Assessment of 5G RAN Features for Integrated Services Provisioning in Smart Cities

Kandoi, Ayushi ; Raftopoulou, Maria; Litjens, Remco

DOI
10.1109/WiMob55322.2022.9941612

Publication date
2022

Document Version
Final published version

Published in
Proceedings of the 2022 18th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)

Citation (APA)
Kandoi, A., Raftopoulou, M., & Litjens, R. (2022). Assessment of 5G RAN Features for Integrated Services Provisioning in Smart Cities. In *Proceedings of the 2022 18th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)* (pp. 81-87). IEEE. https://doi.org/10.1109/WiMob55322.2022.9941612

2022 18th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob) | 978-1-6654-6975-3/22/$31.00 ©2022 IEEE | DOI: 10.1109/WIMOB55322.2022.9941612

Assessment of 5G RAN Features for Integrated Services Provisioning in Smart Cities

Ayushi Kandoi
Performance and Optimisation
KPN
The Hague, The Netherlands
ayushi.kandoi@kpn.com

Maria Raftopoulou
Quantum and Computer Engineering
Delft University of Technology
Delft, The Netherlands
M.Raftopoulou@tudelft.nl

Remco Litjens
Department of Networks
TNO
The Hague, The Netherlands
Quantum and Computer Engineering
Delft University of Technology
Delft, The Netherlands
remco.litjens@tno.nl

***Abstract*—In a 5G Radio Access Network (RAN), different features are offered as solutions to serve traffic with diverse characteristics and requirements, including flexible numerology, (non-)pre-emptive mini-slot based scheduling and network slicing. In this paper, we present an extensive simulation-based assessment of the relative merit of these distinct 5G features in the context of a smart city environment. We further derive the optimal feature combination and associated configuration which best handles the services related to the smart city environment given their performance requirements. The obtained insights confirm the commonly argued potential of slicing, emphasizing that the optimal configuration of the slice-specific numerology depends not only on the nature of the handled services but also on the selected RAN features. Among these features, non-preemptive mini-slot based scheduling and idle resource sharing reveal significant performance potential.**

***Index Terms*—5G, RAN features, RAN slicing, flexible numerology, bandwidth parts, mini-slots, smart cities**

I. INTRODUCTION

In smart cities, people, objects and machines are connected via wireless technologies to exchange data and collectively improve sustainability, traffic and safety, among others [1]. The cellular network will support services like Virtual Reality (VR), video surveillance and environment monitoring and thus the network should be designed to simultaneously support services of the Ultra-Reliable Low-Latency Communications (URLLC), enhanced Mobile Broadband (eMBB) and massive Machine Type Communication (mMTC) categories.

The RAN should be properly configured to best support the performance requirements associated with the mix of handled services. *Flexible numerology* has been introduced in 5G to allow for shorter Orthogonal Frequency Division Multiplex (OFDM) symbol times and thus for a shorter Transmission Time Interval (TTI) at the cost of a lower number of Physical Resource Blocks (PRBs) for a given carrier bandwidth [2]. The newly introduced concept of *BandWidth Parts* (BWPs) allows to configure multiple distinct numerologies on a given carrier, enabling the use of a tailored numerology for different service categories [3]. However, the split or radio resources among the BWPs leads to trunking losses which can be (partly) compensated by inter-BWP *idle radio resource sharing*.

Furthermore, on *Time Division Duplexing* (TDD) based carriers, the resource split between the DownLink (DL) and UpLink (UL) channels can be flexibly configured thus allowing for a better adaptation to the actual DL and UL traffic [3]. Also, *packet scheduling* i.e. the assignment of the available radio resources to the active Quality of Service (QoS) flows in the network, can be performed using *mini-slots*, which are contiguous sets of either 2, 4 or 7 OFDM symbols within a normal time slot, enabling latency-optimised and resource-efficient scheduling for e.g. URLLC-type services [4].

To support services with diverse requirements, *network slicing* has been introduced in 5G, which allows to configure the network with multiple independent virtual networks or slices that share the same physical infrastructure. Each slice is configured to serve traffic with a specific Service Level Agreement (SLA) [5] and thus, in the RAN, each slice can have its own numerology (or multiple ones) as well as packet scheduler such that it best serves the intended traffic. Similarly to BWPs, there are inherent trunking losses due to the radio resource split [6] which can be (partially) compensated by inter-slice idle radio resource sharing.

Several papers study the multi-numerology networks [7], the use of mini-slots [8], [9] or combinations of them [10], [11] to serve traffic with diverse QoS requirements. Most papers on RAN slicing focus on the resource assignment problem [12]–[14], neglecting the possibility that sliced networks may be outperformed by non-sliced networks due to the trunking losses which may be caused by slicing [6]. The purpose of this paper is to evaluate the merit of the various RAN features as well as combinations of these features in a smart city environment. In addition, the feature combination which best handles the services from all three service categories (eMBB, URLLC, mMTC) of the smart city environment simultaneously in regards to their QoS requirements, is found.

The remainder of the paper is organised as follows. In Section II the different RAN features are discussed in more detail. The modelling aspects, traffic characteristics and RAN configurations considered for the smart city environment are

The work was done while Ayushi Kandoi was with the Delft University of Technology, The Netherlands.

presented in Section III. The simulation results are analysed in Section IV and finally, the conclusions and recommendations for future work are given in Section V.

II. RAN Features

This section discusses the RAN features considered.

A. Flexible Numerology

The numerology μ defines the SubCarrier Spacing (SCS) of the OFDM symbols, which is fixed to 15 kHz in 4G networks, and the corresponding symbol duration, as shown in Table I. The decrease of the OFDM symbol duration, due to the increase of the SCS, leads to shorter slot duration and thus shorter TTIs. The shortened TTIs allow for faster transmissions and thus URLLC services benefit from higher numerologies. Considering that a PRB comprises twelve subcarriers regardless of the numerology, the number of PRBs within a given carrier bandwidth reduces with an increase of the SCS. This reduction in the number of PRBs, can lead to lower throughput as the gains obtained from frequency-domain channel-adaptive scheduling are reduced. Therefore, eMBB/mMTC services are best served with lower numerologies and thus there is a trade-off between delay and throughput [6]. Furthermore, the choice of numerology is also limited by whether the carrier frequency is in Frequency Range 1 (FR1) ($<$ 6 GHz) or in FR2 ($>$ 6 GHz), as also shown in Table I.

B. Bandwidth Parts

A BWP is a sub-band of a given carrier that is configured with its own numerology [3]. Thus, multiple numerologies can be configured on a single carrier to support multi-service traffic e.g. by configuring BWPs with both low and high numerologies to serve eMBB and URLLC traffic, respectively.

Unless appropriate measures are taken, the multiplexing of numerologies on the same carrier introduces Inter-Numerology Interference (INI) because the subcarriers of distinct numerologies are not orthogonal to each other. INI can be eliminated by using windowing, filtering and guard bands between the numerologies, whose size depends on the adjacently used numerologies [15]. Additionally, edge guard bands are used at either edge of the carrier, whose size depends on the numerology used at the carrier's edge [16].

C. Duplexing

In 5G, the TDD configuration, defining the frequency of the DL and UL channels, can be flexibly configured to adapt to the traffic intensity of the DL and UL channels. Note that the same TDD configuration should apply in all cells to avoid interference and that the regulator may impose a fixed TDD configuration [17]. To switch from the DL to the UL resources, a Guard Time (GT) is necessary and we denote as a Special Slot (SS), the slot that contains the GT. The remaining OFDM symbols within the given TDD periodicity can be assigned to the DL and UL channels based on the expected DL and UL traffic [18]. Fig. 1 shows the derived TDD configuration, assuming a network where the DL to UL traffic ratio is 2:1. The DL channel consists of 66% of the resources and a periodicity of five slots and a GT of two symbols is assumed.

TABLE I
5G Numerologies [2].

μ	SCS (kHz)	OFDM Symbol Duration (μs)	Slot Duration (ms)	Frequency Range
0	15	71.35	1	FR1
1	30	35.68	0.5	FR1
2	60	17.84	0.25	FR1 and FR2
3	120	8.92	0.125	FR2
4	240	4.46	0.0625	FR2

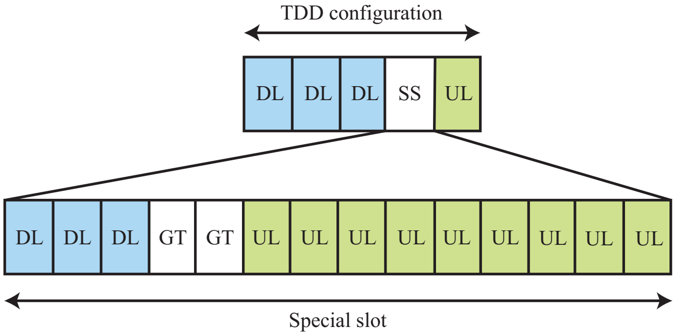

Fig. 1. TDD configuration example with a periodicity of five slots.

D. Packet Scheduling

The packet scheduler S determines at every TTI t and for every PRB f, which packets of the active QoS flows will be served based on the metric $M_{S,q}(t, f)$, calculated for every QoS flow q. Assigning the resources on a per-PRB level exploits frequency-selective fading and interference, and thus achieves frequency diversity gains. In general, for a network with N flows, at a given TTI t, with a queue of packets maintained for each flow, the deployed scheduler S assigns PRB f to QoS flow q^* which has the highest $M_{S,q}(t, f)$ value:

$$i^* = \underset{1 \leq q \leq N}{\operatorname{argmax}} M_{S,q}(t, f).$$

Finally, the packet scheduler can decide to not serve (and hence drop) a head-of-line packet if it determines that it cannot be delivered within an imposed delay constraint.

E. Mini-slots

While a regular scheduling slot comprises 14 OFDM symbols, a mini-slot consist of 2, 4 or 7 OFDM symbols, and the selected mini-slot size depends on the transport block size and the channel quality. Thus, the transmission of a small transport block may be shorter than a full slot duration and thus lead to delay improvements and lower inter-cell interference. Also, because of the shorter transmission times, multiple transport blocks can be transmitted within a single slot, rather than wasting otherwise unused symbols, which enhances resource efficiency and, consequently traffic handling capacity and service performance. The only limitation is that a mini-slot should not span over two adjacent slots, i.e. it should be aligned to the symbol boundaries of the slots [10]. Furthermore, a transport block can be transmitted at any time within a slot which improves the delay. In this study, we consider three different mini-slots based scheduling approaches:

- *Default:* At the start of a regular slot, the scheduler decides which packets from the buffer are served in the upcoming slot, based on the $M_{S,q}$ metric, and determines the minimum number of symbols needed per packet.
- *Non-pre-emptive:* There are two distinct types of scheduling moments: *(i)* at the start of a regular slot, transmissions are scheduled as under the default scheme; *(ii)* during a regular slot, and given on-going transmissions, the scheduler continuously checks for new URLLC packet arrivals, which are then transmitted using a mini-slot, assuming sufficient unused resources.
- *Pre-emptive:* This approach is largely the same as the non-pre-emptive scheme. However, in case of insufficient unused resources for a new URLLC packet, the scheduler pre-empts an ongoing eMBB or mMTC transmission to schedule the URLLC packet. The transmission to be pre-empted is the one whose resources maximize the $M_{S,q}$ metric of the new URLLC packet, as this minimizes the used resources and, consequently, the degree of pre-emption of the ongoing transmissions. We assume that the pre-empted transmission has failed.

F. RAN Slicing

RAN slicing is introduced in 5G to support services with diverse requirements, e.g. by configuring each slice with its own numerology and packet scheduler, to best serve the intended traffic. Each RAN slice can be dedicated to a particular service category, e.g. eMBB, URLLC and mMTC or to a vertical customer with specific QoS requirements [5]. The use of BWPs in non-sliced networks allows for a similar configuration to sliced networks with the key difference that BWPs cannot have their own packet scheduler. Because there are no advantages of BWPs over slicing when configuring different numerologies, we only consider slicing in our analysis.

Assigning the radio resources to each slice in such a way that the QoS requirements for each slice are guaranteed is a non-trivial task as the traffic can be very dynamic, the QoS requirements may be very demanding and the amount of available radio resources in the network is limited. This can be resolved by dynamically assigning the resources to each slice. In this study, the average traffic demand is constant; with variations only at finer timescale, which are handled by the scheduling mechanism and/or by inter-slice idle resource sharing, as explained in the following subsection. Thus, dynamic slice resource assignment is out of focus for this study. Additionally, splitting the radio resources among the slices leads to trunking losses compared to non-sliced networks [6] and it can be compensated by inter-slice idle resource sharing.

G. Idle Radio Resource Sharing

Idle radio resource sharing allows slices (denoted as source slices) to use idle resources of other slices (denoted as target slices), and thus it improves the performance and the spectral efficiency of the network. However, there are limitations when sharing the radio resources, based on the configuration of the source and target slices [18]. In particular, a source slice with a different numerology compared to the target slice can only use the resources of the target slice when the scheduling times at both slices are aligned. This drawback is mitigated with mini-slots which allow scheduling of data at any time.

III. Models and Scenarios

This section describes the modeling aspects, provides the scenario configurations and defines the Key Performance Indicators (KPIs).

A. System Model

An urban macro-cellular environment is considered with 19 three-sectorised sites in a hexagonal layout with an inter-site distance of 500 m [19]. Three types of User Equipments (UEs) are uniformly distributed in space, with each type of UE having one data session related to either eMBB or mMTC or URLLC traffic. Table II shows the configuration of the base stations (gNBs) and the UEs while the antenna diagram used at the gNB is shown in [20].

The propagation environment has been generated with the QuadRiGa 3GPP Urban Macro-cell Non-Line-of-Sight (NLoS) model, including lognormal shadowing with standard deviation σ_{SF} = 6 dB and Rayleigh multipath fading [21].

We assume that each cell is assigned a 15 MHz wide carrier in the 3.5 GHz band (FR1) and we consider both DL and UL transmissions [18]. In each cell, the packet scheduler determines, in both the DL and UL, which time-frequency resources are assigned to each of the active UEs, based on channel quality estimates/reports. The DL channel quality of the UEs is reported to the gNB via sub-band Channel Quality Indicators (CQIs), while the UL channel quality of the UEs is measured at the gNB based on the Sounding Reference Signal (SRS), both applying a 5 ms periodicity. [22].

Based on the scheduling decisions, the gNB selects for the DL or UL transmission the highest attainable Modulation and Coding Scheme (MCS) (up to 64-QAM) with an estimated BLock Error Rate (BLER) not exceeding 0.001% and 10% for URLLC and eMBB/mMTC services, respectively. MCS-specific BLER-vs-Signal-to-Interference-plus-Noise Ratio (SINR) curves have been derived using the Vienna 5G Link Level Simulator [23]. To map a set of PRB specific SINRs to a single effective SINR value for the full set of PRBs, we use the Mutual Information Effective SINR Mapping (MIESM) method [24]. Finally, an Outer-Loop Link Adaptation (OLLA) scheme is used to modify the mapping of the SINR to an MCS due to imperfections in channel quality reporting [25].

B. Traffic Model

The services in smart cities are BroadBand access everywhere (BB), Virtual Reality (VR) relating to gaming sessions, Video Surveillance (VS) used by security officials to improve the safety of the city and Sensors for Monitoring (SM) e.g.

TABLE II
CONFIGURATION OF THE gNB AND THE UE [19], [20].

Parameter	gNB	UE
Height	25 m	1.5 m
Maximum Antenna Gain	17 dBi	0 dBi
Transmit Power	49 dBm	23 dBm
Noise Figure	3 dB	9 dB
Electrical Tilt	10^o	N/A

environmental conditions and smoke in public buildings and houses [26]. The QoS requirements for each service are shown in Table III based on [18], [26]–[28], with reliability defined as the fraction of packets delivered within the delay budget.

The BB and VR sessions follow a spatially uniform Poisson process with arrival rates $\lambda_{BB,DL}$ (for DL) and $\lambda_{BB,UL}$ (for UL) and λ_{VR} (for both DL and UL) sessions per sec, respectively. The BB sessions are active until the file of size S_{BB} is fully transferred. The VR sessions are modeled to be active for a relatively short period of 5 seconds yet with an increased VR session arrival rate, to ensure both a realistic VR traffic load as well as a sufficient number of handled VR sessions to allow reliable performance assessment within a reasonable simulation time. For the other two services i.e. VS and SM, there are a total of N_i persistent sessions per service, where i denotes the service name, which are spatially uniformly distributed. The packets of each session have size $S_{i,j}$, where j denotes the channel direction, and they arrive with a constant period of $T_{i,j}$ seconds. For the persistent flows, the arrival time of the first packet is randomly chosen at $[0, T_{i,j}]$. Table III shows all of the modeling parameters based on [18], [26], [29]–[32].

C. Scenario Configurations

Each RAN configuration consists of a number of the previously explained features i.e. numerology, duplexing, scheduler, mini-slots, slicing and resource sharing. The TDD configuration is derived based on the expected DL to UL traffic ratio network-wide. Based on the previously discussed traffic model, the TDD configuration consists of two DL slots, followed by a special slot and then by two UL slots, thus considering a TDD periodicity of five slots. The special slot consists of ten DL symbols, followed by two symbols for the GT and then by two UL symbols [18].

Additionally, for all considered RAN configurations, the chosen packet scheduler is the *Modified-Largest Weighted Delay First* (M-LWDF) scheduler which aims to serve URLLC, eMBB and mMTC flows simultaneously, featuring both channel-adaptive and delay-oriented aspects [6]. The applied scheduling metric is given by:

$$M_{S,q}(t,f) = \begin{cases} -\frac{W_q(t) log(\delta_q)}{\tau_q} \frac{R_q(t,f)}{\overline{R}_q(t-1)} & q \in \text{URLLC}, \\ \frac{R_q(t,f)}{\overline{R}_q(t-1)} & q \in \text{eMBB, mMTC}, \end{cases}$$

where τ_q denotes the latency constraint of flow q, $\delta_q \in [0, 1]$ is the maximum allowed packet drop rate for flow q, $W_q(t)$ and $\overline{R}_q(t) = (1-\frac{1}{t_c})\overline{R}_q(t-1)+\frac{1}{t_c}R_q(t-1)$ denote the head-of-line packet latency and the exponentially smoothed experienced bit rate, respectively, of flow q up to and including TTI t and t_c is the smoothing parameter [6].

For the configurations using slices, the radio resources are distributed among the slices. This resource assignment is based on the resource utilisation of each service, which is derived via simulations, as shown in [18]. Also, when multiple numerologies are configured, the relevant guard bands are applied. Fig. 2 shows an example with two slices configured with numerologies 0 and 1, respectively. The size of the edge guard bands as well as the inter-numerology guard band are dependent on the applied numerologies. Finally, we consider two slicing options: *(i)* slicing per numerology i.e. a slice for each distinct numerology used and *(ii)* slicing per service category i.e. a total of three slices, with each slice related to eMBB, URLLC and mMTC services, respectively.

In this paper we analyse the respective (dis)advantages of the flexible numerology, mini-slots, slicing and resource sharing features, as well as sensible combinations of these 5G features in mixed-traffic scenarios. For example, one potential way of handling a mix of URLLC and eMBB/mMTC traffic, which is intrinsically best served with higher and lower numerologies, respectively [6], is to configure a single slice with numerology (in support of eMBB/mMTC services) in combination with mini-slot based scheduling to ensure good URLLC performance. An alternative approach could be to configure distinct URLLC and eMBB/mMTC slices with tailored numerologies, noting however the resource inefficiencies due to the required inter-numerology guard bands and (potentially) trunking losses.

D. Key Performance Indicator Definitions

Distinct KPIs are defined for the eMBB, URLLC and mMTC services. For the eMBB services the KPI of relevance is the 5th throughput percentile and its target level per service is shown in Table III under the QoS requirements. For the URLLC services, the KPI of relevance is the fraction of URLLC sessions per service experiencing the required degree of reliability shown in Table III. We define reliability for a URLLC session as the fraction of packets that are successfully received within the given delay budget. The target level for this KPI is 90% for all URLLC services. Finally, the KPI related to the mMTC services is the 95th delay percentile. Because the delay budget for mMTC services varies between seconds and minutes, a specific target value is not set.

IV. SIMULATION RESULTS

The above-described models and KPIs have been implemented in a Python-based dynamic system-level simulator which has been used to conduct a series of scenario-based assessments. This section presents the evaluation of the flexible numerology, mini-slot based scheduling, slicing and resource sharing features, as well as combinations of these features. To describe each RAN configuration, we adapt a notation, consisting of the features, combined with an underscore as:

$$\{\text{NS,S}\}\_(\mu_1/\cdots/\mu_Y)\_\{\text{NM,M(x)}\}\_\{\text{NR,R}\},$$

TABLE III
SUMMARY OF TRAFFIC MODELING AND THEIR RESPECTIVE QOS REQUIREMENTS.

Service	Service Category	Channel	QoS Requirements	Session Arrival	Packet Arrival	File/Packet Size (KB)
BroadBand Access Everywhere (BB)	eMBB	DL and UL	50 Mbps (DL) and 10 Mbps (UL)	Non-persistent with $\lambda_{BB,DL} = 100$ and $\lambda_{BB,UL} = 40$	Full Buffer	$S_{BB} = 2000$ (both DL and UL)
Virtual Reality (VR)	URLLC	DL and UL	90% of sessions with 96% reliability (delay budget: 10ms) (both DL and UL)	Non-persistent with $\lambda_{VR} = 40$ (both DL and UL)	Poisson Process with $\lambda_{VR,DL} = 200$ and Periodic with $T_{VR,UL} = 0.004s$	$S_{VR,DL} = 1.25$ and $S_{VR,UL} = 0.5$
Video Surveillance (VS)	eMBB	UL	25 Mbps	Persistent with $N_{VS} = 676$	Periodic with $T_{VS,UL} = 0.036s$	$S_{VS,UL} = 4.5$
Sensors for Monitoring (SM)	mMTC	UL	delay budget: seconds to minutes	Persistent with $N_{SM} = 675600$	Periodic with $T_{SM,UL} = 60s$	$S_{SM,UL} = 0.2$

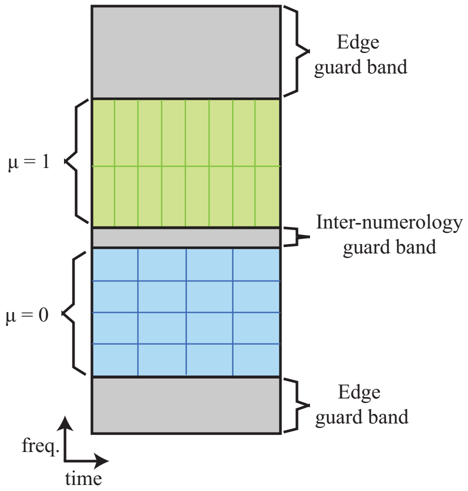

Fig. 2. Edge and inter-numerology guard bands for a carrier bandwidth with two slices.

to indicate no slicing (NS) or slicing (S), the use of Y numerologies and their respective values $\mu_1, \cdots, \mu_Y$, no mini-slot based scheduling (NM) or mini-slot based scheduling with approach x (M(x)), where x $\in$ {D, NP, P} denotes the default, non-pre-emptive and pre-emptive approaches, respectively, and no resource sharing (NR) or resource sharing (R). The abbreviations of the considered services are shown in Table III. Fig. 3 illustrates, for each feature combination, the results obtained for each service and their corresponding KPIs separated in three sub-figures, one for each service category (eMBB, URLLC, mMTC). All scenarios have been simulated with a range of distinct random seeds and the correspondingly obtained 95%-confidence intervals are shown in the sub-figures, revealing a reasonable degree of attained statistical accuracy. Furthermore, the dashed lines indicate the target KPI level for the URLLC and eMBB services.

A. Feature Evaluation

The impact of *numerology* on the performance of the services is shown in Fig. 3 under configurations 1-3. As expected based on the qualitative arguments given above, URLLC services benefit from higher numerologies, in particular $\mu = 2$, due to the shortened TTIs while the eMBB and the mMTC services benefit from lower numerologies i.e. $\mu = 0$ due to the higher frequency-domain channel-adaptive scheduling gains.

Because eMBB and mMTC services benefit from numerology 0, configurations 4-6, in Fig. 3, consider the use of numerology 0 in combination with the three approaches of *mini-slot based scheduling*, respectively, to enhance the performance of the URLLC service. All sub-figures show that configuration 4, which uses the *default approach* for mini-slots, performs better compared to normal slot-based scheduling with numerology 0 (configuration 1). All services benefit from the faster, more resource-efficient and less interfering transmissions enabled by the mini-slots, as qualitatively argued above. In particular, multiple packets can be transmitted in distinct mini-slots within a given slot, whereas under configuration 1 each of those packet transmissions would utilise a full slot by themselves. However, the URLLC sub-figure shows that configuration 4 performs worse than when numerology 2 is used (configuration 3), illustrating the importance of configuring the appropriate numerology.

Configuration 5 considers the use of the *non-pre-emptive mini-slot based scheduling* and the URLLC sub-figure illustrates the gains for the URLLC services compared to the default mini-slot base scheduling (configuration 4). Specifically, the performance is improved because the URLLC packets are immediately transmitted after their arrival in the buffer, assuming that there are enough available resources. Consequently, the performance of the eMBB and mMTC services, which are scheduled only at the start of each slot, is also improved because the buffer is kept short, as shown in the respective sub-figures. Note that the performance of the URLLC service is almost the same compared to only using numerology 2 (configuration 3) while the performance of the eMBB and mMTC services is better than only using numerology 0 (configuration 1).

The use of *pre-emptive mini-slot based scheduling* is considered in configuration 6. The URLLC sub-figure shows that the pre-emptive approach provides further gains compared to the non-pre-emptive approach (configuration 5) because the URLLC packets are transmitted immediately upon arrival, pre-empting an on-going transmission if there are no available

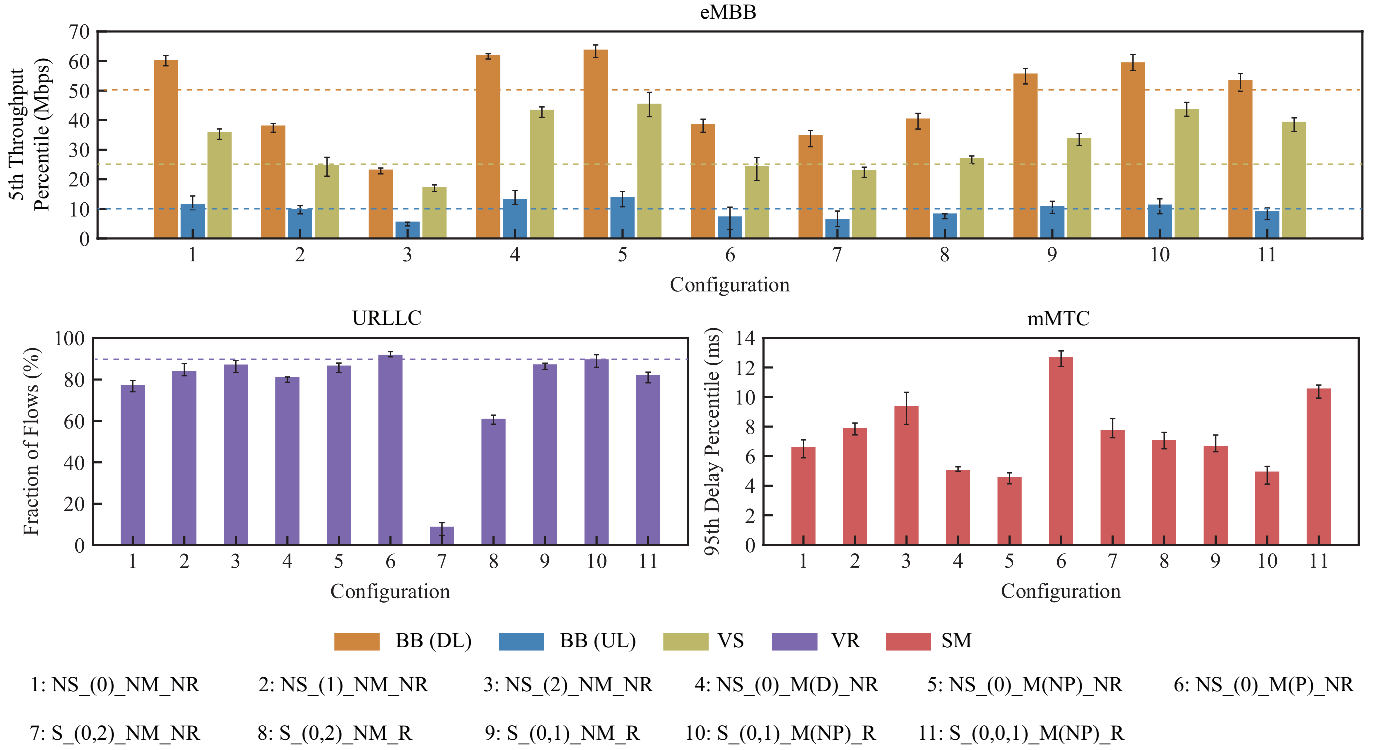

Fig. 3. Performance of the considered services for configurations related to the numerology, mini-slot based scheduling, slicing and resource sharing features and combinations of these features. The dashed lines indicate the target KPI level for each URLLC and eMBB service.

resources. For the same reason, the eMBB and mMTC sub-figures show a significant performance degradation, which is also reflected by the number of eMBB and mMTC packet re-transmissions which is increased by 49% compared to the non-pre-emptive approach. Therefore, among the three mini-slot based scheduling approaches, the non-pre-emptive approach is considered best when taking all services into consideration.

Based on the outcome of the numerology comparisons done in configurations 1-3, configuration 7 assumes two *slices*, one with numerology 0 for the eMBB and mMTC services, and one with numerology 2 for the URLLC service. Fig. 3 shows that the resource split causes an immense performance degradation compared to an unsliced RAN, due to trunking losses, the reduction of usable resources due to the needed inter-numerology guard band (further causing higher interference levels), and the decrease of frequency-domain channel-adaptive scheduling gains as the eMBB data are only scheduled on the resources of the assigned slice.

To decrease the effects of the resource split, configuration 8 in Fig. 3 allows the *sharing of idle radio resources* between the slices. All sub-figures show performance improvement compared to configuration 7, however, configuration 8 performs worse than simply configuring the RAN with numerology 0 (configuration 1), which is due to the guard band effects.

B. Optimization of Feature Combinations

Based on the previous results, an evaluation of some promising candidate feature combinations to find the optimal configuration is presented. First, similarly to configuration 8, we consider configuration 9 but the slice related to the URLLC service is now configured with numerology 1 instead of 2 to reduce the size of the middle and edge guard bands by 28%. In Fig. 3, all sub-figures show that the guard band reduction brings gains to all services, mainly because more resources are now available for data transmissions (in both slices) and secondly because the frequency-selective channel-adaptive scheduling gains are increased in the slice related to the URLLC service. The comparison between configurations 8 and 9 also reveals that choosing the numerology per slice solely based on the service type may not yield optimal performance. Rather the effects of all features should be taken into account when choosing the numerologies. Also, the URLLC sub-figure shows that the performance with configuration 9 is equivalent to that obtained when only numerology 2 is configured (configuration 3) while the eMBB and mMTC sub-figures show that the performance with configuration 9 is similar compared to only configuring numerology 0 (configuration 1). Additionally, the performance of configuration 9 is similar to the performance of configuration 5, with both configurations achieving the targets for the eMBB services and providing a similar performance for the URLLC service.

Subsequently, the combination of two slices, with numerology 0 and 1, idle resource sharing and non-pre-emptive mini-slot based scheduling, which was found to be the best performing mini-slot approach, is considered as configuration 10. Fig. 3 shows that the performance of all services improves, compared to configuration 9, because of the gains that mini-slots bring. Furthermore, with configuration 10, the targets for all services are achieved.

To evaluate the way of slicing, i.e. per numerology or per service category, configuration 11 considers the same feature combination as configuration 10 but with three slices i.e. eMBB, mMTC and URLLC slices, with numerologies 0, 0 and 1, respectively. Fig. 3 shows that configuration 11 performs worse compared to configuration 10 because of the additional trunking losses introduced by the extra slice.

C. Summary

The best overall performance is provided with configuration 10, achieving the KPI target values of all the services. This result supports the commonly argued potential of RAN slicing and highlights that its inherent trunking losses can be compensated by smartly applying available RAN features.

V. Concluding Remarks

In this study we have assessed the merit of distinct 5G RAN features to support an integrated services smart city environment, particularly concentrating on flexible numerology, mini-slots, slicing and idle resource sharing. Performance was shown to be optimised when all RAN features are appropriately combined and configured. This highlights the commonly argued potential of RAN slicing, for which it was shown that the slice-specific numerology should not be solely based on the service type in the respective slices, but rather on the combined effects of all involved RAN features.

In this study we considered traffic with variability on a very fine time scale, which is handled by the scheduler and/or by the idle resource sharing feature. In the future, traffic with high variability over time should be considered and the concept of dynamic resource assignment between slices should be investigated. Furthermore, the design of a more advanced differentiating scheduler is recommended to challenge the need for slicing as a differentiating mechanism.

Acknowledgment

This work is part of NExTWORKx, a collaboration between TU Delft and KPN on future telecommunication networks.

References

[1] B. N. Silva, M. Khan and K. Han, “Towards sustainable smart cities: A review of trends, architectures, components, and open challenges in smart cities”, *Sustainable Cities and Society*, vol. 38, 2018.

[2] 3GPP, TS 38.211, *“NR; Physical channels and modulation”*, v16.3.0, 2020.

[3] 3GPP, TS 38.213, *“NR; Physical layer procedures for control”*, v17.1.0, 2020.

[4] 3GPP, TS 38.912, *“Study on New Radio (NR) access technology”*, v17.0.0, 2022.

[5] S. E. Elayoubi, S. B. Jemaa, Z. Altman and A. Galindo-Serrano, “5G RAN slicing for verticals: Enablers and challenges”, *IEEE Communications Magazine*, vol. 57, no. 1, 2019.

[6] M. Raftopoulou and R. Litjens, “Optimisation of numerology and packet scheduling in 5G networks: To slice or not to slice?”, *Proceedings VTC ’21 (Spring)*, Helsinki, Finland, 2021.

[7] A. Akhtar and H. Arslan, “Downlink resource allocation and packet scheduling in multi-numerology wireless systems”, *Proceedings of WCNC ’18*, Barcelona, Spain, 2018.

[8] K. I. Pedersen, G. Pocovi, J. Steiner and S. R. Khosravirad, “Punctured scheduling for critical low latency data on a shared channel with mobile broadband”, *Proceedings of VTC ’17 (Fall)*, Toronto, Canada, 2017.

[9] H. Yin, L. Zhang, S. Roy, “Multiplexing URLLC traffic within eMBB services in 5G NR: fair scheduling”, *IEEE Transactions on Communications*, vol. 69, no. 2, 2021.

[10] A. A. Zaidi, R. Baldemair, V. Molés-Cases, N. He, K. Werner and A. Cedergren, “OFDM numerology design for 5G new radio to support IoT, eMBB and MBSDN”, *IEEE Communications Standards Magazine*, vol. 2, no. 2, 2018.

[11] T. Bag, S. Garg, Z. Shaik, A. Mitschele-Thiel, “Multi-numerology based resource allocation for reducing average scheduling latencies for 5G NR wireless networks”, *Proceedings of EuCNC ’19*, Valencia, Spain, 2019.

[12] T. Ma. Y. Zhang, F. Wang, D. Wang and D. Guo, “Slicing resource allocation for eMBB and URLLC in 5G RAN”, *Hindawi Wireless Communications and Mobile Computing*, 2020.

[13] L. Feng, Y. Zi, W. Li, F. Zhou, P. Yu and M. Kadoch, “Dynamic resource allocation with RAN slicing and scheduling for uRLLC and eMBB hybrid services”, *IEEE Access*, vol. 8, 2020.

[14] J. Li, W. Shi, P. Yang, Q. Ye, X. S. Shen, X. Li and J. Rao, “A hierarchical soft RAN slicing framework for differentiated service provisioning”, *IEEE Wireless Communications*, vol. 27, no. 6, 2020.

[15] Ericsson, “Mixed numerology in an OFDM system”, R1-165833 3GPP TSG RAN WG1 meeting, 2016.

[16] 3GPP, TS 38.101, *“User Equipment (UE) radio transmission and reception”*, v15.3.0, 2019.

[17] GSMA, *“5G TDD Synchronisation”*, white paper, 2020.

[18] A. Kandoi, “Assessment of Key 5G RAN Features for Integrated Services Provisioning in a Smart City Environment”, MSc thesis, Delft University of Technology, 2022.

[19] 3GPP, TS. 38.901, “Study on channel model for frequencies from 0.5 to 100GHz”, v14.0.0, 2019.

[20] F. Gunnarsson, M. N. Johansson, A. Furuskar, M. Lundevall, A. Simonsson, C. Tidestav and M. Blomgren, “Downtilted base station antennas - A simulation model proposal and impact on HSPA and LTE performance”, *Proceedings of VTC ’08 (Fall)*, Calgary, Canada, 2008.

[21] S. Jaeckel, L. Raschkowski, K. Borner, K. Thiele, F. Burkhardt and E. Eberlein, “QuaDRiGa - Quasi deterministic radio channel generator: user manual and documentation”, *Fraunhofer Heinrich Hertz Institute Technical Report* v.2.4.0, 2020.

[22] 3GPP, TS 38.214, *“NR; Physical layer procedures for data”*, v16.3.0, 2020.

[23] S. Pratschner, B. Tahir, L. Marijanovic, M. Kirev, R. Nissel, S. Schwarz and M. Rupp, “Versatile mobile communications simulation: The Vienna 5G link level simulator”, *EURASIP Journal on Wireless Communications and Networking*, 2018.

[24] K. Sayanam, J. Zhuang and K. Stewart, *“Link performance abstraction based on mean mutual information per bit (MMIB) of the LLR channel”*, 2007.

[25] S. N. Anbalagan, R. Litjens, K. Das, A. Chiumento, P. Havinga and J. L. van den Berg, “A sensitivity analysis on the potential of 5G channel quality prediction”, *Proceedings of VTC ’21 (Spring)*, Helsinki, Finland, 2021.

[26] NGMN *“5G”*, white paper, v1.0, 2015.

[27] J. Morais, S. Braam, R. Litjens, S. Kizhakkekundil and J. L. van den Berg, “Performance modelling and assessment for social VR conference services in 5G radio networks”, *Proceedings of WiMob ’21*, Bologna, Italy, 2021.

[28] NGMN, *“Perspectives on vertical industries and implications for 5G”*, white paper, 2016.

[29] J. Navarro-Ortiz, P. Romero-Diaz, S. Sendra, P. Ameigeiras, J. J. Ramos-Munoz and J. M. Lopez-Soler, “A survey on 5G usage scenarios and traffic models”, IEEE *IEEE Communication Surveys & Tutorials*, vol. 22, no. 2, 2020.

[30] P. Sarigiannidis, M. Louta and A. Michalas, “On effectively determining the downlink-to-uplink sub-frame width ratio for mobile WiMax networks using spline extrapolation”, *Proceedings of Panhellenic Conference on Informatics*, Kastoria, Greece, 2011.

[31] Qualcomm, *“Traffic models for XR”*, R1-2101493 3GPP TSG RAN WG1 e-meeting, 2021.

[32] S.-C. Tseng, Z.-W. Liu, Y.-C. Chou and C.-W. Huang, “Radio resource scheduling for 5G NR via deep deterministic policy gradient”, *Proceedings of ICC ’19*, Shanghai, China, 2019.