RESEARCH ARTICLE

Joint Q-Learning Based Resource Allocation and Multi-Numerology B5G Network Slicing Exploiting LWA Technology

NOHA A. ELMOSILHY[1], **MAHMOUD M. ELMESALAWY**[2], (Member, IEEE),
IBRAHIM I. IBRAHIM[2], **AND AHMED M. ABD EL-HALEEM**[2], (Member, IEEE)

[1]Electronics and Communications Department, Canadian International College, Cairo 11865, Egypt
[2]Electronics and Communications Engineering Department, Helwan University, Cairo 11795, Egypt

Corresponding author: Mahmoud M. Elmesalawy (melmesalawy@h-eng.helwan.edu.eg)

ABSTRACT The emergence of the sixth generation (6G) era has highlighted the importance of Network Slicing (NS) technology as a promising solution for catering the diverse service requests of users. With the presence of a large number of devices with different service requests and since each service has different goals and requirements; efficiently allocating Resource Blocks (RBs) to each network slice is a challenging task to meet the desired Quality of Service (QoS) standards. However, it is worth noting that the majority of research efforts have primarily concentrated on cellular technologies, leaving behind the potential benefits of utilizing unlicensed bands to alleviate traffic congestion and enhance the capacity of existing LTE networks. In this paper we propose a novel idea by exploiting LTE-WLAN Aggregation technology (LWA) in Multi-Radio Access Technology (RAT) Heterogeneous Networks (HetNet), aiming to solve radio resource allocation problem based on the Radio Access Network (RAN) slicing and 5G New Radio (NR) scalable numerology technique. A joint optimization problem is proposed by jointly finding an efficient resource allocation ratio for each slice in each Base Station (BS) and by finding the optimum value of scalable numerology with the objective of maximizing users' satisfaction. In order to solve this problem, a novel three-stage framework is proposed which is based on channel state information as a pre-association stage, Reinforcement Learning (RL) algorithm as finding the optimum value of slice resource ratio and scalable numerology, and finally Regret Learning Algorithm (RLA) as users' re-association phase. Furthermore, a comprehensive performance evaluation is conducted against different baseline approaches. The simulation results show that our proposed model balances and achieves improvement in users' satisfaction by deploying the proposed Multi-RAT Het-Net architecture that leverages LWA technology.

INDEX TERMS HetNet, LWA, multi-RAT, network slicing, numerology, regret matching, WLAN.

I. INTRODUCTION

The Sixth Generation (6G) is required to support more challenging scenarios than the Fifth Generation (5G), including the presence of a large number of devices with different service requirements. Starting from 5G and Beyond 5G (B5G); a New Radio (NR) supports heterogeneous services which are classified into Massive Machine-type Communication (mMTC), Ultra- Reliable Low Latency (URLLC),

The associate editor coordinating the review of this manuscript and approving it for publication was Nafees Mansoor.

and Enhanced Mobile Broadband (eMBB) services [1]. Each service has different requirements in a common wireless network infrastructure, thus Network Slicing (NS) technology is introduced. An End-to-End (E2E) network slice is divided into the Radio Access Network (RAN), transport, and Core Network (CN) sub-network slices in between the end user devices [2]. The NS which is based on Software-Defined Networks (SDN) and Network Function Virtualization (NFV) [3] presents the ability to divide the network into several logical networks; each logical network can support different services with different requirements. These logically independent

networks are called network RAN slices [4], [5]. Each RAN slice can support a service with diverse requirements, for example, a RAN slice for URRLC service requires a very low latency and high reliability as in autonomous driving, remote surgery, etc. On the other hand, a slice for eMBB service requires high data rate to support services as video streaming and virtual reality, etc. The big challenge here is how to satisfy the required Quality of Service (QoS) in each slice with these diverse requirements and with the existing physical infrastructure's limited resources. As each Base Station (BS) can consist of several slices and each slice support **s** different service **s** and serve several users, an efficient resource allocation mechanism is necessary to manage the Resource Blocks (RBs) allocation efficiently between these slices to satisfy the heterogeneous users' demands and services. Disposing of an efficient resource slicing framework that can satisfy each service requirements exposed in eMBB and URLLC services is a major challenge in 5G and B5G.

Moreover, to facilitate supporting service diversity requirements, 3rd Generation Partnership Project (3GPP) has standardized multi-numerology and a mini-slot approach to enhance the adaptability of physical layer in NR. Unlike Long Term Evolution (LTE), which supports only a single numerology in the downlink, 5G NR allows a scalable numerology technique that provides different subcarrier spacing, symbol durations and a number of symbols per time slot [6]. Exploiting the 3GPP multi-numerology technology will add the foundation of a flexible Orthogonal Frequency-Division Multiple Access (OFDMA) frame structure. Where, the downlink transmission is divided into frames, each frame is 10ms divided into ten sub-frames each is a 1ms duration. In NR, the number of slots in each sub-frame is scalable and can be evaluated using numerology μ where each time slot duration is $2^{-\mu}$. The division of time slots into mini-slots, without the need to wait for slot boundaries, allows for the quick delivery of low-latency payloads [6]. Efficiently optimizing the scalable numerology value for each slice to cater the scheduling needs of both eMBB and URLLC users in multi-user scenarios, by selecting suitable subcarrier spacing and symbol length values that meet the requirements of each service while allocating resources appropriately, is still a significant challenge.

Furthermore, since previous studies have primarily focused on developing a balanced and efficient approach to address the co-existence issue between eMBB and URLLC services [7], [8], [9], [10], [11], [12], there is a requirement for a Reinforcement Learning (RL) algorithm to obtain a globally optimal solution for selecting the most suitable scalable numerology value for each slice and allocating slice resources in the problem of RAN slicing. The RL is a methodology that aims to maximize the learning of utility functions in the context of dynamic resource allocation decisions. It is particularly effective in tackling unknown stochastic dynamic environments with either significantly large or continuous state spaces. Machine Learning (ML) enabled 5G network architecture has recently been applied

and investigated to resource allocation of slices as in [13], [14], [15], [16], and [17]. However, with the increase of users' service demands and needs to satisfy each user's service requirement, the LTE limited resources will still impact on user's satisfaction to meet each service diver requirement. Thus, exploiting unlicensed band to leverage the traffic load and enhance the LTE capacity to maintain the required QoS in each slice will be considered as an efficient solution. Exploiting both the licensed and unlicensed spectrum was the motivated key to standardizing the LTE-Wireless Local Area Network (WLAN) Aggregation (LWA) technology by 3GPP in Release 13 [18].

In LWA, the data traffic is aggregated at the RAN level, where the eNodeB (eNB) resolves to steer the data traffic either on a switched bearer or split bearer. In the split bearer, the user is considered as both an LTE user and a WiFi user; thus, the user can access WiFi channel with channel contention probability. Comparing LWA technology to LTE-Unlicensed (LTE-U) technology that can utilize also the unlicensed band; Although LTE-U can enhance network capacity by using the unlicensed band, however the LTE-U must apply techniques for fair coexistence as Clear Channel Assessment (CCA) and Listen Before Talk (LBT) techniques to avoid any harm on WiFi co-existence users. LWA not only increases system capacity but also introduces network deployment cost reduction compared to LTE-U that need modification on eNB design out of 3GPP standard [19].

Moreover, in LWA-based Multi-Radio Access Technology (RAT) Heterogeneous Networks (HetNet), a user who can access both LTE and WiFi RATs has the capability to choose between three transmission modes defined as LTE mode, where user can associate with LTE-BS using LTE RAT, or WiFi mode, where user associates with WiFi-BS access WLAN RAT or LWA aggregation mode where user can associates with both LTE and WiFi BSs. Therefore, a new user association approach must be formulated to efficiently associate users with proper BS that can guarantee user's Service Level Agreement (SLA) for requested service.

In this work, we use Regret Learning Algorithm (RLA) to efficiently associate user with proper BS that fulfills the user's requested service requirements in Multi-RAT HetNet network architecture considering LWA technology. The RLA [20] almost converges to a set of correlated equilibrium, where each user is an optimal response to the actions of other users and also a response to its environment.

Motivated by the previous observation and to overcome the aforementioned challenges, this paper proposes a novel idea of deploying a Multi-RAT Het-Net, leveraging the LWA technology and 5G NR scalable numerology aiming to maintain each user's satisfaction which is defined in achieving SLA of each requested service. Specifically, our main contributions are:

- A radio resource allocation scheme regarding the eMBB and URLLC slices in Multi-RAT HetNet architecture is studied. The Multi-RAT Het-Net architecture consists of an LTE Macro cell and Integrated Small Cells (ISCs) that

support three different transmission modes (LTE mode, WiFi mode, or LWA mode for aggregation). To the best of our knowledge, such architecture that combines LWA technology jointly with LTE-network slicing supporting eMBB and URLLC services has not been studied.

- A user association problem is formulated and defined as a joint optimization problem with the objective of maximizing user's satisfaction. The joint optimization problem aims to guarantee each user's SLA.
- We decompose our optimization problem as two sub-problems; the first sub-problem involves the allocation of slice resources (Slice Ratio) and determining the optimal scalable numerology value. To address this, we propose a Reinforcement Learning-based algorithm (Q-Learning) that dynamically adjusts the resource allocation per base station and identifies the ideal scalable numerology value for each slice. This approach ensures the maintenance and fulfillment of user satisfaction requirements in eMBB and URLLC service.
- The second sub-problem is defined as a user-association problem. A Regret Matching Learning algorithm is proposed with the aim of efficiently solve users' association problem in Multi-RAT Het-Net environment with unified user utility function defined as user's satisfaction. To the best of our knowledge, this paper is the first to use Regret Learning algorithm in integrated network architecture considering LWA technology.
- We study and analyze the computational time complexity of proposed algorithm compared to other RL-based algorithm. In addition, we analysis optimality and convergence of proposed algorithm.
- The performance of our proposed algorithm is compared against different network architectures as single RAT deployment, Multi-RAT deployment without LWA to validate our proposed algorithms.

The rest of the paper is organized as follows: Section II, represents Literature review, Section III represents a detailed description of the proposed system model, the Problem formulation and proposed framework are presented in Section IV. Section V of the paper introduces the proposed RL-based resource slicing algorithm, which outlines the methodology for optimizing resource allocation in network slicing. This is followed by Section VI, which presents the proposed Regret-Learning User Re-association algorithm. The simulation results are then presented in Section VII, providing a detailed analysis and evaluation of the proposed algorithms. Finally, the paper concludes in Section VIII.

II. LITERATURE REVIEW

Resource allocation in RAN slicing is a critical aspect, as it defines how the physical resources in each BS will be portioned among different slices. Several approaches have been proposed in literature considering different schemes and algorithms to address this challenge. One of these approaches is static resource allocation. In static resource allocation, the network resources are allocated to different slices based on a pre-defined scheme. In [21] and [22], the authors provide a comprehensive survey on various resource allocation schemes, including static resource allocation. They highlight the limitations of using static resource allocation in a multi service network. Additionally, in [23] and [24], the authors propose a static resource allocation scheme that takes into account the expected traffic load. While they consider the QoS requirements of each service, this approach still falls short in evaluating the dynamic changes in the network environment and user's traffic patterns. Consequently, it may not effectively adapt to real-time fluctuations and variations in the network. Other studies considered dynamic resource allocation, in multi-service network. In [7], the authors present a dynamic resource allocation scheme considering two different services, eMBB and URLLC. The authors proposed a joint resource allocation problem to satisfy the eMBB user data rate while other users requested URLLC service. The dynamic radio resource allocation in multi-service network also studied in [8], where the network was logically divided into low delay slice and high rate slice; however, the authors' main objective was to minimize delay tolerance. In [9], a risk-sensitive resource allocation formula is proposed aiming to allocate resources to URLLC traffic, while ensuring the eMBB service requirement. The literature [10], [11], [12] further explores the dynamic resource allocation problem in the presence of multiple slices. Specifically, the authors have recently directed their attention to the coexistence of URLLC and eMBB traffic within a RAN to fulfill each service's diverse requirement.

Based on this, adopting the 5G NR standard scalable numerology technique will enable an efficient fulfillment of diverse qualities like low latency. In [25], the authors proposed a scalable numerology approach aiming to solve scheduling problems for different user services. The authors tend to maximize user satisfaction while taking into consideration latency demand and data transmission. Also, a numerology-based resource allocation scheme is presented in [26]; the authors proposed an energy efficient algorithm for eMBB and URLLC services, the proposed algorithm is formulated as mixed-integer non-linear problem. The authors also studied the impact of URLLC users on eMBB users. In addition, authors in [27] applied NR scalable numerology technique to serve users who requested URLLC services. The authors aim to increase the minimum achievable data rate for eMBB users while applying fairness among URLLC and eMBB using a one-to-one matching game.

The aforementioned studies strive to strike a balance between efficient resource utilization and meeting the QoS requirements of co-existing slices. In order to obtain a global optimum solution for resource allocation in network slicing while finding an optimum scalable numerology value that efficiently fulfills the users' service requests, a RL approach must be applied. Recently, many researches have used RL to solve the resource allocation problem and decision- making in network slicing architecture.

The authors in [28], deployed NR scalable numerology technique and mini-slot-based transmission to support both eMBB and URLLC network slices, a RAN slicing problem is formulated and solved using a hierarchical RL framework. In [29], a deep reinforcement learning approach was proposed to allocate the URLLC traffic, the authors aim to maximize the eMBB data rate while considering the URLLC reliability constraint. In [30], the authors proposed an approach for slice resource management using a Mixed Integer Linear Program (MILP). This approach takes into account the concept of scalable numerology as well as the requirements for both latency and throughput. The proposed approach was solved using Deep Reinforcement Learning (DRL) algorithm. In [31], a resource allocation framework DRL-based for network slicing in a RAN is proposed with the aid of massive Multiple-Input Multiple-Output (MIMO). The proposed framework in [31] aims to maximize the resource allocation efficiency through slices and the user's quality of experience. In [32], an RL-based framework is proposed focusing mainly on eMBB aiming to solve resource allocation problems that are defined with Channel State Information (CSI) uncertainty. The authors model the CSI uncertainty by three methods: worst-case, probabilistic, and hybrid.

All the previous researches aim to address the challenges associated with efficiently allocating resources to accommodate the diverse requirements and co-existence of these two types of traffic (URLLC and eMBB). However, in high-load scenarios where the number of users with diverse service requests increases, the LTE-limited capacity may still face challenges even with the implementation of network slicing. Despite the benefits of network slicing in enabling resource isolation and customization for different services, there are inherent limitations to the capacity of LTE networks. These limitations can lead to congestion and degradation of performance when the network is heavily loaded. Based on this, exploiting both the licensed band and unlicensed band using LWA technology was the motivated key to leverage the traffic load and enhance the LTE capacity. In [33], a multi-RAT HetNet scenario is considered adopting LTE-WLAN aggregation technology. The authors proposed a user association algorithm that considered different RAT selections with the aim of maximizing total network throughput. The simulation results showed significant performance in enhancing the network's throughput as a result of exploiting the LWA technology.

Indeed, the presence of LWA technology introduces a multi-RAT environment. With LWA, users have the option to associate with LTE-RAT, WLAN-RAT or even aggregate both technologies simultaneously. A new user association approach must be formulated in order to efficiently associate users with proper RAT that will fulfill the user's service requirement. Many studies that adopted user association problem while considering LWA technology [34], [35], [36], [37], to the best of our knowledge, this paper is the first to use the Regret Learning algorithm to efficiently associate users with proper BS (RAT) that fulfill user's requested service
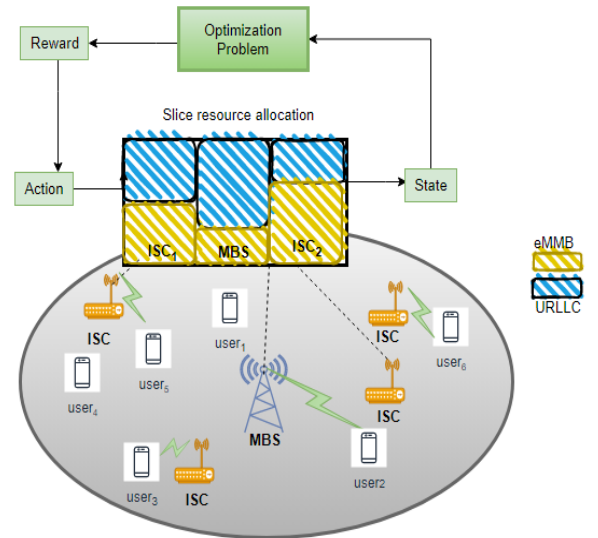

FIGURE 1. System model.

requirement in network slicing architecture considering LWA technology.

III. SYSTEM MODEL

In this paper, a Multi-RAT Het-Net is considered as shown in Fig.1, consisting of a Macro Base Station (MBS) and LTE-WLAN-ISCs overlaid under the coverage of MBS. All the ISCs support both RATs (LTE and WiFi) technology, which allow three access options can be denoted by three transmission modes of operation: M1 (LTE mode), M2 (WiFi mode), and M3 LTE-WLAN aggregation (LWA mode). Since ISC supports both LTE and WiFi RATs, for this we define a set that represents MBS and small cells (LTE and WiFi) denoted by $\mathcal{K} = \{0, 1, .., k .. , K\}$ where, $k = 0$ represents the MBS, followed by $k = \{1, 2, 3, \ldots K\}$ that represents small cells. Without the loss of generality, we define two subsets to denote each RAT, where the number of small LTE base stations (LBS) is denoted by L, while the number of WiFi-BSs is denoted by $\mathcal{W}\}$. The set of LBSs is denoted by $B = \{1, \ldots, L\}$, where $B \subset \mathcal{K}$. In addition, the set of WiFi-BSs is denoted by $\mathcal{W} = \{L + 1, L + 2, \ldots, K\}$ with cardinality $\mathcal{W} = K - L$, where $\mathcal{W} \subset \mathcal{K}$. Moreover, a set that represents ISCs is denoted by $= \{L, L + 1\}$, where $k = L$ denotes LBSs and $k = L + 1$ denotes WiFi-BSs. LTE RAT adopts the accessing scheme of OFDMA, while WiFi adopts the IEEE 802.11 Distributed Coordination Function (DCF) mechanism that is based on the Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) protocol for channel access [38].

Moreover, a set of slices $j \in \acute{j} = \{0, 1, 2, .., J\}$ is introduced to support RAN slicing in LTE BSs, which is defined in MBS and ISC-LTE-RAT. Noted that $j=0$, means WiFi mode (M2) is selected. Moreover, a number of users are uniformly distributed under the coverage of MBS and ISCs. In addition, as ISCs may operate on three different transmission modes; a set of users $u \in \mathcal{U} = \{1, 2, \ldots, U\}$ is defined, where the

TABLE 1. Table of notation.

Notation	Description
μ_j	Scalable Numerology
ΔF_j	Sub-carrier Spacing
T_{slot}^L	Time Slot Interval
$N_{k,j}^{RBmax}$	Maximum Number of RBs assigned for a slice
$N_{k,j}^{RB}$	Number of RBs assigned for a slice
α_j	Slice Resource Ratio
$R_{u,j,k}^L$	Achievable downlink data rate achieved by user u from LTE-BS
$\mathfrak{R}_{uk}^W$	Achievable downlink data rate achieved by user u from WiFi-BS
Rt_{uk}^{phy}	WLAN physical rate
U_j	Total number of users per slice j associated to BS $k \in \mathcal{K}/\mathcal{W}$
$\mathcal{U}_w$	Total number of users associated with WiFi-BS $k \in \mathcal{W}$
$R_{u,k}^{ISC}$	Achievable downlink data rate achieved by user u from ISC
$D_{u,j,k}^{Ach}$	Achieved Average Delay
$ST_{u,j,k}$	The satisfaction of user u in slice j associated with BS k
$\mathcal{R}^{eq}$	User's Service Request
r_t	Reward
$V^*(s)$	State value function
R_t	Cumulative Discounted Reward
$Q^*(s,a)$	Action value Function
$\beth$	Discount Factor
M^{epi}	Number of Episodes
$U_u(s^R)$	User's Utility Function
$\rho_u^t(z,y)$	Regret of player u is to choose strategy y instead of z at a time t
$D_u^t(z,y)$	Payoff for player u
$p_u^{t+1}(y)$	Probability distribution of player u chooses an action at time t
$\bar{z}_t$	Empirical Distribution

TABLE 2. 5G nr scalable numerology [6].

μ_j	ΔF_j	T_{slot}^L
0	15 kHz	1ms
1	30 kHz	0.5ms
2	60 kHz	0.25ms
3	120 kHz	0.125ms
4	240 kHz	62.5µs

various channel conditions. The NR frame structure introduced mixed numerology where each numerology represents different multicarrier modulation parameters which define different subcarrier spacing, Cyclic Prefix (CP) duration, and slot duration.

Based on the above, a scalable numerology set $\mathcal{M}_j = \{0, 1, 2, \dots 4\}$ as shown in Table 2. is considered, where $\mu_j = 0$, $\mu_j \in \mathcal{M}_j$ will be used as the eMBB slice, while $\mu_j = \{1, 2, 3, 4\}$ will vary to support URRLC slice. A variable μ_j will result in a variable sub-carrier spacing ΔF_j and time slot interval T_{slot}^L given by (1) & (2) respectively [6]:

$$\Delta F_j = \Delta F_0 \times 2^{\mu_j}, \tag{1}$$

$$T_{slot}^L = \frac{T_{slot_0}^L}{2^{\mu_j}}, \tag{2}$$

where, $\Delta F_0 = 15$kHz and $T_{slot_0}^L = 1$ms

On the other side, in order to support eMBB service, which requires a high data rate, we adopt 3GPP standardized LWA technology to maintain the user's QoS by allowing the user's data to be transmitted either on LTE (M1 mode), or WiFi (M2 mode), or LWA (M3 mode). The LWA technology will exploit the unlicensed band to support the LTE network in high-load scenarios. Adopting both NR scalable numerology and LWA technology can efficiently strike a balance between the diverse user service requests.

B. NETWORK SLICING MODEL
The network slicing main goal is to fulfill the QoS requirements of different network slices. To be able to achieve this, we need to consider three main components, first to consider the number of users' slice preferences, second to model eMBB slice requirement which is defined as rate and third to model URLLC slice requirement modeled as transmission delay.

1) SLICE RESOURCE ALLOCATION
We will start with the user's request slice preference phase, as the number of users' requests $\mathcal{R}^{eq}$ preference to a slice in a BS k will affect the number of logical resource blocks in each slice in each BS $k \in \mathcal{K}/\mathcal{W}$. Based on this, the maximum number of RBs assigned for each slice j from BS $k \in \mathcal{K}/\mathcal{W}$ can be calculated as follows:

$$N_{k,j}^{RBmax} = \frac{B}{\Delta F_j}, \tag{3}$$

where, B is BS $k \in \mathcal{K}/\mathcal{W}$ bandwidth. In addition, the number of RBs assigned for a slice j considering other slices allocated

number of users associated with slice j is denoted by U_j and the number of users associated with ISC WiFi-RAT is denoted by U_w. In addition, a table of notation is provided in Table 1.

A. SLICE DESCRIPTION
Our system model considers URLLC and eMBB services, without taking into account the mMTC service. Generally, the proposed system supports two services where each service provides a different aim, the URLLC services require low latency combined with high reliability, while eMBB services require high data rates combined with a moderate reliability level. To elaborate both services, a user u service's request, can be denoted by $\mathcal{R}^{eq}$. When user u requests $\mathcal{R}^{eq}$ a service requires a high data rate it falls under eMBB slice and when user u requests $\mathcal{R}^{eq}$ a service requires a minimum delay it falls under URLLC slice.

In order to support both services with their different requirements, our proposed model, adopts 5G NR frame structure [6]. The NR frame structure flexibly supports the heterogeneous services while allows the adaption to user's

resources in BS k is presented as:

$$N_{k,j}^{RB} = \alpha_j N_{k,j}^{RBmax}, \tag{4}$$

where, α_j is the ratio of RBs allocated to a slice which is dependent on other slices ratios and ranges $0 \le \alpha_j \le 1$. The value of α_j must give $N_{k,j}^{RB}$ an integer value. Moreover, if BS $k \in \mathcal{K}/\mathcal{W}$ supports two slices and if the number of RBs assigned for slice one can be calculated using (4), then the number of RBs assigned for the second slice can be calculated as follows:

$$N_{k,j',j' \neq j}^{RB} = (1 - \alpha_j) N_{k,j}^{RBmax}, \tag{5}$$

2) eMBB SLICE MODEL

In order to achieve eMBB slice requirement, we need to calculate the average SINR/SNR received by any user u from an MBS or ISC as (LBS &WiFi). For this, the average received SINR by user u from a MBS $k \in \mathcal{K}$ and $k = 0$ can expressed as:

$$SINR_{0,u}^L = \frac{P_0 G_{0,u}}{\sum_i P_i G_{i,u} + \sigma 2}, u \in \mathcal{U}_j, i \in B, \tag{6}$$

where, P_0 denotes the transmitted power from MBS while, P_i denotes the average transmitted power from interfering LBS $i \in B$; $G_{0,u}$ and $G_{i,u}$ represent the average channel gains between the MBS $k \in \mathcal{K}$ and user u, and between the interfering LBS $i \in B$ and user u, respectively; while $\sigma 2$ denotes the additive noise power.

Moreover, the average SINR received by user u from LBS $k \in B$ can be calculated as follows:

$$SINR_{k,u}^L = \frac{P_k G_{k,u}}{P_0 G_{0,u} + \sum_{i \in B, i \neq k} P_i G_{i,u} + \sigma 2}, u \in \mathcal{U}_j, k \in B, \tag{7}$$

where, P_k is the transmitted power from LBS $k \in B$; $G_{k,u}$ represents the average channel gain between the LBS $k \in B$ and user u, while $G_{i,u}$ and $G_{0,u}$ denote average channel gain between interfering LBS $i \in B, i \neq k$ and user u and interfering MBS $k \in \mathcal{K}$ and user u, respectively. On the other side, regarding WiFi-BSs in ISCs, the average SNR received by user u, from WiFi-BS $k \in \mathcal{W}$ can be represented as follows:

$$SNR_{k,u}^W = \frac{P_k G_{k,u}}{\sum_{k'} P_{k'} G_{k',u} + \sigma 2}, u \in \mathcal{U}_w, k' \in \mathcal{W}/\{k\}, \tag{8}$$

where, P_k is the transmitted power from WiFi-BS $k \in \mathcal{W}$ and $G_{k,u}$ denotes the average channel gain of WiFi-BS $k \in \mathcal{W}$ to user u. While $P_{k'} G_{k',u}$ are the power transmitted and channel gain of interfering WiFi-BSs to user u, respectively. As defined before, ISC supports LTE and WLAN RAT, the average SINR received by user u from LTE-ISC (LBS) will be calculated using equation (7), while the average SNR received from user u from WiFi-ISC (WiFi-BS) is calculated using (8). Based on the above, the achievable downlink data

rate achieved by user u from $k \in \mathcal{K}/\mathcal{W}$ in slice j is represented as follows:

$$R_{u,j,k}^L = \frac{N_{RB-SUB}^L N_{slot}^L N_{u,k}^{L_{bits}} N_{k,j}^{RB} CR_{u,k}^L}{U_j T_{slot}^L}, j \in \dot{j}, u \in \mathcal{U}_j \tag{9}$$

where, N_{RB-SUB}^L represents the number of sub-carriers per one resource block, N_{slot}^L represents the number of slots per one sub-frame; $N_{u,k}^{L_{bits}}$ represents the number of bits per symbol, $N_{k,j}^{RB}$ is the total number of RB in slice j in BS $k \in \mathcal{K}/\mathcal{W}$; $CR_{k,u}^L$ is the user u coding rate. $N_{u,k}^{L_{bits}}$ and $CR_{k,u}^L$ can be both calculated using Channel Quality Indicator (CQI) that can be determined by SINR achieved at user u. U_j is the total number of users per slice j associated with BS $k \in \mathcal{K}/\mathcal{W}$, while T_{slot}^L is time slot interval and can be determined using (2). On the other hand, the achievable downlink data rate achieved by user u, from WiFi-BS $k \in \mathcal{W}$ can be represented as follows taking into consideration the WLAN MAC layer effect [38],

$$\mathcal{R}_{uk}^W = \frac{\tau(1-\tau)^{\mathcal{U}_w} D}{\left(T + \left(\frac{D\tau(1-\tau)^{\mathcal{U}_w}}{Rt_{uk}^{phy}} \right) \right)}, \forall k \in \mathcal{W}, \tag{10}$$

where , τ represents the channel contention probability; $\mathcal{U}_w$ denotes the total number of users associated with WiFi-BS $k \in \mathcal{W}$. While D, is the maximum allowed size of user u packets. T can be calculated as follows [38]:

$$T = (1-\tau)^{\mathcal{U}_w+1} e + \left(1 - (1-\tau)^{\mathcal{U}_w+1} \right) (T_{RTS} + T_{DIFS})$$
$$+ (\mathcal{U}_w + 1) \tau (1-\tau)^{\mathcal{U}_w} (T_{CTS} + T_{ACK} + 3T_{SIFS}) \tag{11}$$

where, T_{CTS}, T_{ACK}, and T_{SIFS} are the duration of Clear to Send (CTS) short frame, Acknowledgment short frame (ACK), and Short Frame Inter Space (SIFS), respectively. While T_{RTS} and T_{DIFS} denote the duration of Request to Send (RTS) short frame and DCF Inter-Frame Space (DIFS) short frame, respectively. In addition, e represents the duration of an empty slot time.

In (10), the nominator represents the average per-user data transferred in a time slot, while the denominator denotes the average length of time slot which defines the duration of successful data transmission and is represented in terms of WLAN physical rate that can be calculated as follows:

$$Rt_{uk}^{phy} = \frac{N_{u,k}^{sp} \times N_{Sub} \times N_{u,k}^{Wbits} \times CR_{u,k}^W}{T_{sym}}, \forall k \in \mathcal{W}, \tag{12}$$

where, $N_{u,k}^{sp}$ denotes the number of available spatial streams, $N_{u,k}^{Wbits}$ is the number of bits in one symbol; while N_{Sub} represents the total number of sub-carriers and $CR_{u,k}^W$ represents the coding rate. In addition, T_{sym} denotes the symbol duration.

Accordingly, if user u is associated to ISC using LWA (M3 mode), the achievable data rate of user u will be expressed as the summation of achieved rate from LBS BS and WiFi BS as follows:

$$R_{u,k}^{ISC} = \gamma R_{u,j,k}^L + \beta \mathcal{R}_{uk}^W, \forall k \in \daleth, \tag{13}$$

where, γ and β are two binary variables that represent the user's u association which reflects the transmission mode selection and can be expressed as follows:

$$\gamma, \beta = \begin{cases} 1, 0 \text{ if } ISC\ k \text{ operates in } M1, k \in B \\ O, 1 \text{ if } ISC\ k \text{ operates in } M2, k \in \mathcal{W} \\ 1, 1 \text{ if } ISC\ k \text{ operates in } M3, k \in \daleth \end{cases} \quad (14)$$

3) URLLC SLICE MODEL

On the other hand, in order to support URLLC service and to satisfy the requirements of URLLC service critical latency, a mini-slot based frame structure is adopted, where each time slot is divided into mini-slots based on selected scalable numerology μ_j. Based on this, the average delay to transmit a packet which is defined as the number of transmission time intervals can be calculated as follows:

$$D_{u,j,k}^{Ach} = \left[ceil\left(\left(\frac{Packet_{size}}{N_{RB-SUB}^L \times N_{slot}^L \times N_{k,j}^{RB} \times \psi} \right) \times U_j \right) \right] T_{slot}^L,$$
$$k \in \mathcal{K}/\mathcal{W} \quad (15)$$

where, ψ is defined as $N_{u,k}^{L_{bits}} \times CR_{u,k}^L$. Moreover, the average delay $D_{u,j,k}^{Ach}$ is dependent on $N_{k,j}^{RB}$ assigned for each slice which is based on α_j and μ_j.

Our objective is to fulfill the service request requirements of each user, which can be in the form of high rate for eMBB services or minimum delay for URLLC service. The satisfaction of user u in slice j associated with BS $k \in \mathcal{K}$ is represented by an increasing concave function, expressed as follows:

$$ST_{u,j,k} = \mathcal{R}^{eq}\left(1 - e^{-\frac{R_{u,j,k}^L}{R_{ref}}} \right) + \left(1 - \mathcal{R}^{eq} \right)\left(1 - e^{-\frac{D_{ref}}{D_{u,j,k}^{Ach}}} \right), \quad (16)$$

where, $\mathcal{R}^{eq}$ represents the user's service request. $\mathcal{R}^{eq}$ is a binary variable, in which $\mathcal{R}^{eq} = 1$ when a user u requested eMBB service and $\mathcal{R}^{eq} = 0$, when user u requested URLLC service. R_{ref} and D_{ref} is defined as minimum average rate and the minimum average delay a user u can achieve respectively.

IV. PROBLEM FORMULATION AND FRAMEWORK

As our objective is to meet the service request requirements of users, ensuring that users requesting eMBB services achieve the requested data rate, and besides at the same time, users requesting URLLC services meet the service requirements in terms of low latency. To address this, a joint optimization problem is formulated and defined to maximize users' satisfaction in network. The joint optimization problem aims to find an efficient allocation ratio α_j for slice resource allocation, determine the optimal scalable numerology value μ_j, and solve the user association problem.

A. PROBLEM FORMULATION

Consequently, the user's satisfaction maximization problem is formulated as follows;

$$OPT: \quad max_{x,\mu_j,\alpha_j} \sum_{k=0}^{\mathcal{K}} \sum_{j=0}^{J} \sum_{u}^{U} x_{u,j,k} \left(ST_{u,j,k} \right), \quad (17a)$$

$$S.t \sum_{k \in k} x_{u,j,k} = 1, \forall k \in \mathcal{K}, u \in U \quad (17b)$$

$$\sum_{j \in J} x_{u,j,k} = 1, \forall j \in J, u \in U_j \quad (17c)$$

$$x_{u,j,k} = \{0, 1\}, \forall k \in \mathcal{K}, u \in U \quad (17d)$$

$$R_{u,j,k}^L \geq R_{ref}, \forall u \in U \quad (17e)$$

$$D_{u,j,k}^{Ach} \leq D_{ref}, \forall u \in U \quad (17f)$$

$$\sum_j^J x_{u,j,k} R_{u,k,j}^L \geq R_{ref}, \forall j \in \dot{j} \quad (17g)$$

where, constrain (17b) denotes that user u can only associate with only one BS either MBS or ISC; while constrain (17c) also denotes that user u can only associate with only one slice per BS; constrain (17d) implies that association index $x_{u,j,k}$ can only have the value of $\{0,1\}$, where 0 denotes that user u is not associated with BS k while 1 denotes that user u is associated to BS k. While constrain (17e) and (17f) ensure that the satisfaction of user u of requested service is achieved; in terms of achievable data rate or delay, respectively. Constrain (17g) ensures the isolation between the eMBB and URLLC slices.

B. FRAMEWORK

We proposed a framework in order to find an optimal solution to the proposed optimization problem OPT. Which jointly solves the user association problem while finding the optimum slice resource allocation ratio and optimum numerology value that tends to maximize the user's satisfaction and guarantee each user's SLA. The proposed framework as shown in Fig.2 is divided into multiple phases as follows:

- **Phase I** (Pre-association): The primary goal of Phase I is to determine the number of user requests for each slice in each BS. This phase occurs only once at the beginning in order to find each user's service preference, where each user u creates a ranked preference list of all BSs that would satisfy its requested service requirement (eMBB or URLLC) to associate with. The user's ranked preference list will evaluated based on the channel state condition between user u and each BS k in the network. After each user u creates its preference list, the user associates with the first BS in the list. Consequently, the number of users U_j in slice j associated to BS k, $k \in \mathcal{K}/\mathcal{W}$ and the number of users $\mathcal{U}_w$ associated to BS k, $k \in \mathcal{W}$ can be calculated.

- **Phase II**: a dynamic RL-based slice resource allocation scheme is presented adapting Q-Learning algorithm in order to adjust slice's resource ratio α_j per each slice in each BS and to find the optimum scalable numerology μ_j value that will maintain and guarantee each user's service. In the Q-Learning algorithm [39], an agent learns

through iterative interactions with its environment. The agent selects actions and receives rewards based on its chosen actions. The objective of the agent is to maximize its cumulative reward over time. To address this, based on phase I calculation that occurs once, the proposed Q-Learning algorithm will iteratively check users' satisfaction in all BSs, if there is an increase in total users' satisfaction, the algorithm will re-adjust slice resource ratio α_j and scalable numerology value μ_j. If there is no enhancement in total user's satisfaction, the algorithm exits and last α_j and μ_j will be ignored.

- **Phase III** (Re-association): in this phase, users are re-associated with BSs based on the α_j and μ_j values resulted from previous phase. A regret learning matching algorithm is developed in order to re-associate user u to slice j in BS k that guarantees user's satisfaction. In regret matching, the user takes actions based on a specified utility function; all the actions may be chosen, with probabilities that are proportional to the apparent gains, as measured by the user's regrets.

The *OPT* problem (17a), is a binary non-linear NP-hard with the multi-objective problem. To find a global optimum solution, we decompose our optimization problem as two sub-problems; the first sub-problem involves the allocation of slice resources (Slice Ratio α_j) and determining the optimal scalable numerology value μ_j and the second sub-problem, defined as user Re-association problem. The proposed approaches to solve the two sub-problems are described in detail in the next two sections.

V. RL-BASED RESOURCE SLICING (Q-LEARNING) ALGORITHM

In order to satisfy each user's u service request requirement, we need to evaluate efficiently slice resource ratio α_j of each slice per BS jointly evaluating the optimum value of scalable numerology μ_j in each slice. To address this, our proposed Q-Learning algorithm is modeled as a finite Markov Decision Problem (MDP), where the states and the actions spaces are defined as finite. Markov Decision Problem quaternion [39] is defined as $(\mathcal{S}, \mathcal{A}, \mathcal{R}, \mathcal{S}')$, where $\mathcal{S}$ represents the discrete set of environment space, $\mathcal{A}$ represents the discrete set of possible actions of an agent, $\mathcal{R}$ represents the reward function of agent and S' is the state transition probability.

State ($\mathcal{S}$): the discrete set of states can be obtained from two factors: the ratio of logical RBs α_j assigned to a slice j in a BS $k \in \mathcal{K}/\mathcal{W}$ assuming two slices per BS, where $j = 1$ represents eMMB slice while $j = 2$, denotes URLLC slice. While, the second factor is the scalable numerology value $\mu_j \in \mathcal{M}_j = \{0, 1, 2, \ldots 4\}$; which ensures the service requirement SLA of the URRLC slice. By obtaining the value of α_j for one slice, the $\alpha_{j', j' \neq j}$ for the other slice can be obtained using the following:

$$\alpha_{j', j' \neq j} = \left\lfloor \frac{B - \alpha_j \Delta F_0}{\Delta F_j} \right\rfloor, \tag{18}$$
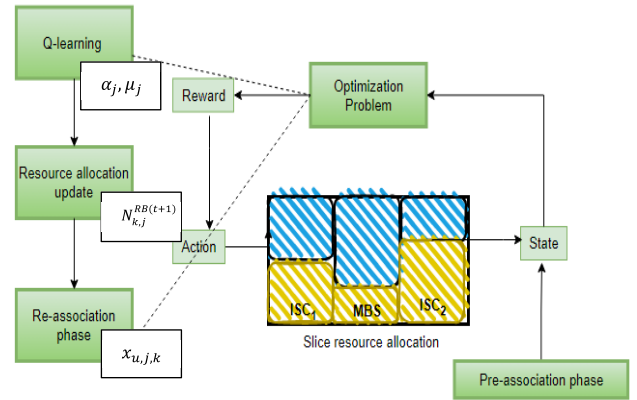

FIGURE 2. Proposed framework.

Action ($\mathcal{A}$): based on the current state $s_t \in \mathcal{S}$, the learning agent selects the best action $a_t \in \mathcal{A}$ aiming to maximize reward $r_t \in \mathcal{R}$ over time. The set of possible actions to be selected is represented as $\mathcal{A} = \{(\alpha_j \pm 1, 1 \leq \alpha_j \leq 100, \mu_j \pm 1, 0 \leq \mu_j \leq 4\}$. Moreover, when the agent selects an action, the agent can transfer to next state $\mathcal{S}'$ with a certain transition probability. The transition probability is defined as [39]:

$$P = \{p_{s,s'}^a \mid s, s' \in \mathcal{S}, a \in \mathcal{A}\}, \tag{19}$$

In addition, the reward $r_t \in \mathcal{R}$ is calculated as the sum of the satisfaction of users associated with BSs $k \in \mathcal{K}$ based on user's service preference as follows:

$$r_t = \sum_{u,j,k} ST_{u,j,k}, j \in \dot{j}, u \in U, \tag{20}$$

where, $ST_{u,j,k}$ is user's satisfaction and can be calculated using (16). In Q-learning algorithm, the agent aims to learn the optimal strategy π^* which corsseponding to state value function $V^*(s)$ and the action value function $Q^*(s, a)$ as follows:

$$V^*(s) = max_a Q^*(s, a), \tag{21}$$

The formula of updating the Q value is represented as follows [39]:

$$Q(s, a) = Q(s, a) + \delta_L[\mathcal{R} + \beth + max_{a*}Q^*(s', a) - Q(s, a), \tag{22}$$

where, δ_L is the learning rate, the agent over time, learns optimal policy to maximize the discounted rewards, this cumulative discounted reward can be represented as follows [39]:

$$R_t = E\left[\sum_{j=0}^{\infty} \beth^j r_{t+j}\right], \tag{23}$$

where, $0 \leq \beth \leq 1$ represents the discount factor and the operation E[.] is the expectation with respect to the reward distribution.

Furthermore, Algorithm.1 proposed the details of resource slicing Q-learning algorithm. First, (Line 1) is the initialization process in each episode, in a defined number of episodes M^{epi}. The $N_{k,j}^{RB}$ per slice is dependent on α_j and μ_j values, where their values can be determined either by with initial

values or valued from the previous Q-learning algorithm iteration (line 3). In each state, an action is performed based on the ε-greedy policy [40] that represents both the exploration (random action selection) and the exploitation (action with the maximum Q-value) trade-off (lines 4,5). r_t is then calculated according to (20) based on the resulted state s_{t+1}, and the state-action table is updated (lines 6-8). The algorithm is terminated after Q-table convergence or a predefined number of iteration T is reached.

Algorithm 1 Q-Learning based Resource-Slicing

1: **Initial**: Calculate $N_{k,j}^{RB}$ using (4), and update the environment states $\mathcal{S}$. For each state-action pair (s, a), initialize the state-action value function $Q(s, a)$ arbitrarily to zero. Initialize the value of the discount factor $\beth$ and the learning rate δ_L.
2: **For** *episode* := 1 *to* M^{epi} **do**
3:    Get initial state s_1.
4:   **For** $t := 1$ *to* T **do**
5:

$$a_t = \begin{cases} \text{random,} & \text{with } \varepsilon \text{ probability} \\ \underset{a}{\text{argmax }} Q(s_t, a), & \text{otherwise} \end{cases}$$

6:    Execute action a_t, and obtain s_{t+1}.
7:    Calculate r_t using (20).
8:    Select an action a_{t+1} based on state s_{t+1}, and update state-action table $Q(s, a)$ based on (22).
9:    Replace $s_t \leftarrow s_{t+1}$.
10:   **End for**
11: **End for**
12: **Result:** Optimal state with r_{max}, and optimal policy Q*.

VI. MATCHING GAME-BASED USER RE-ASSOCIATION ALGORITHM

The final phase is the re-association phase. Regret- Matching Learning algorithm is adopted in order to solve the users' association problem after finding the optimum value for slice resource allocation ratio α_j and scalable numerology μ_j value. In this algorithm, a generalized optimal form of Nash Equilibrium which is known as the Correlated Equilibrium (CE) is considered [41]. It's known by its bounded payoffs for any finite game, where players' actions (users and BSs) can be correlated resulting in an equilibrium border set in which a deduction can occur in a better payoff for the players. Based on this, the regret matching game is formulated to model users' association problem and can be mathematically represented as follows:

$$\mathcal{G} = \left(\mathcal{U}', \left(\mathcal{S}_u^R \right), (U_u) \right), \qquad (24)$$

where, $\mathcal{U}'$ represents the set of players (i.e., the users $\mathcal{U}$), $\mathcal{S}_u^R \in \mathcal{S}^R$ denotes the set of strategies for player u (i.e., the BSs $\mathcal{K}$), while $\mathcal{S}^R$ is the set of strategies for all players, and $U_u : \mathcal{S}^R \longrightarrow \mathbb{R}$ is the payoff function for player u when

the action taken by all players is $s^R \in \mathcal{S}^R$. From this, the utility function of player (user) u to associate with BS k can be represented as follows which is represented as user's satisfaction:

$$U_u(s^R) = ST_{u,j,k}, \qquad (25)$$

In addition, the overall payoff for player u taking into consideration the random actions with overall Probability Mass Function (PMF) $n \in \aleph$ can be represented as follows:

$$U_u(n) = \sum_{s^R \in \mathcal{S}^R} n\left(s^R\right) . U_u(s^R), \qquad (26)$$

Definition 1: A probability distribution π^R on a set of strategies $\mathcal{S}^R$ is defined to be correlated equilibrium for a defined game if for every player $u \in \mathcal{U}'$, and for every pair of action $z, y \in \mathcal{S}_u^R$ it holds that:

$$\sum_{s^R \in S^R: u=z} \pi^R(s^R)\left(U_u\left(y, s_{-u}^R\right) - U_u\left(z, s_{-u}^R\right)\right) \leq 0, \qquad (27)$$

This equation proved that when a strategy z is recommended. Changing the strategy to a new strategy $y \neq z$ will lead to no regret. The main objective of no-regret matching algorithms is to provide no regret for a player u, in which the probability of choosing a strategy is proportional to the regret for not choosing any other strategies. If there are two strategies $y \& z$, where $y \neq z$, thus the regret of player u is to choose strategy y instead of z at a time t is defined as follows [42]:

$$\rho_u^t(z, y) \triangleq max(D_u^t(z, y), 0), \qquad (28)$$

where $D_u^t(z, y)$ represents the payoff for player u if he had played action y instead of z every time in the past, and it can be calculated as follows [42]:

$$D_u^t(z, y) \triangleq \frac{1}{t} \sum_{T \leq t} \left(U_u^T\left(y, s_{-u}^R\right) - U_u^T\left(z, s_{-u}^R\right)\right), \qquad (29)$$

Thus, the probability distribution of player u chooses an action at time t is [42]:

$$p_u^{t+1}(y) = \begin{cases} \frac{1}{\mu} \rho_u^t(z, y), & y \neq z \\ 1 - \sum_{y \in \mathcal{S}_u^R, y \neq z} p_u^{t+1}(y), & y = z \end{cases}, \qquad (30)$$

where $\mu > 2MG$ is a constant which guarantees that $p_u^{t+1}(y) > 0$ at $y = z$ and G is the upper bound of $|U(s^R)|$ for all $s^R \in \mathcal{S}^R$. In addition, when t = 1, the initial probability is distributed uniformly over the set of all possible actions. Moreover, it states in [42], that the empirical distribution $\bar{z}_t$ of joint actions s of all players until t:

$$\bar{z}_t(s) = \frac{1}{t} N(t, s^R), \qquad (31)$$

where, $N(t, s^R)$ represents the number of periods before t where action s^R is chosen.

Furthermore, Algorithm. 2 proposed the details of regret matching algorithm for user re-association phase to find optimal S^{R*}. Starting by initialization (line 1-4), where for each

user u, the utility $U_u(s^R)$ is calculated, then user's Payoff $D_u^t(z, y)$ is evaluated to update the regrets $\rho_u^t(z, y)$. Followed by, updating the probabilities $p_u^{t+1}(y)$ for each user u, where the strategy s_u^R is chosen based on ε^R-greedy policy (lines 5,6). The algorithm is repeated until $(\rho_u^t(z, y)) < \xi$, where ξ should be properly chosen as in [43].

Algorithm 2 Regret based learning algorithm for users Re-association phase

---

1: **Initial**: for each user u, generate random uniform probability $p_u^1(y)$ for all base stations $y \in \mathcal{K}$.
2: **While** sup $(\rho_u^t(z, y)) < \xi$ **do**
3:    Calculate users utilities $U_u(s^R)$ using (24).
4:    Update regrets $\rho_u^t(z, y)$ using (27)
5:    Use (29) to update the probabilities $p_u^{t+1}(y)$.
6:    Use $p_u^{t+1}(y) \, \forall y \in \mathcal{M}$ to select action $s_u^R \in \mathcal{S}_u^R$ as follow:

$$s_u^R = \begin{cases} \text{random,} & \text{with } \varepsilon^R \text{ probability} \\ \underset{y}{\arg\max} \, p_u^{t+1}(y), & \text{otherwise} \end{cases}$$

8: $t = t + 1$.
9: **End While**
10: **Result:** Optimal S^{R*}

VII. PERFORMANCE EVALUATION

In this section, we present our simulation results to evaluate our proposed framework. We begin by describing the simulation setup, and providing details on the parameters and configurations used. Following that, we present the obtained simulation results.

A. SIMULATION SETUP

In order to evaluate the proposed framework, the following simulation setup was adopted, where a multi-RAT HetNet is considered with one MBS and 2 ISCs. The ISCs support both RATs (LTE and Wi-Fi) technology, which allows three access options can be denoted by three transmission modes of operation: M1 (LTE mode), M2 (Wi-Fi mode), and M3 LTE-WLAN aggregation (LWA mode). Based on this, each physical ISC node can be represented by three virtual nodes (ISC-LTE, ISC-WiFi, and ISC-LWA) corresponding to its three modes of operation (M1, M2, and M3). Each ISC has a radius of 50m deployed in an indoor hotspot area of 350×225 m2 and under coverage of MBS of a radius of 1000m. A number of users are uniformly distributed under the coverage of MBS and inside the indoor hotspot area. In addition, each ISC-LTE is divided into 2 slices which are eMBB and URLLC.

Moreover, the path loss model for WLAN technology is presented as follows [44]:

$$P_{loss} = 20log_{10}\left(f^W\right) + \eta^w log_{10}(d) + P_f\left(n_{walls}\right) - 28 \tag{32}$$

where, f^W denotes WLAN transmission frequency in MHZ; d denotes the distance in meters; η^w is the distance power loss

coefficient and its assumed to be equal 30; n_{walls} is the number of walls which is assumed to be 3; in addition, $P_f(n_{walls})$ represents the penetration loss facture and can calculated as $n_{walls} + 13$.

B. SIMULATION RESULTS

In our proposed scenario, we deployed Multi-RAT BSs, which include a MBS and ISCs. The concept behind this deployment is to leverage different technologies to accommodate both eMBB users and URLLC users in the network. This requires striking a balance between providing high data rates for eMBB users and low latency for URLLC users at the same time. In order to evaluate the performance of this idea, we compare it with other different scenarios in terms of architecture and deployed algorithm. For the baseline approaches, compared to our proposed framework

- LTE-only model [17]: where all the ISCs proposed in our framework is replaced by LBSs, resulting in all BSs operating in LTE technology only. The idea behind comparing our proposed model to a scenario of deploying only LTE base stations is to highlight the benefit of deploying LWA technology. Deploying LWA technology to exploit the unlicensed band, leads to an improvement in network capacity especially in high demands. In addition, concerning the cost, where deploying a number of WiFi BSs (unlicensed band) plus LTE-BSs will definitely be less in cost compared to deploying the entire architecture with LTE-BSs.
- LTE-WiFi model: where all ISCs proposed in our framework are replaced by LBSs and WiFi BSs with no aggregation transmission mode, results in a number of BSs supporting only LTE technology while others support only WiFi technology.

For both, the LTE-only model and the LTE-WiFi model, we ensure the same number of BSs that will be equivalent to ISCs in our proposed framework.

- Heuristic-Genetic Algorithm (GA): The proposed framework is compared to three different scenarios based on the GA as follows
1. **Genetic All**: In this scenario, both the Q-learning algorithm and the Regret-Matching user Re-association algorithm in our proposed framework are replaced with the Genetic Algorithm.
2. **Q-learning and GA**: This scenario keeps the Q-learning phase of our proposed framework as it is, but replaces the Regret-matching User Re-association algorithm with the Genetic Algorithm.
3. **GA and Regret-Matching**: In this scenario, the Q-learning algorithm in our proposed framework is replaced by the Genetic Algorithm, while the Regret-matching algorithm is retained in the User Re-association phase.

These scenarios serve as comparisons to evaluate the performance and effectiveness of our proposed framework

TABLE 3. Simulation paramters [45].

LTE Parameters	
Transmit power of LTE MBS	46 dBm
Transmit power of LTE SBS	20 dBm
Path loss between MBS and a user	$128.1+37.6 \log_{10}d(km)$
Path loss between SBS and a user	$140.7+36.7 * \log_{10}(d)$
LTE bandwidth	20 MHz
Channel bandwidth	180 KHz
Transmission Frequency	2GHz
Noise Power	-174 dBm/Hz
Packet Size	1500 byte
WLAN Parameters	
WLAN bandwidth	20MHz
Transmit power of WAP	200 mW
Spatial Streams	4
Minimum Contention Window (W)	32
Maximum Number of back-off stage	5
Maximum Number of Re-transmission (μ)	7
Transmission Frequency (f^W)	2.4GHz
Slot time	9 μs
DIFS	50 μs
SIFS	10 μs
ACK	160 bits
RTS	208 bits
CTS	160 bits
Network Parameters	
Number of Slices (eMBB, URLLC)	2
5G numerology in eMBB slice	$\mu_j = 0$
5G numerology in URLLC slice	$\mu_j = \{1,2,3,4\}$
RB ratio range	$0 \leq \alpha_j \leq 1$
Epsilon Decay	0.003
Min. Rate demand R_{ref}	1M bps
Min. Delay demand D_{ref}	1ms

against different combinations of Genetic Algorithm-based approaches.

- Different RL-based approaches [17]: Our proposed Q-Learning algorithm is compared to two different RL-based algorithms which are Deep Q-Network (DQN) and Dueling-DQN (DDQN) algorithms.

Furthermore, we evaluate and analyze our proposed framework by comparing it to these different scenarios and algorithms ensuring they have the same total capacities and simulation parameters. The rest of LTE, WLAN and Network parameters are summarized in Table. 3 [45].

To start, we evaluate the effectiveness of our proposed framework by deploying Multi-RAT BSs utilizing LWA technology. We compared the performance of this deployment to a traditional LTE-only model. Fig.3, illustrates the variation in the users' average satisfaction with the increase of the number of users in the network (N=40 to N=80). The users' average satisfaction can be defined as the sum of users' satisfaction in each slice in each BS, divided by the total number of users in the network (N). The evaluation of our proposed model compared to LTE-only model is analyzed over different values $\mathcal{R}^{eq}$. Specifically, when the user's request ($\mathcal{R}^{eq}$) is set to 0.2, it indicates that 20% of the total number of users is requesting eMBB service, while the remaining 80% are requesting a delay service (URLLC). Similarly, if $\mathcal{R}^{eq}$ is set to 0.5, it means
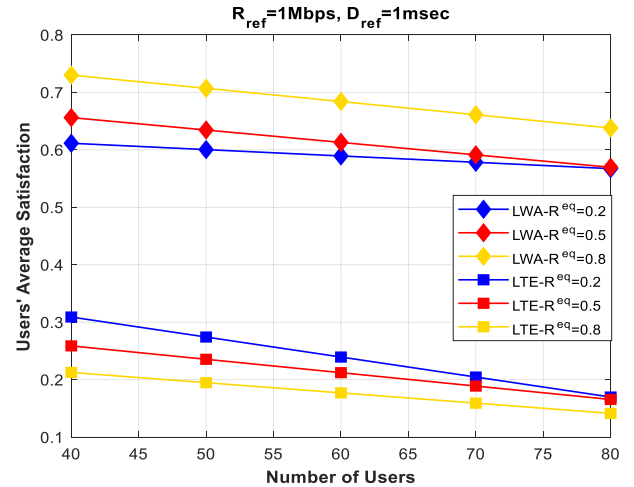

FIGURE 3. Evaluation of users' average satisfaction between our proposed model and LTE-only model with different $\mathcal{R}^{eq}$ values.

that an equal number of users are requesting eMBB service and URLLC service, etc.

It can be noticed that our proposed framework outperforms LTE-only model in all $\mathcal{R}^{eq}$ values. As our proposed framework guarantees to satisfy each user's service request through finding the optimum slice resource allocation ratio and selecting the optimum numerology value. Additionally, as the number of users who request eMBB service ($\mathcal{R}^{eq}$ =0.8) increases, our proposed model keeps to maintain average users' satisfaction high compared to LTE-only model. This is due to; the utilization of LWA technology effectively maintains average user's satisfaction, especially in high demand rate-based services. Moreover, by identifying the optimal value for scalable numerology, our framework successfully satisfies users who require URLLC services, thereby striking a balance between the diverse users' service requests.

In contrast, in the LTE-only model, the users' average satisfaction decreases with the increase in the number of users who are requesting eMBB service ($\mathcal{R}^{eq}$ = 0.8). For example, when the number of users requesting URLLC service is greater than eMBB service ($\mathcal{R}^{eq}$ = 0.2), the user's average satisfaction is the greatest. This is due to, when $\mathcal{R}^{eq}$ value is small, the number of users requesting high-rate based service is small but when $\mathcal{R}^{eq}$ value increases the LTE-only model can't maintain to satisfy all users results in a great degradation in average users' satisfaction due LTE limited resources. In the LTE-only model, users can only benefit from deploying network slicing with scalable numerology, and this showed that when $\mathcal{R}^{eq}$ = 0.2, as 80% of users are requesting URRLC the users average satisfaction is the best at the expense of eMBB users.

Without loss of generality, Fig.(4&5) capture the difference between our proposed framework and LTE-only model in terms of striking a balance between the diverse users' service requests. As the user's average satisfaction is divided into user's rate satisfaction and user's delay satisfaction according to (16), Fig. 4 shows the change of users' average rate

satisfaction with the increase of the number users. The comparison is evaluated over two values of $\mathcal{R}^{eq}$, which are the minimum ($\mathcal{R}^{eq} = 0.2$) and the maximum ($\mathcal{R}^{eq} = 0.8$) to evaluate the performance of both models. It can be noticed that our proposed framework, always achieves better performance in terms of user rate satisfaction, even with the increase of the number of users who request eMBB service that requires a high data rate.

This shows the effect of deploying the LWA technology which leverages the traffic load and supports LTE limited capacity. In addition, our proposed regret-matching re-association phase algorithm is designed to follow each user's intentions (utility function (24)) to choose the most suitable transmission mode (M1, M2, and M3), that tends to maximize the user's satisfaction and guarantee each user' SLA. In comparison, LTE-only model led to a significant degradation in users' average rate satisfaction levels compared to our proposed framework with an average of only 35% of users are satisfied, this can be attributed to deploying only LTE-BSs that suffer from a limited capacity, especially in high demand rate-based services.

Moreover, Fig. 5 shows the change in users' average delay satisfaction with the increase in the number of users. The comparison is also evaluated over two values of $\mathcal{R}^{eq}$ (0.2&0.8). It can be noticed that, the LTE-only model reaches high values in users delay satisfaction compared to our proposed framework. This can be attributed to the number of LTE-BSs that deploy network slicing with scalable numerology is greater than the number in our proposed framework. In our proposed system model, we deploy 2 ISCs, which are divided into 2 LBSs and 2 WiFi-BSs, while in LTE- only model these 2 ISCs are replaced 4 LBSs to maintain the same number of BSs. Increasing the number of LTE-BSs that support NR scalable numerology will lead to an enhancement in users' average delay satisfaction at the expense of users' average rate satisfaction. In contrast, our proposed framework maintains a guaranteed total average user's satisfaction compared to LTE-only model.

Our proposed model is capable of maintaining user satisfaction while achieving a balance between rate and delay users' satisfaction even when there are changes in user's service requests percentage. Disposing our proposed framework which efficiently strike a balance between the diverse user's service requests, enhance LTE-RAT limited capacity while considering NR scalable numerology technique can achieve better performance in users' average satisfaction.

Furthermore, the performance of our proposed framework is evaluated and compared with the LTE-WiFi model as shown in Fig.6. In the LTE-WiFi model, 2 ISCs deployed in our proposed framework are replaced by 2 LSBs and 2 WiFi- BSs to maintain the same number of BSs.

The comparison is performed at the change in the user's average satisfaction with the increase in the number of users considering different values ($\mathcal{R}^{eq}$), as shown in Fig.6. It can be noticed that, in the LTE-WiFi model, the users' average satisfaction is enhanced compared to LTE-model
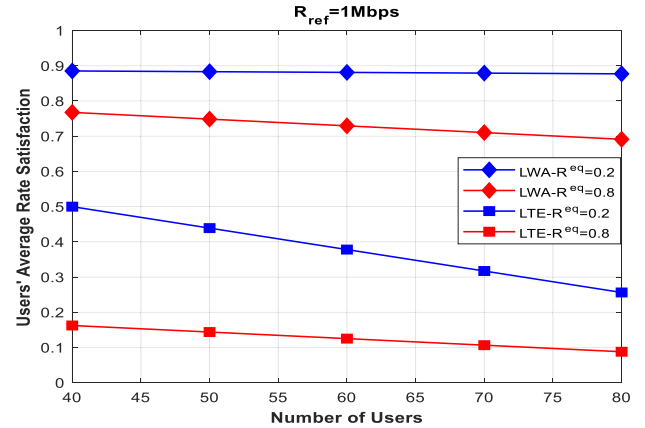

FIGURE 4. Evaluation of users' average rate satisfaction between our proposed model and LTE-only model with different $\mathcal{R}^{eq}$ values.
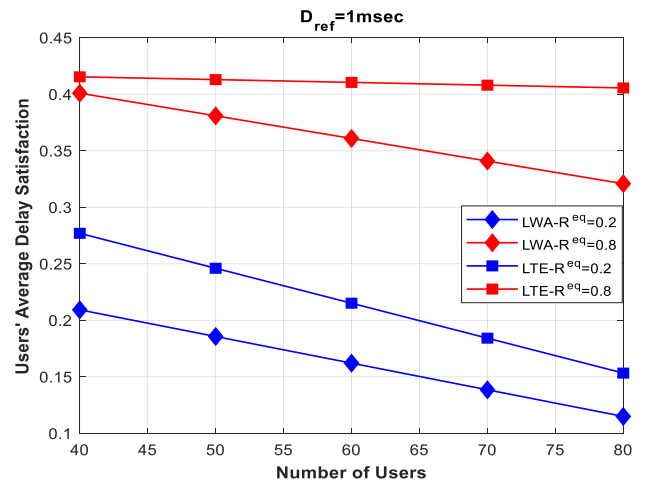

FIGURE 5. Evaluation of users' average delay satisfaction between our proposed model and lte-only model with different $\mathcal{R}^{eq}$ values.
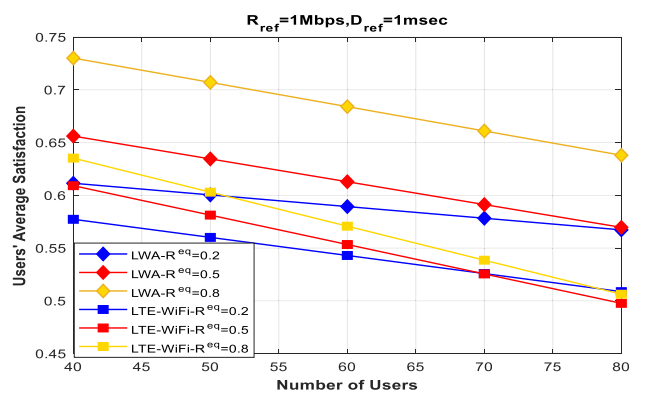

FIGURE 6. Evaluation of users' average satisfaction between our proposed model and LTE-WiFi model with different $\mathcal{R}^{eq}$ values.

as the presence of unlicensed band (WiFi) leverages the traffic load and supports the LTE limited capacity. By deploying WiFi-BSs, eMBB users' opportunity to achieve their requested high data rate service requirement increased. However, when the number of users in the network increased to reach N = 80, and the percentage of users requesting,

eMBB service also increased ($\mathcal{R}^{eq} = 0.8$), there is a noticeable degradation in LTE-WiFi model. As only of an average 64% of users are being satisfied. This degradation can be subjected to congestion in both LTE and WiFi BSs as they become overwhelmed with the increase of the number of users. On the other hand, our proposed model offers users an opportunity to enhance their achievable data rate by exploiting the LWA transmission mode.

Additionally, in order to show the effectiveness of our proposed framework, Fig.7 shows the effect of changing the R_{ref} values on users' average satisfaction. In Fig.7, we compared our proposed framework with the LTE-only model and LTE-WiFi model ranging the R_{ref} from 1Mbps to 2Mbps, and analyzed how this will affect the users' average satisfaction. We took an average rate request ($\mathcal{R}^{eq} = 0.5$) where the number of users requesting eMBB service is equal to the number of users who are requesting URLLC service with a total number of N = 80 users. Generally, as shown the users' average satisfaction decreases with the increase of R_{ref} value in all models, however, our proposed framework can still maintain an average of 70% of users' average satisfaction compared to the other two models.

Moreover, Fig.8 shows the effect of changing the D_{ref} values on users' average satisfaction. In Fig.8, we compared our proposed framework with LTE-only model and LTE-WiFi model ranging the D_{ref} from 0.5msec to 1.5msec, and analyzed how this will affect the users' average satisfaction. We took an average rate request ($\mathcal{R}^{eq} = 0.5$) as the number of users who requesting eMBB service is equal the number of users who are requesting URLLC service with a total number of users N = 80 users.

Generally, as shown the users' average satisfaction decrease with the decreases of D_{ref} value in all models, however, our proposed framework can still maintain an average of 60% of users' average satisfaction compared to the other two models. It may concluded, that our proposed framework has the capability to achieve and maintain users' satisfaction regardless of the type of service requested compared to LTE-only model or LTE-WiFi model. This is accomplished by as our proposed framework leveraging LWA technology incorporating NR scalable numerology technique. By proposing such a framework we can increase capacity and improve the overall user experience for both high data rate and lower latency based-service.

Furthermore, in Fig.9 our proposed algorithm is compared with three Genetic-based algorithms which are Genetic all, Q-learning &Genetic and Genetic &Regret-matching. In order to solve the optimization problem, in terms of maximizing total users' satisfaction, GA-all is first used to jointly find an efficient resource allocation ratio α_j for slice j in each BS along with finding the optimum value of scalable numerology μ_j, then GA is used in re-association phase instead of regret matching algorithm. In Q-learning &Genetic, the GA only substitutes the regret matching algorithm in re-association phase in our proposed model, in contrast in Genetic &Regret-matching, the GA
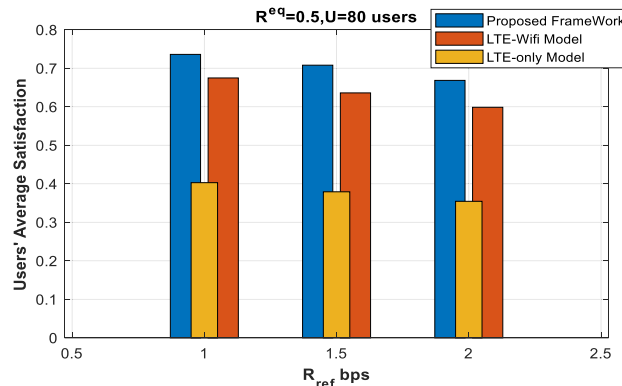

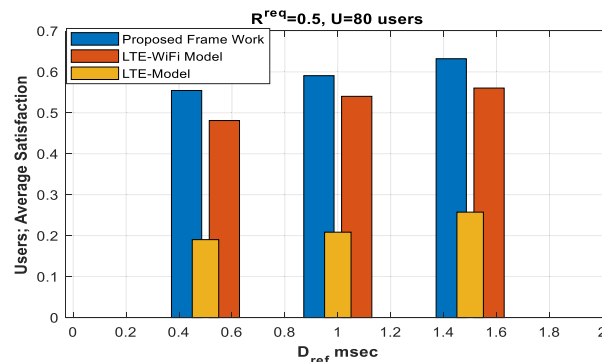**FIGURE 7.** The effect of changing the R_{ref} on users' average satisfaction.

FIGURE 8. The effect of changing the D_{ref} on users' average satisfaction.
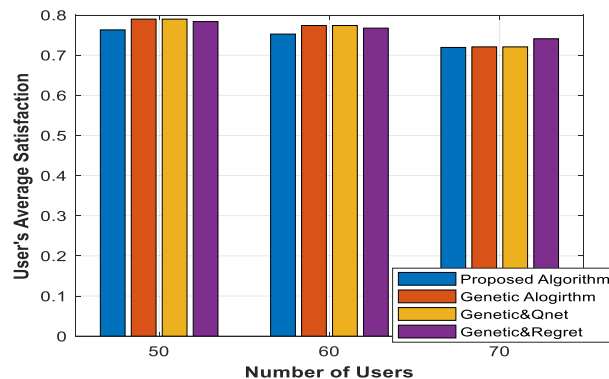

FIGURE 9. Comparing our proposed model with heuristic-genetic algorithm (GA).

substitutes the Q-learning algorithm in our proposed model. Our proposed framework has demonstrated notable success in achieving high average user satisfaction values, reaching approximately 74%. This performance is comparable to the results obtained by three other Genetic Algorithm (GA)-based algorithms, which achieve around 77% satisfaction. While GA is renowned for finding optimal solutions in large search spaces, it is important to consider the convergence time complexity. This complexity is affected by factors such as the population size (P), the number of generations (G), and the runtime complexity of the fitness function (F). Overall, the complexity of GA can be expressed as O(GPF).

C. COMPUTATIONAL COMPLEXITY, OPTIMALITY AND CONVERGENCE ANALYSIS

We conducted a comparison between our proposed Q-Learning algorithm and two other RL-based algorithms, namely DQN and DDQN. We analyzed the variations in achieving users' average satisfaction as the number of users in the network increased. As shown in Fig.10 the three RL-based algorithms' performances are equally the same, they all almost achieve an average of 75% of user's satisfaction.

However, from the perspective of the time complexity of each algorithm, the Q-Learning can be easily converging after I iterations and T slot time. Therefore, the time complexity of Q-Learning algorithm can be expressed as O(IT). On the other side, the time complexity of DQN can rely on more factors, I iterations, T slot time, number of layers L of DQN neural network and the number of neurons n in each layer l, $l \in$ L. Therefore, the time complexity of DQN to converge can be expressed as $O(IT \sum_{l=0}^{L-1} n_l n_{l+1})$. In addition, although the convergence of DDQN algorithm depends more on the number of states and action sets associated with the learning process, however, once the convergence state is achieved, any updating states will not need a learning process [46]. Concluded from this, Q-learning has the highest learning efficiency and less time complexity than other RL-based algorithm, however, if the state and action space are increased and become too large, other RL-based algorithm will be more efficient.

Furthermore, In Q-learning algorithm, in each state, an action is performed based on the ε-greedy policy that represents both the exploration (random action selection) and the exploitation (action with the maximum Q-value) trade-off. Where the agent randomly selects an action with probability epsilon ε and otherwise selects with probability $1 - \varepsilon$ the action greedily. The value of epsilon decays is based on a decay rate called epsilon decay and can be calculated as follows [47]:

$$Epsilon = Epsilon(1 - EpsilonDecay), \qquad (33)$$

The greater the value of Epsilon, the greater the agent randomly explores the action space. We present an analysis of the impact of different Epsilon values on the optimality of our proposed optimization problem. We specifically examine the effect of Epsilon Decay values on user satisfaction, which is a reward in our Q-learning algorithm. We evaluate three different Epsilon Decay values as shown in Fig 11, ranging from 0.03 to 0.0003. In our proposed algorithm (as described in Table 3), we utilize an Epsilon Decay value of 0.003, which is determined based on the chosen Maximum number of steps index. Our analysis as shown in Fig.11 reveals that the Epsilon Decay value of 0.003 consistently outperforms the other Epsilon values across different N values (number of users), resulting in an average user satisfaction of approximately 65%.

Furthermore, in this simulation, the convergence of our proposed Q-Learning Algorithm is evaluated as shown in Fig.12. The result of this simulation is based on the maximum
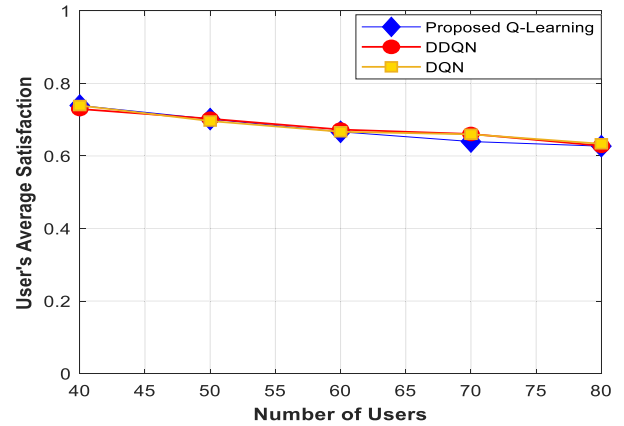

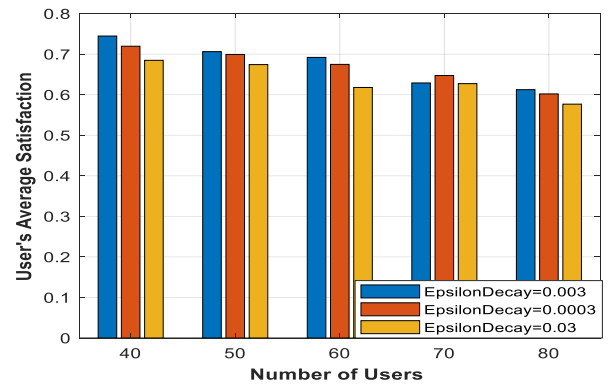**FIGURE 10.** A comparison between our proposed Q-Learning algorithm and other RL-based algorithms.

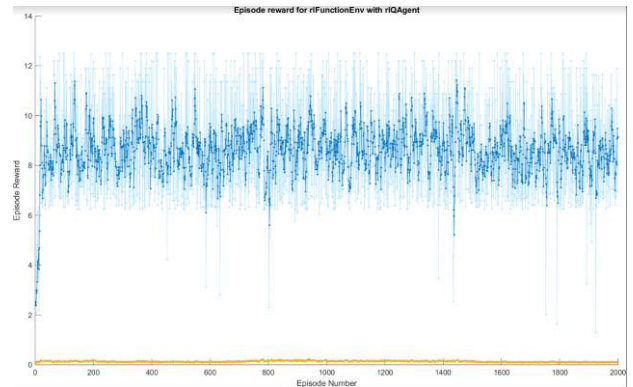**FIGURE 11.** A graphical demonstration of variation of epsilon decay.

FIGURE 12. Convergence process of proposed q-learning algorithm.

value of every 2000 episodes. It can be observed that the Q-Learning algorithm starts at low reward first until it starts to converge at episodes 200.
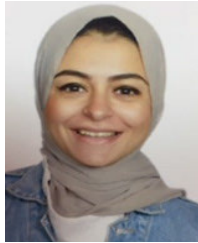
VIII. CONCLUSION

A radio resource allocation scheme regarding the eMBB and URLLC slices in Multi-RAT HetNet architecture is studied. The proposed framework leverages the 5G New Radio NR scalable numerology technique and LWA, aiming to solve the radio resource slicing allocation problem for serving URLLC

and eMBB users. The user association problem has been formulated as an optimization problem jointly with finding an efficient resource allocation ratio for each slice in each base station and finding the optimum value of scalable numerology in URLLC slice in each BS with the objective of maximizing users' satisfaction. To solve this problem a Q-learning and Regret-Matching algorithms are formulated. Our simulation results show that our proposed framework is capable of catering the diversity in users' services requests and maintaining users' satisfaction.

REFERENCES

[1] A. Dogra, R. K. Jha, and S. Jain, "A survey on beyond 5G network with the advent of 6G: Architecture and emerging technologies," *IEEE Access*, vol. 9, pp. 67512–67547, 2021.

[2] D. Gligoroski and K. Kralevska, "Expanded combinatorial designs as tool to model network slicing in 5G," *IEEE Access*, vol. 7, pp. 54879–54887, 2019.

[3] F. Z. Yousaf, M. Bredel, S. Schaller, and F. Schneider, "NFV and SDN—Key technology enablers for 5G networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 11, pp. 2468–2478, Nov. 2017.

[4] J. Mei, X. Wang, and K. Zheng, "An intelligent self-sustained RAN slicing framework for diverse service provisioning in 5G-beyond and 6G networks," *Intell. Converged Netw.*, vol. 1, no. 3, pp. 281–294, Dec. 2020.

[5] Z. Shu and T. Taleb, "A novel QoS framework for network slicing in 5G and beyond networks based on SDN and NFV," *IEEE Netw.*, vol. 34, no. 3, pp. 256–263, May 2020.

[6] N. Patriciello, S. Lagen, L. Giupponi, and B. Bojovic, "5G new radio numerologies and their impact on the end-to-end latency," in *Proc. IEEE 23rd Int. Workshop Comput. Aided Modeling Design Commun. Links Netw. (CAMAD)*, Sep. 2018, pp. 1–6.

[7] X. Zhang, X. Guo, and H. Zhang, "RB allocation scheme for eMBB and URLLC coexistence in 5G and beyond," *Wireless Commun. Mobile Comput.*, vol. 2021, pp. 1–7, Oct. 2021.

[8] T. Dang and M. Peng, "Delay-aware radio resource allocation optimization for network slicing in fog radio access networks," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, May 2018, pp. 1–6.

[9] M. Alsenwi, N. H. Tran, M. Bennis, A. K. Bairagi, and C. S. Hong, "eMBB-URLLC resource slicing: A risk-sensitive approach," *IEEE Commun. Lett.*, vol. 23, no. 4, pp. 740–743, Apr. 2019.

[10] A. K. Bairagi, M. S. Munir, M. Alsenwi, N. H. Tran, S. S. Alshamrani, M. Masud, Z. Han, and C. S. Hong, "Coexistence mechanism between eMBB and uRLLC in 5G wireless networks," *IEEE Trans. Commun.*, vol. 69, no. 3, pp. 1736–1749, Mar. 2021.

[11] P. Yang, X. Xi, T. Q. S. Quek, J. Chen, X. Cao, and D. Wu, "How should I orchestrate resources of my slices for bursty URLLC service provision?" *IEEE Trans. Commun.*, vol. 69, no. 2, pp. 1134–1146, Feb. 2021.

[12] W. Wu, "Dynamic RAN slicing for service-oriented vehicular networks via constrained learning," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 7, pp. 2076–2089, Jul. 2021.

[13] V. Sciancalepore, X. Costa-Perez, and A. Banchs, "RL-NSB: Reinforcement learning-based 5G network slice broker," *IEEE/ACM Trans. Netw.*, vol. 27, no. 4, pp. 1543–1557, Aug. 2019.

[14] H. Xiang, S. Yan, and M. Peng, "A realization of fog-RAN slicing via deep reinforcement learning," *IEEE Trans. Wireless Commun.*, vol. 19, no. 4, pp. 2515–2527, Apr. 2020.

[15] G. Sun, Z. T. Gebrekidan, G. O. Boateng, D. Ayepah-Mensah, and W. Jiang, "Dynamic reservation and deep reinforcement learning based autonomous resource slicing for virtualized radio access networks," *IEEE Access*, vol. 7, pp. 45758–45772, 2019.

[16] X. Chen, Z. Zhao, C. Wu, M. Bennis, H. Liu, Y. Ji, and H. Zhang, "Multi-tenant cross-slice resource orchestration: A deep reinforcement learning approach," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 10, pp. 2377–2392, Oct. 2019.

[17] R. Li, "Deep reinforcement learning for resource management in network slicing," *IEEE Access*, vol. 6, pp. 74429–74441, 2018.

[18] *3GPP/WLAN RAN Interworking, Release 12*, document TS 37.834, 3rd Generation Partnership Project, 2014.

[19] Intel. *Alternative LTE Solutions in Unlicensed Spectrum*. Accessed: Nov. 11, 2023. [Online]. Available: https://www.intel.com/content/dam/www/public/us/en/documents/white-papers/unlicensed-lte-paper.pdf.Accessed

[20] J. Chen, Q. Yu, B. Chai, Y. Sun, Y. Fan, and X. Shen, "Dynamic channel assignment for wireless sensor networks: A regret matching based approach," *IEEE Trans. Parallel Distrib. Syst.*, vol. 26, no. 1, pp. 95–106, Jan. 2015.

[21] S. Zhang, "An overview of network slicing for 5G," *IEEE Wireless Commun.*, vol. 26, no. 3, pp. 111–117, Jun. 2019.

[22] S. Wijethilaka and M. Liyanage, "Survey on network slicing for Internet of Things realization in 5G networks," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 2, pp. 957–994, 2nd Quart., 2021.

[23] W. Li, "A static resource allocation scheme for virtualized radio access networks," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 74–79, 2016.

[24] X. Costa-Perez, J. Swetina, T. Guo, R. Mahindra, and S. Rangarajan, "Radio access network virtualization for future mobile carrier networks," *IEEE Commun. Mag.*, vol. 51, no. 7, pp. 27–35, Jul. 2013.

[25] N. Correia, F. Al-Tam, and J. Rodriguez, "Optimization of mixed numerology profiles for 5G wireless communication scenarios," *Sensors*, vol. 21, no. 4, p. 1494, Feb. 2021.

[26] A. Esmaeily, H. V. K. Mendis, T. Mahmoodi, and K. Kralevska, "Beyond 5G resource slicing with mixed-numerologies for mission critical URLLC and eMBB coexistence," *IEEE Open J. Commun. Soc.*, vol. 4, pp. 727–747, 2023.

[27] Y. Prathyusha and T.-L. Sheu, "Coordinated resource allocations for eMBB and URLLC in 5G communication networks," *IEEE Trans. Veh. Technol.*, vol. 71, no. 8, pp. 8717–8728, Aug. 2022.

[28] M. Setayesh, S. Bahrami, and V. W. S. Wong, "Resource slicing for eMBB and URLLC services in radio access network using hierarchical deep learning," *IEEE Trans. Wireless Commun.*, vol. 21, no. 11, pp. 8950–8966, Nov. 2022.

[29] M. Alsenwi, N. H. Tran, M. Bennis, S. R. Pandey, A. K. Bairagi, and C. S. Hong, "Intelligent resource slicing for eMBB and URLLC coexistence in 5G and beyond: A deep reinforcement learning based approach," *IEEE Trans. Wireless Commun.*, vol. 20, no. 7, pp. 4585–4600, Jul. 2021.

[30] K. Boutiba, M. Bagaa, and A. Ksentini, "Optimal radio resource management in 5G NR featuring network slicing," *Comput. Netw.*, vol. 234, Oct. 2023, Art. no. 109937.

[31] D. Yan, B. K. Ng, W. Ke, and C.-T. Lam, "Deep reinforcement learning based resource allocation for network slicing with massive MIMO," *IEEE Access*, vol. 11, pp. 75899–75911, 2023.

[32] A. Gharehgoli, A. Nouruzi, N. Mokari, P. Azmi, M. R. Javan, and E. A. Jorswieck, "AI-based resource allocation in end-to-end network slicing under demand and CSI uncertainties," *IEEE Trans. Netw. Service Manage.*, vol. 20, no. 3, pp. 3630–3651, Sep. 2023.

[33] N. A. Elmosilhy, M. M. Elmesalawy, and A. M. A. Elhaleem, "User association with mode selection in LWA-based multi-RAT HetNet," *IEEE Access*, vol. 7, pp. 158623–158633, 2019.

[34] S. Singh, M. Geraseminko, S.-P. Yeh, N. Himayat, and S. Talwar, "Proportional fair traffic splitting and aggregation in heterogeneous wireless networks," *IEEE Commun. Lett.*, vol. 20, no. 5, pp. 1010–1013, May 2016.

[35] S. Singh, S.-p. Yeh, N. Himayat, and S. Talwar, "Optimal traffic aggregation in multi-RAT heterogeneous wireless networks," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC)*, May 2016, pp. 626–631.

[36] I. Balan, E. Perez, B. Wegmann, and D. Laselva, "Self-optimizing adaptive transmission mode selection for LTE-WLAN aggregation," in *Proc. IEEE 27th Annu. Int. Symp. Pers. Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2016, pp. 1–6.

[37] B. Liu, Q. Zhu, and H. Zhu, "Delay-aware LTE WLAN aggregation in heterogeneous wireless network," *IEEE Access*, vol. 6, pp. 14544–14559, 2018.

[38] H. Wu, S. Cheng, Y. Peng, K. Long, and J. Ma, "IEEE 802.11 distributed coordination function (DCF): Analysis and enhancement," in *Proc. IEEE Int. Conf. Commun.*, vol. 1, Apr. 2002, pp. 605–609.

[39] C. J. C. H. Watkins and P. Dayan, "Q-learning," *Mach. Learn.*, vol. 8, nos. 3–4, pp. 279–292, 1992, doi: 10.1007/BF00992698.

[40] Y. Wu, G. Zhao, D. Ni, and J. Du, "Dynamic handoff policy for RAN slicing by exploiting deep reinforcement learning," *EURASIP J. Wireless Commun. Netw.*, vol. 2021, no. 1, pp. 1–17, Dec. 2021.

[41] S. Hart and A. Mas-Colell, "A simple adaptive procedure leading to correlated equilibrium," *Econometrica*, vol. 68, no. 5, pp. 1127–1150, Sep. 2000.

[42] L. Chen, "A distributed access point selection algorithm based on no-regret learning for wireless access networks," in *Proc. IEEE 71st Veh. Technol. Conf.*, May 2010, pp. 1–5.

[43] D. Athukoralage, I. Guvenc, W. Saad, and M. Bennis, "Regret based learning for UAV assisted LTE-U/WiFi public safety networks," in *Proc. Global Commun. Conf. (GLOBECOM)*, Dec. 2016, pp. 1–7.

[44] S. T. V. Pasca, R. A. Srivats, A. A. Franklin, and B. R. Tamma, "Optimal placement of colocated and non-colocated LWA nodes in dense deployments," in *Proc. IEEE Int. Conf. Adv. Netw. Telecommun. Syst. (ANTS)*, Dec. 2017, pp. 1–6.

[45] M. Anany, M. M. Elmesalawy, and A. M. A. El-Haleem, "Matching game-based cell association in multi-RAT HetNet considering device requirements," *IEEE Internet Things J.*, vol. 6, no. 6, pp. 9774–9782, Dec. 2019.

[46] G. Sun, K. Xiong, G. O. Boateng, G. Liu, and W. Jiang, "Resource slicing and customization in RAN with dueling deep Q-network," *J. Netw. Comput. Appl.*, vol. 157, May 2020, Art. no. 102573.

[47] MathWorks. *Options for Q-Learning Agent MATLAB, RlQAgentOptions*. Accessed: Dec. 1, 2023. [Online]. Available: https://www.mathworks.com/help/reinforcement-learning/ref/rl.option.rlqagentoptions.html.Accessed

IBRAHIM I. IBRAHIM received the B.Sc. degree in electronics and communications engineering from Helwan University, Cairo, Egypt, in 1976, the M.Sc. degree in electronics and communications engineering from Cairo University, Egypt, in 1982, and the Ph.D. degree from the Queen's University of Belfast, Belfast, U.K., in 1987. He is currently the former Head of the Electronics, Communication and Computers Engineering Department, and a Professor of wireless communication with the Faculty of Engineering, Helwan University. His current research interests include massive MIMO, 5G and 6G radio access networks wireless communication, heterogeneous cellular networks, satellite communications, the Internet of Things, visible light communications, cloud radio access networks, AI-based applications in wireless communication systems, unmanned aerial vehicle (UAV) communications, and reconfigurable intelligent surface (RIS). He is a former member of the National Communication and Electronics Engineering Promotion Committee.

NOHA A. ELMOSILHY received the B.S. degree in electronics and communication engineering from Canadian International College, Cairo, Egypt, in 2011, and the M.S. degree in electronics and communication engineering from the Arab Academy for Science, Technology and Maritime Transport, Cairo, in 2016. She is currently pursuing the Ph.D. degree in electronics and communications engineering with Helwan University, Cairo. Since 2011, she has been a Teacher Assistant with the Electronics and Communication Department, Canadian International College. Her research interests include 5G, B5G, and 6G access networks, the IoT, LWA, AI-based applications in wireless communication systems, and WLAN resource management. She is a member of the Research Team that awarded several applied research projects funded by national and international funding agencies in the field of wireless communication and smart education systems.

MAHMOUD M. ELMESALAWY (Member, IEEE) was born in Cairo, Egypt, in 1981. He received the B.S., M.S., and Ph.D. degrees in electronics and communications engineering from Helwan University, Cairo, in 2002, 2005, and 2010, respectively. He is currently a Professor of wireless communications with the Department of Electronics and Communications Engineering and the Dean of the Faculty of Engineering, Helwan University. He led the effort to build the cellular-based communication infrastructure of the Egyptian wide area monitoring system (EWAMS) that is developed on the Egyptian grid and operated by Helwan University. His current research interests include B5G and 6G mobile communication technologies, the Internet of Things (IoT) communications, unmanned aerial vehicle (UAV) communications, reconfigurable intelligent surface (RIS), smart grid communication, and AI-based applications in wireless communication systems.

AHMED M. ABD EL-HALEEM (Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees in electronics and communications engineering from Helwan University, Cairo, Egypt, in 2001, 2006, and 2012, respectively. He is currently an Associate Professor with the Electronics and Communications Engineering Department, Faculty of Engineering, Helwan University, and also a secondment with the Electrical Department, Faculty of Engineering, The British University in Egypt. He has supervised undergraduate and postgraduate students and involved on projects in areas related device-to-device communication, the IoT, and software defined network (SDN), including mobility and resource management. His current research interests include mobile/vehicular ad-hoc communication networks, 5G and 6G radio access networks, cognitive radio networking, device to device communication, the Internet of Things (IoT), reconfigurable intelligent surface (RIS), AI application in wireless communication, mobility management techniques and routing schemes for mobile ad-hoc networks (MANET), security and secure routing algorithms for ad-hoc wireless networks, and routing protocols for cognitive radio communication networks. His research was well received and has been published in several prestigious journals and conferences in the field of electrical and communication engineering, including, IEEE, Springer, *IJNSA*, *IJCSI*, and Hindawi and Wiley journals, with more than 27 journal and conference papers. He is a member of a research teams that awards several applied research projects funded by National and International funding agencies in the field of wireless communication, the IoT and its applications, and smart education systems.

● ● ●