Deep Learning-based Multi-Connectivity Optimization in Cellular Networks

J. J. Hernández-Carlón, J. Pérez-Romero, O.Sallent, I. Vilà, F. Casadevall
Signal Theory and Comunications Department of Universitat Politècnica de Catalunya (UPC)
Barcelona, Spain
[juan.jesus.hernandez, jordi.perez-romero, irene.vila.munoz]@upc.edu, [sallent, ferranc]@tsc.upc.edu

***Abstract*— Multi-connectivity emerges as a useful feature to handle the traffic in heterogeneous cellular scenarios and fulfill the demanding requirements in terms of data rate and reliability. It allows a device to be simultaneously connected to multiple cells belonging to different radio access network nodes from a single or multiple radio access technologies. This paper addresses the problem of optimally splitting the traffic among cells when multi-connectivity is used. For this purpose, it proposes the use of deep learning to determine the optimum amount of traffic of a device that needs to be sent through one or another cell depending on the current traffic and radio conditions. Obtained results reveal a promising capability of the proposed Deep Q Network solution to select quasi optimum traffic splits in the considered scenario.**

Keywords—Multi-connectivity, deep learning, Deep Q Network, Heterogeneous Networks

I. Introduction

Mobile Network Operators (MNO) are aggressively trialing and rolling-out the fifth generation (5G) networks, intending to support new vertical-driven use cases and enhanced user experiences leveraging 5G performance enhancements [1]. These 5G deployments further increase heterogeneity with different cell types (e.g. macrocells, indoor and outdoor small cells) based on multiple radio access technologies (RATs) (e.g., 2G, 3G, 4G and 5G New Radio (5G NR)), operating in different spectrum bands (e.g. sub 6 GHz bands used by all RATs and millimeter wave (mmW) bands used by 5G New Radio).

In this context, multi-connectivity (MC) arises as a vital technology that will support simultaneous access via LTE and 5G networks [2]. The basic principle of MC is that a User Equipment (UE) has simultaneous connectivity to multiple nodes of the Radio Access Network (RAN), e.g. eNodeBs (eNB) operating with LTE and/or gNodeBs (gNB) operating with 5G NR [3]. There is one master node (MN) responsible for the radio-access control plane and one, or in the general case multiple, secondary node(s) (SN) that provide additional user-plane links. In this way, a UE can aggregate the radio resources from multiple eNBs/gNBs, which provides an efficient way to fulfil the 5G requirements of high data rate and ultra-reliability. In the context of the Third Generation Partnership Project (3GPP), MC is specified through the Multi-Radio Dual Connectivity (MR-DC) feature defined in [4] that considers different options depending on the technology used by the MN and by the SN and on the core network technology (i.e. 5G core or Evolved Packet Core (EPC)). Moreover, currently, there are some 3GPP Rel-17 study items that address the MR-DC with multiple cells operating in different bands (see e.g. [5]). The operation of MR-DC relies on the use of three different types of radio bearers [4], namely the Master Cell Group (MCG) bearers in which data is transmitted through the MN, the Secondary Cell Group (SCG) bearers in which data is transmitted through the SN, and the Split bearers, in which data is split between the SN and the MN at the Packet Data Convergence Protocol (PDCP) layer of the radio interface protocol stack.

The literature has considered different problems in relation to MC, such as the resource allocation in [6][7] or the traffic split [8][9][10][11]. Concerning resource allocation, a Smart Aggregated RAT Access (SARA) strategy is proposed in [6] for joint RAT selection and resource allocation in a scenario with cellular base stations and WiFi access points. The solution makes use of a Semi Markov Decision Process (SMDP)-based hierarchical decision framework (HDF). In [7] the optimization problem of resource allocation in a MC scenario with 5G and LTE is formulated. Then, a solution based on two heuristic algorithms is proposed, namely a base station selection algorithm performed by the UE and a resource block algorithm executed by the base station. The problem of traffic split between different RATs is considered in [8] focusing on an LTE/Wi-Fi scenario and assuming a single coordinating node that decides on the best choice of RAT for all users, and advices on the actual amount of radio resources that every user may utilize on each technology. The problem is formulated analytically and a solution based on the weighted max-min algorithm is proposed. In [9] the problem of traffic splitting between the master and the secondary eNB in LTE with dual connectivity is modelled as a Constrained Markov Decision Process and a solution based on the Lagrangian approach is proposed. Similarly, [10] considers a scenario with 5G-LTE multi-connectivity and makes use of Lagrangian Dual Decomposition to determine the fraction of traffic transmitted through each cell to maximize the goodput. In turn, [11] formulates a PDCP split bearer decision problem that decides whether and how to split the traffic across multiple cells in order to meet the bandwidth requirements of user services and proposes a heuristic solution to solve the problem.

With all the above, this paper addresses the traffic split multi-connectivity problem in multi RAT scenarios. The target is to determine a policy to optimally distribute the traffic of a UE across the different RATs and cells by fulfilling the QoS requirements while minimising the resource consumption of the UE and ensuring that no overload situations arise in the involved cells. For this purpose, the paper relies on the use of Deep Reinforcement Learning (DRL), and in particular on the Deep Q Network (DQN) technique [12], in order to learn the traffic split policy to be applied on a per UE basis. DRL techniques are useful for optimizing dynamic decision-making problems in the absence of an accurate mathematical model of the operational environment. Moreover, thanks to their capability of generalizing from past experience, DRL techniques are efficient in problems that depend on a large number of input variables and in which both the inputs and the decision making outputs can take a large range of possible values, as it is the case of the MC problem considered here. To the authors' best knowledge, the use of DQN has not been considered yet by previous works in the context of MC, and therefore this constitutes the main novelty of this paper.

The rest of the paper is organised as follows. Section II presents the system model and formulates the considered multi-connectivity problem. The proposed DQN-based solution is presented in Section III and different performance results are provided in Section IV. Finally, Section V summarises the conclusions.

II. System Model and Problem Definition

Let us consider a HetNet scenario where different UEs with multi-connectivity capabilities are camping. A given UE u is able to get connected up to a maximum of M different RATs and N different cells per RAT. Let us denote as $A_u=\{C_{m,n}\}$ the set of cells detected by the UE u, which constitute the candidate cells to which the UE can be connected to. $C_{m,n}$ denotes the n-th cell of the m-th RAT with n=1, ..., N and m=1,...,M. It is worth mentioning that, due to the mobility of the UE, the specific cells that the UE detects in a given RAT may change with time. In this respect, it is assumed that the N cells of a RAT correspond to the best N cells detected by the UE at a certain time based on measurements averaged during a time window ΔT.

Through the use of multi-connectivity the traffic of the u-th UE is split across multiple RATs/cells of the set A_u. The multi-connectivity configuration for the u-th UE can be expressed as the $M \times N$ matrix $\mathbf{B}=\{\beta_{m,n}\}$ where $\beta_{m,n}\in[0,1]$ defines the fraction of total traffic of UE u that is delivered through the n-th cell of the m-th RAT. Then, the objective is to find the policy that determines the optimal configuration $\mathbf{B}=\{\beta_{n,m}\}$ to be applied in a time window of ΔT s that allow ensuring the QoS requirements with minimum resource consumption and avoiding overload situations in the different RATs/cells. In this respect, it is assumed that the Quality of Service (QoS) requirements of the user u are expressed in terms of a minimum bit rate R_u (b/s) to be provided to the user.

To formalize the problem, let us denote as $T_u(\mathbf{B})$ the total throughput obtained by user u during the last time window period ΔT with the multi-connectivity configuration $\mathbf{B}$. Let also denote $\phi_{m,n}(\beta_{m,n})$ as the fraction of the available resources in the m-th cell and n-th RAT that is allocated to the u-th UE to transmit the traffic corresponding to $\beta_{m,n}$. This fraction will be measured differently depending on the specific RAT. For example, for 4G/5G RATs this can be the fraction of Physical Resource Blocks (PRBs) allocated to the UE while for a Wi-Fi RAT this can be the fraction of allocated airtime. In turn, the total fraction of occupied resources in a RAT/cell accounting for all the UEs connected to that cell is denoted as $\rho_{m,n}(\beta_{m,n})$.

Then, the considered problem to be solved for the u-th UE is formally given as:

$$\begin{aligned} &\mathbf{B} = \arg\min_{\mathbf{B}} \sum_{m=1}^{M}\sum_{n=1}^{N}\phi_{m,n}\left(\beta_{m,n}\right) \\ &\text{s.t.}\quad T_u\left(\mathbf{B}\right)\geq R_u\ ,\quad \rho_{m,n}\left(\beta_{m,n}\right)\leq\rho_{\max}\quad \forall m,n \\ &\qquad \sum_{m=1}^{M}\sum_{n=1}^{N}\beta_{m,n}=1 \end{aligned} \tag{1}$$

where $\rho_{max}\in(0,1]$ is the maximum threshold established to avoid overload situations in a cell.

Fig. 1 depicts the architectural components to enforce in the network the multi-connectivity configuration $\mathbf{B}$ obtained as a result of the above problem. The figure illustrates an example for the downlink traffic transmitted to a UE served by two cells of RAT m=1 (e.g. 5G). The cell n=1 is at the MN while the cell n=2 is at the SN. The traffic between these cells is split at the PDCP layer of the MN using dual connectivity feature. The multi-connectivity configuration is determined by an MC controller that takes as an input different measurement from the RATs/cells as it will be explained in Section III.A. As an implementation example, the MC controller could be hosted at the so-called non-real time RAN Intelligent Controller (non-RT RIC) in the O-RAN Alliance [13] reference architecture. The non-RT RIC is associated to the service management and orchestration and supports intelligent RAN optimization with operation time scales larger than 1s. The output of the MC controller is the configuration $\mathbf{B}=\{\beta_{m,n}\}$ with the weights $\beta_{m,n}$ to be configured at the PDCP layer of the MN to split the traffic between cells 1 and 2. Finally, the MAC scheduler in a 5G NR or LTE cell will allocate the necessary amount of resources $\phi_{m,n}(\beta_{m,n})$ to the UE to transmit the fraction of traffic $\beta_{m,n}$ corresponding to the cell. The specific design of the per-RAT resource allocation mechanisms is out of the scope of this work, but in general it will take into account aspects such as the propagation and interference conditions observed by the UE, the QoS requirements, the amount of UEs in the cell, etc.

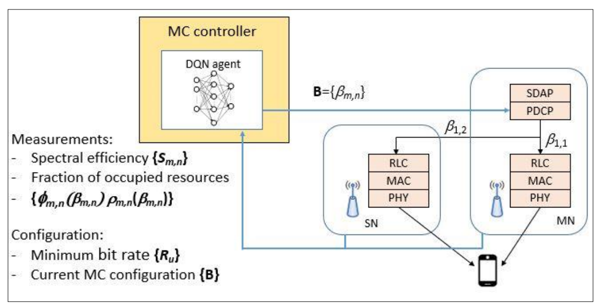

Fig. 1. Architectural components of the considered approach

III. DQN-based Solution

The solution of the problem defined by (1) depends on a large number of variables, including the propagation conditions and interference experienced by the UE in the links with the different cells/RATs, the existing load in each cell, the QoS metrics, etc. Moreover, it also depends on the behavior of the MAC layer in each cell/RAT that determines the amount of resources allocated to the UE. However, since the MC controller operates at the management systems on top of the different RATs, in general it will not have a precise model of how these resource allocation techniques work to determine the function

$\phi_{m,n}(\beta_{m,n})$ and its impact on the QoS metrics. Based on these considerations and on the capabilities of the DRL techniques discussed in Section I, this paper proposes the use of DQN in order to address the considered problem.

In the proposed DQN approach, the learning process is conducted dynamically by a DQN agent at the MC controller that makes decisions for the different UEs. The agent operates in discrete times with granularity equal to the time window duration ΔT. These discrete times are denoted as t, t+1, ..., t+k,... At time t the DQN selects an action $\boldsymbol{a(t)}$ that contains the MC configuration to be applied for a given UE in the next time window. The action selection is based on the current state at time t, denoted as $\boldsymbol{s(t)}$ and on the decision-making policy available at this time. Then, as a result of applying the selected MC configuration a reward signal $r(t+1)$ is provided to the DQN agent at the end of the time window. This reward signal measures how good or bad was the last performed action and therefore it is used to improve the decision making policy. The different components of this process are detailed in the following.

A. State, action and reward specification

The state $\boldsymbol{s}(t)$ is a vector that includes the following components for a given UE u:

- Requirements of UE u: R_u
- Spectral efficiency per RAT/cell $\{S_{m,n}\}$ of UE u
- Current configuration $\mathbf{B}=\{\beta_{m,n}\}$, which corresponds to the configuration applied at time t-1.
- Fraction of occupied resources by the UE u in each RAT/cell $\{\phi_{m,n}(\beta_{m,n})\}$
- Fraction of total occupied resources in each RAT/cell $\{\rho_{m,n}(\beta_{m,n})\}$

All the values $S_{m,n}$, $\phi_{m,n}(\beta_{m,n})$ and $\rho_{m,n}(\beta_{m,n})$ are average values measured during the last time window of duration ΔT, i.e. between discrete times t-1 and t. Notice that the state has a total of $1+4\cdot N\cdot M$ components.

Each action $\boldsymbol{a(t)}\in\mathcal{A}$ represents a matrix $\mathbf{B}=\{\beta_{m,n}\}$ that corresponds to the MC configuration to be applied during the next time window ΔT. The action space $\mathcal{A}$ includes all the MC configurations and is defined considering that the possible $\beta_{m,n}$ values are discretized with granularity $\Delta\beta$ and the aggregate of all $\beta_{m,n}$ values in matrix $\mathbf{B}$ equals 1.

The reward $r(t+1)$ intends to measure how good or bad was the performance obtained by the last action $\boldsymbol{a(t)}$ for the state $\boldsymbol{s}(t)$ in relation to the target of the optimization. Then, considering the optimization problem (1), and that the last action $\boldsymbol{a(t)}$ is given by MC configuration $\mathbf{B}=\{\beta_{m,n}\}$, the reward is defined as:

$$r(t+1)=\left(1-\frac{1}{M\cdot N}\sum_{m=1}^{M}\sum_{n=1}^{N}\phi_{m,n}(\beta_{m,n})\right)\cdot\min\left(1,\frac{T_u(\mathbf{B})}{R_u}\right)\cdot \cdot\prod_{\substack{m,n\\ \beta_{m,n}>0}}\min\left(1,\frac{\rho_{\max}}{\rho_{m,n}(\beta_{m,n})}\right) \tag{2}$$

The first term in $r(t+1)$ captures the total resources assigned to the UE u in all the cells/RATs, so the lower the amount of resources assigned the higher will be the reward. The second term represents a penalty introduced when the achieved throughput $T_u(\mathbf{B})$ is lower than the minimum requirement R_u, The last term introduces a penalty for each cell/RAT in which the UE has transmitted traffic (i.e. $\beta_{m,n}>0$) and the cell is overloaded. The fractions of used resources $\phi_{m,n}(\beta_{m,n})$ and $\rho_{m,n}(\beta_{m,n})$ as well as the total throughput $T_u(\mathbf{B})$ correspond to the average values obtained during the time window ΔT between discrete times t and t+1.

B. Policy learning process

The DQN agent dynamically learns the decision-making policy π used to select the different actions based on the rewards obtained from previous decisions. This is done by means of the DQN algorithm of [12] particularised to the state, action and reward signals presented above. In summary, the algorithm aims at finding the optimal policy that maximises the discounted cumulative future reward by approximating the optimum action-value function with an evaluation deep neural network (DNN) denoted as $Q(\boldsymbol{s},\boldsymbol{a},\theta)$, where $\boldsymbol{s}$ is the observed state, $\boldsymbol{a}$ is one of the possible actions that can be selected and θ are the weights of the interconnections between the different neurons. Given the evaluation DNN, the decision making policy consists in selecting the action $\boldsymbol{a}$ with the highest value of $Q(\boldsymbol{s},\boldsymbol{a},\theta)$ for a given state.

The decision-making policy is updated progressively by modifying the weights θ based on the experiences gathered by the DQN agent. For this purpose, at a certain time t the DQN agent observes the state of the environment $\boldsymbol{s}(t)$ for a given UE and it triggers an action by selecting with probability 1- ε the action $\boldsymbol{a}(t)$ with the highest value of $Q(\boldsymbol{s},\boldsymbol{a},\theta)$ and with probability ε a random action. As a result, the DQN agent gathers the obtained reward and the new state at time t+1 and stores this information (i.e. $\boldsymbol{s}(t)$, $\boldsymbol{a}(t)$, $r(t+1)$, $\boldsymbol{s}(t+1)$) in an experience dataset. The information collected in this dataset is then used to update the weights θ of the evaluation DNN using the expressions detailed in [12].

IV. Performance Evaluation

This section evaluates the performance of the proposed solution by means of system level simulations.

A. Scenario description

The considered scenario is a square area of 500 m x 500 m composed by four 5G NR cells and two LTE cells. The relevant parameters of the cells are presented in Table I. The scenario assumes a non-homogeneous traffic distribution with UEs that support MC and other UEs that generate additional background traffic. The UEs that support MC follow specific trajectories moving at 1m/s along the scenario and have an active session during the whole simulation duration with a required bit rate R_u=50 Mb/s. These UEs can connect to up to M=2 RATs and N=1 cell per RAT. The background traffic generation assumes Poisson session arrivals with aggregate generation rate 0.6 sessions/s and exponentially distributed session duration with average 120s. A background UE remains static during a session. 50% of the background UEs are randomly located inside a square hotspot of 250 m x 250 m centred at the middle of the scenario. The rest of background UEs are randomly distributed in the whole scenario. Background UEs connect to the RAT/cell with the highest Signal to Interference and Noise Ratio (SINR).

To capture the different bit rates achievable in the two technologies, when a background UE is connected to LTE, its serving cell allocates the needed resource blocks to achieve a bit rate of 2.5 Mb/s, and when it is connected to 5G NR, the allocation is to achieve a bit rate of 40 Mb/s.

The DQN model parameters are detailed in Table II. The DQN model has been developed in Python using the *TF-agents* library [14]. The presented results correspond to the performance obtained by the DQN algorithm with the MC configuration policy learnt by the DQN agent after a total of 1E6 policy updates according to the procedure of Section III.B.

TABLE I. CELL CONFIGURATION PARAMETERS

Parameter	Value	
Cells position [x, y] m	[62, 250] [437,250]	[187, 125] [187,375] [312,125] [312,375]
Type of RAT	LTE	5G-NR
Frequency	2100 MHz	26 GHz
Channel bandwidth	20 MHz	50 MHz
Base station transmitted power	49 dBm	21 dBm
Base station antenna gain	5 dB	26 dB
Base station height	25 m	10 m
UE antenna gain	5 dB	10 dB
Number of available PRBs	100	66
Subcarrier separation	15 kHz	60 kHz
Overload threshold ρ_{max}	0.95	0.95
UE noise figure	9 dB	
UE height	1.5 m	
Path loss model	Model of Sec 7.4 of [15]	

TABLE II. DQN ALGORITHM PARAMETERS

Parameter	Value
Initial collect steps	5000
Number of policy updates during learning	1e6
Experience Replay buffer maximum length (l)	1e5
Mini-batch size (J)	256
Discount factor(γ)	0.9
Learning rate (τ)	0.0003
ε value (ε-Greedy)	0.1
Neural network architecture	Single layer of 100 neurons
Time window (ΔT)	1 sec
Granularity $\Delta\beta$	0.1

B. *Performance evaluation of the DQN-based strategy*

In order to assess the benefits brought by the proposed DQN-based MC strategy, this sub-section compares the performance obtained by the proposed approach against two reference approaches, namely the *optimum strategy*, which applies an exhaustive search process to select in each time step the MC configuration with the maximum reward, and a classical *SINR-based strategy*, in which all the traffic of a UE is served by the cell with the highest SINR. It is worth mentioning that the optimum strategy is just considered here as an upper bound of the DQN algorithm operation, but it would not be a practical strategy for its implementation in real scenarios, mainly due to the large execution time for assessing all the possible MC configurations.

The comparison is performed by simulating a UE of interest following one hundred different trajectories of duration 400 s in the evaluation scenario and applying in each time window the MC configuration according to each of the evaluated schemes. Fig. 2 shows the obtained average reward for each one of the trajectories with all the considered strategies. It is observed that the DQN-based strategy achieves a performance very close to the optimum one in all the studied cases, which confirms the good behavior of the proposed approach. In turn, Fig. 2 also shows that the DQN-based strategy outperforms the classical SINR-based strategy in all the studied cases thanks to the better distribution of the traffic of the UE among the cells that avoids overload and enhances the obtained bit rate.

In order to assess the throughput performance, Fig. 3 plots, for each strategy, the cumulative distribution function (CDF) of the instantaneous throughput (T_u) values obtained by the UE of interest with all the considered trajectories. Again, it is observed that the DQN-based strategy achieves a close performance to the optimum strategy and clearly outperforms the SINR-based strategy.

Fig. 4 shows the CDF of the number of radio resource blocks assigned in the set of base stations that the UE is connected to at each moment while following each one of the studied trajectories. Again, it is observed how the DQN-based clearly outperforms the SINR-based strategy, as it is able to provide the required bit rate with less resource blocks.

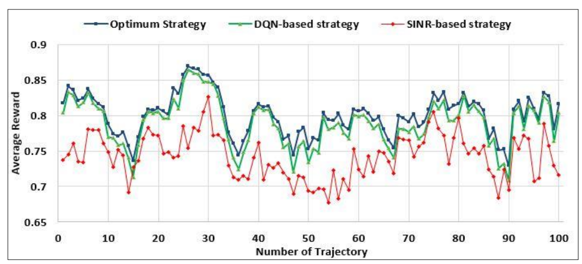

Fig. 2. Average reward for different trajectories.

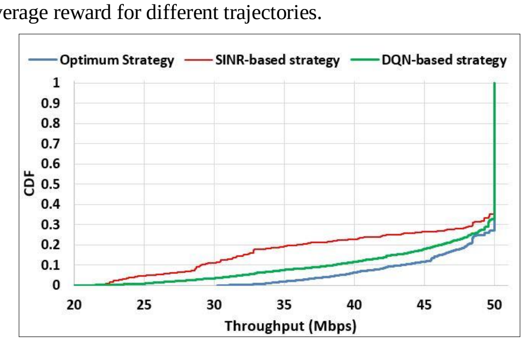

Fig. 3. CDF of the throughput achieved by the UE of interest.

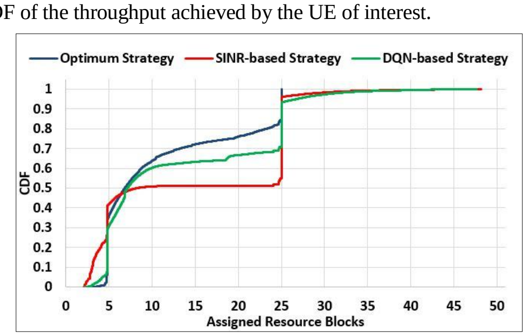

Fig. 4. CDF of assigned resource blocks to the UE of interest.

Aiming to further assess the behavior of the DQN-based strategy, we have carried out a more detailed analysis focusing on a specific period of time within one of the trajectories and comparing the SINR-based and the DQN-based strategies. In particular, during the selected period the UE moves following a straight trajectory passing close to the position of one of the LTE cells. Fig. 5 represents the evolution of the SINR experienced by the UE in the LTE and 5G NR cells during this period. It is clear that LTE cell has a higher SINR value, so this will be the selected cell with the SINR-based strategy during the whole analyzed period. In contrast, the DQN-based strategy is able to split the traffic through the LTE and 5G NR cells in accordance with the experienced conditions in terms of signal and load. In this case, Fig. 6(a) plots the evolution of the value $\beta_{1,1}$ selected by the algorithm for the LTE cell (the corresponding value for the 5G NR cell will be $\beta_{2,1}$=1-$\beta_{1,1}$). During most of the time the selected value $\beta_{1,1}$ corresponds to transmitting only a small fraction of the traffic through the LTE cell while the remaining traffic is sent through the 5G NR cell. In contrast, with the SINR-based strategy all the traffic will be sent through the LTE cell. As a result, Fig. 6(b) plots the evolution of the reward obtained with both strategies, and it can be observed that the DQN-based strategy outperforms the SINR-based one.

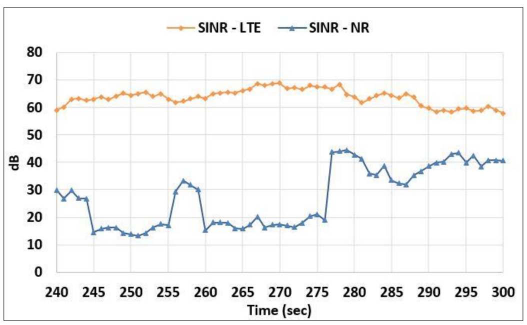

Fig. 5. SINR evolution of the LTE and NR cells in the analyzed period.

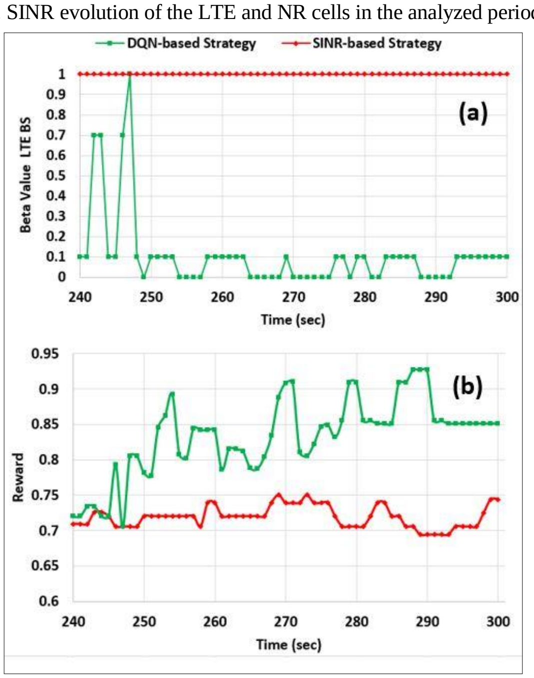

Fig. 6. Evolution of (a) $\beta_{1,1}$ for the LTE cell and (b) reward in the analyzed period.

V. Conclusions

This paper has presented a novel approach based on Deep Q-Network for splitting the traffic of a UE among cells when using multi-connectivity depending on the current traffic and radio conditions experienced by the UE and the involved cells. The strategy intends to minimize the resource consumption, the overload situations in the involved cells and enhancing throughput. Different performance results have been presented to compare the proposed approach against the optimum case and against a classical SINR-based approach. Results have shown the capability of the DQN agent to learn a quasi-optimal policy that outperforms the SINR-based approach in up to 13% in terms of reward, obtaining as a result better throughput performance with a more reduced number of allocated resource blocks.

Acknowledgement

This paper is part of ARTIST project (ref. PID2020-115104RB-I00) funded by MCIN/AEI/10.13039/501100011033. The work is also funded by the Spanish Ministry of Science and Innovation under grant ref. PRE2018-084691.

References

[1] B. B. Haile, E. Mutafungwa, H. Hämäläinen, "A Data-Driven Multiobjective Optimization Framework for Hyperdense 5G Network Plannin", IEEE Access, September, 2020.

[2] A. Maeder et al. "A Scalable and Flexible Radio Access Network Architecture for Fifth Generation Mobile Networks", *IEEE Communications Magazine*, November, 2016.

[3] E.Dahlman, S. Parkvall, J. Sköld, "5G NR the Next Generation Wireless Access Technology", Academic Press Elsevier, 2018

[4] 3GPP TS 37.340 v16.3.0 "Evolved Universal Terrestrial Radio Access (E-UTRA) and NR; Multi-connectivity; Stage 2 (Release 16)", September, 2020

[5] 3GPP TR 37.717-41-11 v0.6.0, "Rel-17 Dual Connectivity (DC) of 4 bands LTE inter-band CA (4DL/1UL) and 1 NR band (1DL/1UL) (Release 17)", August, 2021

[6] M. Yan, G. Feng, J. Zhou and S. Qin, "Smart Multi-RAT Access Based on Multiagent Reinforcement Learning," in *IEEE Transactions on Vehicular Technology*, vol. 67, no. 5, pp. 4539-4551, May 2018.

[7] V. F. Monteiro, et al., "Distributed RRM for 5G Multi-RAT Multiconnectivity Networks", IEEE Systems Journal, Vol. 13, No. 1, March, 2019.

[8] M. Gerasimenko et al. "Adaptive Resource Management Strategy in Practical Multi-Radio Heterogeneous Networks", IEEE Access, February, 2017

[9] P. K. Taksande, A.Roy, A. Karandikar, "Optimal Traffic Splitting Policy in LTE-based Heterogeneous Network", *IEEE Wireless Communications and Networking Confernece (WCNC)*,2018

[10] B. Zhang et al. "Goodput-Aware traffic Splitting Scheme with Non-ideal Backhaul for 5G-LTE Multi-Connectivity", *IEEE Wireless Communications and Networking Confernece (WCNC)*, 2019

[11] J. Elias, F. Martignon and S. Paris, "Optimal Split Bearer Control and Resource Allocation for Multi-Connectivity in 5G New Radio," 2021 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit), 2021, pp. 187-192.

[12] V. Mnih, et al. "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[13] O-RAN Alliance, "O-RAN Architecture Description v4.0", March, 2021.

[14] S. Guadarrama et al., TF-Agents: A Library For Reinforcement Learning in TensorFlow (2018).

[15] 3GPP TS 38.901 v16.1.0 "Study on Channel Model for Frequencies From 0.5 to 100 GHz (Release 16)", Dec. 2019