DRL-Driven Intelligent Access Traffic Management for Hybrid 5G-WiFi Multi-RAT Networks

Xueqing Zhou, Haiyuan Li, Anderson Bravalheri, Amin Emami,
Reza Nejabati, Shuangyi Yan, and Dimitra Simenidou
High Performance Network Group, Smart Internet Lab, Faculty of Engineering, University of Bristol, U.K.
E-mail: xueqing.zhou@bristol.ac.uk

***Abstract*—Integrating mobile networks with Non-3GPP networks provides a promising solution to mitigate the wireless RF spectrum scarcity. Despite the maturity of integration technologies, a comprehensive approach for radio resource allocation in highly dynamic and complex multiple Radio Access Technologies (multi-RAT) networks is still lacking. To tackle this challenge, this paper proposes an Access Traffic Management (ATM) system that enhances radio resource allocation during access, transmission, and handover processes. The system features a scalable and concise ATM-supported multi-RAT network architecture, supported by a Deep Deterministic Policy Gradient (DDPG) based Intelligent ATM (IATM) algorithm. To evaluate the proposed system, a Network Simulator 3 (NS3) based network simulation is built with realistic 5G and WiFi modules, interacting with the IATM algorithm in real time for decision making and policy improvement. Numerical improvements of our solution demonstrate a 45% and 70% increase in resource utilization efficiency, respectively, compared to two traditional traffic steering modes. Additionally, our solution improves link quality by a factor of three and doubles throughput with the same cost. Moreover, session stability is significantly enhanced under conditions of network size dynamics and UE mobility. Notably, we explore a set of universal UE-side parameters and verify their effectiveness, which further enhances the scalability of the ATM architecture and facilitates the development of UE-led IATM algorithms.**

***Index Terms*—multi-RAT, radio resource management, DRL.**

I. Introduction

The future cellular networks are expected to meet increasingly stringent demands, providing ubiquitous and uninterrupted connections for User Equipments (UEs) in a fully connected world. This compels cellular networks to expand their scarce radio spectrum into higher frequencies to achieve greater capacity, incurring significant costs in terms of increased power consumption and infrastructure development. On the other hand, the coexistence of multiple Radio Access Technologies (multi-RAT) is common to both current and future networks, especially in indoor environments. Hence, integrating with the Non-3GPP Radio Access Technologies (RATs) turns to be a potential solution to release more unlicensed radio resources [1]. However, each RAT currently employs its own spectrum management strategy, leading to inefficient utilization of radio resources. Therefore, it is imperative to develop a unified Access Traffic Management (ATM) system for the future multi-RAT network.

The ATM in multi-RAT scenarios involves several wireless network management issues, including network selection, network handover, and network aggregation. Optimising these issues has been researched independently. The optimization algorithms for RAT selection are proposed to improve network utilization and user experience by selecting a less-congested, better-quality, and lower-cost link [2, 3]. Enhanced vertical handover strategies have been widely investigated to minimize interruption when switching between different RATs [4]. Traffic splitting, which diverts traffic and transmits it through several RATs simultaneously, is another promising approach to provide more bandwidth with better link utilisation [5]. Besides, some works have jointly addressed several ATM issues, such as [6] and [7], which both propose two-tier Deep Reinforcement Learning (DRL) approaches. They use two separate DRL algorithms to handle network selection and bandwidth allocation, respectively, running on different network elements. However, addressing individual ATM issues as separate or sequential network functions can result in a more complex system architecture and decreased information utilization efficiency. Therefore, a holistic approach that jointly handles the three major ATM issues is necessary for effective multi-RAT radio resource management.

In this regard, the Access Traffic Steering, Switching, Splitting (ATSSS) function was proposed by 3GPP to serve as an unified solution, providing more comprehensive and efficient multi-RAT ATM [8]. Some inspiring preliminary explorations for the control procedures and solutions of ATSSS function has been illustrated in [8] and [9]. However, the ATSSS function currently only provides the low-layer rules based on coarse-grained traffic classification and limited network state information, hindering its adaptation to the complexity and the dynamics of 5G multi-RAT networks. Moreover, DRL-based algorithms have shown superiority for UE access and resource allocation in complex and the dynamic environments, as demonstrated in [6] and [7]. Therefore, it is worthwhile to further explore DRL-based algorithms for ATM in multi-RAT networks.

Furthermore, for better wireless resource management, it is important to effectively utilize various types of network information. In [4], user preferences, service requirements, network attributes and mobility trajectories are used for seamless vertical handover. However, some user and service information may not be available due to privacy concerns. Additionally, some RAT attributes are not common due to different vendors or standardization bodies developing them for various communication scenarios, leading to varying protocols, interfaces, configurations, management mechanisms, and performance metrics. These uncommon RAT attributes can hinder the scalability of ATM systems. Therefore, it is crucial to explore universal and

This work is supported partly by EU H2020 5G-CLARITY project (No 871428), UK funded Future Open Networks Research Challenge project: REASON, and China Scholarship Council.

representative network parameters for ATM system to facilitate interoperability and compatibility between different RATs.

In light of the imperative demand for ATM and the shortcoming of the current solutions, this paper reports the development of an intelligent ATM system for multi-RAT networks that provides a unified solution for better access, whether it is achieved by a single link or aggregated multiple links, while also enabling seamless vertical handover. The main contributions of this work are summarized as follows:

1) Architecture Design: We design a highly scalable ATM-support network architecture that includes various features such as multi-RAT integration, multi-path transmission, telemetry collection, and near real-time model-training and real-time decision-making for IATM.
2) Scalable Algorithm: We propose an adaptive Deep Deterministic Policy Gradient (DDPG) based algorithm with parameter sharing to accommodate the varying number of UEs. Moreover, a set of network parameters that are universal across RATs is selected and utilized. These parameters can be easily collected by UEs, which reduces interaction overhead for UE-led real-time ATM decisions.
3) Corroborating Simulation: We construct a multi-RAT network using Network Simulator 3 (NS3), featuring validated 5G New Radio (NR) and WiFi 802.11 modules that accurately represent the complexity and dynamics of real wireless networks [10, 11]. Its live interactions with DRL algorithms are supported by standard interfaces.
4) Comparative Case-Study: We demonstrate quantitatively that our proposed algorithm outperforms existing heuristic-based ATSSS rules, achieving nearly a twofold increase in resource utilization efficiency, tripling link quality, and doubling throughput at the same cost. Furthermore, it is prove that our algorithm can rapidly adapt to dynamic network changes, such as variations in UE size or mobility.

The rest of the paper is organized as follows. Section II describes the Multi-RAT ATM system architecture. The system model and the optimization problem are given in Section III. Section IV introduces the DDPG based IATM algorithm. Section V presents the performance evaluation of our approach, followed by the concluding remarks and future directions in Section VI.

II. Design of Multi-RAT ATM System

In this section, we discuss the architecture design of the Multi-RAT ATM system, considering multi-RAT integration, multi-path transmission, telemetry collection, and the ATM workflow. We exemplify the multi-RAT architecture using an integrated WiFi and 5G NR network, as depicted in Fig. 1. This architecture is supported by a set of standard network interfaces, including N6, N2, N3, NWu, F2, Y1, Uu, and E2, while also incorporating the design principle of separating the Centralized Unit (CU) and Distributed Unit (DU) in 5G NR [12].

Various interworking technologies have been proposed for integrating WiFi with the mobile networks, such as LTE-WLAN Aggregation (LWA) and LTE-WLAN Radio Level Integration with IPsec Tunnel (LWIP). In 5G networks, Non-3GPP Interworking Function (N3IWF) has been standardized to associate the Non-3GPP access with the 5G Core (5GC) [12]. Our approach employs this, as illustrated in Fig. 1. Notably, the architecture is scalable, as any other integration technologies can be applied to converge the corresponding RATs.

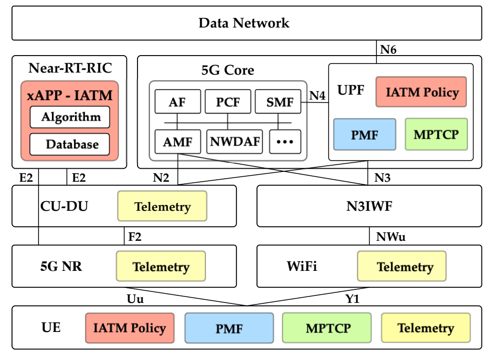


Fig. 1. Architecture of ATM system in hybrid 5G-WiFi network

Moreover, the multi-path transmission technologies, such as Multi-path TCP (MPTCP) and Multi-path QUIC (MPQUIC), allow UEs to transceive packets via multiple links concurrently, and the prevalent multi-home feature of UEs allows them to connect to multiple Radio Access Networks (RANs) simultaneously [13]. In our design, we utilize MPTCP to enable multi-homed UEs and User Plane Function (UPF) to transmit via the available 5G NR and WiFi access interfaces simultaneously.

The telemetry system comprises a set of telemetry modules located in both UEs and RANs, providing regular updates of the global signal state. The Performance Measurement Function (PMF) introduced by [9] measures the access performance, such as the round trip delay, while the Application Function (AF) provides application information. Hence, the Network Data Analytics Function (NWDAF) can collect and analyze the abundant information gathered by telemetry system, which is then shared with the RAN Intelligent Controller (RIC) to support the training of IATM algorithm.

For the ATM operation, our IATM algorithm is trained as an xAPP in the near real-time RIC (Near-RT-RIC), designed based on Open RAN framework [14]. The trained model undergoes periodic updates, and is processed by Policy Control Function (PCF) and Session Management Function (SMF) before being dispatched to the UE or UPF as the real-time IATM policy. The distributed execution of the IATM policy allows it to adapt to changing UE size. Moreover, the IATM policy dynamically allocates the traffic between several RATs based on current network state and user requirements to improve network resources utilization and Quality of Service (QoS).

III. System Model and Problem Formulation

A. System Model

We consider a multi-RAT network where M RATs coexist to assist U UEs in obtaining services from a remote server. The UEs and RATs are indexed by $\{1, \ldots, u, \ldots, U\}$ and $\{1, \ldots, m, \ldots, M\}$, respectively. UEs move independently and generate service requests randomly, which can be served by several RANs simultaneously or separately thanks to multi-path transmission. Packets are assumed to be transmitted in a weighted round-robin pattern across all RATs, with each RAT independently using its own rate control algorithm to schedule packets transmitted over it. Let $H_u^m(t)$ and $D_u^m(t)$ indicate the amount of data transmitted by the server and received by UE u via RAT m during time slot t, respectively. Thus, the transmission and reception rate via all RATs can be represented by $H_u(t)$ and $D_u(t)$. We assume that UE u requires a fixed transmission rate $H_u(t)$, and α_u^m denotes the diversion weight of the sub-flow via RAT m. Hence, the transmission rate $H_u^m(t)$ via RAT m can be expressed as $H_u^m(t) = \alpha_u^m H_u(t)$. Besides, the reception rate $D_u^m(t)$ is influenced by the packet loss and corruption due to network conditions. The total service time for all served requests of UE u is denoted as T_u, and the capacity of RAT m is denoted as B^m, which is unfixed based on the selected Modulation and Coding Scheme (MCS) and its physical bandwidth. Thus, the average system transmission and reception rates $\overline{H}$ and $\overline{D}$ can be respectively expressed as follows:

$$\overline{H} = \frac{1}{U \sum\limits_{u \in U} T_u} \sum_{u \in U} \sum_{m \in M} \sum_{t \in T_u} H_u^m(t) \tag{1}$$

$$\overline{D} = \frac{1}{U \sum\limits_{u \in U} T_u} \sum_{u \in U} \sum_{m \in M} \sum_{t \in T_u} D_u^m(t) \tag{2}$$

where $\overline{D}$ is also known as the average throughput. The average cost could be expressed as:

$$\overline{C} = \frac{1}{U \sum\limits_{u \in U} T_u} \sum_{u \in U} \sum_{m \in M} \sum_{t \in T_u} \lambda^m H_u^m(t) \tag{3}$$

where λ^m is the adjustable cost ratio coefficient of each RAT, which can be tailored to fit specific cost definitions.

The multi-RAT network is modeled in NS3, which operates at the packet level, allowing for a more detailed modeling of network performance and behavior [15]. Accurate and valid 5G NR and WiFi 802.11ax modules, which adhere to the specifications defined by 3GPP Release-15 NR and IEEE 802.11 standard, are used for 5G and WiFi network modeling, respectively [10, 11, 15, 16]. These modules include physical layer and Media Access Control (MAC) layer models, propagation loss and delay models, packet error models, and rate control algorithms. Besides, UEs' mobility is modeled using the Gauss Markov mobility model, which provides near-realistic movement patterns [16]. More information about the NS3 modeling is presented in Table I.

B. Problem Formulation

With the objective to maximize the average system throughput per cost, we formulate the optimization problem as follows:

$$\mathcal{P}: \quad \max \quad \frac{\overline{D}}{\overline{C}} \tag{4a}$$

$$\text{s.t.:} \quad C1: \quad \sum_{u \in U} D_u^m(t) \leq B_m, \quad \forall m \tag{4b}$$

$$C2: \quad D_u^m(t) \leq H_u^m(t), \quad \forall u, m \tag{4c}$$

where the Constraint 4b refers to the bandwidth limitation of each link, and the Constraint 4c indicates that the amount of data received by UE u should be equal to or less than the amount of data transmitted by the server.

The non-convex nature of this problem and the highly dynamic and complex network conditions make it computationally expensive using conventional methods [17]. Therefore, we will reformulate the problem to suit the DRL-based solution.

IV. DDPG Based Multi-RAT IATM Framework

A. MDP Modeling

The Markov decision process (MDP) representation is required to solve the non-convex problem using a DRL-based solution. The typical elements of the MDP are defined as follows.

State Space is denoted as $\mathcal{S}$. Here, we define any state $s \in \mathcal{S}$ of UE u at the start of time slot t as:

$$s_u(t) = [p_u^g(t), n_u^g(t), n_u^w(t), d_u^g(t), d_u^w(t)] \tag{5}$$

where p_u^g, n_u^g, and n_u^w correspond to 5G Reference Signal Received Power (RSRP), 5G Signal to Interference plus Noise Ratio (SINR), and WiFi SINR sensed by UE u. Plus, d^g and d^w represent one-way delay (OWD) from the remote server to UE u via 5G and WiFi links, respectively. Remarkably, only telemetry information from the requesting UE is used as state information, which effectively reduces the overhead during the distributed ATM decision-making phase.

Action Space, denoted as $\mathcal{A}$, represents the diversion weight α_u^m of the split sub-flow via RAT m. As only two RATs are considered as example, the continuous action $a \in \mathcal{A}$ of UE u can be represented by a single action value:

$$a_u(t) \in [0, 1] \tag{6}$$

When $a_u(t) = 1$ or $a_u(t) = 0$, all the traffic is exclusively transmitted through either 5G or WiFi link. For $a_u(t) \in (0, 1)$, the traffic is split between the two RATs, with $a_u(t)$ and $1 - a_u(t)$ representing the percentage of traffic transmitted to 5G and WiFi links, respectively, during the given time slot. Notably, by dynamically adjusting the diversion weight α_u^m in real time, the multi-path transmission can provide smoothly seamless handover and better session continuity.

Reward Function, as the evaluator of the action $a_u(t)$, is designed to maximize the throughput per cost for each UE in every time slot according to the optimization objective $\mathcal{P}$:

$$r_u(t) = \frac{D_u^g(t) + D_u^w(t)}{(H_u^g(t) + H_u^w(t))(\lambda^g a_u(t) + \lambda^w(1 - a_u(t)))^\rho} \tag{7}$$

where D_u^g, H_u^g, D_u^w, and H_u^w stand for the amount of data received and transmitted via 5G or WiFi, respectively. λ^g and λ^w are the cost ratio coefficients of 5G and WiFi, respectively. $\rho \in [0, 1]$ is a smoothing coefficient of the reward function for better training performance. When $\rho = 1$, the denominator and numerator of (7) are equal to the cost sum $C_u(t)$ and throughput sum $D_u(t)$ of all RATs, respectively. These physical meanings are not ambiguous, as the trend of denominator is not affected by ρ. This reward function is continuous within $[0, 1]$.

Algorithm 1. DDPG based IATM Algorithm

Initial critic network $Q(s,a|\theta^Q)$ with weights θ^Q
Initial actor network $\mu(s|\theta^\mu)$ with weights θ^μ
Initial target network Q' with weights $\theta^{Q'} \leftarrow \theta^Q$
Initial target network μ' with weights $\theta^{\mu'} \leftarrow \theta^\mu$
Initial replay buffer $\mathcal{M}$
for episode = 1…E **do**
  Initialize OU noise $\mathcal{N}$ for action exploration
  while $\sum_{u \in U} T_u \leq \mathcal{T}$ **do**
    while UE u requests for service **do**
      while Service is not completed **do**
        Collect state $s_u(t)$ of UE u
        Choose action $a_u(t)$ using (8)
        Send action $a_u(t)$ to UE u for execution
        Receive next state $s_u(t+1)$ and reward $r_u(t)$
        Store $(s_u(t), a_u(t), r_u(t), s_u(t+1))$ in $\mathcal{M}$
        Randomly sample a batch of N tuples from $\mathcal{M}$
        for i = 1…N **do**
          Compute TD target $y(i)$ using (9)
        end for
        Update critic network using (10)
        Update actor network using (11)
        Update target networks using (12)
      end while
    end while
  end while
end for

B. DDPG Based Multi-RAT IATM Algorithm

The DDPG algorithm is chosen due to the continuity of the defined action space. It is an actor-critic framework, where the actor part performs as a policy to select actions based on current state using state value function $\mu(s) = E[R(t)|s(t) = s]$, while the critic part evaluates the selected action using the state-action value function $Q(s|a) = E[R(t)|s(t) = s, a(t) = a]$ and guides the actor to improve its policy accuracy. The optimal policy $\mu^*(s) = \arg\max_a Q^{\mu^*}(s,a)$ aims to achieve a maximum long-term accumulated reward $R(t) = E(\sum_{t\in T} \gamma r(t))$, where $\gamma \in [0,1]$ is the discount factor and $\mathcal{T}$ is the maximum iteration times of each episode. An episode ends when $\sum_{u\in U} T_u \geq \mathcal{T}$. Besides, assuming regular and similar behavior patterns of UEs, their policies can be trained more efficiently using parameter sharing [18]. The proposed algorithm will be introduced in detail, and its pseudo-code is summarized in Algorithm 1.

In our DDPG based IATM algorithm, we firstly initialize a actor network $\mu(s|\theta^\mu)$, a target actor network $\mu'(s|\theta^{\mu'})$, a critic network $Q(s,a|\theta^Q)$, a target critic network $Q'(s,a|\theta^{Q'})$ and a replay buffer $\mathcal{M}$. In each episode, when an idle UE u makes a new request, the agent receives the observation $s_u(t)$ from the UE's telemetry module and selects action $a_u(t)$ according to:

$$a_u(t) = clip(\mu(s_u(t)|\theta^\mu) + \mathcal{N}(t), a_{Low}, a_{High}) \tag{8}$$

where $\mathcal{N}$ is the Ornstein–Uhlenbeck (OU) noise. Next, the traffic is proportionally divided into 5G and WiFi flows according to $a_u(t)$ within the time slot. If the service is incomplete after the given time slot, the new observation state $s_u(t+1)$ and the reward $r_u(t)$ are collected. Then, the 4-tuple $(s_u(t), a_u(t), R_u(t), s_u(t+1))$ is saved to the buffer $\mathcal{M}$. When one service is completed, the UE will wait for a random time before proposing a new request, and the process repeats until the maximum number of serving times slots $\mathcal{T}$ of the episode is reached. Notably, to eliminate the asynchronous effects of the random UE requests, a buffer indexed by the UE identifiers (IDs) is used. This enables each UE's MDP to be processed independently, without being impacted by the requests timing or order from other UEs.

Besides, the DDPG policy optimization process in [19] is adopted in Algorithm 1 and performed concurrently with experience collection process. Using a mini-batch of N experience tuples $(s(i), a(i), r(i), s(i+1))$ sampled from the buffer $\mathcal{M}$, the Temporal Difference (TD) target value is calculated as:

$$y(i) = r(i) + \gamma Q'(s(i+1), \mu'(s(i+1)|\theta^{\mu'})|\theta^{Q'}) \tag{9}$$

Then, gradient descent is applied to minimize the critic loss:

$$\nabla_{\theta^Q} \frac{1}{|N|} \sum_i (Q(s(i), a(i)|\theta^Q) - y(i))^2 \tag{10}$$

The actor part updates with one-step of gradient ascent:

$$\nabla_{\theta^\mu} \frac{1}{|N|} \sum_i (Q(s(i), \mu(s(i)|\theta^\mu)|\theta^Q) \tag{11}$$

In addition, the weights of target critic and actor networks are renewed by soft update:

$$\theta^{Q'} \leftarrow \tau\theta^Q + (1-\tau)\theta^{Q'} \tag{12a}$$

$$\theta^{\mu'} \leftarrow \tau\theta^\mu + (1-\tau)\theta^{\mu'} \tag{12b}$$

where $\tau \in (0,1)$ is the soft updating factor.

Using this framework, the trained IATM model can select actions with maximum expected rewards, leading to an optimal policy that resolves the formulated optimization problem $\mathcal{P}$.

V. Simulation and Performance Evaluation

In this section, we introduce the setup of the NS3-Gym simulation system, as well as the performance evaluation of the proposed ATM system.

A. Simulation System Setup

Our simulation is built with NS3-Gym, which combines NS3 for network simulation and Open-AI Gym for DRL training [20]. Fig. 2 shows the architecture of our NS3-Gym simulation platform, where actions from DRL agent, and states and rewards from network simulation can be exchanged via standard interfaces in real time. Remarkably, the built-in functions accurately simulate network attributes and transmission performance due to the utilization of the validated 5G NR Lena and WiFi modules [10, 11]. Besides, the 5G NR adopts a error model-based adaptive MCS model according to Channel Quality Indicator (CQI) feedback.

In NS3 simulation, a variable number of UEs are placed within a 70-meter squared area. The UEs are multi-homed and MPTCP-enabled, allowing they to access services via both 5G and WiFi in a multi-path manner. The cost ratio coefficient of

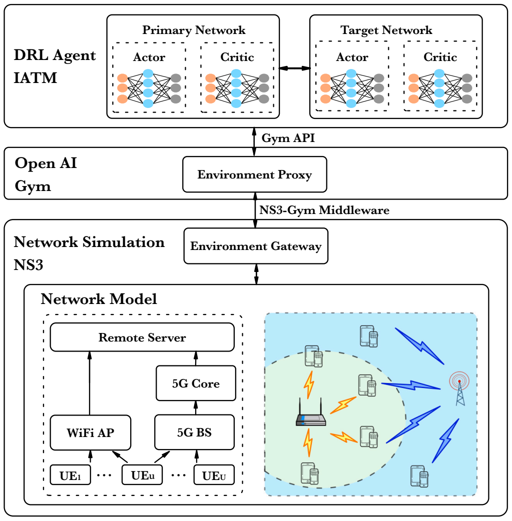


Fig. 2. Architecture of NS3-Gym simulation platform

5G and WiFi are initially set to $\lambda^g = 7$ and $\lambda^w = 1$, respectively, but can be adjusted based on the specific cost definition. Besides, only one WiFi AP and 5G Base Station (BS) are used, as each UE can connect to only one AP per RAT using the single interface at a time, and the handover process within the same RAT can also be reflected by the variation of telemetry data, such as SINR. Thus, this simulation scenario are still sufficient to validate the ATM system performance. The NS3 simulation models and parameters used for UEs, services, 5G NR, and WiFi are shown in Table I, partly following [2] and [4].

The IATM algorithm is configured with an episode length of 1500 iterations and a batch size of 128. The learning rates μ for the actor and critic network are given as 10^4 and 10^5, respectively. The target network update coefficient τ and the reward discount factor γ are set to 0.001 and 0.97, respectively, while the reward smoothing coefficient ρ is 0.25. Fig. 3a presents the convergence of average reward, actor loss, and critics loss of our algorithm, which have been scaled and shifted to the range of $[0, 1]$ according to the formula shown in the legend.

To ensure general performance, all results here are the average outputs from 10 repeated tests with different random seeds, and the tests are performed with 5, 10 and 15 UEs to represent different network congestion levels.

B. Numerical Result

To examine the performance of the IATM algorithm, we compare it with two traditional ATSSS traffic steering modes namely Active-Standby and Load-Balance, as described in [8]. In the Active-Standby mode, WiFi and 5G RATs are defined as active and standby access, respectively. All traffic is transmitted via the active access until it becomes unavailable, after which traffic is transferred to the standby access. The Load-Balance mode splits traffic between 5G and WiFi access with a weight factor of 2:1, which is determined based on the approximate ratio of their maximum capacities tested in our experiments. A set of metrics are defined for performance evaluation.

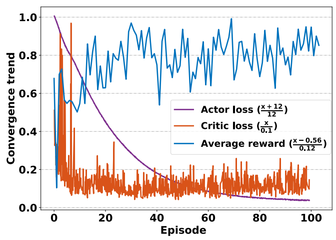
Firstly, the average reception rate $\overline{K}$ is defined as the average percentage of received data to transmitted data, denoted as $\overline{K} = \frac{\overline{D}}{\overline{H}}$. A higher reception rate means greater throughput under the same network capacity and transmitted data volume, indicating higher utilization efficiency of network resources. Fig. 3b shows that our proposed algorithm outperforms the Active-Standby and Load-Balance methods in terms of a higher and more stable average reception rate $\overline{K}$ as network congestion level changes. With 15 UEs, our algorithm achieves 45% and 70% improvements over the other two methods, respectively.

TABLE I
NS3 SIMULATION SETTINGS

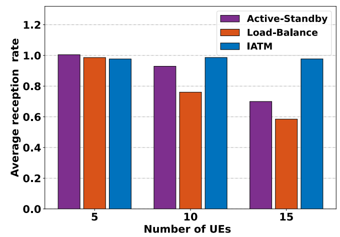
	Parameter	Value
UE	Number	5, 10, 15
	Transmission power	23 dBm
	Noise figure	9 dB
	Antenna height	1.5 m
Service	Packet size	1000 Bytes
	Packet count	725-7500
	Data rate	4096 Kbps
	Loss rate threshold	30%
5G NR	Central frequency	700 MHz
	Channel bandwidth	20 MHz
	Transmission power	49 dBm
	Noise figure	5 dB
	Antenna height	35 m
	Channel model	ThreeGppChannelModel
	Propagation loss model	UMi_StreetCanyon_LoS
	Shadow fading	4 dB
	Error model	NrEesmIrT1
WiFi	Standard	IEEE 802.11ax
	Central frequency	5.15 GHz
	Channel bandwidth	20 MHz
	Channel number	36
	Transmission power	23 dBm
	Noise figure	4 dB
	Antenna height	2.5 m
	Antenna type	Isotropic antenna
	Propagation loss model	$30.2 + 36.7log(d)$
	Fading model	Nakagami
	MCS index	3
	Guard interval	3200 ns

Besides, the average QoS satisfaction rate is defined as the average percentage of QoS satisfaction times to the total QoS checks, expressed as $\overline{L} = \frac{\sum_{u \in U} \sum_{t \in T_u} L_u(t)}{\sum_{u \in U} T_u}$, where L_u is the QoS indicator for UE u at time slot t. The QoS satisfaction is measured by the packet loss rate with a threshold of 30%. When the packet loss rate exceeds the threshold, $L_u = 0$; otherwise, $L_u = 1$. The proposed algorithm maintains excellent transmission quality in terms of QoS satisfaction rate $\overline{L}$ as the congestion level increases, as illustrated in Fig. 3c. Our algorithm also achieves a QoS satisfaction rate almost three times higher than the Active-Standby and Load-Balance methods under congested conditions.

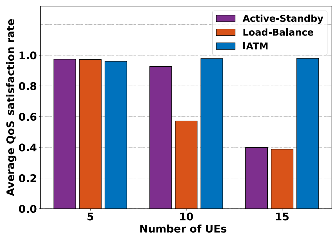
Furthermore, the average throughput per cost is defined as $\overline{P} = \frac{\overline{D}}{\overline{C}}$, based on the optimization objective $\mathcal{P}$. Fig. 3d illustrates that the IATM algorithm outperforms two traditional algorithms in terms of cost-effectiveness throughput regardless of network congestion level. Despite their fixed policy avoiding raising more 5G traffic expenditure, our model achieves a higher throughput with the same cost. Notably, our algorithm improves the throughput by more than double that of the Active-Standby policy, and nearly three times that of the Load-Balance mode



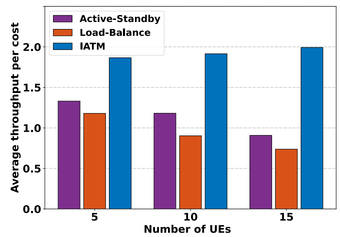
(a) Convergence of IATM algorithm



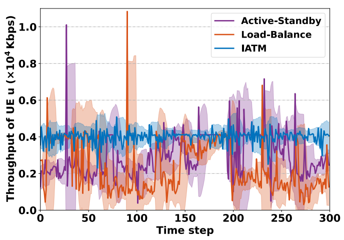
(b) Average reception rate $\overline{K}$



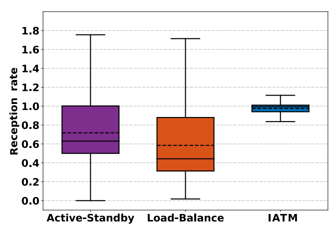
(c) Average QoS satisfaction rate $\overline{L}$



(d) Average throughput per cost $\overline{P}$



(e) Variation of throughput $D_u(t)$



(f) Stability of reception rate $K_u(t)$

Fig. 3. Simulation results include the convergence process of IATM algorithm in Fig. 3a, and the transmission performance achieved by IATM algorithm compared to two current ATSSS steering modes in Fig. 3b, 3c, 3d, 3e, and 3f.

at the same cost.

Lastly, the transmission stability is crucial for use experience. We collect the throughput sum $D_u(t)$ of UE u for 300 service time slots using each ATM method when the UE size is 15. In Fig. 3e, a rolling average with a window size of 10 time-steps was used to smooth the collated data and obtain a curve with less fluctuations, and the 95% confidence interval is presented as a shaded region around the curve. As demonstrated in Fig. 3e, our algorithm achieves higher and more stable throughput. This is intuitively revealed in the box plot of Fig. 3f, where the dispersion of the reception rate $K_u(t) = \frac{D_u(t)}{H_u(t)}$ is represented by the height of the box and distance between the outer lines. The IATM policy shows a much smaller box and shorter range of outer lines, indicating a more stable transmission performance compared to the other two methods.

VI. Conclusion and Future Work

In this paper, we have proposed an intelligent ATM system for 5G-WiFi hybrid multi-RAT network to improve the system throughput cost-effectively. Firstly, an concise multi-RAT network architecture was designed to achieve network integration, multi-path transmission, and telemetry collection for ATM function performing. Our DDPG based IATM algorithm is trained in near real-time RIC to maximize system throughput while minimizing costs, and is subsequently deployed in the UE and UPF for real-time execution and to support dynamic UE size. Besides, our approach reduces the overhead by requiring only the UE-side information during execution. Furthermore, using the NS3-Gym simulation, we have demonstrated that our solution is vastly superior in radio resource utilization efficiency, service quality, cost effectiveness and connection stability, particularly in congested networks. Our ATM system maximizes the advantages of the Non-3GPP and 5G mobile networks integration and contributes to future multi-RAT networks with enhanced connections at a lower cost.

In future work, we will incorporate cost modeling for RATs and investigate the performance of our algorithm under varying cost ratio coefficients. Moreover, we will explore more scenarios with multiple types of RATs, multiple APs for each RAT, and more service types to expand the scope of our research. The exploration of the IATM management in testbed is undergoing.

References

[1] O. Galinina, A. Pyattaev, S. Andreev, M. Dohler, and Y. Koucheryavy, "5G multi-RAT LTE-WiFi ultra-dense small cells: Performance dynamics, architecture, and trends," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 6, pp. 1224–1240, 2015.

[2] A. Roy, P. Chaporkar, A. Karandikar, and P. Jha, "Online radio access technology selection algorithms in a 5G multi-RAT network," *IEEE Transactions on Mobile Computing*, 2021.

[3] M. A. Habib, H. Zhou, P. E. Iturria-Rivera, M. Elsayed, M. Bavand, R. Gaigalas, Y. Ozcan, and M. Erol-Kantarci, "Hierarchical reinforcement learning based traffic steering in multi-RAT 5G deployments," *arXiv preprint arXiv:2301.07818*, 2023.

[4] H. Tong, T. Wang, Y. Zhu, X. Liu, S. Wang, and C. Yin, "Mobility-aware seamless handover with MPTCP in software-defined HetNets," *IEEE Transactions on Network and Service Management*, vol. 18, no. 1, pp. 498–510, 2021.

[5] H. Alshaer and H. Haas, "Software-defined networking-enabled heterogeneous wireless networks and applications convergence," *IEEE Access*, vol. 8, pp. 66 672–66 692, 2020.

[6] M. S. Allahham, A. A. Abdellatif, N. Mhaisen, A. Mohamed, A. Erbad, and M. Guizani, "Multi-agent reinforcement learning for network selection and resource allocation in heterogeneous multi-RAT networks," *IEEE Transactions on Cognitive Communications and Networking*, vol. 8, no. 2, pp. 1287–1300, 2022.

[7] Alwarafy, Abdulmalik and Ciftler, Bekir Sait and Abdallah, Mohamed and Hamdi, Mounir and Al-Dhahir, Naofal, "DRL-Based joint RAT association, power and bandwidth optimization for future HetNets," *IEEE Wireless Communications Letters*, 2022.

[8] 3GPP, "Study on access traffic steering, switch and splitting support in the 5G System (5GS) architecture," 3rd Generation Partnership Project (3GPP), Technical report (TR) 23.793, 12 2018, version 16.0.0.

[9] 3GPP, "5G System; Access Traffic Steering, Switching and Splitting (ATSSS)," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 24.193, 06 2021, version 16.4.0.

[10] K. Koutlia, B. Bojovic, Z. Ali, and S. Lagén, "Calibration of the 5G-LENA system level simulator in 3GPP reference scenarios," *Simulation Modelling Practice and Theory*, p. 102580, 2022.

[11] NS-3 Wi-Fi module testing documentation. [Online]. Available: https://www.nsnam.org/docs/models/html/wifi-testing.html

[12] 3GPP, "System architecture for the 5G system," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 23.501, 09 2020, version 16.6.0.

[13] M. Wasserman and P. Seite, "Current practices for multiple-interface hosts," Internet Engineering Task Force, RFC 6419, 2011.

[14] M. Polese, L. Bonati, S. D'Oro, S. Basagni, and T. Melodia, "Understanding O-RAN: Architecture, interfaces, algorithms, security, and research challenges," *IEEE Communications Surveys & Tutorials*, 2023.

[15] Network simulator-3. [Online]. Available: http://code.nsnam.org/ns-3-dev/

[16] T. Camp, J. Boleng, and V. Davies, "A survey of mobility models for ad-hoc network research," *Wireless communications and mobile computing*, vol. 2, no. 5, pp. 483–502, 2002.

[17] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[18] J. K. Gupta, M. Egorov, and M. Kochenderfer, "Cooperative multi-agent control using deep reinforcement learning," in *Autonomous Agents and Multiagent Systems: AAMAS 2017 Workshops, Best Papers, São Paulo, Brazil, May 8-12, 2017, Revised Selected Papers 16*. Springer, 2017, pp. 66–83.

[19] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," *arXiv preprint arXiv:1509.02971*, 2015.

[20] P. Gawłowicz and A. Zubow, "NS-3 meets OpenAI Gym: The playground for machine learning in networking research," in *ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM)*, November 2019.