Reinforcement Learning-Based Radio Access Network Slicing for a 5G System with Support for Cellular V2X

Haider Daami R. Albonda(✉) and J. Pérez-Romero

Universitat Politècnica de Catalunya (UPC), Barcelona, Spain
{haider.albonda, jorperez}@tsc.upc.edu

Abstract. 5G mobile systems are expected to host a variety of services and applications such as enhanced mobile broadband (eMBB), massive machine-type communications (mMTC), and ultra-reliable low-latency communications (URLLC). Therefore, the major challenge in designing the 5G networks is how to support different types of users and applications with different quality-of-service requirements under a single physical network infrastructure. Recently, Radio Access Network (RAN) slicing has been introduced as a promising solution to address these challenges. In this direction, our paper investigates the RAN slicing problem when providing two generic services of 5G, namely eMBB and Cellular Vehicle-to-everything (V2X). We propose an efficient RAN slicing scheme based on offline reinforcement learning that allocates radio resources to different slices while accounting for their utility requirements and the dynamic changes in the traffic load in order to maximize efficiency of the resource utilization. A simulation-based analysis is presented to assess the performance of the proposed solution.

Keywords: Vehicle-to-everything (V2X) · Network slicing · Reinforcement learning

1 Introduction

The 5G system has the ambition to meet the widest range of service and applications in the history of mobile and wireless communications. Supported services are classified as (a) enhanced Mobile Broad Band (eMBB) that include services that require high bandwidth requirements, such as high definition (HD) video and Virtual Reality (VR); (b) Ultra Reliable and Low Latency Communications (URLLC) that aim to support low-latency transmissions of small payloads with extremely high reliability for a range of active terminals and (c) massive Machine Type Communications (mMTC) that aim to meet the demands of a large number of Internet Things (IoT). In responding to the very different requirements of these services and applications, the 5G system aims to provide a flexible platform to enable new business cases and models to integrate vertical industries, such as, automotive, manufacturing, and entertainment [1, 2].

In order to realize the above vision, network slicing is one of the key capabilities that will provide the required flexibility, as it allows multiple logical networks to be created on top of a common shared physical infrastructure. Each one of these logical

A. Kliks et al. (Eds.): CrownCom 2019, LNICST 291, pp. 262–276, 2019.
https://doi.org/10.1007/978-3-030-25748-4_20

networks is referred to as network slice and can be used to serve a particular service category (e.g. applications with different functional requirements) through the use of specific control plane (CP) and/or user plane (UP) functions [3]. Network slicing will help new services and new requirements to be quickly addressed, according to the needs of the industries [4].

Different works in the literature have investigated different aspects of network slicing, addressing both the slicing of the core network and the slicing of the Radio Access Network (RAN). For example, a low complexity heuristic algorithm and slicing for joint admission control in virtual wireless networks is proposed in [5]. In turn, the deployment of function decomposition and network slicing as a tool to improve the Evolved Packet Core (EPC) is presented in [6]. In [7], a model for orchestrating network slices based on the service requirements and available resources is introduced. They proposed a Markov decision process framework to formulate and determine the optimal policy that manages cross-slice admission control and resource allocation for the 5G networks.

Focusing on the RAN, some research studies have dealt with managing the split of the available radio resources among different slices to support different services (e.g. eMBB, mMTC, and URLLC) with main focus on the Packet Scheduling (PS) problem through different approaches. For example, a novel radio resource slicing framework for 5G networks with haptic communications is proposed in [8] based on virtualization of radio resources. The author adopted a reinforcement learning (RL) approach for dynamic radio resource slicing in a flexible way, while accounting for the utility requirements of different vertical applications. Similarly, a network slicing strategy based on an auction mechanism is introduced in [9] to decide the selling price of different types of network segments in order to maximize the network revenue and to optimally satisfy the resource requirements. A network slicing scheme based on game theory for managing the split of the available radio resources in a RAN among different slice types is proposed in [10] to maximize utility of radio resources. Similarly, an adaptive algorithm for virtual resource allocation based on Constrained Markov Decision Process is proposed in [11]. An online network slicing solution based on multi-armed bandit mathematical model to maximize network slicing multiplexing gains and achieving the accommodation of network slice requests in the system with an aggregated level of demands above the available capacity is proposed in [12].

Although the above works have proposed different approaches for RAN slicing, none of them has dealt with scenarios including slices for supporting Vehicle-to-Vehicle (V2V) communications, which constitute the focus of this paper. V2V communications are a particular type of the so-called Vehicle-to-everything (V2X) services in which vehicles can communicate between them through two operational modes, namely sidelink (i.e. direct communication between vehicles via PC5 interface) and cellular mode (i.e. communication between vehicles in two hops with the support of the base station via the Uu interface). These different options have impact on the resource consumption in the different links of the radio interface and thus they have to be taken into account when devising a RAN slicing strategy that distributes the radio resources among different slices if one of them supports V2X services. It is worth mentioning that, although the support for V2X sidelink communications was already standardized in 3GPP in the context of LTE [13], V2X sidelink is not yet included in the current release 15 of 5G New Radio (NR) specifications, but it is subject to study for future release 16 [14]. Based on all the above considerations, the key contributions of this paper can be summarized as follows. Firstly, the paper formulates the RAN slicing problem to support one slice for eMBB and another one for cellular V2X services on the same RAN infrastructure. The problem considers the split of radio resources assigned to each slice considering the characteristics of the different involved links, i.e. uplink and downlink for the eMBB services and uplink, downlink and sidelink for V2V. Secondly, the paper proposes a novel strategy based on offline Q-learning and softmax decision-making to determine the amount of radio resources assigned to each slice. The proposed solution is evaluated through extensive simulations to demonstrate its capability to perform efficient resource allocation in terms of network utilization, latency, data rate and congestion probability.

The rest of the paper is organized as follows. Section 2 presents the system model assumptions and the RAN slicing problem formulation. Section 3 presents the proposed RL approach for splitting the radio resources among the involved RAN slices. This approach is evaluated through simulations in Sect. 4 and compared against a reference scheme. Finally, conclusions and future work are summarized in Sect. 5.

2 System Model and Problem Formulation

2.1 System Model

The considered scenario assumes a cellular Next Generation Radio Access Network (NG-RAN) with a gNodeB (gNB) [15] composed by a single cell. A roadside unit (RSU) supporting V2X communications is attached to the gNB. A set of eMBB cellular users (CUs) numbered as $m = 1,\ldots, M$ are distributed randomly around the gNB and a flow of several independent vehicles move along a straight highway, as illustrated in the right part of Fig. 1. The highway segment is divided into sub-segments (clusters) by sectioning the road into smaller zones according to the length of the road. It is assumed that each vehicle includes a User Equipment (UE) that enables communication with the UEs in the rest of vehicles in the same cluster. Clusters are numbered as $j = 1,\ldots, C$, and the vehicles in the j-th cluster are numbered as $i = 1,\ldots, V(j)$.

The vehicles in the highway are assumed to enter the cell coverage following a Poisson process with arrival rate λ_a. The association between clusters and vehicles is managed and maintained by the RSU based on different metrics (e.g. position, direction, speed and link quality) through a periodic exchange of status information.

Regarding the V2X services, this paper assumes V2V communication between vehicles. They can be performed either in cellular or in sidelink mode. In cellular mode each UE communicates with each other through the Uu interface in a two-hops transmission via the gNB while in sidelink mode, direct V2V communications can be established over the PC5 interface. We assume that, when sidelink transmissions are utilized, every member vehicle can multicast the V2V messages directly to multiple member vehicles of the same cluster $1 \leq i \leq \mathrm{V}(j)$ using one-to-many technology. The decision on when to use cellular or sidelink mode is done based on [16].

To simultaneously support the eMBB and the V2X services, the network is logically divided into two network slices, namely RAN_slice_ID = 1 for V2X and RAN_slice_ID = 2 for eMBB. The whole cell bandwidth is organized in Resource Blocks (RBs) of bandwidth B. Let denote as N_{UL} the number of RBs in the UpLink (UL) and N_{DL} the number of RBs in the DownLink (DL). The RAN slicing process should distribute the UL and DL RBs among the two slices. For this purpose, let denote $\alpha_{s,UL}$ and $\alpha_{s,DL}$ as the fraction of UL and DL resources, respectively, for the RAN_slice_ID = s with s = 1, 2. Regarding sidelink communications, and since the support for sidelink has not been yet specified for 5G in current 3GPP release 15, this paper assumes the same approach as in current LTE-V2X system, in which the SL RBs are part of the total RBs of the UL. For this reason, the slice ratio $\alpha_{s,UL}$ is divided into two slice ratios, namely $\bar{\alpha}_{s,UL}$, which corresponds to the fraction of UL RBs that are used for uplink transmissions, and, $\alpha_{s,SL}$, which corresponds to the fraction of UL RBs used to support sidelink transmissions.

Each vehicle is assumed to generate packets randomly with rate λ_v packets/s according to Poisson arrival model. The length of the messages is S_m. When the vehicles operate in sidelink mode, the messages are transmitted using the SL resources allocated to the slice. Instead, when the vehicles operate in cellular mode, the messages are transmitted using the UL and DL resources. The average number of required RBs from V2X users of RAN_slice_ID = 1 per Transmission Time Interval (TTI) in UL, DL and SL, denoted respectively as $\Gamma_{1,UL}$, $\Gamma_{1,DL}$, $\Gamma_{1,SL}$ can be estimated as follows:

$$\Gamma_{1,x} = \frac{\sum_{t=1}^{T}\sum_{j=1}^{C}\sum_{i=1}^{V(j)} m(j,i,t)\cdot S_m}{T\cdot SP_{eff,x}\cdot B\cdot F_d} \tag{1}$$

where x denotes the type of link, i.e. $x \in$ {UL, DL, SL}, $m(j, i, t)$ is the number of transmitted messages by the vehicles of the j-th cluster in the t-th TTI and $SP_{eff,x}$ is the spectral efficiency in the x link, F_d is the TTI duration, which is 0.1 ms and T is the number of TTIs that defines the time window used to compute the average.

Regarding the eMBB service, the average number of required RBs for eMBB users of RAN_slice_ID = 2 in UL and DL in order to support a certain bit rate R_b is denoted as $\Gamma_{2,UL}$, $\Gamma_{2,DL}$, respectively, and can be statistically estimated as follows:

$$\Gamma_{2,x} = \frac{\sum_{t=1}^{T}\sum_{m=1}^{M} \rho_x(m,t)}{T} \tag{2}$$

where x denotes the type of link, and $\rho_x(m, t)$ is the number of required RBs by the m-th user in the link x and in the t-th TTI in order to get the required bit rate R_b. It is given by $\rho_x(m, t) = R_b/(SP_{eff,x} \cdot B)$. The values $\Gamma_{2,UL}$, $\Gamma_{2,DL}$ are computed within a time window T TTIs. Note also that $\Gamma_{2,SL} = 0$, since the eMBB slice does not generate sidelink traffic.

2.2 Problem Formulation for RAN Slicing

The focus of this paper is to determine the optimum slicing ratios $\alpha_{s,UL}$, $\alpha_{s,DL}$ in order to maximize the overall resource utilization under the constraints of satisfying the resource requirements for the users of the two considered slices.

The total utilization of UL resources U_{UL} is given by the aggregate of the required RBs in the UL and SL for each slice, provided that the aggregate of a given slice s does not exceed the total amount of resources allocated by the RAN slicing to this slice, i.e. $\alpha_{s,UL} \cdot N_{UL}$. Otherwise, the utilization of slice s will be limited to $\alpha_{s,UL} \cdot N_{UL}$ and the slice will experience outage. Correspondingly, the optimization problem for the uplink is defined as the maximization of the UL resource utilization subject to ensuring an outage probability lower than a maximum tolerable limit p_{out}. This is formally expressed as:

$$\max_{\alpha_{s,UL}} U_{UL} = \max_{\alpha_{s,UL}} \sum_{s} \min\left(\Gamma_{s,SL} + \Gamma_{s,UL}, \alpha_{s,UL} \cdot N_{UL}\right) \tag{3}$$

$$\text{s.t.} \quad \Pr\left[\Gamma_{s,SL} + \Gamma_{s,UL} \geq \alpha_{s,UL} \cdot N_{UL}\right] < p_{out} \quad s = 1,2 \tag{3a}$$

$$\sum_{s} \alpha_{s,UL} = 1 \tag{3b}$$

Following similar considerations, the optimization problem to maximize the resource utilization U_{DL} in the DL subject to ensuring a maximum outage probability is given by:

$$\max_{\alpha_{s,DL}} U_{DL} = \max_{\alpha_{s,DL}} \sum_{s} \min\left(\Gamma_{s,DL}, \alpha_{s,DL} \cdot N_{DL}\right) \tag{4}$$

$$\text{s.t.} \quad \Pr\left[\Gamma_{s,DL} \geq \alpha_{s,DL} \cdot N_{DL}\right] < p_{out} \quad s = 1,2 \tag{4a}$$

$$\sum_{s} \alpha_{s,DL} = 1 \tag{4b}$$

3 Reinforcement Learning-Based RAN Slicing Solution

The problems in (3) and (4) with their constraints are nonlinear optimization problems. Such an optimization problem is generally hard to solve. The complexity of solving this problem is high for a network of realistic size with fast varying traffic conditions. For this reason, we propose the use of an offline reinforcement learning approach to solve the problem in a more practical way.

The general approach is depicted in Fig. 1. Specifically, a slicing controller is responsible for determining the slicing ratios $\alpha_{s,UL}$, $\alpha_{s,DL}$ for each slice by executing the RL algorithm. It is assumed that two separate RL algorithms are executed for the UL and the DL to determine respectively $\alpha_{s,UL}$ and $\alpha_{s,DL}$. In the general operation of RL, the optimum solutions are found based on dynamically interacting with the environment based on trying different actions $a_{k,x}$ (i.e. different slicing ratios) selected from a set of possible actions numbered as $k = 1,\ldots, A_x$, where $x \in \{\text{UL, DL}\}$. As a result of the selected action, the RL process gets a reward $R_{TOT,x}(a_{k,x})$ that measures how good or bad the result of the action has been in terms of the desired optimization target. Based on this reward, the RL algorithm adjusts the decision making process to progressively learn the actions that lead to highest reward. The action selection is done by balancing the trade-off between exploitation (i.e. try actions with high reward) and exploration (i.e. try actions that have not been used before in order to learn from them). In case this interaction with the environment was done in an on-line way, i.e. by configuring the slicing ratios on the real network and then measuring the obtained performance, this could lead to serious performance degradation since, during the exploration process, wrong or unevaluated decisions could be made at certain points of time due to the exploration, and affecting all the UEs of a given slice. To avoid this problem, this paper considers an off-line RL, in which the slicing controller interacts with a network model that simulates the behavior of the network and allows testing the performance of the different actions in order to learn the optimum one prior to configuring it in the real network. The network model is based on a characterization of the network in terms of traffic generation, propagation modelling, etc.

The specific RL algorithm considered in this paper is the Q-learning based on soft-max decision making [17], which enables an exploration-exploitation traversing all possible actions in long-term. In turn, the reward should be defined in accordance with the optimization problem, which in this paper intends to maximize the resource utilization subject to the outage probability constraint. The details about the reward function and the detailed operation of the Q-learning algorithm are presented in the following.

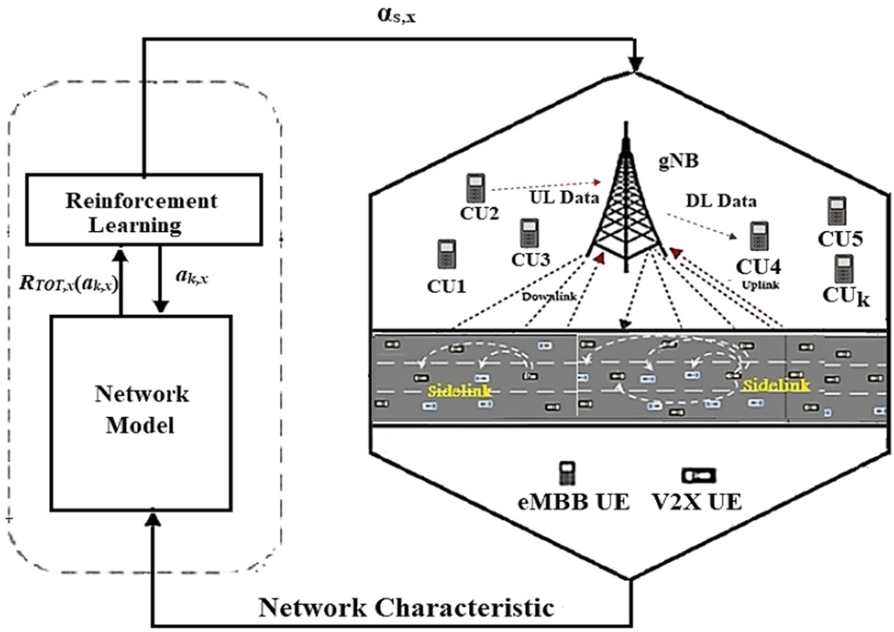

Fig. 1. General approach for the proposed RAN slicing solution.

3.1 Reward Computation

The reward function should reflect the ability of the taken action to fulfill the targets of the optimization problems (3) and (4). Based on this, and for a given action $a_{k,x}$ with associated slicing ratios $\alpha_{s,x}(k)$ the reward is computed as function of the normalized resource utilization $\Psi_{s,x}(a_{k,x})$ of slice s in link $x \in \{\text{UL, DL}\}$ defined as the ratio of used resources to the total allocated resources by the corresponding action. For the case of the V2X slice ($s = 1$), it is defined as:

$$\Psi_{1,UL}\left(a_{k,UL}\right) = \frac{\Gamma_{1,UL} + \Gamma_{1,SL}}{\alpha_{1,UL}(k) \cdot N_{UL}} \tag{5}$$

$$\Psi_{1,DL}\left(a_{k,DL}\right) = \frac{\Gamma_{1,DL}}{\alpha_{1,DL}(k) \cdot N_{DL}} \tag{6}$$

In turn, for the case of eMBB slice ($s = 2$), it is defined as:

$$\Psi_{2,UL}\left(a_{k,UL}\right) = \frac{\Gamma_{2,UL}}{\alpha_{2,UL}(k) \cdot N_{UL}} \tag{7}$$

$$\Psi_{2,DL}\left(a_{k,DL}\right) = \frac{\Gamma_{2,DL}}{\alpha_{2,DL}(k) N_{DL}} \tag{8}$$

Based on these expressions, the reward $R_{s,x}(a_{k,x})$ for the slice s in link $x \in \{\text{UL, DL}\}$ as a result of action $a_{k,x}$ is defined as

$$R_{s,x}\left(a_{k,x}\right) = \begin{cases} e^{\Psi_{s,x}\left(a_{k,x}\right)} & \Psi_{s,x}\left(a_{k,x}\right) \leq 1 \\ 1/\Psi_{s,x}\left(a_{k,x}\right) & otherwise \end{cases} \tag{9}$$

In (9), whenever $\Psi_{s,x}(a_{k,x})$ is a value between 0 and 1, the reward function will increase exponentially to its peak at $\Psi_{s,x}(a_{k,x}) = 1$. Therefore, the actions that lead to higher value of $\Psi_{s,x}(a_{k,x})$ (i.e. higher utilization) provide larger rewards and therefore this allows approaching the optimization target of (3) and (4). In contrast, if the value of $\Psi_{s,x}(a_{k,x}) > 1$, it means that the slice s will be in outage and thus the reward decreases to take into consideration constraints (3a) and (4a). Consequently, the formulation of the reward function per slice in (9) takes into account the constraints of the optimization problem. In addition, since the total reward has to account for the effect of the action on all the considered slices $s = 1,\ldots, S$, it is defined in general as the geometric mean of the per-slice rewards, that is:

$$R_{TOT,x}\left(a_{k,x}\right) = \left(\prod_{s=1}^{S} R_{s,x}\left(a_{k,x}\right)\right)^{\frac{1}{s}} \tag{10}$$

3.2 Q-Learning Algorithm

The ultimate target of the Q-learning scheme at the slicing controller is to find the optimal action (i.e. the optimal slicing ratios for a given link $x \in \{UL, DL\}$) that maximizes the expected long-term reward to each slice. To achieve this, the Q-learning interacts with the network model over discrete time-steps of fixed duration and estimates the reward of the chosen action. Based on the reward, the slice controller keeps a record of its experience when taking an action $a_{k.x}$ and stores the action-value function (also referred to as the Q-value) in $Q_x(a_{k,x})$. Every time step, the $Q_{UL}(a_{k,UL})$ and $Q_{DL}(a_{k,DL})$ values are updated following a single-state Q-learning approach with a null discount rate [17] as follows:

$$Q_x(a_{k,x}) \leftarrow (1-\alpha)\, Q_x(a_{k,x}) + \alpha \,.\, R_{TOT,x}(a_{k,x}) \tag{11}$$

where $\alpha \in (0, 1)$ is the learning rate, and $R_{TOT,x}(a_{k,x})$ is the total reward accounting for both V2X and eMBB slices after executing an action $a_{k,x}$. At initialization, i.e. when action $a_{k,x}$ has never been used in the past, $Q_x(a_{k,x})$ is initialized to an arbitrary value.

The selection of the different actions based on the $Q_x(a_{k,x})$ is made based on the softmax policy [17], in which the different actions are chosen probabilistically. Specifically, the probability $P_x(a_{k,x})$ of selecting action $a_{k,x}$, $k = 1,\ldots, A_x$, is defined as

$$P_x\left(a_{k,x}\right) = \frac{e^{Q_x\left(a_{k,x}\right)/\tau}}{\sum_{j=1}^{A_x} e^{Q_x\left(a_{j,x}\right)/\tau}} \tag{12}$$

where τ is a positive integer called temperature parameter that controls the selection probability. With high value of τ, the action probabilities become nearly equal. However, low value of τ causes a greater difference in selection probabilities for actions with different Q-values. Softmax decision making allows an efficient trade-off between exploration and exploitation, i.e. selecting with high probability those actions that have yield high reward, but also keeping a certain probability of exploring new actions, which can yield better decisions in the future. The pseudo-code of the proposed RL-based RAN slicing algorithm is summarized in Algorithm 1. Once the offline RL algorithm has converge, i.e. the selection probability of one of the actions is higher than 99.99%, the selection ratios $\alpha_{s,x}$ associated to this action are configured on the network, as illustrated in Fig. 1.

Algorithm 1: RAN slicing algorithm based on RL

1. **Inputs:** N_{UL}, N_{DL}: Number of RBs in UL and DL. S: number of slices, Set of actions $a_{k,x}$ for link $x \in \{UL,DL\}$
2. **Initialization of Learning**: $t \leftarrow 0$, $Q_x(a_{k,x}) = 0$, $k=1,\ldots,A_x$, $x \in \{UL,DL\}$
3. **Iteration**
4. **While** learning period is active do
5. **for** each link $x \in \{UL,DL\}$
6. Apply softmax and compute $P_x(a_{k,x})$ for each action $a_{k,x}$ according to (12);
7. Generate an uniformly distributed random number u ∈ {0,1}
8. Select an action $a_{k,x}$ based on u and probabilities $P_x(a_{k,x})$
9. Apply the selected action to the network and evaluate $\Psi_{s,x}(a_{k,x})$ based on (5)-(8).
10. **If** $\Psi_{s,x}(a_{k,x}) \leq 1$ then
11. $R_{s,x}\left(a_{k,x}\right) = e^{\Psi_{s,x}\left(\, _{k,x}\right)}$
12 **else**
13. $R_{s,x}\left(a_{k,x}\right) = 1/\Psi_{s,x}\left(a_{k,x}\right)$
14. **End**
15. **Compute** $R_{TOT,x}(a_{k,x})$ based on equation (10)
16. **Update** $Q_x(a_{k,x})$ based on equation (11)
17. **End**
18. **End**

4 Performance Analysis

In this section, we evaluate the performance of the proposed RAN slicing solution through system level simulation performed in MATLAB. Our simulation model is based on a single-cell hexagonal layout configured with a gNB. The model considers vehicular UEs communicating through cellular mode (uplink/downlink) and via side-link (direct V2V) and use slice (RAN_slice_ID = 1) and eMBB UEs operating in cellular mode (uplink/downlink) and using slice (RAN_slice_ID = 2) based on the assumptions described in Sect. 2. Note that the slice ratio $\alpha_{1,UL} \cdot N_{UL}$ is divided into two ratios ($\bar{\alpha}_{1,UL}$ = 65% of $\alpha_{1,UL} \cdot N_{UL}$ PRBs for V2X users in sidelink and $\alpha_{1,SL}$ = 35% of $\alpha_{1,UL} \cdot N_{UL}$ PRBs PRBs for V2X service in uplink direction). The traffic generation associated to each eMBB UE at a random position assumes that services generate sessions following a Poisson process with rate λm, required bit rate R_b = 1 Mb/s and average session duration of 120 s. The gNB supports a cell with a channel organized in 200 RBs composed by 12 subcarriers with subcarrier separation Δf = 30 kHz, which corresponds to one of the 5G NR numerologies defined in [18]. The actions specify the fraction of resources for

Table 1. Simulation parameters

Parameter	Value	Parameter	Value
Cell radius	500 m	Path loss model	The path loss and the LOS probability for cellular mode are modeled as in [20]. In sidelink mode, all V2V links are modeled based on freeway case (WINNER+B1) with hexagonal layout [ITU-R] [21]
Number of RBs per cell	$N_{UL} = N_{DL} = 200$ RBs		
Base station antenna gain	5 dB		
GBR (Rreq)	1 Mb/s	Number of actions	20 $\alpha_{1,x}$: varies from 0.05 to 1 in steps of 0.05 $\alpha_{2,x}$: varies from 1 to 0.05 in steps of 0.05
λ_a	1 UE/s		
Shadowing standard deviation	3 dB in LOS and 4 dB in NLOS	Length of the street	Freeway length = 1 km
Frequency	2.6 GHz	Lane width	4 m
Average session duration	120 s	Size of cluster	250 m
		Number of lanes	3 in one direction
Vehicle speed	80 km/h	Number of clusters	4
Learning rate α	0.1	Vehicular UE height	1.5 m
Temperature parameter τ	0.1	Safety message size (Sm)	300 bytes
		Time window T	30 s
Spectral efficiency model to map SINR	Model in section A.1 of [19]. The maximum spectral efficiency is 8.8 b/s/Hz	Average generation rate	Slice 1: $\lambda v = 1$ [packets/s] Slice 2: λm = varied from 0.2 to 1.2 sessions/s

V2X and eMBB and they are defined such that action $a_{k,x}$ corresponds to $\alpha_{1,x}(k) = 0.05 \cdot k$ and $\alpha_{2,x}(k) = (1 - 0.05 \cdot k)$ for $k = 1,\ldots, 20$, $x \in \{UL, DL\}$. All relevant system and simulation parameters are summarized in Table 1. The presented evaluation results intend to assess and illustrate the performance of the proposed solutions in terms of network capacity, throughput, and network congestion. As a reference for comparison, we assume a simpler RAN slicing strategy denoted as "Proportional Scheme", in which the ratio of RBs for each slice is proportional to its total traffic rate (in Mb/s). Figure 2 presents the RB utilization for V2X and eMBB slices in the UL as a function of the session generation rate (λ_m) for eMBB users.

From the presented results, we notice that our proposed model with off-line Q-learning maintains high resource utilization compared to the proportional strategy in different load scenarios. This is due to the RL-based slicing strategy that inherently tackles slice dynamics by selecting the most appropriate action considering the resource utilization in the reward. It is clearly observed that, as the arrival rate of requests increases, the RB utilization of the system increases gradually. For the proposed off-line Q-learning, when the arrival rate for the traffic of slice_ID = 2 is 1.2, the system utilizes around 79% of radio resources. For the proportional approach, the utilization is only about 73% of radio resources. Figure 3 depicts the throughput delivered in Mbits/sec for both eMBB and V2X slices in the sidelink and uplink. The figures illustrate with two lines the behavior of the proposed solution and the proportional scheme. Here, we can observe that the off-line Q-learning outperforms the proportional scheme in terms of throughput. The proposed scheme with off-line Q-learning achieved maximum throughput of 120 Mb/s in uplink when the eMBB arrival rate is 1.2 sessions/s, whereas in case of the proportional strategy model, the maximum throughput is only reached 114 Mb/s in uplink. The reasons are two-fold. First, when the arrival rate λ_m of eMBB UEs is increased, more users will use the network and this will increase the number of eMBB sessions and request more RBs to be used in transmissions. Second, as the number of eMBB sessions increases, requiring more radio resources, the proposed off-line Q- learning approach ensures more RBs which can be used to transmit data, while the proportional approach provides a lower number of available RBs for use in data transmissions.

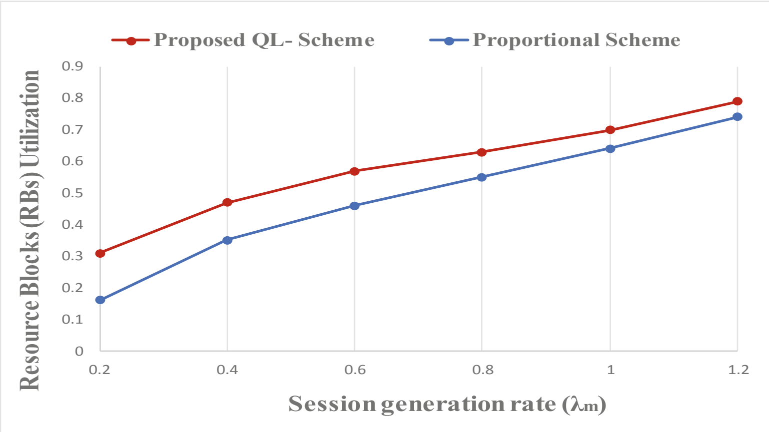

Fig. 2. Uplink RB utilization as a function of the eMBB session generation rate λ_m (sessions/s).

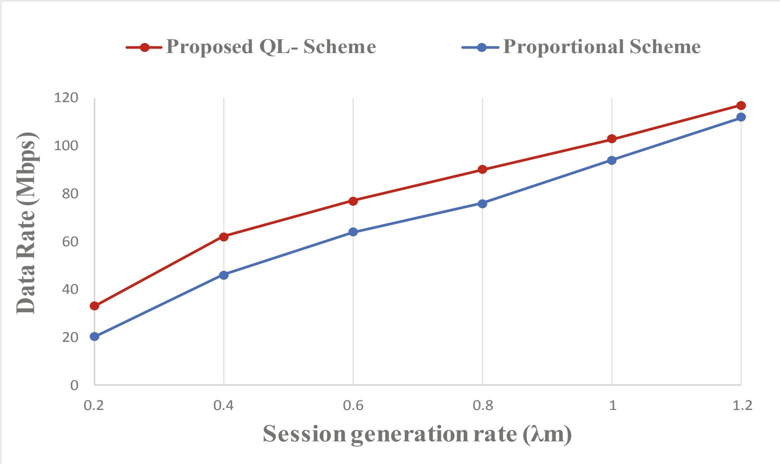

Fig. 3. Aggregated throughput experienced by both slices in uplink as a function of the eMBB session generation rate λ_m (sessions/s).

Figure 4 presents the RB utilisation for the sidelink. It shows that the proposed scheme with off-line Q-learning is able to improve the resource utilization compared to the reference model in different load scenarios. The proposed scheme with off-line Q-learning achieved maximum RB utilization of 93% in sidelink when the V2X arrival rate is 5 packets/s while in case of the proportional strategy model, the maximum utilization of RBs is only reached 72%.

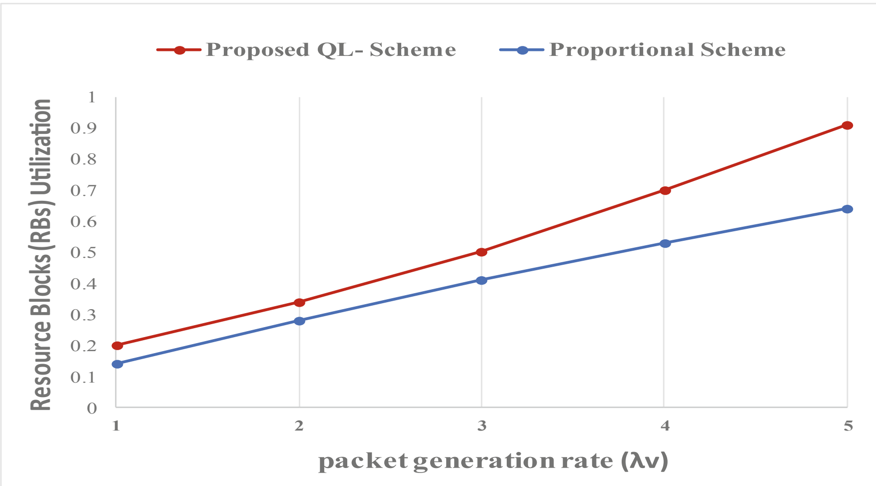

Fig. 4. Resource blocks utilization for sidelink transmissions as a function of the V2X UEs packet generation rate λ_v (packets/s).

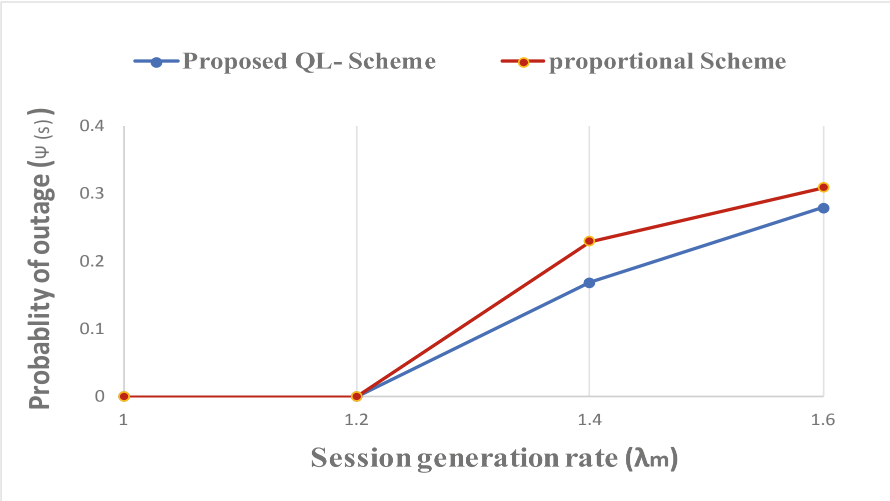

Fig. 5. Outage probability as a function of the eMBB session generation rate λ_m (UEs/s).

In Fig. 5, we investigate the probability of having congestion due to the lack of radio resources at a certain point of time. The outage probability of the proposed and proportional strategy is plotted against the eMBB session generation rate λ_m. As shown in the figure, increasing the traffic load leads to an increase in the outage probability of service. It can be also noted that our proposed scheme with off-line Q-learning can substantially reduce the congestion probability.

Figure 6 depicts the average latency for V2X service caused by channel access delay and the transmission delay. We can clearly observe that when packet generation

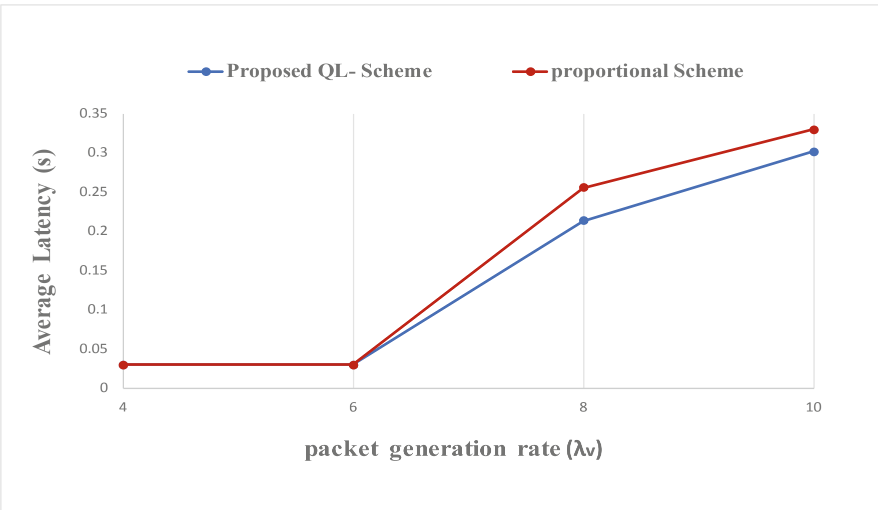

Fig. 6. Average latency as a function of the V2X UEs packet generation rate λ_v (packets/s).

rate λ_v is increased, more vehicles will use the network and request RBs to be used for the transmissions. This cause an increase in the waiting time and therefore increase the latency. We notice that our proposed model with off-line Q-learning approach reduces the latency compared to the proportional model and this is due to the fact that the proposed solution guarantees higher availability of resources avoiding outage situations.

5 Conclusions and Future Work

In this paper, we have investigated the splitting of radio resources into multiple RAN slices allocated to support V2X and eMBB services in uplink, downlink and sidelink (direct V2V) communications. We proposed a new RAN slicing strategy based on off-line Q-learning to determine the split of resources assigned to eMBB and V2X slices. This strategy has been compared against a reference scheme that makes an allocation of resources in proportion to the traffic rate of each slice. Extensive simulations were conducted to validate and analyze the performance of our proposed solution. Simulation results show the capability of the proposed algorithm to allocate the resources efficiently and improve the network performance. From the presented results, we notice that our proposed scheme outperforms the proportional scheme in terms of resource utilization, data rate, latency and congestion probability.

Acknowledgements. This work has been supported by the Spanish Research Council and FEDER funds under SONAR 5G grant (ref. TEC2017-82651-R).

References

1. Report ITU-R M.2410-0: Minimum Requirements Related to Technical Performance for IMT - 2020 Radio Interface(s), November 2017
2. 3GPP TR 38.912 V15.0.0: Study on New Radio (NR) access technology (Release 15), June 2018
3. 5G Americas: Network Slicing for 5G Networks and Services. Technical report, November 2016. http://www.5gamericas.org/files/3214/7975/0104/5GAmericasNetworkSlicing11.21Final.pdf
4. 3GPP TS 22.261 v16.0.0: Service requirements for the 5G system; Stage 1 (Release 15), June 2018
5. Soliman, H.M., Leon-Garcia, A.: QoS-aware frequency-space network slicing and admission control for virtual wireless networks. In: IEEE GLOBECOM (2016)
6. Sama, M.R., An, X., Wei, Q., Beker, S.: Reshaping the mobile core network via function decomposition and network slicing for the 5G era. In: Proceedings of IEEE Wireless Communications Networking Conference Workshops (WCNCW), pp. 90–96, April 2016
7. Hoang, D.T., Niyato, D., Wang, P: Optimal cross slice orchestration for 5G mobile services. arXiv:1712.05912v1, 16 December 2017
8. Aijaz, A.: Hap-SliceR: a radio resource slicing framework for 5G networks with haptic communications. IEEE Syst. J. **12**, 2285–2296 (2017)
9. Jiang, M., Condoluci, M., Mahmoodi, T.: Network slicing in 5G: an auction-based model. In: IEEE ICC (2017)

10. Caballero, P., Banchs, A., de Veciana, G., Costa-Pérez, X.: Network slicing games: enabling customization in multi-tenant networks. In: IEEE INFOCOM (2017)
11. Tang, L., Tan, Q., Shi, Y., Wang, C., Chen, Q.: Adaptive virtual resource allocation in 5G network slicing using constrained Markov decision process. IEEE Access **6**, 61184–61195 (2018)
12. Sciancalepore, V., Zanzi, L., Costa-Perez, X., Capone, A.: (PDF) ONETS: online network slice broker from theory to practice, January 2018
13. 3GPP TS 36.300 v15.0.0: Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2 (Release 15), December 2017
14. 3GPP TR 38.885 v1.0.0: NR; Study on Vehicle-to-Everything (Release 16), November 2018
15. 3GPP TS 38.401 v15.2.0: NG-RAN; Architecture description (Release 15), June 2018
16. Albonda, H.D.R., Pérez-Romero, J.: An efficient mode selection for improving resource utilization in sidelink V2X cellular networks. In: IEEE (CAMAD) Workshops, Barcelona, Spain, September 2018
17. Sutton, R.S., Barto, A.G.: Reinforcement Learning: An Introduction. MIT Press, Cambridge (1998)
18. 3GPP TS 38.211 v15.2.0: NR; Physical channels and modulation (Release 15), June 2018
19. 3GPP TR 36.942 v15.0.0: Radio Frequency (RF) system scenarios, September 2018
20. Report ITU-R M.2135: Guidelines for evaluation of radio interface technologies for IMT-Advanced (2009)
21. WINNER II Channel Models, D1.1.2 V1.2. http://www.cept.org/files/1050/documents/winner2%20%20final%20report.pdf