Low-Complexity Centralized Multi-Cell Radio Resource Allocation for 5G URLLC

Ali Karimi[1], Klaus I. Pedersen[1,2], and Preben Mogensen[1,2]
[1]*Wireless Communications Networks (WCN) Section, Department of Electronic Systems, Aalborg University, Denmark.*
[2]*Nokia-Bell Labs, Aalborg, Denmark.*
alk@es.aau.dk

***Abstract*—This paper addresses the problem of downlink centralized multi-cell scheduling for ultra-reliable low-latency communications in a fifth generation New Radio (5G NR) system. We propose a low-complexity centralized packet scheduling algorithm to support quality of service requirements of URLLC services. Results from advanced 5G NR system-level simulations are presented to assess the performance of the proposed solution. It is shown that the centralized architecture significantly improves the URLLC latency. The proposed algorithm achieves gains of 99% and 90% URLLC latency reduction in comparison to distributed scheduling and spectral efficient dynamic point selection.**

***Index Terms*—5G New Radio, URLLC, Centralized scheduling, Radio resource allocation, Dynamic point selection.**

I. Introduction

Supporting ultra-reliable low-latency communications (URLLC) is one of the major goals of the fifth generation New Radio (5G NR) systems [1], [2]. The third generation partnership project (3GPP) has defined several quality of service (QoS) requirement levels for URLLC. The most extreme one has a short latency budget of one millisecond (msec) for transmission of small payloads with 99.999% reliability [3].

Recently, lots of attention has been focused on developing the theoretical bases and engineering-related protocols for URLLC. In [4] and [5], the authors investigate the principles of finite-block length (FBL) wireless data transmission. Motivated by the results from FBL communications, multi-user resource allocation and optimization of data and metadata transmission for URLLC are studied in [6] and [7], respectively. The research in [8] outlines several challenges that 5G NR faces to enable URLLC. Especially, extreme latency requirements of URLLC services need network design to support low processing/transmission times, short hybrid automatic repeat-request (HARQ) delay, and also to accomplish with unprecedented queuing delay [9], [10]. The use of flexible numerology with dynamic transmission times and user-centric adjustment of control information are investigated in [11]. The studies in [12] and [13] present performance analysis of HARQ retransmission for URLLC.

Several studies have found that URLLC performance is overshadowed even with temporary queuing delays [14], [15]. This is further exacerbated with growing traffic. The authors in [16] present a method to reduce the network traffic by dropping delayed payloads. Dynamic link adaptation is exploited in [14] to effectively allocate the resources. An efficient packet scheduling algorithm is proposed in [17] to reduce the tail of queuing delay for URLLC.

In this paper, we investigate the potentials of centralized radio access network (C-RAN) to reduce the queuing delay and enhance URLLC performance in the downlink (DL). Motivated by previous study in [15] we discuss C-RAN architecture with multi-cell multi-user scheduling, where each user can be scheduled from a cluster of connected cells. An optimization problem is formulated to maximize the network capacity of supporting URLLC payloads. As the optimal solution requires high computational complexity, a heuristic, low-complexity algorithm is proposed. In comparison to [15], this includes an improved scheduling metric and an enhanced frequency-selective multi-user resource allocation.

We present numerical results from dynamic advanced system-level simulations with high degree of realism. The results show that proposed centralized scheduling provides outstanding URLLC performance in comparison to those of distributed scheduling [17] and the spectral-efficient dynamic point selection (DPS) [18].

The paper is organized as follows: We present the system model in Section II. The optimization problem and the proposed solution are discussed in Section III and IV, respectively. Simulation methodology and performance results are presented in Section V. Finally, Section VI concludes the paper.

II. Setting the Scene

A. System Model

As outlined in [15], [18], we focus on the C-RAN architecture for the DL transmission operating in frequency division duplexing (FDD) mode. A C-RAN is connected to seven radio remote heads (RRHs). Each RRH has three cells with sectorized deployment and is

978-1-7281-3106-1/20/$31.00 ©2020 IEEE

responsible for physical layer functionalities. The upper layer network protocols are hosted in the C-RAN.

A set of U URLLC user equipments (UEs) are randomly placed over the entire geographical area covered by the network with uniform distribution. Each UE is subject to the DL transmission of B bytes data packets that need to be successfully received within a short latency target of Υ^{tar}. Packet arrival is modelled as a Poisson point process with average of λ [payload/sec/UE]. The average offered load is thus equivalents to $L = C^{-1} \times U \times \lambda \times B \times 8$ [bps/cell], where $C = 21$ is the total number of cells.

The UEs are multiplexed on a shared channel of 20 MHz bandwidth using orthogonal frequency division multiple access (OFDMA) with 30 kHz sub-carrier spacing. The transmission time interval (TTI) equals two or four OFDM symbols (≈ 0.072 and 0.143 msec). A physical resource block (PRB) of 12 sub-carriers is assumed as the minimum physical allocation unit.

B. Channel Measurement and Cell Connectivity

Each UE periodically performs measurements by estimating the received signal reference power (RSRP) of the cells. The UE connects to one/multiple serving cell(s) as follows:

1) Baseline (Distributed scheduling): The UE connects to only the cell with the highest average RSRP value. Channel state information (CSI) measurement is performed for the connected cell and the corresponding post-receiver channel quality indicator (CQI) is reported to the network.

2) DPS: As outlined in [18], the UE connects to a cluster of maximum Q cells that are within a RSRP power window of W dB from the cell with the highest RSRP value. The UE periodically measures the CSIs of the connected cells and reports the CQI for the cell with the highest instantaneous estimated throughput (TP).

3) Centralized Scheduling: The UE performs CSI measurements for the cells in its cluster. The CQIs are reported to the network.

C. URLLC Latency Components

Focussing on the one-way latency (Υ) in the DL transmission, the latency components are given by:

$$\Upsilon = d_{fa,q} + d_{bsp} + d_{tx} + d_{uep} + d_{HARQ}, \quad (1)$$

where $d_{fa,q}$ denotes the frame alignment and queuing delay of the initial transmission. The processing times required for scheduling the payload at the network and that for decoding the data at the UE are presented by d_{bsp} and d_{uep}, respectively. The transmission time is denoted by d_{tx}. The imposed delay by HARQ retransmission(s) is denoted by d_{HARQ}. We set $d_{HARQ} = 0$ if the packet is received correctly within the first transmission. In case of failure, a minimum HARQ delay equal to 12 OFDM symbols is assumed [14]. The processing times d_{bsp} and d_{uep} are constant, equal to $d_{bsp} = 2.75$ and $d_{uep} = 3.25$ OFDM symbols, respectively [19].

III. Problem Formulation

Let us assume D^t and $\mathbf{\Gamma}_u$ represent the total number of PRBs and the set of cells in the measurement cluster of UE u (those the UE can be scheduled from). The set of assigned PRBs to UE u from cell c ($c \in \mathbf{\Gamma}_u$) is denoted by $\mathbf{P}_{uc}$ and the corresponding achievable data rate is shown by $R_{\mathbf{P}_{uc}}$. For each TTI, we formulate an optimization problem for cell and PRB allocation as follows:

$$\begin{aligned}
&\max_{x_{uc}, p_u^j} \sum_u \sum_c a_u x_{uc} R_{\mathbf{P}_{uc}} \\
&\text{Sub. to C1: } \sum_c x_{uc} \leq 1, \qquad \forall u, \\
&\qquad \text{C2: } \sum_j p_u^j \geq x_{uc} p_{uc}^{\min}, \quad \forall u, c, \\
&\qquad \text{C3: } R_{\mathbf{P}_{uc}} \leqslant Q_u, \qquad \forall u, \\
&\qquad \text{C4: } \sum_u x_{uc} p_u^j \leq 1, \quad \forall c, j, \\
&\qquad \text{C5: } x_{uc}, p_u^j \in \{0, 1\}, \quad \forall u, c, j, \qquad (2)
\end{aligned}$$

where x_{uc} is a binary variable accounting for the association of UE u to cell c. Binary variable p_u^j is the PRB allocation indicator equals one if j-th PRB ($1 \leq j \leq D^t$) is assigned to UE u, otherwise it is zero. Parameters a_u is a weighting factor responsible for guaranteeing latency requirements of UE u. $p_{uc}^{\min}$ is the number of required PRBs to transmit the control information if UE u is scheduled by cell c. The buffered data of u-th UE is denoted by Q_u.

The first constraint in (2) indicates that each UE can be scheduled from maximum one cell per TTI. The second constraint ensures the minimum allocated resources for sending the scheduling grants. Constraint C3 imposes that the achievable rate of each UE is bounded by that of the buffered data. Finally, the orthogonality of resource allocation is ensured by C4. The problem (2) is a non-convex combinatorial optimization that is difficult to solve in polynomial time and within the stringent processing time requirement of URLLC services [15], [17]. Assuming each UE is connected to Q cells, the cell selection has the complexity of $\mathcal{O}((Q+1)^U)$. The complexity further increases by considering PRB allocation in frequency domain.

IV. Proposed Solution

We present a heuristic low-complexity solution for (2). The following describes the steps of the proposed solution.

In line with [17], we first prioritize scheduling HARQ retransmissions. The C-RAN schedules HARQ retransmissions immediately from the cells with the best link

quality (highest TP) and over a set of PRBs with the highest CQI values, aiming to avoid additional queuing delay and minimize the need for more retransmission(s). After allocating HARQ retransmissions, UEs with pending data are scheduled. We split the problem (2) into two sub-problems of cell and PRB allocation. First, the C-RAN determine the UEs and cells allocations. Afterwards, frequency domain (FD) PRB allocation is performed for the UEs allocated to the same cell.

For a given UE, the achievable data rate depends on CSI and the assigned PRBs. Dealing with small URLLC payloads, the overhead of transmitting control information is not negligible and can not be ignored [14], [20]. To manage the cost of sending multiple physical downlink control channels (PDCCHs), we aim for scheduling full data payloads in one transport block and not to split them over multiple TTIs (aka as packet segmentation). As detailed in [14], the PDCCH size is adjusted based on the wide-band CQI to ensure low-failure probability. We define pair(s) of each UE with pending data and the cell(s) in its measurement set. For a given pair (u, c) of UE u and cell c, we denote P_{uc} as the estimated number of required PRBs for sending both the PDCCH and data. The length of the data block can be estimated from the wide-band CQI.

The UE/cell allocation is performed sequentially by selecting a UE/cell pair with the highest scheduling metric. For a pair (u, c), the scheduling metric m_{uc} is defined as:

$$m_{uc} = \Upsilon_u + \frac{TP_{uc}}{TP_{\mathbf{\Gamma}u}}, \tag{3}$$

where Υ_u [TTI] is the head-of-line delay for UE u. We denote TP_{uc} as the estimated full-bandwidth TP of UE u if served by cell c. Variable $TP_{\mathbf{\Gamma}_u}$ is the sum TPs of all serving cells in the measurement set of UE u that equals $TP_{\mathbf{\Gamma}_u} = \sum_{c \in \mathbf{\Gamma}_u} TP_{uc}$. The role of Υ_u in (3) is to minimize the outage latency by prioritizing UEs which are closer to the latency target. The term TP_{uc} exploits channel aware scheduling. For a given UE, a cell with the best CSI (primary cell) receives higher priority for being selected. Finally, the term $TP_{\mathbf{\Gamma}_u}$ acts as normalizing factor. Cell-centred UEs (which usually connect to one cell) will get a higher chance for being scheduled from the primary cells. On the other hand, cell-edge UEs (that usually have multiple connections with similar signal strengths) are assigned relatively lower values as they have a chance for being served by the secondary cells (if the primary cells are overloaded).

The C-RAN sequentially selects a pair (u, c) with the highest scheduling metric m_{uc} and assign the UE to the cell if the number of available PRBs at cell c (D_c) is enough for scheduling a full payload (i.e. $D_c \geq P_{uc}$). Afterwards the C-RAN removes the other pairs associated with the selected UE and update the number of available PRBs at the serving cell. The procedure continues until all UEs are scheduled or there are not enough resources at the serving cells. The approximated complexity of proposed cell allocation is $\mathcal{O}(Q{\cdot}\,U \log{(Q{\cdot}\,U)})$.

For each cell, the selected UEs are multiplexed in FD by comparing per-PRB metric. Throughput to average (TTA) [17] is adopted. j-th PRB is assigned to UE $\hat{u}$ (i.e. $p^j_{\hat{u}} = 1$) that has the highest metric of

$$\hat{u} = \arg \max_{u \in \mathit{\Pi}_c} \frac{r^j_u}{\bar{\mathbf{r}}_u}, \tag{4}$$

where $\mathit{\Pi}_c$ denotes the set of UEs allocated to cell c. Variables r^j_u and $\bar{\mathbf{r}}_u$ are the u-th UE's achievable TP of PRB j and the instantaneous full-bandwidth TP, respectively.

Finally, the C-RAN checks if it is possible to schedule more UEs on the remaining PRBs after FD allocation. If there are not sufficient PRBs to schedule a full payload, at most one payload is segmented and transmitted over the available PRBs. A UE with the lowest PDCCH overhead is prioritized. Algorithm 1 summarizes details of the proposed solution.

Algorithm 1 Proposed centralized scheduling.

1: Schedule the HARQ retransmissions. Update the available PRBs at the serving cells.
2: Create pairs of the buffered UEs and the cells in their measurement sets.
3: For each pair, calculate P_{uc} and m_{uc}.
4: Create list $\mathbf{S}$ of the pairs, sort $\mathbf{S}$ in decreasing order of the scheduling metric (3).
5: **while** enough PRBs at cells and $\mathbf{S}$ is not empty **do**
6: Select pair (u, c) as the first element of $\mathbf{S}$.
7: **if** $P_{uc} \leq D_c$ **then**
8: Assign UE u to cell c.
9: Update UE $D_c = D_c - P_{uc}$.
10: Remove pairs associated with UE u from $\mathbf{S}$.
11: **else**
12: Remove pair (u, c) from $\mathbf{S}$, put it in the unscheduled list $\overline{\mathbf{S}}$
13: **end if**
14: **end while**
15: **for** $c = 1 : C$ **do**
16: Calculate the FD scheduling metric (4) for the available PRBs and the users assigned to cell c.
17: Sort PRBs in decreasing order of (4).
18: Allocate PRBs to UEs. Remove if there is segmented payload.
19: **if** there are unallocated PRBs **then**
20: Schedule a full payload from $\overline{\mathbf{S}}$ or segment a UE with low PDCCH usage if PRBs are not enough.
21: **end if**
22: **end for**

TABLE I
DEFAULT SIMULATION ASSUMPTIONS.

Description	Assumption
Environment	3GPP Urban Macro (UMa); 3-sector RRHs with 500 meters inter-site distance. 21 cells.
Propagation	Urban Macro-3D
Carrier	2 GHz (FDD), 20 MHz carrier bandwidth
PHY numerology	30 kHz sub-carrier spacing. PRB size of 12 sub-carriers (360 kHz).
TTI sizes	0.072 and 0.143 msec corresponding to 2 and 4 OFDM symbols mini-slot, respectively.
MIMO	Single-user 2x2 closed loop single-stream (Rank-1) configuration. LMMSE-IRC receiver.
CSI	Periodic CSI measurement every 5 msec, with 2 msec latency for being applied in DL transmission. One CQI per a sub-channel of eight PRBs.
MCS	QPSK to 64QAM, with same encoding rates as specified for LTE.
Link adaptation	Dynamic MCS with outer-loop link adaptation to achieve 1% BLER of initial transmission.
HARQ	Asynchronous HARQ with Chase-combining [21]. HARQ delay equals 12 OFDM symbols.
User distribution	2100 URLLC users (Average 100 users per cell).
Traffic model	FTP3 downlink traffic with payload sizes of $B = 32$ and $B = 50$ bytes.
Link-to-system (L2S) mapping	Based on MMIB mapping [22].

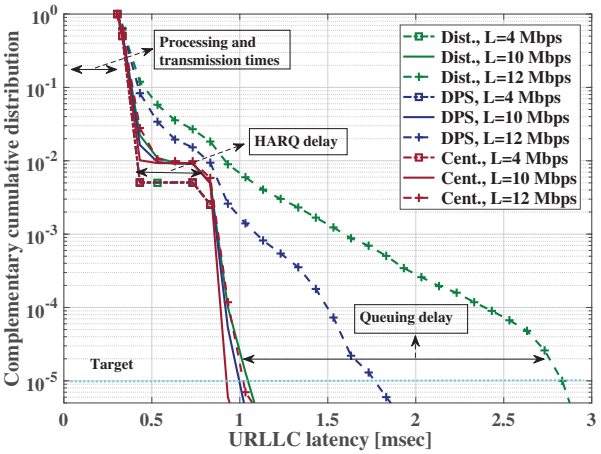

Fig. 1. URLLC latency for different scheduling methods assuming $B = 32$ and two OFDM symbols TTI size.

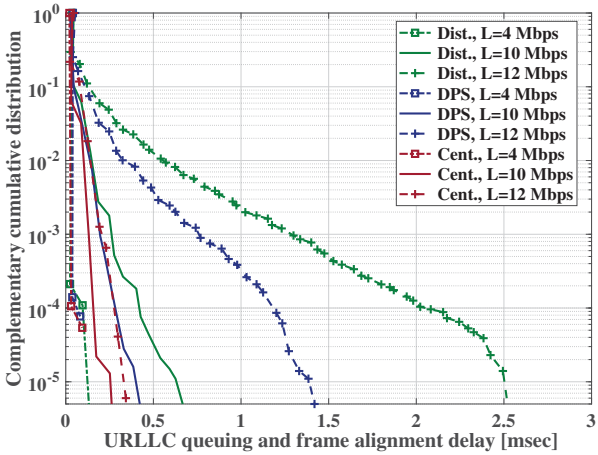

Fig. 2. Queuing and frame alignment delay for different scheduling methods assuming $B = 32$ and two OFDM symbols TTI size.

V. NUMERICAL EVALUATIONS

Performance results are obtained by running advanced 5G NR system-level simulations in line with the 3GPP NR guidelines as outlined in Section II [1], [23]. The summarized assumptions are presented in Table I. The simulation time is set so at least five million packets to obtain reliable results. We compare the proposed C-RAN scheduling against the distributed allocation [17] and DPS [18]. In line with [18], for centralized scheduling and DPS we set $Q = 2$ cells and $W = 3$ dB. Applying these settings, only 32% of UEs have two cells in their measurement set while 68% of them connect to only one cell [18]. This further reduces the complexity of cell allocation at the C-RAN as well as the overhead and the complexity of channel measurement at the user-side.

A. One millisecond latency performance

First we focus on achieving one msec latency at the outage probability of 10^{-5}, assuming $B = 32$ bytes and a short TTI size of two OFDM symbols. The complementary cumulative distribution function (CCDF) of the latency is depicted in Fig. 1. At low-offered loads, one msec latency at 10^{-5} outage is fulfilled for all the scheduling methods. The reason is that the probability of experiencing queuing delay is very low. Therefore, the performance is mainly impacted by the transmission time, processing delays, and HARQ delay.

The probability of the queuing increases with the offered load. The centralized scheduling shows clear latency benefits by instantly offloading the congested cells. The results in Fig. 1 show that one msec latency can not be achieved for the distributed and DPS if the traffic exceeds 10 Mbps, whereas the C-RAN can tolerate up to 12 Mbps load while fulfilling URLLC requirements. This is equivalent to 20% improvement in the network capacity.

Fig. 2 illustrates the CCDF of the queuing and frame alignment. As expected, the queuing delay is negligible at low-load so that most of the packets are immediately

TABLE II
URLLC PERFORMANCE AT THE OUTAGE OF 10^{-5} FOR DIFFERENT NUMEROLOGIES AND OFFERED LOADS.

Scenario		TTI = 2 symbols, B = 32 bytes Load [Mbps/cell]				TTI = 4 symbols, B = 50 bytes Load [Mbps/cell]			
		4	8	10	12	12	16	18	20
Distributed [msec]		0.83	0.92	1.05	2.8	1.23	7.8	45	500
DPS [msec]		0.83	0.92	1	1.75	1.23	3.83	16.4	43
Centralized [msec]		0.83	0.89	0.91	1	1.13	1.23	1.53	4.95
Relative gain to	Distributed	0%	3%	13%	64%	8%	84%	97%	99%
	DPS	0%	3%	9%	42%	8%	68%	90%	88%

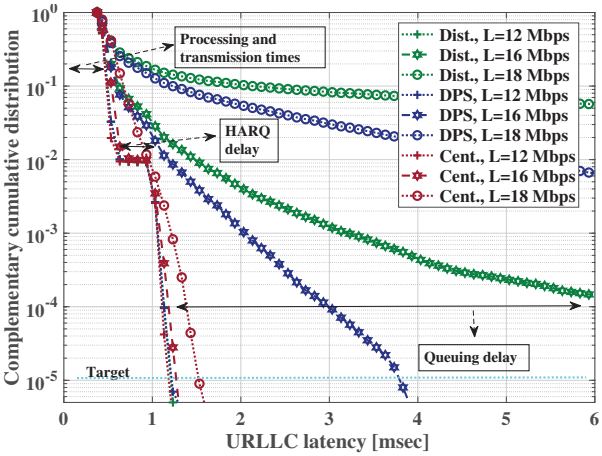

Fig. 3. URLLC latency for different scheduling methods assuming $B = 50$ and four OFDM symbols TTI size.

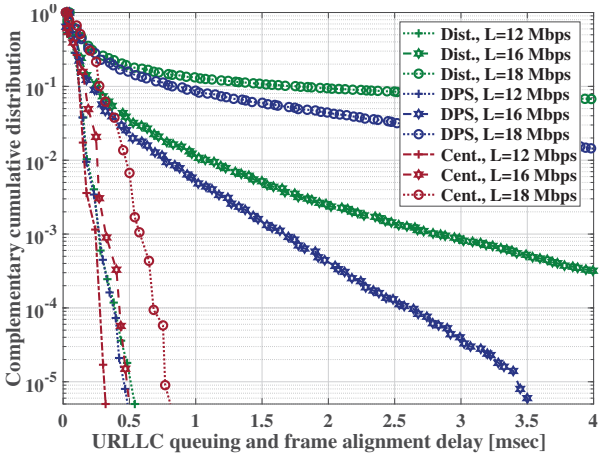

Fig. 4. Queuing and frame alignment delay for different scheduling methods assuming $B = 50$ and four OFDM symbols TTI size.

allocated without waiting for receiving resources. We observe a longer tail of queuing when increasing the offered traffic, indicating that some of cells become temporarily overloaded. For 12 Mbps, we observe that in 99.999% of the cases the queuing delay is below 0.3 msec for the C-RAN. Whereas, this increases to 1.4 and 2.5 msec for DPS and the distributed scheduling, respectively.

B. Beyond one millisecond latency

For cases with larger payload sizes, higher offered loads, and more relaxed latency requirements (e.g. two or five msec), we configure the TTI size to four OFDM symbols. So that a full payload of UEs with low CSIs can fit into one TTI without segmentation. Fig. 3 and Fig. 4 plot the CCDFs of the latency and the queuing plus frame alignment delay for $B = 50$ bytes. In Fig. 3, the outage reliability of 10^{-5} is achieved at 1.25 msec for distributed scheduling, whereas it is 2.8 msec in Fig. 1. That is because for the same volume of traffic, the packet arrival rate for $B = 50$ bytes case is 36% less than that of $B = 32$ bytes. This leads to lower overhead of sending PDCCHs as well as the queuing delay and overall latency.

In comparison to the distributed case, the DPS provides better channel quality specially for the cell-edge UEs. This results in lower latency. For 18 Mbps load, the latency at the outage level of 10^{-5} is 45 msec for the distributed case. It is reduced to 16 msec by applying DPS. We observe that the C-RAN further boosts the performance by reducing the latency to only 1.5 msec. The proposed centralized solution not only benefits from fast-fading channel variations but also exploits the gain of fast load balancing. The superior performance of the C-RAN in improving the queuing delay is highlighted in Fig. 4. At 18 Mbps offered traffic, 20% and 15% of the payloads for the distributed and DPS scheduling experience more than 0.5 msec queuing delay. As compared to then, only 0.5% of the packets experience queuing delay when applying the proposed centralized algorithm.

Table II summarizes the URLLC latency at the outage level of 10^{-5} for different resource allocation methods. The results show the advantage of the C-RAN at high offered traffic. As an example, for 20 Mbps load gains of 99% and 88% latency reduction is achieved in comparison to the distributed scheduling and DPS, respectively.

VI. CONCLUSION AND FUTURE STUDIES

This paper studied DL URLLC performance optimization through the use of C-RAN. We proposed an attractive low-complexity multi-cell scheduling algo-

rithm to improve the network performance of supporting URLLC in 5G NR. Extensive numerical results from 3GPP compliant advanced system-level simulator were presented. The results confirm the high potential of the C-RAN to tackle the undesired queuing delay of URLLC payloads. Performance results show that the proposed solution can achieve remarkable gains of up to 99% and 90% URLLC latency improvements in comparison to the distributed scheduling and DPS, respectively.

As future work, it is of high interest to incorporate dynamic inter-cell interference coordination techniques to further boost URLLC capacity.

REFERENCES

[1] 3GPP Technical Specification 38.300, "NR and NG-RAN overall description; stage-2," Version 15.5.0, March 2019.

[2] IMT Vision, "Framework and overall objectives of the future development of IMT for 2020 and beyond," International Telecommunication Union (ITU), Document, Radiocommunication Study Groups, February 2015.

[3] 3GPP Technical Specification 23.501, "Technical specification group services and system aspects, system architecture for the 5G system," Release 15, December 2017.

[4] W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy, "Quasi-static multiple-antenna fading channels at finite blocklength," *IEEE Transactions on Information Theory*, vol. 60, no. 7, pp. 4232–4265, July 2014.

[5] P. Popovski *et al.*, "Wireless access in ultra-reliable low-latency communication (URLLC)," *IEEE Transactions on Communications*, pp. 1–1, 2019.

[6] W. R. Ghanem, V. Jamali, Y. Sun, and R. Schober, "Resource allocation for multi-user downlink URLLC-OFDMA systems," *CoRR*, vol. abs/1901.05825, 2019. [Online]. Available: http://arxiv.org/abs/1901.05825

[7] A. Karimi *et al.*, "On the multiplexing of data and metadata for ultra-reliable low-latency communications in 5G," *Submitted to IEEE Transactions on Vehicular Technology*, 2019.

[8] M. Bennis, M. Debbah, and H. V. Poor, "Ultra-reliable and low-latency wireless communication: Tail, risk, and scale," *Proceedings of the IEEE*, vol. 106, no. 10, pp. 1834–1853, October 2018.

[9] C. She, C. Yang, and T. Q. S. Quek, "Cross-layer optimization for ultra-reliable and low-latency radio access networks," *IEEE Transactions on Wireless Communications*, vol. 17, no. 1, pp. 127–141, January 2018.

[10] A. Anand and G. de Veciana, "Resource allocation and HARQ optimization for URLLC traffic in 5G wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 11, pp. 2411–2421, November 2018.

[11] K. I. Pedersen *et al.*, "A flexible 5G frame structure design for frequency-division duplex cases," *IEEE Communications Magazine*, vol. 54, no. 3, pp. 53–59, March 2016.

[12] J. P. B. Nadas *et al.*, "Performance analysis of hybrid ARQ for ultra-reliable low latency communications," *IEEE Sensors Journal*, pp. 1–1, 2019.

[13] A. Avranas, M. Kountouris, and P. Ciblat, "Energy-latency tradeoff in ultra-reliable low-latency communication with retransmissions," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 11, pp. 2475–2485, November 2018.

[14] 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. 28 912–28 922, 2018.

[15] A. Karimi *et al.*, "5G centralized multi-cell scheduling for URLLC: Algorithms and system-level performance," *IEEE Access*, vol. 6, pp. 72 253–72 262, 2018.

[16] E. Khorov, A. Krasilov, and A. Malyshev, "Radio resource and traffic management for ultra-reliable low latency communications," in *2018 IEEE Wireless Communications and Networking Conference (WCNC)*, April 2018, pp. 1–6.

[17] A. Karimi *et al.*, "Efficient low-complexity packet scheduling algorithm for mixed URLLC and eMBB traffic in 5G," *In Proc. 2019 IEEE 89th Vehicular Technology Conference - VTC2019-Spring*, May, 2019.

[18] A. Karimi, K. I. Pedersen, and P. Mogensen, "5G URLLC performance analysis of dynamic-point selection multi-user resource allocation," *in Proceeding of 16th International Symposium on Wireless Communications Systems*, August, 2019.

[19] 3GPP Technical Documents R1-1813120, "Discussion on the RAN2 LS on TSN requirements evaluation," November 2018.

[20] N. H. Mahmood *et al.*, "On the resource utilization of multi-connectivity transmission for URLLC services in 5G new radio," *in Proc. 2019 IEEE Wireless Communications and Networking Conference (WCNC)*, April 2019.

[21] D. Chase, "Code combining - a maximum-likelihood decoding approach for combining an arbitrary number of noisy packets," *IEEE Transactions on Communications*, vol. 33, no. 5, pp. 385–393, May 1985.

[22] T. L. Jensen, S. Kant, J. Wehinger, and B. H. Fleury, "Fast link adaptation for MIMO OFDM," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 8, pp. 3766–3778, October 2010.

[23] 3GPP Technical Report 38.913, "Study on scenarios and requirements for next generation access technologies," Version 14.1.0, March 2016.