Md M. Rashid, M.C. Lucas-Estañ, M. Sepulcre, J. Gozalvez, "5G RAN Slicing to Support Reliability in Industrial Applications", *Proceedings of the IEEE 27th International Conference on Emerging Technologies and Factory Automation (ETFA 2022)*, Stuttgart, Germany, 6-9 September 2022. doi: 10.1109/ETFA52439.2022.9921528. https://ieeexplore.ieee.org/document/9921528

5G RAN Slicing to Support Reliability in Industrial Applications

Md Mamunur Rashid, M. Carmen Lucas-Estañ, Miguel Sepulcre and Javier Gozalvez
UWICORE Laboratory, Universidad Miguel Hernández de Elche (UMH), Elche 03202, Spain
Email: {mrashid, m.lucas, msepulcre, j.gozalvez}@umh.es

Abstract—**Industry 4.0 and 5.0 applications will contribute towards safer, zero-defect and customized production environments. Such applications (e.g. digital twins, collaborative robotics and extended reality) require communication networks capable to satisfy stringent latency, bandwidth, and reliability requirements. Such requirements can be sustained with 5G networks and their evolution that offer unprecedented communications performance and flexibility thanks to the softwarization of networks and the use of network slicing. Network slicing creates different logical partitions or slices of the common network infrastructure and configures each slice to the requirements of the applications it will support. RAN (Radio Access Network) slicing is a fundamental part of network slicing in 5G as the radio channel is prone to errors and this impacts the capacity to support stringent reliability requirements. To date, RAN slices have been created considering the number of radio resources that must be reserved to guarantee the transmission rate or bandwidth demanded by the applications they will serve. This study demonstrates that this design approach cannot guarantee satisfying the reliability requirements of industrial applications and proposes a novel RAN slice descriptor that takes into account both the reliability and transmission rate requirements of the applications.**

***Keywords*— 5G, RAN slicing, Beyond 5G, Industry 4.0, slice creation, slice design, reliability, deterministic, non-deterministic.**

I. INTRODUCTION

Factories are evolving towards more reconfigurable and resilient environments that exploit the digital domain and promote more human-centric industries under the Industry 4.0 and 5.0 paradigms [1]. Applications such as digital twins, mobile and collaborative robots, and extended reality (XR) will facilitate safer, zero-defect and customized production environments, but require communication networks capable to satisfy their latency, bandwidth, and reliability requirements. Such requirements can be sustained with 5G networks and their future evolution that are considered critical enablers for Industry 4.0 and 5.0 applications [2]. 5G and beyond networks offer unprecedented communications performance and flexibility through the softwarization and virtualization of the network and the use of network slicing [3][4]. Network slicing creates different logical partitions or slices of the common network infrastructure, with a slice formed by a set of network functions, computing, storage, networking and radio resources. Each slice can be tailored and configured to support specific applications with distinct QoS (Quality of Service) requirements. Network slices can be created in both the Core Network (CN) and Radio Access Network (RAN). RAN slicing handles the distribution and management of radio resources among slices and is critical to support stringent latency and/or reliability requirements as the radio channel is prone to interferences and errors [5].

RAN slicing designs, creates, and manages the RAN slices along their lifecycle [6]. The lifecycle includes the preparation and commissioning of the RAN slices. The preparation phase evaluates the application requirements to be supported, and designs the slices to support specific QoS profiles. The commissioning phase creates the slices and allocates the radio resources among the slices; this phase is also referred to as partitioning. The design of the slices during the preparation phase is crucial since it is in charge of identifying the number of radio resources that should be allocated to each slice in order to adequately serve the users or applications demanding a particular QoS profile. Most of the existing RAN slicing solutions design RAN slices considering only the users/applications' bandwidth or transmission rate demands [7][8]. However, this approach may not be suitable for serving applications with strict latency and reliability requirements. [9] demonstrated the importance of considering latency requirements to design RAN slices, and proposed a novel latency-sensitive RAN slice descriptor for serving latency-sensitive or time-critical applications. The study reported in [10] also highlights the importance to consider reliability requirements, and proposes a slice admission control and a scheduler to partition the available radio resources among the slices. However, it does not propose a solution to design RAN slices (prior to the partitioning) that accounts for the reliability requirements. In this context, this paper progresses the state-of-the-art by proposing a novel RAN slice descriptor for designing RAN slices also considering the reliability requirements. The descriptor establishes the number of radio resources that should be allocated to each RAN slice (i.e. the size of the RAN slice) considering the transmission rate and reliability requirements of the applications to be served. The conducted evaluations show that the proposed RAN slice descriptor improves the capacity of 5G networks to satisfy the reliability requirements of industrial applications compared to when designing RAN slices only taking into account the application transmission rate requirements. In addition, this paper discusses and illustrates the need to also consider the reliability requirements when defining the shape of a slice. The shape of a slice is a descriptor introduced in [9] that identifies in which slots the RBs reserved for a RAN slice should be located to satisfy the latency requirements of the traffic.

II. RAN SLICE DESCRIPTOR

RAN slices need to be designed considering the characteristics of the traffic generated by the applications to be supported. RAN slices are commonly defined in terms of the number of radio resources that should be allocated to the slice. This slice descriptor is referred to as the size of the RAN slice. To date, the size of the slices is generally defined based on the transmission rate or bandwidth demanded by the applications. In this section, we present a novel analytical model to calculate the size of a RAN slice considering not only the transmission rate requirement but also the reliability requirements of the traffic to be supported in the slice.

Following [9], we estimate the size K_s of a slice s as the sum of the number of radio resources required by all the UEs that will be served by the slice. Without loss of generality, we consider a 5G network, and a radio resource is given by an RB (12 subcarriers in frequency) in one transmission slot. Let's consider that the RAN slice should serve a set of M UEs with similar QoS requirements. Each UE u (with u=1,…,M) requires $J_u(\gamma_u)$ RBs, and this number depends on the experienced signal-to-interference-plus-noise ratio (SINR) represented by γ_u. This is the case because low SINR conditions require higher error protection levels than high SINR conditions, and hence more RBs are needed to transmit a packet of a given size. The size of the slice s can then be expressed as:

$$K_s=\sum_{u=1}^{M} J_u(\gamma_u) \tag{1}$$

$J_u(\gamma_u)$ in (1) depends on the transmission rate R_u demanded by the UE u and the effective transmission rate R_u^{eff} per assigned RB that will be experienced by the UE u. R_u is the amount of data generated by the application that the UE u has to transmit per unit of time measured in bits per second (bps). R_u^{eff} represents the amount of application data that the UE u can transmit per RB expressed in bits and considering that it depends on γ_u. $J_u(\gamma_u)$ is estimated in [9] as:

$$J_u(\gamma_u) = \left\lceil R_u / R_u^{eff}(\gamma_u) \right\rceil \tag{2}$$

R_u and R_u^{eff} are estimated considering the characteristics and requirements of the traffic class. As explained next, we consider non-deterministic and deterministic periodic traffic classes that are characteristic of industrial applications [11]. In addition, while [9] only considers the transmission rate requirements, this study proposes to also include the reliability demanded by the application.

A. Non-deterministic traffic

Non-deterministic traffic does not have latency requirements and is usually characterized by a demanded transmission rate. This traffic class is characteristic, for example, of applications related to software updates or file downloads in process automation or motion control among others [11].

The size of the slice for non-deterministic traffic is calculated as the number of RBs that must be reserved within a time window of duration T_w in order to satisfy the transmission rate required by the UEs that will be served by the slice. For non-deterministic traffic, R_u is the minimum transmission rate requested by the UE u. To estimate R_u^{eff}, we consider: 1) the transport block size or number of data bits that can be transmitted in one RB based on the experienced γ_u, which is represented as $TBS(\gamma_u)$, and 2) the number of times a packet has to be retransmitted to achieve the reliability required by the application. First, we obtain the $TBS(\gamma_u)$ that depends on the modulation and coding scheme (MCS) used to achieve a target BLER as a function of the experienced γ_u for the UE u. With high values of γ_u, it is possible to use an MCS with low error protection and high spectral efficiency, i.e., that allows to transmit a high number of data bits per RB. Once we know $TBS(\gamma_u)$, we calculate the amount of data transmitted per RB in the time window T_w as $TBS(\gamma_u)/T_w$; this represents the transmission rate of UE u per assigned RB. To estimate R_u^{eff}, we also consider the number of retransmissions that need to be performed to achieve the required reliability P_{rel}. The reliability is defined as the percentage of packets that are correctly received. The reliability achieved after n retransmissions of a packet is given by $1\text{-}BLER^{(n+1)}$. This reliability must be higher than the reliability requirement P_{rel} to satisfy the requirements of the applications. For a given $BLER$, we can then calculate the maximum number of retransmissions, n_{max}, that a packet would require to satisfy the required reliability (i.e. $1\text{-}BLER^{(n_{max}+1)} \geq P_{rel}$) using the following equation:

$$n_{max}=\left\lceil \frac{\log_{10}(1\text{-}P_{rel})}{\log_{10} BLER} \right\rceil \text{-}1 \tag{3}$$

where $\lceil\cdot\rceil$ denotes the ceil operator. The average number n_{avg} of retransmissions performed for each packet is given by:

$$n_{avg}=\sum_{i=1}^{n_{max}} BLER^i \tag{4}$$

R_u^{eff} is finally defined as:

$$R_u^{eff}(\gamma_u)=\frac{1}{1+n_{avg}} \cdot \frac{TBS(\gamma_u)}{T_w} \tag{5}$$

B. Deterministic periodic traffic

Deterministic traffic requires data to be delivered to the receiver within a maximum latency deadline. Deterministic periodic traffic is the most common industrial traffic [11], and can be found in motion control, process automation or control to control communication applications among others.

Deterministic periodic traffic is characterized by a transmission period T_p that represents the time between two consecutive packets, the size of the packet to be transmitted by UE u given by L_u, the latency deadline D, and the reliability requirement P_{rel}. P_{rel} is defined as the percentage of packets that are correctly received within the latency deadline. The size of the slice for deterministic periodic traffic is defined as the number of RBs within a transmission period that must be reserved to satisfy the transmission rate required by the UEs. For this traffic class, the transmission rate required by a UE u can be calculated based on the packet size L_u and the latency deadline D as:

$$R_u = L_u / D \tag{6}$$

To estimate the effective transmission rate R_u^{eff}, we consider the transport block size $TBS(\gamma_u)$ and the average number of retransmissions needed to achieve the required reliability P_{rel}. R_u^{eff} is calculated like in (5) but considering the latency deadline D instead of T_w. R_u^{eff} for deterministic periodic traffic is then estimated as:

$$R_u^{eff}(\gamma_u)=\frac{1}{1+n_{avg}} \cdot \frac{TBS(\gamma_u)}{D} \tag{7}$$

n_{avg} is calculated based on the reliability requirement P_{rel} using (3) and (4). From (6) and (7), the number of RBs $J_u(\gamma_u)$ required by user u can be expressed for deterministic periodic traffic as:

$$J_u(\gamma_u) = \left\lceil \frac{L_u}{TBS(\gamma_u)} \cdot (1+n_{avg}) \right\rceil \tag{8}$$

Finally, the size of the slice for deterministic periodic traffic is calculated using (1) and (8).

III. EVALUATION

This section evaluates by simulation the reliability that can be achieved when RAN slices are designed using the descriptor proposed in this study that jointly considers the transmission rate and reliability requirements of the traffic to be supported to define the size of a slice. The performance is compared against that achieved with a reference descriptor that defines the size of a RAN slice only based on the required transmission rate as proposed in studies such as [12], [13] and [7]. We emulate an industrial scenario with a control-to-control application covered by a single 5G cell. Following [11], control-to-control communications typically generate deterministic periodic traffic[1]. We consider that there are 10 nodes or UEs in the cell, and each one generates periodic packets with L_u=1 KByte with a latency requirement of 10 ms [11]. We consider an MCS that results in that a UE u requires 10 RBs to transmit one packet, i.e., $\lceil L_u/TBS(\gamma_u)\rceil=10$.

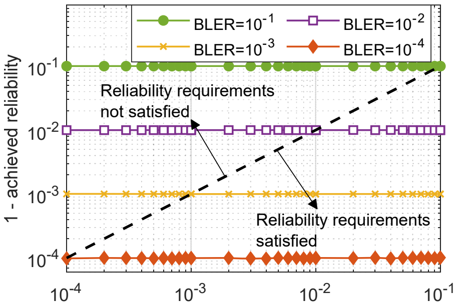
Fig. 1 shows the reliability that can be achieved when the size of the RAN slice is designed with the reference descriptor (Fig. 1a) and the proposed one (Fig. 1b). In this figure, the achieved reliability is depicted as a function of the target reliability and for different values of the BLER. Fig. 1a shows that it is not possible to satisfy the target reliability when the BLER is higher than 1-P_{rel} when the slice is designed only based on the transmission rate requirement (i.e. the reference descriptor). This is because the number of RBs reserved for the slice is not sufficient to perform the number of retransmissions needed to achieve the target reliability. In this case, the achievable reliability is bounded by the BLER experienced in the radio channel. Fig. 1b shows that the target reliability can always be met when using the proposed descriptor (that jointly considers the transmission rate and reliability requirements). This result is achieved at the expense of reserving a higher number of resources for the slice. Fig. 2 shows the relative increase (in percentage) of the number of RBs reserved for the slice with the proposed RAN slice size descriptor with respect to the reference descriptor. Results in Fig. 1 and Fig. 2 show that, for example, the reference descriptor is not capable to satisfy a reliability requirement of 0.9999 when the BLER is 0.1. On the other hand, the proposed descriptor can satisfy it by reserving just 20% more resources for the slice than the reference descriptor. When the BLER decreases to 0.01, the percentage of additional resources reserved for the RAN slice decreases to 10% while the 0.9999 target reliability is still satisfied by the proposed descriptor but not the reference one.

IV. DISCUSSION & FUTURE WORK

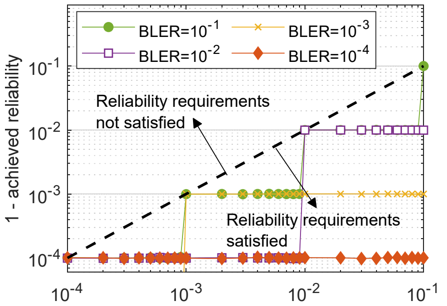
This study has demonstrated that it is possible, and necessary, to design 5G RAN slices that effectively account for the reliability requirements of industrial applications. This study has focused on the design of the size of a RAN slice (i.e. the number of radio resources included in a slice), and our future work will look into embedding reliability requirements in the definition of the shape of a slice. The shape of a slice is a descriptor introduced in [9] that identifies in which slots the RBs reserved for a RAN slice should be located to satisfy the latency requirements of the traffic. In this section, we discuss about the need to also consider the reliability requirement of the traffic in the definition of the shape of a slice.

Non-deterministic traffic does not have latency requirements. In this context, the K_s RBs that need to be reserved for the slice can be located in any slot within the time window T_w, and the shape of the RAN slice for non-deterministic traffic is defined as in [9]:

$$\sum_{t=1}^{T_w} L_t = K_s \tag{9}$$



(a) Reference descriptor.



(b) Proposed descriptor.

Fig. 1. Achieved reliability as a function of the target reliability for a RAN slice designed with the reference (a) and proposed (b) RAN slice descriptors.

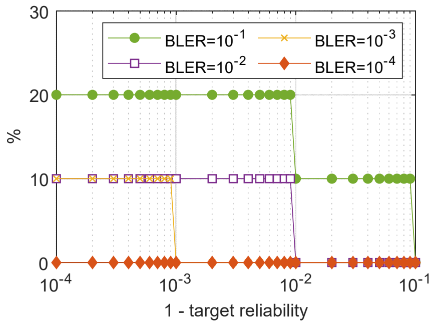


Fig. 2. Relative increase of the number of radio resources reserved for the slice with the proposed descriptor with respect to the reference descriptor.

where L_t is the number of RBs allocated to the slice in slot t. Since non-deterministic traffic does not have latency requirements, the reliability will not affect the shape of the slice.

For deterministic traffic, packets need to be received before a latency deadline D. The shape of a slice must then be such that the RBs reserved for the slice are located in slots between the generation of a packet and the latency deadline.

---

[1] The evaluation scenario considers deterministic periodic traffic. However, we should note that similar trends and conclusions as those reported in this section have been observed for non-deterministic traffic since the size descriptor follows a similar expression.

If t_z represents the time at which a packet z is generated and D_{slots} represents the latency deadline expressed as an integer number of slots, the shape of a slice for deterministic periodic traffic was defined in [9] as:

$$\sum_{t=t_z}^{t_z+D_{slots}-1} L_t = K_s \qquad (10)$$

where $t_z \in \{t_1, t_1{+}T_p, t_1{+}2T_p, \ldots\}$. The shape descriptor in [9] does not consider the reliability requirement of the traffic, and slices designed following the size and shape descriptors in (1) and (10) may not be able to satisfy the reliability requirement of an application. To illustrate this, let's consider the examples shown in Fig. 3. Fig. 3 illustrates the RBs allocated to 3 slices that serve 10 UEs that generate deterministic periodic traffic with a transmission period of T_p. The 10 UEs generate packets in t_z with a latency deadline D and required reliability of 90%. The 3 slices have a size of 12 RBs (calculated using (1) and (8)) to serve the 10 UEs based on their transmission rate and reliability requirements. The 3 slices also satisfy the shape constraint that establishes that the RBs must be reserved before the latency deadline D within the transmission period. The stripped shaded squares represent the RBs where the packets of the 10 UEs are transmitted. In each example, two packets are received with error (those with a red mark) and need to be retransmitted. In the first slice, the packets received with error cannot be retransmitted because the latency deadline cannot be satisfied. In the second slice, the packets received with error cannot be retransmitted because there are no RBs reserved after the time needed to request and prepare the retransmission (represented as t_{rtx} in the figure). Only the design of the third slice can handle the retransmission of packets received with errors using the RBs represented as shaded squares with vertical stripes. Although the 3 slices had sufficient resources to allocate the required packet retransmissions, only the third slice meets the latency and reliability requirements of the traffic. These examples have highlighted the need to also consider the required reliability when designing the shape of the slice.

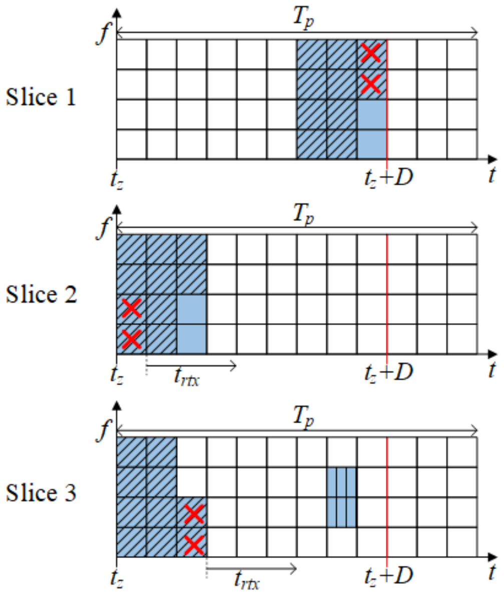


Fig. 3. Example of 3 RAN slices for deterministic periodic traffic with a size of K_s=12 RBs and a shape that satisfies a latency deadline D. Each square represents an RB in one transmission slot, and shaded squares represent RBs reserved for the slice but not used. Stripped shaded squares represent RBs that are part of the slice and are used for packet transmissions while stripped shaded squares with a red cross represent packets received with error. Finally, shaded squares with vertical stripes represent RBs used for packet retransmissions.

V. CONCLUSIONS

This paper has presented a novel slice descriptor that incorporates reliability requirements into the design of 5G RAN slices. The proposed RAN slice descriptor establishes the size or number of RBs that need to be reserved for a RAN slice considering both the reliability and transmission rate required by the traffic to be supported in the RAN slice. This study has shown that the use of the proposed descriptor allows to design 5G RAN slices that can satisfy the strict reliability requirements of industrial traffic. In addition, this paper has discussed and illustrated the importance of also considering reliability requirements to decide the shape of a RAN slice, i.e. how the reserved resources within a slice are distributed in time. This will be critical for the support of industrial applications that simultaneously require stringent reliability and latency requirements, and is hence the focus of our future work.

ACKNOWLEDGMENT

This work has been funded by MCIN/AEI/10.13039/501100011033 through the project PID2020-115576RB-I00, and FSE funds through the grant PRE2018-084743, by the Generalitat Valenciana through the project CIGE/2021/096, and by a research grant awarded by the Vicerrectorado de Investigación of the UMH (2022).

REFERENCES

[1] P. K. R. Maddikunta, et al. "Industry 5.0: A survey on enabling technologies and potential applications", *Journal of Industrial Information Integration*, vol. 26, 2022, 100257.

[2] 5G-ACIA, *Our view on the Evolution of 5G towards 6G*, 5G-ACIA Position Paper, May 2021.

[3] 3GPP TS 22.261 v16.8.0, "Service requirements for the 5G system; Stage 1 (Rel. 16)", June 2019.

[4] A. Ksentini, N. Nikaein, "Toward enforcing network slicing on RAN: Flexibility and resources abstraction", *IEEE Communications Magazine*, vol. 55, no. 6, pp. 102–108, June 2017.

[5] B. Khodapanah, et al., "Framework for Slice-Aware Radio Resource Management Utilizing Artificial Neural Networks", *IEEE Access*, vol. 8, pp. 174972-174987, 2020.

[6] 3GPP TS 28.530 v15.1.0, "Management and Orchestration; Concepts, Use Cases and Requirements (Rel. 15)", Dec. 2018.

[7] I. Vilà, et al., "An Analytical Model for Multi-Tenant Radio Access Networks Supporting Guaranteed Bit Rate Services", *IEEE Access*, vol. 7, pp. 57651-57662, 2019.

[8] B. Khodapanah, et al., "Framework for Slice-Aware Radio Resource Management Utilizing Artificial Neural Networks", *IEEE Access*, vol. 8, pp. 174972-174987, 2020.

[9] J. García-Morales, M. C. Lucas-Estañ, J. Gozalvez, "Latency-Sensitive 5G RAN Slicing for Industry 4.0", *IEEE Access*, vol. 7, pp. 143139-143159, 2019.

[10] T. Guo, A. Suárez, "Enabling 5G RAN Slicing With EDF Slice Scheduling", *IEEE Transactions on Vehicular Technology*, vol. 68, no. 3, pp. 2865-2877, March 2019.

[11] 3GPP TR 22.804 v16.2.0, "Study on Communication for Automation in Vertical Domains (Rel. 16)", December 2018.

[12] R. Ferrus, et al., "On 5G Radio Access Network Slicing: Radio Interface Protocol Features and Configuration", *IEEE Communications Magazine*, vol. 56, no. 5, pp. 184-192, May 2018.

[13] R. Kokku, et al., "CellSlice: Cellular wireless resource slicing for active RAN sharing", in *Proc. of Fifth International Conference on Communication Systems and Networks (COMSNETS)*, 2013, pp. 1-10