Low-latency Communications for Community Resilience Microgrids: A Reinforcement Learning Approach

Medhat Elsayed*, Melike Erol-Kantarci*, *Senior Member, IEEE*, Burak Kantarci*, *Senior Member, IEEE*,
Lei Wu‡, *Senior Member, IEEE*, Jie Li†, *Member, IEEE*

*School of Electrical Engineering and Computer Science
University of Ottawa, ON
‡Department of Electrical and Computer Engineering
Stevens Institute of Technology, NJ
† Department of Electrical and Computer Engineering
Clarkson University, NY
Emails: {melsa034, melike.erolkantarci, burak.kantarci}@uottawa.ca, lei.wu@stevens.edu, jieli@clarkson.edu

***Abstract*—Machine learning and Artificial Intelligence (AI) techniques can play a key role in resource allocation and scheduler design in wireless networks that target applications with stringent QoS requirements such as near real-time control of Community Resilience Microgrids (CRMs). Specifically, for integrated control and communication of multiple CRMs, a large number of microgrid devices need to coexist with traditional mobile User Equipments (UEs), which are usually served with self-organized and densified wireless networks with many small cell base stations (SBSs). In such cases, rapid propagation of messages becomes challenging. This calls for a design of efficient resource allocation and user scheduling for delay minimization. In this work, we introduce a resource allocation algorithm, namely, Delay Minimization Q-learning (DMQ) scheme, which learns the efficient resource allocation for both the macro cell base stations (eNB) and the SBSs using reinforcement learning at each Time-To-Transmit Interval (TTI). Comparison with the traditional Proportional Fairness (PF) algorithm and an optimization-based algorithm, namely Distributed Iterative Resource Allocation (DIRA) reveals that our scheme can achieve 66% and 33% less latency, respectively. Moreover, DMQ outperforms DIRA and PF in terms of throughput while achieving the highest fairness.**

***Index Terms*—Community resilience microgrid, Low-latency communications, Reinforcement learning, Resource allocation, Small cells, Smart grid.**

I. Introduction

Community Resilience Microgrids (CRMs) allow sharing distributed energy resources among multiple owners. The dynamic nature of CRMs calls for a reconfigurable control which inherently relies on low-latency and reliable communication networks [1]. In CRMs, all Micro Grid Devices (MGDs) including distributed energy resources, loads, controllers, and distribution equipment generate data. In addition, control information needs to propagate fast for islanding from the grid or allowing energy transactions among multiple consumers within the CRM [2]. Wireless mobile networks provide ubiquity while the heterogeneous network environment with dense deployment of small cells provides flexibility [3], and enables effective communication among heterogeneous assets within a CRM. However wireless mobile networks, such as LTE-based fourth generation (4G) networks or the future 5G networks, are not dedicated, and as a result, a large number of MGDs need to coexist with traditional mobile User Equipments (UEs). Therefore, resource allocation and scheduler design play a key role for communications and efficient control of CRMs, in particular, when MGDs are aimed to be scheduled with a minimum delay. The dynamic nature of CRMs and the possibility of dynamically configuring small cells in accordance with the CRMs call for intelligent techniques. Reinforcement learning, more specifically Q-learning, is a machine learning technique that has potential to improve the performance of dense small cell networks via efficient self-organization.

In literature, various resource allocation schemes have been proposed such as traditional maximum rate and Proportional Fairness (PF) or more recent reinforcement learning algorithms. In [4], the authors propose a multi-agent Q-learning based resource allocation algorithm to improve network capacity. Their results show that throughput is improved when using centralized Q-learning over distributed Q-learning. In [5], the authors propose a joint resource allocation and power allocation scheme using cooperative Q-learning on femtocell networks with an objective of capacity maximization. In [6], the authors use Q-learning to enhance performance of Device-to-Device networks, again with the objective of maximizing throughput. In our previous work in [7], we proposed a throughput maximization Q-learning algorithm that aims at learning an efficient resource block allocation to improve the throughput of data intensive devices. Many other variants of the Q-learning algorithm have been proposed [8]–[14] in general for throughput maximization. Minimizing delay has been studied in a few works. In [15], the authors performed resource allocation for packet delay minimization in multi-layer unmanned aerial vehicles using gradient descent with bisection method which does not involve machine learnign. Most recently, in [16], we further utilized deep reinforcement learning to perform resource allocation for latency minimization in small-cell LTE networks.

In this paper, we propose a Q-learning-based resource allocation scheme by aiming at lower latency and improved fairness. A salient feature of our Delay Minimization Q-learning (DMQ) scheme is its ability to capture CRM network dynamics without *a priori* information. In addition, the proposed algorithm is fully distributed which promotes independent learning, and lowers signaling overhead on the network. Moreover, the design of the reward function achieves both low-latency and high fairness among MGDs and UEs. Finally, the integrated design of Q-learning and two-tier small cell network allows for great flexibility in deployment and fast network adaptability. Performance results show 33% and 66% latency reduction for MGDs when compared to previously proposed Distributed Iterative Resource Allocation (DIRA) which is an optimization based solution and the traditional Proportional Fairness (PF) algorithms, respectively. Meanwhile, a significant improvement in throughput and fairness is also achieved by the proposed scheme which makes DQM tailored for the connected microgrids of the future smart grid. It is worth mentioning that our approach can be adopted to other latency critical applications.

The paper is organized as follows: Section II presents the CRM communication and control model, covering state of art in microgrid communications and motivation of the paper. Section III presents the small cell network architecture and problem formulation. The proposed Q-learning-based algorithm along with the baseline algorithms are presented in section IV. Next, we present the performance of our proposed scheme and compare with traditional and optimization-based schemes in Section V. Finally, Section VI concludes the paper.

II. Background

A. Community Resilience Microgrid (CRM)

An increasing frequency of catastrophic weather events has been observed recently in the United States and globally, which has brought serious social and economic impacts. A critical issue associated with such catastrophic events is the availability of electricity for recovery efforts [17]. The smart grid is expected to heal itself under extreme circumstances [18]. In response to this, CRMs have been sought for enhancing resilient electricity supply to critical loads in a community during such disruption events. A CRM is a microgrid that is expected to supply electricity uninterruptedly during the damage phase and initial recovery period of a resiliency event. As shown in Fig. 1, a CRM includes multiple distributed energy resources and critical loads that are owned/controlled by different entities within a clearly defined electrical boundary, which are connected via primary distribution lines owned by a local regulated power company [2], [19], [20].

However, CRMs, as complex networked systems, exhibit unique structure and bring new challenges for the operation and control. The methods used for control of standalone microgrids, such as droop control [21], need to be tailored to CRMs. Specifically, as CRMs present a variety of dynamical behaviors ranging from minutes and hours to milliseconds, such as real-time uncertain loads and renewable generation outputs, maintaining reliable operation of CRMs in terms of voltage and frequency stability calls for real-time response and control. Consequently, a three-level hierarchical control architecture, including hour-to-minute-level tertiary control, second-level secondary control, and millisecond-level primary control, is usually deployed to realize secure and cost-effective operation and coordinate multiple partners in CRMs.

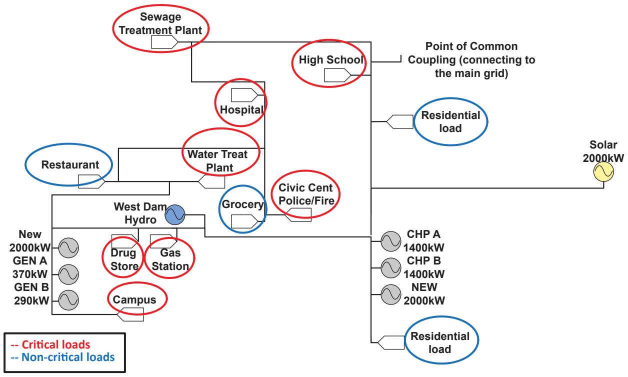

Fig. 1. Conceptual design of a CRM with critical and non-critical loads.

1) A Minute-Level Unbalanced AC Optimal Power Flow (ACOPF) Based Tertiary Control: is used to determine the most economic set points of DERs, loads, and voltage/reactive power regulation devices. The unbalanced ACOPF model in general is non-convex, nonlinear, and non-deterministic polynomial-time hard (NP-hard), because of the quadratic relationship among voltages and real/reactive power injections of three phases at individual buses as shown in [22]. Early works solve ACOPF via different mathematical models including linear programming, quadratically-constrained quadratic programming, and nonlinear programming, as well as various solution algorithms including Lagrange relaxation, interior point, heuristic, and convex relaxation approaches. However, many of these approaches rely on strong assumptions (e.g., single-phase radial or weak mesh networks), which are invalid for practical distribution networks of community microgrids. Recently, in [23], ACOPF has been formulated as a moment relaxation based semidefinite programming (SDP) model, which demonstrated that considering sparsity of the distribution network can drastically improve computational capacity by 10-100 times while guaranteeing solution quality of the same level. Certainly, low-latency communications will be needed for messaging between controllers as well as other entities.

2) A Second-Level Secondary Control: is used to restore frequency and voltage levels and enable islanding and resynchronization. The secondary control compensates for deviations induced by the primary control at the level of seconds, while also performing the synchronization control to seamlessly island and resynchronize with the main grid. Specifically, with a sudden change in demands followed by adjustment of DERs to achieve balance, if frequency and/or voltage deviates from the rated value (i.e., [0.95, 1.05] pu and [59.3, 60.5] Hz per ANSI 84.1-2016 standards), the secondary control generates a compensation signal to restore the rated frequency/voltage. At this level of controllers, communication latency becomes even more of a concern as the latency requirements becomes more stringent.

3) Millisecond-Level Primary Control: is used to stabilize voltage/frequency in real-time and provide plug-and-play capabilities. The primary control realizes active and reactive load sharing among parallel-connected DERs with plug-and-play capabilities, and stabilizes community microgrid frequency and voltage in real-time. This is especially important subsequent to islanding events, when community microgrid loses its voltage/frequency stability due to power mismatches. In recognizing that most DERs are interfaced via power electronic converters, droop control is widely used to determine adjustments of real/reactive power in response to frequency /voltage deviations. Primary control is the most delay-sensitive domain and benefits the most from low-latency communication techniques.

The key of the hierarchical control strategy is to effectively integrate the three control levels at different timescales. Indeed, frequency of interactions between different control levels can be optimized to achieve the trade-off between optimal power output tracking and economic operation. To facilitate the coordination, tertiary and second controls will receive minute-to-second level system information to support demand-supply balance calculations and determine optimal set points of local device controllers. On the other hand, certain time-critical control tasks, such as fault coordination and clearing as well as adjustments on presets of protective relays, are primarily performed at the individual controller level due to the speed of response required. In summary, different control levels would have distinct communication delay tolerance, and an efficient resource allocation and user scheduling approach is needed to optimally customize communication traffics, resource allocation, and delays of different needs. In this paper, we focus on the primary control as it poses the most stringent latency requirements. For example, in [24], latency requirements of substation automation is stated to be less than 100 ms. Our simulation results verify the suitability of the proposed algorithm by achieving latency values less than 50 ms for the worst-case scenario.

B. State-of-art in Microgrid Communications

The literature on microgrid communications has several notable works that address many metrics such as delay, reliability, security, etc. The research in [25] provides an overview of applying game-theoretic techniques for smart grid applications. In [26], the authors propose a security network architecture for data confidentiality and authentication, while taking the microgrid real-time communication into consideration. Authors in [27] consider an energy-aware optimization for aggregation of smart grid data packets. The optimization is formulated as a mixed integer non-linear problem that aims at finding the optimal data aggregator, optimal transmit power, and optimal number of concatenated packets. In [28], the authors address the microgrid demand-response management maximization alongside with communication spectrum management. A joint optimization problem is formulated to balance demand-response management performance and the cost of imperfect communications. On the other hand, cellular networks are becoming an attractive solution for carrying microgrid traffic due to its significant performance. The authors in [29] conduct a survey on the evolution of cellular networks utilized for microgrid applications. The paper highlights the challenges and gains of LTE networks, with a focus on LTE Device-to-Device discussion.

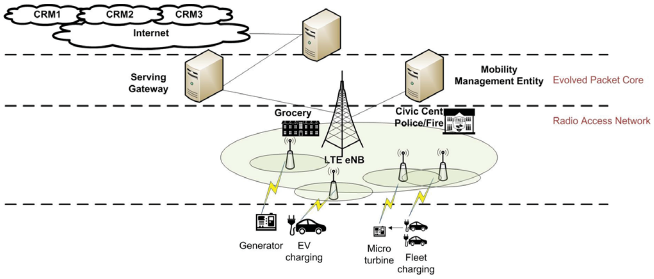

Fig. 2. A minimalist illustration of CRM communications over small cell wireless networks. A two-tier wireless network of an eNB underlaid with SBS covering users and the evolved packet core.

III. System Model

A. Small Cell Network

Our network model considers a two-tier network of an eNB underlaid with J Small-cell Base Stations (SBSs) covering N_J user devices. In general, users can be classified as either MicroGrid Devices (MGDs) such as smart meters, micro-phasor measurement units, etc, or conventional LTE User Equipments (UEs) such as smart phones, tablets, etc. All nodes follow the downlink and uplink communication according to LTE release 12 standard. The LTE frame structure consists of 10 subframes of 1 ms duration, whereas each subframe has two time slots of 0.5 ms each. According to Orthogonal Frequency Division Multiple Access (OFDMA), the LTE resource block grid is divided in time and frequency into a number of resource blocks (RBs) denoted by N_{RBs}. Each RB has N_{SC} subcarriers and spans N_{sym} OFDM symbols. The resource allocation and communication is performed in each subframe, namely Time-To-Transmit (TTI). We assume that all nodes transmit with the same amount of transmit power which means the problem translates into resource block allocation where spectrum and time allocation is tackled.

In Fig. 2, all nodes conform to Frequency Division Duplex with single antenna transmission. To remove cross-tier interference, we decompose the uplink (downlink) transmissions of both tiers. This is explained as follows. Uplink (downlink) of the links (users-SBS) and (SBS-eNB) use interleaving subframes to transmit their data. For example, users' uplink uses subframes (1, 3, 5, .. $2i+1$, .., $2n+1$), while SBS' uplink uses subframes (0, 2, 4, ...i,...$2n$). Although this succeeds to remove the cross-tier interference, co-tier interference still remains due to the dense deployment of SBSs, which causes users attached to one SBS and lying in the range of other SBS to cause interference in the adjacent cells.

Resource allocation process is performed by identifying the best resource block (RB) in time and frequency domains for active users in the network. Both the eNB and individual

SBSs perform resource allocation to allocate RBs to their attached users in each TTI. In each TTI, the users report their scheduling request to their attached base station (i.e., users report to SBSs, and SBSs report to eNB). The base station performs the resource allocation algorithm and informs the users with the allocated RBs to use in the next TTI. Performing the resource allocation on the two-tier network (i.e., eNB, and SBSs) reduces the burden on the eNB, as well as facilitates small cells for capacity and coverage improvement.

B. Channel Model

The wireless channel between nodes can be prone to multiple fading sources. We use the 3GPP pathloss model following [30]–[32]. $PL_{3GPP} = 128.1{+}37.6{*}log_{10}(d)$, where PL_{3GPP} is pathloss in dB, and d is distance between the base station and the user in km [33]. The shadowing effect is modeled as a log-normal distribution with zero-mean and 10 dB variance.

C. Traffic Model

Following the traffic model proposed by 3GPP TR 37.868 for machine-type communication (MTC) [34], we apply a Beta/M/1 queue on each node. Therefore, the corresponding average queue delay (waiting time) for user i attached to base station j can be formulated as in [35]:

$$D^{q}_{i,j} = \frac{B_{i,j}}{\mu_{i,j}\ (1 - B_{i,j})} \tag{1}$$

$$B_{i,j} = M_A(\mu_{i,j}\ B_{i,j} - \mu_{i,j}) \tag{2}$$

where $\mu_{i,j} = R_{i,j}/L_{i,j}$ is the service rate on link (i, j), $R_{i,j}$ is the instantaneous rate, $L_{i,j}$ is the packet size, $M_A(\mu_{i,j}) = \int_{-\infty}^{\infty} f(x)e^{ux}dx$ is the moment generating function of the arrival process with PDF $f(x)$, inter-arrival time A, and mean $1/\lambda$, and Bi, j is the Markovian transition probability solution as shown in [36] for the G/M/1 queue. In 3GPP [34], PDF of the inter-arrivals of MTC devices follows a Beta distribution with shape parameters $a = 3$ and $b = 4$, and the moment generating function of Beta/M/1 can be derived as in [35]:

$$M_A(u) = 15\ \frac{d^2}{du^2}\Big[\frac{4!}{u^4}\ (e^u - 1 - u - \frac{u^2}{2} - \frac{u^3}{6})\Big] \tag{3}$$

D. Problem Formulation

Delay on the link between user i and base station j (i.e., link (i, j)) can be formulated as in Eq. 4:

$$D_{i,j} = D^{tr}_{i,j} + D^{q}_{i,j} \tag{4}$$

where $D^{tr}_{i,j}$ is transmission delay on link (i, j), and $D^{q}_{i,j}$ is queuing delay on link (i, j). Equation 5 formulates the transmission delay on link (i, j), where $L_{i,j}$ is packet size and $R_{i,j}$ is transmission rate. Delay and transmission rate can be formulated as follows:

$$D^{tr}_{i,j} = \frac{L_{i,j}}{R_{i,j}} \tag{5}$$

$$R_{i,j} = \sum_{k=1}^{K} x_{i,j,k}\ r_{i,j,k} \tag{6}$$

and

$$r_{i,j,k} = W_k\ log_2\Big(1 + \frac{x_{i,j,k}\ p_{i,j,k}\ h_{i,j,k}}{W_k\ N_0\ + \sum\limits_{\substack{m \neq i \\ m \in N_J}} x_{m,j,k}\ p_{m,j,k}\ h_{m,j,k}}\Big) \tag{7}$$

where $r_{i,j,k}$ is rate on RB k, $x_{i,j,k}$ indicates if RB k is allocated to user i, W_k is bandwidth of RB k, N_0 is Additive White Gaussian Noise (AWGN) single-sided power spectral density, $p_{i,j,k}$ is transmit power of i^{th} node on k^{th} RB, $h_{i,j,k}$ is channel coefficient of k^{th} RB, and $p_{m,j,k}$ is transmit power of interfering node m on the k^{th} RB of the j^{th} node.

IV. Proposed Delay-Minimizing Resource Allocation

In this section, we provide a detailed description of our proposed algorithm, namely delay minimization using Q-learning (DMQ) for RB allocation. Furthermore, we provide the details of the baseline algorithms that are used in comparisons.

A. Delay minimization using Q-learning (DMQ)

DMQ is a decentralized algorithm where multiple agents (i.e., SBSs/eNB) aim at learning a sub-optimal decision policy by taking actions and computing feedback from the environment. The algorithm is represented by the tuple {agents, states and actions} and *reward function (i.e., the Q-value)*. The convergence point is reached when each agent learns a state-action pair that maximizes its reward over infinite time horizon. This can be realized by having a Q-value representing the agents' reward over iterations. Hence, the optimal decision would be actions corresponding to the maximum reward (i.e., max Q-value). We define the DMQ tuple as follows:

1) ***Agents****:* The macro-cell / small-cell base stations (i.e., eNB / SBSs) form the set of agents in the DMQ tuple.

2) ***States****:* The Q-learning agents perform a search to find the best resource allocation vector for its attached users. To this end, the number of states is limited to one.

3) ***Action space****:* Each base station (eNB/SBS) performs RB-to-user mapping for its attached users on uplink. Hence, we denote $a_{j,t}$ as the decision of base station j at TTI t regarding the set of RBs allocated to its attached users. Consequently, dimension of the action space is $m = N^K$, where K is total number of RBs in a subframe and N is number of users. In order to avoid high dimensionality, it is viable to allocate users in a group of contiguous RBs, namely Resource Block Group (RBG). This reduces the action space and helps the algorithm converge fast. Lastly, ϵ-greedy is used to account for action space exploration.

4) ***Reward/Cost function****:* We define the reward function as follows:

$$RC_j(S_{j,t}, a_{j,t}) = \beta\ \tau_{j,c} + (1 - \beta)\ \tau_{j,n} \tag{8}$$

where, β is a scalar weight to control priorities of individual loads (i.e., traffic from MGD and UE), $\tau_{j,c}$ and $\tau_{j,n}$ are defined as follows:

$$\tau_{j,c} = \Big(\frac{-2}{\pi}\Big)\ \arctan(D_{j,c}) \tag{9}$$

$$\tau_{j,n} = \Big(\frac{-2}{\pi}\Big)\ \arctan(D_{j,n}) \tag{10}$$

where, $D_{j,c}$ and $D_{j,n}$ are the average delays of critical and non-critical loads, respectively. This function rewards the critical load delay with a positive reward as long as the achieved average delay is low. At the same time, it aims at minimizing delay of non-critical loads in order to maintain fairness among users.

*5) **Q-Value**:* The *Q-Value* is calculated using the Temporal Difference equation as in [37]:

$$Q(S_{j,t}, a_{j,t}) = (1-\alpha)Q(S_{j,t}, a_{j,t}) + \alpha[RC_j(S_{j,t}, a_{j,t}) + \gamma \max_{a_{j,t}} Q(S_{j,t+1}, a_{j,t})] \tag{11}$$

*6) **Policy**:* The policy of base station j at TTI t is denoted by $\pi_{j,t}$ and aims to maximize the Q-value:

$$\pi_{j,t} = arg \max_{a_{j,t}} Q(S_{j,t}, a_{j,t}) \tag{12}$$

The algorithm works in a two-tier scheduling fashion on both the eNB and SBSs. That is, the eNB represents the first tier agent, and its attached SBSs are considered as its *environment*. Each SBS constitutes the second tier agent, with its attached users as the environment. SBS/users report their channel state information to eNB/SBS, respectively, on the uplink transmission. The channel state information enables the eNB/SBS to estimate the link quality of the allocated RBs thanks to the Channel Quality Indicator (CQI), Signal to Interference plus Noise Ratio (SINR), and total delay of the previous packet included in the channel state information. Algorithm 1 presents the Q-learning steps performed by each j^{th} agent. Algorithm 1 is repeated for the entire simulation time (T), during which the algorithm either performs exploration (i.e., random action selection) or exploitation (i.e., select actions with max Q-value). The same algorithm runs on both the eNB and SBSs, where the eNB is responsible of scheduling SBSs on the uplink.

Algorithm 1 Delay minimization using Q-learning (DMQ)

```
Initialization: Q-Table ← 0, $\epsilon$, and $T$ (Simulation time).
while $i < T$ do
    Generate a uniform random variable, $M$.
    if $M \geq \epsilon$ then
        $a_{j,t} \leftarrow arg\ rand\ \{a_{j,t}\}$.
    else if $M < \epsilon$ then
        // Exploit using Q-learning policy
        $a_{j,t} \leftarrow arg \max\{Q(S_{j,t}, a_{j,t})\}$.
    end if
    Calculate the Reward using Equation (8).
    Update $Q(S_{j,t}, a_{j,t})$ using Equation (11).
    Increment i.
end while
```

B. Baseline algorithms

To evaluate the effectiveness of our proposed solution, DMQ, we compare its performance to two baseline algorithms, which are briefly introduced in this subsection.

1) Proportional Fairness (PF): PF is a well-known algorithm that aims to give priority to the user having the maximum relative channel condition. The utility function is formulated as:

$$u_i^* = arg \max_{i=1,\ldots,N_j} \frac{(R_{i,k}(t))}{(T_{i,k}(t))}, \tag{13}$$

where $R_{i,k}(t)$ is the instantaneous rate of user i at RB k on TTI t, $T_{i,k}(t)$ is the moving average rate of user i [38], [39], and u_i^* is the user achieving the highest relative channel conditions. The moving average rate can be computed as in [40]:

$$T_{i,k}(t+1) = \begin{cases} (1-\frac{1}{t_w})\ T_{i,k}(t) + \frac{1}{t_w}\ R_{i,k}(t), i^* = i \\ (1-\frac{1}{t_w})\ T_{i,k}(t), i^* \neq i \end{cases} \tag{14}$$

where t_w is the history window length.

2) Distributed Iterative Resource Allocation (DIRA): We compare our proposed scheme with an optimization-based solution that targets delay-sensitive users similar to our work [41]. To make DIRA comparable to our scheme, we slightly modify the original algorithm to consider only RB allocation and omit power allocation. Furthermore, to have a fair comparison, we run DIRA on both network tiers (i.e., at the eNB and SBSs). To the best of our knowledge, the literature lacks an algorithm that both considers two-tier architecture and targets low-latency while tackling the resource block allocation problem. Therefore, DIRA is chosen to compare our results to a baseline solution that aims to provide low latency. Resource block allocation with DIRA can be formulated as:

$$\max_{x_{i,j,k}} \quad \sum_{j=1}^{J}\sum_{i=1}^{N_j}\sum_{k=1}^{K} x_{i,j,k}\ r_{i,j,k} \tag{15}$$

Subject to:

$$\sum_{k=1}^{K} x_{i,j,k}\ p_{i,j,k} \leq P_{max}, \forall j, i \tag{15a}$$

$$p_{i,j,k} \geq 0, \forall i, k \tag{15b}$$

$$\sum_{k=1}^{K} x_{i,j,k}\ r_{i,j,k} \geq R_u, \forall i \in MGDs, \forall j \tag{15c}$$

$$\sum_{i=1}^{N_j} x_{i,j,k} \leq 1, \forall j, i \tag{15d}$$

$$x_{i,j,k} \in 0, 1, \forall j, i, k \tag{15e}$$

where R_u is the aggregate capacity threshold of MGDs in each base station, P_{max} is the maximum transmission power of each user i. (15) aims to maximize the aggregate network rate through RB allocation. (15a) limits the power allocation of each user i to P_{max} on all of its RBs. (15c) guarantees a minimum achievable spectral efficiency, R_u, to each user i. (15d) and (15e) guarantee that each RB can only be assigned to one user within each cell. Following the same derivation methodology presented in [41], the following formula can be obtained:

$$H_{i,j,k} = (1+\hat{\nu}_{i,j})r_{i,j,k} - \theta_{i,j}\ p_{i,j,k} - (1+\hat{\nu}_{i,j})\ \frac{1}{ln(2)}\ \Big(\frac{p_{i,j,k}\ h_{i,j,k}}{p_{i,j,k}\ h_{i,j,k} + I_{i,j,k}}\Big) \tag{16}$$

where $I_{i,j,k} = p_{m,j,k}\ h_{m,j,k} + \sigma^2$ is the interference on link (i,j) on RB k.

$$\hat{\nu}_{i,j} = \begin{cases} \nu_{i,j}, & \forall i, j \in MGDs \\ 0, & Otherwise \end{cases} \tag{17}$$

where $\nu_{i,j}$, and $\theta_{i,j}$ are Lagrangian multipliers obtained using the subgradient method. Hence, RB k is assigned to the user with the largest $H_{i,j,k}$ as follows:

$$\hat{x}_{i^*,j,k} = 1|_{i^*=\max\limits_{i} H_{i,j,k}}\ , \quad \forall j, k \tag{18}$$

where $\hat{x}_{i^*,j,k}$ is the RB allocation decision to selected user i^*.

V. Performance Evaluation

We use the LTE system toolbox in Matlab to design a discrete-level simulator for our network setup. Table I summarizes simulation settings used in the evaluation of the proposed and baseline algorithms. The simulation considers one eNB covering 20 SBSs, where eNB and SBS radii are 800m and 50m [42], [43], respectively. The pathloss model is 3GPP model, penetration loss is 20 dB, and receivers noise figure is 9 dB [44]. The DMQ uses a learning rate α of 0.5, a discount factor γ of 0.9, and ϵ of 0.8 [45]. All results are averaged over 5 testing runs, where each run is 500 subframes (i.e., 500 msec). A 95% confidence interval is provided in all our simulation results.

Fig. 3 presents the average packet delay versus the number of MGDs with 10 SBSs and 5 UEs per SBS. DMQ achieves the lowest transmission latency for both MGDs and UEs. Although the delay increases with the increase in the number of MGDs, as expected, DMQ still achieves the lowest delay compared to the other algorithms. Fig. 4 presents the average queuing delay for MGDs and UEs. This accounts for time that the packets have to wait until the resource block allocated to them becomes available. DMQ achieves the lowest queuing delay, with some degradation when increasing the number of MGDs. However, it still has the lowest delay trend. It is also observed that most of the end-to-end delay is due to queuing delay.

Fig. 5 presents the average throughput versus number of MGDs. The results show that DMQ outperforms DIRA and PF. However, increasing the number of MGDs/SBS degrades the DMQ's throughput. The main reason behind this is that DMQ's main aim is to decrease the end-to-end latency of MGDs while maintaining fairness among MGDs and UEs. Therefore, as can be seen in Fig. 3, both MGDs and UEs delays are decreased, whereas this comes on the price of higher throughput degradation, especially in dense scenarios. In Fig. 6, we show the top-10 users' throughput, which again shows a better performance of DMQ. Yet, throughput of DMQ is impacted by the number of MGDs more than the other algorithms. Once again, for denser networks, throughput results converge since the available resources are limited.

To study the fairness of DMQ, Jain's fairness index is plotted in Fig. 7. Since the reward function of DMQ aims to minimize UEs delay as well, it provides fairness among users. Our results show fairness values that exceed PF fairness by about 2%.

Fig. 8 presents the impact of a longer learning phase on both the delay and throughput results under the proposed DMQ scheme. It can be seen that performing more action-space exploration allows the algorithm to learn better resource allocation actions, hence the delay decreases and throughput increases at the same time. However, this also leads to the requirement of longer training time for improving performance.

In summary, DMQ performs better than DIRA and PF, in terms of delay, throughput and fairness. However, its throughput degrades in a faster trend than DIRA and PF. As a trade-off DMQ favors delay and fairness over throughput, which can be observed from the reward design in eq. (8). Note that, the average latency and throughput performance of MGDs and UEs is close for DMQ, as well PF, since both algorithms have fairness in their objective. Yet, DMQ results in lower latency and higher throughput for both types of devices than the compared algorithms.

TABLE I
Simulation settings

General parameters	
Time-to-Transmit Interval (TTI)	1 msec
Resource allocation algorithms	DMQ, PF and DIRA.
eNB radius	800 m
SBS radius	50 m
Min distance between SBSs	30 m
Number of eNBs	1
Number of SBSs per eNB	10
Number of MGDs per SBS	4:2:12
Number of UEs per SBS	5
Speed of users	Fixed positions
MGDs Traffic model	Beta ($\alpha = 3, \beta = 4$) [34]
UEs Traffic model	Poisson
Packet mean Inter-arrival time	5 milli-seconds.
Packet size	Exponential (mean = 25 Bytes)
Transmission bandwidth	10 MHz
Number of RBs	50 (12 subcarriers / RB)
Number of RBGs	5 (10 RBs/RBG)
eNB Tx power	40 dBm [46]
SBS Tx power	20 dBm [46]
Pathloss model	3GPP $PL_{dB} = 128.1 + 37.6 * log_{10}(d)$
Penetration loss	20 dB
Noise Figure	9 dB
Shadowing	$\sim$ LOGN(0, 10(dB))
Proportional Fairness (PF)	
ϕ	1 (PF)
ζ	1 (PF)
t_c (window)	2
DMQ	
Learning rate (α)	0.5
Discount factor (γ)	0.9
Exploration probability (ϵ)	0.8
Priority weight of MGDs (β)	0.9
DIRA	
R_u	9 bps/Hz [41]

TABLE II
Comparison among the three algorithms

Criteria	PF	DIRA	DMQ
Resource allocated	Spectrum	Spectrum (Power removed)	Spectrum
Objective	Rate and Fairness	Rate and delay constraint	Delay
Network Model	two-tier	Adapted to two-tier	two-tier
Complexity (per TTI per BS)	$O(N_j)$	$O(N_j)$	$O(N_j)$

Lastly, a comparison among the three algorithms is presented in Table II, where N_j is the number of users per base station. The table presents the modeling assumptions as well as the complexity and drawbacks. DIRA was adopted to work on both tiers, furthermore, we revised the optimization to account for spectrum allocation only - removing the power allocation. The complexity is presented in Big-O notation, evaluated per base station per TTI.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2019.2931753, IEEE Transactions on Smart Grid

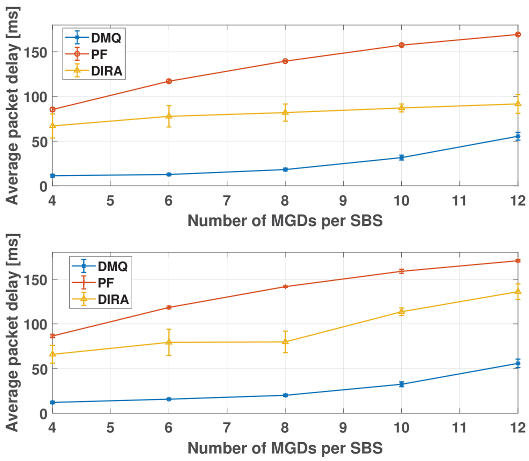

Fig. 3. Average packet delay [ms] for (top) MGDs and (bottom) UEs vs number of MGDs; number of SBS is 10 and number of UEs is 50.

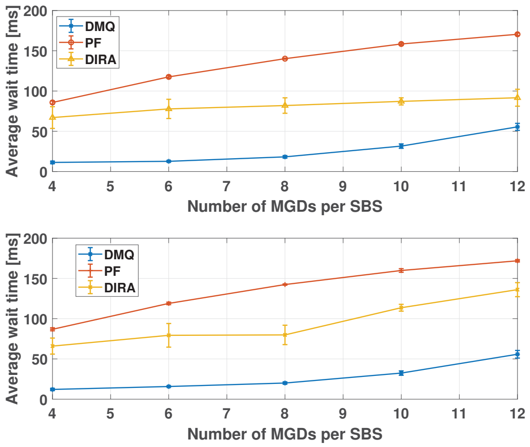

Fig. 4. Average queuing delay [ms] for (top) MGDs and (bottom) UEs vs number of MGDs; number of SBS is 10, and number of UEs is 50.

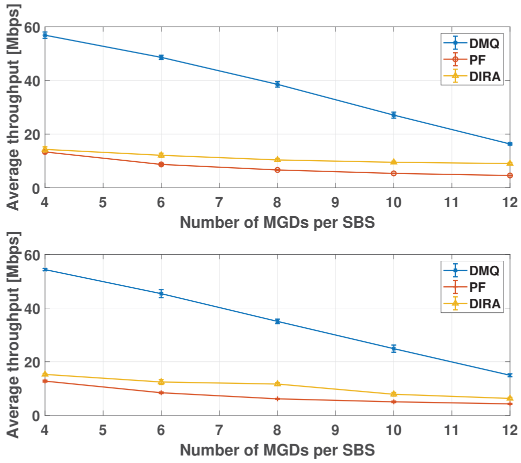

Fig. 5. Average throughput [Mbps] for (top) MGDs and (bottom) UEs vs number of MGDs; number of SBS is 10, and number of UEs is 50.

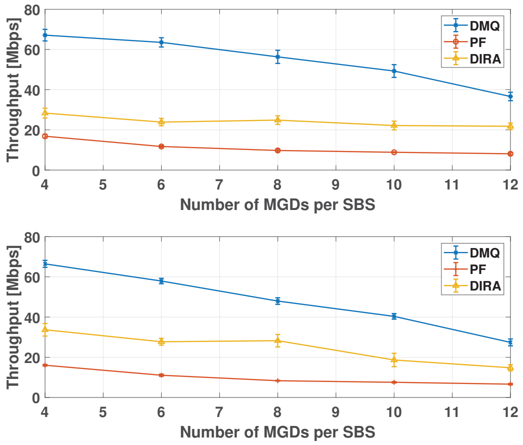

Fig. 6. Top-10 throughput [Mbps] for (top) MGDs and (bottom) UEs vs number of MGDs; number of SBS is 10, and number of UEs is 50.

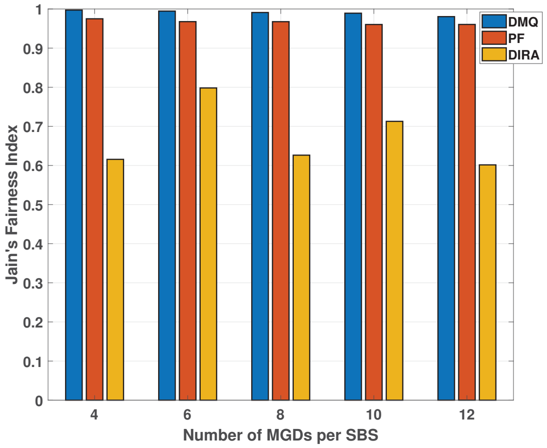

Fig. 7. Jain's Fairness Index (JFI) vs MGDs; number of SBS is 10, and number of UEs is 50.

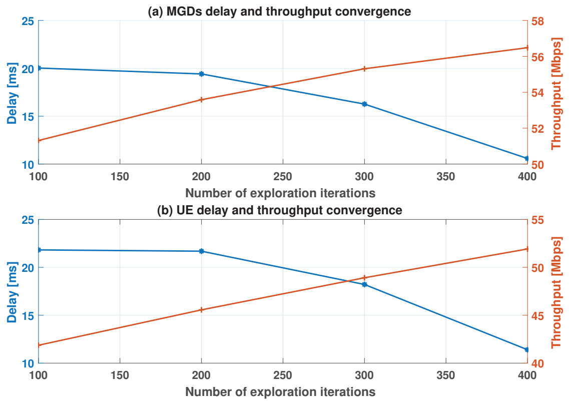

Fig. 8. Average delay and throughput convergence for (a) MGDs, and (b) UEs vs number of exploration iterations (in TTIs); 10 SBSs, 8 MGDs and 5 UEs per SBSs.

VI. Conclusion

In this paper, we have proposed a resource allocation algorithm with the purpose of providing low-latency communications for primary control of Community Resilience Microgrids (CRMs) that use small cell networks. Our proposed resource allocation algorithm, namely the Delay minimization using Q-learning (DMQ) aims at low delay for microgrid devices (MGD) while concurrently achieving low delay for UEs through an effective reward function. Furthermore, DMQ runs in a decentralized fashion on both the SBSs and eNB for two-tier scheduling, which facilitates the network agility and self-organization. We have compared DMQ with the well-known proportional fairness (PF) algorithm as well as an algorithm with delay-sensitive users, namely Distributed Iterative Resource Allocation (DIRA). The results show delay reduction of 66% and 33% is obtained for MGDs when compared to PF and DIRA, respectively. In addition, higher throughput is achieved. Meanwhile, DMQ has the highest fairness values among the other schemes where it exceeds the fairness index of PF by 2%. As a future work, we plan to integrate an online learning approach in order to further reduce latency and enhance the performance of training.

Acknowledgment

This research is supported by the U.S. National Science Foundation (NSF) under Grant Number CNS-1647135 and the Natural Sciences and Engineering Research Council of Canada (NSERC) under RGPIN-2017-03995.

References

[1] M. Erol-Kantarci, B. Kantarci, and H. T. Mouftah, "Reliable overlay topology design for the smart microgrid network," *IEEE Network, Special issue on Communication Infrastructures for Smart Grid*, vol. 25, pp. 38–43, September 2011.

[2] L. Wu, J. Li, M. Erol-Kantarci, and B. Kantarci, "An integrated reconfigurable control and self-organizing communication framework for community resilience microgrids," *The Electricity Journal*, vol. 30, no. 4, pp. 27 – 34, 2017. Special Issue: Contemporary Strategies for Microgrid Operation and Control.

[3] M. Peng, C. Wang, J. Li, H. Xiang, and V. Lau, "Recent advances in underlay heterogeneous networks: Interference control, resource allocation, and self-organization," *IEEE Communications Surveys Tutorials*, vol. 17, pp. 700–729, Secondquarter 2015.

[4] M. Chen, Y. Hua, X. Gu, S. Nie, and Z. Fan, "A self-organizing resource allocation strategy based on q-learning approach in ultra-dense networks," in *2016 IEEE International Conference on Network Infrastructure and Digital Content (IC-NIDC)*, pp. 155–160, Sept 2016.

[5] A. Shahid, S. Aslam, H. S. Kim, and K.-G. Lee, "A docitive q-learning approach towards joint resource allocation and power control in self-organised femtocell networks," *Transactions on Emerging Telecommunications Technologies*, vol. 26, no. 2, pp. 216–230, 2015.

[6] Y. Luo, Z. Shi, X. Zhou, Q. Liu, and Q. Yi, "Dynamic resource allocations based on q-learning for d2d communication in cellular networks," in *2014 11th International Computer Conference on Wavelet Actiev Media Technology and Information Processing(ICCWAMTIP)*, pp. 385–388, Dec 2014.

[7] M. Elsayed and M. Erol-Kantarci, "Learning-based resource allocation for data-intensive and immersive tactile applications," in *IEEE 5G World Forum*, July 2018.

[8] H. Saad, A. Mohamed, and T. ElBatt, "A cooperative q-learning approach for distributed resource allocation in multi-user femtocell networks," in *2014 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1490–1495, April 2014.

[9] Y. Hu, R. MacKenzie, and M. Hao, "Expected q-learning for self-organizing resource allocation in lte-u with downlink-uplink decoupling," in *European Wireless 2017; 23th European Wireless Conference*, pp. 1–6, May 2017.

[10] I. S. Comsa, S. Zhang, M. Aydin, P. Kuonen, and J. F. Wagen, "A novel dynamic q-learning-based scheduler technique for lte-advanced technologies using neural networks," in *37th Annual IEEE Conference on Local Computer Networks*, pp. 332–335, Oct 2012.

[11] M. Dirani and Z. Altman, "A cooperative reinforcement learning approach for inter-cell interference coordination in ofdma cellular networks," in *8th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks*, pp. 170–176, May 2010.

[12] A. Asheralieva and Y. Miyanaga, "Multi-agent q-learning for autonomous d2d communication," in *2016 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, pp. 1–6, Oct 2016.

[13] Z. Li, Z. Lu, X. Wen, W. Jing, Z. Zhang, and F. Fu, "Distributed power control for two-tier femtocell networks with qos provisioning based on q-learning," in *2015 IEEE 82nd Vehicular Technology Conference (VTC2015-Fall)*, pp. 1–6, Sept 2015.

[14] M. Elsayed and M. Erol-Kantarci, "Deep Q-Learning for Low-Latency Tactile Applications: Microgrid Communications," in *IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids Workshops*, 2018.

[15] J. Li and Y. Han, "Optimal resource allocation for packet delay minimization in multi-layer uav networks," *IEEE Communications Letters*, vol. 21, pp. 580–583, March 2017.

[16] M. Elsayed and M. Erol-Kantarci, "Deep Reinforcement Learning for Reducing Latency in Mission Critical Services," in *IEEE Global Communications Conference*, 2018.

[17] , "Billion-dollar weather/climate disasters," Tech. Rep. 36.104, National Oceanic and Atmospheric Administration, January 2014.

[18] T. Jiang, H. Wang, M. Daneshmand, and D. Wu, "Cognitive radio-based smart grid traffic scheduling with binary exponential backoff," *IEEE Internet of Things Journal*, vol. 4, pp. 2038–2046, Dec 2017.

[19] L. Wu, T. Ortmeyer, and J. Li, "The community microgrid distribution system of the future," *The Electricity Journal*, vol. 29, no. 10, pp. 16 – 21, 2016.

[20] T. Ortmeyer, L. Wu, and J. Li, "Planning and design goals for resilient microgrids," in *2016 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, pp. 1–5, Sept 2016.

[21] R. Hidalgo-León, C. Sanchez-Zurita, P. Jácome-Ruiz, J. Wu, and Y. Muñoz-Jadan, "Roles, challenges, and approaches of droop control methods for microgrids," in *2017 IEEE PES Innovative Smart Grid Technologies Conference - Latin America (ISGT Latin America)*, pp. 1–6, Sept 2017.

[22] J. Lavaei and S. H. Low, "Zero duality gap in optimal power flow problem," *IEEE Transactions on Power Systems*, vol. 27, pp. 92–107, Feb 2012.

[23] Y. Liu, J. Li, L. Wu, and T. Ortmeyer, "Chordal relaxation based acopf for unbalanced distribution systems with ders and voltage regulation devices," *IEEE Transactions on Power Systems*, vol. 33, pp. 970–984, Jan 2018.

[24] V. C. Gungor, D. Sahin, T. Kocak, S. Ergut, C. Buccella, C. Cecati, and G. P. Hancke, "A survey on smart grid potential applications and communication requirements," *IEEE Transactions on Industrial Informatics*, vol. 9, pp. 28–42, Feb 2013.

[25] W. Saad, Z. Han, H. V. Poor, and T. Basar, "Game-theoretic methods for the smart grid: An overview of microgrid systems, demand-side management, and smart grid communications," *IEEE Signal Processing Magazine*, vol. 29, pp. 86–105, Sept 2012.

[26] V. Kounev, D. Tipper, A. A. Yavuz, B. M. Grainger, and G. F. Reed, "A secure communication architecture for distributed microgrid control," *IEEE Transactions on Smart Grid*, vol. 6, pp. 2484–2492, Sept 2015.

[27] F. Uddin, "Energy-aware optimal data aggregation in smart grid wireless communication networks," *IEEE Transactions on Green Communications and Networking*, vol. 1, pp. 358–371, Sept 2017.

[28] C. Yang, J. Yao, W. Lou, and S. Xie, "On demand response management performance optimization for microgrids under imperfect communication constraints," *IEEE Internet of Things Journal*, vol. 4, pp. 881–893, Aug 2017.

[29] C. Kalalas, L. Thrybom, and J. Alonso-Zarate, "Cellular communications for smart grid neighborhood area networks: A survey," *IEEE Access*, vol. 4, pp. 1469–1493, 2016.

[30] G. Pocovi, K. I. Pedersen, and P. Mogensen, "Multiplexing of latency-critical communication and mobile broadband on a shared channel," in

IEEE Wireless Communications and Networking Conference (WCNC), pp. 1–6, April 2018.

[31] H. Liao, P. Chen, and W. Chen, "An efficient downlink radio resource allocation with carrier aggregation in lte-advanced networks," *IEEE Transactions on Mobile Computing*, vol. 13, pp. 2229–2239, Oct 2014.

[32] D. López-Pérez, X. Chu, A. V. Vasilakos, and H. Claussen, "On distributed and coordinated resource allocation for interference mitigation in self-organizing lte networks," *IEEE/ACM Transactions on Networking*, vol. 21, pp. 1145–1158, Aug 2013.

[33] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Frequency (RF) requirements for LTE Pico Node B," Technical Specification (TS) 36.931, 3rd Generation Partnership Project (3GPP), 05 2014. Version 12.0.0.

[34] 3GPP, "Technical Specification Group GERAN; GERAN Improvements for Machine-type Communications," Technical Specification Group GERAN 43.868, 3rd Generation Partnership Project (3GPP), 11 2011. Version 0.5.0.

[35] X. Jian, X. Zeng, Y. Jia, L. Zhang, and Y. He, "Beta/m/1 model for machine type communication," *IEEE Communications Letters*, vol. 17, pp. 584–587, March 2013.

[36] O. C. Ibe, "7 - markovian queueing systems," in *Markov Processes for Stochastic Modeling (Second Edition)* (O. C. Ibe, ed.), pp. 178 – 182, Oxford: Elsevier, second edition ed., 2013.

[37] E. Alpaydin, *Introduction to Machine Learning*. MIT Press, 2014.

[38] T. B. Sorensen and M. R. Pons, "Performance evaluation of proportional fair scheduling algorithm with measured channels," in *VTC-2005-Fall. 2005 IEEE 62nd Vehicular Technology Conference, 2005.*, vol. 4, pp. 2580–2585, Sept 2005.

[39] J. Yang, Z. Yifan, W. Ying, and Z. Ping, "Average rate updating mechanism in proportional fair scheduler for hdr," in *Global Telecommunications Conference, 2004. GLOBECOM '04. IEEE*, vol. 6, pp. 3464–3466 Vol.6, Nov 2004.

[40] G. Miao, J. Zander, K. W. Sung, and S. Ben Slimane, "Scheduling," in *Fundamentals of Mobile Data Networks*, ch. 4, p. 65–94, Cambridge University Press, 2016.

[41] H. Zhang, C. Jiang, N. C. Beaulieu, X. Chu, X. Wen, and M. Tao, "resource allocation in spectrum-sharing ofdma femtocells with heterogeneous services," *IEEE Transactions on Communications*.

[42] A. Saeed, E. Katranaras, M. Dianati, and M. A. Imran, "Dynamic femtocell resource allocation for managing inter-tier interference in downlink of heterogeneous networks," *IET Communications*, vol. 10, no. 6, pp. 641–650, 2016.

[43] A. Saeed, E. Katranaras, M. Dianati, and M. A. Imran, "Dynamic femtocell resource allocation for managing inter-tier interference in downlink of heterogeneous networks," *IET Communications*, vol. 10, no. 6, pp. 641–650, 2016.

[44] C. C. Coskun and E. Ayanoglu, "Energy- and spectral-efficient resource allocation algorithm for heterogeneous networks," *IEEE Transactions on Vehicular Technology*, vol. 67, pp. 590–603, Jan 2018.

[45] Y. Y. Liu and S. J. Yoo, "Dynamic resource allocation using reinforcement learning for lte-u and wifi in the unlicensed spectrum," in *2017 Ninth International Conference on Ubiquitous and Future Networks (ICUFN)*, pp. 471–475, July 2017.

[46] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA); Base Station (BS) radio transmission and reception," Technical Specification 36.104, 3rd Generation Partnership Project (3GPP), 10 2014. Version 12.5.0.

Medhat Elsayed is a Ph.D. candidate at the University of Ottawa. He obtained his BSc and MSc degrees from Cairo University, Egypt, in 2009 and 2013 respectively. His research interests are wireless networks, 5G and beyond, artificial intelligence, and smart grids.

Melike Erol-Kantarci is an associate professor at the School of Electrical Engineering and Computer Science at the University of Ottawa. She is the founding director of the Networked Systems and Communications Research (NETCORE) laboratory. She is also a courtesy assistant professor at the Department of Electrical and Computer Engineering at Clarkson University, Potsdam, NY, where she was a tenure-track assistant professor prior to joining University of Ottawa. She received her Ph.D. and M.Sc. degrees in Computer Engineering from Istanbul Technical University in 2009 and 2004, respectively. She is an editor of the IEEE Communications Letters and IEEE Access. Her main research interests are wireless communications, AI-enabled wireless networks, smart grid, cyber-physical systems, electric vehicles, Internet of things and wireless sensor networks. She is a senior member of the IEEE.

Burak Kantarci (S'05 M'09 SM'12) is an Associate Professor with the School of Electrical Engineering and Computer Science at the University of Ottawa. From 2014 to 2016, he was an assistant professor at the ECE Department at Clarkson University, where he currently holds a courtesy appointment. Dr. Kantarci received the M.Sc. and Ph.D. degrees in computer engineering from Istanbul Technical University, in 2005 and 2009, respectively. He has co-authored over 130 papers in established journals and conferences, and contributed to 11 book chapters. He is an Editor of the IEEE Communications Surveys and Tutorials.

Lei Wu received the B.S. degree in electrical engineering and the M.S. degree in systems engineering from Xi'an Jiaotong University, Xi'an, China, in 2001 and 2004, respectively, and the Ph.D. degree in electrical engineering from the Illinois Institute of Technology, Chicago, IL, USA, in 2008. From 2008 to 2010, he was a Senior Research Associate with the Robert W. Galvin Center for Electricity Innovation, IIT. He worked as summer Visiting Faculty at NYISO in 2012. Currently, he is an Associate Professor with the Electrical and Computer Engineering Department, Stevens Institute of Technology, NJ, USA. His research interests include power systems operation and planning, energy economics, and community resilience microgrid.

Jie Li received the B.S. degree in information engineering from Xi'an Jiaotong University in 2003 and the M.S. degree in system engineering from Xi'an Jiaotong University, China in 2006, and the Ph.D. degree from the Illinois Institute of Technology (IIT), Chicago, in 2012. From 2006 to 2008, she was a Research Engineer with IBM China Research Lab. From 2012 to 2013, she was a Power System Application Engineer with GE Energy Consulting. Presently, she is an Assistant Professor in the Electrical and Computer Engineering Department at Clarkson University. Her research interests include green data center, power systems restructuring, and bidding strategy.