On the Configuration of Radio Resource Management in a Sliced RAN

J. Pérez-Romero, O. Sallent, R. Ferrús, R. Agustí

Universitat Politècnica de Catalunya (UPC)

[jorperez, sallent, ferrus, ramon]@tsc.upc.edu

***Abstract*— Network slicing is a fundamental feature of 5G systems that facilitates the provision of particular system behaviours adapted to specific service/application domains on top of a common network infrastructure. A network slice is in general composed by a core network slice and a Radio Access Network (RAN) slice. The realization of RAN slices is particularly challenging because it requires configuring and operating traffic differentiation and protection mechanisms to simultaneously deliver multiple and diverse RAN behaviors over a given pool of radio resources. In this context, this paper proposes to characterize the behavior of a RAN slice through the specification of a set of control parameters that are used to dictate the operation of the packet scheduling function at Layer 2 and the radio admission control function at Layer 3. An evaluation of the suitability of these parameters for achieving efficient radio resource sharing and isolation between RAN slices is presented when configuring a network for supporting a slice with multiple enhanced Mobile BroadBand services and another slice for providing Mission Critical services. The analysis reveals the different impact of the Layer 3 and Layer 2 parameters for isolating services of different slices depending on whether they require guaranteed or non-guaranteed bit rates.**

Keywords—Network slicing; 5G New Radio; RAN slice

I. INTRODUCTION

5G systems target the simultaneous support of a wide range of application scenarios and business models (e.g. automotive, utilities, smart cities, high-tech manufacturing) [1]. This expected versatility comes with a high variety of requirements on network functionalities (e.g. security, mobility, policy control features) and expected performance (e.g. peak rates above 10 Gbps, latencies below 1 ms with 10^{-5} reliability, 500 km/h mobility target) that cannot always be met through a common network setting. In this respect, support for network slicing in 5G has become a foundational requirement to allow operators to compose and manage dedicated logical networks with specific functionality, without losing the economies of scale of a common infrastructure [2].

Each one of these logical networks is referred to as *network slice*, and can be tailored to provide a particular system behavior (i.e. slice type) through the use of specific control plane (CP) and/or user plane (UP) functions to best support specific service/applications domains. For instance, a User Equipment (UE) for smart metering applications can be served through a network slice with radio access tailored to very small, infrequent messages and with no need to implement unnecessary functions (e.g. no mobility support), while a UE for enhanced Mobile BroadBand (eMBB) applications can be served through a network slice tailored to high data rate transmissions. Similarly, a network slice can also be used to provide a particular tenant (i.e. an organization or business entity entitled to use the network slice) with a given level of guaranteed network resources and isolation with regard to the operation of other concurrent slices. For instance, UEs/subscribers of a Public Safety agency can be served through a network slice that guarantees a minimum capacity during network congestion periods.

The normative specifications regarding service and operational requirements to support network slicing in 5G New Radio (NR) have already been completed by 3GPP [3] and current work is focusing on both system architecture aspects [4] and related management and orchestration capabilities [5]. Simultaneously, the network slicing concept has been addressed in the 5G architectures of different research projects such as 5G-NORMA [6], METIS-II [7] or SESAME [8]. Indeed, a complete solution for network slicing combines multiple facets, ranging from virtualization techniques for the abstraction and sharing of radio resources (e.g. network virtualization substrate concept in [9]) up to network slice lifecycle management solutions enabling the delivery of *Network Slice as a Service* (e.g. 5G network slice broker concept in [10]).

The realization of network slices considers, in the most general case, support for specific features and resources both in the 5G Core (5GC) network part, referred to as Core Network slice, and in the New Generation Radio Access Network (NG-RAN) part, referred to as Radio Access Network (RAN) slice. The realization of RAN slices is particularly challenging because it requires addressing how the pool of radio resources (i.e. RF bandwidth) available to one NG RAN node (referred to as gNB) can be configured and operated to simultaneously deliver multiple and diverse RAN behaviors [11].

Different works in the recent literature have proposed solutions for managing the split of the available radio resources in a RAN among different slice types (e.g. eMBB, massive Machine Type Communications [mMTC], Ultra-Reliable Low Latency Communications [URLLC]) and/or among different tenants. As discussed in [8], the split of radio resources can be performed at different levels with the support of different Radio Resource Management (RRM) functions. Most of the works have focused on the Packet Scheduling (PS) problem to determine the amount of resources available to each slice, making use of different approaches such as reinforcement learning [12], game theory [13], auction-based models [14] or generalized rate scheduling [15]. Similarly, the Network Virtualization Substrate (NVS) is proposed in [16] that includes a slice scheduler for defining bandwidth-based and resource-based reservations in a cell. In [17], the NVS is extended for multi-cell scenarios with the inclusion of a network-level scheduler that decides the sharing ratios for each cell. There exist

978-1-5386-3416-5/18/$31.00 ©2018 IEEE

less works that have considered Radio Admission Control (RAC) as a supporting technique for RAN slicing. In this respect, a multi-tenant admission control for cellular networks was proposed in [18], while in [19] the NVS concept based on packet scheduling is used but applying on top of it a tenant-specific admission control. Similarly, a joint admission control and network slicing approach is proposed in [20] by means of a heuristic algorithm that integrates spectrum allocation, admission control and spatial multiplexing. Very recently, in [21] a 5G network slice brokering solution has been presented that incorporates traffic forecasting, admission control and scheduling. Our recent work [11] proposes a general framework for the specification of a set of generic RAN slice configuration parameters, denoted as RAN Slice Descriptors, which can be used to characterize the features, policies and resources to be put in place across the radio protocol layers of a NG-RAN node for the realization of RAN slices.

Based on [11], the novelty of this paper is to further develop the specification of the subset of RAN slice configuration parameters intended to dictate the operation of both PS and RAC functions with regard to traffic differentiation and protection among the RAN slices offered though a common pool of radio resources. These control parameters are conceived as generic parameters that can be understood and enforced through any vendor-specific algorithmic implementation of a PS and/or RAC function. In addition, this paper takes into consideration the existence of multiple traffic classes in each slice and includes in the analysis the different Quality of Service (QoS) parameters identified by 3GPP for 5G NR. This constitutes another novelty with respect to most of the previous works, which either consider a single traffic class or take the slice traffic as an aggregate. Only the recent approach of [21] assumes multiple classes, but the focus of that paper is placed on the traffic forecasting algorithms.

Under the above framework, the contribution of this paper is to propose and analyze different options for configuring RAN slices by controlling certain parameters at radio protocol Layer 2 (L2) and Layer 3 (L3). These parameters allow configuring the RAC and PS functionalities that specify how the different radio resources are allocated to the slices. The considered approaches are evaluated in a multi-service scenario with one slice providing eMBB services and another one providing Mission Critical (MC) services for public safety.

The rest of the paper is organized as follows. Section II presents the approach for specifying RAN slices, detailing the considered control parameters at L2 and L3. Section III presents in detail the example scenario considered for RAN slice deployment, and Section IV provides the performance results for the different configuration options. Finally, Section V summarizes the conclusions.

II. Specification of RAN Slices

In order to support RAN slicing in a NG-RAN, [11] identifies that a set of new blocks of information, configuration descriptors and protocol features has to be introduced across the protocol layers of the radio interface. In particular, from a functional perspective, [11] proposes to specify the operation of each RAN slice through a set of configuration descriptors of the underlying radio protocol layers L3, L2 and L1.

L3 comprises the Radio Resource Control (RRC) protocol and RRM functions such as Radio Bearer Control (RBC), Radio Admission Control (RAC) and Connection Mobility Control (CMC) for the activation and maintenance of Radio Bearers (RB), which are the data transfer services delivered by the radio protocol stack. For each UE, one or more user plane RBs, denoted as Data RBs (DRBs), can be established per Protocol Data Unit (PDU) session, which defines the connectivity service provided by the 5GC [22].

When multiple RAN slices are realized over shared radio resources, the RRM functions for RBC, RAC and CMC have to assure that each RAN slice gets the expected amount of resources and, in case, handle any resource conflicts that might appear across slices. This concept is illustrated in Fig. 1, which represents a set of DRBs belonging to different slices. Whenever a Guaranteed Bit Rate (GBR) DRB is established in a cell, the RAC process is executed to check the availability of resources in the cell to provide the requested bit rate guarantees. As for non-GBR DRBs, no admission control has to be performed since there are no bit rate guarantees.

Based on the above aspects and as a practical realization of the L3 slice descriptor in [11], this paper considers as the main means of RAN slices' control related to L3 a parameter that specifies the maximum percentage of Physical Resource Blocks (PRBs) that can be considered in the RAC for the admission of GBR DRBs within the slice.

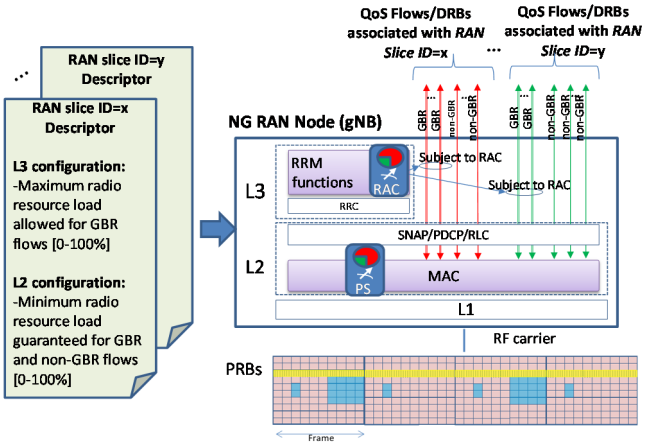


Fig. 1. Control of RAN slices at L3 and L2

L2 comprises a Medium Access Control (MAC) sub-layer for multiplexing and scheduling the packet transmissions of the DRBs over a set of transport channels exposed by L1. Moreover, L2 embeds a number of processing functions configurable on a per-DRB basis for e.g. segmentation, Automatic Repeat reQuest (ARQ) retransmissions, compression and ciphering (i.e. Radio Link Control [RLC] and Packet Data Convergence Protocol [PDCP]). In the NR specifications, an additional L2 sub-layer named Service Data Adaptation Protocol (SDAP) is included to map the DRBs and the traffic flows managed by the 5GC, referred to as QoS Flows [22]. Therefore, considering that the current MAC operation is based on individual UE and DRB - specific QoS profiles, it is necessary to define the Packet Scheduling (PS) behaviors to be enforced on the traffic aggregate of DRBs of the same slice and to specify the capability set of the applicable L2 sub-layers processing functions. The operation of the PS under multiple slices is illustrated in Fig. 1, reflecting that it determines the amount of resources that are assigned to both the GBR and non-GBR DRBs of the slices.

As a practical realization of the L2 slice descriptor in [11], this paper considers as the main means of RAN slices' control related to L2 a parameter that specifies the minimum % of PRBs that the PS guarantees to the slice for allocating transmissions of both GBR and non-GBR bearers. The limit is expressed as a minimum % because the actual value of PRBs used by a RAN slice can exceed this value at specific points of time provided that other RAN slices are not consuming all their PRBs.

III. RAN SLICES DEPLOYMENT SCENARIO: AN EXAMPLE

The considered scenario assumes a commercial operator that has deployed a NG-RAN and uses it to provide eMBB services to its customers. At the same time, the operator leases capacity from its NG-RAN infrastructure to public safety communications operator that provides Mission Critical (MC) services. For this purpose, the NG-RAN is configured in two RAN slices, namely RAN_slice_ID=1 for the eMBB services and RAN_slice_ID=2 for the MC services.

The considered services of each RAN slice are summarized in Table I, indicating for each one the QoS parameters following the QoS model of [23]. The parameters considered here include: (i) 5G QoS Identifier (5QI): it is a scalar that is mapped to specific QoS characteristics in terms of a priority in the scheduling process (indicated in parenthesis in Table I and meaning that the lower the value the higher the priority), a packet delay budget and a packet error rate. (ii) Allocation Retention and Priority (ARP): it defines the relative importance of a resource request and allows deciding whether a new QoS flow may be rejected in case of resource limitations. Lower values of ARP represent higher priorities. (iii) Guaranteed Flow Bit Rate (GFBR): it specifies the bit rate to be provided to a GBR QoS flow.

TABLE I. SERVICES OF EACH RAN SLICE

RAN Slice ID	Service	Type	5QI (Priority)	ARP	GFBR
1	Premium - Video HD	GBR	2 (40)	2	10 Mb/s
	Premium – Data	Non-GBR	6 (60)	2	N/A
	Basic – Video	GBR	2 (40)	3	1 Mb/s
	Basic – Data	Non-GBR	8 (80)	3	N/A
2	MC Video	GBR	2 (40)	2	2 Mb/s
	MC PTT	GBR	65 (7)	1	10 kb/s
	MC Data	Non-GBR	70 (55)	3	N/A

As shown in Table I, the commercial operator includes two different user profiles, the Premium and the Basic profile. Each of them includes a GBR video service and a non-GBR data service. The Premium video service provides High Definition (HD) quality while the Basic video service provides standard quality. In turn, the public safety operator provides two GBR services, namely Mission Critical Video (MC Video) and Mission Critical Push To Talk (MC PTT), and a non-GBR Mission Critical Data (MC Data) service

The deployment assumes a gNB with a single cell configured with a channel of 100 MHz organized in 275 Physical Resource Blocks (PRBs) composed by 12 subcarriers with subcarrier separation Δf=30 kHz, corresponding to one of the numerologies defined for 5G NR in [24]. Only the downlink direction is considered.

A system-level simulator is used to assess the performance. Table II presents the considered simulation parameters. Traffic generation assumes that services generate sessions following a Poisson process with the average rate indicated in the table for each slice. Each session corresponds to one QoS flow for one DRB associated to a UE at a random position following a uniform distribution within the cell radius. The service mix is such that, for RAN Slice ID=1, 10% of sessions are of Premium Video HD, 20% of Premium -Data, 30% of Basic Video and 40% of Basic-Data. In turn, for RAN Slice ID=2, 10% of the sessions are of MC Video, 50% of MC PTT and 40% of MC Data. In all the services the session duration is exponentially distributed with the average value indicated in Table II. During a session, the GBR services always have data in the buffer to send, while a non-GBR session has data in the buffer according to an activity factor of 0.2. A UE remains static for the duration of the session.

Based on the above traffic mixes and simulation parameters, the cell supports a reference GBR planned load level of $L_P(1)$=200 Mb/s for RAN Slice ID=1 and a GBR planned load level of $L_P(2)$=80 Mb/s for RAN Slice ID=2. In this respect, and to illustrate the performance for different offered load levels in relation to this planned load, this paper defines the normalized offered load for RAN slice s as $L_{norm}(s)=L(s)/\ L_P(s)$, where $L(s)$ is the GBR offered load of RAN slice s. It is worth mentioning that the offered load only accounts for GBR because non-GBR bearers do not have specific guarantees on the bit rate they can obtain.

TABLE II. SIMULATION PARAMETERS

Parameter	Value
Cell radius	115m
Path loss and shadowing model	Urban micro-cell model with hexagonal layout (see details in [25])
Shadowing standard deviation	3 dB in Line Of Sight (LOS) and 4 dB in Non Line Of Sight (NLOS) [25]
Base station antenna gain	5 dB
Frequency	3.6 GHz
Transmitted power per PRB	16.6 dBm
Number of PRBs	275
UE noise figure	9 dB
Link-level model to map Signal to Interference and Noise Ratio and bit rate	Model in section A.1 of [26] with maximum spectral efficiency 8.8 b/s/Hz.
Average session generation rate	Slice 1: varied from 0.5 to 3 sessions/s Slice 2: 2 sessions/s
Average session duration	120 s
Activity factor of NonGBR services	0.2
Average number of UEs with an active session	Slice 1: varies from 60 to 360 Slice 2: 240
Averaging period for measuring PRB occupation	30 s
Simulation duration	20000 s

The analysis considers the following configurations of the slices, as shown in Table III.

- Configuration #0 (Slice-agnostic): The L3 RAC function does not make distinctions among slices. The RAC establishes a limit of 70% of PRBs for all the GBR DRBs of the two RAN slices. In turn, at L2 the PS operates on the basis of the 5QI parameter and does not make distinctions among slices.
- Configuration #1 (Slice-aware L3): The RAC function for the RAN Slice ID=1 supporting eMBB services is

TABLE III. RAN SLICE CONFIGURATIONS

Control parameters	Configuration #0 (Slice agnostic)		Configuration #1 (Slice-aware L3)		Configuration #2 (Slice-aware L2&L3)	
	RAN Slice ID=1	RAN Slice ID=2	RAN Slice ID=1	RAN Slice ID=2	RAN Slice ID=1	RAN Slice ID=2
L3 - Maximum % of PRBs for the admission of GBR DRBs	70%		50%	20%	50%	20%
L2 - Minimum % of PRBs that the PS guarantees to the slice	N/A	N/A	N/A	N/A	70%	30%
L1 - Number of PRBs	275 PRBs		275 PRBs		275 PRBs	
L1- Numerology: Subcarrier separation (Δf)	30 kHz		30 kHz		30 kHz	

configured with a 50% of capacity for GBR bearers while for the RAN Slice ID=2 supporting MC services it is configured with 20% of capacity. In turn, the PS will not make differentiations among RAN slices when allocating PRBs to the different QoS flows.

- Configuration #2 (Slice-aware L3 & L2): This configuration considers the same at L3 as Configuration #1 to limit the % of PRBs of GBR DRBs in each slice at the RAC function. Additionally, the PS will ensure at least 70% of PRBs for RAN Slice ID=1 and 30% of PRBs for RAN Slice ID=2.

The RAC is configured based on the L3 control parameter of each slice and takes into consideration the different priorities associated with the ARP of each DRB. More specifically, with Configurations #1 and #2 a GBR DRB of RAN slice *s* requesting a guaranteed bit rate $GFBR_i$ with ARP value equal to ARP_i is admitted provided that the following condition is fulfilled:

$$\rho_{occ}\left(ARP_i,\ s\right)+\Delta\rho\left(GFBR_i\right)\le\rho_{\max}\left(s\right) \qquad (1)$$

where $\rho_{occ}(ARP_i,s)$ measures the average % of PRBs (with respect to the total number of PRBs in the cell) occupied by the GBR bearers of slice *s* that have an ARP lower or equal than ARP_i, $\Delta\rho(GFBR_i)$ is the estimated % of PRBs needed to provide a bit rate equal to $GFBR_i$ and $\rho_{max}(s)$ is the admission control limit, which is set equal to the maximum % of PRBs for GBR bearers indicated in the L3 control parameter.

In the case of Configuration #0, the same general formula as in (1) is used but without making distinctions between RAN slices, i.e. the PRB occupation $\rho_{occ}(ARP_i)$ considers all the GBR bearers of any slice with ARP lower or equal than ARP_i, and the limit ρ_{max} is the same for all RAN slices.

The resource allocation process at the PS operates by distributing first the PRBs among the admitted GBR DRBs in order to provide them with the requested GFBR value. After this process, the remaining PRBs are allocated among the active non-GBR DRBs in such a way that each DRB gets a number of PRBs inversely proportional to the priority level associated with its 5QI (see Table I). More specifically, let assume that, after assigning the PRBs to GBR bearers, there are *N* available PRBs that have to be distributed among *K* non-GBR DRBs, and that p_i is the priority level of the *i*-th bearer based on its 5QI. Then, the average number of PRBs that will be assigned to the *i*-th DRB along a certain time period (1s in the considered simulations) is:

$$N_i = N\frac{1/p_i}{\sum_{j=1}^{K} 1/p_j} \qquad (2)$$

In case of Configurations #0 and #1 the general formula (2) is applied by considering that *K* is the number of non-GBR bearers of all slices and *N* the remaining PRBs computed as the total number of PRBs in the cell minus the PRBs allocated to the GBR DRBs of all slices. On the contrary, in case of Configuration #2, expression (2) is applied separately for each slice. Then, *K* refers to the non-GBR DRBs of a given slice *s* and *N* the remaining PRBs of this slice, computed as the minimum number of PRBs guaranteed to the slice *s* (i.e. the L2 control parameter of slice *s*) minus the PRBs allocated to the GBR DRBs of this slice. Besides, in case that a slice has no active non-GBR DRBs, the remaining PRBs of this slice are distributed among the non-GBR bearers of the other slices according to their priority level.

IV. PERFORMANCE EVALUATION

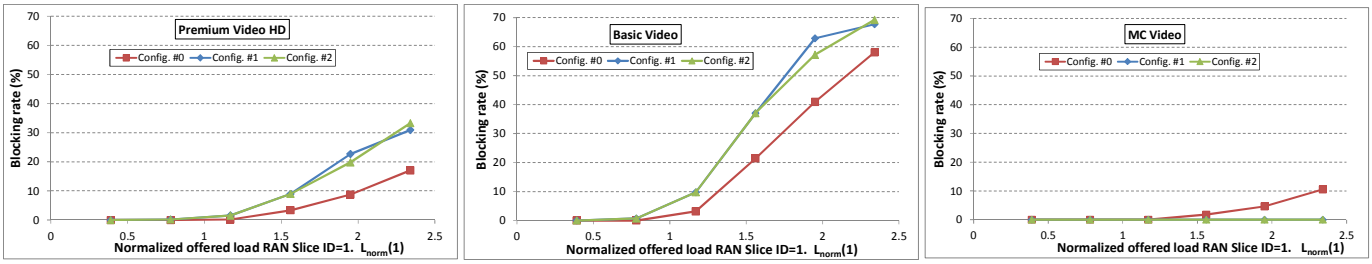
This section presents the comparison between the different Configurations #0, #1 and #2 in terms of the achieved performance for the different services when keeping constant the offered load of RAN Slice 2 at a normalized level of $L_{norm}(2)=0.6$ and varying the total offered load of the RAN Slice 1 considering both underload (i.e. $L_{norm}(1)<1$) and overload situations (i.e. $L_{norm}(1)>1$).

A. GBR services

For GBR services the main Key Performance Indicator (KPI) considered in this analysis is the blocking rate, which measures the percentage of GBR DRBs that are rejected by the admission control. Fig. 2a and Fig. 2b depict, respectively, the blocking rate for the Premium Video HD and the Basic Video services of RAN Slice 1 as a function of the normalized offered load of this RAN slice, $L_{norm}(1)$. In turn, Fig. 2c depicts the blocking rate of the MC Video service of RAN Slice 2. The blocking rate of the MC PTT service is not shown explicitly because it is 0% for all the considered traffic loads.

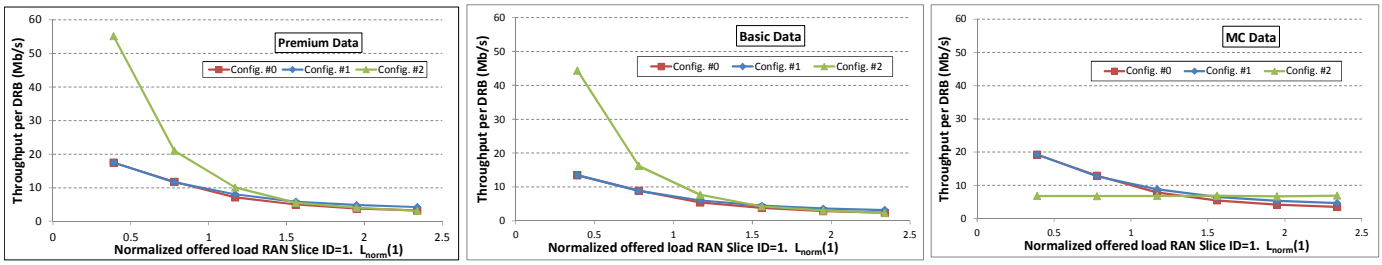
Focusing on the services of the RAN Slice 1 it is observed in Fig. 2a and Fig. 2b that the blocking rate is higher for the Basic Video than for the Premium Video HD, because the latter has a lower ARP value and therefore it has more priority during the admission process. As expected, the blocking rate of these services increases with the total offered load of their slice, being the increase significant when the normalized load is higher than 1 (i.e. during overload situations).

The comparison between the different configurations for the Basic Video and Premium Video HD reveals that the blocking rate is basically the same with both Configuration #1 and Configuration #2. The reason of this behaviour is that the blocking rate is primarily affected by the configuration of the admission control for each slice given by the L3 control parameter (i.e. maximum % of PRBs for GBR slices), which takes the same value (50% of PRBs for RAN Slice 1) in both configurations. Similarly, the blocking rate of the MC Video service of RAN Slice 2 is kept equal to 0% with both Configuration #1 and Configuration #2, and the performance is not affected by the increase in the offered load of the RAN Slice 2. The reason is that in these configurations the RAC is adjusted



(a) (b) (c)

Fig. 2 Blocking rate of (a) Premium Video HD service, (b) Basic Video service, (c) MC Video service



(a) (b) (c)

Fig. 3 Average throughput per DRB of (a) Premium Data service, (b) Basic Data service, (c) MC Data service

based on a maximum PRB consumption of 20% for this slice, and this amount is sufficient to serve both the MC Video and the MC PTT DRBs.

When the reference Configuration #0 is used, it is observed in Fig. 2a and Fig. 2b that the blocking rate of both Premium Video HD and Basic Video services of RAN Slice 1 decreases in relation to the other configurations. However, this is at the expense of degrading the blocking rate of the MC Video service of RAN Slice 2 (see Fig. 2c) when there is overload in RAN Slice 1 (i.e. when the normalized offered load is higher than 1). Specifically, the blocking rate can be as high as 10% for the highest value of the RAN Slice 1 offered load considered here. This is explained by the fact that Configuration #0 does not make distinctions between RAN Slices at L3 and just configures a total admission control limit of 70% of PRBs for GBR DRBs of both RAN Slices. Correspondingly, during overload situations of RAN Slice 1, the GBR bearers of Premium Video HD and Basic Video can consume more than the limit of 50% of PRBs imposed by Configurations #1 and #2 and, therefore, the GBR DRBs of RAN Slice 2 will have less than the 20% of PRBs that they can get with these configurations. This leads to some blockings of MC Video sessions. It is worth mentioning that this effect does not impact on the blocking rate of MC PTT DRBs because they have more priority (i.e. lower ARP) than all the other DRBs at the admission control stage. Looking at Fig. 2a and Fig. 2c with Configuration #0, it is also noticed that the blocking rate of Premium Video HD is higher than that of MC Video although both of them have the same ARP and 5QI. The reason is that MC Video requires a lower GFBR and, therefore, less PRBs, so it is less likely to block a request of this service.

The above observations reflect that the L3 control of the maximum % of PRBs for GBR bearers in a slice as done in Configurations #1 and #2 is appropriate to avoid that overload situations in one slice may affect the performance of GBR services in other slices, thus providing an adequate isolation.

The performance of GBR DRBs in terms of other KPIs such as the bit rate per DRB is not included here because the analysis reveals that in all the considered cases the PS is able to ensure the guaranteed GFBR values of Table I to the admitted RABs.

B. Non-GBR services

Non-GBR services do not pass any admission control check and they make use of the PRBs that are available after having performed the PRB allocation to the GBR services. Based on this, the main KPI considered for non-GBR services is the average throughput obtained by each DRB that has data to transmit in the buffer. Fig. 3a and Fig. 3b plot the corresponding values of this KPI for the Premium and Basic data services of RAN Slice 1, respectively, while Fig. 3c plots the throughput for the MC Data service of RAN Slice 2.

It is observed in Fig. 3a and Fig. 3b that, with all the configurations, the throughput per DRB of the Premium and Basic data services of RAN Slice 1 decreases when increasing the load of RAN Slice 1. The reasons are two-fold. First, because the load increase leads to an increase in the number of GBR sessions of this slice, thus leaving less PRBs available for the non-GBR services. Second, because the load increase also implies an increase in the number of active non-GBR data sessions, so the available PRBs have to be distributed among a larger number of non-GBR DRBs.

By comparing Fig. 3a and Fig. 3b it is also observed that Premium Data service achieves higher throughput than the Basic Data service. This is due to the higher priority associated to the 5QI value of Premium Data bearers in comparison with Basic Data bearers (see Table I). Indeed, it is noticed that the ratio between the Premium Data bearer throughput and the Basic Data bearer throughput matches the ratio between priorities of Table I for these services (i.e. 80/60).

The comparison between Configurations #0 and #1 for non-GBR services reveals that there are almost no differences in terms of throughput. The reason is that none of these configurations makes use of the L2 control to ensure a minimum amount of PRBs per slice at the PS. Therefore, the remaining PRBs in the cell after having allocated the GBR DRBs are shared between the Premium and Basic data services of RAN Slice 1 and the MC Data service of RAN Slice 2. On the contrary, when Configuration #2 is used, the throughput of both Premium and Basic data services is substantially increased, particularly for

low loads of RAN Slice 1. The reason is that in this case the L2 control ensures that the PS provides to this RAN slice at least the 70% of the PRBs in the cell for serving all its DRBs (both GBR and non-GBR). As a result, after having allocated the PRBs to the GBR DRBs, the PS will distribute the remaining PRBs from this 70% among the Premium and Basic data DRBs, without including in this distribution the MC Data service.

For RAN Slice 2, Fig. 3c reveals that the performance of the MC Data service with Configuration #2 is insensitive to the increase of offered load in RAN Slice 1, because in this case the PS always ensures at least 20% of the PRBs for this slice. On the contrary, with Configurations #0 and #1 the throughput of MC Data decreases with the load of RAN Slice 1, because the available PRBs are shared among the non-GBR services of the two RAN slices. Then, while for low loads of RAN slice 1 MC Data benefits from higher throughput than with Configuration #2, this throughput is progressively reduced when increasing the load of RAN slice 1, and eventually becomes lower than with Configuration #2 for the largest considered load.

V. CONCLUSIONS

This paper has proposed the specification of RAN slice configuration parameters at L2 and L3 to control the operation of PS and RAC functions in order to provide traffic differentiation and protection among RAN slices. In particular, the L3 control is performed by specifying the maximum percentage of PRBs to be considered in the admission control of GBR DRBs of a slice. In turn, the L2 control is performed by regulating the minimum percentage of PRBs that the packet scheduling guarantees to a slice for serving all its DRBs.

Different slice configurations have been analyzed by means of system-level simulations in a multi-service scenario with one slice providing eMBB services and another one providing MC services. A reference slice-agnostic configuration that does not make use of the abovementioned L2 and L3 control parameters has been compared against a slice-aware configuration based on L3 control and another one based on L2 and L3.

Results have revealed that: (i) The L3 control performed by the two slice-aware configurations is appropriate to isolate the GBR services of the different slices and to avoid that overload situations in one slice may affect the blocking rate of the other slice. (ii) The performance of GBR services obtained with both slice-aware configurations is very similar, reflecting that the introduction of L2 control parameters has little influence on the performance of GBR services. (iii) From the perspective of non-GBR traffic, very similar performance is obtained with a slice-agnostic configuration and with a slice-aware L3 control, meaning that the L2 control is needed to properly isolate the non-GBR DRBs of different slices, avoiding that the non-GBR traffic of one slice impacts negatively on the throughput experienced by the other slice. (iv) In certain cases the ARP and 5QI parameters used for controlling the QoS are able to provide some degree of isolation among services of different slices even with the slice agnostic configuration (e.g. in the case of an MC PTT service which has higher priority than other services). However, this is not sufficient in the most general case in which services of different slices may be configured with the same ARP values.

ACKNOWLEDGEMENT

This work has been supported by the EU funded H2020 5G-PPP project 5G ESSENCE under grant agreement 761592 and by the Spanish Research Council and FEDER funds under SONAR 5G grant (ref. TEC2017-82651-R).

REFERENCES

[1] NGMN Alliance, “5G White Paper”, February 2015.

[2] 3GPP TR 22.864: "Feasibility Study on New Services and Markets Technology Enablers - Network Operation; Stage 1 (Release 15)", September 2016.

[3] 3GPP TS 22.261 v15.0.0, “Service requirements for the 5G system; Stage 1 (Release 15)”, March 2017.

[4] 3GPP TS 23.501 v1.0.0, “System Architecture for the 5G System; Stage 2 (Release 15)”, June 2017.

[5] 3GPP TS 28.530 v0.3.0 “Management of network slicing in mobile networks; Concepts, use cases and requirements (Release 15)”, November, 2017.

[6] P. Rost et al., "Mobile network architecture evolution toward 5G," in IEEE Communications Magazine, vol. 54, no. 5, pp. 84-91, May 2016.

[7] METIS II White Paper “Preliminary Views and Initial Considerations on 5G RAN Architecture and Functional Design”, March 2016.

[8] O. Sallent, J. Perez-Romero, R. Ferrús, R. Agusti, "On Radio Access Network Slicing from a Radio Resource Management Perspective", IEEE Wireless Communications, October, 2017, pp. 166-174.

[9] X. Costa-Perez, J. Swetina, T. Guo, R. Mahindra and S. Rangarajan, "Radio access network virtualization for future mobile carrier networks," in IEEE Communications Magazine, vol. 51, no. 7, pp. 27-35, July 2013

[10] K. Samdanis, X. Costa-Perez and V. Sciancalepore, "From network sharing to multi-tenancy: The 5G network slice broker," in IEEE Communications Magazine, vol. 54, no. 7, pp. 32-39, July 2016.

[11] R. Ferrús, O. Sallent, J. Pérez-Romero, R. Agustí, “On 5G Radio Access Network Slicing: Radio Interface Protocol Features and Configuration”, to appear in IEEE Communications Magazine, September, 2017.

[12] A. Aijaz, “Hap-SliceR: A Radio Resource Slicing Framework for 5G Networks with Haptic Communications”, IEEE Systems Journal, 2017.

[13] P. Caballero, A. Banchs, G. de Veciana, X. Costa-Pérez, “Network Slicing Games: Enabling Customization in Multi-Tenant Networks”, IEEE INFOCOM, 2017.

[14] M. Jiang, M. Condoluci, T. Mahmoodi, “Network slicing in 5G: an Auction-based model”, IEEE ICC 2017.

[15] I. Malanchini, S. Valentin, O. Aydin, “Generalized Resource Sharing for Multiple Operators in Cellular Wireless Networks”, Int.Wireless Comm. and Mob. Comp. Conf. (IWCMC), Nicosia, Cyprus, August, 2014.

[16] R. Kokku, R. Mahindra, H. Zhang, S. Rangarajan, “NVS: A substrate for Virtualizing Wireless Resources in Cellular Networks”, IEEE/ACM Transactions on Networking, Vol. 20, No. 5, October, 2012.

[17] R. Mahindra, M. Khojastepour, H. Zhang, S. Rangarajan, “Radio Access Networks Sharing in Cellular Networks”, 21st IEEE Int. Conference on Network Protocols (ICNP), Göttingen, Germany, October, 2013

[18] J. Pérez-Romero, O. Sallent, R. Ferrus, R. Agustí, “Admission Control for Multi-tenant Radio Access Networks”, IEEE Int. Conference on Communicatins (ICC) - workshops, Paris, France, May, 2017.

[19] T. Guo, R. Arnott, “Active LTE RAN Sharing with Partial Resource Reservation”, IEEE 78th Vehicular Technology Conference (VTC Fall), Las Vegas, NV, USA, September, 2013.

[20] H. M. Soliman, A. Leon-Garcia, “QoS-Aware Frequency-Space Network Slicing and Admission Control for Virtual Wireless Networks”, IEEE GLOBECOM, 2016.

[21] V. Sciancalepore, K. Samdanis, X. Costa-Perez, D. Bega, M. Gramaglia, A. Banchs, “Mobile Traffic Forecasting for Maximizing 5G Network Slicing Resource Utilization”, IEEE INFOCOM 2017.

[22] 3GPP TS 38.300 V0.4.1, “NR; NR and NG-RAN Overall Description; Stage 2 (Release 15)”, June 2017.

[23] 3GPP TS 23.501 v1.4.0 “System Architecture for the 5G System; Stage 2 (Release 15)”, September, 2017.

[24] 3GPP TS 38.211 v1.0.0 “NR; Physical channels and modulation (Release 15)”, September, 2017.

[25] 3GPP TR 36.814 v9.0.0, “E-UTRA; Further advancements for E-UTRA physical layer aspects (Release 9)”, March, 2010.

[26] 3GPP TR 36.942 v12.0.0, "Radio Frequency (RF) system scenarios", September, 2014.