Learning-Based Resource Allocation for Data-Intensive and Immersive Tactile Applications

Medhat Elsayed and Melike Erol-Kantarci, *Senior Member, IEEE*

School of Electrical Engineering and Computer Science
University of Ottawa, Ottawa, ON
Emails: {melsa034, melike.erolkantarci}@uottawa.ca

***Abstract*—The immersive tactile applications that are emerging in the entertainment, education and health industries are anticipated to be available for mobile users in the close future. These applications are data-intensive and delay-sensitive due to the nature of information that is being exchanged. With today's mobile networks, the throughput and latency challenges are the major roadblocks for mobile users. In this paper, we propose a resource allocation technique with the aim of increasing throughput and reducing latency of Data Intensive Devices (DIDs). We consider the coexistence of DIDs with traditional User Equipments (UEs) on a two-tier, densely deployed network of Small cell Base Stations (SBSs) and eNBs. We propose a Q-learning-based resource allocation scheme, namely, Throughput Maximizing Q-Learning (TMQ) that learns the efficient resource allocation of both SBSs and eNB. The proposed technique is compared with well-known Proportional Fairness (PF) algorithm in terms of average throughput, delay, and fairness. Simulation results show significant improvement in throughput, 80% reduction in delay, and 6% increase in fairness.**

***Index Terms*—Immersive communications, Resource allocation, Small cell networks, Q-Learning, tactile applications.**

I. Introduction

Immersive communications is concerned with the real-time exchange of natural social signals between people at different locations in a way that mimics face-to-face interactions [1]. This new form of information exchange can be visual or even tactile. Tactile refers to transmission of signals to deliver the feeling of touch, while in general Tactile Internet refers to an extremely reliable, secure and ultra low-latency network. Tactile information can be used to train students and medical staff on clinical skills [2], anatomical evaluation [3], or guide surgeons in intraoperative surgery [4]. In parallel to tactile applications, augmented and virtual reality (AR/VR) applications are emerging to serve health, entertainment and education [5]. Most of these technologies have not been initially developed for mobile users. Only recently, they are being considered for mobile users. However all of the mentioned, immersive tactile applications require a high-throughput and low-latency network performance which is beyond the capacity of the state-of-the-art wireless mobile networks.

In this paper, we consider a small cell network of SBSs and an eNB where a large number of DIDs coexist with UEs. This calls for efficient use of time and frequency resources which is essentially a resource allocation problem. We address resource block allocation using reinforcement learning, more specifically Q-Learning. Q-Learning is a machine learning algorithm, which offers fast and sub-optimal results in model-free environments [6]. Our proposed Throughput Maximizing Q-Learning (TMQ) algorithm performs resource allocation in dense small cell networks with an objective of maximizing throughput and minimizing delay.

In the literature, resource allocation schemes span from traditional Round-Robin (RR) and Proportional Fairness (PF) to recent reinforcement learning based schemes. Authors in [7] provide a Universal Software Radio Peripheral (USRP)-based Q-Learning implementation for femtocell interference mitigation. The algorithm aims to maximize the femtocells aggregate capacity without sacrificing macrocell capacity. In [8], the authors propose a heuristic power and resource block allocation algorithm for haptic communications. The algorithm follows a steepest descent approach to decrease the difference between the uplink and downlink rates. In [9], the authors utilize distributed and centralized Q-Learning algorithm to improve the system performance in ultra-dense heterogeneous cellular network.

On the other hand, low-latency is a critical QoS requirement for tactile applications [10]–[13]. Therefore, in this paper, we address the resource block allocation for network throughput maximization while achieving low packet delay for tactile communications. We develop a throughput-maximization algorithm based on Q-Learning. We show that by combining the two-tier network model and careful design of TMQ reward, the algorithm can improve multiple network metrics simultaneously. In summary, our algorithm increases the throughput notably, achieves 80% reduction in delay, and 6% increase in fairness.

The paper is organized as follows. We first discuss the related work in section II. The network model and problem formulation are presented in Section III. The proposed Q-Learning algorithm and the comparison algorithm are discussed in Section IV. Section V demonstrates the simulation results to show our scheme's performance. And Section VI concludes the paper.

II. Related work

In the literature, several studies have used reinforcement learning techniques for enhancing network performance. In [14], the authors proposed a distributed co-operative Q-Learning algorithm for power allocation with the aim of maximizing femtocells capacity, while guaranteeing a certain macrocell interference level. In [15], the authors employ Q-Learning to perform spectrum and power allocation to improve capacity of D2D communications in cellular networks. The paper [16] addresses the minimization of energy consumption in heterogeneous cellular networks while maintaining the Quality-of-Service (QoS) of the mobile users. To solve the curse of dimensionality problem, the authors propose to use centralized and decentralized Q-Learning algorithms. Meanwhile, in [17], the authors use Q-Learning to improve the femtocells spectral efficiency. The number of subchannels are dynamically adjusted by different frequency reuse factors. Furthermore, the authors in [18] consider the resource allocation problem in LTE-U systems. They perform user association, spectrum allocation, and load balancing using a decentralized expected Q-Learning algorithm.

In [19], the authors propose three variations of the Q-Learning algorithm, one to capture accurate information about the channel, another to consider Signal to Interference plus Noise Ratio (SINR) of different cells, and a third one to jointly consider the behaviour of users and channel conditions to perform spectrum allocation. Results demonstrate the spectrum allocation efficiency and SINR values during transmission. In [20], the authors address the improvement of the Q-Learning slow convergence through the introduction of smart initialization procedure. Hence, they utilize this to perform power control for throughput improvement in LTE femtocell networks. In [21], the authors use non-deterministic Q-Learning scheme to perform the spectrum allocation to secondary users in cognitive radio networks, in which aging can remove the starvation of low-priority users.

Previous works either focus on improving one metric at a time (e.g., spectral efficiency, throughput, etc.), or use the Q-Learning approach with special initialization conditions. In our work, we aim to jointly improve latency and throughput; We do not enforce initialization conditions; And we use a two-tier heterogeneous network where the Q-Learning algorithm works at both tiers. Our results show that network throughput, delay, and fairness can be improved compared to a well-known resource allocation schemes. This is facilitated through careful design of the action-space and reward function as explained in the following sections. As with other learning based approaches Q-learning needs time for exploring. Therefore, the performance is expected to improve over time.

III. Network model

The system model constitutes a two-tier network with one eNode-B (eNB), J Small-cell Base Stations (SBSs), and N_j user devices per SBS (Figure 1). Network users can be classified as either Data-Intensive Devices (DIDs), or User Equipments (UEs). DIDs can be haptics gadgets, AR/VR devices, mobile ultrasound, etc., whereas UEs are conventional users of the mobile network such as smart phones. All nodes comply with LTE standard's Downlink (DL) and Uplink (UL) communication release 12 [22]. According to the standard, each frame consists of 10 subframes of 1 milli-seconds duration. LTE resource grid consists of N_{RB} resource block (RBs), where the RB is a collection of frequency subcarriers and spans two time slot durations (i.e., time slot = 0.5 msec). Each M contiguous RBs are combined to form one RB Group (RBG). The resource allocation process is performed each Time-To-Transmit Interval (TTI) (TTI = one subframe) by allocating RBGs to the covered nodes. On the other hand, we consider uniform power allocation. We utilize the Almost-Blank SubFrame (ABSF) to reduce the cross-tier interference. In particular, each tier performs its uplink transmission in different subframes in an interleaved manner.

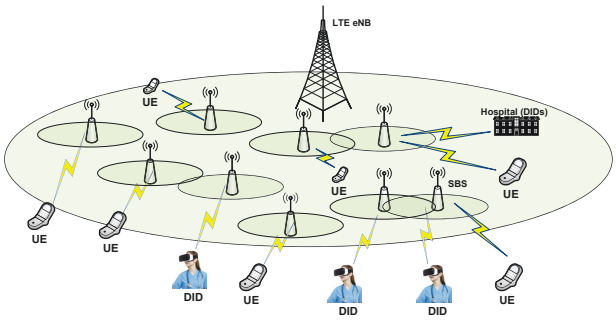

Fig. 1. Data-intensive and tactile application users over small cell wireless networks.

A. Two-tier Allocation in the Small Cell Network

In this paper, resource block allocation is performed to maximize the DIDs throughput. The proposed algorithm is a Q-Learning-based approach, performed on the two-tiers (i.e., on both eNB and SBSs). The algorithm utilizes the Channel Quality Indicator (CQI) and the recent packet rate sent to each Base Station (BS) to update its state and reward function.

B. Problem formulation

The transmission rate of the link between user i connected to BS j (i.e., link (i, j)) can be formulated as:

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

$$C_{i,j,k} = W_k\, log_2\Big(1+\frac{P_{i,j,k}\, g_{i,j,k}}{W_k\, N_0 \;+ \sum\limits_{\substack{m\neq i \\ m\in N_J}} x_{m,j,k}\, P_{m,i,k}\, g_{m,i,k}}\Big) \tag{2}$$

Here, $R_{i,j}$ denotes the total rate of i^{th} user attached to j^{th} BS. $x_{i,j,k}$ is a binary variable and it is 1 if k^{th} RB is allocated to i^{th} user that is attached to j^{th} BS, otherwise 0. $C_{i,j,k}$ denotes the user rate on RB k. W_k is the bandwidth of RB k. N_0 is the Additive White Gaussian Noise (AWGN) single-sided power spectral density. $P_{i,j,k}$ is the transmission

power. $g_{i,j,k}$ is the channel coefficient on link (i,j) at RB k. $P_{m,j,k}$ and $g_{m,i,k}$ are the transmission power and the channel coefficient of interfering user m, respectively.

Resource allocation can be formulated as an optimization problem aiming to maximize network rate:

$$\underset{x_{i,j,k}}{\text{Maximize}} \quad z = \sum_{j=1}^{J} \sum_{i=1}^{N_j} R_{i,j} \tag{3}$$

However, optimization-based centralized approaches do not scale well with the dynamic nature of the network. Therefore, we propose to use a Q-Learning-based approach. Q-Learning has the potential to reduce the computational complexity and improve the convergence speed. Moreover, the proposed algorithm is decentralized and independent. In other words, it can achieve the desired performance, in terms of throughput and delay, without having to share Q-Learning information between the active nodes. In the following section, we present the internal design of the algorithm and highlight its main features.

IV. Throughput-Maximizing Resource Allocation using Q-Learning (TMQ)

TMQ utilizes the Q-Learning algorithm, which is a Reinforcement Learning algorithm that aims to achieve a sub-optimal decision policy by selecting the actions having the maximum expected reward. In our model, we apply the TMQ on both network tiers of our network model (i.e., eNB, and SBSs). The eNB performs TMQ to allocate RBs to its attached SBSs, while SBSs perform it to allocate RBs to its attached users. Each tier agent can estimate the link quality by utilizing the CSI feedback from its users.

To facilitate a distributed and self-organizing approach, a multi-agent scenario is considered, in which agents are eNB and SBSs. The actions are the set containing all the RB allocation possibilities to active users ($a_{i,j,t}$). Hence, $a_{i,j,t}$ is the action of user i attached to BS j at time (TTI) t. Obviously, this leads to a curse of dimensionality as the action-space dimension will increase dramatically with the number of users. For instance, a SBS covering 10 users and performing allocation using 50 RBs will have 10^{50} actions to choose among. Instead, allocation can be performed in a group of contiguous RBs, namely Resource Block Groups (RBG). In this paper, we use a RBG of size 10 RBs. This significantly reduces the action-space to become $N^{(K/10)}$, where N is the number of users and K is the number of RBs. This improves the convergence of the algorithm significantly.

In TQM, the reward function aims at maximizing the rates of both DIDs and UEs. The priority given to traffic types of devices are managed by Ψ_{DID} and Ψ_{UE} which are defined as follows:

$$\Psi_{DID} = \left(\frac{2}{\pi}\right) \arctan(R_{DID}) \tag{4}$$

$$\Psi_{UE} = \left(\frac{2}{\pi}\right) \arctan(R_{UE}) \tag{5}$$

where, R_{DID}, and R_{UE} are the peak rates for DIDs and UEs, respectively. Then, the reward function can be defined as follows:

$$RC_j(a_{j,t}) = \beta\ \Psi_{DID} + (1-\beta)\ \Psi_{UE} \tag{6}$$

where $RC_j(a_{j,t})$ is the reward function of BS j for the action $a_{j,t}$ of its covered users. $a_{j,t}$ represents the RB allocation to users attached to BS j at TTI t. β is a scalar weight to control the priority among devices and their traffic types. Thus, the reward function aims at maximizing both the DID and UE rates while giving higher priority to the critical load by increasing the parameter β.

The Q-Value is updated using Bellman's equation as follows [23]:

$$Q(a_{j,t}) = (1-\alpha)Q(a_{j,t}) + \alpha[RC(a_{j,t}) + \gamma \max_{a_{j,t}} Q(a_{j,t})] \tag{7}$$

where α represents the learning rate and γ is a discount factor that determines the importance of future rewards.

The Q-Learning policy is to select the action that maximizes the Q-value as in eq. 8:

$$\pi_{j,t} = arg \max_{a_{j,t}} Q(a_{j,t}) \tag{8}$$

Figure 2 presents the flow-chart of the TMQ algorithm performed by each BS j. To account for action exploration, Q-Learning utilizes the ϵ-greedy algorithm to select actions randomly with probability (ϵ) (Exploration), or using policy in eq. 8 with probability ($1-\epsilon$) (Exploitation). Afterwards, the agent observes the new reward and state, it updates the Q-value, and reports the selected actions to the users so that they perform UL/DL communications. Last, the throughput is estimated by the agent to be used in the next iteration.

Performance comparison is conducted between our algorithm and a well-known algorithm from the literature: Proportional Fairness (PF). PF algorithm allocates the RBs to the users having a maximum relative channel conditions, with an intent to have fairness on the long-run [24].

V. Performance Evaluation

Our simulations are performed using the Matlab LTE toolbox. Our settings incorporate one eNB with a radius of 800 meters, covering 10 SBSs, each with 50 meters radius [25]. In all simulation figures, a fixed number of 5 UEs per SBS is considered, while number of DIDs is changed from 4 to 12 with a step of 2. The 3GPP pathloss model is used [26]: $PL_{3GPP} = 128.1 + 37.6 * log(d_{Km})$, where d_{Km} is the distance in Km between the BS and the user. Shadowing is drawn from a log-normal distribution of zero-mean and 8 dB variance while penetration loss is set to 20 dB [27], and the receiver noise is set to 5 dB. TMQ learning rate α is 0.5, discount factor γ is 0.9 [28], and exploration probability (ϵ) of 0.2. We consider two traffic types. For devices running tactile applications we adopt the Beta distribution as defined in 3GPP for Machine-Type Communications (MTC) [29], and

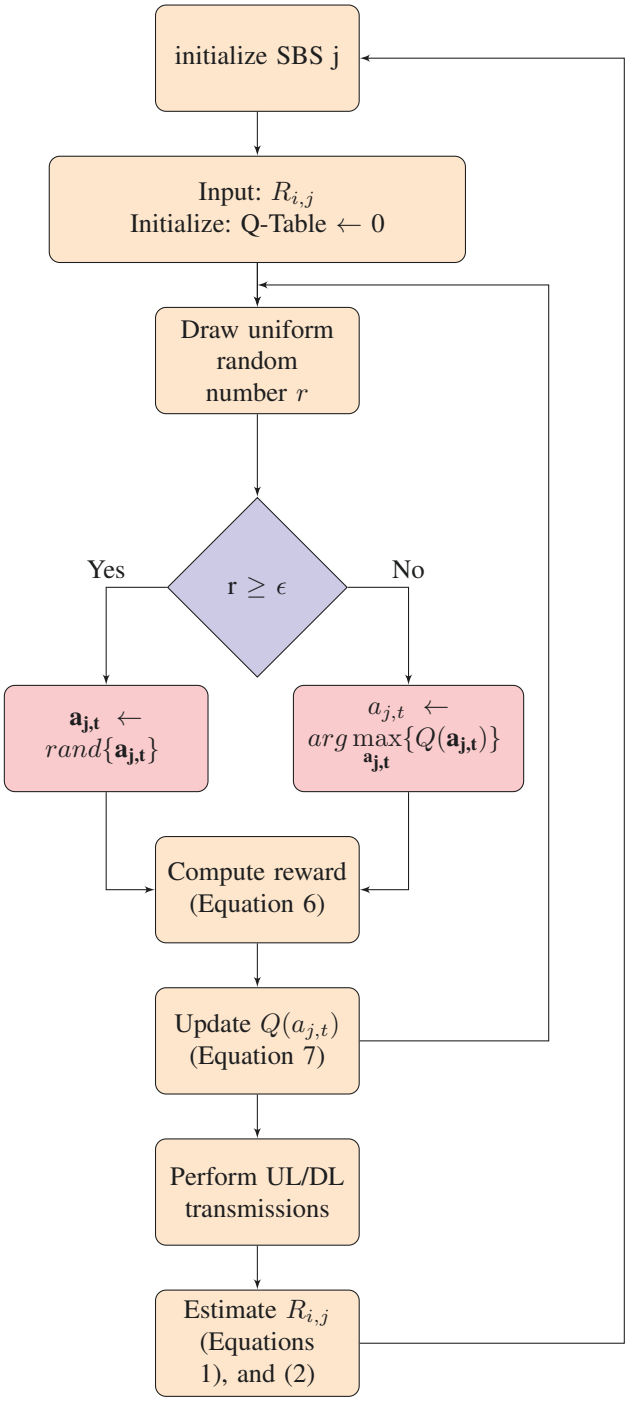

Fig. 2. TMQ algorithm running on each SBS j (also works for eNB)

the traditional user traffic is modeled as Poisson distribution with interarrival time 5ms. Simulations are performed and averaged for 5 runs. Table I summarizes the simulation parameters.

The performance is analyzed in terms of average throughput, average packet delay, average queuing delay, and fairness. We define delay as the total transmission delay from a user to an eNB, starting from the packet generation time. The queue waiting delay is the aggregate waiting time the

TABLE I
SIMULATION SETTINGS

Number of eNBs	1
Number of SBSs per eNB	10
Number of DIDs per SBS	4:2:10
Number of UEs per SBS	5
eNB radius	800 m
SBS radius	50 m
Min distance between SBSs	30 m
Traffic arrival model	DIDs: Beta [29]
	UEs: Poisson
Packet mean Inter-arrival time	5 milli-seconds
Packet size	Exponential (mean = 25 Bytes)
Modulation Schemes	QPSK, 16-QAM, 64-QAM
Bandwidth	10 MHz
Number of RBs	50
Resource Block Groups	10
eNB power transmit	46 dBm
SBS power transmit	20 dBm
Pathloss model	3GPP
	$PL_{dB} = 128.1 + 37.6 * log(d)$
Penetration loss	20 dB
Noise Figure	5 dB
Shadowing	$\sim$ LOGN(0, 8 dB)
Simulation time	500 TTIs
α (Learning rate)	0.5
γ (Discount factor)	0.9
β (Priority weight of DIDs)	0.9
ϵ (Exploration probability)	0.2
Confidence Interval	95%

packet experiences throughout its transmission (i.e., waiting in the device queue, and waiting in SBS queue).

Figures 3 and 4 present the average and peak throughput versus number of DIDs. Figures 3a and 4a shows the DID throughput. It can be seen that TMQ outperforms PF both for average and peak throughput. Meanwhile, TMQ improves DIDs throughput without compromising the UEs throughput. As seen in Figures 3b and 4b, UE throughput is also higher than the case with PF.

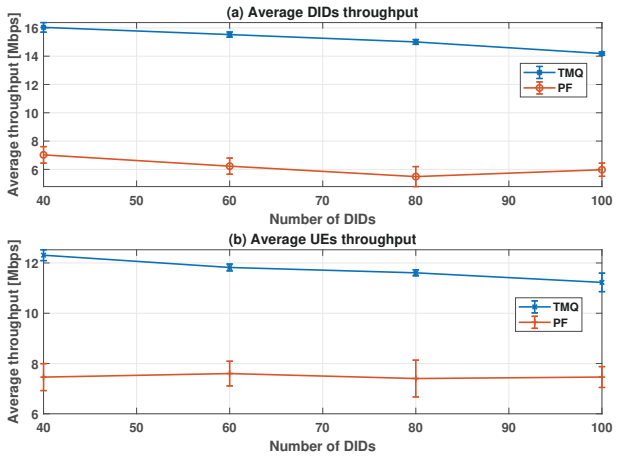

Fig. 3. Average throughput for (a) DIDs and (b) UEs (10 SBS, 5 UEs per SBS)

Figure 5 presents the average packet delay in milli-seconds versus the number of DIDs. As seen from the figure, TMQ

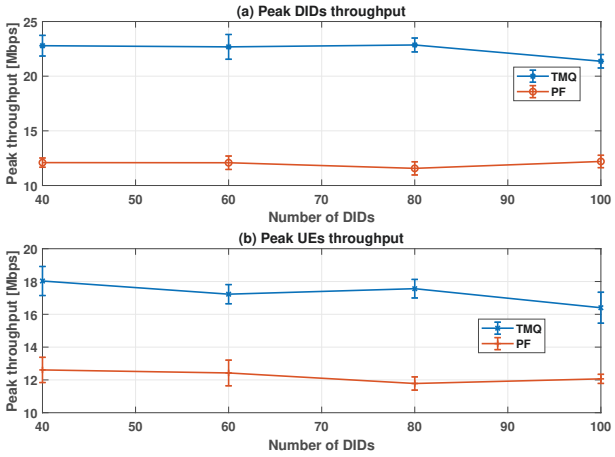

Fig. 4. Max-user throughput for (a) DIDs and (b) UEs (10 SBS, 5 UEs per SBS)

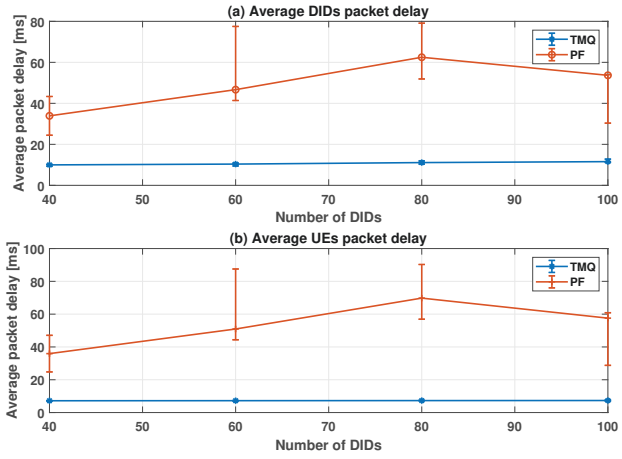

Fig. 5. Average packet delay [ms] for (a) DIDs and (b) UEs (10 SBS, 5 UEs per SBS)

achieves the lowest total packet delay. On the other hand, Figure 6 shows the average queuing delay experienced by both algorithms. The waiting time is a direct outcome of the scheduling time, where it constitutes the time the user waits for getting a RB allocation from its BS. This result reflects that most of the packet delay comes from the scheduling time, which was significantly improved using the TMQ algorithm. As seen from the figures, TMQ achieves 80% decrease in delay for the highly-dense scenario (i.e., number of DIDs = 100). In addition, both algorithms do not incur any outage. It is worth noting that the achieved delay is still higher than QoS requirements of tactile applications. The reason is that the minimum scheduling unit in LTE networks is one TTI, which is 1 msec duration. Therefore, the achieved delay can be reasonable, particularly when adding signaling delay. In our future work, we plan to use flexible slot durations fo fifth-generation networks to combat this constraint.

To study fairness of TMQ, Jain's fairness index is plotted in Figure 7. As the figure shows, TMQ achieves better results

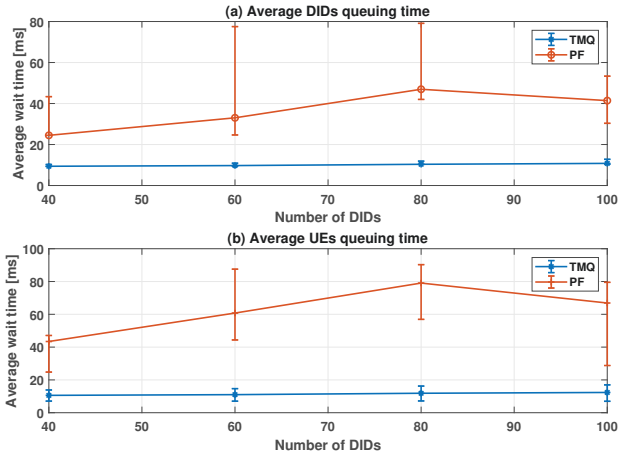

Fig. 6. Average queuing time [ms] for (a) DIDs and (b) UEs (10 SBS, 5 UEs per SBS)

that PF, which is a direct product of applying reward function that maintains fairness between DIDs and UEs.

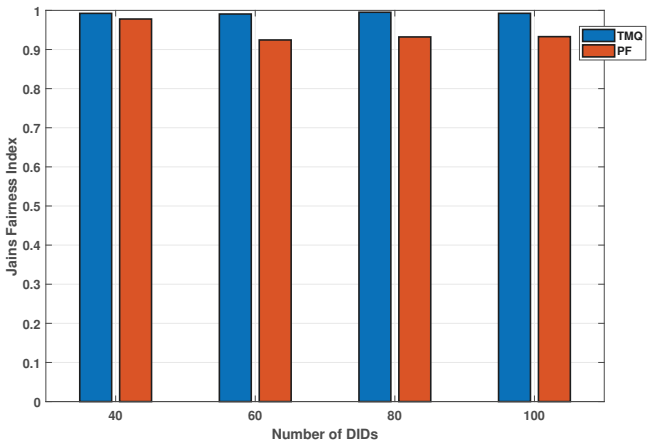

Fig. 7. Jain's Fairness Index (JFI) (10 SBS, 5 UEs per SBS)

Finally, the results presented here have converged after 200 TTIs (i.e., 200 msec). For tactile applications, this is a high convergence time, however note that this only happens for network initialization. In other words, if network dynamics changes after the convergence, the TMQ will need less number of iterations to reach the new decision due to the nature of exploration and exploitation of Q-learning algorithms.

VI. Conclusion

In this paper, we proposed a Throughput-Maximization Q-Learning (TMQ) algorithm to provide high-throughput to mobile Data Intensive Devices (DIDs) that can run tactile applications. The number of DIDs will increase in the future mobile networks with the emerging AR/VR and tactile applications. The proposed TMQ algorithm is based on Q-Learning and it aims to maximize throughput of DIDs. TMQ is a distributed algorithm running on both eNB and SBSs.

Performance results are compared to both proportional fairness (PF) algorithm in terms of throughput, delay and fairness. TMQ is shown to have the higher throughput and fairness and the lower delay than PF. In our on-going work, we are aiming to improve training time for the Q-learning approach.

VII. Acknowledgement

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] P. A. Chou, "Advances in Immersive Communication: (1) Telephone, (2) Television, (3) Teleportation," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 9, p. 41:141:4, 10 2013.

[2] S. Persky, "Employing immersive virtual environments for innovative experiments in health care communication," *Patient Education and Counseling*, vol. 82, pp. 313–317, 3 2011.

[3] W. S. Khor, B. Baker, K. Amin, A. Chan, K. Patel, and J. Wong, "Augmented and virtual reality in surgery-the digital surgical environment: applications, limitations and legal pitfalls.," *Annals of translational medicine*, vol. 4, p. 454, 12 2016.

[4] L.-M. Su, B. P. Vagvolgyi, R. Agarwal, C. E. Reiley, R. H. Taylor, and G. D. Hager, "Augmented Reality During Robot-assisted Laparoscopic Partial Nephrectomy: Toward Real-Time 3D-CT to Stereoscopic Video Registration," *Urology*, vol. 73, pp. 896–900, 4 2009.

[5] M. Erol-Kantarci and S. Sukhmani, "Caching and Computing at the Edge for Mobile Augmented Reality and Virtual Reality in 5G," *Proc. of ADHOCNETS*, 2017.

[6] O. Sigaud and F. Garcia, "Reinforcement Learning," in *Markov Decision Processes in Artificial Intelligence*, pp. 39–66, Hoboken, NJ USA: John Wiley & Sons, Inc., 3 2013.

[7] M. H. M. Elsayed and A. Mohamed, "Distributed interference management using Q-Learning in cognitive femtocell networks: New USRP-based implementation," in *2015 7th International Conference on New Technologies, Mobility and Security (NTMS)*, pp. 1–5, IEEE, 7 2015.

[8] A. Aijaz, "Towards 5G-enabled Tactile Internet: Radio resource allocation for haptic communications," in *2016 IEEE Wireless Communications and Networking Conference*, pp. 1–6, IEEE, 4 2016.

[9] 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.

[10] A. Aijaz, M. Dohler, A. H. Aghvami, V. Friderikos, and M. Frodigh, "Realizing the tactile internet: Haptic communications over next generation 5g cellular networks," *IEEE Wireless Communications*, vol. 24, pp. 82–89, April 2017.

[11] M. Simsek, A. Aijaz, M. Dohler, J. Sachs, and G. Fettweis, "5g-enabled tactile internet," *IEEE Journal on Selected Areas in Communications*, vol. 34, pp. 460–473, March 2016.

[12] M. A. Lema, A. Laya, T. Mahmoodi, M. Cuevas, J. Sachs, J. Markendahl, and M. Dohler, "Business case and technology analysis for 5g low latency applications," *IEEE Access*, vol. 5, pp. 5917–5935, 2017.

[13] M. Dohler, T. Mahmoodi, M. A. Lema, M. Condoluci, F. Sardis, K. Antonakoglou, and H. Aghvami, "Internet of skills, where robotics meets ai, 5g and the tactile internet," in *2017 European Conference on Networks and Communications (EuCNC)*, pp. 1–5, June 2017.

[14] H. Saad, A. Mohamed, and T. ElBatt, "Distributed cooperative Q-Learning for power allocation in cognitive femtocell networks," in *2012 IEEE Vehicular Technology Conference (VTC Fall)*, pp. 1–5, Sept 2012.

[15] 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.

[16] X. Chen, C. Wu, Y. Zhou, and H. Zhang, "A learning approach for traffic offloading in stochastic heterogeneous cellular networks," in *2015 IEEE International Conference on Communications (ICC)*, pp. 3347–3351, June 2015.

[17] X. Ji, Z. Qi, and Z. Su, "Spectrum allocation based on q-learning algorithm in femtocell networks," in *2012 IEEE International Conference on Computer Science and Automation Engineering (CSAE)*, vol. 1, pp. 381–385, May 2012.

[18] 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.

[19] L. R. Faganello, R. Kunst, C. B. Both, L. Z. Granville, and J. Rochol, "Improving reinforcement learning algorithms for dynamic spectrum allocation in cognitive sensor networks," in *2013 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 35–40, April 2013.

[20] M. Simsek, A. Czylwik, A. Galindo-Serrano, and L. Giupponi, "Improved decentralized q-learning algorithm for interference reduction in lte-femtocells," in *2011 Wireless Advanced*, pp. 138–143, June 2011.

[21] S. Bhattacharjee, A. Bhar, and R. Saha, "Channel allocation in a cognitive radio network using non deterministic q learning algorithm," in *2012 Third International Conference on Emerging Applications of Information Technology*, pp. 327–330, Nov 2012.

[22] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA); Medium Access Control (MAC) protocol specification," Technical Specification (TS) 136.321, 3rd Generation Partnership Project (3GPP), 04 2015. Version 12.5.0.

[23] E. Alpaydin, *Introduction to machine learning*. MIT Press, 2010.

[24] A. Ghosh, J. Zhang, J. G. Andrews, and R. Muhamed, *Fundamentals of LTE*. Upper Saddle River, NJ, USA: Prentice Hall Press, 1st ed., 2010.

[25] M. A. Imran, E. Katranaras, M. Dianati, and A. Saeed, "Dynamic femtocell resource allocation for managing inter-tier interference in downlink of heterogeneous networks," *IET Communications*, vol. 10, pp. 641–650, 4 2016.

[26] 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 2011. Version 9.0.0.

[27] C. C. Coskun and E. Ayanoglu, "Energy-Spectral Efficient Resource Allocation Algorithm for Heterogeneous Networks," *IEEE Transactions on Vehicular Technology*, pp. 1–1, 2017.

[28] 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, IEEE, 7 2017.

[29] 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.