Data Duplication for High Reliability: A Protocol-Level Simulation Assessment

Diomidis S. Michalopoulos*, and Volker Pauli†
* Nokia Bell Labs, Munich, Germany, Email: diomidis.michalopoulos@nokia-bell-labs.com
†Nomor Research, Munich, Germany. Email: pauli@nomor.de

***Abstract*—Among the key 5G targets is to offer "ultra-reliable" communication services with packet loss rates not exceeding 10^{-5}. It is widely acknowledged that this objective can be achieved only in multi-cellular setups with a special use of multi-connectivity, known as data duplication. This paper examines the use of data duplication at layer 2, i.e., at the Packet Data Convergence Protocol (PDCP) layer of the protocol stack. Unlike many alternative schemes, it does not rely on detailed channel state information or coordination in layer 1, but relies on feedback and coordination at layer 2 only, including a new level of packet acknowledgments. Using system-level simulations the efficiency of the proposed scheme in a real-world environment is illustrated. Besides packet loss rate, the system performance against throughput and delay at the application layer is assessed.**

I. Introduction

Reliability is one of the major requirements of 5G networks. This particularly applies to mission critical applications associated with minimal tolerance to failures. In this context, the notion of Ultra Reliable Low Latency Communications (URLLC) has been developed [1], which represents a 5G approach specially designed for achieving packet error rates that are as low as 10^{-5}.

Research has shown that such stringent reliability requirements cannot be achieved with existing methods, unless new approaches are deployed [1], [2]. The main method towards this direction is the so-called data duplication (referred to also as packet duplication), which in fact represents a special case of multi-connectivity solutions for 5G [3], [4]. In fact, data duplication entails that redundant packets are transmitted for the sake of reducing the probability of missing packets to the lowest possible extent.

A. Data Duplication in the Literature

Although in principle a known technique, data duplication has recently gained special interest due to its relation to 5G solutions. A comparison of data duplication approaches involving different radio access technologies is presented in [2], where specific system design aspects are discussed. An information-theoretic reliability analysis of URLLC with data duplication was conducted in [5], with particular focus on short packet transmission. As regards reliability enhancements of the control plane, recent solutions involving data duplication are available as well (see, e.g., [6]).

On the Radio Access Network (RAN) protocol level, works that address specific challenges of implementing data duplication in 5G setups have also appeared. An overview of a relevant theoretical framework is available in [7], including also useful architectural aspects. Recently, resource efficient data duplication approaches gained interest, attempting to mitigate the negative effect of excessive resource usage for the sake of duplication. Such works include [8], [9]. In [8] the concept of cancelling downlink (DL) Protocol Data Units (PDUs) at the lower layers of RAN protocol stack is introduced; in [9] coordination aspects of uplink (UL) duplication mechanisms are addressed, including a heuristic solution to the problem of optimally configuring the available resources over multiple links.

B. Contribution and Structure

This paper advances existing considerations on data duplication by proposing and assessing a new packet coordination scheme. It particularly addresses New Radio (NR) architecture setups where the base station is split into a Central Unit (CU) and several Distributed Units (DUs), where a User Equipment (UE) may be simultaneously connected to multiple DUs.

Special attention is paid to a careful design of a feedback and coordination scheme between the involved network elements. This facilitates a quick recovery of missing packets and thus increases the overall system reliability. In this respect, acknowledgment messages are introduced to the PDCP layer of the RAN protocol stack, accounting thus for a layer 2 coordination of data duplication. The benefits of this scheme are particularly visible in scenarios where the involved links between UE and DUs are imbalanced (that is, they are associated with considerably different relative strengths), yielding interesting insights.

The remainder of this paper is structured as follows. Section II presents the considered architecture. Section III introduces the proposed packet duplication scheme. Sections IV and V present the simulation assumptions and results to assess the value of the proposed scheme, respectively. Section VI summarizes the main conclusions.

II. Considered Architecture

Next, the network architecture considered in the analysis of data duplication and the corresponding simulation campaign is presented. We first provide an overview of the architectural issues relevant to NR, followed by a discussion on the particular architectural characteristics needed for employing data duplication in scenarios with imbalanced links.

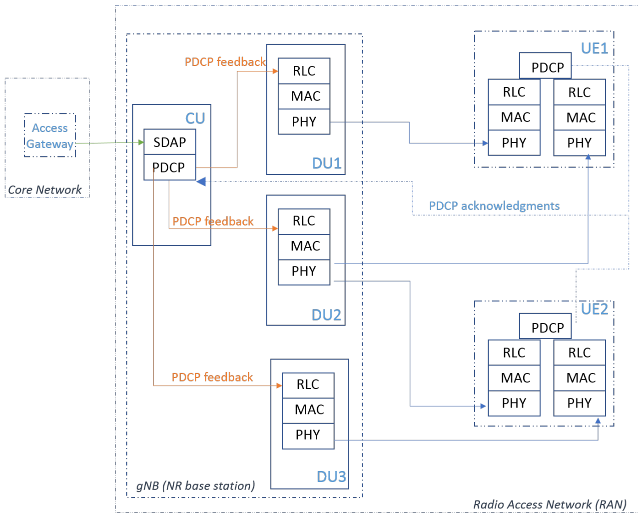

Fig. 1. The considered network architecture involving gNB split

A. Split Architecture

The network architecture considered throughout this work is in accordance with the NR architecture in 5G standards, as described in [10]. This architecture involves the use of two separate network entities that consist the base station (referred to as "gNB" in NR). These components of the gNB are known as the CU and DU, where different layers of the protocol stack are carried out. These two entities are connected with each other via a fronthaul interface, referred to as F1 interface in the specifications [11].

Fig. 1 provides an elaborated view of the considered architecture. As seen in Fig. 1, a single CU and multiple respective DUs have been included, where multiple DUs are connected to one CU and multiple CUs are directly connected to the access gateway at the core network. The Service Data Adaptation Protocol (SDAP) and PDCP functionalities are carried out at the CU, while the Radio Link Convergence (RLC), Medium Access Control (MAC), and Physical layer (PHY) functionalities are executed at the respective DUs.

B. Introducing PDCP-Level Coordination

The architectural approach described so far is the typical state-of-the-art approach adopted in 5G [10], [11]. It is highlighted that, in the context of this work, a new feature is proposed and assessed, which accounts for *introducing a new level of acknowledgment messages (ACKs)*. This new level of acknowledgment is applied at the PDCP layer, located at the CU.

It is noted that in the LTE standard, two levels of packet acknowledgments are employed, which are embedded in the Automatic Repeat Request (ARQ) and Hybrid ARQ (HARQ) protocols carried out at the RLC and MAC layers, respectively. Nevertheless, the existence of such acknowledgment procedures in the DU does not facilitate the coordination of duplicate packets across different DUs. This is because, due to the different channel conditions involved, the RLC and MAC packets are DU-specific, in the sense that different DUs are associated with different payload added to the RLC PDUs. As a result, the packet numbering of RLC PDUs of e.g. DU1 cannot be interpreted from DU2, and vice versa. Consequently, the only protocol layer which can identify missing packets at both the involved links is the PDCP layer, located at the CU. A detailed description of the process of coordinating duplicate packets is provided in the ensuing section.

III. Considered Data Duplication Protocol

With reference to the aforementioned need for a PDCP-level coordinated procedure, the data duplication method operates as follows. Each time the CU receives a downlink packet from the access gateway for a given UE, it directs a replica of this packet to all DUs which are connected to this UE. The DUs then apply the respective RLC layer processing to their replicas and transmit the packets independently from one another. Accordingly, at the UE side, the packets are received in separate RLC entities and are passed to the receiver PDCP entity, where the replicas are combined, as shown in Fig. 1. Then, if a PDCP packet has been successfully received, an acknowledgement message is generated and fed back to the CU, which then informs the RLC entities at the different DUs to proceed to the next packet whose reception has not been acknowledged by the receiving PDCP entity, yet.

In order to find an adequate set of serving cells for each UE an additional link selection mechanism is applied, which compares the Reference Signal Received Power (RSRP) values of the nearby cells with that of the primary serving cell. As such, cells are added into the serving set of the data duplication mode only if they are associated with an RSRP measurement which is at least as large as the RSRP from the serving cell minus a given margin. This margin is subsequently referred to as "link imbalance threshold"and allows for a tradeoff between efficiency and reliability.

A. Benefits of PDCP-Level Coordination

For a better understanding of the benefit of the proposed method for enhancing the efficiency of data duplication, let us consider the following example. Suppose that the UE is simultaneously connected to DU1 and DU2; DU1 has a strong link to the UE, while the corresponding link from DU2 to the UE is relatively weak. This implies that different modulation and coding schemes (MCS) are deployed in the PHY layer of the two links, such that the link between DU1 and the UE conveys more information per unit time than the link from DU2 to the UE. This further implies that packets which have been already correctly delivered to the UE via DU1 are likely to still be pending for transmission from DU2.

The proposed method that involves the use of PDCP packet acknowledgments increases the efficiency of data duplication in utilizing the available resources. As such, a PDCP packet which has been correctly received by the UE via DU1, will generate an acknowledgment message to

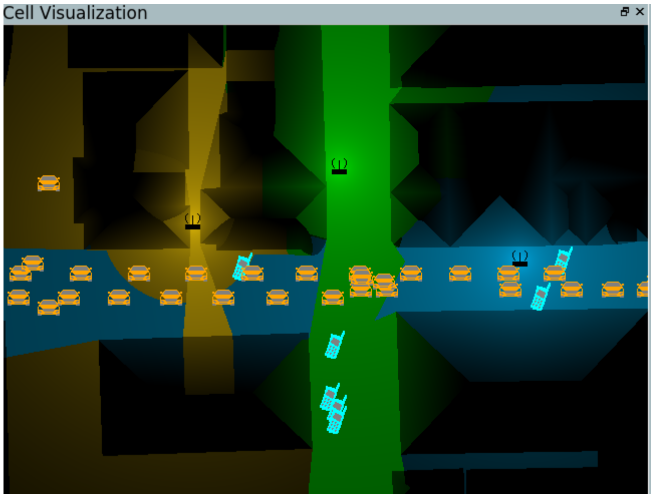

Fig. 2. Visualization of the considered simulation scenario

the CU, which will then notify DU2 to discard this packet from the corresponding RLC entity's transmission buffer. This in fact means that the weak PHY link (i.e., the PHY link between DU2 and UE) will primarily be used for those packets which did not succeed to be transferred via the strong link (i.e., between DU1 and UE).

Such advantage of the proposed data duplication technique as described above reflects in the overall delay and reliability in delivering PDCP packets to the recipient, as will be manifested in Section V where the respective simulation results are shown. Before proceeding to presenting and discussing the obtained results, a detailed description of the simulation setup is provided in the ensuing section.

IV. Simulation Setup

For assessing the performance of the proposed data duplication scheme, a RAN protocol layer simulator was developed. That is, the PDCP, RLC, MAC and PHY layers of the protocol stack were simulated in respective order, using the architecture of Fig. 1. The application layer is also included in the simulator, comprising traffic sources and sinks of given type and UDP protocol agents.

A. Considered Scenario

The simulated scenario involves three outdoor cells, where the transmit power is set to 30dBm each. A transmit time interval (TTI) length of 0.25ms was assumed, with 14 OFDM symbols per TTI. The carrier frequency was set to 3.5GHz, with a system bandwidth of 100MHz. The number of Physical Resource Blocks (PRBs) was set to 10, with a subcarrier spacing of 60kHz.

Fig. 2 provides a snapshot of the considered Manhattan-like simulation scenario. Specifically, the three cells are associated with distinct colors, and are assumed to extend to areas that resemble streets in an urban environment. The shade of the respective colors denotes the RSRP level and the color itself indicates the coverage areas of the different cells. Buildings can be identified based on very dark shades reflecting the penetration loss.

B. Channel, Traffic, and Mobility Model

Within the coverage area of the three cells, 56 UEs are assumed to move in a wrap-around fashion. In particular, the considered UEs are grouped into two major categories, namely pedestrians (marked with light blue cell-phone symbol in Fig. 2), and vehicles (marked with orange car symbol in Fig. 2). The pedestrians are assumed to move along the sidewalks with a speed of 3Km/h, whereas vehicular mobility is controlled by traffic lights at the street crossings and applies random turning models with a maximum vehicle speed of 50Km/h. Whenever a UE reaches the coverage area of neighbouring cells and if certain handover conditions are satisfied, the UE is handed over, i.e., it switches its connection to the new strongest cell. When packet duplication is active, such a hard handover from one cell to another is typically not performed. Instead, as described above in Section III, new cells are added to the serving set based on a link-imbalance criterion *before* the UE is disconnected from its current serving cell.

The considered channel model is that of [12]. It is modified in the sense that the line of sight (LoS) state of any link between two devices is determined explicitly based on the presence of the buildings between them rather than based on distance dependent statistical models. Various data traffic models are considered in order to simulate different use-case and load scenarios, as explained in the ensuing section.

V. Obtained Results

This section summarizes the results obtained from the simulation campaign. The performance assessment of data duplication refers to the downlink, and is carried out with respect to the following Key Performance Indicators (KPIs), namely i) the ability to recover lost packets at the PDCP layer, where duplication is executed; ii) the delay of delivering the packet, measured at the application layer; iii) the overhead in terms of throughput. In each case, a direct comparison with the reference case of no duplication, referred to as "single connectivity", is provided.

Since the performance of data duplication highly depends on the level of offered load to the simulated system, we distinguish between three different scenarios, namely the low, medium, and high load scenario. The results for the corresponding scenarios are presented in Sections V-A and V-B below. Moreover, in Section V-C we investigate the performance of data duplication in restricted areas, i.e., in mission critical areas. This provides insight on the ability of data duplication to reach the ambitious 5G requirements of 99,999% of correct packet reception [13], [14].

A. Low Load Scenario

We first concentrate on the scenario where the generated traffic of the served users corresponds to a relatively low load. Specifically, the traffic in all 56 UEs is assumed exponential with an average of 128Kbps per device. The average

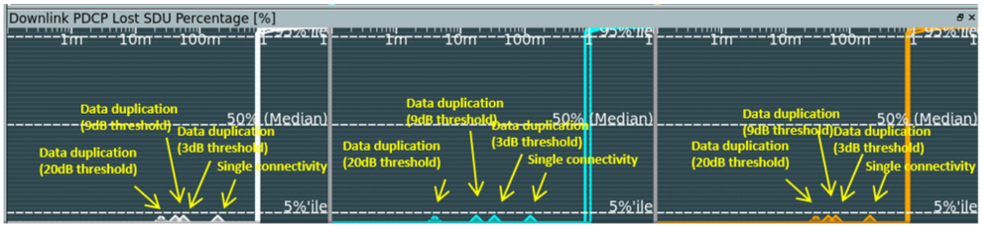

Fig. 3. Low Load Scenario: Percentage of lost PDCP packets for single connectivity (no duplication) and data duplication, under different assumptions on the link imbalance threshold

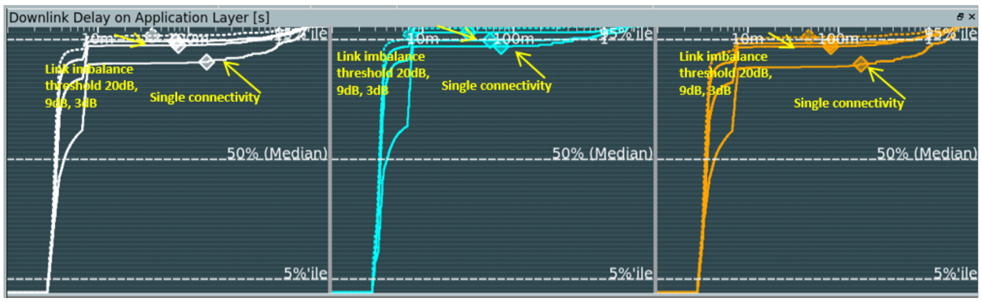

Fig. 4. Low Load Scenario: CDF of packet delivery delay at the application layer

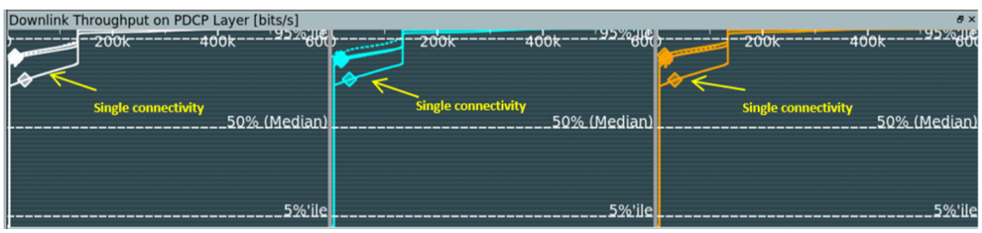

Fig. 5. Low Load Scenario: CDF of throughput for single connectivity and data duplication, for variable values of the link imbalance threshold

burst duration equals 5sec and the idle duration equals 15sec. This corresponds to an overall load of $128 \times 56/4$=1.8Mbps across the entire simulated area.

1) Packet Recovery via Data Duplication: The anticipated benefit of data duplication with respect to reliability at the RAN level is reflected into the percentage of PDCP packets which fail to be successfully transmitted to the UE. This is illustrated in Fig. 3, where the Cumulative Distribution Function (CDF) of the lost PDCP Service Data Units (SDUs) is depicted. In Fig. 3, the light blue lines correspond to the pedestrian UE, the orange to the vehicle UEs, while the white color corresponds to data aggregated over both UE classes. The mean values per group are highlighted and marked with the diamond symbols and presented in the respective color.

In Fig. 3, the signle connectivity case as well as data duplication with different values of the link imbalance threshold (namely 3dB, 9dB and 20dB) have been considered. As can be seen, increasing the link imbalance threshold results in fewer lost packets, since for this case the inclusion of additional links in the multi-connectivity setup is facilitated. In particular, it is observed that without data duplication (single connectivity) approximately 0,25% of PDCP packets are lost, while with data duplication a loss percentage of 0,07% to 0,03% can be achieved, depending on the threshold value (20dB). Nonetheless, as will be shown later, this reduction on the lost packets comes at the cost of decreased throughput, since more resources are utilized for transmitting replicas of the same packet, decreasing thus the overall spectrum utilization efficiency.

2) Delay reduction at application layer: Similar observations with regards to the performance of data duplication are obtained from the application layer packet delivery delay, as depicted in Fig. 4. In particular, it is noticed that a considerable reduction in the packet delivery delay is attained with the activation of data duplication. As expected, such reduction increases with the link imbalance threshold, since an increased value of such threshold via macro diversity results in higher chances that additional links are included, which leads to an overall faster packet delivery.

The observed average values (white marks) of packet delivery are in the range of 170ms for single connectivity, while such values drop to approximately 80ms to 40ms for a link imbalance threshold ranging from 3dB to 20dB. That is, by activating the data duplication mode a decrease on the packet delivery delay of approximately 50% can be achieved, even with relatively small values of the link imbalance threshold.

3) Throughput Overhead: As expected, data duplication introduces a throughput overhead. Such overhead stems from the utilization of redundant radio resources for the sake of reliability, thereby leaving less resources for new data transmission, which ultimately reduces the overall capacity.

The throughput reduction caused by data duplication is quantified in Fig. 5. The main observation is that the use of data duplication decreases the throughput by approximately 50% (that is, a decrease from 32KBps to 15KBps on average). It is further observed that such reduction does not highly depend on the link imbalance threshold. This is anticipated, since low load scenario implies that network resources are scarcely fully occupied, and hence plenty of resources are available to transmit duplicate packets.

B. Medium and High Load Scenario

We now examine the performance of data duplication in scenarios with higher load. This is expected to lead to a deteriorated throughput performance, since in a highly loaded system the additional resource consumption caused by duplicate transmissions has a stronger impact to system performance. In particular, the investigated scenarios pertain to the cases of i) constant bit rate traffic of 200Kbps per device in all 56 devices, corresponding to an overall system load of 11.2Mbps; ii) file transfer protocol (ftp) traffic of 5Mbytes every second per device, for all 56 devices, corresponding to an overall load of 2.2Gbps. The above two cases are referred to as "medium load scenario" and "high load scenario", respectively.

Medium Load Scenario: Figs. 6, 7 and 8 illustrate the percentage of lost PDCP packets, the delay at the application layer and the mean throughput, respectively, in the medium load scenario. In principle, as regards the relative performance of data duplication with respect to single connectivity, similar observations can be made as with the case of low load, in the sense that higher threshold leads to better packet loss and delay performance, yet to higher throughput.

In particular, Fig. 6 shows that, with the exception of the 20dB threshold case, the medium load scenario leads to a larger percentage of lost packets than the low load scenario. Interestingly, we observe a high dependence of the application layer delay (c.f. Fig. 7), as well as of the mean throughput (c.f. Fig. 8) on the link imbalance threshold. Such effect is less visible in the low load scenario (c.f. Fig. 4 and Fig. 5), since in that case that the additional resources used for duplication lead to a saturation of the available resources only in rare cases.

High Load Scenario: Figs. 9, 10 and 11 depict the percentage of lost PDCP packets, the delay at the application layer and the mean throughput, respectively, of the high load scenario. The main observations are as follows. First, data duplication demonstrates a limited capacity to recover lost packets, corresponding to a packet loss drop from approximately 0.2% to 0.1%. As shown in Fig. 9, this effect hardly depends on the value of the applied link imbalance threshold.

More importantly, the observed application layer delay does not improve with the use of data duplication in the high load scenario (c.f. Fig. 10); it is further deteriorated as the value of link imbalance threshold grows large. This effect is explained by the resource saturation due to the high load, resulting in an inefficient use of resources when data duplication is activated. In a similar context, the mean throughput drops when data duplication is active, yet the effect of the link imbalance threshold is less visible.

C. On the Performance Limits of Data Duplication

So far, the simulation results correspond to the entire simulation area, as this is depicted in Fig. 2. From a close observation of the obtained results, one can infer that the strict requirements of ultra reliable services in 5G, corresponding to 99.999% of correct packet reception, are not met. Nevertheless, given that such strict requirements correspond to mission critical services, it is natural to consider that such services are supported in limited areas only, i.e., in areas of critical importance.

In view of this, we repeat the simulation campaign in a similar way as before, yet now we capture the performance of the UEs located within a restricted geographical area, corresponding to the black box in Fig. 13. For the same reason, only the low load scenario is considered, as an attempt to investigate the performance limits of data duplication in special areas. The packet size has been set to 32 bytes to match the assumptions of [13]. The results pertaining to the considered KPIs are depicted in Fig. 12 and explained as follows.

Packet Recovery via Data Duplication: In certain restricted areas with sufficient coverage, data duplication leads to a substantial improvement of lost packets. The corresponding reliability levels can even exceed the target of 99.999% with proper configuration of the link imbalance threshold, as shown in Fig. 12.[1]

Delay Reduction at Application Layer: Similar to the packet loss KPI, a proper configuration of data duplication can lead to a considerable reduction of the application layer delay as compared to the single connectivity (no duplication) case. As demonstrated in Fig. 12, the 95%ile of the delay CDF can be as low as 3ms to 4ms for the case of 20dB link imbalance threshold. It is noted that while this is a relatively low value, it is still beyond the ambitious target of 1ms for 32 byte packets, as set in [13].

Throughput Overhead: It is observed from Fig. 12 that the relative throughput reduction due to data duplication is at approximately the same levels as with the case of non-restricted simulation area, shown in Figs. 5, 8, and 11.

VI. Conclusions

A data duplication scheme that applies to the CU-DU NR architecture was proposed, that involves a coordination process of the duplicate packets at layer 2 (PDCP)

[1]For pedestrian UEs (blue lines) and link imbalance threshold 20dB, the number of lost packets was smaller than the measurement capability of the deployed simulation. This case is therefore not included in Fig. 12.

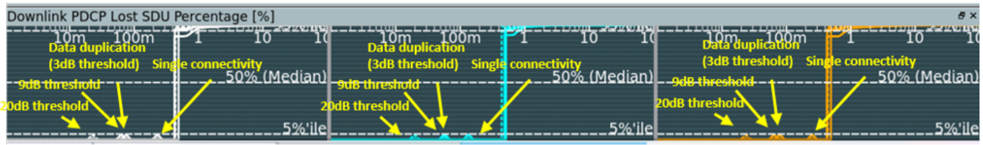

Fig. 6. Medium Load Scenario: Percentage of lost PDCP packets for single connectivity (no duplication) and data duplication, under different assumptions on the link imbalance threshold

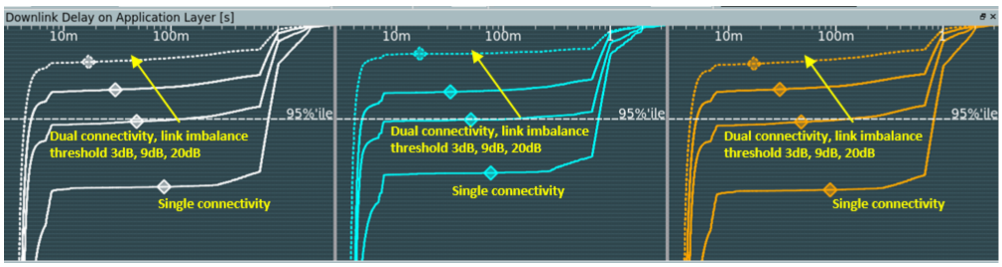

Fig. 7. Medium Load Scenario: CDF of packet delivery delay at the application layer

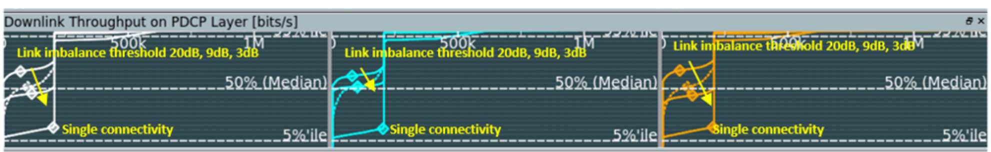

Fig. 8. Medium Load Scenario: CDF of throughput for single connectivity and data duplication, for variable values of the link imbalance threshold

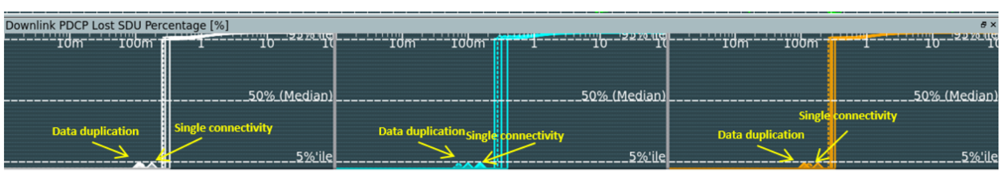

Fig. 9. High Load Scenario: Percentage of lost PDCP packets for single connectivity (no duplication) and data duplication, under different assumptions on the link imbalance threshold

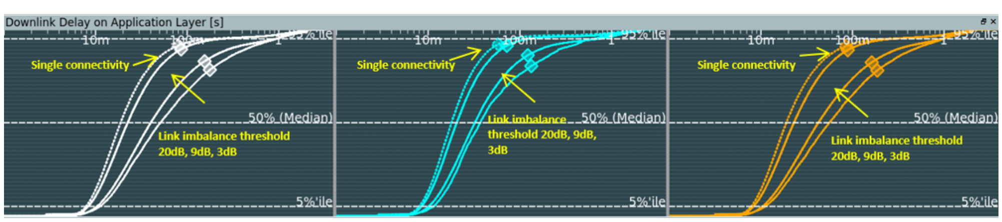

Fig. 10. High Load Scenario: CDF of packet delivery delay at the application layer

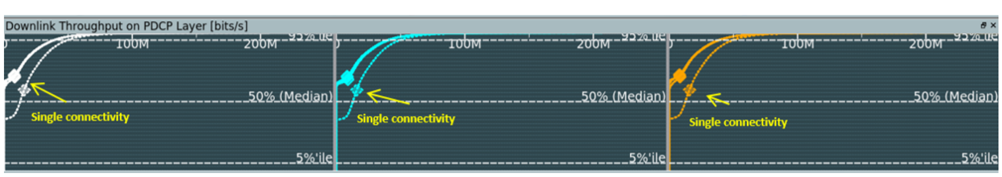

Fig. 11. High Load Scenario: CDF of throughput for single connectivity and data duplication, for variable values of the link imbalance threshold

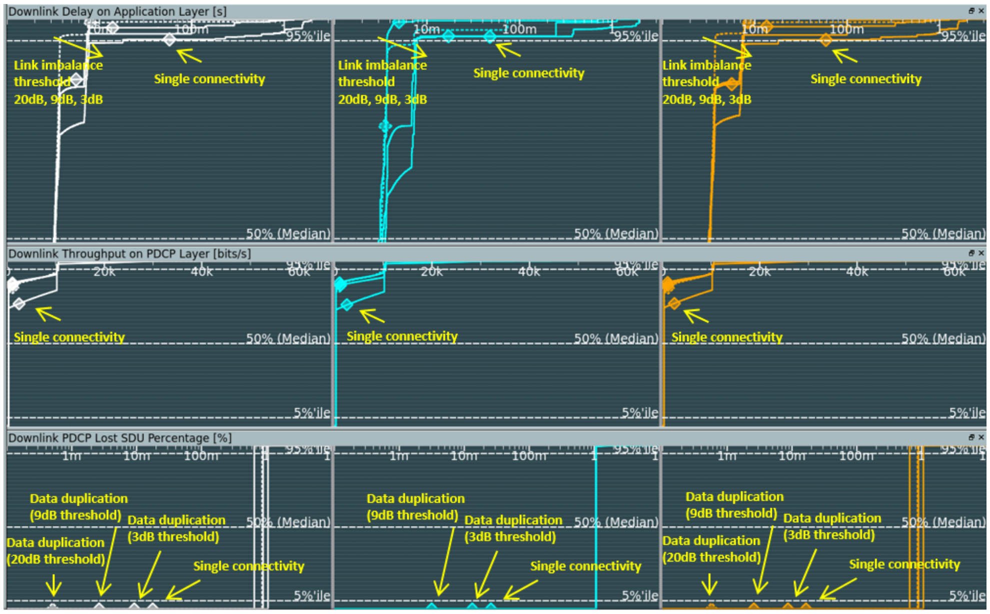

Fig. 12. Performance in terms of the KPIs of interest within a restricted area

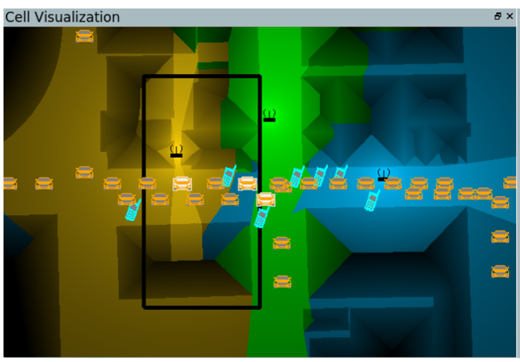

Fig. 13. The restricted area of the simulation scenario where the KPIs of interest are captured

of RAN protocol stack. It was demonstrated via protocol level simulation that network reliability can be improved if properly configured. This, however, comes at a cost in throughput, rendering data duplication suitable for mission-critical applications where reliability is of higher importance than throughput.

The sensitivity of data duplication to network setups with different load was investigated, and was shown that high reliability levels can be reached in low loaded scenarios. As regards the stringent requirements of 5G, it was shown that packet loss rates of 10^{-5} and below can be achieved only in confined multi-cell areas, where conventional single connectivity fails to guarantee the targeted packet loss rate.

ACKNOWLEDGMENT

This work has been performed in the framework of the H2020-ICT-2016-2 project 5G-MoNArch. The authors would like to acknowledge the contributions of their colleagues. This information reflects the view of the consortium, but the consortium is not liable for any use that may be made of any of the information contained therein.

REFERENCES

[1] M. Bennis, M. Debbah and H. V. Poor, *"Ultra-Reliable and Low-Latency Wireless Communication: Tail, Risk and Scale"*, Proceedings of the IEEE (in press), Oct 2018. Online: https://arxiv.org/abs/1801.01270

[2] P. Popovski, J. J. Nielsen, C. Stefanovic, E. de Carvalho, E. Stroem, 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 (URLLC): Principles and Building Blocks"*, submitted to IEEE Network, Dec 2017. Online: https://arxiv.org/abs/1708.07862

[3] A. Ravanshid *et al.*, *"Multi-Connectivity Functional Architectures in 5G"*, IEEE International Conference on Communications Workshops (ICC), 2016

[4] D. S. Michalopoulos, I. Viering, and L. Du, *"User-Plane Multi-Connectivity Aspects in 5G"*, International Conference on Telecommunications (ICT), 2016

[5] G. Durisi, T. Koch and P. Popovski, *"Toward Massive, Ultrareliable, and Low-Latency Wireless Communication With Short Packets"*, Proceedings of the IEEE Network, vol. 104, no. 9, Sep 2016.

[6] T. Fehrenbach, R. Datta, B. Gktepe, T. Wirth and C. Hellge, *"URLLC Services in 5G - Low Latency Enhancements for LTE"*, IEEE Vehicular Technology Conference (VTC-Fall), Oct 2018.

[7] J. Rao and S. Vrzic, *"Packet Duplication for URLLC in 5G: Architectural Enhancements and Performance Analysis"*, IEEE Network, vol. 32, no. 2, March-April 2018

[8] M. Lopez Lechuga, *"Multi-Connectivity in 5G New Radio: Configuration Algorithms and Performance Evaluation"*, MSc Thesis, Catalunya Polytechnic University and Aalborg University, May 2018

[9] J. Rao and S. Vrzic, *"Packet Duplication for URLLC in 5G Dual Connectivity Architecture"*, IEEE Wireless Communications and Networking Conference (WCNC), 2018

[10] 3GPP TR 38.801, *"Technical Specification Group Radio Access Network; Study on new radio access technology: Radio access architecture and interfaces. (Rel. 14)"*, Mar, 2017

[11] 3GPP TR 38.473, *"NG-RAN; F1 application protocol (Rel. 15)"*, Sep, 2018

[12] 3GPP TR 38.901, *"Study on Channel Model for Frequencies from 0.5 to 100 GHz (Rel. 14)"*, Jul, 2017

[13] 3GPP TR 38.913, *"Study on scenarios and requirements for next generation access technologies (Rel. 15)"*, Aug 2017

[14] 3GPP TS 22.261, *"Service requirements for the 5G system; Stage 1 (Rel. 16)"*, Sep 2018