

Strategies to meet the configured repetitions in URLLC Uplink Grant-Free transmission

Trung-Kien Le[†], Umer Salim^{*}, Florian Kaltenberger[†]

[†]EURECOM, Sophia-Antipolis, France

^{*}TCL Mobile, Sophia-Antipolis, France

Trung-Kien.Le@eurecom.fr

Abstract—In Ultra-Reliable Low-Latency Communication (URLLC), the user (UE) can be configured to transmit in grant-free/configured-grant (GF/CG) resources for uplink (UL) transmission that does not require the UE to transmit scheduling request (SR) and receive UL grant to reduce latency. In addition, the UE is also configured to transmit automatically a specific number of repetitions without waiting feedback. However, these repetitions are only allowed to carry out in an interval with period P to avoid identity (ID) confusion in a Hybrid automatic repeat request (HARQ) process. Thereby, there is a chance that the UE cannot transmit all repetitions as configured if data arrives late and it leads to a drop of reliability. Two approaches are proposed in this paper to cope with this problem. This first approach requires an usage of the explicit HARQ feedback structure and the second one is related to an additional SR transmitted by the UE in parallel with data. The numerical results show the benefit of these two methods in increasing system performance in case of less configured repetitions made when they help the system to avoid or reduce packet loss due to Demodulation Reference Signal (DMRS) miss-detection.

Index Terms—5G, URLLC, grant-free repetitions, uplink scheduling scheme, explicit HARQ feedback, scheduling request

I. INTRODUCTION

The advent of new applications such as remote surgery, vehicle-to-everything communication, etc. with high demands of latency and reliability requires that 5G supports URLLC. The strict requirements of URLLC are given in [1]: “A general URLLC reliability requirement for one transmission of a packet is 10^{-5} for 32 bytes with a user plane latency of 1 ms”. These requirements continue to rise in the recent meetings of the 3rd Generation Partnership Project (3GPP): “Higher reliability (up to 10^{-6}), higher availability, short latency in the order of 0.5 to 1 ms, depending on the use cases (factory automation, transport industry and electrical power distribution)” [2].

A. Techniques accepted in 3GPP Release 15

In order to facilitate the implementation of URLLC, 3GPP made three important decisions about physical layer design.

One of the aspects is a permission to use larger subcarrier spacings (SCS). In 5G, SCS is allowed to have a value up to 240 kHz instead of an unique value of 15 kHz as in LTE [3]. The next aspect is about mini-slot based transmission that allows a UE to be scheduled in a period of one or several symbols rather than a whole slot [4]. These two decisions bring down time alignment of arriving packets.

The third aspect is related to the UL transmission in GF region to lessen time consumption of SR and UL grant [5].

B. Repetition problem in URLLC GF UL transmission

As mentioned in Section I-A, in UL transmission, the base station (called gNB) can configure a set of GF resources to one or more UEs with a periodicity defined by parameters in Radio Resource Control (RRC) from higher layer. When a UE is configured to transmit in GF resources, it can transmit data immediately instead of sending SR and waiting for UL grant as grant-based (GB) transmission.

To further reduce latency as well as increase reliability of the transmission, the URLLC UEs transmitting in the GF regions are also configured to transmit automatically a number of repetitions (called K repetitions with $K \in \{1, 2, 4, 8\}$ [6]) defined by a parameter repK from higher layer instead of waiting an UL grant to reschedule data if necessary after each repetition. Nevertheless, the UEs are only permitted to carry out the repetitions in one HARQ process with a periodicity P as shown in Fig. 1 (a set of allowed periodicity P are defined in [6]). Thereby, the UEs stop to do the repetitions at the boundary of a HARQ process even if they still do not reach the configured number of repetitions. In Fig. 1, an interval P contains 4 GF occasions, the UE is configured to do 4 repetitions. In the first period, data comes before all the 4 GF occasions so the UE is able to do 4 repetitions for the first packet as configured. However, when data comes in the second period, there are only 3 GF occasions left in that period. This means that the UE only can carry out 3 repetitions that are less than the configured number. Similarly, in the fourth period, the UE only can transmit the packet 2 times.

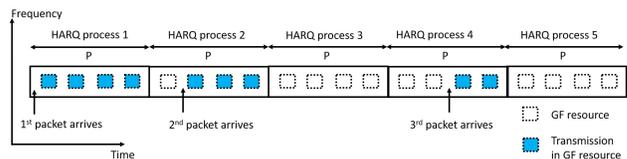


Fig. 1. Less than K repetitions in GF UL transmission.

This constraint is to help the gNB avoid a confusion in HARQ IDs of different HARQ processes that preventing its from recognizing the first repetition to do soft combining or determining the UE IDs to send an UL grant. However, this constraint also makes the system face a degradation

of UL transmission's performance due to a smaller number of repetitions that is catastrophic to the URLLC UEs with high reliability requirement. It also impacts latency when the gNB cannot decode the packet because the configured number of repetitions are not transmitted, it must schedule a retransmission by an UL grant and the UE only carries out more retransmissions after receiving this UL grant.

C. Prior art

In 3GPP Release 15, an option for the UE is to wait until the next HARQ process to transmit the whole K repetitions if it cannot make K repetitions as configured in the current process because data comes late. In other word, data is only allowed to be transmitted at the beginning of a period. This method leads to huge latency and might not suitable for URLLC.

In [7] and [8], the UE does not stop at the boundary of a period P if the configured number of repetitions have not been reached. The UE is allowed to continue to transmit the repetitions across the boundary and only stops after transmitting all K repetitions. To avoid a confusion of HARQ ID as mentioned in Section I-B, a new mechanism needs to be defined to communicate explicitly HARQ ID from the UE to the gNB and results in big overhead and effort in standardization.

Multiple configurations in GF region are proposed in [9] and [10]. A UE is configured with configurations having different starting time offsets so it can choose the nearest configuration to start a transmission without having to wait long time until the next period and guarantee the number of repetitions. However, this method causes big overhead from downlink control information (DCI) used to configure different configurations to the UEs. Resource consumption is another concern, if the configurations are configured separately without any overlap.

[11] and [12] propose to use shared resource for URLLC repetitions to improve resource utilization. However, they do not count a constraint that the UE cannot do repetitions crossing the border of a period. If this constraint is not solved, it might lead to a number of repetitions smaller than the number configured and transmission reliability is degraded.

In this work, two methods are presented to enhance the performance of URLLC UL GF transmission in case the UE cannot make K repetitions as configured. The first method described in Section II applies an explicit HARQ feedback structure to the transmissions with less than K repetitions. The second method in Section III asks the UEs with less than K repetition to transmit a SR in parallel with data packet. These two methods guarantee a retransmission in case of DMRS miss-detection due to a smaller number of repetitions. Section IV shows numerical results and performance evaluation. Finally, Section V provides the concluding remarks.

II. IMPROVING RELIABILITY-LATENCY BY FLEXIBLY MOVING TO EXPLICIT HARQ FEEDBACK STRUCTURE

A. Operation of explicit HARQ feedback structure

The HARQ structure in UL GF transmission is timer-based. This means that there is no explicit acknowledgement (ACK)

feedback sent from the gNB to the UE if data is decoded correctly. Instead, the UE uses a timer. If it does not receive an UL grant to schedule a retransmission at the end of time configured in the timer, the transmission is considered as successful. This structure has a drawback because the UE cannot differentiate between a successful transmission and a miss-detection when the gNB cannot decode DMRS to identify the UE and transmits UL grant. Therefore, in case of miss-detection, the UE does not receive any signal from the gNB, it assumes a successful transmission and drops data in buffer. This behavior impacts reliability of a transmission and becomes more severe in less-than- K -repetitions situation because the gNB has a smaller number of configured repetitions to detect successfully DMRS. Once the UE is not able to carry the configured number of repetitions, the QoS of transmission is badly affected and there is a higher probability that the gNB cannot decode DMRS sequence. It leads to a degradation of URLLC transmission due to packet dropped by the UE. To handle the issue with GF transmissions with less than K repetitions, we propose to use explicit HARQ feedback from the gNB.

An UL transmission results in three scenarios: one for correct decoding and two scenarios for incorrect decoding. In explicit HARQ feedback structure, the UE's behavior in each scenario will be analyzed as follows.

The first scenario is about correct data decoding. The gNB tries to combine all the repetitions of a transport block (TB) to facilitate data decoding, and the number of these repetitions can be less than K as per the previous discussion. For a normal operation, the gNB is capable of identifying the repetitions concerning a specific TB. Thus, whenever the gNB is able to correctly decode a TB, and it sees that it was sent with less than K repetitions, it will send an explicit ACK for this TB to the transmitting UE.

The second scenario is about a failure of data decoding with a successful UE identification. When the UE transmits less than K repetitions, it is possible that the data decoding is not successful but the gNB is able to identify the UE transmitting the TB with less than K repetitions through identification of UE specific DMRS sequence which it was configured with as part of CG configuration. In this case, the gNB will reschedule a retransmission of the previously transmitted TB.

The third scenario is related to UE identification failure. It is where the proposed explicit feedback becomes pivotal. The bad quality of received data may lead to a situation when the gNB is unable to identify the transmitting UE. This situation is the most damaging for the URLLC UEs/applications due to their tight constraints on latency and reliability. With a timer-based HARQ structure, which is currently used for URLLC transmissions in 3GPP Release 15, this situation leads to different understanding at the gNB and at the UE. The gNB, being unable to identify the UE, cannot schedule the retransmission. The UE, upon receiving no UL grant for retransmission, considers the packet successfully decoded at the gNB and discards the buffer upon the expiry of HARQ timer.

Although the situation when the gNB cannot identify the

UE may be caused by a number of reasons such as the very bad channel conditions, large amount of interference or insufficient number of actual repetitions, the configuration parameters of CG transmission, in particular MCS and the number of repetitions K , are designed to combat most of these adverse effects. On the other hand, if the configured number of repetitions cannot be made, this brings the CG operation point to a lower QoS target than the desired operating point.

In the proposed technique, whenever the UE transmits less than K repetitions, the TB in question is supposed to operate with explicit HARQ based feedback. In general, the gNB can identify transmissions with less than K repetitions thanks to DMRS detection and CG window boundary knowledge. When the gNB fails to identify the transmitting UE and sends no ACK or UL grant to this UE, the UE upon expiry of configured HARQ timer re-transmits automatically the TB.

The retransmission timing and resources can be configured as part of the explicit HARQ feedback configuration. One suitable option is to retransmit in the closest CG periodic window after the expiry of HARQ feedback timer. The HARQ feedback timer should include the time for the gNB to decode the data and find the suitable occasion for potential DL transmission of HARQ ACK or UL grant.

B. Design of explicit HARQ feedback

In the proposed strategy, the UEs are allowed to request explicit HARQ feedback for certain TBs. Thus, a design for the explicit HARQ feedback in general may be needed. One strategy can be to define a channel where HARQ ACK can be transmitted. This can be similar to Physical HARQ Indicator Channel (PHICH) as specified in 4G LTE but this requires a lot of specification effort and high resource overhead. The rationale is that in typical operation mode, the UEs will not request explicit HARQ feedback to reduce overhead and only in exceptional cases it will be required.

With this in view, the proposal is to use the UL grant (which is DCI) as an explicit HARQ feedback. This DCI can be sent with UE specific configured scheduling-radio network temporary identifier (CS-RNTI) which is used with configured GB transmissions. If the gNB is able to successfully decode the data, it sends DCI to this UE with the same HARQ process number (HARQ ID) as of the successfully received TB. To avoid any confusion between DCI used as feedback and DCI used as UL grant, new data indicator (NDI) field can be set to 0. Further, some of the fields in the DCI such as the time and frequency resource assignment fields are set to 0 to help the UE differentiate DCI used as feedback and DCI used for other purposes.

III. IMPROVING RELIABILITY-LATENCY BY FLEXIBLY TRANSMITTING AN ADDITIONAL SCHEDULING REQUEST

In this section, another scheme is proposed to deal with the problem of less-than- K repetitions. In this scheme, to improve the reliability of UL GF transmissions, whenever the UE transmits less than the configured number of repetitions for a TB, it sends SR to the gNB in parallel to transmission

of TB with less than K repetitions. This SR provides another mean for the gNB to detect UE ID and compensates the drop of reliability.

3GPP Release 15 does not allow transmission of Physical uplink control channel (PUCCH) and Physical uplink shared channel (PUSCH) simultaneously. The UE transmits UCI encoding SR, HARQ feedback, etc on PUCCH. Therefore, the UE multiplexes UCI and PUSCH if it wants to transmit uplink control information (UCI) while sending PUSCH. This strategy allows the UE to transmit SR in case of less than K repetitions. However, in UCI and PUSCH multiplexing, if the gNB cannot detect DMRS of PUSCH due to bad channel, there is high probability that the gNB also cannot decode UCI (SR) to find the UE ID. Thus, multiplexing strategy might not enhance the performance of UE ID detection. For this reason, SR should be configured to be transmitted on the configured PUCCH resources. The gNB upon receiving PUCCH and PUSCH from the same UE will understand that the SR in PUCCH is for the same TB sent in PUSCH for the UE that is only able to make less than K repetitions.

Table I shows in tabular format the UE and gNB actions for strategy of SR transmission in parallel to TB transmission.

The parallel transmission of PUCCH and PUSCH maybe slightly onerous for certain UEs because the UEs need to create the gap in the resource grid between PUCCH and PUSCH to protect them from interference. Nevertheless, considering that the main focus is here on URLLC type of UEs with strict latency and reliability targets, this overhead may be acceptable.

If the UE is transmitting different types of traffic at the same time, the gNB can differentiate the proposed SR transmitted in parallel with PUSCH in GF resources from a standalone classic SR which is sent to the gNB to have the UL resources scheduled through the HARQ ID of the transmitted TB included in the SR. As upon receiving the TB, the gNB will know its HARQ ID from its timing window, it will be able to conclude that the SR concerns the same TB or not.

Under certain situations, it may be beneficial to allow a hybrid scheme where the UEs can flexibly choose between an explicit HARQ feedback structure or sending an SR in parallel to the transmission of a TB, when they are able to make less than K repetitions. The simplest scheme would be the one that the gNB configures the UEs to follow one of these two schemes. Alternatively, the UEs can be configured to choose one these two schemes. In that case, it would make sense to have an explicit indication in the TB for the explicit HARQ feedback. If the UEs choose to transmit SR in parallel to the transmission of the TB, they do not trigger explicit feedback with the TB transmission. This can be advantageous in the situations when there is at least a suitable SR transmission occasion available where the UEs can transmit SR for the TB in question. In the contrary situation, the UEs do not transmit SR in parallel to the transmission of the TB but send an indication to trigger explicit HARQ feedback. This can be more advantageous if there is no suitable SR transmission occasion and a SR transmission may harm the latency budget.

For the traffic with extremely stringent latency-reliability

TABLE I
SR TRANSMISSION WITH TB AND ACTIONS FOR THE GNB AND THE UE

Case	GF PUSCH	SR in PUCCH	gNB understanding	gNB action	UE action
1	Correctly decoded at the gNB	Correctly decoded at the gNB	The gNB knows that SR is for the decoded TB	Indicate a correct detection (ex: using UL grant with the same HARQ ID)	Discard data upon receiving the gNB indication
2	Correctly decoded at the gNB	Incorrectly decoded at the gNB	The gNB upon correctly decoding the data and seeing less than K rep knows about missing SR. This case should be rare as SR is sent with strong coding	Indicate correct detection (ex: using UL grant with same HARQ ID)	Discard the data upon the gNB indication
3	Incorrectly decoded at the gNB but UE Identified through DMRS	Correctly decoded at the gNB	the gNB understands that UE sent SR along with the TB that it failed to decode	The gNB sends the UL grant for re-transmission	The UE follows the UL grant for re-transmission
4	Incorrectly decoded and UE Identification Failure at the gNB	Correctly decoded at the gNB	The gNB completely misses the CG transmission due to failure in UE identification but it receives SR. From the timing of SR and CG configurations, the gNB knows its decoding failure	The gNB sends the UL grant for transmission	The UE follows the UL grant for transmission and re-transmits the data
5	Incorrectly decoded at the gNB but UE Identified through DMRS	Incorrectly decoded at the gNB	The gNB identifies the UE from PUSCH. If it can identify the case of less than K repetitions, it knows also about SR detection failure	The gNB sends the UL grant for re-transmission	The UE follows the UL grant for re-transmission
6	Incorrectly decoded at the gNB and UE Identification Failure at the gNB	Incorrectly decoded at the gNB	The gNB has no indication about UE transmission	No action	The UE can be configured to retransmit in the subsequent CG resources or SR

constraints, it can be foreseen that both mechanisms, explicit HARQ feedback and transmission of SR, are triggered in parallel to maximize the reliability within a short time interval when the UE transmits the TB with less than K repetitions.

IV. NUMERICAL RESULTS AND PERFORMANCE EVALUATION

TABLE II
SIMULATION PARAMETERS

Parameters	Values
Waveform	CP-OFDM
Subcarrier spacing	60kHz
Channel model	Rician
K factor	1
Number of allocated PRB	8
DMRS detection mechanism	Time-domain correlation

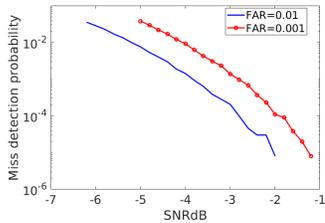


Fig. 2. DMRS detection performance.

Simulation parameters for the performance of DMRS detection of each repetition in Fig. 2 are shown in Table II. For each DMRS detection, the correlation result is compared with a threshold to determine whether DMRS exists or not. This threshold is chosen according to a target false alarm rate (FAR) indicating the cases that the gNB determines the existence of DMRS while in reality there is no DMRS transmitted.

A higher threshold is required for a lower FAR but also results in more missed detection. Because channel estimation to decode data as well as recognizing UE ID to reschedule a retransmission if necessary in conventional scheme depend on DMRS detection, a degradation of DMRS detection due to a smaller number of repetitions than configured makes the system not be able to support reliability URLLC requirement.

TABLE III
PERFORMANCE COMPARISON OF DIFFERENT SCHEMES WHEN DATA COMES AFTER THE FIRST OCCASION IN A PERIOD AT $SNR = -5dB$ AND $FAR = 0.001$

Case	Starting time offset (ms)	Number of repetitions	UE ID miss-detection probability	Retrans in ID miss-detection	Total UE ID miss-detection probability
Conventional transmission	0	3	5.5×10^{-5}	0	5.5×10^{-5}
Conventional transmission with the UE waiting the next period	0.75	1	0.038	0	0.038
Transmission with explicit ACK	0	3	5.5×10^{-5}	1	2.1×10^{-6}
Transmission with SR	0	3	2.1×10^{-6}	0	2.1×10^{-6}

In considered system, periodicity P of HARQ process is 4 slots spreading in 1ms with SCS of 60kHz. The UE is configured to transmit 4 repetitions. As specified in [3], each slot only contains one repetition so one repetition consumes 0.25ms and the UE needs 4 slots to carry out all 4 configured repetitions and satisfies URLLC latency budget of 1ms.

Table III, Table IV and Table V show the performance of DMRS detection at SNR of -5dB and FAR of 0.001 in

TABLE IV
PERFORMANCE COMPARISON OF DIFFERENT SCHEMES WHEN DATA COMES AFTER THE SECOND OCCASION IN A PERIOD AT $SNR = -5dB$ AND $FAR = 0.001$

Case	Starting time offset (ms)	Number of repetitions	UE ID miss-detection probability	Retrans in ID miss-detection	Total UE ID miss-detection probability
Conventional transmission	0	2	10^{-3}	0	10^{-3}
Conventional transmission with the UE waiting the next period	0.5	2	10^{-3}	0	10^{-3}
Transmission with explicit ACK	0	2	10^{-3}	2	2.1×10^{-6}
Transmission with SR	0	2	5.5×10^{-5}	0	5.5×10^{-5}

TABLE V
PERFORMANCE COMPARISON OF DIFFERENT SCHEMES WHEN DATA COMES AFTER THE THIRD OCCASION IN A PERIOD AT $SNR = -5dB$ AND $FAR = 0.001$

Case	Starting time offset (ms)	Number of repetitions	UE ID miss-detection probability	Retrans in ID miss-detection	Total UE ID miss-detection probability
Conventional transmission	0	1	0.038	0	0.038
Conventional transmission with the UE waiting the next period	0.25	3	5.5×10^{-5}	0	5.5×10^{-5}
Transmission with explicit ACK	0	1	0.038	3	2.1×10^{-6}
Transmission with SR	0	1	10^{-3}	0	10^{-3}

various schemes and arrival time of data. As can be seen with conventional transmission, due to the constraint of boundary of a period P , the UE cannot transmit 4 configured repetitions and it affects DMRS detection's performance. The later the packet comes, the worse DMRS detection is. In the timer-based feedback of the conventional scheme, the packet is lost because the UE assumes a successful transmission in case of DMRS miss-detection. Even if the UE waits for the next period with the intention of carrying of K repetitions configured as the second scheme, it also cannot achieve that intention because of URLLC latency requirement of 1ms. The more time the UE waits, the less time it has to transmit the repetitions. In consequence, it also cannot transmit K repetitions and DMRS miss-detection grows.

The presence of explicit feedback solves the problem of an increase of DMRS-miss detection because it helps the UE differentiate between a successful transmission and DMRS-miss detection. Therefore, the UE is able to retransmit packet even if the gNB does not detect UE ID by DMRS detection

and reliability is guaranteed (total miss-detection probability of 2.1×10^{-6} smaller than the conventional schemes) as shown in Table III, Table IV and Table V.

These three tables also show that an additional SR transmitted in parallel with data provides another chance for the gNB to identify the UE and compensates for the missing repetitions due to boundary constraint so system performance is enhanced.

The usages of an explicit feedback or an additional SR increase resource consumption. However, these approaches are only used in the critical situation as less than K repetitions made by the UE so a growth of resource consumption is limited. In addition, in case of less than K repetitions, the URLLC performance is degraded. Regarding the priorities of latency and reliability, resources used for the explicit feedback or the SR are necessary to guarantee those strict requirements.

V. CONCLUSION

In URLLC GF UL transmission, a packet is configured with K repetitions to achieve target QoS. However, because of boundary of a period P and arrival time of data, the UEs might transmit less than K repetitions so the target QoS is not attained. This paper presents two strategies to help the UEs achieve the target QoS in case of less than K repetitions as shown in the results. These approaches relating to an explicit HARQ structure and an additional SR can be used individually or combined together based different scenarios.

ACKNOWLEDGMENTS

The work described in this paper has been partially supported by TCL and the H2020 project 5GENESIS.

REFERENCES

- [1] 3GPP TR 38.913 v15.0.0, "Study on scenarios and requirements for next generation access technologies."
- [2] Huawei, HiSilicon, Nokia, Nokia Shanghai Bell, "New SID on Physical Layer Enhancements for NR URLLC". 3GPP RP-182089, TSG-RAN#81, Gold Coast, Australia, Sept 10–13, 2018.
- [3] 3GPP TS 38.211 v15.3.0, "Physical channels and modulation."
- [4] 3GPP TR 38.802 v14.2.0, "Study on new radio access technology physical layer aspects."
- [5] 3GPP TS 38.214 v15.3.0, "Physical layer procedures for data."
- [6] 3GPP TS 38.331 v15.3.0, "Radio Resource Control (RRC) protocol specification."
- [7] Ericsson, "Enhancement of Configured Grant for NR URLLC". 3GPP R1-1812162, RAN1#95, Spokane, USA, November 12–16, 2018.
- [8] Huawei, HiSilicon, "Enhanced UL configured grant transmissions", 3GPP R1-1812226, RAN1#95, Spokane, USA, November 12–16, 2018.
- [9] Intel Corporation, "On enhanced Configured Grant PUSCH for eURLLC", 3GPP R1-1812506, RAN1#95, Spokane, USA, November 12–16, 2018.
- [10] Sony, "Discussion on enhanced UL grant-free transmissions", 3GPP R1-1812746, RAN1#95, Spokane, USA, November 12–16, 2018.
- [11] R. breu, G. Berardinelli, T. Jacobsen, K. Pedersen and P. Mogensen, "A Blind Retransmission Scheme for Ultra-Reliable and Low Latency Communications", 2018 IEEE 87th Vehicular Technology Conference (VTC Spring), June 2018.
- [12] Z. Zhou, R. Ratasuk, N. Mangalvedhe and A. Ghosh, "Resource Allocation for Uplink Grant-Free Ultra-Reliable and Low Latency Communications", 2018 IEEE 87th Vehicular Technology Conference (VTC Spring), June 2018.
- [13] Trung-Kien Le, Umer Salim, Florian Kaltenberger, "Optimal reserved resources to ensure the repetitions in Ultra-Reliable Low-Latency Communication Uplink Grant-free transmission", 2019 European Conference on Networks and Communications (EuCNC), June 2019.