IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received December 5, 2019, accepted December 17, 2019, date of publication December 23, 2019, date of current version January 2, 2020.

Digital Object Identifier 10.1109/ACCESS.2019.2961658

System-Level Study of Data Duplication Enhancements for 5G Downlink URLLC

MARCO CENTENARO[1], (Member, IEEE), DANIELA LASELVA[2], JENS STEINER[2], KLAUS PEDERSEN[2,3], (Senior Member, IEEE), AND PREBEN MOGENSEN[3,2], (Member, IEEE)

[1]Fondazione Bruno Kessler, 38123 Trento, Italy
[2]Nokia Bell Labs, 9220 Aalborg, Denmark
[3]Department of Electronic Systems, Aalborg University, 9220 Aalborg, Denmark

Corresponding author: Marco Centenaro (mcentenaro@fbk.eu)

This work was supported in part by the framework of the H2020 project ONE5G, funded by the European Union under Grant ICT-760809.
M. Centenaro started contributing to this work when he was with by the Department of Electronic Systems, Aalborg University.

ABSTRACT Data duplication is studied as a fundamental enabler for ultra-reliable low-latency communication (URLLC) in fifth-generation cellular systems. It entails the simultaneous usage of multiple radio links delivering redundant data between a terminal and the network to boost the transmission reliability. However, the improved reliability comes at a cost of reduced spectral efficiency, since the transmission of multiple instances of the data message on different links occupies more radio resources as compared to sending only one instance using a single link. It is therefore crucial to improve the performance of data duplication schemes, with the aim of reducing the radio resource consumption without degrading the reliability gain provided by this transmission paradigm. In this paper, we propose several methods to increase the downlink URLLC capacity supported by data duplication in fifth-generation cellular networks based on the New Radio standard. A single-user analytical model is derived to evaluate a combination of the proposed enhancements. The most promising solution, namely selective data duplication upon failure which entails a massive reduction of the overall number of duplicate transmissions, is finally evaluated by means of extensive multi-user system-level simulation campaigns. The simulation results with background mobile broadband traffic show that, in the investigated scenario, the proposed solution with 4 Mbps offered URLLC traffic outperforms the baseline approach for data duplication with 1 Mbps offered URLLC traffic, thus increasing the amount of URLLC user equipments that can be effectively sustained by the network.

INDEX TERMS URLLC, downlink, PDCP duplication, dual connectivity, resource efficiency, system-level simulations, standardization, 3GPP, NR.

I. INTRODUCTION

The fifth-generation (5G) cellular systems [1], [2] standardized by the Third Generation Partnership Project (3GPP)[1] aim at supporting a wide variety of traffic types, each of which has different quality of service (QoS) requirements characterized by distinct key performance indicators (KPIs). We can divide the envisioned traffic types in three well-established categories [3]: i) enhanced mobile broadband (eMBB), ii) massive machine-type communication (mMTC), and iii) ultra-reliable low-latency communication (URLLC).

The associate editor coordinating the review of this manuscript and approving it for publication was Mostafa Zaman Chowdhury.

[1]https://www.3gpp.org

In this paper, we focus on URLLC, which is expected to be the key enabler of industrial use cases, commonly referred to as industrial Internet of Things (IIoT). The QoS requirements of URLLC consist of an upper bound on packet errors of 10^{-5} and a delivery delay up to 1 ms. These demands are satisfied by the New Radio (NR) standard [2] of 3GPP, which overcomes the limitations of the legacy Long-Term Evolution (LTE) standard [1]. The NR air interface is characterized by a redesigned and flexible transmission time interval (TTI) structure and reduced processing times to meet the low-latency constraint as well as methods to achieve the reliability requirement like, e.g., antenna diversity [4] and multi-connectivity [5]. The aim of this study is to investigate the optimal design of reliability-oriented multi-connectivity radio protocols; they entail the transmission of replicas of the same data across multiple wireless links between a mobile terminal and the network to maximize the probability of successful packet delivery over the radio interface. These links may involve distinct base stations or component carriers (CCs) within the same base station, or combinations of the two approaches. In particular, here we address the case of multi-connectivity in which downlink (DL) data are duplicated at higher layers, i.e., at the Packet Data Convergence Protocol (PDCP) layer, and transmitted by two distinct base stations towards terminals in an uncoordinated fashion; this scheme is referred to as dual connectivity (DC)-based DL PDCP data duplication. An obvious gain in terms of packet reliability at no latency cost, as compared to other techniques such as, e.g., hybrid automatic repeat request (HARQ), and with limited computational effort with respect to, e.g., coordinated multi-point (CoMP) transmissions as defined for LTE systems [6] or multiple transmission-reception-point (multi-TRP) schemes being defined for 5G systems [7], is provided by this functionality. On the other hand, such reliability gain can be severely limited by the capacity reduction due to the transmission of redundant packets which increase the radio-resource consumption as well as interference, potentially outweighing the benefits of data duplication. Therefore, enhancements aiming at resource-efficient data duplication should be adopted in order to avoid wasting radio resources, e.g., in unnecessary redundant packet transmissions. This study proposes several methods that improve the spectral efficiency of DL data duplication in 5G NR systems, with the aim of increasing the supported URLLC load. A thorough system-level analysis of the most promising solution, namely *selective data duplication upon failure*, shows significant radio efficiency gains with respect to the baseline approach in the considered reference scenario.

The rest of this paper is organized as follows. In Section II, we present the related work on URLLC and reliability-oriented DC, highlighting the novel contributions of this study and the adopted evaluation methodology. In Section III, the system model based on a state-of-the-art reference scenario and 3GPP-compliant data duplication is introduced. The problem formulation is provided in Section IV, while three proposals to improve the spectral efficiency of data duplication are described in Section V. A single-user analytical model to evaluate the performance of the proposed enhancements is derived in Section VI; the model is exploited to identify the most promising enhancement among the proposed ones. Extensive results from system-level simulations campaigns are then provided in Section VII to validate the gains of the identified primary enhancement. Finally, we draw the conclusions of this work and formulate recommendations for the possible future research in Section VIII.

II. RELATED WORK AND NOVEL CONTRIBUTIONS

A large body of work has appeared on URLLC in literature, after the seminal work by Popovski [8]. Research efforts have been devoted to enhancing various communication aspects, yielding, e.g., grant-free wireless access protocols [4] and scheduling algorithms based on mini-slots and preemption [9]–[11] of radio resources that provide effective support of URLLC in single connectivity (SC). However, the exploitation of different sources of diversity was identified as a promising enabler to achieve the challenging reliability and latency requirements [12]. One of these sources is multi-connectivity, which enables the transfer of redundant data through multiple wireless links belonging to the same radio-access technology (RAT) (either operating on licensed or unlicensed frequency bands [13]–[15]) or to different ones [16], [17]. Here, we focus on the former type of multi-connectivity leveraging 3GPP technologies by considering two active connections in the DL direction, i.e., DC towards terminals from two serving base stations. There are several possible implementations of reliability-oriented DC, depending on the layer which performs the duplication. Indeed, data can be duplicated at the physical layer, exploiting CoMP- or multi-TRP-based approaches like joint transmission [18], or at higher layers [19]. The main difference between these two approaches resides in the degree of coordination between the two active base stations. In physical-layer duplication the base stations precisely synchronize their physical transmissions of a packet, while the terminal combines the received packets at the physical layer. Such a tight synchronization between transmission-reception points (TRPs) requires a high-speed backhaul connection, making the network deployment more expensive. On the other hand, in higher-layer duplication, a packet is duplicated at the PDCP layer and undergoes independent processing at each active base station. The layers below PDCP operate in an independent fashion, thus the degree of coordination between TRPs in this case is lower. In fact, the duplicated packets may eventually be transmitted at different time instants, over different frequency resources, and with different modulation and coding schemes (MCSs). The network deployment cost is also lower than having physical-layer duplication. At the receiver side, the user equipment (UE) discards replicas in case the same packet is correctly received multiple times.

In this study, we focus on DC-based PDCP data duplication enhancements, which are being standardized by 3GPP for an effective support of IIoT applications [20] – as further discussed in Sec. III-C. Various studies in literature raised several technical challenges, mainly due to the increased radio-resource consumption caused by redundant packet transmissions [21], [22], but also dealing with further network management aspects [23]. Inefficiencies of DC-based data duplication with respect to physical-layer duplication have been also proven via measurement campaigns in industrial scenarios [24]. The observed negative impact of data duplication on system-level performance triggered the investigation of suitable system configurations [21], [25], dynamic link activation heuristics [26], and predictive flow control strategies [27], with the aim of achieving as wide as possible support of URLLC exploiting DC-based data duplication. Nevertheless, there is margin for further improvements of reliability-oriented DC in 5G cellular systems.

The contribution of this study consists of several novel enhancements for DC-based data duplication in DL with the aim of reducing the radio-resource utilization, thus increasing the amount of URLLC terminals that can be supported by the network. We incorporate the state of the art in the matter of optimal DC configuration, and focus on designing new approaches that inherently reduce the impact of duplicate packet transmission from the involved base stations towards the terminals. We derive an analytical model to evaluate the proposed approaches in a single-user scenario and identify the most promising one, which we refer to as selective data duplication upon failure. The identified approach is finally evaluated by means of system-level simulation campaigns conducted under highly realistic conditions, where all the major performance-determining factors are included for a dynamic multi-cell/multi-user system with time-variant traffic.

III. SYSTEM MODEL

In this section, we first describe the reference scenario, then we introduce the terminology and notation of DC-based data duplication in current 3GPP systems, including a brief discussion on the ongoing standardization activities of multi-connectivity for IIoT.

A. REFERENCE SCENARIO

We consider a heterogeneous network (HetNet) scenario as reference scenario. It consists of a small-cell network layer assisting the macro network layer in enhancing the coverage and capacity of the UEs, most of which will detect two possible connections, one towards a macro base station and one towards a small-cell base station. In particular, UEs near the edge within a small cell are likely to experience a similar received signal strength as from the macro cell if high cell-range extension (CRE) offsets are adopted for load balancing between the macro and the small-cell network layer. Therefore, these UEs are expected to benefit from the DC-based data duplication [23]. On the other hand, UEs that already have a good-quality connection towards the serving cell do not need to operate with DC because the redundant transmission on the worse link would be useless, thus they can rely on SC and save precious radio resources.

In order to properly tune the amount of URLLC UEs in DC with the aim of reducing the amount of radio resources spent for duplicated packet transmissions, the `DC_Range` parameter was introduced in [21]. This parameter conditions the association of a UE attached to a small cell also to a node within the macro layer only if the inequality

$$\mathtt{RSRP}_{\text{macro}}^{[\text{dBm}]} + \mathtt{DC\_Range}^{[\text{dB}]} > \mathtt{RSRP}_{\text{small}}^{[\text{dBm}]} + \mathtt{CRE}^{[\text{dB}]}, \tag{1}$$

is satisfied, where $\mathtt{RSRP}_x$ is the reference-signal received power (RSRP) by the UE from the best macro and small cell, respectively. In other words, the UE is allowed to have two active connections only for an RSRP difference between the macro layer and the small-cell layer in the interval $[\mathtt{CRE}^{[\text{dB}]} - \mathtt{DC\_Range}^{[\text{dB}]}, \mathtt{CRE}^{[\text{dB}]}]$. This results in allowing data duplication only for UEs that are at the small cell edge, providing the network with the control of the extension of such "data-duplication area" through the `DC_Range` parameter. The results of [21] show that this approach effectively reduces the amount of overall duplicate transmissions by enforcing data duplication only to a selected group of UEs.

An illustration of the considered reference scenario, explaining also how the `DC_Range` parameter works, is provided in Fig. 1. It can be seen that, as `DC_Range` increases, a greater amount of UEs falls in the highlighted area, where DC is enabled.

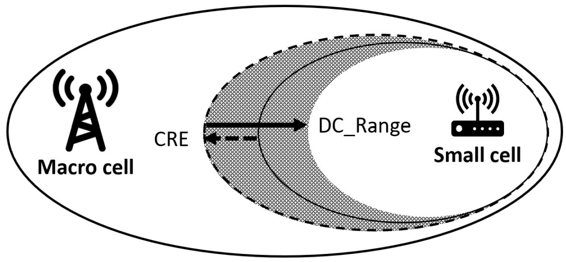


FIGURE 1. Configuration of DC exploiting the `DC_Range` parameter [21].

B. DATA DUPLICATION BASED ON DUAL CONNECTIVITY

DC-based data duplication can be enabled in the considered HetNet scenario for all those UEs which are in the coverage range of small cells, because they may exploit the assistance of the macro layer to improve their transmission reliability via DC. The activation of this feature is determined on a per-UE basis depending on whether (1) is fulfilled or not. In the following, we described the considered 3GPP-compliant architecture for DC-based data duplication exploiting NR links, i.e., NR/NR DC (NR-DC).[2]

The duplication of data packets takes place at the PDCP layer [29], [30] as shown in Fig. 2, which provides an illustration of the NR-DC user plane architecture in Release-15 3GPP systems for the DL transmission case. The picture shows a target UE, which is connected simultaneously to two next-generation NodeBs (gNBs), namely a master gNB (MgNB), which is in charge of managing the DC operations, and a secondary gNB (SgNB) that assists the MgNB. In the considered HetNet scenario, we can map the MgNB and the SgNB to the serving macro cell and small cell, respectively. The wireless link between the MgNB (SgNB) and the UE is denoted as primary link (secondary link).[3] We can see that a PDCP protocol data unit (PDU) created by the MgNB contains as service data unit (SDU) a datagram coming from the serving user-plane function (UPF). The PDCP PDU is

[2]We remark that DC-based PDCP duplication exploiting LTE and NR links jointly is supported in 3GPP as well [28], but is not treated here.

[3]In other contexts, the MgNB (SgNB) is also called "PDCP hosting node" or "primary node" ("assisting node" or "secondary node"). The term "legs" may also be used in place of "links."

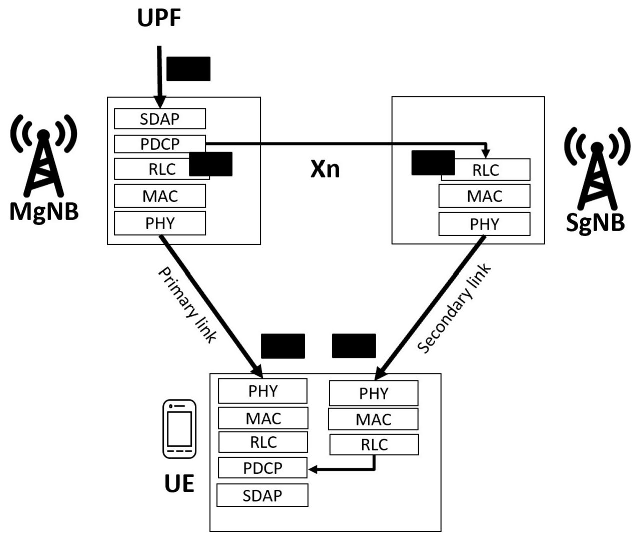


FIGURE 2. DC-based PDCP duplication in Release-15 3GPP NR systems.

duplicated and the duplicate is forwarded via the Xn interface to the SgNB over a tunnel between bearers with a URLLC QoS target. In this way, two copies of the same PDCP PDU may be sent down to lower layers on distinct gNBs, for the physical data transmission of the corresponding transport block (TB) which are independently handled by the two schedulers. At the UE side, the received data from the two gNBs are treated independently until they reach the PDCP layer, when the device eventually realizes that multiple copies of the same PDCP PDU were transmitted based on their sequence number, and can eventually discard redundant received copies. Let us observe that, according to the latest 3GPP specifications regarding the configurations suitable to URLLC support, a 30-kHz subcarrier spacing (SCS) and 2-symbol TTI (i.e., mini-slot) are needed for a 5G NR system in order to accommodate two HARQ transmission attempts per link within the 1-ms maximum latency [34], [35]. Therefore, in the best case of PDCP duplication both the first transmission attempts from the two legs are successful, and two transmission attempts are performed overall.

We remark that the UE has no means to identify packet duplicates at lower layers. Moreover, at the time of its transmission the MgNB may be unaware of the transmission at the SgNB and its outcome, and vice versa, because of the independent scheduling of transmissions on the primary link and secondary link. Thus, it may happen that a leg that experienced a transmission failure is not aware that the transmission over the other leg was successful; in these cases, it would continue its regular operations unnecessarily. For example, the leg would try a HARQ retransmission to attempt recovering the failed packet despite the same packet was already successfully received by the UE. To prevent an inefficient utilization of resources when a PDCP PDU is successfully transmitted by one of the two gNBs, some mechanisms for duplicate discarding based on signaling exchange over the Xn network interface between the MgNB and the SgNB are present in Release-15 3GPP systems [29], [30]. A first mechanism entails the SgNB providing the MgNB with the flow control report indicating, for instance, the PDU with the highest sequence number among all the PDUs successfully delivered to the UE, and the PDU with the highest sequence number among the PDUs successfully delivered in sequence. This allows to determine the PDUs that can be safely discarded at the primary node. With a second mechanism, the MgNB provides the SgNB with an explicit discard indication for those sequence numbers that have been acknowledged by the UE.

C. RELATED STANDARDIZATION ACTIVITIES

DC as a feature of 3GPP systems was already introduced in previous releases of the LTE standard [1] to boost the network capacity for eMBB traffic thanks to data split in multiple simultaneous traffic flows. With the introduction of the NR standard, Release-15 3GPP systems can exploit this feature to improve the data reliability thanks to PDCP duplication. Currently, maximum two Radio Link Control (RLC) entities can be associated to the PDCP entity in control of duplication, thus two copies of the same PDCP PDU can be transmitted. The secondary RLC entity can belong to a distinct base station as in DC-based PDCP duplication, or to the same base station, yielding carrier aggregation (CA)-based PDCP duplication. In the latter case, the two RLC entities are assigned to distinct CCs to ensure that the two copies are transmitted on distinct wireless links, preserving the diversity gain.

At the time of writing, 3GPP is investigating enhancements to PDCP duplication in the context of Release-16 activities, including both DC- and CA-based PDCP duplication, in both DL and uplink (UL). A 3GPP study item on IIoT [31], [32] was dedicated to researching improvements targeting URLLC, considering both sub-6-GHz frequency bands and millimeter waves (frequency range (FR) 1 and FR 2, respectively [33]), as well as time-division duplex (TDD) and frequency-division duplex (FDD), with the Release-15 solutions as a baseline. As an outcome of the study phase, various PDCP duplication enhancements have been identified as beneficial for Release-16 3GPP systems in order to effectively support IIoT in NR [20]. The specific areas which are currently being investigated/discussed include i) the support of up to 4 copies combining DC with CA to further boost reliability, ii) the support of more dynamic control of duplication in the UL, and iii) radio efficiency enhancements of DL PDCP duplication. The contribution of this manuscript is related to the third objective, with specific focus on the DC configuration of DL PDCP duplication.

IV. PROBLEM DESCRIPTION AND KPIS

In the rest of this paper, let us denote Release-15 PDCP duplication as "blind" PDCP duplication, since the two copies of each PDCP PDU for a UE configured with DC are managed by the two serving cells without exchanging timely information regarding the respective transmission status.

In the previous section, we observed that at least two transmission attempts are performed in blind PDCP duplication. On the other hand, if one of the two first transmissions fails, the overall amount of transmitted TBs grows to three. Finally, four transmissions are needed in case both the first transmission attempts from the two legs cannot be decoded by the UE. In essence, most of the radio-resource utilization of blind PDCP duplication is due to the two initial first transmissions, one on the primary link and one on the secondary link. In fact, both PDCP PDU duplicates are almost always successfully conveyed to the UE, especially if a low block error rate (BLER) target is utilized for the link adaptation. For instance, if a 1% BLER target is set for a TB transmission on a given wireless link, a very reliable MCS is selected such that on average only a single packet is lost on that link out of one hundred transmitted packets. Therefore, a large amount of radio resources is wasted in unnecessary transmissions of redundant data.

Another issue of blind PDCP duplication is the *queuing delay* caused by duplicated PDCP PDUs at the SgNB to other high-priority data flows that also need to be transmitted from the SgNB. Indeed, the packet scheduler of the SgNB has to multiplex its own URLLC packets with duplicated PDCP PDUs forwarded by the MgNB, likely increasing the average queuing time of URLLC packets and eventually causing some of these packets to exceed the URLLC maximum latency budget.

The solutions proposed in the related work to tackle the aforementioned issues are not satisfactory. The `DC_Range` parameter represents a way to limit the number of UEs in DC, so that only a fraction of the UEs can benefit from PDCP duplication, but does not bring any enhancement to the blind PDCP duplication mechanism itself. On the other hand, in-network discarding mechanisms proposed in 3GPP tackle the reduction of the number of duplicate transmissions of blind PDCP duplication, but may be easily ineffective due to the non-idealities of network interfaces. In fact, such a *two-hop approach* with an Xn interface having realistic delays may not be able to prevent unnecessary packet transmissions from another gNB in a timely manner.

Thus, we need to mitigate the resource consumption required by blind PDCP duplication in order to increase the amount of URLLC UEs that can be supported by the network. In the following, this study provides three methods to maximize the system-level KPI of supported URLLC load while ensuring the URLLC KPIs of latency (1 ms) and outage (10^{-5}). In particular, the proposed methods address the following objectives with the aim of radio resource efficiency improvements of PDCP duplication operations to boost spectral utilization:

1) minimization of queuing delay at the assisting node,
2) timely removing TBs containing redundant PDCP PDUs from the transmission queue to avoid resource wastage, and
3) reducing the overall amount of transmitted PDCP PDU copies.

V. PROPOSED DATA DUPLICATION ENHANCEMENTS

The aforementioned objectives are addressed by three novel enhancements, namely i) *differentiated scheduling of duplicates*, ii) *fast PDU discarding*, and iii) *selective data duplication upon failure*, which are presented in the following of this section. The approaches are presented in the order of increasing expected impact and effectiveness in reducing the radio-resource wastage when PDCP duplication is enabled.

A. DIFFERENTIATED SCHEDULING OF DUPLICATES

The proposed method to reduce the risk of queuing delays at the medium-access control (MAC) layer of the assisting node consists of applying differentiated scheduling policies entailing, e.g., differentiated prioritization and/or link adaptation of duplicated PDCP PDUs between the MgNB and the SgNB. For example, the MgNB may utilize an aggressive scheduling policy, comprising, e.g., the prioritization of duplicated URLLC packets with respect to packets belonging to other traffic flows (including even other URLLC flows) and/or a low BLER target for the transmission of the TB. On the other hand, the SgNB may employ a more relaxed policy for PDCP PDU duplicates forwarded by the MgNB, like, e.g., a lower priority and/or higher BLER target. We remark that the choice of the specific differentiated policy in terms of scheduling and/or link adaptation depends on the offered URLLC load, whose grow impacts the average queue length.

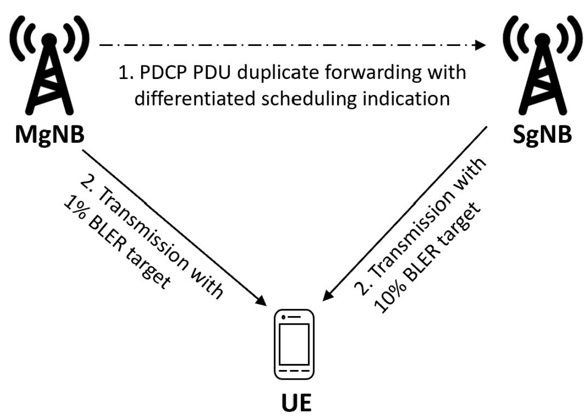


FIGURE 3. Differentiated scheduling of PDU duplicates.

A picture explaining the proposed approach is provided in Fig. 3 for the DC-based case. It can be seen that the MgNB forwards the PDCP PDU duplicate to the SgNB with differentiated scheduling indication, instructing the SgNB to transmit with 10%-BLER target, which results in a lower radio-resource consumption. The transmission from the MgNB, instead, utilizes a much more reliable MCS thanks to the 1% BLER target. In this way, the MgNB maximizes the transmission success probability on the primary link, while the SgNB minimizes the impact of PDCP PDU duplicates on its own transmission queue, thus achieving objective 1).

B. FAST PDU DISCARDING INDICATION

For a PDCP PDU discarding indication to be effective, such an indication needs to be received and processed *fast*, i.e., early enough to prevent an unnecessary HARQ retransmission. Thus, we propose a *one-step approach* in which a successful UE reception triggers HARQ acknowledgement (ACK) on the leg which performed the transmission and simultaneous discarding indication to the other leg containing the corresponding PDCP PDU sequence number that has just been acknowledged, as shown in Fig. 4. We remark that this indication must be notified by the PDCP layer, which is aware of the sequence numbers.

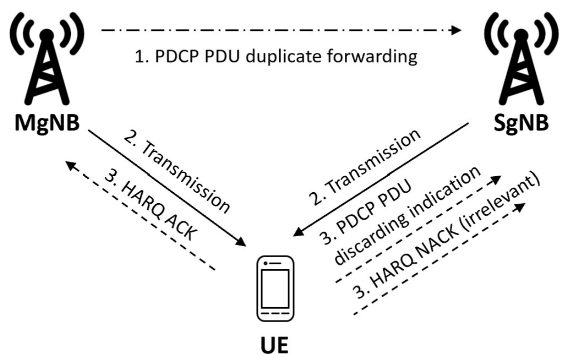


FIGURE 4. Fast PDU discarding indication for DC-based PDCP duplication. The transmission on the secondary link fails, but the HARQ NACK indication from the UE can be disregarded by the SgNB thanks to the fast PDU discarding indication.

A feasibility analysis, comparing the proposed approach and the current two-hop approach is provided in Fig. 5, accounting for the various delay components introduced at the MgNB and SgNB: the gNB layer-2 (L2) and layer-1 (L1) processing delays, the UE processing delay, the average alignment to the next frame/control opportunity, TB and HARQ feedback transmission times, and the transport latency on network interfaces. The values of these parameters were set according to the latest 3GPP specifications mentioned in Section IV. We observe that the UE-triggered discarding is more effective in canceling pending packets before a node actually transmits them, because the MgNB, whose first transmission is assumed failed, becomes aware of the outcome of SgNB's first transmission earlier by skipping the Xn hop.

Therefore, thanks to the proposed fast PDU discarding mechanism, we can effectively reduce the overall amount of DL transmission attempts from three to two when only one transmission failed, thus achieving objective 2). We remark that by removing these unnecessary transmissions we free precious radio resources for other transmissions.

C. SELECTIVE DATA DUPLICATION UPON FAILURE

Finally, if we enable the assistance of the SgNB only if actually needed by the MgNB, that is, only upon a transmission failure on the primary link, we can effectively reduce the average amount of transmission attempts to slightly more than a single one. In particular, as shown in picture Fig. 6, we propose that once the MgNB receives a HARQ NACK of its first transmission, it forwards the PDCP PDU duplicate with urgent scheduling indication to the SgNB, triggering its first transmission attempt along with the second transmission attempt on the primary link. Due to the URLLC latency constraint of 1 ms, a single transmission attempt can be performed by the SgNB under the assumptions on URLLC support mentioned in Section IV. Moreover, a urgent scheduling indication from the MgNB urges the SgNB to schedule such transmission as soon as possible. It is apparent that in most of the cases (99%, if the BLER target of 1% is set) a single transmission attempt is performed, thus achieving objective 3). On the other hand, when the primary leg has a failure, three attempts are performed, two on the primary link and one on the secondary link.

As already discussed previously, we observe that the non-ideality of network interfaces between gNBs and processing delays within gNB protocol stacks may impact the feasibility of this scheme, i.e., the time remaining after the MgNB forwards the PDCP PDU duplicate to the SgNB may not be sufficient to schedule the transmission from the SgNB within the URLLC latency budget. To cope with that, the MgNB may send the PDCP PDU duplicate to the SgNB as soon as it is created, together with a *hold-on flag*. Once the MgNB receives the HARQ NACK, it sends just a transmit command urging the SgNB to transmit the duplicated PDU as soon as possible. In this way, at the SgNB side the PDU could be pre-processed (including TB preparation upfront), granting more flexibility to the entire system; if the urgent scheduling indication is not eventually received (meaning that the MgNB received an HARQ ACK), the SgNB just drops the PDU from the transmission queue. The feasibility of selective PDCP duplication is analyzed in the timing chart provided in Fig. 7, obtained accounting for the same delay components of Fig. 5.

D. SUMMARY OF THE PROPOSALS

The proposed enhancements for PDCP duplication are summarized in Table 1, along with the respective expected impact on radio-resource consumption reduction and implementation complexity.

As for the envisioned method for differentiated scheduling of duplicates, we expect it to be effective in reducing the queuing delays at the assisting node, thus increasing the amount of supported URLLC. Its implementation complexity is likely to be low, comprising signaling exchange for the initial setup of the scheduling policy between MgNB and SgNB as well as the enforcement of the differentiated policy at the SgNB. On the other hand, the impact of spectrum utilization is limited, as it does not help in reducing the amount of overall duplicate PDCP PDU transmissions.

Fast PDU discarding indication instead provides an effective means to reduce the impact of unnecessary PDCP PDU retransmissions, in those cases where the first transmission attempt is not successful from one of the two gNBs. This method is expected to outperform the discarding mechanisms

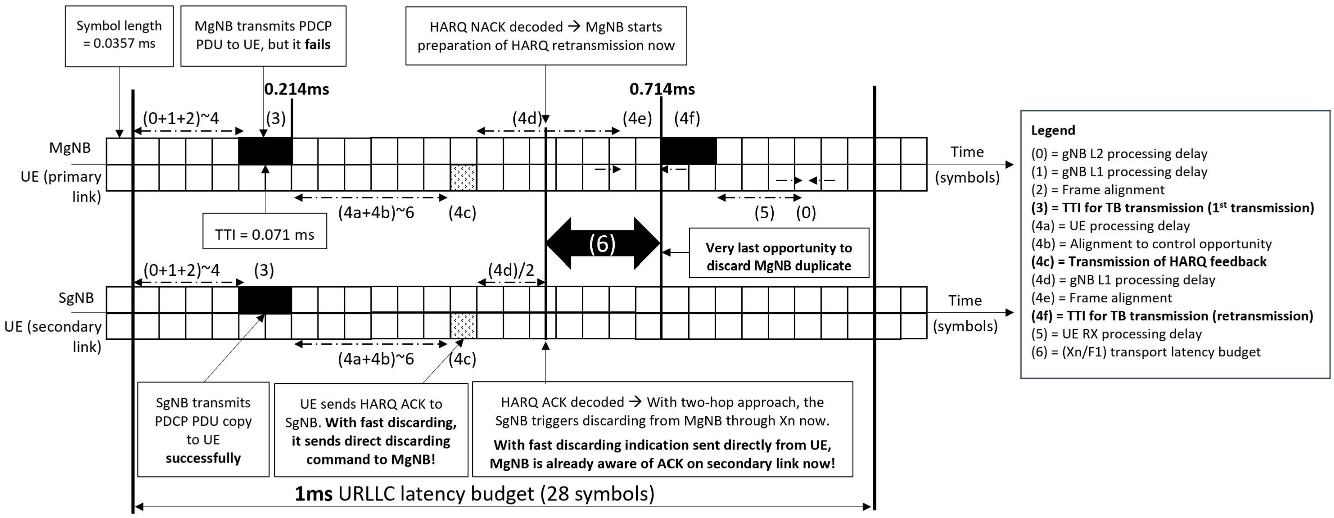


FIGURE 5. Feasibility analysis of fast PDU discarding indication, with 30 kHz SCS, 2-symbol TTI, and processing times provided by [34, Tab. 2.1-1].

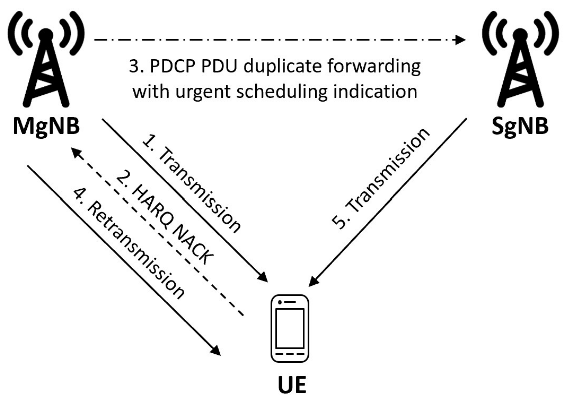


FIGURE 6. Selective PDCP PDU duplication.

TABLE 1. Summary of the proposed enhancements. For each approach, the rationale, the impact on reduction of spectrum utilization for data duplication, and implementation complexity are provided.

Approach	Rationale	Impact	Complexity
Differentiated scheduling of duplicates	Reduce the impact of duplicated PDU transmissions on SgNB's transmission queues	Low	Low
Fast PDU discarding indication	Effectively prevent unnecessary HARQ retransmissions from a serving gNB	Medium	Medium
Selective data duplication upon failure	Condition duplicated PDU transmission from SgNB on MgNB failure to reduce the amount of initial PDU copies transmissions	High	Medium

specified by 3GPP thanks to its one-step approach. The implementation of this enhancement causes a slight computational complexity to the UE, which is now instructed to generate and send an explicit PDCP PDU discarding indication other than the HARQ feedback.

Finally, selective data duplication upon failure is expected to massively reduce the amount of duplicated PDCP PDU transmissions, releasing a lot of radio resources. The implementation of this approach relies on a signaling message from MgNB to SgNB to trigger the first transmission attempt from the assisting node.

VI. SINGLE-USER ANALYSIS OF PROPOSED DUPLICATION ENHANCEMENTS

In the following, we derive a simplified mathematical framework to i) evaluate the gains of fast PDU discarding and ii) compare the performance of blind PDCP duplication and selective PDCP duplication, utilizing differentiated scheduling policies of PDCP PDU duplicates – in particular in terms of link adaptation strategies. The analysis is carried out in the *single-user case*.

A. NOTATION AND ASSUMPTIONS

Let us denote with p_1 and p_2 the probabilities of transmission failure of the primary leg and secondary leg respectively. The values of p_1 and p_2 equal the utilized BLER target on each link, according to the link adaptation of the differentiated scheduling policy. Moreover, since two transmission attempts can be afforded within the maximum URLLC latency of 1 ms, HARQ is exploited to maximize the decoding probability by applying packet combining, e.g., maximum ratio combining (MRC). In our model, we account for packet combining by defining a new pair of parameters, $p_1^{\text{comb}} \triangleq p_1^2$ and $p_2^{\text{comb}} \triangleq p_2^2$, which denote the decoding failure probability after two transmission attempts on the primary leg and on the

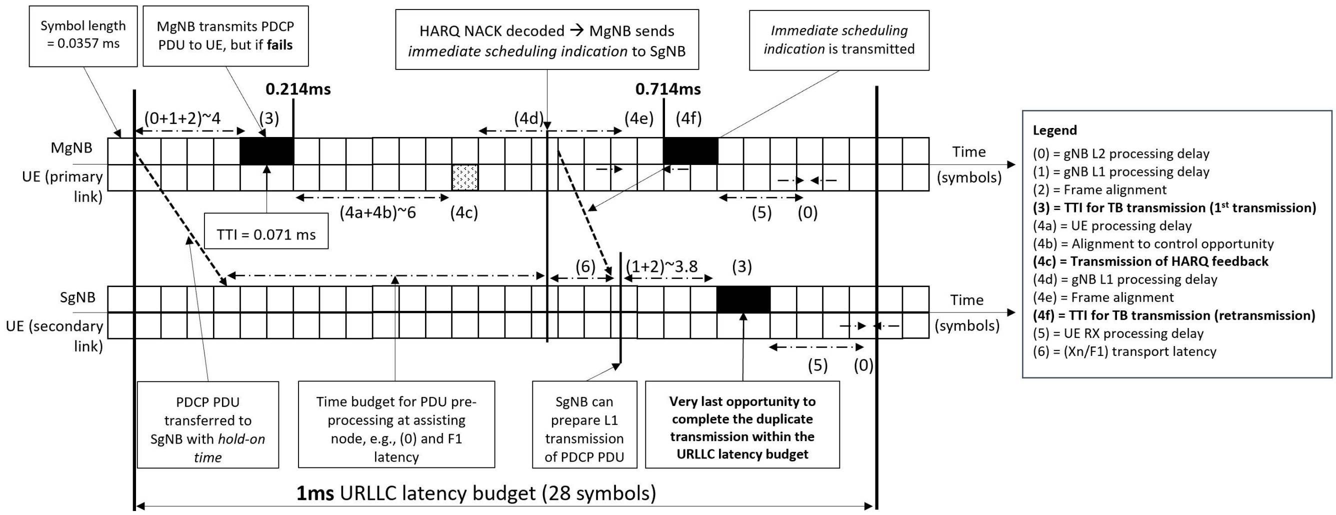


FIGURE 7. Selective duplication timing analysis, assuming a SCS of 30 kHz, 2-symbol TTI, and the processing times provided by [34, Tab. 2.1-1].

secondary leg, respectively. An independent-error model is assumed [36].

Moreover, we define c_1 and c_2 as the costs of each transmission attempt on the primary leg and secondary leg respectively. These variables are functions of p_1 and p_2, since they depend on the utilized BLER on each link; intuitively, a low BLER target corresponds to a robust MCS, which consumes more physical resources for the same payload. Let us define $c_1 \triangleq \log(1/p_1) = -\log p_1$ and $c_2 \triangleq \log(1/p_2) = -\log p_2$.

B. MODEL OF BLIND DUPLICATION

Regardless whether fast PDU discarding is enabled or not, the packet outage probability for blind duplication is

$$\begin{aligned} p_{\text{outage,BD}} &= \mathbb{P}[\text{all tx attempts fail}] \\ &= p_1\, p_1^{\text{comb}} p_2\, p_2^{\text{comb}} = p_1^3 p_2^3. \end{aligned} \tag{2}$$

The aggregate average transmission cost differs based on the presence or lack of fast PDU discarding as follows:

$$\bar{c}_{\text{BD}} = \begin{cases} c_1 + c_2 + c_1 p_1 + c_2 p_2, & \text{w/o discarding;} \\ c_1 + c_2 + c_1 p_1 p_2 + c_2 p_1 p_2, & \text{w/ discarding.} \end{cases} \tag{3}$$

C. MODEL OF SELECTIVE DUPLICATION

The packet outage probability of selective duplication is given by

$$p_{\text{outage,SD}} = p_1\, p_1^{\text{comb}} p_2 = p_1^3 p_2, \tag{4}$$

while the average transmission cost can be computed as

$$\bar{c}_{\text{SD}} = c_1 + (c_1 + c_2) p_1. \tag{5}$$

D. ANALYTICAL RESULTS

The beneficial effect of an effective fast PDU discarding mechanism over blind PDCP duplication resource efficiency is evident from the results in Fig. 8. The plot shows the trend of the difference between the two expressions for the average transmission cost in (3), denoted by $\Delta\bar{c}_{\text{BD}}$, for a grid of possible values of p_1 and p_2 in $[10^{-3}, 10^{-1}]\times[10^{-3}, 10^{-1}]$ which represent the available choices for the link adaptation policy. It can be seen that the cost saving is larger when the BLER targets on the two links are set high because, in this case, transmission failures are more likely to happen thus triggering the discarding mechanism.

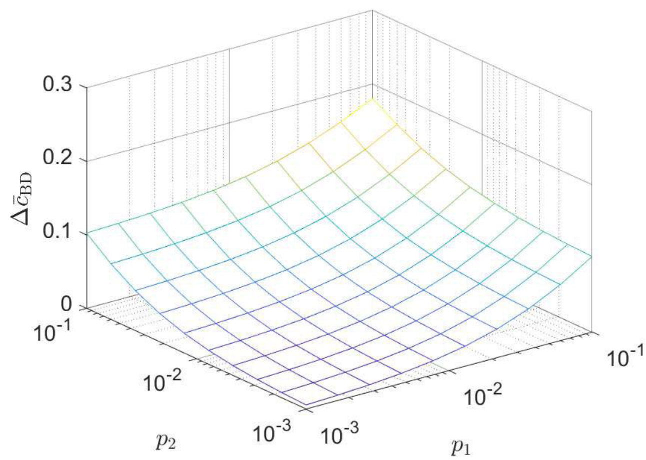
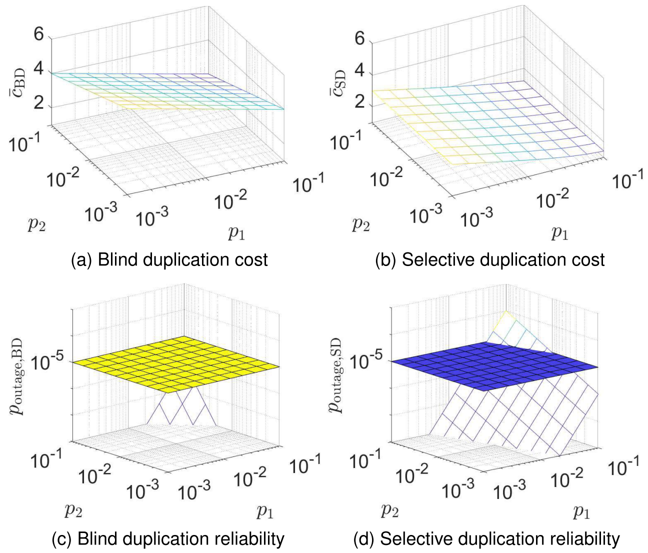


FIGURE 8. Cost saving for fast PDU discarding compared to baseline blind PDCP duplication, for various link adaptation policies.

For cases where fast PDU discarding is enabled, Fig. 9 compares the performance of blind PDCP duplication (9a and 9c) and selective PDCP duplication (9b and 9d) in terms of average transmission cost $\bar{c}$ (9a and 9b) and outage probability (9c and 9d) for the usual grid of values of p_1 and p_2. We observe that the blind PDCP duplication can always achieve an outage lower than 10^{-5} for all possible values of p_1 and p_2 – see Fig. 9c, where all operating points stand below the horizontal plane representing the URLLC reliability requirement. Nevertheless, the optimal operating point in terms of link-adaptation policy should be chosen in order to



(a) Blind duplication cost (b) Selective duplication cost

(c) Blind duplication reliability (d) Selective duplication reliability

FIGURE 9. **Performance comparison between blind PDCP duplication and selective PDCP duplication for various link adaptation policies. The two horizontal planes in (c) and (d) highlight the URLLC reliability target of 10^{-5}.**

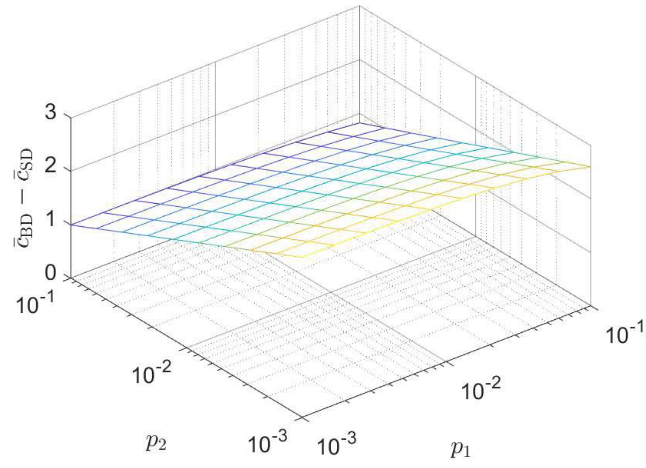


FIGURE 10. **Cost saving of selective PDCP duplication with respect to blind PDCP duplication for various link adaptation policies.**

minimize the cost in Fig. 9a, therefore one should set $p_1^\star = p_2^\star = 10^{-1}$ (that is, 10% BLER target), yielding $\bar{c}_{\mathrm{BD}}^\star \simeq 2$. As for the selective PDCP duplication, from Fig. 9d we notice that it cannot satisfy the URLLC reliability requirement for all possible values of p_1 and p_2; in particular, we infer that, contrary to the previous case, we should *not* choose their values so that p_1 and p_2 are both high. Instead, from Fig. 9b we note that setting an *aggressive* primary leg ($p_1^\star = 10^{-1}$) and a *robust* secondary leg ($p_2^\star = 10^{-2}$) yields $\bar{c}_{\mathrm{SD}}^\star \simeq 1.3$ while achieving the 10^{-5} outage probability target given by URLLC.

Finally, it is straightforward to derive the following expression of cost savings provided by selective PDCP duplication with respect to blind PDCP duplication:

$$\bar{c}_{\mathrm{BD}} - \bar{c}_{\mathrm{SD}} = \begin{cases} c_2(1 + p_2 - p_1), & \text{w/o discarding;} \\ c_2 - p_1(1 - p_2)(c_1 + c_2), & \text{w/ discarding.} \end{cases} \tag{6}$$

The trend of cost savings provided by the proposed approach against blind PDCP duplication with discarding enabled is graphically shown in Fig. 10. We notice that the largest cost difference is obtained when $p_1 = p_2 = 10^{-3}$, i.e., when selective PDCP duplication almost surely exploits a single transmission from the MgNB while blind PDCP duplication is very likely to employ two redundant transmission attempts (one from the MgNB and one from the SgNB).

VII. SYSTEM-LEVEL PERFORMANCE OF SELECTIVE DATA DUPLICATION UPON FAILURE

From the analytical study we learned that, in blind PDCP duplication, similar and rather high BLER targets may be set on both legs to minimize the aggregate transmission cost. On the other hand, the selective PDCP duplication scheme is asymmetrical, possibly benefiting from robust SgNB transmissions if the bandwidth allows. However, the derived models neither consider bandwidth limitations nor the interference caused by URLLC and eMBB traffic to UEs. Traffic variations may correlate with interference and block error probability (BLEP) dynamics, and queuing delay at the MAC scheduler may occur; these aspects are not covered by the analytical framework. Therefore, in this section we will compare blind PDCP duplication and selective PDCP duplication, both based on DC, by means of advanced multi-user, multi-cell dynamic system-level simulations, where the performance determining effects are modeled in more detail, providing more realistic results.

A. SIMULATION SETUP

We consider the 3GPP-compliant HetNet scenario described in [37, Sec. 4.2], and previously utilized in [21]. It includes 7 macro gNBs with 3 sectors per macro, and 4 clustered small cells per macro cell, counting overall $n_{\mathrm{m}} = 21$ macro cells and $n_{\mathrm{s}} = 84$ small cells. A FDD configuration is assumed; the carrier frequency equals 2 GHz for the macro layer and 3.5 GHz for the small-cell layer, while the system bandwidth is 10 MHz for both. $n_{\mathrm{u}} = 840$ UE are dropped in the scenario. The association between UE and serving cells is based on RSRP measurements, assuming a 15-dB CRE for load balancing between the macro layer and the small-cell layer. This leads the 70% of UEs to fall within the coverage range of small cells, while the remaining 30% of UEs are covered only by a macro gNB.

Those UEs which are covered only by the macro gNB are in single connectivity (SC). On the other hand, UEs in the coverage range of small cells may exploit the assistance of the macro layer to improve their transmission reliability via DC-based PDCP duplication. In case of blind PDCP duplication, the activation of DC depends whether (1) is fulfilled by a given UE. In our simulations, we set `DC_Range` $= 15$ dB, which provides DC activation to 1/3 of the UEs attached to the small-cell layer. As discussed in [21], this configuration of PDCP duplication results in the highest reliability gain, despite less than $1/3$ of all UEs in the scenario are allowed to enable DC. On the other hand, in case of selective PDCP duplication we disable the limitation on the amount of UE in DC via the `DC_Range` parameter. We do that because we expect a massive reduction of resource utilization thanks to the selective PDCP duplication in the multi-user case, allowing to enable DC for any UE in the coverage range of small cells.

The traffic offered to the network consists of two QoS types: URLLC and eMBB. In particular, a fraction r of all deployed UEs are of URLLC type, i.e., they receive packets with payload P (in bytes) generated by a Poisson-distributed arrival process of rate λ packets per second. The offered URLLC traffic per macro cell area can be computed as

$$G_{\text{URLLC}} = \frac{r \cdot n_{\text{u}} \cdot \lambda \cdot 8P}{n_{\text{m}}} \text{ [bps]}, \tag{7}$$

where "bps" stands for bits per second. We remark that a given value for the aggregate offered URLLC traffic can be alternatively achieved by keeping a constant packet arrival rate λ and increasing the number of URLLC UEs n_{u}, or vice versa. In our simulations, we adopt the latter approach, i.e., keeping n_{u} constant while increasing λ, for computational efficiency. The remaining $(1-r)$ fraction of the UEs receive full-buffer User Datagram Protocol (UDP) packets modeling the eMBB QoS type.

We assume the NR frame structure previously utilized by [9], [21], [38], entailing a short TTI duration of 0.143 ms (obtained with 2-symbol TTI at 15-kHz SCS) and a HARQ round-trip time (RTT) of 4 TTIs, which comprises 1 TTI for DL TB transmission, UE processing delay, HARQ feedback transmission, and gNB processing. With this configuration, two transmission attempts from each leg can be performed within 1 ms. We recall that, according to [34], two transmission attempts can be accommodated only with 30-kHz SCS, thus halving the TTI duration we considered in our simulations. However, it is easy to check that our modeling of the HARQ RTT reflects closely the timeline under consideration by 3GPP.

For each transmission we first calculate the signal-to-interference-plus-noise ratio (SINR) per subcarrier symbol, then we obtain the mean mutual information per coded bit (MMIB), which is used to determine the BLEP, and the TB is received correctly or not accordingly. Dynamic link adaptation with outer-loop link adaptation (OLLA) is considered, setting a BLER target of 1% for URLLC packets. Reliable HARQ feedback is assumed, i.e., the Physical Uplink Control Channel (PUCCH) is always error-free. The Xn interface latency and packet loss rate is neglected since we assume a high-speed low-latency fiber connectivity for this interface.

Table 2 summarizes the simulation parameters.

TABLE 2. System-level simulation parameters.

PARAMETER	MACRO LAYER	SMALL-CELL LAYER
Number of cells	$n_{\text{m}} = 21$ (7 gNBs with 3 sectors each)	$n_{\text{s}} = 84$ (with 15-dB CRE)
Inter-gNB distance	500 m	clusters of 4 per macro
Antenna height	32 m	10 m
Tx. antennas (gNB)	2	2
Rx. antennas (UE)	2	2
Carrier frequency	2 GHz	3.5 GHz
Bandwidth	10 MHz	10 MHz
Tx power	46 dBm	30 dBm
Channel model	Urban macro (UMa)	Urban micro (UMi)
Number of UEs n_{u}	840	
URLLC ratio r	$r = 1$ (840 URLLC UEs, 0 eMBB UEs), $r = 2/3$ (560 URLLC UEs, 280 eMBB UEs)	
BLER target	1%	
Subcarrier spacing	15 kHz	
Symbol time	0.072 ms	
TTI length	2 symbols	
UE proc. delay	1 TTI	
HARQ feedback tx.	1 TTI	
gNB proc. delay	1 TTI	
HARQ RTT	4 TTIs	
Xn latency	0	
URLLC traffic	FTP Model 3 (Poisson) [41], $\lambda \in \{658, 938\}$ pkt/s if $r = 1$, $\lambda \in \{94, 188, 376\}$ pkt/s if $r = 2/3$	
URLLC payload P	50 B	
eMBB traffic	Full buffer, UDP	

B. RESULTS WITH URLLC-ONLY TRAFFIC

We first present the results in presence of URLLC-only offered traffic (i.e., with $r = 1$) in Fig. 11, which shows the complementary cumulative distribution function (CCDF) of the packet delivery latency (see Fig. 11a) and the CCDF of the radio resource utilization (see Fig. 11b) for SC, blind PDCP duplication, and selective PDCP duplication under medium and high URLLC offered loads, i.e., $G_{\text{URLLC}} \in \{10.5, 15\}$ Mbps, obtained using $\lambda \in \{658, 938\}$ pkt/s. We remark that such values of the parameter λ may not refer to a specific IIoT use case, rather they are determined with the aim of demonstrating how much traffic the network is capable to serve.

As stated previously, we observe good alignment between the timing of our system-level simulations and the timing chart in Fig. 7. Indeed, considering the curve of selective PDCP duplication at $G_{\text{URLLC}} = 10.5$ Mbps (solid line, no marker) we notice that the first packet transmission from the MgNB (corresponding to the first drop of the curve) is delivered approximately within 0.3 ms, while the second transmission attempt (corresponding to the second drop of the curve) is triggered at around 0.7 ms, pretty much aligned with our timing analysis. Observing that the URLLC packet reliability can be inferred from the vertical axis, the curve shows that the 10^{-5} requirement can be met within the 1-ms latency threshold, as envisioned by the timing chart in Fig. 7.

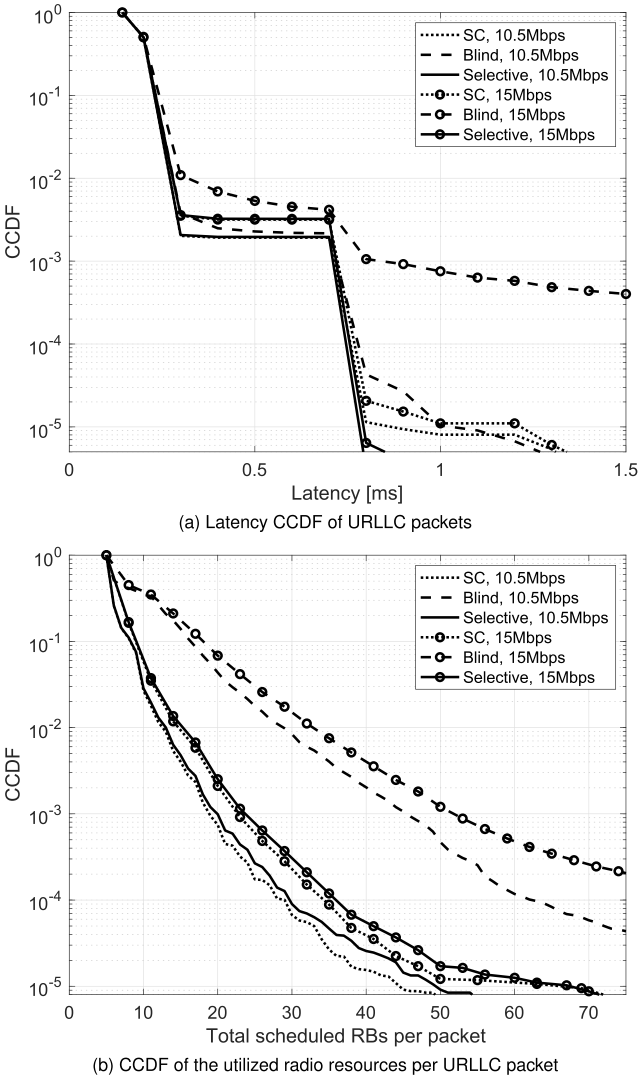


FIGURE 11. Performance comparison between SC, blind PDCP duplication, and selective PDCP duplication in presence of URLLC-only offered traffic.

We notice that selective PDCP duplication and SC achieve the same reliability (even lower than the BLER target 10^{-2}) after the first transmission at both offered URLLC loads; this is as expected, because if SINRs are very good both approaches transmit a single TB, provided that there is no packet segmentation. However, selective PDCP duplication outperforms SC when the retransmission is triggered, by exploiting the first transmission from the SgNB which allows an additional transmission attempt compared to SC. Nevertheless, both approaches are able to meet the 10^{-5} within the 1-ms latency at both offered loads. On the other hand, it can be seen that blind PDCP duplication is always outperformed by both SC and selective PDCP duplication. In particular, at all offered URLLC loads, we notice that the queuing delay effect prevents blind PDCP duplication from achieving the same reliability values of SC and selective PDCP duplication after the first transmission attempt. Upon the second transmission attempt, blind PDCP duplication can hardly meet the URLLC requirements at $G_{\text{URLLC}} = 10.5$ Mbps, while it is clearly not able to sustain a load G_{URLLC} of 15 Mbps.

The radio resource utilization results in Fig. 11b show the CCDF of the total number of scheduled resource blocks (RBs) per URLLC packet. We observe a clear shift of blind PDCP duplication curves, heading to the right part of the chart with respect to SC and selective PDCP duplication. This means that blind PDCP duplication requires much more RBs than the other two approaches. The waste of resources due to unnecessary first transmissions of duplicates causes increased interference and risk of queuing delays at the gNBs, severely limiting the gain that was theoretically expected from a single-user scenario. On the other hand, the curves associated with selective PDCP duplication are very close to those associated with SC, up to and above 90-percentile (corresponding to 10^{-1} on the y-axis). After this point, SC is more cost-efficient than selective PDCP duplication due to the additional packet transmission from the SgNB required by the latter approach. However, such additional radio resources utilized by selective PDCP duplication allow to outperform SC in terms of packet reliability, as shown earlier.

C. RESULTS WITH EMBB BACKGROUND TRAFFIC

The performance of the schemes is now compared in presence of eMBB background traffic, besides URLLC traffic, assuming $r = 2/3$. The results are provided in Fig. 12. G_{URLLC} is varied in the set $\{1, 2, 4\}$ Mbps, that is a much lower level than in the previous analysis because the background traffic significantly increases the interference in the system, hence decreases the UE SINRs. The curve associated to SC at $G_{\text{URLLC}} = 1$ Mbps is added as a reference.

At $G_{\text{URLLC}} = 1$ Mbps, SC is outperformed by blind PDCP duplication, almost achieving the required 10^{-5} reliability at 1 ms; this result proves the effectiveness of data duplication in high-interference conditions, where the packet duplicates do not increase the interference but 'just' replace eMBB traffic. The proposed selective PDCP duplication is capable of reducing the packet error rate even further, well below the 10^{-5} threshold at 1 ms, being more effective than blind PDCP duplication in supporting the offered URLLC load. It is interesting to observe also how the respective reliability level at 1-ms latency is achieved by the various approaches. After the first transmission attempt, selective PDCP duplication and SC achieve the same reliability (10^{-2}, equal to the BLER target), as observed in the URLLC-only system-level simulations. Blind PDCP duplication outperforms both of them, achieving a higher reliability at this stage. However, with the second transmission from the MgNB the selective duplication of PDUs outperforms blind duplication.

Let us now compare the performance of blind and selective PDCP duplication at the different offered URLLC load levels. We notice that blind PDCP duplication can support $G_{\text{URLLC}} = 4$ Mbps with the same reliability achieved by SC at $G_{\text{URLLC}} = 1$ Mbps; nevertheless, as the offered URLLC load increases, blind PDCP duplication becomes less and

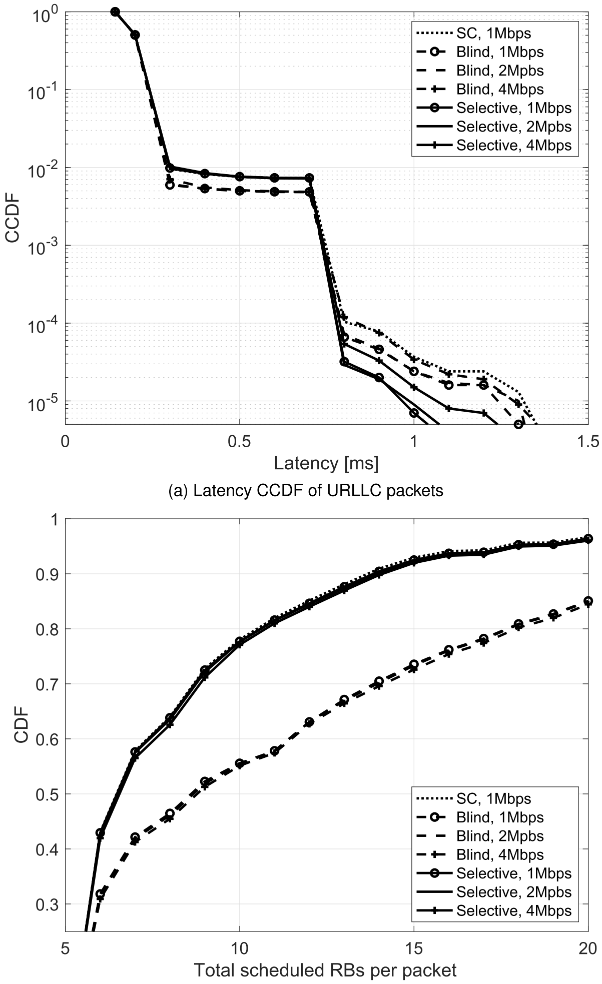


FIGURE 12. Performance comparison between SC, blind PDCP duplication, and selective PDCP duplication in presence of URLLC offered traffic and eMBB background traffic.

less effective in closing the gap with the 10^{-5} reliability target at 1 ms. On the other hand, selective PDCP duplication provides a lower-than-10^{-5} packet error rate at $G_{\mathrm{URLLC}} \in \{1, 2\}$ Mbps and is close to meet the requirement even at $G_{\mathrm{URLLC}} = 4$ Mbps. At all offered loads, the proposed approach outperforms blind PDCP duplication at $G_{\mathrm{URLLC}} = 1$ Mbps. In particular, we notice that the error rate at 1-ms latency of selective PDCP duplication is approximately 60% lower than blind PDCP duplication, with a slightly decreasing trend as the offered URLLC load increases. We recall that the number of UEs in DC when the selective PDCP duplication is enabled is much higher than when we enable blind PDCP duplication, due to the deactivation of the `DC_Range` parameter for the former approach. Therefore, when the offered URLLC load increases, selective PDCP duplication is expected to be impacted more than blind PDCP duplication; on the other hand, the massive reduction of radio resources utilized for the proposed approach results in an approximately constant relative gain.

Fig. 12b shows the CDF of the total number of scheduled RBs per URLLC packet to better appreciate the difference in terms of radio resource usage between the various approaches. We can clearly see that selective PDCP duplication curves are bundled together with the reference SC curve, and they are clearly spaced apart from those of blind PDCP duplication. We recall that such a gain in radio resource usage is due to the reduction of the amount of PDU duplicate transmissions of 99%, assuming a BLER of 1%.

VIII. CONCLUSION AND FUTURE RESEARCH

This paper dealt with data duplication for URLLC in 3GPP systems, proposing several enhancements for achieving radio efficiency in downlink PDCP duplication. A mathematical framework was derived to assess the performance of a selected set of the proposed enhancements in the single-user case. The multi-user case was addressed via realistic system-level simulation campaigns to evaluate the performance of the most promising proposal, namely the *selective data duplication upon failure*. The simulation results showed that, for the investigated scenario and assumptions, the proposed approach is effective in reducing the radio resource wastage, as it provides the same reliability as the baseline approach for data duplication at four-times higher offered URLLC load, in presence of eMBB best effort background load. It also yields a 60% reduction of the packet error rate at 1-ms latency with respect to the baseline. Moreover, while baseline PDCP duplication requires careful tuning of the "data-duplication area," selective PDCP duplication enables a more robust performance as such careful tuning is no longer needed.

For future research directions in the field of PDCP duplication in 3GPP systems, it is worth investigating the performance of data duplication with more than two copies and how to increase the PDCP duplication performance in the uplink by a more dynamic selection of transmission legs, in line with the objectives of ongoing 3GPP standardization activities. More research effort will be needed also in respect of new enhanced URLLC (eURLLC) and time-sensitive communication (TSC) requirements [40], entailing an even lower latency (0.5 ms) and higher packet reliability (10^{-6}), exploiting multi-connectivity protocols with more than two copies.

ACKNOWLEDGMENT

The authors would like to thank Nurul H. Mahmood (Oulu University, Finland) and Gilberto Berardinelli (Aalborg University, Denmark) for precious help and useful discussions.

REFERENCES

[1] *E-UTRA and E-UTRAN; Overall Description; Stage 2 (Release 15)*, document 3GPP, TSG-RAN 36.300 V15.5.0, Mar. 2019.

[2] *NR; NR ad NG-RAN Overall Description; Stage 2 (Release 15)*, document 3GPP, TSG-RAN 38.300 V15.5.0, Mar. 2019.

[3] *Minimum Requirements Related to Technical Performance for IMT-2020 Radio Interface(s)*, ITU-R M.2410-0, Nov. 2017.

[4] P. Popovski, J. J. Nielsen, C. Stefanovic, E. de Carvalho, E. Strom, K. F. Trillingsgaard, A.-S. Bana, D. M. Kim, R. Kotaba, J. Park, and R. B. Sorensen, "Wireless access for ultra-reliable low-latency communication: Principles and building blocks," *IEEE Netw.*, vol. 32, no. 2, pp. 16–23, Mar./Apr. 2018.

[5] G. Pocovi, H. Shariatmadari, G. Berardinelli, K. Pedersen, J. Steiner, and Z. Li, "Achieving ultra-reliable low-latency communications: Challenges and envisioned system enhancements," *IEEE Netw.*, vol. 32, no. 2, pp. 8–15, Mar./Apr. 2018.

[6] R. Fantini, W. Zirwas, L. Thiele, D. Aziz, and P. Baracca, "Coordinated multi-point transmission in 5G," in *5G Mobile and Wireless Communications Technology*, A. Osseiran, J. Monserrat, and P. Marsch, Eds. Cambridge, U.K.: Cambridge Univ. Press, 2016, ch. 9, pp. 248–276.

[7] J. Liu, K. Au, A. Maaref, J. Luo, H. Baligh, H. Tong, A. Chassaigne, and J. Lorca, "Initial access, mobility, and user-centric multi-beam operation in 5G new radio," *IEEE Commun. Mag.*, vol. 56, no. 3, pp. 35–41, Mar. 2018.

[8] P. Popovski, "Ultra-reliable communication in 5G wireless systems," in *Proc. 1st Int. Conf. 5G Ubiquitous Connectivity*, Akaslompolo, Finland, Nov. 2014, pp. 146–151.

[9] K. Pedersen, G. Pocovi, J. Steiner, and A. Maeder, "Agile 5G scheduler for improved E2E performance and flexibility for different network implementations," *IEEE Commun. Mag.*, vol. 56, no. 3, pp. 210–217, Mar. 2018.

[10] A. Karimi, K. I. Pedersen, N. H. Mahmood, J. Steiner, and P. Mogensen, "5G centralized multi-cell scheduling for URLLC: Algorithms and system-level performance," *IEEE Access*, vol. 6, pp. 72253–72262, 2018.

[11] M. Alsenwi, N. H. Tran, M. Bennis, A. K. Bairagi, and C. S. Hong, "EMBB-URLLC resource slicing: A risk-sensitive approach," *IEEE Commun. Lett.*, vol. 23, no. 4, pp. 740–743, Apr. 2019.

[12] D. Öhmann, A. Awada, I. Viering, M. Simsek, and G. P. Fettweis, "Diversity trade-offs and joint coding schemes for highly reliable wireless transmissions," in *Proc. IEEE Veh. Tech. Conf. (VTC2016-Fall)*, Montreal, QC, Canada, Sep. 2016, pp. 1–6.

[13] M. Mohiuddin, M. Popovic, A. Giannakopoulos, and J.-Y. Le Boudec, "Experimental validation of the usability of Wi-Fi over redundant paths for streaming phasor data," in *Proc. IEEE Int. Conf. Smart Grid Commun.*, Sydney, NSW, Australia, Nov. 2016, pp. 533–538.

[14] G. Cena, S. Scanzio, and A. Valenzano, "Improving effectiveness of seamless redundancy in real industrial Wi-Fi networks," *IEEE Trans. Ind. Informat.*, vol. 14, no. 5, pp. 2095–2107, May 2018.

[15] M. C. Lucas-Estañ, J. L. Maestre, B. Coll-Perales, J. Gozalvez, and I. Lluvia, "An experimental evaluation of redundancy in industrial wireless communications," in *Proc. IEEE 23rd Int. Conf. Emerg. Technol. Factory Automat. (ETFA)*, Turin, Italy, Sep. 2018, pp. 1075–1078.

[16] J. J. Nielsen, R. Liu, and P. Popovski, "Optimized interface diversity for ultra-reliable low latency communication (URLLC)," in *Proc. IEEE Global Commun. Conf.*, Singapore, Dec. 2017, pp. 1–6.

[17] P. K. Taksande, P. Jha, and A. Karandikar, "Dual connectivity support in 5G networks: An SDN based approach," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Marrakesh, Morocco, Apr. 2019, pp. 1–6.

[18] P. W. Baier, M. Meurer, T. Weber, and H. Troger, "Joint transmission (JT), an alternative rationale for the downlink of time division CDMA using multi-element transmit antennas," in *Proc. IEEE 6th Int. Symp. Spread Spectr. Techn. Appl.*, Parsippany, NJ, USA, Sep. 2000, pp. 1–5.

[19] J. Rao and S. Vrzic, "Packet duplication for URLLC in 5G: Architectural enhancements and performance analysis," *IEEE Netw.*, vol. 32, no. 2, pp. 32–40, Mar./Apr. 2018.

[20] *New WID: Support of NR IIoT*, document RP-190728, 3GPP TSG-RAN meeting #83, Nokia and Nokia Shanghai Bell, Shenzhen, China, Mar. 2019.

[21] N. Mahmood, M. Lopez, D. Laselva, K. Pedersen, and G. Berardinelli, "Reliability oriented dual connectivity for URLLC services in 5G New Radio," in *Proc. 15th Int. Symp. Wireless Commun. Syst. (ISWCS)*, Lisbon, Portugal, Aug. 2018, pp. 1–6.

[22] D. S. Michalopoulos and V. Pauli, "Data duplication for high reliability: A protocol-level simulation assessment," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Shanghai, China, May 2019, pp. 1–7.

[23] A. Aijaz, "Packet duplication in dual connectivity enabled 5g wireless networks: Overview and challenges," *IEEE Commun. Standards Mag.*, vol. 3, no. 3, pp. 20–28, Sep. 2019. [Online]. Available: https://arxiv.org/pdf/1804.01058.pdf

[24] E. J. Khatib, D. A. Wassie, G. Berardinelli, I. Rodriguez, and P. Mogensen, "Multi-connectivity for ultra-reliable communication in industrial scenarios," in *Proc. IEEE 89th Veh. Technol. Conf.*, Kuala Lumpur, Malaysia, Apr./May 2019, pp. 1–6.

[25] N. Mahmood, K. Pedersen, A. Karimi, G. Berardinelli, and D. Laselva, "On the resource utilization of multi-connectivity transmission for URLLC services in 5G New Radio," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Marrakech, Morocco, Apr. 2019, pp. 1–6.

[26] N. H. Mahmood and H. Alves, "Dynamic multi-connectivity activation for ultra-reliable and low-latency communication," in *Proc. 16th Int. Symp. Wireless Commun. Syst. (ISWCS)*, Oulu, Finland, Aug. 2019, pp. 112–116.

[27] D. Guzman, R. Schoeffauer, and G. Wunder, "Predictive network control in multi-connectivity mobility for URLLC services," in *Proc. IEEE 24th Int. Workshop Comput. Aided Modeling Design Commun. Links Netw.*, Limassol, Cyprus, Sep. 2019, pp. 1–7.

[28] *E-UTRA and NR; Multi-Connectivity; Stage 2 (Release 15)*, document 3GPP, TSG-RAN 37.340 V15.5.0, Mar. 2019.

[29] *E-UTRA; Packet Data Convergence Protocol (PDCP) Specification (Release 15)*, document 3GPP, TSG-RAN 36.323 V15.3.0, Mar. 2019.

[30] *NR; Packet Data Convergence Protocol (PDCP) Specification (Release 15)*, document 3GPP, TSG-RAN 38.323 V15.5.0, Mar. 2019.

[31] *Revised SID: Study on NR Industrial Internet of Things (IoT)*, document RP-182090, 3GPP TSG-RAN meeting #81, Nokia and Nokia Shanghai Bell, Gold Coast, Australia, Sep. 2018.

[32] *Study on NR Industrial Internet of Things (IIoT) (Release 16)*, document 3GPP, TSG-RAN 38.825 V16.0.0, Mar. 2019.

[33] *NR; User Equipment (UE) Radio Transmission and Reception; Part 1: Range 1 Standalone (Release 15)*, document 3GPP, TSG-RAN 38.101-1 V15.5.0, Mar. 2019.

[34] *Discussion on the RAN2 LS on TSN Requirements Evaluation*, document R1-1900935, 3GPP TSG-RAN WG1 Ad-Hoc Meeting, Nokia and Nokia Shanghai Bell, vol. 1901, Taipei, Taiwan, Jan. 2019.

[35] *NR; Physical Layer Procedures for Data (Release 15)*, document 3GPP, TSG-RAN 38.214 V15.6.0, Jun. 2019.

[36] R. Sassioui, E. Pierre-Doray, L. Szczecinski, and B. Pelletier, "Modelling decoding errors in HARQ," Dec. 2015, *arXiv:1512.02511*. [Online]. Available: https://arxiv.org/abs/1512.02511

[37] *Small Cell Enhancements for E-UTRA and E-UTRAN—Physical Layer Aspects (Release 12)*, document 3GPP, TSG-RAN 36.872 V12.1.0, Dec. 2013.

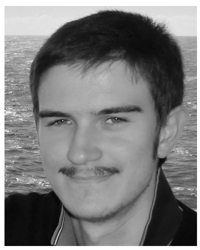
[38] G. Pocovi, K. I. Pedersen, and P. Mogensen, "Joint link adaptation and scheduling for 5G ultra-reliable low-latency communications," *IEEE Access*, vol. 6, pp. 28912–28922, 2018.

[39] *Study on Scenarios and Requirements for Next Generation Access Technologies*, document 3GPP 38.913 V15.0.0, Jun. 2018.

[40] *Study on Enhancement of 5G System (5GS) for Vertical and Local Area Network (LAN) Services (Release 16)*, document 3GPP, TSG-SA 23.734 V16.2.0, Jun. 2019.

[41] *E-UTRA; Base Station (BS) Radio Transmission and Reception (Release 16)*, document 3GPP, TSG-RAN 36.104 V16.2.0, Jun. 2019.

[42] *NR; Base Station (BS) Radio Transmission and Reception (Release 16)*, document 3GPP, TSG-RAN 38.104 V16.0.0, Jun. 2019.



MARCO CENTENARO received the B.S. degree in information engineering and the M.S. and Ph.D. degrees in telecommunication engineering from the University of Padova, Italy, in 2012, 2014, and 2018, respectively.

He was on leave as a Visiting Researcher at Nokia Bell Labs, Stuttgart, Germany, and Yokohama National University, Japan, in 2016 and 2017, respectively. From 2018 to 2019, he was Postdoctoral Research Fellow at the University of Padova, doing research on projects funded by Huawei Technologies, and at Aalborg University, Denmark, working in tight collaboration with Nokia Bell Labs Aalborg. At the time of writing, he is an Expert Researcher at Fondazione Bruno Kessler (FBK), Trento, Italy. His research is focused on telecommunication standards and the Internet of Things technologies.

Dr. Centenaro was a recipient of the Ph.D. Award from the Gruppo Telecomunicazioni e Tecnologie dell'Informazione (GTTI – Italian Association of Telecommunications and Information Technologies), in 2018, in recognition of the best Ph.D. theses defended at an Italian University in the areas of communications technologies, and the Best Paper Award at IEEE PIMRC 2018.

DANIELA LASELVA received the M.Sc. degree in electrical engineering from the Polytechnic of Bari, Italy, in 2002.

From 2002 to 2006, she was active in COST273 and EU FP6 project WINNER on MIMO radio channel modeling with Elektrobit, Finland, and worked as a Senior Design Engineer with Nokia, Finland. Since 2006, she has been with Nokia Aalborg, Denmark. She is currently with Nokia Bell Labs, as a Senior Researcher. At present, she is engaged with 5G New Radio Design and Standardization, including multiconnectivity solutions, support of mission-critical applications, and end-to-end performance optimization. She has coauthored more than 30 peer-reviewed publications and four book chapters, and the inventor of numerous patents on a wide range of topics.

JENS STEINER received the M.Sc. degree in electrical engineering from Aalborg University, Denmark, in 1996, with specialty in software engineering.

Since 1996, he has been working for different companies mainly in the telecommunications sector. Since 2005, he has been working at Nokia Bell Labs, Aalborg. First as an External Consultant and subsequently as a permanent member of staff. At Nokia Bell Labs, he is currently involved in 5G radio access network system-level simulator development. He also contributes to radio research beyond software development.

KLAUS PEDERSEN received the M.Sc. degree in electrical engineering and the Ph.D. degree from Aalborg University, Aalborg, Denmark, in 1996 and 2000, respectively.

He is currently leading the Nokia Bell Labs Research Team (with focus on radio systems and protocol research), Aalborg, and also a part-time Professor with the Wireless Communications Network Section, Aalborg University. He has authored or coauthored approximately 200 peer-reviewed publications on a wide range of topics and is an inventor of several patents. His current research interests include 5G new radio, including radio resource management aspects, ultrareliable low latency communications, and flexible duplexing, having special emphasis on mechanisms that offer improved end-to-end (E2E) performance delivery and overall systems performance aspects.

PREBEN MOGENSEN received the M.Sc. and Ph.D. degrees from Aalborg University, in 1988 and 1996, respectively.

Since 1995, he has been part-time associated with Nokia Bell Labs. Since 2000, he has been a Full Professor with Aalborg University, where he is currently leading the Wireless Communication Networks Section, Department of Electronic Systems. He is also a Principal Scientist with Nokia Bell Labs. He has coauthored more than 400 articles in various domains of wireless communication. His current research interest includes 5G evolution towards 6G.