System Level Analysis of Uplink Grant-Free Transmission for URLLC

Jacobsen, Thomas; Abreu, Renato Barbosa; Berardinelli, Gilberto; Pedersen, Klaus I.; Mogensen, Preben Elgaard; Kovacs, Istvan; Madsen, Tatiana Kozlova

Published in:
IEEE 2017 Globecom Workshops

DOI (link to publication from Publisher):
10.1109/GLOCOMW.2017.8269137

Publication date:
2017

Document Version
Accepted author manuscript, peer reviewed version

System Level Analysis of Uplink Grant-Free Transmission for URLLC

Thomas Jacobsen*, Renato Abreu*, Gilberto Berardinelli*, Klaus Pedersen*†, Preben Mogensen*†, István Z. Kovács*†, Tatiana K. Madsen*
*Dept. of Electronic Systems, Aalborg University; †Nokia Bell Labs, Aalborg, Denmark
Email: {tj,rba}@es.aau.dk

***Abstract*—In the context of 5'th Generation (5G) New Radio (NR), new transmission procedures are currently studied for supporting the challenging requirements of Ultra-Reliable Low-Latency Communication (URLLC) use cases. In particular, grant free (GF) transmissions have the potential of reducing the latency with respect to traditional grant-based (GB) approaches as adopted in Long Term Evolution (LTE) radio standard. However, in case a shared channel is assigned to multiple users for GF transmissions, the occurrence of collisions may jeopardize the GF potential. In this paper, we perform a system analysis in a large urban macro network of several transmission procedures for uplink GF transmission presented in recent literature. Specifically, we study K-Repetitions and Proactive schemes along with the conventional HARQ scheme referred to as Reactive. We evaluated their performance against the baseline GB transmission as a function of the load using extensive and detailed system level simulations. Our findings show that GF procedures are capable of providing significant lower latency than GB at the reliability level of $1 - 10^{-5}$, even at considerable network loads. In particular, the GF Reactive scheme is shown to achieve the latency target while supporting at least 400 packets per second per cell.**

I. Introduction

Ultra-Reliable Low-Latency Communication (URLLC) represents the most challenging set of services/use cases [1] for upcoming 5th Generation (5G) New Radio (NR), with ambitious latency and reliability targets ($1\,\mathrm{ms}$ with $1 - 10^{-5}$ reliability) for small packet transmissions [2]. A number of technology components including spatial diversity [3], frame structure [4], [5], resource allocation [6] including link adaptation and transmission schemes, all need to be redesigned when dealing with requirements that are beyond current Long Term Evolution (LTE) capabilities [7].

In particular, the transmission procedures, including Hybrid Automatic Repeat Request (HARQ) retransmissions, play a major role in achieving the URLLC requirements [8]. LTE utilizes dynamic scheduling as a basic transmission mode, which is referred to as Grant Based (GB) scheduling (specified in [9]). A traditional GB transmission requires the User Equipment (UE) to be scheduled by the base station (BS). The scheduling procedure is initiated by the UE with a scheduling request which the BS can respond by issuing a scheduling grant.

Grant-Free (GF) transmission schemes are also well known solutions that are meant for fast uplink access, by removing the phases of scheduling request and grant issuing [10]. With Semi-Persistent-Scheduling (SPS), the BS can configure the UE to have pre-allocated periodic radio resources available for transmissions [11], [12]. For periodic traffic, SPS is expected to be a valid solution to meet the URLLC requirements. However, in case of aperiodic (sporadic) traffic, pre-allocating dedicated resources may lead to a large waste and will scale poorly with the number of URLLC users. A possible solution to overcome this limitation, is to pre-schedule shared resources for contention-based transmissions [4].

Conventional HARQ operations in LTE allows for retransmissions only upon reception of a negative acknowledgement. This requires the BS to have first received the payload, processed it and issued the feedback. Such HARQ scheme is often referred to as *Reactive* since retransmissions are triggered based on the knowledge about the previous transmission.

However, the reactive HARQ scheme can only support a limited number of retransmissions before the URLLC requirements is no longer met. Therefore different HARQ strategies to further reduce latency and improve reliability have been recently studied. One technique that has been considered for 5G, is to run a number of blind transmissions of the same payload. The BS can then perform soft combining of the transmissions to improve the decoding reliability [13]. Such kind of solution is already part of the recent 3GPP agreements for NR and are referred to as *K-Repetitions* (K-Rep) [14].

In a proactive version of the HARQ scheme mentioned above, the UE can still transmit in consecutive frames (like K-Rep), but it will stop when it has received and decoded a positive feedback from the BS. Such scheme is known as repetition scheme with early termination, and is mentioned in [15] and [16]. This scheme is more computational heavy for the UE, which needs to monitor the feedback. However, it is also likely to be more resource efficient than K-Rep if the number of blind repetitions is overestimated and more reliable if the number is underestimated.

The theoretical foundation of the transmission procedures mentioned above is already well established. However, to the best of our knowledge their suitability for URLLC has been so far evaluated in simplified scenarios, such as single cell (and therefore no inter-cell interference impact), basic abstraction models for contention-based transmissions and throughput mapping. In this paper, we perform a detailed system level evaluation of the identified transmission procedures in an outdoor 3GPP urban macro setup with 21 cells, including realistic traffic and radio propagation models, receiver types and open

loop power control. GB with conventional HARQ scheme is used as performance baseline. The transmission schemes are then evaluated in terms of the latency and reliability and as a function of the load imposed by URLLC devices in the network. Our aim is to assess the effective system benefits of the identified techniques and their potential in a network of URLLC devices.

The paper is structured as follows. The considered URLLC UL transmission schemes are described in section II. The simulation assumptions are outlined in section III, while the results are presented in section IV. The work is discussed in section V and concluded in section VI.

II. URLLC UL Transmission Schemes

This section provides a general description of the transmission schemes considered in this paper. A frame-based system alike LTE is assumed, meaning that transmissions can start on a frame basis. The transmissions occur when the UE is already synchronized and in connected state. We consider both GB and GF solutions.

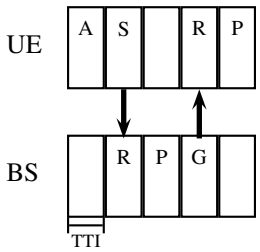


Fig. 1. Scheduling request model used for Grant-Base access. Legend: A = Frame alignment, S = Scheduling Request, R = Reception, P = Processing, G = Scheduling Grant.

The GB approach is the common method to perform an UL transmission in cellular networks, and is evaluated with the usual LTE scheduling grant procedure as illustrated in illustrated in Fig. 1 and with the conventional HARQ scheme (reactive Fig. 2(a)).

When using the GB approach, each UL transmission is coordinated by the base-station (BS). Upon a packet arrival on layer 3 (L3), a UE waits for the next subframe occurrence for transmitting a scheduling request (SR) signal (S). After processing the SR, the BS transmits a scheduling grant (G) which indicates the time-frequency resources among other settings that the UE should use for its uplink data transmission (T). Only after receiving (R) and processing (P) the scheduling grant, the UE can perform the data transmission. This procedure allows the BS to assign resources in a very flexible manner, leading to a high spectral efficiency. Further, the transmissions are collision-free.

The scheduling process comes with a number of drawbacks; it is time consuming, which makes it harder to make the URLLC requirements, it introduces a large signalling overhead for small packets which might be a limiting factor for scalability and the signalling is error prone. The cost is that the transmissions becomes prone to collisions and intra-cell interference.

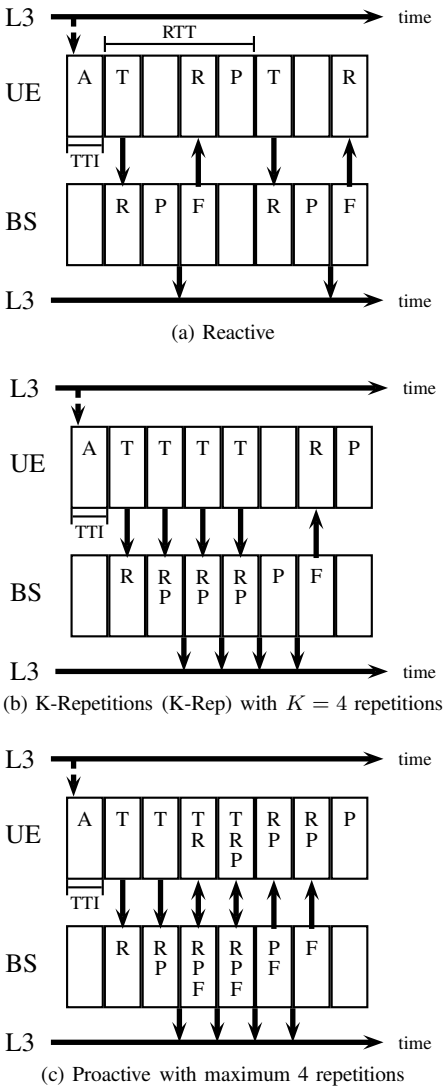


Fig. 2. The considered Uplink HARQ Schemes for URLLC. Shown for Grant-Free transmissions. Legend: A = Frame alignment, T = transmission, R = Reception, P = Processing, F = Feedback.

Three HARQ schemes are considered for GF transmissions, namely a Reactive, K-Rep and Proactive scheme. The Reactive scheme is illustrated in Fig. 2(a). When the UE has finalized its initial uplink data transmissions, its signal is processed at the BS, which will transmit a positive or negative acknowledgement. Upon processing the feedback, the UE can transmit a new payload or retransmit the same payload again. The time duration of the cycle from the beginning of a transmission until the processing of its feedback is called the HARQ Round-Trip-Time (RTT). In the illustration it is assumed that the BS spends 1 transmission time interval (TTI) for processing and 1 TTI for transmitting the feedback. These assumptions are similar to the ones used by the authors in [8].

The K-Rep scheme is illustrated in Fig. 2(b). The UE is configured to autonomously transmit the same packet K times before waiting for feedback from the BS. Each repetition can be identical, or be a different redundancy versions of the encoded data. This method can reduce the delay in the HARQ

process, with a potential waste of resources if the number of repetitions is overestimated.

The last HARQ scheme considered for GF transmissions is the Proactive scheme which is illustrated in Fig. 2(c). Similarly to the K-Rep scheme, the UE aims at repeating the initial transmission for a number of times, however, it will receive a feedback at every repetition. This allows the UE to stop the chain of repetitions earlier in case of a positive feedback. A reduction of the overall transmission resources can be obtained compared to the K-Rep scheme in case the time spent for the K'th transmission is higher than the HARQ RTT. Further it might enhance the reliability compared to the K-Rep, in case K is underestimated.

Note that both GB and GF transmissions can be subject to queuing delays. This occurs due to the limit that a UE can only transmit one packet per TTI or if the UE runs out of Stop-And-Wait (SAW) channels. A SAW channel is occupied throughout the entire transmission, meaning from the initial transmission until the stopping criteria determined by the HARQ RTT from the last transmission.

III. Simulation assumptions

The simulation assumptions and parameters used for this study are in line with the guidelines for NR performance evaluations presented in [17] and are summarized in Table I.

TABLE I
SIMULATION ASSUMPTIONS

Parameter	Value
Network layout	3GPP Urban Macro (UMa) [17] with 21 cells, $500\,\mathrm{m}$ inter-site distance
UE deployment	Uniformly distributed outdoor, speed of $3\,\mathrm{km\,h^{-1}}$, without handover
Carrier and Bandwidth	$10\,\mathrm{MHz}$ at $4\,\mathrm{GHz}$
PHY numerology	2 OFDM symbols per TTI, subcarrier spacing of $15\,\mathrm{kHz}$, 12 subcarriers/PRB
Uplink receiver	MMSE-IRC
Uplink antenna	1x2 antenna configuration
Channel model	3D UMa propagation model, noise density of $-174\,\mathrm{dBm\,Hz^{-1}}$
HARQ configuration	4 TTI RTT and 1 TTI processing (for both UE and BS), 4 SAW channels
Frame alignment model	Uniform random variable up to 1 TTI
Traffic model	FTPModel3 with $32\,\mathrm{B}$ packet size and Poisson arrival of 10 packets per second (PPS) per UE
Link-Adaptation	Conservative modulation and coding scheme fixed to QPSK 1/8
Power control	Open Loop Power Control (OLPC) with $\alpha = 0.8$ and $P_0 = -85\,\mathrm{dBm}$
SR configuration	SR periodicity of 1 TTI
Shared channel configuration	48 RB contention based channel, all UEs can transmit in any TTI

The system level simulation of the multi-cell synchronous network includes inter-cell interference, realistic propagation models, link-to-system mapping and modeling of major radio resource management (RRM) functionalities in accordance with the evaluation methodology of recent 3GPP standardization agreements.

In this work we compare the GF schemes with a baseline GB scheme. As in [8], we assume here 1 TTI for transmitter and receiver processing time. It is worth mentioning that a higher processing time directly translates to a higher delay on the scheduling procedure and HARQ schemes. To ensure a fair comparison between GF and GB schemes we use the same amount of resources for the uplink shared channel used by GF and GB. Uplink and downlink is separated in frequency (FDD), where the uplink shared channel has 48 resource blocks (RBs) in the $10\,\mathrm{MHz}$ bandwidth. The shared channel is assumed to be available in all subframes for GF transmission. For the GB procedure, the configured SR periodicity of 1 TTI permits the UE to ask to be scheduled at every TTI. No additional control overhead is assumed. In this work, we assume the control signalling to be error free, meaning that particular the GB results can be optimistic.

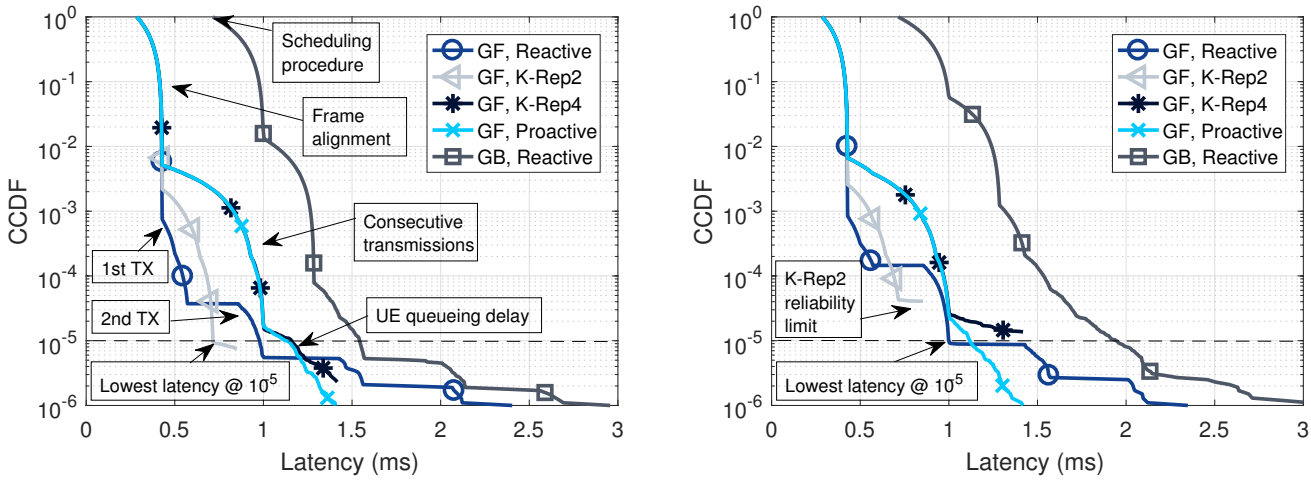
The scenario used in our study is slightly deviating from the one specified in [17], since here all UEs are deployed outdoor. Indoor users showed an tendency to get power limited and were hence unable reach URLLC reliabilities.

Open loop power control is used in this study by the UE to compensate the coupling loss and is configured with $\alpha = 0.8$ and $P_0 = -85\,\mathrm{dBm}$. In the considered deployment this configuration permits the UEs to operate mostly below the maximum transmit power ($23\,\mathrm{dBm}$).

It is assumed that the URLLC UEs are pre-configured with 48 RB for contention based uplink transmissions. The modulation and coding scheme (MCS) is also pre-configured as very conservative (QPSK with coding rate 1/8), which permits the UE to transmit the $32\,\mathrm{B}$ packet (in accordance with baseline in [2]) in 1 TTI using the full band.

The adopted Minimum Mean Square Error Interference Rejection Combining (MMSE-IRC) receiver is assumed to be able to ideally estimate the interference covariance matrix for suppressing intra-cell and inter-cell interference. Given the 2 receive antennas, up to one interfering stream can be suppressed. This also means the decoding of two simultaneously transmitting UEs in the same cell is still possible and depends on the post-detection Signal-to-Noise Plus Interference Ratio (SINR) and the selected MCS.

We focus on the user plane latency and reliability for small packet transmissions assuming the UE is in connected mode. The latency is measured as a one-way latency from when the packet leaves the L3 buffer at the UE until it enters L3 layer at the BS. Throughout the study it has been observed that the packet generation rate per UE impacts the queuing delay and hence forces an upper bound of the load. In order to circumvent this limitation, a variable cell load is simulated by varying the number of UEs per cell, while their packet generation rate is maintained constant. However this comes at the penalty of increased computational complexity of the simulation when more UEs are added. In order to have an acceptable simulation time for different number of UEs, we chose a mean packet generation rate of 10 packets per second giving a theoretical lower bound probability (depending on the HARQ scheme) of a packet being queued at $\approx 10^{-6}$.



(a) Low load (10 UE / cell)

(b) High load (40 UE / cell)

Fig. 3. CCDF of the latency for GF and GB baseline for low (a) and high (b) load.

IV. Results

The evaluation of the UL transmission schemes is carried out with Monte Carlo simulations. More than 5×10^6 samples per simulations are acquired to ensure sufficient statistical confidence in the 10^{-5} quantile [8]. The transmission schemes are evaluated at different loads, determined by URLLC densities. Results are presented in terms of one-way latency for a packet transmission, as well as number of transmissions per packet. Unsuccessful packets are represented as void samples and are used to reflect the achievable reliability.

In Fig. 3(a) the emperical Complementary Cumulative Distribution Function (CCDF) of the one-way latency for the different GF HARQ transmission schemes is shown along with the GB baseline with low load (10 UEs / cell). On the horizontal axes the latency is shown in ms and on the vertical axes the outage probability quantiles are shown. The GF schemes clearly provide a better latency for the same reliability compared to the GB reference. One of the main differences between these are the unavoidable delay offsets from the scheduling procedure. The first slope from $\approx 0.3\,\mathrm{ms}$ to $\approx 0.4\,\mathrm{ms}$ corresponds to the uniformly distributed frame alignment delay.

The Reactive HARQ scheme is the one providing the best reliability for the the first transmission. The stair behaviour is caused by the HARQ RTT. K-Rep scheme with 2 repetitions follows the initial transmission with a similar slope for the second consecutive transmission, and is capable of providing $l1\,\mathrm{ms}$ latency with the target $1 - 10^{-5}$ reliability. The curve has a tail caused by low probability events corresponding the probability of packet buffering at the UE.

The K-Rep scheme with 4 repetitions and Proactive scheme have a similar latency and reliability performance until $1\,\mathrm{ms}$. This can be explained from the fact that the Proactive scheme earliest determination time depends on the HARQ RTT which here it is assumed to be 4 TTIs. Since more than 4 repetitions is rarely needed in this scenario, K-Rep4 and Proactive perform almost identically. The schemes shows different tail tendencies, where the Proactive scheme is better on handling the low probability events where more than $K = 4$ repetitions is needed.

Comparing the HARQ Reactive transmission scheme for GF and GB transmission, they show a similar stair behaviour, with the initial step occurring at different latency and reliability combinations (e.g. $0.6\,\mathrm{ms}$ and $1.6\,\mathrm{ms}$ for GF and GB respectively). The reason for the reliability difference for the initial transmission is due to the impact of intra-cell interference. Further the GB curve shows tendencies for higher packet queuing probability due to the longer pre-transmission time caused by the scheduling procedure.

Performance at a higher load (40 UE / cell) is shown in Fig. 3(b). The impact of a higher load is clearly visible for the Reactive HARQ schemes. The CCDF of the Reactive HARQ scheme shows an increase in the probability of needing multiple retransmissions and causing its tail to be longer compared with the low load. The GF K-Rep schemes reach a reliability floor around $\approx 1 - 4 \times 10^{-5}$ instead of $\approx 1 - 10^{-5}$. With this load, only the Proactive and Reactive HARQ schemes for GF transmissions are able to achieve the $1 - 10^{-5}$ reliability and only the Reactive HARQ scheme is capable of doing within the $1\,\mathrm{ms}$ latency target.

Figure 4 illustrates the impact of the load on the achievable latency with $1 - 10^{-5}$ reliability. At low load, the Reactive scheme and the K-Rep scheme with 2 repetitions meet the URLLC performance target, where the latter has the lowest latency. For more than 40 UEs / cell no GF or GB scheme is capable of achieving the URLLC target. However, at high load the GF Proactive scheme leads to the lowest latency.

Figure 5 shows the empirical Cumulative Distribution Function (CDF) of the average SINR per RB for the case of 40 UE / cell. Here it is possible to see that the GB transmissions

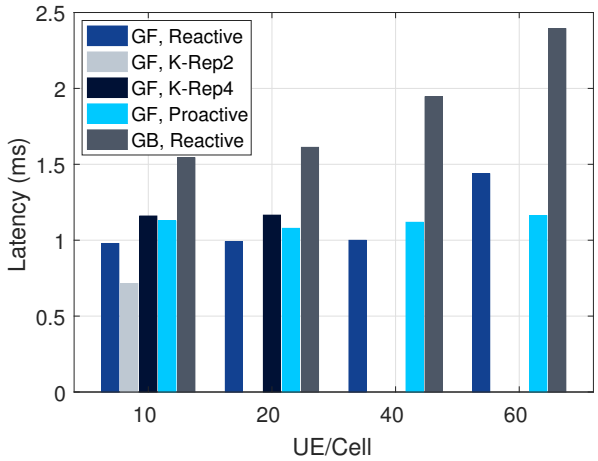


Fig. 4. Achieved latency at $1 - 10^{-5}$ reliability as a function of load.

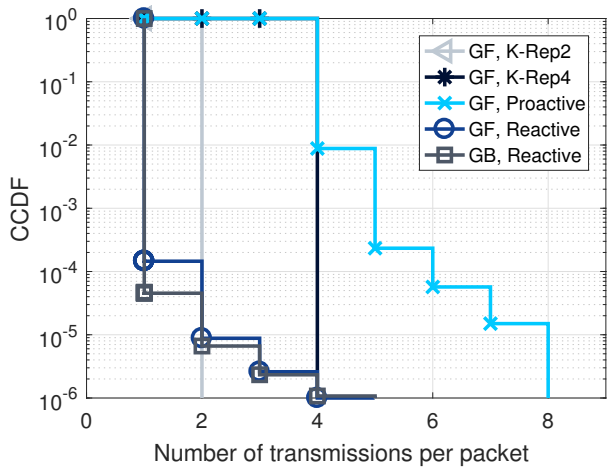


Fig. 6. CCDF of the number of transmissions per packet (40 UE / cell).

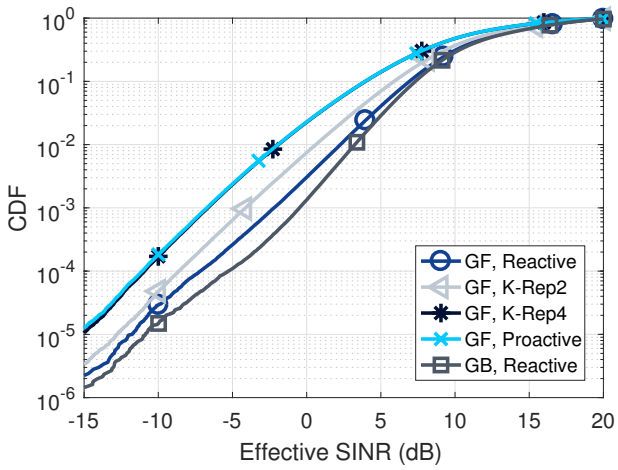


Fig. 5. Average effective SINR per RB for GF and GB (40 UE / cell).

presents the best SINR condition since intra-cell interference is avoided in this procedure. GF with the K-Repetitions and Proactive scheme on the other hand presents the worst SINR due to the extra intra-cell interference caused by the blind repetitions. The GF Reactive scheme presents a better SINR then the other GF schemes given that it avoids unnecessary retransmissions. This explains why each transmission of the Reactive scheme presents a higher reliability, compared to the cases with blind repetitions. In this case, for GF Reactive, a $1 - 10^{-5}$ reliability can be achieved with 2 transmission attempts. While, for instance, in the Proactive or K-Rep after 4 attempts the achieved reliability is even lower.

As showed in [7], achieving low latency and high reliability has a cost in terms of resource utilization and therefore spectral efficiency. Figure 6 shows the empirical CCDF of the number of transmissions used for successfully delivering a packet for the different schemes, assuming a load of 40 UEs / cell. The GB scheme presents, not surprisingly, the lower probability of requiring multiple channel accesses for transmitting a packet. The curve for the GF Reactive scheme is slightly higher compared to the GB Reactive. This is likely due to the presence of collisions. The K-Rep schemes are very deterministic in terms of channel usage, while GF Proactive occupies the channel at least during the RTT. The two former schemes, besides not meeting the baseline requirement, also presents the lowest spectral efficiency at this scenario and with this load.

V. Discussion

The evaluated GF solutions clearly show better latency performance than GB transmission at $1 - 10^{-5}$ reliability, despite the impact of collisions. Our results also showed that GF schemes are not outperformed by GB even in the case of 40 devices per cell. This section discussions the dominating factors impacting our results.

GB avoids intra-cell interference by ensuring a single transmit UE per TTI, but also causes a latency increase by waiting for the channel to become available. The GF schemes have no such limitation, but are instead affected by the intra-cell interference from competing UEs. Therefore GB has the potential to achieve the $1 - 10^{-5}$ reliability when the latency requirement is relaxed, to e.g. $2\,\mathrm{ms}$ for the referred loads, causing a lower interference in the network.

The reasoning behind the usage of GF K-Rep schemes, is to cope with tight time constraints by allowing a number of consecutive transmissions in a short time. Our findings show, however, that the additional intra-cell interference due to the multiple transmissions is the major impacting factor and surpasses the benefits of the combining gain. One way to lower the average intra-cell interference with K-Rep schemes is to use a faster reconfiguration cycle that sets higher number of repetitions only for the UEs in worse channel condition, though requiring additional RRC signalling.

In the studied scenario with GF, the use of a robust MCS (QPSK 1/8) ensures a high decoding probability even under a potentially high intra-cell interference. Another aspect is the

benefit of HARQ which adds combining gain and diversity, given also that a packet has lower probability of colliding.

As mentioned in Section III, results are obtained with a MMSE-IRC receiver with 2 antennas, which is able to resolve two simultaneous transmissions from two different UEs. It is left for future analysis to investigate the impact of other receiver types and antenna configurations, whose capabilities of resolving the interference may affect the trade-off between GB and GF transmissions. The use of a successive Interference Cancellation (SIC) receiver is also considered.

With GF transmissions the BS has to conduct blind decoding as every connected UE has the possibility to transmit in every TTI. The BS should be able to identify a UE before attempting to decode it. This assumes a system design where the UE identity is mapped over e.g. preambles and header at each transmission [18]. The impact on the preamble and header design on the GF performance is left for future work.

Moreover, in this work the control channel is assumed to be ideal and not introducing any overhead. While the control signalling is typically designed to be very robust, the potential errors may not be negligible for the range of reliability expected for URLLC. Errors in control signalling can significant impact the schemes relying on feedback, such as the Proactive and particular the Reactive schemes, as well as the scheduling procedure for GB. These are also the scheme relying on the most DL resources due to the signalling. The impact of error-prone control signalling is left for further analysis.

The GF analysis can also be extended with the adoption of other enhancements, as a Non-Orthogonal Coded Access scheme like proposed in [19], that increases the capacity and reduce collisions with additional spreading codes.

VI. Conclusion

In this paper, we studied the performance of uplink GF schemes in a large outdoor urban macro scenario and compared its performance with a traditional GB scheme. In particular, the schemes referred to as GF Reactive, K-Rep and Proactive, are evaluated. The results are obtained using extensive system level simulations to include the complexity of the receiver, inter-cell interference, power control and HARQ operations including soft combining. The main findings of this work together with the recommendations for a 5G NR design are:

- GF in general outperforms GB transmission procedures in terms of latency at the target reliability ($1 - 10^{-5}$). This makes them valuable candidates for achieving the baseline URLLC requirements in an outdoor scenario.
- The GF Reactive scheme is strongly recommended as it is capable of supporting the largest load among the GF schemes. The maximum achieved load is found to be 400 packets per second per cell (40 UEs per cell generating 10 packets per second on average). This scheme is also the most uplink resource efficient next to the GB baseline.
- The GF Proactive scheme gives the smallest latency performance degradation for loads higher than 400 packets per second.
- GB transmissions can achieve the target reliability if the latency requirements is relaxed to e.g. $2\,\mathrm{ms}$.

The presented results are obtained by relying on a robust MCS (QPSK 1/8) for packet transmission, interference suppression by IRC receiver and HARQ combining gain from repetitions and retransmissions. Future work will investigate the impact on the GF performance of factors such as dynamic link adaptation, power boosting, multiple receiver types and antenna configurations.

Acknowledgment

This research is partially supported by the EU H2020-ICT-2016-2 project ONE5G. The views expressed in this paper are those of the authors and do not necessarily represent the project views.

References

[1] International Telecommunication Union (ITU), "IMT Vision - Framework and overall objectives of the future development of IMT for 2020 and beyond," Telecommunication Standardization Sector of ITU, Tech. Rep., Sep. 2015.

[2] 3GPP TR 38.913 v14.1.0, "Study on Scenarios and Requirements for Next Generation Access Technologies," Mar. 2017.

[3] G. Pocovi, B. Soret, M. Lauridsen, K. I. Pedersen, and P. Mogensen, "Signal Quality Outage Analysis for Ultra-Reliable Communications in Cellular Networks," in *2015 IEEE Globecom Workshops*, Dec. 2015.

[4] 3GPP TR 36.881 v14.0.0, "Study on latency reduction techniques for LTE," Jul. 2016.

[5] K. I. Pedersen, G. Berardinelli, F. Frederiksen, P. Mogensen, and A. Szufarska, "A Flexible 5G Frame Structure Design for Frequency-Division Duplex Cases," *IEEE Communications Magazine*, vol. 54, no. 3, pp. 53–59, Mar. 2016.

[6] C. She, C. Yang, and T. Q. S. Quek, "Radio Resource Management for Ultra-Reliable and Low-Latency Communications," *IEEE Communications Magazine*, vol. 55, no. 6, pp. 72–78, Jun. 2017.

[7] B. Soret, P. Mogensen, K. I. Pedersen, and M. C. Aguayo-Torres, "Fundamental Tradeoffs among Reliability, Latency and Throughput in Cellular Networks," in *2014 IEEE Globecom Workshops*, Dec. 2014.

[8] G. Pocovi, B. Soret, K. I. Pedersen, and P. Mogensen, "MAC Layer Enhancements for Ultra-Reliable Low-Latency Communications in Cellular Networks," in *IEEE International Conference on Communications (ICC) 2017 Workshop*, May 2017.

[9] 3GPP TS 36.213 V14.2.0, "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures," Mar. 2017.

[10] P. Schulz, M. Matthe, H. Klessig, M. Simsek, G. Fettweis, J. Ansari, S. A. Ashraf, B. Almeroth, J. Voigt, I. Riedel, A. Puschmann, A. Mitschele-Thiel, M. Muller, T. Elste, and M. Windisch, "Latency Critical IoT Applications in 5G: Perspective on the Design of Radio Interface and Network Architecture," *IEEE Communications Magazine*, Feb. 2017.

[11] D. Jiang, H. Wang, E. Malkamaki, and E. Tuomaala, "Principle and Performance of Semi-Persistent Scheduling for VoIP in LTE System," in *2007 International Conference on Wireless Communications, Networking and Mobile Computing*, Sep. 2007.

[12] RP-161788, "V2V Work Item Completion," Sep. 2016.

[13] R1-1705246, "UL grant-free transmission for URLLC," Apr. 2017.

[14] 3GPP TSG RAN WG1, "RAN1 Chairman's Notes," Jan. 2017.

[15] R1-1612246, "Discussion on HARQ support for URLLC," Nov. 2016.

[16] 3GPP TSG RAN WG1, "RAN1 Chairman's Notes," Feb. 2017.

[17] 3GPP TR 36.802 v14.0.0, "Study on New Radio Access Technology," Mar. 2017.

[18] K. M. Rege, M. A. Kocak, K. Balachandran, J. H. Kang, and M. K. Karakayali, "On the Design of Preamble for Autonomous Communications with Extended Coverage," in *VTC Spring 2017 - IEEE 85th Vehicular Technology Conference*, Jun. 2017.

[19] Z. Zhao, D. Miao, Y. Zhang, J. Sun, H. Li, and K. Pedersen, "Uplink Contention Based Transmission with Non-Orthogonal Spreading," in *VTC Fall 2016 - IEEE 84th Vehicular Technology Conference*, Sep. 2016.