IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received May 11, 2018, accepted June 14, 2018, date of publication July 19, 2018, date of current version August 7, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2850752

Memory-Based User-Centric Backhaul-Aware User Cell Association Scheme

FARRUKH PERVEZ[1], MONA JABER[2], (Student Member, IEEE), JUNAID QADIR[3], (Senior Member, IEEE), SHAHZAD YOUNIS[4], AND MUHAMMAD ALI IMRAN[5], (Senior Member, IEEE)

[1]Department of Avionics Engineering, CAE, National University of Science and Technology, Islamabad 44000, Pakistan
[2]Home of 5G Innovation Centre, Institute for Communication Systems, University of Surrey, Guildford GU2 7XH, U.K.
[3]Department of Electrical Engineering, Information Technology University, Lahore 54000, Pakistan
[4]Department of Electrical Engineering, SEECS, National University of Science and Technology, Islamabad 44000, Pakistan
[5]School of Engineering, University of Glasgow, Glasgow G12 8QQ, U.K.

Corresponding author: Farrukh Pervez (farrukh.pervez@cae.nust.edu.pk)

This work was supported by the EPSRC Global Challenges Research Fund—the DARE Project—under Grant EP/P028764/1.

ABSTRACT Ultra-dense small cell networks represent a key future network solution that can help meet the exponentially rising traffic requirements of modern wireless networks. Backhauling these small cells are an emerging challenge to the extent that various cells are likely to have different backhaul constraints. The user-centric backhaul scheme has been proposed in the literature to jointly exploit the diversity in users' requirement and backhaul constraints. In this paper, we propose a novel scheme, termed the memory-based hybrid scheme, which additionally also exploits the predictability in a user's mobility. We compare the novel scheme to two variants of memory-less user-centric backhaul implementations and show significant gains in convergence time (15%), user-centric KPIs (51% and 82%) at the negligible cost 2% loss in cumulative throughput. The novel scheme requires additional memory in user-devices to store learned values, which is nonetheless well justified in view of the considerable gains achieved.

INDEX TERMS Heterogeneous networks (HetNets), user cell association, user-centric, backhaul, memory-based learning.

I. INTRODUCTION

The proliferation of smartphones in the recent years has led to an astounding growth in the mobile data traffic. It is anticipated that mobile subscriptions will reach 8.9 billion by 2022 [1] thus increasing aggregate required throughput by a factor of 1000 [2]. In addition to this, with the introduction of role-specific disparate devices (ranging from simple sensors to smartphones) in the past few years, the user service requirements have become diverse. Users tend to prefer one performance metric over another due to device capabilities, host application genre, and user preferences. The fifth generation (5G) mobile networks address the aforementioned challenges of the exponential increase in demands for wireless bandwidth as well as users' covetous approach towards high quality service requirements through densification of small cells, known as Heterogeneous Networks (HetNets). However, due to economically infeasible backhaul options that were available at the advent of small cells, the bottleneck has shifted from air interface to the backhaul [3], [4]. With backhaul as the new bottleneck and users' diverse service requirements coupled with their increasing demands for high quality, self optimization has become an inevitable necessity for 5G. Accordingly, self optimization network-based (SON-based) backhaul-aware user cell association techniques with convincing results have been proposed in the literature and are presented in what follows.

Driven by multi-tier heterogeneous cellular networks and growing number of bandwidth-hungry applications, uplink traffic is generated more than downlink traffic in an uncorrelated fashion, and thus state-of-the-art max-received signal strength (RSS)-based UA becomes suboptimal [5]. Hence, Elshaer *et al.* [6] introduce the concept of decoupled downlink/uplink (DUDe) user cell association (UA) for load balancing keeping in view the cell load and backhaul throughput constraints. Simulations demonstrate that DUDe, a heuristics-based UA algorithm, ameliorates throughput for cell edge users as compared to max-RSS-based UA. Similarly, a distributed UA scheme is proposed in [7] that balances the network load while taking into account the backhaul delay and reliability constraints. It is shown through

simulations that the proposed algorithm leads to significant improvements in reducing delay and improving reliability when compared against traditional UA scheme. A joint downlink UA based on sum user rate maximization for two-tier HetNets with regard to wireless backhaul constraints is investigated in [8]. Z and R [9] propose a waterfilling-like UA scheme that is based on users sum rate maximization, while keeping in view the specific backhaul constraints for small cells. A heuristics-based UA algorithm is presented in [10] that addresses backhaul load balancing while imposing constraints on backhaul constraints along with radio conditions. Simulation results signify that the proposed algorithm copes up with backhaul congestion situations better than the traditional UA schemes. Pantisano *et al.* [11] conceive a content-aware game theoretic-based UA algorithm that equips small cells with capabilities of caching multimedia content in order to improve users' quality of experience (QoE) under backhaul throughput constraints.

Although designed to optimize constrained backhaul links, the aforementioned UA schemes nonetheless operate in cell-centric fashion and fail to account for the diverse QoE requirements of users. Keeping in view the dominant role of users' key performance indicators (KPIs) in 5G era [12], a novel concept of user-centric backhaul-aware (UCB) UA is devised in [13]. The UCB scheme is focused on maximizing users' experience, while considering different backhaul attributes such as throughput, latency, reliability, etc. It enables users to associate with a potential cell that could satisfy their QoE requirements. Cells exploit Q-learning to dynamically optimize their cell range expansion offsets (CREOs)—which represent network end-to-end constraints and capabilities—according to network and user indicators. It is shown through simulations that the proposed UCB scheme leads to substantial improvements in users' QoE, while falling fractionally short of the capacity achieved with cell-centric backhaul-aware schemes. In our previous work [3], we presented an alternate way of implementation of UCB scheme, where we equip small cells with fuzzy Q-learning capabilities to virtually tailor their range as per users' service requirements. The inherent flexibility of fuzzy Q-learning coupled with a well devised exploitation-exploration strategy enables fuzzy Q-learning-based UCB scheme to yield superior results and makes it computationally much more efficient as compared to basic UCB scheme.

Though having significantly diverse service requirements, users' exhibit a high level of predictability in their mobility patterns [14]. Users usually visit business areas or working places during daytime or weekdays while return to residential areas at nights or on weekends [15]. Studies reveal that users' mobility depends on historical patterns and can be predicted with an accuracy of 93-95% even for those who travel over long routes/distances [14], [16]. A few recently proposed UA schemes explored the aspect of exploiting predictable users' mobility patterns for delivering high-quality services. For instance, Cacciapuoti [17] propose a distributed mobility-aware UA strategy for 5G mmWave networks that takes into account the load of the small cells. Simulation results validate superior performance of the proposed UA algorithm as compared to the traditional RSS-based UA. In [18], a two-step mobility-aware hybrid scheme is proposed, which involves both user and the network, for radio access technology (RAT) selection in a multi-RAT network. In the first step, a user selects a list of best available networks based on RSS and user mobility profile. In the second step, the network associates the user to an appropriate RAT based on multi criteria. Simulation results indicate high precision achieved with the proposed scheme than that with the traditional approach. The aforementioned mobility-aware schemes, however, do not consider backhaul constraints for UA and hence may lead to unsatisfactory users' performance [10]. Accordingly, we propose a novel memory-based user-centric backhaul-aware UA scheme that exploits high predictability in users' mobility patterns to guarantee user-centric services, while respecting backhaul constraints. In our scheme, cells broadcast multiple optimized bias factors with each bias factor representing an end-to-end network capability and constraint. In our proposed scheme, cells employ a hybrid approach, exploiting strengths of both fuzzy Q-learning and Q-learning, to dynamically tailor their virtual footprints according to the network and users indicators. The novelty of this work is that cells, equipped with cognizant abilities, are able to store and subsequently broadcast optimal bias values corresponding to different users' mobility patterns as well as QoE profiles in order to have a significant reduction in optimization time. On the other hand, users keep a location-based history of normalized measured QoE and utilize the same besides user-centric bias values to select a cell that potentially meets their QoE requirements.

The main contribution of this paper is the novel memory-based user-centric backhaul-aware UA scheme which investigates memory-awareness both from small cells and users' perspective. Our proposed scheme yields remarkable gains in users' Key Performance Indicators (KPIs), which would dominate the 5G era [12], as compared to Memory-Less (ML) Q-learning-based scheme [13] and ML fuzzy Q-learning-based scheme [3]. Furthermore, our proposed scheme is computationally more efficient than ML schemes and thus practically relevant to 5G since it enables dense number of small cells, one of the key enablers of 5G, to simultaneously optimize bias values.

The rest of the paper is organized as follows. Section II presents the system model, whereas distributed SON-based UA problem is formulated in Section III. Our proposed memory-based scheme supported by preliminary results, analysis and insights is described in detail in Section IV. We conclude the paper in Section V.

II. SYSTEM MODEL

A two-tier downlink HetNet comprising N small cells overlaid within C macro cells is considered. The geographic locations of small cells and macro cells are defined by (x_n, y_n), $n \in \{1, 2, \ldots, n, \ldots, N\}$ and (x_c, y_c),

$c \in \{1, 2, \ldots, c, \ldots, C\}$ respectively. Each cell gets an equal share of a set of K resource blocks, where each resource block consists of s subcarriers each having a bandwidth of B Hz, to deliver services to its associated users. Every cell broadcasts O number of CREOs, each CREO corresponds to different end-to-end limitations and capabilities of the network such as throughput, latency, energy efficiency, etc. A high CREO value signifies that the cell is capable to deliver better end-to-end network performance, whereas a low CREO value represents end-to-end constraint of the network for the respective CREO. Heterogeneous technologies including optical fiber, G.fast and microwave are utilized for backhauling small cells, however, each small cell is provisioned with only one last-mile backhaul link to connect to the core network. We impose this restriction in order to inhibit the complexity of the routing algorithm that increases with increase in the number of last-mile backhaul links between a small cell and core network. We finally assume that all the small cells have the same RAN architecture (D-RAN, C-RAN, F-RAN).

We randomly distribute a total of U users with diverse quality of service requirements in the system. Each user associates relative weights W_q with different QoS metrics Q, termed as QoE targets, depending upon its application requirements, preferences, and device specifications. We assume that the number of QoE parameters of each user is equal to the number of CREOs broadcasted by the cells. However, a nil value is assigned to the irrelevant CREO or QoE parameter. Users' mobility is modelled according to random waypoint (RWP) mobility model in which a user reaches a randomly chosen destination waypoint l with a constant speed. We consider different users' mobility patterns $M = \{m_1, m_2, \ldots m_u, \ldots, m_U\}$, where each mobility pattern $m_i, i \in \{1, 2, \ldots u \ldots, U\}$ contains distinguished number of users preferring high quality services. For the scope of this paper, we redefine the term mobility pattern from a small cell's perspective as the unique number of localized users categorized according to their high and/or low weight associated to various performance metrics. For instance, if we take the case of two attributes namely throughput and reliability then mobility pattern is the total number of users, whose locations are known, grouped under the categories of high-high, high-low, and low-high weights for throughput and reliability respectively. For this purpose, we propose a paging system [19] for macro cells that periodically collects information about users' location as well as weights allocated to the attributes and disseminates the same to the small cells. Since the focus of this paper is UA scheme, a descriptive detail of the paging system is beyond the scope of this paper. We illustrate two-tier system model comprising C macro cells and N small cells in Figure 1.

We propose that small cells are able to optimize and subsequently keep bias values against various mobility patterns in their memory. The set of mobility patterns against which the bias offsets are already stored in the memory of a small cell n is referred as $M_n = \{m_{1,n}, m_{2,n}, \ldots m_{u,n}, \ldots, m_{U,n}\}$ with an

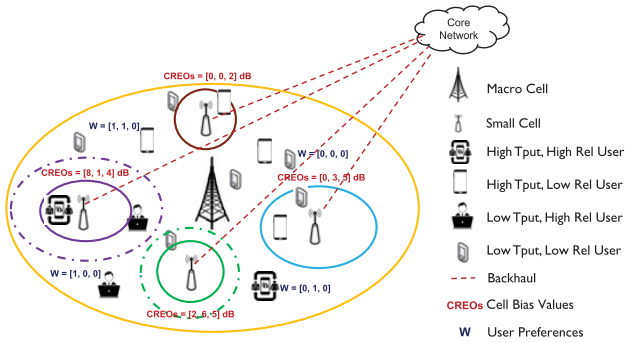


FIGURE 1. System model with C macro cells and N small cells.

associated parameter a_{M_n} such that

$$a_{m_i} = \begin{cases} 1, & if\ m_{i,n} \in (M \cap M_n) \\ 0, & if\ m_{i,n} \notin M_n \end{cases} \tag{1}$$

To exploit the predictability in mobility pattern, each user keeps a location-based history of its measured QoE when served by the cell n and utilizes this information while ranking potential candidate cells for association purpose. We now elaborate the system model in the ensuing paragraphs with the help of related equations.

The signal strength received on the downlink by a user u from a small cell n over a resource block $k, where k \in \{1, 2, \ldots, k, \ldots, K\}$, is given by (2).

$$RSS_{n,u,k} = \frac{P_n}{U_n} \times H_{n,u,k} \tag{2}$$

$$H_{n,u,k} = \chi_n \times d_{n,u}^{-\alpha_n} \times \varepsilon_{n,u,k} \tag{3}$$

where U_n represents the total number of users served by the cell n and P_n is the total transmitted power of the cell n. To emphasize upon the association problem and highlight the potential gains of our proposed scheme, we have assumed that only one resource block can be assigned to a user and, hence, P_n instead of $P_{n,k}$—the power transmitted by cell n over resource block k—is utilized in (2). The channel gain between the transmitter, that is small cell n in this case, and the receiver, the user u, is represented by $H_{n,u,k}$ and is given by the mathematical expression in (3). The power of the received signal deteriorates in accordance with log-distance path loss model, where α_n and χ_n donate the path loss exponent and propagation constant that are specific to the cell n. The Euclidean distance between the cell n and the user u, with geographic locations represented by (x_n, y_n) and (x_u, y_u) respectively, is given by $d_{n,u}$. Log-normal shadowing, represented by $\varepsilon_{n,u,k}$, is assumed to be similar on all sub-channels allocated to the single user u by the small cell n in a resource block k. It is pertinent to mention that we do not include multi-path fading in the channel gain since the subject fading varies rapidly than the time required to adjust a CREO and hence gets averaged out.

Each user ranks all the cells including macro cells according to the criteria given by (4)

$$R_{u,n} = \begin{cases} RSS_{n,u,k} + \sum_{q=1}^{|O|} w_{u,q} v_{n,q} \\ +w_u(\sum_{q=1}^{|O|} w_{u,q} \overline{Q}'_{l,n,u,q} \bar{t}), & if\ n \in N \\ RSS_{n,u,k}, & otherwise \end{cases} \tag{4}$$

where $w_{u,q}$ is the weight the user u allocates to the QoE attribute q and $v_{n,q}$ is the respective bias value broadcasted by the cell n. denotes The time-averaged normalized perceived QoE of the user u served by the cell n at a location l is denoted by $\overline{Q}'_{l,n,u,q}$, and $\bar{t}$ is the time-averaged number of associations occurred between user u and cell n during last d visits to a location l. To avoid overshadowing of user-centric values, a factor of w_u is multiplied with QoE history values in (4). The user attempts to get associated with the highest ranked to the lowest ranked candidate cell until it finds a cell, known as server, with available resource blocks or is declared out of coverage. In case, a user u successfully associates to a cell n over $\{1, 2, \ldots, k, \ldots K\}$ resource blocks, the corresponding SINR is given as follows:

$$SINR_{n,u,k} = \frac{RSS_{n,u,k}}{\sigma^2 + \sum_{i=1,i\neq n}^{N} RSS_{i,u,k} + \sum_{j=1,j\neq n}^{M} RSS_{j,u,k}} \tag{5}$$

where σ^2 defines the noise power in the received signal. The interference caused by macro cells and remaining small cells is expressed by $\sum_{j=1,j\neq n}^{M} RSS_{j,u,k}$ and $\sum_{i=1,i\neq n}^{N} RSS_{i,u,k}$ respectively. The theoretical bound on the throughput achieved by the user u corresponding to the above-mentioned SINR is represented by Shannon's theorem in (6).

$$T_{n,u,k} = B \log_2(1 + SINR_{n,u,k}) \tag{6}$$

where the bandwidth of resource block allocated to the user u is defined by B. The required backhaul throughput BH_n is computed on the basis of the over-the-air throughput achieved by the associated users of cell n and is given by (7).

$$BH_n = \sum_{u=1}^{|U_n|} T_{u,n,k} \times G_n \tag{7}$$

where the factor G_n represents the signalling overhead that varies based on the RAN architecture, backhaul topology and, technology of the cell n. Since the bottleneck has shifted from radio to the backhaul, the effective throughput achieved by the user u now depends on the constrained backhaul capacity δ_n of the serving cell n. We propose that all users associated with a cell n suffer a uniform reduction in their effective throughput, according to (9), if the required backhaul throughput becomes larger than the available backhaul capacity thus resulting in congested backhaul links.

$$S = \delta_n - BH_n \tag{8}$$

$$T'_{n,u,k} = \begin{cases} T_{n,u,k}, & if\ S \geq 0 \\ T_{n,u,k} - \frac{S}{|U_n|}, & otherwise \end{cases} \tag{9}$$

III. PROBLEM FORMULATION

The centralized optimization techniques aim to maximize the system-level objective function based on a given set of constraints and hence may become intractable with an increase in the number of CREOs, the number of intelligent cells per cluster or the number of available last-mile backhaul links. To this end, we propose the distributed optimization that enables cells to individually maximize their throughput while respecting the identified constraints. In other words, the objective is to maximize the network capacity while exploiting the predictability in users' mobility patterns to keep QoE shortage experienced by users to the minimum possible under backhaul constraints. Distributed optimization has the following benefits over centralized optimization.

Firstly, distributed optimization reduces complexity from $O(V.O.N)$ to $O(V.O)$, where V defines the range of bias values, a CREO can be assigned, O is the total number of CREOs, and N represents the number of small cells included in the cluster and aiming to achieve optimization objective.

Secondly and more importantly, the complexity of distributed optimization does not depend on the number of intelligent small cells available in the network. This point is critical, since practical deployments of the 5G network would likely deploy ultra dense small cells, for the following reasons. 1) Unlike centralized optimization, which requires to collect data on a network level through a lengthy and sometimes cumbersome process, distributed optimization does not need a network-wide view or data collection and hence converges at a faster rate than the centralized optimization. 2) Distributed optimization dynamically adapts to network changes such as addition of nodes, modification of topology, introduction of radio features etc.

We articulate the distributed optimization problem in the following equations

$$\max_{s_n} T_n(s_n), \quad \forall n \in N \tag{10}$$

$$T_n = \sum_{u=1}^{|U_n|} T'_{n,u,k} \tag{11}$$

subject to the following identified constraints:

$$\delta_n - BH_n \geq 0 \quad n \in N \tag{12}$$

$$UQ'_{n,q} = \sum_{u=1}^{U_n} w_{u,q} \frac{Q'_{u,q} - Q_{u,q}}{Q_{u,q}} \leq \phi_q \quad Q_q \in Q_u,\ n \in (N \cup C) \tag{13}$$

$$a_{m_i} = 1 \quad \forall m_i \in M_n,\ n \in (N \cup C) \tag{14}$$

$$v_{n,q,m_i} \times a_{m_i} = v^*_{n,q,m_i} \quad n \in (N \cup C), q \in Q \tag{15}$$

where T_n represents the capacity of the cell n and is computed through sum of effective throughput achieved by the associated users of cell n, please refer to (11) for its expression. $s_n = [V_{n,1}, V_{n,2}, \ldots, V_{n,o}, \ldots, V_{n,O}]$ represents the possible combinations of bias values for cell n. Cells consider the backhaul constraint, defined by (12), while aiming to meet the minimum QoE requirements of the users. Furthermore,

cumulative QoE gap (the difference between perceived and target QoE) of all users, $UQ'_{n,q}$, is required to be kept below the minimum preset QoE threshold ϕ_q, as represented by (13). $Q'_{u,q}$ is the measured QoE, whereas QoE $Q_{u,q}$ is the target QoE of a user u for an attribute q. A users' mobility pattern encountered more than once by the cell n should be stored in its memory with associated parameter a_{m_i} having value as 1. Moreover, the bias values kept against the stored mobility patterns should be the optimal ones v^*_{n,q,m_i}, as mentioned in (15).

IV. PROPOSED MEMORY-BASED USER-CENTRIC BACKHAUL-AWARE UA SCHEME

It has been shown that users, though having significantly diverse service requirements, exhibit high predictability in their mobility patterns. In 5G era, where user satisfaction would be one of the crucial parameters, the regularity in users' mobility patterns may be exploited by small cells to deliver content-aware services. The main challenge in this regard, however, stems from considerable diversity that exists in 5G in terms of heterogeneous backhaul solutions, transmit power disparity radio access cells, disparate devices and applications, in addition to the new backhaul bottleneck. Hence, it becomes inevitable for small cells to have intelligence in order to guarantee mobility-aware user-centric services keeping in view the capabilities and constraints of the backhaul network. ML Q-learning-based UCB scheme, when compared against cell-centric backhaul-aware schemes, improves users' QoE significantly while falling fractionally short (0.064%) in terms of network capacity [13]. However, it suffers with impractical increase in iteration time with an increase in the number of small cells per cluster and thus becomes computationally limiting [3]. ML fuzzy Q-learning-based UCB [3] scheme not only overcomes this limitation through a well-devised exploration-exploitation strategy but also improves users QoE significantly (12%), while lagging behind negligibly (1.96%) in terms of overall system capacity as compared to Q-learning-based UCB scheme.

Our proposed scheme, exploiting regularity in users' mobility patterns, is a two-phase scheme—memory-less UCB and memory-aware UCB—that captures the benefits of both ML fuzzy Q-learning-based (substantial gains in users' QoE and computational efficiency) and Q-learning-based schemes (comparable network capacity). Memory-less UCB enables cells to learn bias values using fuzzy Q-learning for any new users' mobility pattern, whereas memory-aware UCB exploits Q-learning to optimize already learnt bias values for any mobility pattern. Both memory-less UCB and memory-aware UCB phases of our proposed scheme have been expounded upon in the subsequent paragraphs.

Memory-less UCB deals with learning of CREOs whenever a mobility pattern is encountered for the very first time. We exploit the following strengths of fuzzy Q-learning to address the problem of memory-less UCB. *Firstly*, fuzzy Q-learning is suitable to the problems where state (input and action) space are continuous in nature. It associates fuzzy sets to the real-valued state space variables thus, sometimes, resulting in reduced state-action pairs. *Secondly*, the adjacent linguistic variables of a fuzzified input overlap partially to render more flexibility to fuzzy Q-learning besides smoothness and robustness [20]. *Thirdly* and more significantly, employing fuzzy Q-learning in memory-less UCB leads to convincing improvements in users' KPIs than that of Q-learning [3]. The optimized CREO values learnt in this step are stored in the memory against respective mobility patterns for subsequent use.

In this phase, a cell acting as fuzzy Q-learning agent learns to map CREOs values (actions) against articulated fuzzy rules through repeated interaction with its environment, which includes dynamic nature of radio conditions, backhaul throughput variations, and users' traffic patterns. To this end, we fuzzify the two inputs namely **required backhaul throughput (BH_n, req_bh_tput)** and **cumulative users' QoE gap ($UQ'_{n,q}$, cum_users'_qoe _shortage)**. Based upon the first two constraints of the identified optimization problem, we assign two linguistic variables with trapezoidal membership functions to each of the fuzzified inputs. We keep the number of linguistic variables to two for each fuzzified input due to the following reasons. *Firstly*, based on linguistic variables, fuzzy Q-learning builds fuzzy control rules and stores a quality matrix in memory for depicting appropriateness of each fuzzy rule with respect to each conclusion [21], [22]. Hence, less number of linguistic variables means less number of fuzzy rules that ultimately leads to reduced memory requirements. *Secondly*, to keep computation time that increases exponentially with increase in the number of fuzzy rules within reasonable bounds. Furthermore, a trial and error approach is applied to reach the selected core width and boundary regions of the membership functions; too much overlapping of adjacent linguistic variables lead to frequent activation of multiple fuzzy rules at a time thus increasing computation time drastically without having a significant effect on performance, while too less overlapping of adjacent linguistic variables compromises smoothness and flexibility. The salient details regarding memory-less UCB are presented in Table 1.

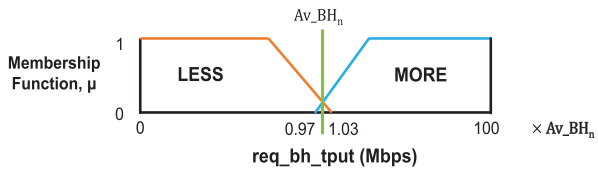
Memory-aware UCB phase gets activated whenever a mobility pattern, for which a cell already keeps optimal values learnt through the memory-less phase in its memory, is observed again. In this part, we employ Q-learning to verify the optimality of stored CREOs values. The motivation behind using Q-learning for this step is its simplicity and efficient handling of problems having discrete input and/or action space. In the context of this phase, cells equipped with Q-learning capability tend to reach optimal CREOs values (actions) for various states of its dynamic radio environment. For the current step, we modify action space by appending an additional action X_M, named as memory check, to it that represents that a cell would straightaway look for the CREOs values in its memory instead of indulging itself in exploitation and exploration phase. We present the salient

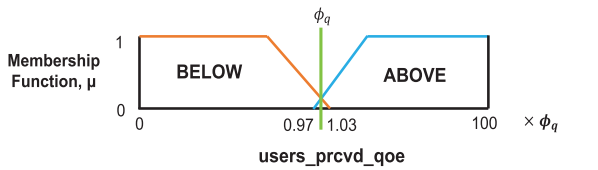
TABLE 1. Salient details of the memory-less UCB phase of our proposed scheme.

Aim – To learn assigning optimal conclusions (CREOs) against each of the articulated fuzzy rules in dynamic radio environment

Fuzzy Q-Learning Agents – All small cells n ∈ N

Fuzzified Inputs – req_bh_tput and $cum\_users'\_qoe\_shortage$





Fuzzy Rule Base
IF req_bh_tput is **LESS AND** $cum\_users'\_qoe\_shortage$ is **BELOW THEN** conc is $\cdots$
IF req_bh_tput is **MORE AND** $cum\_users'\_qoe\_shortage$ is **BELOW THEN** conc is $\cdots$
IF req_bh_tput is **LESS AND** $cum\_users'\_qoe\_shortage$ is **ABOVE THEN** conc is $\cdots$
IF req_bh_tput is **MORE AND** $cum\_users'\_qoe\_shortage$ is **ABOVE THEN** conc is $\cdots$

Conclusions – The set of possible CREO values A_n represented by the vector $a_n = \{0, 1, 2, 3, 4, 5, 6\}$

Optimization Parameters – CREOs representing end-to-end throughput and reliability of the system

Reward – The immediate reward R is computed as follows

$$R = \begin{cases} 1000 \cdot \frac{\delta_n - BH_n}{\delta_n}, & if \quad \delta_n > BH_n \quad and \quad UQ'_{n,q} < \phi_q \\ -1000 \cdot (BH_n - \delta_n), & if \quad BH_n > \delta_n \\ -100 \cdot \sum_{u=1}^{U_n} w_{u,q} \frac{Q'_{u,q} - Q_{u,q}}{Q_{u,q}}, & if\, \delta_n > BH_n \, and \, UQ'_{n,q} > \phi_q \end{cases} \tag{16}$$

Cumulative Rule-Conclusion Reward – The cumulative reward for each rule–conclusion pair, $q(\phi_i, A_i)$, is updated using (17)

$$q(\phi_i, A_i) = q(\phi_i, A_i) + \eta \Delta q(\phi_i, A_i) \tag{17}$$

where

$$\Delta q(\phi_i, A_i) = \frac{w_i}{\sum_{k=1}^{m} w_k} \cdot (r(x, H(x)) + \gamma FQ^*(\hat{x}, H(\hat{x})) - FQ(x, H(x))) \tag{18}$$

η and γ represent learning rate and discount factor respectively

TABLE 2. Salient details of the memory-aware UCB phase of our proposed scheme.

Aim – To learn mapping optimal actions (CREOs) to each of the states in dynamic radio environment

Q-Learning Agents – All small cells n ∈ N

States – A cell may present in one of the following states

$$\Delta_n = \begin{cases} 1, & if \quad BH_n > \delta_n \\ 2, & if \quad UQ'_{n,q} \geq \phi_q \quad Q_q \in Q_u, \quad n \in (N \cup C) \\ 3, & if \quad \delta_n > BH_n \quad and \quad UQ'_{n,q} < \phi_q \end{cases} \tag{19}$$

$\Delta_n = 3$ is the objective state

Actions – The set of possible CREOs values A_n along with an additional action X_M represented by the vector $a_n = \{0, 1, 2, 3, 4, 5, 6, X_m\}$

Optimization Parameters – CREOs values that reflect end-to-end network performance in terms of throughput and reliability

Reward – The immediate reinforcement signal is computed as follows

$$R = \lceil \frac{100}{t_a} \rceil + \begin{cases} 1000 \cdot \frac{\delta_n - BH_n}{\delta_n}, & if \Delta_n = 3 \\ -1000 \cdot (BH_n - \delta_n), & if \Delta_n = 1 \\ -100 \cdot \sum_{u=1}^{U_n} \frac{Q'_{u,q} - Q_{u,q}}{Q_{u,q}}, & if \Delta_n = 2 \end{cases} \tag{20}$$

where $\lceil \cdot \rceil$ represents the ceiling function and t_a is the total time a cell takes to complete an iteration; the cell goes through the exploration-exploitation phase to select an action a, executes it, and receives a reward for the respective action

Cumulative State-Action Reward – Cumulative state-action reward for respective state and action is updated using (21)

$$Q_{t+1}(s, a) = (1 - \phi) \cdot Q_t(s, a) + \phi \cdot (r_t + \gamma \max_{a' \in A} Q(s', a')) \tag{21}$$

TABLE 3. List of simulation parameters.

Parameter	Value/Description
Number of macro cells	1
Number of small cells	9
Number of users	180
Total subcarriers	600
Number of subcarriers in a single resource block	12
Subcarrier bandwidth	180 KHz
Number of mobility patterns	2
Backhaul attributes	Throughput, reliability
Backhaul options	G.fast, microwave, optical fiber
Simulation platform	MATLAB 8.2

details of memory-aware UCB in Table 2. Employing fuzzy Q-learning and Q-learning for memory-less and memory-aware phases respectively enables our proposed scheme to exploit the strengths of both these techniques and makes our proposed scheme a hybrid one on the whole.

The optimal policy in Q-learning and fuzzy Q-learning is governed by choosing the action that maximizes cumulative reward over a time. However, applying optimal policy too early in the learning phase may lead to local minima since Q or FQ values are not significant in the beginning. Hence, it becomes necessary that all possible actions against every state must be evaluated before sticking to optimal policy; these phases are called as exploration and exploitation respectively. Nonetheless, in multi-agent environment (small cells) with a large set of actions (CREO values) complete exploration may consume ample amount of time, whereas on the other hand, implementing optimal policy too early may yield undesirable results. Though statistical methods like epsilon greedy and Boltzmann selection address this issue [23], we implement a simple reward-based exploration and exploitation strategy in our proposed scheme, for both the phases, without compromising on its optimality. Our proposed scheme frequently alternates between exploration and exploitation. In other words, the proposed scheme sequentially explores and immediately shifts to exploitation if the last exploration had yielded negative reward. Besides this, we also ensure that cells in memory-less phase get enough number of iterations to complete exploration phase and reach optimal strategy before moving to the next phase.

Based upon the discussions in the above-mentioned paragraphs, we present basic working of our proposed scheme for a single cell in Table 4.

A. SIMULATION RESULTS

To provide a proof of concept, we implement the proposed memory-based UCB on a simulated downlink HetNet,

TABLE 4. Basic working of memory-based UCB scheme.

Memory-Based UCB Scheme

1. Initialize memory Z for storing CREOs against mobility patterns
2. Initialize Q and FQ tables
3. **for** run=1,total_runs **do**
4. Get info about current mobility pattern m_i from the proposed paging system
5. **if** m_i in Z **do**
6. $a_{m_i} = 1$
7. **else do**
8. $a_{m_i} = 0$
9. **end if**

10. **if** $a_{m_i} == 0$ **do**
11. Observe current input parameters BH_n and $UQ'_{n,q}$ and respective current fuzzy state Γ_n
12. For each activated fuzzy rule ϕ_i, select a conclusion a_n based on implemented exploration-exploitation methodology
13. Compute the inferred action $H(\Gamma_n)$ using (22) and execute it

$$H(\Gamma_n) = \frac{\sum_{i=1}^{m} w_i \cdot A_i}{\sum_{i=1}^{m} w_i} \tag{22}$$

where w_i is the rule firing strength
14. **if** run == nth_run **do**
15. Store action $H(\Gamma_n)$ in Z against pattern m_i
16. **end if**
17. Compute FQ value for the current state $FQ(\Gamma_n, H(\Gamma_n))$

$$FQ(\Gamma_n, H(\Gamma_n)) = \frac{\sum_{i=1}^{m} w_i \cdot q(\phi_i, A_i)}{\sum_{i=1}^{m} w_i} \tag{23}$$

18. Observe new fuzzy state Γ'_n and obtain reward R_n
19. Compute FQ value for the current state based upon optimal policy

$$FQ^*(\hat{\Gamma}_n, H(\hat{\Gamma}_n)) = \frac{\sum_{i=1}^{m} w_i \cdot q(\phi_i, A_i^*)}{\sum_{i=1}^{m} w_i} \tag{24}$$

where $A_i^* = arg\max_{A_j}\{q(\phi_i, A_j)\}$
20. Update the quality values in FQ table using (17)
21. **else if** $a_{m_i} == 1$ **do**
22. **if** run $>$ nth_run and run $\leq$ nplusk_runs **do**
23. Observe current state Δ_n
24. Select an action a_n according to implemented exploration-exploitation policy
25. Perform a_n, observe new state Δ'_n and receive immediate reinforcement signal R_n
26. **if** run == nplusk_runs **do**
27. Store action a_n in Z against pattern m_i
28. **end if**
29. Update Q values corresponding to Δ_n and a_n in Q table using (21)
30. **else if** run $>$ nplusk_runs **do**
31. Execute action stored in Z against m_i
32. **end if**
33. **end if**
34. **end for**
35. Repeat this process for other mobility patterns

consisting of 9 small cells deployed alongside 1 macro cell, using MATLAB 8.2. A set of 50 resource blocks is distributed equally amongst small cells and macro cells, whereas each resource block contains 12 subcarriers. Each subcarrier has a bandwidth of 180 KHz. We map users' QoE to two performance attributes namely throughput and reliability. Small cells employ hybrid memory-based scheme to learn and cache optimal CREOs values against different users' mobility patterns in their memory. Heterogeneous backhaul technologies such as G.fast, microwave, and optical fiber are considered for backhauling small cells, however, every small cell is provided with only one last-mile backhaul link. We assume that macro cell is leveraged with an ideal backhaul that carries aggregated traffic from small cells to the core network. Our aim is to push the traffic to lower layers when possible, in line with HetNets functionality. To this end, the CREO of the macrocell is always set to zero as it acts as a fall-back plan in the absence of a suitable small cell. The list of simulation parameters is summarized in Table 3.

We simulate the proposed scheme using Monte Carlo technique comprising 50 simulation runs. We consider two users' mobility patterns in which users' follow RWP mobility model. For simplicity, we make an assumption that patterns neither overlap with each other nor do they repeat frequently in order to reduce memory requirements at small cells; small cells able to reach and subsequently cache optimal values for a pattern before a new pattern is generated and hence obviates the need to keep separate Q and fuzzy Q-learning tables for each pattern. After every 14 runs, the existing pattern is replaced by the other pattern, whereas in each run users having diverse QoE requirements move to a randomly chosen destination waypoint. Each user keeps a record of its average normalized QoE perceived during last 3 visits to all waypoints. Moreover, to capture the realistic dynamically changing radio characteristics over time, users' mobility along with rapidly varying shadowing effects is generated on a random basis in each run. In addition to this, backhaul capacities are also randomly varied in order to imitate the throughput variations normally observed in the realistic transport network.

To analyze the impact of cognizant capabilities of cells and users on network and users' KPIs, we compare the efficacy of the proposed scheme with memory-less (ML) basic Q-learning-based scheme [13] and ML fuzzy Q-learning-based scheme [3] under identical circumstances in terms of users and network. We present the results for different network and users' KPIs in the form of cumulative distribution function (cdf) compiled over the total number of runs in the subsequent paragraphs.

The two KPIs, presented in left and right part of Figure 2, indicate the aggregate gap between measured and target QoE of users putting high weight to throughput and reliability respectively and are computed using (13). Our proposed memory-based scheme, capitalizing on users' predictable whereabouts with known QoE requirements, significantly addresses the throughput shortage for users preferring high throughput services and outperforms ML Q-learning and fuzzy Q-learning-based schemes by a distinct margin of 63.26% and 51.48% respectively. Also, our proposed scheme successfully bridges the gap between the target and perceived QoE for users demanding high reliability as is evident by a remarkable improvement of 87.5% and 82.13%

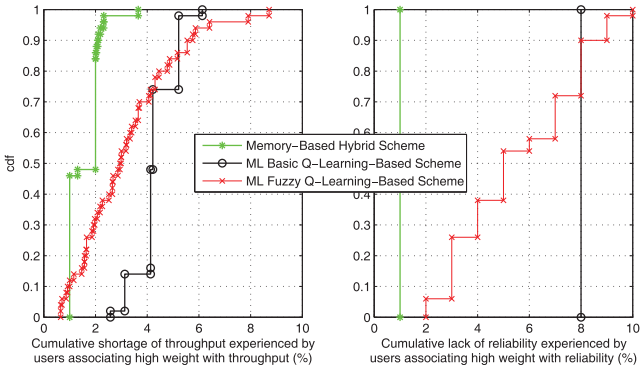


FIGURE 2. (Left) Cumulative shortage of throughput experienced by users associating high weight with throughput. Average throughput shortage attained with memory-based hybrid scheme, ML basic Q-learning-based scheme, and ML fuzzy Q-learning-based scheme is 1.58, 4.30, and 3.19 % respectively. (Right) Cumulative lack of reliability experienced by users associating high weight with reliability. Average lack of reliability with memory-based hybrid scheme, ML basic Q-learning-based scheme, and ML fuzzy Q-learning-based scheme is 1.0, 8.0, and 5.58 % respectively.

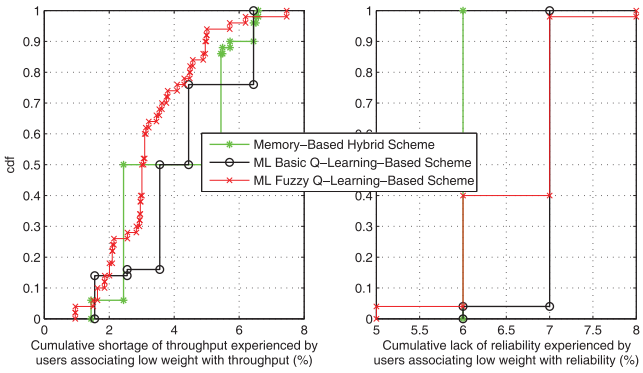


FIGURE 3. (Left) Cumulative shortage of throughput experienced by users associating low weight with throughput. Average throughput shortage attained with memory-based hybrid scheme, ML basic Q-learning-based scheme, and ML fuzzy Q-learning-based scheme is 3.99, 4.30, and 3.29 % respectively. (Right) Cumulative lack of reliability experienced by users associating low weight with reliability. Average lack of reliability with memory-based hybrid scheme, ML basic Q-learning-based scheme, and ML fuzzy Q-learning-based scheme is 6.0, 6.96, and 6.58 % respectively.

when compared to ML Q-learning and fuzzy Q-learning-based schemes respectively.

We now present the results that would reflect the cumulative shortage of measured QoE corresponding to target QoE of users putting low weight on the two performance attributes taken into consideration. The cumulative shortage of throughput is depicted in the left part of Figure 3, whereas the right part of Figure 3 highlights cumulative lack of reliability for users associating low weight with throughput and reliability respectively. It can be seen that our proposed scheme also improves QoE for users, who do not value throughput, by 7.21% over ML Q-learning-based scheme. Our proposed scheme, though, falls behind ML fuzzy Q-learning-based scheme in terms of throughput shortage by 17.55%. Similarly, improvements in QoE for low-reliability users are observed with the proposed scheme as it leads other ML schemes by 13.8% and 9.82% respectively. The proposed scheme, hence, successfully maps quality-rich backhaul resources to users valuing high-quality services regarding any performance attribute and indeed, successfully manages to improve efficient utilization of the realistic backhaul links.

We compute the network achievable throughput, which is equivalent to the aggregate throughput achieved by all the users served by the system, to analyze the performance of the proposed scheme from the perspective of a network's KPI. Figure 4 illustrates that ML basic Q-learning-based successfully maximizes the network throughput and clearly outperforms the proposed scheme by an evident margin of 4.76%. Our proposed scheme also lags behind ML fuzzy Q-learning-based scheme in terms of network throughput by 2.22%. However, since our proposed scheme achieves remarkable gains in regard to users' KPIs, the loss of system capacity may be seen as a negligible degradation.

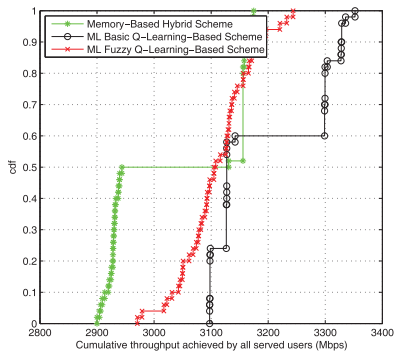


FIGURE 4. Cumulative throughput perceived by all users served by the system. Average cumulative throughput achieved with ML basic Q-learning-based scheme, ML fuzzy Q-learning-based scheme and memory-based hybrid scheme is 3041.7, 3194.6, and 3110.2 Mbps respectively.

Since 5G is expected to rely heavily on small cells deployed in overwhelming numbers, the time acquired by small cells to learn optimal bias values, known as optimization time, plays a crucial role when it comes to gauging viability of a UA scheme. Unlike ML Q-learning-based scheme that evaluates all unexplored actions in the incumbent state before applying optimal policy, our proposed scheme exploits predictable mobility patterns in addition to the simplistic exploration-exploitation strategy implemented in our previous work [3] to deliver optimal performance in less than one-third (precisely 30.65%) of the optimization time of ML Q-learning-based scheme; please refer to Figure 5. Moreover, our proposed scheme, due to memory awareness, is also computationally efficient than ML fuzzy Q-learning-based scheme as the optimization time of the former is 14.54% less than the time taken by the latter.

Look up table is an essential component of both Q-learning and fuzzy Q-learning required in order to learn optimal strategy through mapping of best state-action pairs for every state. Q-learning keeps state-action pairs, whereas fuzzy Q-learning stores quality values to depict appropriateness of each conclusion w.r.t. every fuzzy rule. Our proposed

TABLE 5. Tabulated summary of the simulation results.

KPI	Memory-Based Hybrid Scheme	ML Q-Learning-Based Scheme	ML Fuzzy Q-Learning-Based Scheme	Improvement/Lag wrt ML Q-learning-based scheme	Improvement/Lag wrt ML Fuzzy Q-learning-based scheme	Best Scheme
$\overline{y_{high}}$ (%)	**1.58**	4.30	3.19	63.26% Improvement	51.48% Improvement	Memory-based hybrid scheme
$\overline{z_{high}}$ (%)	**1.0**	8.0	5.58	87.5% Improvement	82.13% Improvement	Memory-based hybrid scheme
$\overline{y_{low}}$ (%)	3.99	4.30	**3.29**	7.21% Improvement	17.55% Lag	ML fuzzy Q-learning-based scheme
$\overline{z_{low}}$ (%)	**6.0**	6.96	6.58	13.8% Improvement	9.82% Improvement	Memory-based hybrid scheme
$\overline{T}$ (Mbps)	3041.7	**3194.6**	3110.2	4.76% Lag	2.22% Lag	ML Q-learning-based scheme
$\overline{t}$ (min)	**2.94**	9.59	3.44	69.35% More efficient	14.54% More efficient	Memory-based hybrid scheme
X_{sc} (bytes)	392	**168**	224	133% Extra	75% Extra	ML Q-learning-based scheme
X_u (bytes)	1800	**Not Required**	**Not Required**	Requires memory	Requires memory	ML Q and fuzzy Q-learning-based schemes

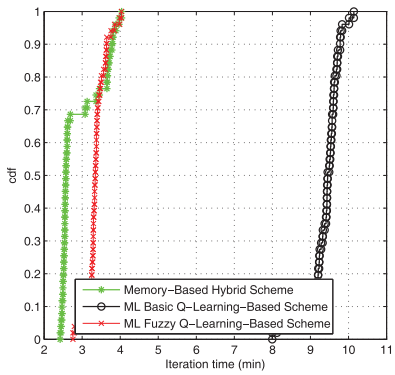


FIGURE 5. Iteration Time. Average iteration time of memory-based hybrid scheme, ML basic Q-learning-based scheme, and ML fuzzy Q-learning-based scheme is 2.94, 9.59 and 3.44 minutes respectively.

scheme, besides managing a look up table each for Q-learning and fuzzy Q-learning, introduces the concept of memory-awareness both from a small cell's as well as a user's iperspective. Since memory is a critical resource, it becomes necessary to compare the three UA schemes vis-ã-vis their memory requirements. Considering all values in look up tables of type double that occupies 8 bytes in memory and ignoring any parity bits, memory space required for Q-learning and fuzzy Q-learning look up tables is 168 bytes (3 states, 7 actions) and 224 bytes (4 fuzzy rules, 7 conclusions) respectively. According to the simulation parameters given in Table 3, a user in our proposed scheme requires 1800 bytes (1.757 MB) of memory to maintain location-based QoE history table. From small cells' perspective, we assume that the information about users' mobility patterns is kept at the proposed paging system and not at the small cells, memory-based hybrid scheme utilizes 392 bytes that is 133% and 75% additional to the memory requirements of ML Q-learning-based scheme (168 bytes) and ML fuzzy Q-learning-based (224 bytes) scheme respectively.

B. ANALYSIS & INSIGHTS

We summarize the above-presented results in Table 5. We represent average aggregate gap between users' measured and target QoE with regards to throughput and reliability by y and z respectively, whereas the subscript *high* or *low* signify users having high or low weight with these performance attributes respectively. The proposed scheme remarkably outperforms ML Q-learning-based scheme in all user-centric KPIs considered, $\overline{y_{high}}$ (users with high throughput weight, 63.26%), $\overline{z_{high}}$ (users with high reliability weight, 87.5%), $\overline{y_{low}}$ (users with low throughput weight, 7.21%), and $\overline{z_{low}}$ (users with low reliability weight, 13.8%). Furthermore, memory-based scheme also exhibits distinct improvements as compared to ML fuzzy Q-learning-based scheme when KPIs for users associating high weight with attributes are taken into account, $\overline{y_{high}}$ (51.48%) and $\overline{z_{high}}$ (82.13%). The proposed scheme, exploiting users' geolocations as well as QoE profiles in addition to capitalizing on flexible solution space of fuzzy Q-learning, intelligently assigns quality-rich backhaul resources to users with high quality requirements and hence yields remarkable gains with regards to users' KPIs. The proposed scheme, however, falls short of the ML Q-learning and fuzzy Q-learning-based schemes in terms of achievable network capacity, T (lag of 4.76% and 2.22% in comparison to ML Q-learning-based scheme and ML fuzzy Q-learning-based scheme respectively). From an operator's perspective, the proposed memory-based UA is the best scheme for delivering user-centric KPIs since it improves users' satisfaction without incurring any additional infrastructure cost. To put it another way, the proposed scheme is appropriate for the situation where operators aim for building extra capacity while retaining their users.

Our proposed memory-based hybrid scheme is computationally more efficient than the ML schemes. Exploiting users' foreseeable diurnal patterns in addition to the well-devised exploration-exploitation strategy, our proposed scheme takes less than one-third (30.65%) of the time taken by ML Q-learning-based scheme to deliver optimal results. Furthermore, cognizant capabilities of cells to optimize bias values corresponding to users' different mobility patterns enable our proposed scheme to reach optimal policy comparatively quicker (14.54%) than ML fuzzy Q-learning-based scheme. Hence, our proposed scheme would start delivering

user-centric services comparatively quicker than ML schemes thus leading to significant reduction in signalling overhead caused by frequent handovers.

In the end, we compare the three schemes with regards to their memory requirements. Our proposed scheme requires 133% and 75% extra memory at small cell, denoted by X_{sc}, than that of ML Q-learning-based scheme and ML fuzzy Q-learning-based scheme respectively. However, employing appropriate function approximators [22] or Artificial Neural Networks (ANNs) [24], [25] to interpolate Q values may save useful memory space. From a user's perspective, our proposed scheme requires 1800 bytes of memory at user, denoted by X_u, to store QoE history pertaining to the scenario presented in this paper.

V. CONCLUSION

In this paper, we investigate memory-based content-aware hybrid scheme that exploits users' predictable mobility patterns to optimize bias values corresponding to specific mobility patterns as well as QoE profiles of the users. The proposed scheme achieves the best user-centric KPIs (improvement of more than 50%) and converges faster than the state-of-the-art. Employing well-devised exploration-exploitation policy besides capitalizing on users' repeated mobility patterns, our proposed scheme renders optimal results in 69.35% and 14.54% less time taken by ML Q-learning-based scheme and ML fuzzy Q-learning-based scheme respectively. Although our proposed scheme utilizes more memory at small cells than that of both ML schemes, employing ANNs or representing Q values through appropriate function approximators may curtail memory usage at small cells to the minimum. Furthermore, unlike ML schemes that do not require any memory from users' perspective, our proposed scheme utilizes 1800 bytes of memory at user for the purpose of keeping location-based perceived QoE history. Our future work will focus on investigating the use of Deep Reinforcement Learning (Deep RL) for dynamic backhaul-aware cell selection scheme, through which we aim to enhance the computational efficiency and overcome exhaustive memory requirements.

REFERENCES

[1] P. Cerwall *et al.* (Nov. 2016). Ericsson mobility report. Ericsson. [Online]. Available: http:www.ericsson.com/res/docs/2016/ericssonmobility-report-november-2016

[2] R. Li *et al.*, "Intelligent 5G: When cellular networks meet artificial intelligence," *IEEE Wireless Commun.*, vol. 24, no. 5, pp. 175–183, Oct. 2017.

[3] F. Pervez, M. Jaber, J. Qadir, S. Younis, and M. A. Imran, "Fuzzy Q-learning-based user-centric backhaul-aware user cell association scheme," in *Proc. 13th Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Jun. 2017, pp. 1840–1845.

[4] M. Khalil, J. Qadir, O. Onireti, M. A. Imran, and S. Younis, "Feasibility, architecture and cost considerations of using TVWS for rural internet access in 5G," in *Proc. 20th Conf. Innov. Clouds, Internet Netw. (ICIN)*, Mar. 2017, pp. 23–30.

[5] *Performance of eICIC With Control Channel Coverage Limitation*, Standard R1-103264 3GPP, N. Docomo, Montreal, QC, Canada, 2010.

[6] H. Elshaer, F. Boccardi, M. Dohler, and R. Irmer, "Load & backhaul aware decoupled downlink/uplink access in 5G systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2015, pp. 5380–5385.

[7] H. Beyranvand, W. Lim, M. Maier, C. Verikoukis, and J. A. Salehi, "Backhaul-aware user association in FiWi enhanced LTE-A heterogeneous networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 2992–3003, Jun. 2015.

[8] N. Wang, E. Hossain, and V. K. Bhargava, "Joint downlink cell association and bandwidth allocation for wireless backhauling in two-tier HetNets with large-scale antenna arrays," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3251–3268, May 2016.

[9] Z. Cui and R. Adve, "Joint user association and resource allocation in small cell networks with backhaul constraints," in *Proc. 48th Annu. Conf. Inf. Sci. Syst. (CISS)*, Mar. 2014, pp. 1–6.

[10] A. De Domenico, V. Savin, and D. Ktenas, "A backhaul-aware cell selection algorithm for heterogeneous cellular networks," in *Proc. IEEE 24th Annu. Int. Symp. Pers. Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2013, pp. 1688–1693.

[11] F. Pantisano, M. Bennis, W. Saad, and M. Debbah, "Cache-aware user association in backhaul-constrained small cell networks," in *Proc. 12th Int. Symp. Modeling Optim. Mobile, Ad Hoc, Wireless Netw. (WiOpt)*, May 2014, pp. 37–42.

[12] B. Bangerter, S. Talwar, R. Arefi, and K. Stewart, "Networks and devices for the 5G era," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 90–96, Feb. 2014.

[13] M. Jaber, M. A. Imran, R. Tafazolli, and A. Tukmanov, "A distributed SON-based user-centric backhaul provisioning scheme," *IEEE Access*, vol. 4, pp. 2314–2330, May 2016.

[14] X. Lu, E. Wetter, N. Bharti, A. J. Tatem, and L. Bengtsson, "Approaching the limit of predictability in human mobility," *Sci. Rep.*, vol. 3, Oct. 2013, Art. no. 2923.

[15] J. Wu, Y. Zhang, M. Zukerman, and E. K. N. Yung, "Energy-efficient base-stations sleep-mode techniques in green cellular networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 2, pp. 803–826, May 2015.

[16] D. Karamshuk, C. Boldrini, M. Conti, and A. Passarella, "Human mobility models for opportunistic networks," *IEEE Commun. Mag.*, vol. 49, no. 12, pp. 157–165, Dec. 2011.

[17] A. S. Cacciapuoti, "Mobility-aware user association for 5G mmWave networks," *IEEE Access*, vol. 5, pp. 21497–21507, Sep. 2017.

[18] N. Zarin and A. Agarwal, "A hybrid network selection scheme for heterogeneous wireless access network," in *Proc. IEEE 28th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Oct. 2017, pp. 1–6.

[19] M. Verkama, "Paging of mobile station in a cellular radio system," U.S. Patent 6 351 641 B2, Feb. 26, 2002.

[20] A. Bonarini, A. Lazaric, F. Montrone, and M. Restelli, "Reinforcement distribution in fuzzy Q-learning," *Fuzzy Sets Syst.*, vol. 160, no. 10, pp. 1420–1443, 2009.

[21] H. Arabnejad, C. Pahl, P. Jamshidi, and G. Estrada, "A comparison of reinforcement learning techniques for fuzzy cloud auto-scaling," in *Proc. 17th IEEE/ACM Int. Symp. Cluster, Cloud Grid Comput. (CCGRID)*, 2017, pp. 64–73.

[22] G. Głowaty, "Enhancements of fuzzy Q-learning algorithm," *Comput. Sci.*, vol. 7, p. 77, Jun. 2013.

[23] M. Coggan, "Exploration and exploitation in reinforcement learning," McGill Univ., Montreal, QC, Canada, Tech. Rep., 2004. [Online]. Available: http://neuro.bstu.by/my/Tmp/Papers-RL/FinalReport.pdf

[24] L.-J. Lin, "Self-improvement based on reinforcement learning, planning and teaching," in *Proc. 8th Int. Workshop Mach. Learn.*, 2014, pp. 323–327.

[25] G. A. Rummery and M. Niranjan, "On-line Q-learning using connectionist systems," Dept. Engineering, Univ. Cambridge, Cambridge, U.K., Tech. Rep. 166, 1994.

FARRUKH PERVEZ received the bachelor's degree in avionics engineering from the National University of Sciences and Technology, Islamabad, Pakistan, in 2009, where he is currently pursuing the M.S. degree in electrical engineering. His research interests include wireless technologies for emergency response, development, and 5G backhaul challenges.

MONA JABER (S'14) received the B.E. degree in computer and communications engineering and the M.E. degree in electrical and computer engineering from the American University of Beirut, Beirut, Lebanon, in 1996 and 2014, respectively. She is currently pursuing the Ph.D. degree with the 5G Innovation Centre, University of Surrey, researching on 5G backhaul solutions. She was a Telecommunication Consultant in various international firms with a focus on the radio design of cellular networks, including GSM, GPRS, UMTS, and HSPA. Her research interests include wireless communications, cellular technologies, backhaul and fronthaul solutions, and self-optimization techniques.

JUNAID QADIR (M'13–SM'14) received the bachelor's degree in electrical engineering from the University of Engineering and Technology, Lahore, Pakistan, in 2000, and the Ph.D. degree from the University of New South Wales, Australia, in 2008. He has been an Associate Professor with Information Technology University (ITU)-Punjab, Lahore, since 2015. He is the Director of the IHSAN (ICTD; Human Development; Systems; Big Data Analytics; Networks Lab) Research Laboratory at ITU. Previously, he has served as an Assistant Professor at the School of Electrical Engineering and Computer Sciences, National University of Sciences and Technology, from 2008 to 2015. His primary research interests are in the areas of computer systems and networking and using ICT for development. He has served on the program committee of a number of international conferences and reviews regularly for various high-quality journals. He is a member of ACM. He is an Award-Winning Teacher who has been awarded the Highest National Teaching Award in Pakistan-the Higher Education Commission's Best University Teacher Award-for the year 2012–2013. He has considerable teaching experience and a wide portfolio of taught courses in the disciplines of systems and networking, signal processing, and wireless communications and networking. He is an Associate Editor of the IEEE Access, *Springer Nature Central's Big Data Analytics*, *Springer Human-Centric Computing and Information Sciences*, and the *IEEE Communications Magazine*.

SHAHZAD YOUNIS received the bachelor's degree from the National University of Sciences and Technology (NUST), Islamabad, Pakistan, in 2002, the master's degree from the University of Engineering and Technology, Taxila, Pakistan, in 2005, and the Ph.D. degree from the University Technology PETRONAS, Perak, Malaysia, in 2009. Before joining NUST, he was the Assistant Manager at a research and development organization named AERO, where he was involved in different signal processing and embedded system design applications. He is currently an Assistant Professor with the Department of Electrical Engineering, School of Electrical Engineering and Computer Science, NUST. He has authored or co-authored over 25 papers in domestic and international journals and conferences. His research interests include statistical signal processing, adaptive filters, convex optimization biomedical signal processing, wireless communication modelling, and digital signal processing.

MUHAMMAD ALI IMRAN (M'03–SM'12) received the M.Sc. (Hons.) and Ph.D. degrees from Imperial College London, U.K., in 2002 and 2007, respectively. He is the Vice Dean of the Glasgow College, UESTC, and a Professor of communication systems with the School of Engineering, University of Glasgow. He is also an Affiliate Professor at the University of Oklahoma, Norman, OK, USA, and a Visiting Professor with the 5G Innovation Centre, University of Surrey, U.K. He has over 18 years of combined academic and industry experience, working primarily in the research areas of cellular communication systems. He has been awarded 15 patents, has authored/co-authored over 300 journal and conference publications, and has been principal/co-principal investigator on over č6 million in sponsored research grants and contracts. He has supervised 30+ successful Ph.D. graduates.

He is a Senior Fellow of the Higher Education Academy (SFHEA), U.K. He has an award of excellence in recognition of his academic achievements, conferred by the President of Pakistan. He was also awarded the IEEE Comsoc's Fred Ellersick Award 2014, the FEPS Learning and Teaching Award 2014, and the Sentinel of Science Award 2016. He was twice nominated for the Tony Jean's Inspirational Teaching Award. He is a shortlisted finalist for the Wharton-QS Stars Awards 2014, the QS Stars Reimagine Education Award 2016 for innovative teaching, and the VC's Learning and Teaching Award from the University of Surrey.

He is the Editor/Co-Editor of two books *Access, Fronthaul and Backhaul Networks for 5G and Beyond* (IET) and *Energy Management in Wireless Cellular and Ad-hoc Networks* (Springer).