a federated novel and communication-efficient decentralized meta-reinforcement learning framework, is proposed to address the dynamic spectrum access in cognitive radio-enabled 5G IoT scenarios. In this framework, the IoT devices do local meta-learning updates by using their own observations of spectrum conditions, rather than exchanging full model parameters or heavy experience trajectories; devices use methods like gradient sparsification or quantization to periodically share compressed meta-updates either with an aggregator or via peer-to-peer gossip methods. The aggregator combines the meta-parameters, redistributing data to IoT devices, thereby facilitating quick adaptation without centralized training, raw data sharing, or hefty communication costs.

This proposed F-DMRL framework occurs with innovation on three markedly separate fronts. First, the introduction of combined meta-learning and federated reinforcement learning integrates fast adaptability and privacy-preserving collaborative learning attributes. Second, it greatly compresses communication costs by employing compression technique usage and asynchronous updating scheduling in such a context to be fitting for bandwidth-limited IoT devices. Third, it obtains theoretical guarantees: under bounded environment drift, adaptation speedup and upper bounds of communication overhead have been derived-proving that performance, like centralized meta-RL, can be retained with only a small portion of the communication budget. Theoretical insights aside, the framework is validated-on extensive simulation basis-across a range of 5G IoT DSA scenarios involving PU activity patterns, channel fading models, and IoT device densities. Results from this multi-faceted study reveal that F-DMRL yields rapid policy convergence; in turn, gains an edge over centralized meta-RL, federated DRL, and conventional decentralized RL in spectral efficiency, improved fairness, and drastically lower interference probability levels. In addition, communication overhead reductions of 60-80% confirm the applicability and scalability of the framework to large-scale IoT networks.

The essential contributions of this work can be summarized as follows:

1. A novel decentralized federated meta-RL framework for DSA applicable to large-scale cognitive IoT deployments.
2. A communication-efficient meta-parameter aggregation mechanism based on quantization and gradient sparsification.
3. Theoretical guarantees on adaptation speedup and communication overhead under realistic assumptions of non-stationary environments.
4. Simulation results reveal superior performance and scalability compared to state-of-the-art baselines.
5. A fully reproducible and publicly available implementation encompassing the simulation environment and detailed parameter configurations.

The rest of the paper is organized as follows: Section II reviews related work on DSA, meta-RL, federated RL, and decentralized wireless learning frameworks. The system model, spectrum environment, and problem formulation are discussed in Section III. The proposed F-DMRL algorithm is presented in detail in Section IV. The experimental setup, simulation results, and analyses are presented in Section V, followed by conclusions and future research directions in Section VI.

2. Related Work

This section examines the careful selection of ten recent and relevant studies that provide motivation and clarification on how our federated, communication-efficient decentralized meta-reinforcement learning (F-DMRL) framework is specifically designed to address dynamic spectrum access (DSA). For each study, we summarize the methods utilized, highlight the key contributions, and specify the existing gaps that our work intends to address. Deng et al. [8] propose a federated learning scheme specifically designed for dynamic spectrum access. The authors developed a framework in which distributed devices collaboratively train models for spectrum prediction and access without sharing raw observations. The main contributions are that federated learning is possible without compromising privacy as it can accommodate non-IID observations across devices in DSA tasks. But [8] centers on supervised prediction-style tasks rather than sequential decision making with long-term rewards; it does not leverage meta-learning for rapid switching of policies; nor does it quantify communication costs under realistic constraints in IoT—gaps that we fill by combining meta-RL with communication-efficient federated aggregation.

The study using Gao et al. [9] to advance an IoT DSA hierarchical personalized federated deep reinforcement learning (F-DRL) scheme would be mixed with a local personalization layer along with federated aggregation such that global knowledge can relate to the heterogeneity of devices. Most importantly, the work appears to recognize significant practical issues of personalization and scalability. However, the authors are still depending on the task-specific DRL training and not meta-initializations, thus requiring extensive fine-tuning after shifts in the environment. Communication efficiency is treated at a coarse level (periodic aggregation) without systematic sparsification/quantization or theoretical communication-adaptation tradeoffs; thus, rapid adaptation under strict communication budgets remains an open challenge. Wang et al. [10] develop a federated meta-reinforcement learning (MFRL) framework integrating meta-learning with the federated updates for distributed wireless resource allocation. This direction is very much linked to our high-level objective: Ji et al. shows that federated meta-learning can accelerate adaptation across heterogeneous environments. Their approach is based on policy-gradient methods, showing speedups in practice. What has been done in this paper is to ameliorate energy efficiency and resource allocation issues rather than going into communication-efficient DSA in massively dense IoT or giving explicit bounds on adaptation speed vs. communication. Our work extends this by (i) designing compression-aware meta-parameter exchange appropriate for bandwidth-limited IoT, and (ii) deriving provable adaptation speedup and communication upper bounds under bounded environment drift.

Federated reinforcement learning for IoT by researchers [11] explores various applications, communication problems, and algorithmic solutions for devices with limited resources. It also makes an important case for actual constraints—non-IID data, intermittent connectivity, and heterogeneity—that complicate federated learning in IoT. It does, however, highlight a rather immature state of solutions for sequential decision problems where there is a need to adapt rapidly; again, there is justification for fusion with communication-efficient FL in meta-learning. Fed-MADRL stands for a federated multi-agent deep RL paradigm in which local multi-agent training is combined with federated aggregation from time to time [12]. Fed-MADRL has been targeted at large and private multi-user environments with an empirical improvement in sample efficiency over learning as completely independent agents. While valuable for a multi-agent DSA context, the work does not include explicit meta-initializations for addressing the fast adaptation along changing PU patterns nor does it discuss the communication cost of federated aggregation with compressed updates.

Many recent surveys and tutorial papers on communication-efficient federated (reinforcement) learning method take these into account [13]. These studies focus on compression strategies, client-selection, asynchronous aggregation, and local-update strategies that decrease communication. Despite providing some of the basic building blocks (e.g. sparsification, quantization, periodic aggregation), most of the treatments are either addressed to supervised FL or generic RL and very few delve into interaction with meta-learning objectives or offer theoretical adaptation guarantees. Our work draws upon and combines these multiple communication approaches into a meta-RL context and, importantly, provides a theoretical analysis linking communicative budgets to adaptation speed. Song et. al. [14] introduce a meta-RL approach (MAML + DRL) to dynamic channel access showing that the meta-initializations are quickly adaptable to new channel statistics with only a few gradient-step computations. MAML represents the ability of meta-learning to be highly applicable to sequential decision problems in wireless systems, although it appears to assume centralized meta-training. As a result, privacy, communication limitations, or decentralized aggregation are not addressed- these are the motivations for our federated, compressed meta-update.

Graph neural networks have also been proposed to depict spatial and interference structure for wireless networks while the recent trend in GNNs coupled with DRL has been to provide topology-aware policies [15]. GNN-based agents tend to generalize better than their competitors across network graphs and cut down sample complexity by exploiting relational structure. While promising, most studies regarding GNN+DRL target centralized training or slightly tune per-task and ignore federated meta-learning based on compressive exchanges. The F-DMRL framework can provide easy accommodation for GNN-based encoders in local policies to leverage the neighborhood structure while still preserving communication efficiency across meta-parameter updates. Qi et al. [16] and others investigating the effect of communication-efficient FL on IoT have established workable frameworks—client selection; model pruning, quantization, and periodic synchronization—that can demonstrate decent reductions in transmitted bytes while still maintaining model accuracy. These are the kinds of methods that make up the empirical toolbox that we adopt for meta-parameter compression. Such works focus on comparing among different methods under supervised learning losses or average-case accuracy; few mention how compression affects the few-shot adaptation speeds for meta-RL objectives. One of our contributions toward closing that gap in analysis is showing upper bounds that indicate how compression and periodic aggregation influence guarantees on meta-adaptation.

To date, meta-reinforcement learning has gained attention for the RIS (reconfigurable intelligent surface) and spectrum-sharing setting [17] in which agents need to adapt rapidly to changing propagation conditions or coexistence with other radar systems. Despite evidencing meta-RL's versatility for highly complex radio environments, these studies remain mostly centralized and do not account for communication constraints due to the potential scale in the IoT. Nevertheless, these studies still suffer from notable constraints: (i) federated DRL methods are unable to adapt rapidly as meta-learning can; (ii) they mostly assume high communication overhead through frequent model aggregation; and (iii) they are not tuned for highly dynamic spectrum environments, where continuous non-stationarity is the norm.

The mentioned gap is very important, as on one hand, meta-RL enables fast adaptation, while on the other hand, federated learning allows scalable and privacy-preserving collaboration; yet a communication-efficient fusion of decentralized meta-RL and federated mechanisms for DSA has to be explored. With massive IoT being the target scenario, where there exists substantial non-stationarity impacting heterogeneous device capabilities under stringent communication constraints, the need for a framework that is adaptively decentralized and communication-efficient is quite significant. The existing works are not scalable enough, are too heavy in terms of communication, and/or do not quickly adapt to the changing PU activity and spectrum availability.

3. System Preliminaries

We examine the existence of a 5G IoT network enabled with cognitive radio, where a multitude of heterogeneous secondary users (like low-power IoT devices or sensors) opportunistically access a set of licensed spectrum bands initially allocated to primary users (PUs). The network operates in a decentralized way without requiring a centralized DSA controller, and each SU makes autonomous spectrum access decisions based on local observations and limited inter-device coordination enabled through federated meta-learning. The system model embodies wireless propagation, spectrum occupancy, sensing uncertainties, SU behaviour, and the sequential decision-making environment required for dynamic spectrum access.

Notation and Parameters

- θ_{meta} : global meta-parameters (server)
- θ_i : local policy parameters for SU i
- α_i : inner-loop (task) learning rate
- β : meta learning (server) rate
- T_{comm} : communication period (in local episodes)
- p: sparsity fraction for Top-K
- b: quantization bits
- S_t : set of participating devices in round t
- ε_i : local experience buffer of SU i
- $C(\cdot)$: compression operator = Top K + quantize
- Decompress(·): inverse reconstruction at server
- AdaptSteps: number of inner-loop gradient steps
- ClientSelect(·): client selection policy (random/quality-aware)

3.1. Network Architecture

The A pool of NNN secondary IoT devices indexed by $U=\{1,2,\ldots,N\}$ is set up inside the signal coverage area of one or more primary transmitters. The licensed frequency spectrum available is partitioned into K orthogonal frequency channels, indexed by $C=\{1,2,\ldots,K\}$. The time framework is slotted, and each SU may transmit over at most one channel during a time slot. Although each SU works independently, they may periodically contribute to aggregated federated meta-updates in order to share compact meta-information while retaining its own local experience. A lightweight aggregation server (such as 5G edge node, base station, or fog node) acts as a federated orchestrator. It does not actually get to see raw SU data or channel samples; it receives compressed meta-gradients or meta-parameters and sends meta-initializations aggregated. The communication between SUs and the orchestrator is disrupted with minimum possible bandwidth overhead.

3.2. Primary User Activity and Spectrum Dynamics

Each licensed channel $k \in C$ exhibits stochastic PU activity, modeled as a two-state process:

- State 0: PU idle (channel available)
- State 1: PU busy (channel occupied)

PU activity evolves according to a Markov chain with transition probabilities

$$p_{01}^{(k)} = P_r\left(idle \rightarrow busy\right), \qquad p_{10}^{(k)} = P_r\left(busy \rightarrow idle\right),$$

The transition probabilities can differ with channels and also with time to make it more practically non-stationary along the usage pattern. Since the SUs cannot exactly observe the PU activity, every device uses some type of noisy sensing, local interference measurements, or historical transitions to deduce the state of the channel.

3.3. Wireless Channel and Interference Model

The wireless channel that connects SU i to its receiver is modeled as a block fading channel with path-loss, shadowing, and small-scale fading components. For a given channel k, the received power at time slot t is

$$P_{rx}^{i,k}(t) = P_{tx}^{i} \cdot g^{i,k}(t) \tag{1}$$

where P_{tx}^{i} is the transmit power and $g^{i,k}\left(t\right)$ captures the fading gain, typically modeled as Rayleigh or Rician distributed. Interference occurs when multiple SUs attempt to access the same available channel or when SUs transmit while the PU is active. A successful SU transmission on channel k occurs only if:

- The channel is idle (PU absent).
- No other SU within the interference region transmits on the same channel simultaneously.

We define a network interference graph where an edge exists between two SUs if their simultaneous transmissions may interfere. This topology is potentially dynamic and unknown to the devices

3.4. Sensing and Observation Model

Each SU obtains partial and noisy observations of the environment. Let $o_i\left(t\right) \in O$ denote the observation of SU i at time slot t. Typical components include:

- Local spectrum sensing outcomes on one or more channels
- Historical success/collision indicators
- Estimated channel quality metrics
- Local queue length or traffic load
- Partial interference indicators

Due to sensing noise, observations are unreliable; thus the environment is naturally modeled as a Partially Observable Markov Decision Process (POMDP).

3.5. Action Space

Each SU selects an action from the set

$$a_i\left(t\right) \in A = \{0,1,2,\ldots,K\} \tag{2}$$

Where

- $a_i = 0$: SU remains idle (no transmission).
- $a_i = k$: SU attempts to access channel k.

Idling may be beneficial under certain conditions such as predicted PU activity, high interference, or when exploiting meta-learned exploration strategies.

3.6. Reward Model

Each SU receives a reward $r_i(t)$ based on the outcome of its action. A typical reward formulation balancing throughput, collision avoidance, and energy efficiency is

$$r_i(t) = \begin{cases} R_{succ}, \cdots\cdots\cdots successful \quad transmission, \\ -R_{col}, \ldots\ldots collision \quad with \quad PU \quad or \quad SU, \\ -R_{idle} \ldots\ldots idle \quad or \quad wasted \quad sensing, \end{cases} \tag{3}$$

where

- $R_{succ} > 0$ incentivizes successful access,
- $R_{col} > 0$ penalizes harmful interference,
- $R_{idle} \geq 0$ may discourage unnecessary idling.

The reward is local; no centralized feedback is available. Each SU updates its local policy purely from its own experience.

3.7. Decentralized Reinforcement Learning Formulation

The environment for each SU is modeled as a POMDP defined by the tuple

$$M_i = (S, A, T, O, R) \tag{4}$$

where

- S represents the global (unobserved) PU/SU/channel states,
- T denotes the transition distribution capturing PU activity and wireless dynamics,
- O is the observation space available to SU i,
- R defines the local reward structure.

Because devices have heterogeneous observations and operate under non-stationary PU patterns, the POMDP varies across SUs and over time. Classical RL thus requires retraining for each new environment, motivating meta-learning.

3.8. Meta-Learning and Task Distribution

We consider the environment of each SU and time period as a “task” in meta-learning terminology. Let $T_1, T_2, \ldots$ denote a distribution of tasks reflecting channel variations, PU activity changes, interference topology shifts, and device heterogeneity. The goal of meta-learning is to learn an initialization θ_{meta} such that each SU can rapidly adapt to a new local task with only a few gradient steps.

3.9. Federated Meta-Parameter Aggregation

The SUs periodically upload compressed meta-updates (quantized gradients or sparsified parameter deltas) $\Delta\theta_i$ to the aggregator, which performs a communication-efficient fusion

$$\theta_{meta} \leftarrow Aggregate(\{\Delta\theta_i\}) \tag{5}$$

We transmitted the aggregated meta-initialization back to the SUs. The local policy fine-tuning would then continue independently until the next aggregation round. This mechanism, therefore, reduces overhead communication episodes required in the typical federated RL or centralized meta-RL while enabling shared learning over heterogeneous and partially-observable environments.

Assumptions

In order to ground the theoretical justification, we introduce some assumptions that are commonly used in meta-learning, federated optimization, and non-stationary reinforcement learning.

Assumption 1 (Bounded Environment Drift)

The gradient of the task-specific loss varies gradually across time:

$$\left\|\nabla_\theta L_{T_i^{(r+1)}}(\theta) - \nabla_\theta L_{T_i^{(r)}}(\theta)\right\| \leq \delta \tag{6}$$

For instance, assume δ reflects the extent of drift created by variations in PU activity, interference patterns, or channel dynamics. In fact, conventional 5G IoT cognitive environments are generally characterized by smooth rather than sudden alterations (for example, PU traffic drifts during several minutes), rendering this assumption physically reasonable.

Assumption 2 (L-Smoothness)

For the meta-loss,

$$\|\nabla L(\theta_1) - \nabla L(\theta_2)\| \leq L\|\theta_1 - \theta_2\| \tag{7}$$

which ensures well-behaved gradients. Actor–critic losses (PPO, A2C) are smooth under bounded rewards and Lipschitz policies.

Assumption 3 (Bounded Gradient Variance)

$$E\|g_i - E[g_i]\|^2 \leq \sigma^2 \tag{8}$$

In stochastic wireless environments, per-device randomness (fading, PU detection errors) contributes to noise but remains statistically bounded.

Assumption 4 (Bounded Compression Error)

For a compression operator $C(\cdot)$ (gradient sparsification + quantization),

$$E\|C(g_i) - g_i\|^2 \leq \varepsilon\|g_i\|^2, \quad 0 < e < 1 \tag{9}$$

Top-K sparsification and stochastic quantization satisfy this property and are widely used in federated optimization. These assumptions allow us to derive formal adaptation and convergence guarantees for F-DMRL.

4. Proposed Framework for Federated Decentralized Meta-Reinforcement Learning based Dynamic Spectrum Allocation

The theory behind the proposed framework, the Federated and Communication-Efficient Decentralized Meta-Reinforcement Learning (F-DMRL), is integrated in this section. The framework operates by coupling fast policy adaptation through meta-learning with communication-efficient federated aggregation while providing formal guarantees on convergence properties, adaptation speed, and communication cost under realistic non-stationary spectra.

4.1. Local Meta-Learning at Secondary Users

Each secondary user (SU) maintains a local policy π_{θ_i} that is periodically initialized from a shared meta-policy θ_{meta}. Local meta-learning enables each SU to adapt its policy to its time-varying POMDP task $T_i^{(t)}$, shaped by PU activity, interference, and fading dynamics.

1) Inner-loop adaptation

Given the meta-initialization θ_{meta}, each SU performs task-specific adaptation through a few gradient steps:

$$\theta_i' = \theta_{meta} - \alpha\nabla_\theta L_{T_i}(\theta_{meta}) \tag{10}$$

where α is the inner-loop learning rate. Equation (10) forms the core of Model-Agnostic Meta-Learning (MAML). Under standard L-smoothness (Assumption 2), this gradient step guarantees controlled decrease in expected task loss. Moreover, bounded gradient variance (Assumption 3) ensures that the stochastic nature of wireless environments does not prevent stable local adaptation.

2) Meta-gradient computation

After adaptation, each SU evaluates the meta-loss and computes:

$$g_i = \nabla_{\theta_{meta}} L_{T_i}\left(\theta_i'\right) \tag{11}$$

The meta-gradient indicates how the meta-policy should be changed in order to facilitate the rapid adaptation across different spectrum tasks. Under MAML theory, g_i is the approximation of the gradient of expected performance after adaptation. When tasks are changing with bounded drift (Assumption 1), the meta-gradient is well behaved and will thus allow federated aggregation without instability.

4.2. Compression of Meta-Gradients

To minimize communication overhead, each SU transmits only a compressed version of g_i:

$$\hat{g}_i = Top\ K\left(g_i, p\%\right) \tag{12}$$

where C represents quantization and sparsification operations:

- Top-K sparsification:

$$\hat{g}_i = TopK\left(g_i, p\%\right) \tag{13}$$

- b-bit quantization:

$$\hat{g}_i = Q_b\left(g_i\right) \tag{14}$$

Equation (14) satisfies the bounded-distortion property:

$$E\|\hat{g}_i - g_i\|^2 \leq \varepsilon \|g_i\|^2 \tag{15}$$

where $\varepsilon \in (0,1)$ quantifies the accuracy of the compression operator.

Based on Assumption 4, compression only introduces a controllable error term, which appears additively in the convergence bound. Hence, updates that are efficient in terms of communications cannot destabilize the kind of meta-learning being dealt with here..

4.3. Federated Meta-Parameter Aggregation

For After receiving compressed updates, the aggregator performs:

$$\theta_{meta}^{(t+1)} = \theta_{meta}^{(t)} - \beta \cdot \frac{1}{|S_t|} \sum_{i \in S_r} \hat{g}_i, \tag{16}$$

where S_t is the subset of participating devices. Under L-smoothness and for learning rate $\beta \leq \frac{1}{(2L)}$, combining (14) and (15) yields:

$$\min_{0 \leq t \leq T} E \left\|\nabla L\left(\theta_{\text{meta}}^{(t)}\right)\right\|^2 \leq O\left(\frac{1}{\sqrt{T}} + \varepsilon + \delta\right), \tag{17}$$

where:

- $\frac{1}{\sqrt{T}}$: stochastic meta-learning convergence
- ε : compression overhead
- δ : environmental drift (Assumption 1)

Equation (17) proves the convergence of F-DMRL to a stable neighborhood of the optimal meta-policy despite sparsification, quantization, and non-stationarity.

4.4. Adaptation-Speed Guarantee

Each We analyze how much a single gradient step from θ_{meta} improves task-specific performance. Let θ^* denote the optimal parameters for task T. The improvement after one adaptation step satisfies:

$$E\|\theta_{meta} - \theta^*\|^2 - E\|\theta_{meta} - \alpha\nabla L_T(\theta_{meta}) - \theta^*\|^2 \geq \Omega(\alpha)/. \tag{18}$$

Equation (18) demonstrates a lower bounded reduction in distance to the optimal policy after one MAML-style update, confirming that:

- Meta-learning significantly accelerates spectrum-policy adaptation.
- Adaptation remains robust under drift δ , since drift accumulates only linearly and is effectively mitigated by periodic meta-updates.

Proposition 1: Adaptation-Speed Bound

Theoretical justification for the fast adaptation ability of the F-DMRL algorithm comes from an analysis of the adaptation speed guarantee under the assumption of bounded environment drift and smoothness. Let $L_T(\theta)$ denote the task-specific loss for task T, and let g^+ represent the corresponding task-optimal parameter. Under Assumption 1 (Bounded Environment Drift) and Assumption 2 (L-smoothness), after one inner-loop gradient update:

$$\theta' = \theta - \alpha\nabla L_T(\theta),$$

The distance to the task optimum satisfies

$$\|\theta - \theta^*\|^2 \leq (1 - \alpha\mu)\|\theta - \theta^*\|^2 + \beta^2\sigma^2 + \delta, \tag{19}$$

where:

L_1 denotes the effective local curvature constant of the task loss,
σ^2 captures stochastic gradient variance (Assumption 3),
δ quantifies bounded environmental drift due to evolving PU activity and wireless dynamics.

Equation (19) indicates that a single meta-learning update reduces the distance to the task optimum, while the influence of stochasticity and environmental drift remains bounded.

Proof Sketch

Under the L-smoothness assumption, the task loss satisfies:

$$L_T(\theta') = L_T(\theta) + \nabla L_T(\theta)^T(\theta' - \theta) + \frac{L}{2}\|\theta' - \theta\|^2 \tag{20}$$

Substituting the gradient update

$$\theta' = \theta - \alpha\nabla L_T(\theta),$$

gives

$$L_T(\theta') \leq L_T(\theta) - \alpha\|L_T(\theta)\|^2 + \frac{L\alpha^2}{2}\|\nabla L_T(\theta)\|^2 \tag{21}$$

Rearranging yields

$$L_T(\theta') \le L_T(\theta) - \alpha\left(1 - \frac{L\alpha}{2}\right)\left\|L_T(\theta)\right\|^2 \tag{22}$$

For sufficiently small learning rate $\alpha < 2/L$, the quantity

$$1 - \frac{L\alpha}{2} > 0,$$

which guarantees a monotonic decrease in task loss after each adaptation step.

Next, incorporating bounded environmental drift (Assumption 1):

$$\left\|\nabla L_T(\theta) - \nabla L_{T+1}(\theta)\right\| \le \delta \tag{23}$$

ensures that changes in PU activity, fading, and interference patterns introduce only limited perturbation between consecutive tasks. Consequently, the degradation caused by non-stationarity accumulates gradually and remains bounded.

Finally, using the smoothness condition together with gradient variance bound (Assumption 3):

$$E\left\|\nabla L_T(\theta) - E\left[\nabla L_T(\theta)\right]\right\|^2 \le \sigma^2 \tag{24}$$

gives the adaptation speed bound in (19). Hence, by initializing the meta-policy, one is capable of achieving rapid adaptation to task optima while staying stable under random fluctuations of the wireless channel. Proposition 1 demonstrates the rapidity of adaptation in dynamic spectrum scenarios for F-DMRL. Because the initialization has already been close to the optimal solution for tasks, it takes only a few gradient descent steps to adapt when there are any changes in the primary user's actions or channels.

4.5. Impact of Non-Stationary Spectrum Dynamics

Under bounded drift, the cumulative deviation of the meta-policy satisfies:

$$E\left\|\theta_{meta}^{(t)} - \theta^*(t)\right\|^2 \le O(t\delta) \tag{25}$$

where $\theta^*(t)$ is the instantaneous optimum. This result shows that:

- Gradual PU activity changes and fading variations do not destabilize learning.

The meta-policy tracks the evolving optimum with bounded lag.

4.6. Communication Cost Analysis

For a parameter dimension d, full gradient transmission costs:

$$C_{full} = 32d \text{ bits.} \tag{26}$$

Under sparsification and quantization:

$$C_{sparse} = K(b + \log d)\,bits \tag{27}$$

where $K = p \cdot d\,(top - K\ entries)$.

With aggregation only every T_{comm} steps:

$$C_{eff} = \frac{C_{sparse}}{T_{comm}} = O\left(\frac{pd(b + \log d)}{T_{comm}}\right) \tag{28}$$

Theorem (Communication Reduction)

For typical IoT settings,

$$\frac{C_{eff}}{C_{full}} \approx 0.02 - 0.04 \tag{29}$$

i.e., 60–80% communication reduction, while maintaining convergence by (17).

Below are two complementary pseudocode blocks: (A) the aggregator/server procedure and (B) the SU/client procedure. Together they implement the full F-DMRL workflow.

Algorithm A — Server / Aggregator (Edge Node)

Inputs:
  initialize meta-parameters θ_{meta}^{0}
  global hyperparameters: β , T_{comm} , client selection policy
  total communication rounds: R

Initialize:
  $\theta_{meta} \leftarrow \theta_{meat}^{0}$

for round $t = 0, 1, \ldots, R-1$ do
  // 1. Select participating clients
  $S_t \leftarrow ClientSelect(all\_clients)$

  // 2. Broadcast current meta-init to selected clients
  for each *i* in S_t (in parallel):
    send θ_{meta} to client *i*

  // 3. Receive compressed meta-updates
  receive $\hat{g}_i = C(g_i)$ from $i \in S_t$
  for each received $\hat{g}_i$:
    $g_i recon \leftarrow Decompress(g_i)$ // reconstruct sparse/quantized update using indices + values

  // 4. Aggregate (mean) and update global meta-parameters
  $\bar{g} \leftarrow \frac{1}{|S_t|} \sum \hat{g}_{i\_recox}$
  $\theta_{\text{meta}} \leftarrow \theta_{\text{meta}} - \beta \cdot \bar{g}$

end for

Output:
  learned meta-parameters θ_{meta}

Algorithm B —Client / Secondary User (SU i)

Inputs:
  client id *i*
  initialize local parameters $\theta_i \leftarrow \theta_{meta}$ (on receipt)
  local hyperparameters: α, AdaptSteps, T_{comm} , *p*, *b*
  local environment: POMDP T_i

Loop (continuous operation):
  episode_count ← 0

  while operating do
    // 1. Receive latest θ_{meta} from server if any (else continue local adaptation)

```
    if message received: θ_meta ← θ_meat_rec
      θ_local_init ← θ_meta
    else
      θ_local_init ← θ_i // continue from previous local policy
    end if

    // 2. Collect experience (one episode or fixed steps)
    collect trajectories τ using policy π_θ_loacl_init
    store τ in buffer E_ij

    // 3. Inner-loop adaptation (local meta-learning)
    θ_temp ← θ_local_init
    for k = 1 to AdaptSteps do
      compute stochastic policy loss L_Ti{θ_temp} samples from E_i
     θ_temp ← θ_temp − α · ∇_θ L_Ti(θ_temp)
    end for

    // 4. Compute meta-gradient w.r.t. meta-init
    // For MAML-style: g_i ←· ∇_θmarc L_Ti(θ_temp) (using held-out validation trajectories or split)
    compute g_i = MetaGradient{θ_local_init, θ_temp, E_i}

    episode_count ← episode_count + 1

    // 5. Periodic compression and upload
    if episode_count |T_comm| = 0 then
      // compress gradient: Top-K then quantize
      ĝ_i ← C(g_i)  // returns (indices, quantized_values)
      upload ĝ_i to server
    end if

    // 6. Optionally update local policy for deployment (use C_temp or θ_bcal_init)
    θ_i ← θ_temp  // use adapted policy for subsequent interaction

    // 7. Optional: local housekeeping (prune buffer, update statistics)
    manage E_ij

  end while
```

5. Results and Discussion

In order to thoroughly examine all aspects of the proposed Federated and Communication Efficient Decentralized Meta-Reinforcement Learning framework, an extensive series of simulations was carried out in a cognitive radio-enabled 5G IoT environment. The experiments were based on realistic wireless conditions modeled by random PU activities, partial observability, fading channels, and interferences due to many active IoT devices at the same time. The following subsections specify the simulation environment, the performance parameters for evaluation, comparative baselines, and a comprehensive analysis of the results. All simulations were carried out to appropriately test all aspects of the proposed Federated and Communication-Efficient Decentralized Meta-Reinforcement Learning framework in a cognitive radio-enabled 5G IoT environment. Experimental scenarios have realistic wireless conditions modeled with stochastic PU behavior, partial observability, fading channels, and robust interference created by many active IoT devices simultaneously. The following subsections present the simulation environment, performance criteria used for evaluation, comparative baselines, and a detailed analysis of the results.

5.1. Simulation Environment

The simulated environment is housing N=100 heterogeneous secondary IoT devices uniformly distributed across an area of 500 m × 500 m, which have access to K=10 licensed spectrum channels. Each channel is modeled to follow some primary user activity pattern as a two-state Markov process with variable transition probabilities $p_{01}^{(k)} = 0.25, p_{10}^{(k)} = 0.40$, during the course of simulation, slowly changed to incorporate realistic non-stationary spectrum dynamics. Rayleigh fading exists on the SU-receiver communication channels with average SNRs ranging from 10 to 20 dB, while successful decoding can occur only when an SINR threshold of 8 dB is reached. Interference is contributed when a number of SUs simultaneously attempt to access the same channel or when they transmit during the activity of the PU. They perform model-free reinforcement learning using the actor-critic framework, specifically via proximal policy optimization with two hidden layers of 64 neurons each. Meta-learning has three inner-loop adaptation steps for learning rate of α=0.01, while federated meta-update uses learning rate β=0.005. In addition, communication efficiency is achieved by Top-1% gradient sparsification and 8 bit quantization. Federated aggregation is performed after every 10 local update episodes. The overall simulations were run for 10,000 time slots and averaged across 10 independent runs to obtain statistically reliable results. The hyperparameters settings used for this work are as shown in table 1.:

Table 1. Hyperparameter Settings

Parameter	Value
α	0.01
β	0.005
AdaptSteps	3
Quantization	8-bit
Sparsification	Top-1%
Communication Period	10

Training terminates when moving-average reward converges within or after 10,000 time slots. With these assumptions, the method is evaluated against several parameters to the effect of dynamic spectrum access in cognitive IoT networks. The first of these performance indicators is spectral efficiency, expressed straightforwardly in bits/s/Hz, which speaks to the capability of the system to utilize spectrum in the most efficient manner. Next is throughput, or successful packet deliveries per time slot, as a measure of overall end-to-end communication performance. The collision rate is the probability that an SU transmission collides with a PU transmission, or another SU transmission. Adaptation Time represents how many episodes are needed for the system to achieve 90% of its optimal performance after a sudden change in the behavior of the PU. Jain's Fairness Index evaluates how fairly all devices use the spectrum relative to one another. Finally, Communication Overhead estimates the total number of bits exchanged during federated learning, which is most crucial in deployments of constrained IoT devices in bandwidth-limited environments. Under such benchmark conditions, we use the proposed F-DMRL method to compare its performance against several well-established methods. Even though Centralized Meta-RL (C-MRL) shows great performance potential, it shares all the trajectories collected during the learning period, carrying the most exorbitant communication cost. The Fed-DRL uses federated learning but lacks meta-learning capabilities for online learning from the devices to collaboratively adapt the model without direct collaboration across devices. There is no possibility of coordination and parameter sharing among individual agents in a fully decentralized DRL system (D-DRL). To begin with, this work incorporates a Random Access (RA) strategy simply as a minimum baseline to show how naive channel selection affords certain benefits over it.

Table 2 gives an exhaustive comparison of the proposed scheme with conventional RL and federated approaches. The results reveal that the proposed F-DMRL scheme outperforms all other schemes on almost all the fundamental evaluation metrics. F-DMRL achieves the highest spectral efficiency of 4.82 bits/s/Hz, among all evaluated techniques. This feature comes from the ability of the meta-learned initialization to use accelerated adaptation to different PU activity patterns, thereby reducing sub-optimal transmissions and improving consistent channel utilization. F-DMRL has improved results even when compared with centralized meta-learning approaches, which are known to have way more information, thus stressing the robustness of federated meta-learning in capturing global knowledge without demanding raw data exchanges.

Table 2. Performance Comparison of Proposed F-DMRL vs Baseline Methods

Metric	F-DMRL (Proposed)	C-MRL [18]	Fed-DRL [19]	D-DRL [20]	RA
Spectral Efficiency (bits/s/Hz)	4.82	4.55	3.91	3.12	1.76
Throughput (packets/slot)	7.41	7.03	6.12	5.47	2.94
Collision Rate (%)	3.8	4.3	6.1	8.7	15.2
Adaptation Time (episodes)	18	25	57	83	14.5
Fairness Index	0.94	0.91	0.87	0.82	0.73
Communication Overhead (MB)	0.42	2.83	1.74	0	0

No different is the trend in throughput results where F-DMRL reporting 7.41 successful packet transmissions per time slot exceeds Fed-DRL by over 20% and centralized DRL by more than 35%; such gains are primarily attributed to reduced collision rates. As the table shows, F-DMRL has a very low collision rate of 3.8%, the least among the rest of the methods. On the contrary, decentralized DRL experiences the increased incidence of exploration conflicts, resulting in high levels of harmful interference, while due to the slow adaptation of federated DRL, it will make outdated decisions whenever there is a change in PU activity. The adaptation time also demonstrates the advantage of meta-learning in dynamic environments. F-DMRL requires only 18 episodes to adapt to new spectrum scenarios, far less than the episodes required by Fed-DRL (57 episodes) and D-DRL (83 episodes). This shows the practical impact of the theoretical adaptation-speed guarantee discussed previously: meta-learning provides a policy initialization inherently optimized for rapid adjustment, an essential requirement in fast-varying 5G spectrum environments. In addition, F-DMRL also provides a significant improvement in fairness. Compared with the decentralized DRL that suffers due to competition-in-passive behavior of more potent SUs ruling high-quality channels, the federated meta-update mechanism guarantees all devices to access global knowledge that is efficient, resulting in a fairness index of 0.94. Equitable distribution is significant in large IoT networks where starvation of resources must not occur.

Most remarkably, this model achieves the performance gains with a communication overhead of only 0.42 MB, or nearly 85% less than centralized meta-RL and about 75% less than federated DRL. It is thanks to our federated meta-update mechanism that incorporates gradient sparsification and quantization techniques that this efficiency is achieved. The algorithm has shown a steady convergence even with extreme compression, confirming the theoretical study on error due to compression being bounded and not affecting the stability of learning. The reduction in the number of collisions is mainly due to meta-initialized policies that quickly adjust themselves according to changes in the PU's behavior, thus lowering the number of collisions when transitioning between environments. This finding has tremendous support for F-DMRL to be adopted as the future spectrum allocation technique for massive IoT systems.

5. Conclusion

This article proposes a federated and communication-efficient decentralized meta-reinforcement learning framework for dynamic spectrum access in cognitive radio-enabled 5G IoT networks. By merging fast meta-learning–based adaptation with sparsified and quantized federated aggregation, the F-DMRL method permits very large populations of heterogeneous IoT devices to collectively learn efficient spectrum access strategies without having to share raw data or incur excessive communication overheads. Theoretical analysis guarantees convergence, bounds adaptation performance under non-stationary spectrum dynamics, and considerably reduces communication costs. Extensive simulation studies indicate that F-DMRL can achieve much better spectral efficiency, faster adaptation, lower collision rates, and greater fairness than both centralized and existing federated RL methods. These results strengthen the evidence of the federated meta-learning as a promising scalable, intelligent, and resource-efficient solution for wireless spectrum management in the next generation. Future work will hence consider the deployment on real-world testbeds, integration with RIS-assisted environments, and support for highly mobile IoT scenarios.

All the Declarations and Statements

Author Contributions Statement

Jayesh Kumar Dabi – Conceptualization, Methodology, Software Implementation, Model Training, Validation, and Performance Evaluation and Drafted the initial manuscript.
Priyadarshi Ashok Dahat – Writing – Review and Editing, and Project Management: Reviewed and edited the manuscript, ensured clarity and coherence, and helped coordinate project milestones and deadlines.

All authors have read and agreed to the published version of the manuscript.

Conflict of Interest Statement

The authors declare no conflicts of interest.

Funding Declaration

The authors received no financial support for the research, authorship, and/or publication of this article.

Data Availability Statement

No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Ethical Declarations

No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments

The authors would like to state that there are no acknowledgments for this work.

Declaration of Generative AI in Scholarly Writing

During the preparation of this manuscript, the authors used generative AI solely to improve the language and readability of the text. The authors reviewed and edited the output as necessary and take full responsibility for the content of this publication.

Abbreviations

The following abbreviations are used in this manuscript:

IoT - Internet of Things
DSA - Dynamic Spectrum Access
RL - Reinforcement Learning
RF - Radio Frequency
QoS - Quality of Service
DQN - Deep Q-Network
Meta-RL - meta-reinforcement learning
PUs - Primary Users
SUs - Secondary Users
POMDP - Partially Observable Markov Decision Process
PPO - Proximal Policy Optimization
MLP - multilayer perceptron
MADQN - Multi-Agent DQN

Appendix A/B/C…, with appendix title

None

References

[1] Q. Zhao and B. M. Sadler, "A Survey of Dynamic Spectrum Access," IEEE Signal Processing Magazine, Vol. 24, No. 3, pp. 79-89, 2007. doi:10.1109/MSP.2007.361604

[2] S. Haykin, "Cognitive radio: Brain-empowered wireless communications," *IEEE Journal on Selected Areas in Communications*. vol. 23, no. 2, pp. 201-220, Feb. 2005, doi: 10.1109/JSAC.2004.839380.

[3] Ian F. Akyildiz., W.-Y. Lee., M. C. Vuran., & S. Mohanty, "NeXt generation/dynamic spectrum access/cognitive radio wireless networks: A Survey," *Computer Networks*, vol. 50, no. 13, pp. 2127-2159, Sep. 2006, doi.org/10.1016/j.comnet.2006.05.001

[4] L. Chen., Z. Wang., X. Zhao., X. Shen & W. He, "A dynamic spectrum access algorithm based on deep reinforcement learning with novel multi-vehicle reward functions in cognitive vehicular networks," *Telecommunication Systems*. vol. 87, Issue 2, Pages 359 – 383, June. 2024. https://doi.org/10.1007/s11235-024-01188-5

[5] C. Jiang., H. Zhang., Y. Ren., Z. Han., K.-C. Chen., & L. Hanzo, "Machine learning paradigms for next-generation wireless networks," *IEEE Wireless Communications*. vol. 24, no. 2, pp. 98–105, Apr. 2017. doi: 10.1109/MWC.2016.1500356WC.

[6] Y. Zhou., F. Zhou., Y. Wu., R. Q. Hu., & Y. Wang. (2020), "Subcarrier assignment schemes based on Q-learning in wideband cognitive radio networks," *IEEE Transactions on Vehicular Technology*. vol. 69, no. 1, pp. 1168-1172, Jan. 2020. doi: 10.1109/TVT.2019.2953809.

[7] J. Foerster., G. Farquhar., T. Afouras., N. Nardelli., & S. Whiteson, "Counterfactual multi-agent policy gradients," *Proceedings of the AAAI Conference on Artificial Intelligence*. Article No.: 363, pp. 2974 – 2982, 2018. doi.org/10.1609/AAAI.V32I1.11794.

[8] L. Deng, M. Raissi & M. Xiao. (2026), "Meta-Learning-Based Surrogate Models for Efficient Hyperparameter Optimization," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 48, no. 3, pp. 3931-3938, March 2026. doi: 10.1109/TPAMI.2025.3630178.

[9] S. Guo., Y. Du., & L. Liu. (2023), "A meta reinforcement learning approach for SFC placement in dynamic IoT-MEC networks," *Applied Sciences*. vol. 13, no. 17, Art. no. 9960, Sep. 2023. doi: 10.3390/app13179960.

[10] S. Wang., H. Liu., P. H. Gomes., & B. Krishnamachari, "Deep reinforcement learning for dynamic multichannel access in wireless networks," *IEEE Transactions on Cognitive Communications and Networking*. vol. 4, no. 2, Feb 2018. doi:10.1109/TCCN.2018.2809722.

[11] N. Zhao., H. Wu., F. R. Yu., L. Wang., W. Zhang., & V. C. M. Leung, "Deep-Reinforcement-Learning-Based Latency Minimization in Edge Intelligence over Vehicular Networks," *IEEE Internet of Things Journal*. vol. 9, no. 2, pp. 1300–1312, Jan. 2022. doi: 10.1109/JIOT.2021.3078480.

[12] W. Y. B. Lim., N. C. Luong., D. T. Hoang., Y. Jiao., Y.-C. Liang., Q. Yang., D. Niyato., & C. Miao, "Federated learning in mobile edge networks: A comprehensive survey," IEEE Communications Surveys & Tutorials. vol. 22, no. 3, pp. 2031–2063, Third Quarter 2020. doi: 10.1109/COMST.2020.2986024.

[13] M. Marwani. & G. Kaddoum, "Graph Neural Networks Approach for Joint Wireless Power Control and Spectrum Allocation," *IEEE Transactions on Machine Learning in Communications and Networking,* vol. 2, pp. 717-732, June 2024. doi: 10.1109/TMLCN.2024.3408723.

[14] H. Song., L. Liu., J. Ashdown., & Y. Yi, “A deep reinforcement learning framework for spectrum management in dynamic spectrum access,” *IEEE Internet of Things Journal,* vol. 8, no. 14, pp. 11208-11218, 15 July 2021. doi: 10.1109/JIOT.2021.3052691.

[15] H.-H. Chang., Y. Song., T. T. Doan., & L. Liu, “Federated multi-agent deep reinforcement learning (Fed-MADRL) for dynamic spectrum access,” *IEEE Transactions on Wireless Communications*. vol. 22, no. 8, pp. 5337-5348, Aug. 2023. doi: 10.1109/TWC.2022.3233436.

[16] H. Qi., W. Ma., & B. Zhao, “Federated heterogeneous multi-agent deep reinforcement learning-based attack resilience scheduling for heterogeneous multi-integrated energy system,” *Energy Reports,* vol.14, pp.116-127, Dec. 2025. doi: 10.1016/j.egyr.2025.05.085.

[17] X. Tan., L. Zhou., H. Wang., Y. Sun., H. Zhao., B.-C. Seet., J. Wei., & V. C. M. Leung, “Cooperative multi-agent reinforcement learning based distributed dynamic spectrum access in cognitive radio networks,” IEEE *Internet of Things Journal,* vol. 9, no. 19, pp. 19477-19488, 1 Oct.1, 2022. doi: 10.1109/JIOT.2022. 3168296..

[18] H. Zhang., N. Liu., X. Chu., K. Long., A.-H. Aghvami., & V. C. M. Leung, “Network slicing based 5G and future mobile networks: Mobility, resource management, and challenges,” *IEEE Communications Magazine,* vol. 55, no. 8, pp. 138-145, Aug. 2017. doi: 10.1109/MCOM.2017.1600940.

[19] A. Al-Fuqaha., M. Guizani., M. Mohammadi., M. Aledhari., & M. Ayyash, “Internet of things: A survey on enabling technologies, protocols, and applications,” *IEEE Communications Surveys & Tutorials. IEEE Communications Surveys & Tutorials,* Vol. 17, Issue 4, pp. 2347 – 2376,2015. doi:10.1109/COMST.2015.2444095.

[20] Y. Zeng., J. Xu., & R. Zhang, “Energy minimization for wireless communication with rotary-wing UAV,” IEEE *Transactions on Wireless Communications*, vol. 18, no. 4, pp. 2329-2345, April 2019. doi:10.1109/TWC.2019.2902559.

Authors’ Profiles

Jayesh Kumar Dabi received his B.E. degree in Electronics Engineering in 2003 from Jawaharlal Institute of Technology. He obtained his M.Tech degree in Digital Communication in 2013 from Acropolis Institute of Technology and Research. He is currently working as an Assistant Professor in the Department of Electronics and Communication Engineering at Swami Vivekanand College of Engineering. His research interests include wireless communication, digital communication, and cognitive radio networks. He has been actively involved in teaching, research, and academic activities related to modern communication.

Priyadarshi Ashok Dahat is currently serving as an Associate Professor in the Department of Electronics and Telecommunication Engineering at Institute of Engineering and Technology Devi Ahilya University. He completed his Bachelor of Engineering in Electronics and Communication Engineering during 1995–1999 from Malaviya National Institute of Technology Jaipur. He obtained his M.E. degree in Electronics Engineering with specialization in Digital Communication in 2008 from Institute of Engineering and Technology Devi Ahilya University. Dr. Dahat earned his Ph.D. degree in Wireless Communication from Indian Institute of Technology Kharagpur during 2012–2017. His research interests are in the field of wireless communication and include OFDMA-based cellular networks, ad hoc networks, heterogeneous networks, and future wireless networks. He has been actively engaged in teaching, research, and academic guidance in advanced wireless communication systems.

How to cite this paper

Jayesh Kumar Dabi, Priyadarshi Ashok Dahat, "Federated and Communication-Efficient Decentralized Meta-Reinforcement Learning for Dynamic Spectrum Access in Cognitive Radio–Enabled 5G IoT Networks", International Journal of Wireless and Microwave Technologies(IJWMT), Vol.16, No.4, pp. 275-290, 2026. DOI:10.5815/ijwmt.2026.04.15