Fuzzy Q-Learning-Based User-Centric Backhaul-Aware User Cell Association Scheme

Farrukh Pervez[1,2], Mona Jaber[3], Junaid Qadir[2], Shahzad Younis[1], Muhammad Ali Imran[4]
[1]National University of Sciences and Technology Pakistan, [2]Information Technology University Lahore Pakistan
[3]University of Surrey UK, [4]University of Glasgow UK
fpervez.msee15seecs@seecs.edu.pk, m.jaber@surrey.ac.uk, junaid.qadir@itu.edu.pk, muhammad.shahzad@seecs.edu.pk, muhammad.imran@glasgow.ac.uk

***Abstract*—Heterogeneous networks are a key solution to serving the exponential surge in data volume and higher quality expectations. Nonetheless, such networks require the ubiquitous presence of fiber-to-the-cell to address the performance demands of 5G and fast-spreading small cells. To this end, innovative ways of optimizing the usage of realistic backhaul links are being investigated. In this work, we propose a fuzzy Q-learning-based user-centric backhaul-aware user cell association scheme.**

The proposed scheme aims at optimizing the user-cell association process in a context-aware and backhaul-aware manner. Complementing the scheme with fuzzy-logic requires 33.3% additional storage memory. On the other hand, it increases the computational efficiency by 60% and improves the users' performance by 12%.

***Index Terms*—Heterogeneous Networks (HetNets), user cell association, user-centric, backhaul, fuzzy Q-learning**

I. Introduction

Heterogeneous Networks (HetNets) is one of the promising solutions for the fifth generation (5G) mobile networks to handle the deluge of smart devices driven data traffic as well as users' acquisitive approach towards high quality of service (high throughput, low latency etc.). HetNets are multi-tiered networks in which low-power small cells are overlaid with high-power macro cells to supplement network capacity and coverage. Small cells, though offer substantial gains in the form of better frequency reuse ratio and improved energy efficiency etc., associate the following challenges to the backhaul links. *Firstly*, ubiquitous extension of incumbent transport links is required to connect these inundated small cells to the core network. *Secondly*, 5G enabled cutting-edge features and services impose multi-constrained stringent performance requirements on the backhaul links. For instance, enhanced Mobile Broadband (eMBB), coordinated multipoint processing (CoMP) and cloud radio access network (C-RAN) act as capacity boosters thus necessitating high-capacity backhaul links. In addition, 5G applications envisioned in the future tactile Internet, require interaction latency to be less than 1 millisecond [1]. Similarly, enabling green networks concerns energy-efficient backhaul links.

The analysis of available wired and wireless backhaul options reveal that point-to-point optical fiber is the only mature backhaul candidate that complies with requisite performance attributes [2]. However, providing fiber-to-the-cell in small cells seems an impractical and cost-ineffective approach for the high density and not easy to reach locations of small cells. In such cases, realistic constrained backhaul links would be used instead of fiber, hence, may indeed be the new 5G bottleneck [3]. Accordingly, ways of optimizing realistic backhaul links are being explored with a view to 'make the extremes' possible in 5G.

The coexistence of disparate transmit power cells makes the state-of-the-art user cell association (UA) scheme—which associates a user to the cell with the strongest downlink signal—incompatible to HetNets, since a major chunk of users would be attracted towards macro cells thus leading to underutilization of the small cells [4]. Besides emerging backhaul bottleneck, there exists a considerable diversity in 5G in terms of service requirements, devices, radio access technologies, backhauls and above all cells and hence, requires an optimal UA scheme that could pragmatically capture user needs while efficiently utilizing coexisting network layers.

Accordingly, we propose a fuzzy Q-learning-based user-centric backhaul-aware UA scheme. The cells advertise their capabilities and constraints through an optimized set of bias factors. In this work, we propose to equip each cell with the capability of fuzzy Q-learning dynamically adjust its bias factors according to network and users indicators in a fully distributed and automated way. As 5G users would have different requirements and priorities, each would make a different cell selection that reflects the user's needs with respect to candidate cells' capabilities.

The main contribution of this paper is the novel implementation methodology for user-centric backhaul-aware UA scheme presented in our previous work [5]. Unlike our basic implementation, which was computationally limiting, our proposed implementation is computationally efficient and hence has significant practical relevance since it allows numerous small cells, one of the key enablers of 5G, to take part in the UA optimization. Our proposed implementation also leads to substantial improvements in users' key performance indicators (KPIs), which would have a crucial role in 5G era [6].

The rest of the paper is organized as follows. Section II covers the related work focusing mainly on backhaul-aware UA schemes. The system model is presented in Section III, followed by problem formulation in Section IV. We present our proposed implementation of the user-centric backhaul-aware scheme with preliminary results, analysis, and insights in Section V. We conclude this paper finally in Section VI. In order to make this paper self-contained, we present basic working of fuzzy Q-learning in Table I.

II. Related Work

In view of the apparent unsuitability of traditional max-received signal strength-based (max-RSS-based) UA approach for HetNets, a concept of Cell Range Expansion (CRE) is devised in [7] that enables small cells to add a positive offset (CREO (Cell Range Expansion Offset) or bias factor) in cell selection process so as to attract more users towards them. Such schemes [8]–[11] succeed in shifting the traffic towards lower layer, thus increasing the network capacity. However, when such schemes are applied without the consideration of the backhaul constraint,

TABLE I
BASIC WORKING OF FUZZY Q-LEARNING

Fuzzy Q-Learning Algorithm

1. Observe the current state s and respective fuzzy state x at time t

2. For each of the activated fuzzy rules ϕ_i choose conclusions A_j using appropriate exploration and exploitation strategy

3. Compute and subsequently execute the inferred action $H(x)$

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

where w_i is the rule firing strength and A_i is the respective conclusion

4. Calculate corresponding FQ value for the current state $FQ(x, H(x))$

$$FQ(x, H(x)) = \frac{\sum_{i=1}^{m} w_i \cdot q(\phi_i, A_i)}{\sum_{i=1}^{m} w_i} \tag{2}$$

where $q(\phi_i, A_i)$ is the quality value representing appropriateness of ϕ_i with A_i

5. Observe the new state x' and receive immediate reward r

6. Compute FQ value for the new state using optimal policy

$$FQ^*(\hat{x}, H(\hat{x})) = \frac{\sum_{i=1}^{m} w_i \cdot q(\phi_i, A_i^*)}{\sum_{i=1}^{m} w_i} \tag{3}$$

where

$$A_i^* = arg \max_{A_j} \{q(\phi_i, A_j)\} \tag{4}$$

7. Update the quality values in look up table using the following equation

$$q(\phi_i, A_i) = q(\phi_i, A_i) + \eta \Delta q(\phi_i, A_i) \tag{5}$$

where

$$\Delta q(\phi_i, A_i) = \frac{w_i}{\sum_{k=1}^{m} w_k} \cdot (r(x, H(x)) + \gamma FQ^*(\hat{x}, H(\hat{x})) - FQ(x, H(x))) \tag{6}$$

η and γ represent learning rate and discount factor respectively

8. Repeat the process unless terminal state is reached or maximum allowed iterations are complete

they may lead to unsatisfactory users' performance (as a result of the backhaul bottleneck) or to inefficient utilization of the network [5], [12]. Hence, various backhaul-aware UA schemes, showing convincing results, have been proposed recently in the literature. For instance, authors in [13] propose balancing the workload, that arises during call admission, among heterogeneous backhauls. The idea of decoupled downlink/uplink UA for load balancing under backhaul capacity and cell load constraints is advocated in [14]. Similarly, load balancing with regard to backhaul latency and reliability limitations is studied in [15]. Authors in [16] propose a UA strategy that conserves the overall network energy, at the cost of the reduction in users' throughput, by minimizing both the number of active cells and backhaul switches. The concept of intelligent small cells, which exploit Q-learning for dynamic optimization of CREO values in view of the joint radio and backhaul capacity constraints, for maximizing users' quality of experience (QoE) is presented in [17]. A game theoretic-based cache-aware UA algorithm that enables small cells, cognizant of their own caching capabilities and users' mobility patterns, to cache multimedia content in order to enhance users' quality of service (QoS) under backhaul capacity limitations is investigated in [18]. A novel user-centric backhaul-aware UA approach is introduced in [5], that enables users to select a cell that potentially meets their respective QoE requirements. Cells leverage Q-learning to virtually tailor their footprints in order to meet users' needs considering both backhaul capacity and delay constraints. Simulations validate that the proposed scheme improves users' QoE significantly while losing a fractional system capacity, in comparison to cell-centric backhaul-aware schemes.

In this work we propose a novel approach to the user-centric backhaul-aware scheme introduced in [5], which aims at maximizing users' QoE through virtually tailored optimal cells ranges with regard to backhaul capacity and reliability constraints. In this work we address two shortcomings of the basic user-centric backhaul-aware scheme: the impractical increase in the process lead-time when the number of cells per cluster increases, and the limitation of the solution space to integer bias values. To this end, we employ fuzzy Q-learning with a well-devised exploitation and exploration strategy to reduce the optimization lead time. In addition, the inherent flexibility of fuzzy Q-learning is exploited to address the solution space limitations through the usage of continuous state-space parameters.

III. SYSTEM MODEL

We consider a two-tier downlink HetNet; the lower tier comprises N small cells having geographic locations defined by (x_n, y_n), whereas the upper tier consists of M macro cells with Cartesian coordinates (x_m, y_m). A set of K resource blocks is shared equally amongst small cells and macro cells in order to render services to the associated users. Each resource block is composed of s subcarriers, with each subcarrier spanning across a fixed bandwidth of B Hz. Every cell broadcasts O number of CREOs that represent different end-to-end constraints of the network such as throughput, resilience, latency etc,. A large CREO value implies high end-to-end network performance and a low CREO value indicates otherwise. Small cells have heterogeneous backhaul network containing technologies like G.fast, direct-fiber, microwave and copper, however, each cell is leveraged with only one last mile backhaul link to connect to the core network. We make this assumption in order to limit the complexity of routing algorithm that increases with increase in the number of backhaul links connecting small cells to the aggregation point. However, we have designed our proposed scheme as transparent to the number of backhaul links for evolution and future development purpose. In the end, we assume that every small cell contains uniform RAN architecture either distributed, centralized or hybrid one.

A total of U users having diverse service requirements are randomly distributed in the system. Each user assigns relative weight W_q to different performance metrics Q, known as QoE parameters, in accordance with its preferences, application requirements, UE features, and capabilities. We assume that every user has QoE parameters equal in number with that of other users' parameters as well as CREOs broadcasted by the cells. We assign a nil value to the undefined QoE parameters or missing CREOs in case total number of QoE parameters of a user does not match with that of CREOs broadcasted by the cells and vice versa. We now describe the system model in detail with the help of related equations.

The strength of the downlink signal received by a user u from a small cell n over a resource block k is given by the following equation

$$RSS_{n,u,k} = \frac{P_n}{|U_n|} \times H_{n,u,k} \tag{7}$$

$$H_{n,u,k} = \chi_n \times d_{n,u}^{-\alpha_n} \times \varepsilon_{n,u,k} \tag{8}$$

where P_n denotes the total power transmitted by the cell n and $|U_n|$ is the total number of users associate with the cell n. $H_{n,u,k}$, given by (8), provides information about the channel quality existing between transmitter (cell n) and the receiver (user u) over the given subframe. The quality of the received signal degrades

according to the log-distance path loss model, with χ_n and α_n as the propagation constant and path loss exponent characteristics of the cell n. The Euclidean distance between the cell n and the user u having Cartesian positions defined by (x_n, y_n) and (x_u, y_u), respectively is given by $d_{(n,u)}$. We assume that the log-normal shadowing, represented by $\varepsilon_{n,u,k}$, is 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 the multi-path (fast) fading gets averaged out since it varies more rapidly than the time required to adjust a CREO and hence is ignored in the channel gain.

Every user employs the criteria, given by (9), to rank all the cells including the macro cells.

$$R_{u,n} = \begin{cases} RSS_{n,u,k} + \sum_{q=1}^{|O|} w_{u,q} \times v_{n,q}, & if n \in N|M \\ RSS_{n,u,k}, & otherwise \end{cases} \tag{9}$$

where $w_{u,q}$ represents the weight assigned by the user u to the QoE q and $v_{n,q}$ is the corresponding CREO value broadcasted by the cell n. The user ranks all candidate cells and tries to associate with the best ranking, down to the worst ranking (but still valid) until it finds a server or is declared out of coverage. The corresponding SINR received by the user u being served by the cell n over $1, 2, ..., k, ...K$ resource blocks is given as follows:

$$SINR_{n,u,k} = \frac{RSS_{n,u,k}}{\sigma^2 + \sum_{i=1, i\neq n}^{N} RSS_{i,u,k} + \sum_{j=1, j\neq n}^{M} RSS_{j,u,k}} \tag{10}$$

where σ^2 defines the noise variance (power) in the received signal. The throughput achieved by the user u corresponding to the SINR mentioned in (10) can be represented by the following equation:

$$T_{n,u,k} = B \log_2(1 + SINR_{n,u,k}) \tag{11}$$

where B denotes the bandwidth of the resource block allocated to the user u by the cell n. Based on the radio throughput generated by the users associated with cell n, the required backhaul throughput can be computed as follows:

$$BH_n = \sum_{u=1}^{|U_n|} T_{u,n,k} \times G_n \tag{12}$$

where the factor G_n reflects the signalling overhead which is determined by the backhaul technology and topology and the RAN architecture of the cell n. The effective throughput attained by a user is dependent on the constrained capacity Av_BH_n offered by the backhaul of the serving cell. The user suffers a reduction in effective throughput, as per (14), if the required backhaul throughput overshoots available backhaul capacity thus leading to congestion in the backhaul links.

$$S = Av\_BH_n - BH_n \tag{13}$$

$$T'_{n,u,k} = \begin{cases} T_{n,u,k}, & if S \geq 0 \\ T_{n,u,k} - \frac{S}{|U_n|}, & otherwise \end{cases} \tag{14}$$

IV. Problem Formulation

Since centralized optimization techniques, which enhance the total system throughput subject to a given set of constraints, face a rapid increase in complexity and may become intractable with an increase in number of cells, the number of bias values or number of backhaul links available with the cell, we propose the distributed optimization in which cells independently maximize their respective throughput while keeping in view the identified constraints. Distributed optimization offers the following significant benefits over centralized optimization.

Firstly, the reduction in complexity from $O(V.O.C)$ to $O(V.O)$, where V is the range of values, a CREO can have, O is the total number of CREOs and C represents the number of intelligent cells taking part in the optimization exercise. Secondly and more importantly, the complexity of distributed optimization is independent of the number of cells deployed in the network. This point becomes critical especially when viewed in the large scale deployments of 5G with densely deployed small cells for many reasons. Distributed optimization does not require the lengthy process of network-wide collection of data, hence is much faster than centralized optimization. Moreover, distributed optimization adapts automatically to dynamic changes in the network such as added nodes, modified topology, introduced radio features etc.

We articulate the distributed optimization problem in the following equations:

$$\max_n (T_n), n \in N \tag{15}$$

$$T_n = \sum_{u=1}^{|U_n|} T'_{n,u,k} \tag{16}$$

subject to the following identified constraints:

$$Av\_BH_n - BH_n \geq 0 \quad n \in N \tag{17}$$

$$UQ'_{n,q} = \sum_{u=1}^{U_n} \frac{Q'_{u,q} - Q_{u,q}}{Q_{u,q}} \leq \phi_q \quad Q_q \in Q_u, \quad n \in (N \cup M) \tag{18}$$

In (15), T_n is the throughput achieved by the cell n and is computed using (16). To render minimum QoE requirements to the users, cells take backhaul throughput constraint, defined by (17), into account. Furthermore, cumulative perceived QoE of all users, $UQ'_{n,q}$, is to be kept below the minimum QoE threshold ϕ_q, as expressed in (18), $Q'_{u,q}$ is the QoE perceived by the user relative to its target QoE $Q_{u,q}$ for an attribute q. It is important to mention here that we measure users' satisfaction relative to various performance targets in terms of QoE, as was introduced in [5].

V. Proposed Fuzzy Q-Learning-Based User-Centric Backhaul-Aware UA Scheme

The possibility to address the problems where input and action state space are real-valued variables with associated fuzzy sets motivates the idea of solving the identified optimization problem with fuzzy Q-learning. The partial overlapping of nearby linguistic variables of a fuzzified input gives fuzzy Q-learning an improved flexibility in addition to robustness and smoothness [19]. In the context of this paper, small cells (RL agents) tend to learn optimal CREO values (actions) for each of the fuzzy rules through iterative interaction with their environment, which includes dynamic radio conditions, temporal and spatial fluctuations in users' traffic, and backhaul capacity variations. We fuzzify the two inputs—required backhaul throughput (BH_n, req_bh_tput) and users' perceived QoE ($UQ'_{n,q}, users\_prcvd\_qoe$)—that define the constraints of the problem. Each fuzzified input is assigned

two linguistic variables. We present the salient details of our proposed scheme in Table III.

We implement a simple exploration and exploitation policy in our proposed scheme without compromising on its optimal performance. 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. The number of membership functions is kept as two for each fuzzified input for the following two reasons. Firstly, to reduce memory requirements to the minimum possible. Secondly, to inhibit an exponential rise in computation time. Furthermore, the chosen core or impact width of the membership functions is reached on a trial and error approach; too much overlapping increases computation time drastically without gaining an extra boost in performance, while too less overlapping compromises flexibility and smoothness.

TABLE II
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
Backhaul attributes	Throughput, reliability
Backhaul options	G.fast, microwave, optical fiber
Simulation platform	MATLAB R2013a

A. Simulation Results

To demonstrate a proof of concept, we implement the proposed fuzzy Q-learning-based user-centric backhaul-aware UA scheme on a simulated downlink HetNet, which comprises 9 small cells overlaid within one macro cell, using MATLAB R2013a. Two users' QoE are considered with respect to two performance attributes namely throughput and reliability. Each resource block comprises 12 subcarriers, with every subcarrier spanning across 180 KHz. Small cells exploit fuzzy Q-leaning to iteratively learn optimal bias values, for these attributes, that reflect end-to-end throughput and reliability of the network, respectively. Every small cell is leveraged with only one of the three available heterogeneous backhaul options such as G.fast, microwave, and optical fiber, while the macro cell is assumed to have ideal backhaul that aggregates the traffic coming from all small cells towards the core network. By ideal backhaul, we mean that macro cell backhaul would not impede the network performance by acting as a bottleneck. As the aim is to capitalize on the deployment of small cells, macro cells are always assigned nil bias factors and act as a fall-back plan in case small cells fail to serve the users adequately. Table II summarizes the list of simulation parameters.

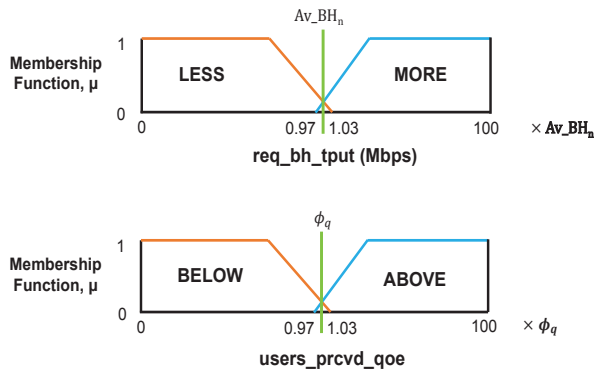
The simulations are conducted using Monte Carlo technique consisting of 50 simulation runs. In each run, users having random QoE preferences are uniformly distributed, mostly around small cells to create traffic hot spots. Moreover, to capture the realistic dynamically changing radio characteristics 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. We compare the efficacy of the proposed scheme with that of Q-learning-based user-centric backhaul provisioning scheme [5] under identical circumstances in terms of users and network, and present the results for different key performance indicators (KPIs) in the form of cumulative distribution function (cdf) compiled over the total number of runs.

TABLE III
SALIENT DETAILS OF THE PROPOSED FUZZY Q-LEARNING-BASED METHODOLOGY

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

RL Agents – All small cells n $\in$ N

Fuzzified Inputs – req_bh_tput and $users\_prcvd\_qoe$



Fuzzy Rule Base
IF req_bh_tput is **LESS AND** $users\_prcvd\_qoe$ is **BELOW THEN** conc is $\cdots$
IF req_bh_tput is **MORE AND** $users\_prcvd\_qoe$ is **BELOW THEN** conc is $\cdots$
IF req_bh_tput is **LESS AND** $users\_prcvd\_qoe$ is **ABOVE THEN** conc is $\cdots$
IF req_bh_tput is **MORE AND** $users\_prcvd\_qoe$ 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

$$R = \begin{cases} 1000 \cdot \frac{Av\_BH_n - BH_n}{Av\_BH_n}, & if Av\_BH_n > BH_n \\ & and\ UQ'_{n,q} < \phi_q \\ -1000 \cdot (BH_n - Av\_BH_n), & if BH_n > Av\_BH_n \\ -100 \cdot \sum_{u=1}^{U_n} \frac{Q'_{u,q} - Q_{u,q}}{Q_{u,q}}, & if Av\_BH_n > BH_n \\ & and\ UQ'_{n,q} > \phi_q \end{cases} \quad (19)$$

Cumulative Rule-Conclusion Reward – The cumulative reward for each rule–conclusion pair, $q(\phi_i, A_i)$, is updated using (5)

The first two KPIs indicate the cumulative difference between the target QoE and the perceived QoE of users associating high weight with throughput and reliability as shown in the left and right part of Figure 1, respectively. We use (18) to calculate these two KPIs. Clearly, our proposed scheme significantly reduces the gap between the target and the achieved throughput for users, who value throughput performance and outperforms Q-learning-based scheme by a margin of 13.87%. Also, our proposed scheme successfully achieves higher QoE for users demanding high reliability as evident from the improvement of 12.69% over its counterpart scheme.

We now present the results that would depict cumulative difference in target and perceived QoE of users associating low weight with the two attributes considered. The left part of

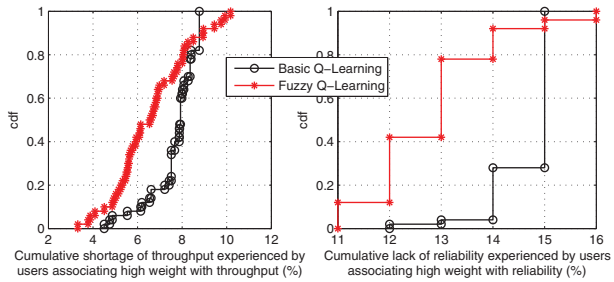


Fig. 1. **(Left)** Cumulative shortage of throughput experienced by users associating high weight with throughput. Average throughput shortage with Q-learning and fuzzy Q-learning is 7.66% and 6.6% respectively. **(Right)** Cumulative lack of reliability experienced by users associating high weight with reliability. The average lack of reliability with Q-learning and fuzzy Q-learning is 14.66% and 12.8% respectively

Figure 2 illustrates cumulative shortage of throughput, whereas the right part of Figure 2 depicts the cumulative lack of reliability experienced by users associating low weight with throughput and reliability, respectively. It can be seen that the proposed scheme improves QoE for such users as well by 7.18% and 4.63% respectively. The proposed scheme, thus, succeeds in matching quality-rich resources to users that associate a high value to the specific quality. As such, the utilization efficiency of the realistic backhaul resources is actively improved.

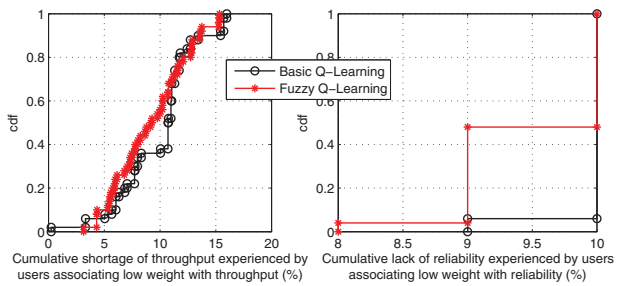


Fig. 2. **(Left)** Cumulative shortage of throughput experienced by users associating low weight with throughput. Average throughput shortage with Q-learning and fuzzy Q-learning is 9.90% and 9.19% respectively. **(Right)** Cumulative lack of reliability experienced by users associating low weight with reliability. The average lack of reliability with Q-learning and fuzzy Q-learning is 9.94% and 9.48% respectively

From a network's KPI perspective, we look at the cumulative achievable throughput, which is computed as the sum of throughput perceived by all users served by the system. Figure 3 shows that the proposed scheme falls short of the system capacity achieved by Q-learning-based scheme by 1.396%. However, this is noticeably a minor degradation in view of the significant gain achieved with respect to the users' KPIs.

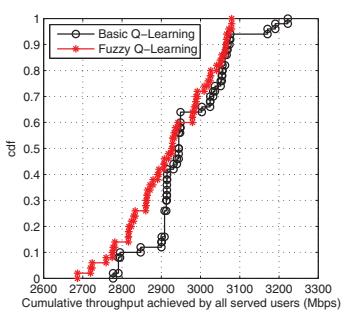


Fig. 3. Cumulative throughput perceived by all users served by the system. Average cumulative throughput achieved with Q-learning and fuzzy Q-learning is 2963.32 and 2921.96 Mbps respectively.

Since small cells would be deployed in dense numbers in 5G, the time taken by small cells to reach optimal strategy, called as lead time, also becomes a crucial factor in gauging the effectiveness of a UA scheme. Unlike Q-learning-based scheme that initially explores all the available actions in the prevalent state before adopting the optimal policy, our proposed scheme employs the simple and effective strategy of exploration and exploitation that enables it to perform optimally in nearly one-third (precisely 36.78%) of the time taken by Q-learning-based scheme, as illustrated by the left part of Figure 4. In addition, though both schemes depict exponential behaviour in iteration time with increase in the number of small cells, the proposed scheme enables cells to learn at a much faster rate thus keeping iteration time within bounds, as can be seen in the right part of Figure 4.

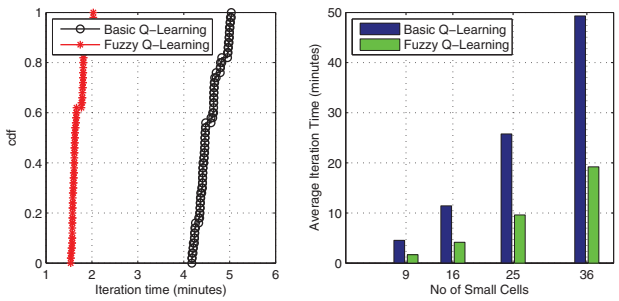


Fig. 4. **(Left)** Iteration time. Average iteration time of Q-learning and fuzzy Q-learning is 4.54 and 1.67 minutes respectively. **(Right)** Effect on computation efficiency with increase in number of small cells. Average iteration time for 36 small cells with Q-learning and fuzzy Q-learning is 49.31 and 19.21 minutes respectively.

Both Q-learning and fuzzy Q-learning techniques stock values in look-up tables in order to map state and action pairs. If the size of state or action set grows considerably, it may lead to exhaustion of memory resources. Hence, it becomes imperative to compare the two schemes in terms of their memory requirements. Unlike Q-learning-based scheme, which stores state-action pairs, our proposed scheme stores quality matrix to map conclusions with fuzzy rules. Ignoring any object overhead bytes and considering all variables as double requiring 8 bits for storage [20], the proposed scheme requires 224 bytes (4 rules each having 7 conclusions), which is 33% more than that of Q-learning-based scheme that needs 168 bytes (3 states each with 7 actions).

B. Analysis & Insights

We present the summary of these results in Table IV. Average cumulative difference between users' perceived and target QoE in terms of throughput and reliability is represented by $\overline{x}$ and $\overline{y}$ respectively, whereas subscript $high$ or low denotes users associating high or low weight with these attributes, respectively. The proposed scheme outperforms the Q-learning-based scheme by a distinct margin in all four user-centric performance metrics considered, $\overline{x_{high}}$ (high throughput users, 13.87%), $\overline{y_{high}}$ (high reliability users, 12.69%), $\overline{x_{low}}$ (low throughput users, 7.18%), and $\overline{y_{low}}$ (low reliability users, 4.63%). The flexible solution space of fuzzy Q-learning enables our proposed scheme to intelligently associate users' having high-quality requirements with quality-rich backhaul resources, and hence leads to noticeable improvements in users' experience. Our proposed scheme lags behind the Q-learning-based scheme in achieving system capacity, denoted by $\overline{T}$, by a negligible difference of 1.396%.

TABLE IV
PERFORMANCE COMPARISON OF THE TWO SCHEMES

KPI	Q-Learning	Fuzzy Q-Learning	Remarks
$\overline{x_{high}}$ (%)	7.66	6.60	13.87% Improvement
$\overline{y_{high}}$ (%)	14.66	12.80	12.69% Improvement
$\overline{x_{low}}$ (%)	9.90	9.19	7.18% Improvement
$\overline{y_{low}}$ (%)	9.94	9.48	4.63% Improvement
$\overline{T}$ (Mbps)	2963.32	2921.96	1.396% Lag
$\bar{t}$ (min)	4.54	1.67	63.22% More efficient
$\overline{t_{36}}$ for 36 small cells (min)	49.31	19.21	61.05% More efficient
X (bytes)	168	224	33.3% Extra

The proposed scheme is computationally more efficient than the Q-learning-based scheme with average iteration time, represented by $\bar{t}$, taken by the former is 63.22% less than that of the latter. Furthermore, our proposed scheme proves to be 61.05% more efficient than the Q-learning-based scheme for higher small cells density; we denote the average iteration time for 36 small cells scenario by $\overline{t_{36}}$. For dense small cells, our proposed scheme having ephemeral optimization time would start delivering high quality in comparatively quick time without incurring significant signalling overhead arising due to frequent unnecessary and undesirable handovers. We expect this optimization time to reduce further considerably with more advanced simulation hardware as compared to the present setup that comprises an ordinary laptop with core 2 duo @2.00 GHz processor and 2 GB RAM.

In the end, we compare the two schemes with regard to memory requirements, denoted by X. Our proposed scheme, though requires 33.3% more memory than Q-learning-based scheme, exploiting interpolation capacity of Artificial Neural Networks (ANNs) [21], [22] or utilizing appropriate function approximators [23] may obviate the need of look up tables and save useful memory space.

VI. CONCLUSION

In this work, we propose an alternative approach to the user-centric backhaul-aware scheme which employs fuzzy Q-learning, and successfully addresses the two main shortcomings: the optimization lead time and the constrained solutions space. Simulation results prove that the proposed scheme leads to notable improvements in user-centric KPIs owing to the inherent continuous state-space of fuzzy logic. Besides, this simple yet effective strategy of exploration and exploitation enables our proposed scheme to reach optimal performance in approximately one-third of the time taken by Q-learning-based scheme. Although the current implementation requires more memory than the Q-learning-based scheme, it may be possible to overcome this issue by integrating ANNs or using appropriate function approximators with the proposed scheme thus saving useful memory space. Our future work will focus on making cells cognizant of the users mobility patterns, as well as their respective QoE requirements, thus envisaged to significantly reduce cells learning time and improve the energy efficiency of the network.

REFERENCES

[1] G. P. Fettweis, "The tactile internet: applications and challenges," *IEEE Vehicular Technology Magazine*, vol. 9, no. 1, pp. 64–70, 2014.

[2] M. Jaber, M. A. Imran, R. Tafazolli, and A. Tukmanov, "5G backhaul challenges and emerging research directions: A survey," *IEEE Access*, vol. 4, pp. 1743–1766, 2016.

[3] 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 *Proceedings of the 20th International conference on Innovations in Clouds, Internet and Networks (ICIN)*. IEEE, 2017.

[4] N. Docomo, "Performance of eICIC with control channel coverage limitation," *R1-103264, 3GPP Std., Montreal, Canada*, 2010.

[5] 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, 2016.

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

[7] R1-103181, "On range extension in open-access heterogeneous networks," *document R1–103181, 3GPP TSG RAN WG1 Meeting,Motorola, Montreal, QC, Canada*, vol. 61, May 2010.

[8] T. Kudo and T. Ohtsuki, "Cell range expansion using distributed Q-learning in heterogeneous networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2013, no. 1, p. 1, 2013.

[9] M. Simsek, M. Bennis, and A. Czylwik, "Dynamic inter-cell interference coordination in HetNets: A reinforcement learning approach," in *Global Communications Conference (GLOBECOM), 2012 IEEE*. IEEE, 2012, pp. 5446–5450.

[10] I. Guvenc, "Capacity and fairness analysis of heterogeneous networks with range expansion and interference coordination," *IEEE Communications Letters*, vol. 15, no. 10, pp. 1084–1087, 2011.

[11] Q. Ye, B. Rong, Y. Chen, M. Al-Shalash, C. Caramanis, and J. G. Andrews, "User association for load balancing in heterogeneous cellular networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 6, pp. 2706–2716, 2013.

[12] A. De Domenico, V. Savin, and D. Ktenas, "A backhaul-aware cell selection algorithm for heterogeneous cellular networks," in *Personal Indoor and Mobile Radio Communications (PIMRC), 2013 IEEE 24th International Symposium on*. IEEE, 2013, pp. 1688–1693.

[13] C. Ran, S. Wang, and C. Wang, "Balancing backhaul load in heterogeneous cloud radio access networks," *IEEE Wireless Communications*, vol. 22, no. 3, pp. 42–48, 2015.

[14] H. Elshaer, F. Boccardi, M. Dohler, and R. Irmer, "Load & backhaul aware decoupled downlink/uplink access in 5G systems," in *2015 IEEE International Conference on Communications (ICC)*. IEEE, 2015, pp. 5380–5385.

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

[16] C. Bottai, C. Cicconetti, A. Morelli, M. Rosellini, and C. Vitale, "Energy-efficient user association in extremely dense small cell networks," in *Networks and Communications (EuCNC), 2014 European Conference on*. IEEE, 2014, pp. 1–5.

[17] M. Jaber, M. Imran, R. Tafazolli, and A. Tukmanov, "An adaptive backhaul-aware cell range extension approach," in *2015 IEEE International Conference on Communication Workshop (ICCW)*. IEEE, 2015, pp. 74–79.

[18] F. Pantisano, M. Bennis, W. Saad, and M. Debbah, "Cache-aware user association in backhaul-constrained small cell networks," in *Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), 2014 12th International Symposium on*. IEEE, 2014, pp. 37–42.

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

[20] "C - Data Types . [online]. available: https://www.tutorialspoint.com/cprogramming/c_data_types.htm."

[21] L.-J. Lin, "Self-improvement based on reinforcement learning, planning and teaching," in *Maching Learning: Proceedings of the Eighth International Workshop*, 2014, pp. 323–327.

[22] G. A. Rummery and M. Niranjan, *On-line Q-learning using connectionist systems*. University of Cambridge, Department of Engineering, 1994.

[23] G. Głowaty, "Enhancements of fuzzy Q-learning algorithm," *Computer Science*, vol. 7, no. 4, p. 77, 2013.