Punctured Scheduling for Critical Low Latency Data on a Shared Channel with Mobile Broadband

Pedersen, Klaus I.; Gerardino, Guillermo Andrés Pocovi; Steiner, Jens; Khosravirad, Saeed R.

Published in:
IEEE 86th Vehicular Technology Conference (VTC-Fall), 2017

DOI (link to publication from Publisher):
10.1109/VTCFall.2017.8287951

Publication date:
2017

Document Version
Accepted author manuscript, peer reviewed version

Punctured Scheduling for Critical Low Latency Data on a Shared Channel with Mobile Broadband

Klaus I. Pedersen[1,2], Guillermo Pocovi[2], Jens Steiner[1], Saeed R. Khosravirad[1]
Nokia – Bell Labs[1], Aalborg University[2]

Abstract— **In this paper, we present a punctured scheduling scheme for efficient transmission of low latency communication (LLC) traffic, multiplexed on a downlink shared channel with enhanced mobile broadband traffic (eMBB). Puncturing allows to schedule eMBB traffic on all shared channel resources, without prior reservation of transmission resources for sporadically arriving LLC traffic. When LLC traffic arrives, it is immediately scheduled with a short transmission by puncturing part of the ongoing eMBB transmissions. To have this working efficiently, we propose recovery mechanisms for punctured eMBB transmissions, and a service-specific scheduling policy and link adaptation. Among others, we find that it is advantageous to include an element of eMBB-awareness for the scheduling decisions of the LLC transmissions (i.e. those that puncture ongoing eMBB transmissions), to primarily puncture eMBB transmission(s) that are transmitted with low modulation and coding scheme index. System level simulations are presented to demonstrate the benefits of the proposed solution.**

I. INTRODUCTION

Research on the 5G New Radio (NR) is gaining further momentum with the closing of the first Study Item on this subject in 3GPP; see especially the following technical reports [1]-[3]. The ambitions for 5G NR are high, aiming for enhanced support for multiplexing of diverse services such as enhanced mobile broadband (eMBB) and low latency communication (LLC) with ultra-reliability constraints [3]-[5]. Simultaneously fulfilling the requirements for a mixture of users with such diverse requirements is a challenging task, given the fundamental tradeoffs known from communication theory [6]. In that respect, the next generation base station (called gNB) scheduler, which orchestrates the allocation of radio resources to different users, plays an important role. The flexible physical layer design [1] - and especially the agile frame structure design [7] - that comes with the 5G NR offers increased degrees of freedom for the scheduler functionality. Facilitating a shift towards a user-centric approach, where the allocation of radio resources for each user is more flexible, and hence can be better optimized in coherence with the users' diverse QoS requirements. Among others, the 5G NR allows to schedule the users with variable transmission time intervals (TTIs) as proposed also in [7]-[9]. Support for variable TTI sizes facilitates matching the radio resource allocations per user in coherence with their radio conditions and QoS requirements. For instance, to schedule LLC users with short TTIs to achieve low latency, accepting the penalty of higher relative control channel overhead; see the recent studies in [10]-[11] on the benefits of variable TTIs for LLC traffic. Similarly, scheduling with variable TTI sizes also provides advantages for eMBB traffic [9], as it offers a powerful instrument to efficiently adapt to different offered load conditions and the internet transport protocols (TCP) [12] closed-loop flow control mechanisms.

However, despite the benefits of scheduling the users with variable TTI sizes, there are still some non-trivial problems that call for more studies. One of those is how to efficiently multiplex eMBB and LLC on a downlink shared channel, especially for scenarios where the eMBB traffic is primarily scheduled with long TTI sizes, while sporadic arriving LLC traffic must be scheduled immediately with a short TTI size when such payloads arrive at the gNB to fulfill the corresponding latency deadline. In our effort to address this problem, our hypothesis is that a promising solution is to allow punctured scheduling, where a longer ongoing eMBB transmission can be partly replaced (i.e. punctured) by an urgent short TTI transmission to a user with LLC traffic. The fundamental principle of punctured scheduling has some similarities to preemptive scheduling principles as studied extensively for computer networks to accommodate real-time services [13]. However, despite those commonalities, there are several differences and open questions for how to best design punctured scheduling for a 5G NR wireless system. In particular, we study how to minimize the impact on the eMBB users that are harmed (i.e. by overriding part of their transmission). For this purpose, we propose recovery mechanisms for the impacted eMBB users, and suggest custom designed radio resource management (RRM) optimizations to most efficiently multiplex eMBB and LLC traffic, when utilizing punctured scheduling. The proposed methods are evaluated in a dynamic multi-user, multi-cell. Due to the complexity of the 5G NR system and the addressed problems, we rely on advanced system-level simulations for results generation to have high degree of realism. Those simulations are based on commonly accepted underlying models, calibrated with 3GPP 5G NR assumptions [1]-[3], making sure that statistical reliable results are generated.

The rest of the paper is organized as follows: Section II further sets the scene for the study by shortly introducing the system model and presenting the problem formulation and related objectives. The proposed punctured scheduling scheme is outlined in Section III, and the corresponding RRM considerations in Section IV. The performance analysis appears in Section V, while concluding remarks are presented in Section VI.

II. SETTING THE SCENE

A. System model

We adopt the 5G NR assumptions as outlined in [1]-[2], focusing primarily on the downlink performance. Users are dynamically multiplexed on a shared channel, using orthogonal frequency division multiple access (OFDMA). We assume the setting with 15 kHz subcarrier spacing. LLC UEs are scheduled

with short TTI of only 2 OFDM symbols, corresponding to a mini-slot of 0.143 ms. eMBB traffic is primarily scheduled with longer TTI sizes of 14 OFDM symbols (1 ms duration), equivalent to two 7-symbol slots (but could also be scheduled with shorter TTI sizes). In the frequency domain, users can be multiplexed on a physical resource block (PRB) resolution of 12 subcarriers. Users are dynamically scheduled, using a user-centric downlink control channel for transmitting the scheduling grant [7]. This includes informing the users on which resources they are scheduled, which modulation and coding scheme (MCS) is used, etc. Asynchronous hybrid automatic repeat request (HARQ) with Chase combining (as also supported for LTE) is assumed. The system is assumed to carry best effort eMBB traffic download, as well as sporadic LLC traffic. The latter is modeled as bursts of small payload size of B bits that arrive for each LLC user in the downlink direction following a uniform Poisson arrival point process with arrival rate λ. Thus, the offered LLC traffic load per cell equals $N \cdot B \cdot \lambda$, where N is the average number of LLC users per cell.

B. Problem formulation and objectives

The objective is to serve the eMBB users with high average data rates (i.e. maximizing the spectral efficiency), while serving the LLC users per their low latency requirement with ultra-high reliability. The LLC traffic takes priority over the best effort eMBB data flows, and needs to be immediately scheduled when it arrives at the gNB. The dilemma, however, is that due to the random unpredictable nature of the LCC traffic, the gNB has no solid a priori knowledge of when LLC traffic arrives, and hence when to reserve radio resources for such transmissions. Reserving radio resources for potentially coming LLC transmissions would be inefficient as it results in capacity loss for the eMBB users. On the other hand, when scheduling the eMBB users, the downlink shared channel will in principle be monopolized by such transmissions, causing unnecessary latency to the LLC users that suddenly have data coming. This is the problem addressed in this study.

III. PUNCTURED SCHEDULING PROPOSAL

A. Basic principle

The basic principle of the proposed punctured scheduling solution is shown in Fig. 1. Here, a UE with eMBB traffic is scheduled by the gNB for transmission on the downlink shared radio channel with a long TTI of 1 ms. The former is facilitated by the gNB sending a scheduling grant (transmitted on the physical layer control channel) followed by the actual transmission of the transport block. During the transmission time of the transport block for the eMBB UE, the shared channel for this transmission is in principle monopolized. However, it may happen that LLC data for another UE arrives at the gNB while the scheduled transmission towards the eMBB UE is ongoing. To avoid waiting for the completion of the transport block transmission to the eMBB UE, we propose to immediately transmit the LLC data by puncturing (i.e. over-riding) part of the ongoing eMBB transmission. The advantage of this solution is that the latency of the LLC data is minimized, at the expense of lower performance of the transmission to the eMBB UE. As some of the resources for the eMBB transmission are corrupted, it essentially results in an error floor, where the performance in terms of block error probability (BLEP) versus SINR for the UE saturates [14]. The impact on the eMBB UE performance from being punctured naturally depends on multiple factors: how many resources have been punctured, whether the eMBB UE is aware of the puncturing, as well as how the information bits for the eMBB transport block (TB) have been encoded, interleaved, and mapped to the physical layer resources [14]. We assume that an eMBB transmission consists of code blocks (CB). The maximum CB size equals Z=6144 bits, and the number of CBs is denoted by C, as for LTE [15]. For the sake of simplicity, we furthermore assume that the CBs are equal size and fully time-frequency interleaved over the assigned resources for the TB.

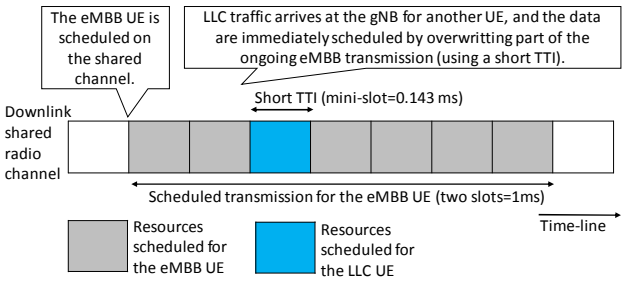

Fig. 1: Basic principle of downlink punctured scheduling.

It should furthermore be noted that the illustration in Fig. 1 is simple in the sense that the eMBB transmission (in this example) experiences one instance of time-domain puncturing only, by one LLC transmission. However, in a loaded multi-user cellular system, an eMBB transmission may in fact experience puncturing by multiple LLC transmissions, and the LLC transmission(s) may have smaller or larger bandwidth than the eMBB transmission. Aspects of how the gNB select which eMBB transmissions to potentially puncture are further addressed in Section IV when outlining the assumed scheduling policy.

B. Recovery mechanisms

As mentioned, the decoding probability (i.e. 1-BLEP) of the punctured eMBB transmission depends on whether the UE is aware of the puncturing. In line with [14], the performance is improved if the eMBB UE is aware of the puncturing. For an initial eMBB transmission we therefore assume that the UE may be informed of the puncturing. The dilemma here is that the gNB does not know when it schedules the eMBB UE if it will be subject to puncturing later. One option is to allow the gNB to append information of the puncturing (if that happens) to the very last part of the eMBB transmission, i.e. embedded in the last part of the data transmission. However, there is of course the risk that if puncturing does happen, it may happen on the last transmission resources of the eMBB transmission, and hence the indication of the puncturing is lost. If the UE fails to correctly decode a punctured eMBB transmission, a HARQ retransmission is triggered. At this point in time, the gNB knows that the previous transmission was punctured, and hence

can inform the UE when scheduling the retransmission by including such information in the downlink scheduling grant. The UE benefits from such information by disregarding the punctured resources of the previous transmission when performing the HARQ soft combining, thereby improving the performance. We assume that HARQ retransmissions consume the same amount of radio resources as the first transmission.

IV. Radio Resource Management Algorithms

A. Scheduling decisions

For scheduling of the eMBB traffic we assume time-frequency domain radio channel aware Proportional Fair (PF) scheduling, based on periodical frequency selective CQI feedback. The PF scheduling metric $M_{u,p}$ is:

$$M_{u,p}[n] = \frac{r_{u,p}[n]}{R_u[n]}, \quad (1)$$

where $r_{u,p}$ is an estimate of the instantaneous supported data rate of user u in the p-th PRB, R_u is its average delivered throughput in the past, and n is the discrete time index for the scheduling interval. eMBB users are scheduled with a TTI size of 1 ms. Pending eMBB HARQ retransmissions are prioritized over new eMBB transmissions as also assumed in [16]. By default, the eMBB traffic is scheduled on all available radio resources, assuming there is enough offered eMBB traffic.

When LLC traffic arrives at the gNB, the scheduler aims at immediately scheduling such traffic with a short TTI size of 0.143 ms (corresponding to 2 OFDM symbols). If there are free (unused) radio resources, the LLC traffic is scheduled on those resources. If not, the LLC traffic is scheduled on radio resources currently allocated to eMBB transmissions, i.e. using punctured scheduling. It should be noted that due to the assumed small payload size for the LLC transmissions, only a fraction of the available PRBs are typically needed for each LLC transmission. The question is now which radio resources currently used for eMBB transmissions are the best to puncture? This is a non-trivial question, to which we propose the following punctured-scheduling metric for the LLC users v:

$$M_{v,p}[n] = \frac{r_{v,p}[n]}{R_v[n]} \cdot W_p^{\alpha}[n], \quad (2)$$

where W_p is the normalized transport block size of the eMBB user per PRB that is currently scheduled on the PRB p; i.e. the basic PF metric is weighted with a function of the MCS employed for eMBB data transmissions on a given PRB. The exponent α controls how much weight is given to W_p. Based on this scheduling framework, we consider the following three options for punctured scheduling:

- **Best Resources (BR):** $\alpha = 0$. In this case, the pending LLC traffic is scheduled on the PRBs where the LLC users experience the best channel quality as per the CQI feedback. Division of resources among competing LLC users is done following the PF rule.
- **Lowest eMBB user (LeU):** $\alpha = -1$. It is prioritized to schedule the LLC traffic on resources that have been allocated to the eMBB user(s) that use the lowest MCS (among the scheduled eMBB users). The rationale here is that eMBB users with low MCS can better tolerate puncturing.
- **Highest eMBB user (HeU):** $\alpha = 1$. It is prioritized to schedule LLC traffic on resources that have been allocated to eMBB users with highest MCS (among the scheduled eMBB users). The rationale here is to protect the cell-edge eMBB users from experiencing puncturing.

The proposed eMBB-aware scheduling (LeU and HeU) for the LLC transmissions tends to favor puncturing the same eMBB transmission(s) in case several LLC transmissions happens during the same 1 ms TTI interval used for scheduling the eMBB users. Our hypothesis is that the eMBB-aware scheduling options therefore are more attractive.

B. Service-specific link adaptation

Dynamic link adaptation (LA) is assumed for both the eMBB and LLC users by setting the MCS for each transmission, based on the users reported CQI. The MCS for the eMBB users is adjusted to reach an average block error rate (BLER) target of 10%. This is achieved by using the well-known outer loop link adaptation (OLLA) algorithm, where the received CQI values are offset by certain factor (a.k.a. the OLLA offset) calculated in accordance to the received HARQ Ack/Nacks from past transmissions [17]. The OLLA-offset for the eMBB users is only adjusted based on Ack/Nack feedback from eMBB transmissions that have not been punctured; i.e. we only aim at controlling the BLER (10% target) for the eMBB transmissions that do not experience any puncturing. The BLER of the punctured eMBB transmissions will naturally be higher, as the error probability increases with the amount of puncturing.

The LA for the LLC transmissions is conducted to have a BLER target of only 1% to have lower latency. The LA for the LLC users is also conduced based on the users CQI, using standard OLLA to reach the 1% BLER target. Single-stream single-user MIMO transmission is assumed, i.e. benefitting from both transmission and reception diversity against fast fading radio channel fluctuations.

V. Performance Analysis

A. Methodology and assumptions

Extensive dynamic system-level simulations are conducted, following the 5G NR methodology in 3GPP [1], [3], assuming a macro-cellular multi-cell scenario. The default simulation assumptions are summarized in Table 1. All the RRM functionalities described in Section IV are modeled. Full buffer traffic is used to model the eMBB best effort traffic. A bursty LLC traffic model is used, with 50-byte packets generated following a Poisson arrival process. Different levels of offered LLC traffic load per cell are considered.

Table 1: Summary of default simulation assumptions.

Description	Assumption
Environment	3GPP Urban Macro (UMa); 3-sector base stations with 500 meters inter-site distance. 21 cells.
Carrier	10 MHz carrier bandwidth at 2 GHz (FDD)
PHY numerology	15 kHz subcarrier spacing configuration [1].
TTI sizes	0.143 ms for LLC (2-symbol mini-slot). 1 ms for eMBB (two slots of 7-symbols).
MIMO	Single-user 2x2 closed loop MIMO and UE MMSE-IRC receiver.
CSI	Periodic CSI every 5 ms, with 2 ms latency, containing CQI, and PMI.
Data channel modulation and coding	QPSK to 64QAM, with same encoding rates as specified for LTE. Turbo codes.
Link adaptation	Dynamic MCS selection. 1% initial BLER target for LLC 10% initial BLER target for eMBB
HARQ	Asynchronous HARQ with Chase Combining soft combining. The HARQ RTT equals minimum 4 TTIs.
Traffic model	In average 5 full buffer eMBB users per cell. In average 10 LLC users per cell with Poisson arrival of B=50 bytes data bursts.
Scheduling	Proportional fair scheduling of eMBB. Punctured scheduling for LLC traffic following BR, LeU, and HeU.
Link-to-system (L2S) mapping	Based on the mean mutual information per coded bit (MMIB) mapping methodology.

Whenever a user is scheduled, the SINR at the receiver is calculated for each subcarrier symbol, assuming a minimum mean square error with interference rejection combining (MMSE-IRC) receiver at the terminal. Inspired by the model in [18]-[19], the SINR values are mapped to the mutual information domain, taking the applied modulation scheme into account. The mean mutual information per coded bit (MMIB) is calculated as the arithmetic mean of the values for the sub-carrier symbols of the transmission [19]. Given the MMIB and the used modulation and coding rate of the transmission, the error probability of a CB is determined from look-up tables that are obtained from extensive link level simulations. For transmissions consisting of more than one CB, we assume identical and independent error performance for all CBs. Thus, the error probability for the transport block is modeled as $P(\varepsilon_{TB}) = 1 - (1-P(\varepsilon_{CB}))^{C}$, where $P(\varepsilon_{CB})$ is the CB BLEP.

The effect of an eMBB transmission that is punctured is captured as follows: The punctured sub-carrier symbols contain no useful information for the receiver, and hence is modelled as information-less. This effect is included in the calculation of the MMIB and the effective coding rate of the transmission prior to using the look-up tables described above to determine the CB BLEP. In other words, a receiver that is aware of the puncturing incident is assumed to be aware of the exact subcarrier symbols that are punctured. Therefore, the receiver can discard the punctured parts of the physical resources prior to the decoding. Hence, the MMIB for such users is calculated only as the mean from transmission resources that were not punctured and the effective coding rate of the transmission is increased accordingly.

On the other hand, if the UE is unaware of the puncturing the punctured part of the transmission will still be taken as useful signal by the UE thus, used in the decoding process. Therefore, in such scenarios we model the punctured resources as interference only, which decreases the overall MMIB, while keeping the effective coding rate unaffected. The setting in all simulations, except where explicitly mentioned, is that the eMBB UEs are fully aware of the puncturing when it happens.

B. Performance results

Fig. 2 shows the cumulative distribution function (cdf) for the ratio of punctured eMBB resources per user allocation for different offered load conditions of LLC traffic. The ratio of punctured eMBB resources per user allocation is defined as the sum of the sub-carrier symbols allocated to LLC within a given eMBB transmission, divided by the total amount of sub-carrier symbols in the eMBB allocation. As expected, the higher the LLC load, the more eMBB allocations are punctured. At 0.1 Mbps LLC load, only around 10% of the eMBB allocations are punctured, whereas more than 70% puncturing ratio can be observed for a high LLC load of 2 Mbps. The scheduling scheme also impacts the distribution. The LeU and HeU schemes tends to favor puncturing the same eMBB transmission multiple times, in case several LLC transmissions happen during the same 1 ms TTI interval. This results in fewer eMBB allocations being punctured as compared to BR, at the expense of higher puncturing ratio for those transmissions. It is observed that with relatively high probability, the puncturing ratio is either 1/7 (~0.14) or 2/7 (~0.28), which corresponds to the case when LLC allocations puncture the entire frequency sub-band of a certain eMBB allocation with one or two short TTI transmissions.

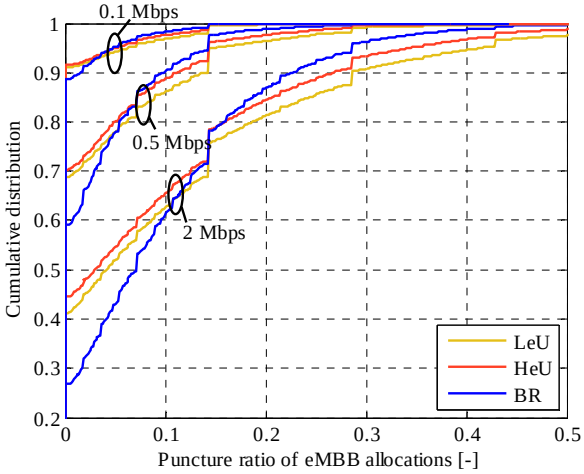

Fig. 2: Cdf of the ratio of punctured resources per eMBB user allocation.

Fig. 3 pictures the average decoding probability of eMBB transmissions for different puncturing ratios, including the case where the eMBB UEs are not made aware of the puncturing. eMBB transmissions without any puncturing achieve the 90% decoding probability (or 10% BLER) in line with the service-specific LA setting described in Section IV-B. For cases where an eMBB allocation is punctured on 1/7 of the resources, the decoding probability drastically decreases. It is observed that LeU scheme achieves the best decoding performance, as users

with low MCS (typically low coding rate) can better tolerate puncturing. The BR scheme tends to equally affect cell-edge and cell-center eMBB UEs. This results in a decoding probability which is in between what is observed for the other two scheduling schemes. As expected, there is some gain from making the eMBB UE aware of the puncturing (indicated by the dashed line on Fig. 3); this gain is, however, generally lower than what is reported in [14], although we still observe a clear benefit of making eMBB UEs aware of the puncturing. The reason for observing differences in performance gain of having such puncturing awareness at the eMBB UEs, is expected to be due to the abstract L2S model applied in our system-level study, while findings in [14] are based on more detailed link level simulations. For eMBB users that experience extensive puncturing of 3/7, there is no visible gain by making the eMBB aware of the puncturing. This is because such a large fraction of "lost" resources can anyway not be compensated at the receiver end, and hence is likely to result in failed decoding independent of whether the UE is made aware of it, or not.

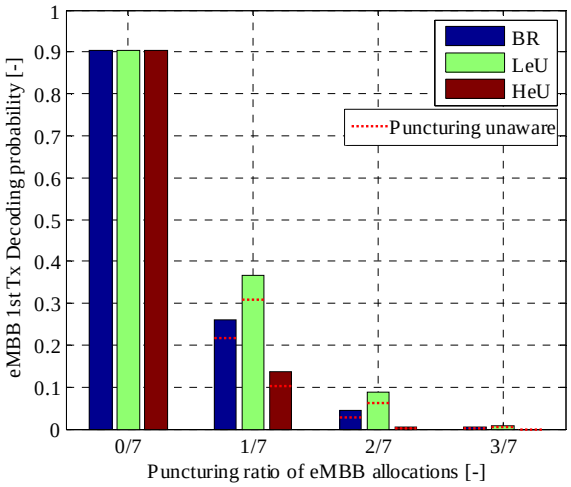

Fig. 3: Average decoding probability of eMBB transmissions with different puncturing ratio.

The impact on the eMBB performance from the puncturing is shown in Fig. 4, where a cdf of the eMBB throughput is plotted. It is observed that the eMBB throughput generally declines as LLC traffic is increased, due to more puncturing. The LeU scheme generally offers the best throughput performance. This is due to the larger robustness against puncturing, as also observed in Fig. 3. The difference in performance between the BR and HeU schemes depend on the offered LLC load. For a LLC offered load of 0.5 Mbps, the performance is as expected: the BR scheme performs better than the HeU, especially in the upper part of the distribution, as HeU favors puncturing of users with high MCS. However, for a higher LLC offered load of 2 Mbps, HeU performs slightly better than BR. This is due to the benefits of concentrating the LLC puncturing in only a few eMBB allocations. Such gain is especially relevant at high LLC load, when the eMBB allocations are more likely to experience puncturing from multiple LLC users.

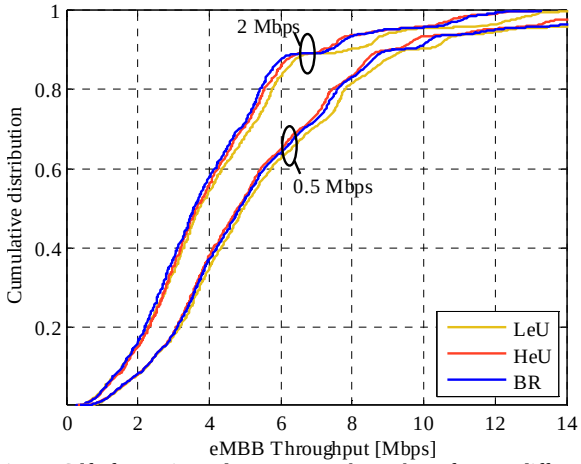

Fig. 4: Cdf of experienced eMBB user throughput for two different offered loads of LLC traffic.

Fig. 5 shows the complementary cdf (ccdf) of latency statistics for the LLC traffic for different load conditions. The service-specific 1% BLER target for LLC transmissions is clearly observed in form of a HARQ delay. Looking at the achievable LLC latency at the 10^{-5} percentile, it is observed that the 1 ms latency requirement for 5G is achieved for both 0.1 Mbps and 2 Mbps load of LLC traffic. Furthermore, no significant difference between the three proposed schedulers is observed at the 10^{-5} level. One of the reasons is that the proposed puncturing scheduling schemes partly accounts for the experienced channel quality of LLC users. Also, sufficiently good channel quality is experienced across the whole frequency band due to the high diversity from using 2x2 closed-loop single-stream MIMO with MMSE-IRC receiver at the UE.

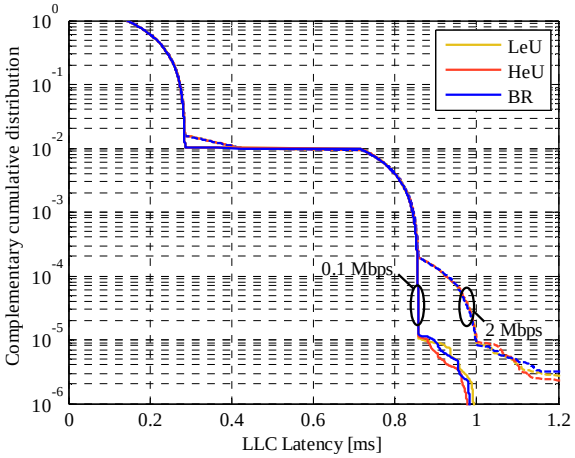

Fig. 5: Latency distribution (ccdf) of the LLC traffic.

Finally, Fig. 6 shows the 50%-ile eMBB throughput (left axis) and the 99.999%-ile latency for the LLC traffic (right axis), for different offered loads of LLC traffic. Only the LLC latency with the BR scheme is shown as there is marginal difference in performance (as seen in Fig. 5). The eMBB throughput follows the trends previously described: The LeU scheme offers the best throughput performance due to larger

robustness to puncturing. At low LLC offered load, the BR scheme performs better than the HeU, as HeU favors puncturing of users with high MCS (sensitive to puncturing). However, at high LLC offered load, HeU performs slightly better than BR. This is due to the benefits of concentrating the LLC puncturing in only a few eMBB allocations - especially relevant at high LLC load, when the eMBB allocations are more likely to experience puncturing from multiple LLC users.

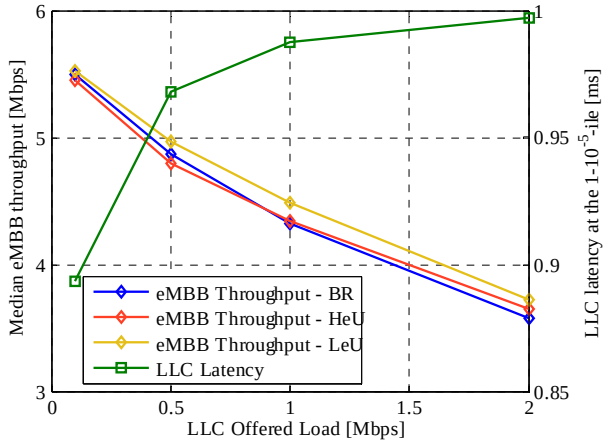

Fig. 6: 50%-ile MBB throughput (left axis), and 99.999%-ile latency for LLC traffic (right axis).

VI. Concluding Remarks

In this paper we have presented a punctured scheduling solution, tailored to efficient transmission of urgent LLC traffic on a shared channel with eMBB transmissions. The scheme does not require any pre-reservation of radio resources for transmission of the randomly arriving LLC payloads. Mechanisms to have such solutions perform efficiently are proposed. Those include recovery mechanisms for the eMBB transmissions that experience puncturing, service-specific and puncturing-aware dynamic link adaptation, as well as eMBB-aware scheduling decisions for LLC traffic to minimize the capacity loss for eMBB due to LLC traffic. The presented system-level performance results document the benefits of such solutions, confirming our hypothesis that punctured scheduling (sometimes referred to as preemptive scheduling) is attractive and worth pursuing in the design of a 5G multi-service systems.

However, despite of those findings, there are still more options and enhancements for punctured scheduling that are worth studying. Among others, we are currently studying the case where so-called variable block-length HARQ retransmissions are applied, only retransmitting the damaged part of the punctured eMBB transmissions.

Acknowledgements

Part of this work has been performed in the framework of the Horizon 2020 project ONE5G (ICT-760809) receiving funds from the European Union. The authors would like to acknowledge the contributions of their colleagues in the project, although the views expressed in this contribution are those of the authors and do not necessarily represent the project.

References

[1] 3GPP Technical Report 38.802, "Study on New Radio Access Technology Physical Layer Aspects", March 2017.

[2] 3GPP Technical Report 38.801, "Study on New Radio Access Technology: Radio Access Architecture and Interfaces", March 2017.

[3] 3GPP Technical Report 38.913, "Study on Scenarios and Requirements for Next Generation Access Technologies", March 2016.

[4] 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.

[5] E. Dahlman, et.al., "5G Wireless Access: Requirements and Realization", IEEE Communications Magazine - Communications Standards Supplement, December 2014.

[6] B. Soret, et.al., "Fundamental Tradeoffs among Reliability, Latency and Throughput in Cellular Networks", IEEE Proc. GLOBECOM, December 2014.

[7] K.I. Pedersen, et.al., "A Flexible 5G Frame Structure Design for Frequency-Division Duplex Cases", in IEEE Communications Magazine, pp. 53-59, March 2016.

[8] Q. Liao, P. Baracca, D. Lopez-Perez, L.G. Giordano, "Resource Scheduling for Mixed Traffic Types with Scalable TTI in Dynamic TDD Systems", in IEEE Proc. Globecom, December 2016.

[9] K.I. Pedersen, M. Niparko, J. Steiner, J. Oszmianski, L. Mudolo, S.R. Khosravirad, "System Level Analysis of Dynamic User-Centric Scheduling for a Flexible 5G Design", in IEEE Proc. Globecom, December 2016.

[10] G. Pocovi, B. Soret, K.I. Pedersen, P.E. Mogensen, "MAC Layer Enhancements for Ultra-Reliable Low-Latency Communications in Cellular Networks", in IEEE Proc ICC (workshop), June 2017.

[11] G. Pocovi, K.I. Pedersen, B. Soret, M. Lauridsen, P.E. Mogensen, "On the Impact of Multi-User Traffic Dynamics on Low Latency Communications", in Proc. International Symposium on Wireless Communication Systems (ISWCS), September 2016.

[12] A.S. Tanenbaum, "Computer networks", fifth edition, Prentice Hall, 2011.

[13] G.C. Buttazzo, M. Bertogna, G. Yao, "Limited Preemptive Scheduling for Real-Time Systems: A Survey", in IEEE Trans. on Industrial Informatics, vol. 9, no. 1, pp. 3-15, Feb. 2013.

[14] Technical contribution to 3GPP, Document R1-1700374, "Downlink Multiplexing of eMBB and URLLC Transmissions", Intel Corporation, January 2017.

[15] 3GPP TS 36.212, "Evolved Universal Terrestrial Radio Access (E-UTRA); Multiplexing and channel coding", January 2017.

[16] H. Holma and A. Toskala (Editors), "LTE-Advanced: 3GPP solution for IMT-Advanced", John Wiley & Sons, 2012.

[17] A. Pokhariyal, et.al., "HARQ Aware Frequency Domain Packet Scheduling with Different Degrees of Fairness for UTRAN Long Term Evolution", in IEEE Vehicular Technology Conference (VTC-Spring), May 2007.

[18] K. Brueninghaus, et.al, "Link performance models for system level simulations of broadband radio access systems," in IEEE Proc. Personal, Indoor and Mobile Radio Communications (PIMRC), pp. 2306-2311, Sept. 2005.

[19] R. Srinivasan, J. Zhuang, L. Jalloul, R. Novak, and J. Park, "IEEE 802.16m Evaluation Methodology Document (EMD)", in IEEE 802.16 Broadband Wireless Access Working Group, Tech. Rep. IEEE 802.16m-08/004r2, http://ieee802.org/16/tgm/docs/80216m-08_004r2.pdf, July 2008.