

Optimal reserved resources to ensure the repetitions in Ultra-Reliable Low-Latency Communication 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—To meet the strict requirements of Ultra-Reliable Low-Latency Communication in the uplink, grant-free uplink transmission has been specified, allowing the UE to transmit data in a random-access fashion without first transmitting a scheduling request and then waiting for an uplink grant from the gNB. To further increase the reliability, these grant-free uplink transmissions can be repeated without waiting for HARQ feedback from the gNB. However, these repetitions have to happen within a certain interval to avoid a confusion in HARQ IDs of different HARQ processes. When a UE starts transmitting late in the interval, it, therefore, can not exploit all the possible repetitions and thus reliability and latency decrease. In this paper, a scheme based on reserved resources is proposed to ensure the number of repetitions in a specific period. The size of each reserved resource is optimized depending on its position so as to reduce resource consumption. The scheme evaluated by theoretical analysis and numerical results shows its benefits to system performance.

Index Terms—5G, URLLC, repetitions, uplink scheduling scheme, reserved resources

I. INTRODUCTION

In 5G New Radio (NR), Ultra-Reliable Low-Latency Communication (URLLC) is one of the key features specified by The 3rd Generation Partnership (3GPP) to serve the applications such as augmented virtual reality, tactile internet and industrial automation. This new service poses a huge challenge due to a high demand of two conflicting factors: reliability and latency. The requirement for URLLC is specified 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”. Besides the baseline URLLC performance, the next release of 3GPP aims to achieve higher requirements: “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 Release 15, new aspects have been made and agreed by 3GPP to support URLLC.

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 a unique value of 15 kHz as in LTE [3]. This decision brings down transmission duration of packets from 1ms down to $62.5\mu\text{s}$. Mini-slot based transmission further

helps to reduce latency and transmission duration of packets [4]. A user (UE) can be scheduled in a period of one or several symbols rather than a whole slot.

The third aspect is related to the uplink (UL) transmission in grant-free (GF) region to reduce the time consumption of scheduling request (SR) and uplink grant (UL grant) [5]. The base station (gNB) can configure a set of GF resources to one or more UEs with a periodicity defined by parameters in 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.

B. Repetition problem in URLLC GF UL transmission

As mentioned in Section I-A, in UL transmission, the gNB configures the UEs with high priority and strict requirements to transmit in the GF regions. In addition, it also configures the number of repetitions K that these UEs need to carry out by a parameter $\text{rep}K$ from higher layer in order to guarantee transmissions reliability and latency. K has values 1, 2, 4 and 8 as standardized in [6]. The UEs transmit the repetitions automatically in the GF regions without waiting for Hybrid automatic repeat request (HARQ) feedback or UL grant from the gNB. However, the UEs are only allowed to do repetitions in one interval with a periodicity P ranging from several symbols to several slots (a set of allowed periodicities P is defined in [6]) and prohibited to retransmit packet as configured by $\text{rep}K$ crossing boundary of that interval. This constraint is to help the gNB avoid a confusion in HARQ identities (IDs) of different HARQ processes. Therefore, depending on the arrival time of data in relation to the periodicity P , the number of repetitions might be smaller than the configured number because the UEs need to stop their transmission at the last transmission occasion in the period P .

Fig. 1 illustrates the situation when the number of configured repetitions is not ensured due to the constraint of boundary of period P . 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.

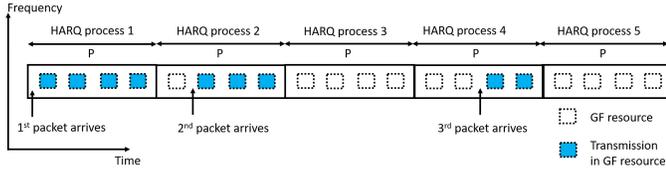


Fig. 1. Less than K repetitions in GF UL transmission.

It is evident that when the packet comes after the first GF occasions in a period, the UE transmits the packet with a smaller number than the number configured by $\text{rep}K$. It degrades the reliability of UL transmission. The situation becomes more severe for the URLLC UEs with high reliability requirement. Moreover, latency of a transmission also increases because with a smaller number of repetitions, the gNB has a higher probability of failing to decode the packet and needs to schedule a retransmission. In that case, the UE needs to wait the gNB to decode the repetitions of the packet and transmit an UL grant to reschedule a retransmission if necessary and it has a huge impact on latency.

C. Prior art

In 3GPP Release 15, the only option for the UE is to wait until the next period to start the transmission and fulfill the configured number of repetitions. Nevertheless, this option causes big delay and prevents the UE from meeting the URLLC latency requirement, especially for low SCS.

In [7] and [8], repetitions are allowed to cross the boundary of a periodicity when the UE cannot make the requiring repetitions. This solution leads to a confusion of HARQ ID that makes the UE not differentiate the initial transmission and the retransmissions in order to combine and decode a specific codeblock. Besides that, a new mechanism needs to be defined to communicate explicitly HARQ ID from the UE to the gNB and results in overhead and effort in standardization.

Multiple configurations in GF region are proposed in [9] and [10]. A UE is configured with configurations having different time offsets so it can choose the nearest configuration to start a transmission and guarantee the number of repetitions. The main concern of multiple configurations is the control overhead and delay because of signals used to configure the UE as well as resource consumption if resources in the configurations are not overlapped.

[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 configured number. In addition, shared resources are allocated periodically with the same size for all transmission occasions. These two drawbacks bring an increase of resource consumption and a drop of reliability.

In this work, an UL GF transmission scheme with reserved resources is described. Reserved resources are used and shared among the UEs when they need to transmit repetitions crossing the border of a transmission period P in order to achieve the target number of repetitions. These reserved resources have different sizes that are optimized based on the positions of their transmission occasions. Firstly, reserved resources and calculations to optimize their sizes are presented in Section II. The process allowing the UEs to access to these resources is also discussed. Section III shows numerical results obtaining from the equations derived in diverse scenarios. Finally, Section IV provides the concluding remarks for this work.

II. STRATEGY TO ENSURE THE CONFIGURED NUMBER OF REPETITIONS IN UL GF TRANSMISSION

A. Reserved resources

To make the URLLC UEs achieve the strict requirements, a strategy to ensure that the UEs can transmit the number of repetitions as configured by $\text{rep}K$ from higher layer is indispensable. Reserved periodic resources are proposed to be created and assigned to multiple UEs by the gNB so that they are likely to retransmit data in case the transmissions in the reserved resources are necessary to ensure the configured number of repetitions. These reserved resources have the same periodicity as the GF resources.

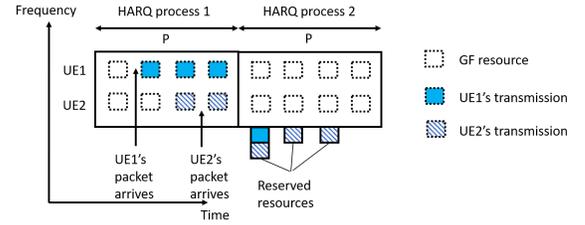


Fig. 2. Reserved resources for repetitions.

The use of reserved resources is shown in Fig. 2. There are 2 UEs considered with GF resources in different bandwidth and each UE is configured with $\text{rep}K = 4$. UE1's data comes after the first GF occasion in the first period so it only can do 3 repetitions in the period. In order to attain 4 repetitions, the UE1 retransmits data in the first reserved resource of the next period. Similarly, UE2's data arrives at the last GF occasion and only one repetition can be made. Thus, the UE2 uses the 3 reserved resources in the next period to achieve the configured number of repetitions.

In the example, 4 repetitions are configured so 3 reserved resources are needed in the next period. To increase the efficiency of resource consumption, the reserved resources are shared among the UEs. The first reserved resource in Fig. 2 with 2 blocks is likely to be shared with more than 2 UEs while still attain the target collision probability being approximate to the collision probability in the GF resources. Similarly, the second and third reserved resources are also shared by a group with more than 2 UEs. The equation showing the relation among the number of UEs, the size of the reserved resources and the collision probability is derived in Section II-C.

B. System model

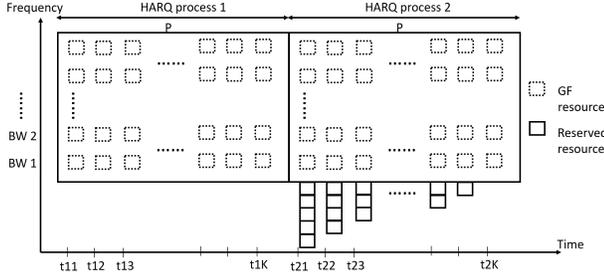


Fig. 3. UL transmission resources' distribution.

The system in Fig. 3 considered to calculate collision probability in the reserved resources contains N UEs. These N UEs are configured by the gNB to transmit in the periodical GF resources with the number of repetitions K set by parameter $\text{rep}K$. The UEs transmit the automatic repetitions of a packet in the consecutive GF resources. The GF resources in one frequency band can be shared among the UEs in a group. A period P corresponding to a HARQ process consists of GF transmission occasions equal to the number of repetitions K . The results derived below are also valid if the number of GF transmission occasions in a period are bigger than K . The reserved resources ($K - 1$ reserved resources) are configured periodically as GF resources in each period P . These resources are shared among the N UEs to retransmit data in case the number of repetitions that they can carry out in one period are less than the configured one. Each reserved resource has size of M_i blocks with index i indicating that the reserved resource is at the i th transmission occasion of a period. The size of each block in the reserved resources is the same as that of the GF resource in one transmission occasion.

The arrival of data for each UE follows a Poisson process with the average number of random access events in an interval λ calculated from the period between the GF resources T and an average packet inter-arrival time μ : $\lambda = T/\mu$.

C. Collision probability in reserved resources

With random access, the N UEs in system are allowed to use any block in the reserved resources of a specific transmission occasion if they need to do the transmissions in order to fulfill the configured number of repetitions.

The collision probability in the reserved resource at the first transmission occasion of a period (at t_{21} in Fig. 3) is calculated as follow by considering 1 UE of interest having a transmission in the reserved resource at t_{21} and the rest of ($N-1$) UEs. In time T between two GF resources, the probability that one UE has one or more random transmissions is

$$P_{data} = 1 - e^{-\lambda}, \quad (1)$$

The GF resources in one bandwidth can be shared between a group of the UEs so collision probability in the GF resources between the UE of interest and other UEs of group is

$$P_{c_GF} = 1 - e^{-\lambda(N_{UE_group}-1)}, \quad (2)$$

where N_{UE_group} is the number of the UEs that use the same frequency band of GF resources.

The reserved resource at t_{21} is used by a UE if its data comes after the first GF transmission occasion at t_{11} . The probability that one UE has transmission after the first GF transmission occasion in a period P is

$$P_d = 1 - (1 - P_{data})^{K-1}. \quad (3)$$

There is no collision in the first reserved resource of a period P at t_{21} if no UE rather than the UE of interest has a transmission after the first GF transmission occasion at t_{11} . The probability that no other UE from the set of $N - 1$ UEs has a transmission after the first GF occasion is calculated by

$$P_0 = (1 - P_d)^{N-1}. \quad (4)$$

In case other UEs in a set of $N - 1$ UEs has a transmission after the first GF occasion, the probability that n UEs have such transmission is

$$P_n = \binom{N-1}{n} P_d^n (1 - P_d)^{N-1-n}. \quad (5)$$

The probability that the UE of interest and n UEs do not access the same resource block in the first reserved resource at t_{21} is

$$P_{a0_n} = \left(\frac{M_1 - 1}{M_1}\right)^n. \quad (6)$$

The probability that the UE of interest does not collide with any other UE in the first reserved resource at t_{21} is calculated by

$$P_{sum} = \sum_{n=1}^{N-1} P_n P_{a0_n}. \quad (7)$$

From (4) and (7), the collision probability in the first reserved resource for the UE of interest is derived as

$$\begin{aligned} P_{c1} &= 1 - P_0 - P_{sum} \\ &= 1 - \left(\frac{M_1 - 1 + e^{-\lambda(K-1)}}{M_1}\right)^{N-1}. \end{aligned} \quad (8)$$

Based on the same calculating process, a general equation of collision probability for the reserved resource at any transmission occasion in a period can be derived as

$$P_{ci} = 1 - \left(\frac{M_i - 1 + e^{-\lambda(K-i)}}{M_i}\right)^{N-1}, \quad (9)$$

where $i \in [1, K - 1]$ is index indicating the position of the reserved resource based on the position of transmission occasion in a period.

D. Optimization of reserved resource size

From (9) with a set of parameters λ, N, K, i , and P_{ci} , the size of the reserved resources at any transmission occasion can be calculated. To guarantee the reliability of transmission in the reserved resources in comparison to that in the GF resources, the target probability P_{ci} is chosen to be approximate to P_{c_GF} in (2). If P_{ci} is set to have the same value in all the reserved resources, the trend of reserved resources' size in terms of their positions in a period is: $M_1 > M_2 > M_3 > \dots > M_{K-1}$. The size of the reserved resources to achieve a target probability decreases gradually from the first to the last reserved resource in a period. The design of the reserved resources following this decreasing

trend instead of using the same size for all of the reserved resource brings an efficiency of resource consumption because the size is optimized particularly for each reserved resource.

E. Group access to the reserved resources

In group access approach, one UE is only allowed to access and uses a specific part of the reserved resource pre-configured by the gNB. One part of the reserved resource is assigned and shared among a group of the UEs. For example, the UE1-3 is only permitted to access the resource blocks as shown in Fig. 4. On the other hand, other UE groups are also prohibited to access the resource blocks pre-configured to the UE1-3.

UE1 UE2 UE3			
	UE1 UE2 UE3		
		UE1 UE2 UE3	

Fig. 4. Group access to the reserved resource.

For a group of the UEs accessing to a part of the reserved resources, the collision probability is calculated by (9) as random access approach. The size of the whole reserved resource in each transmission occasion is sum of all parts assigned to the UE groups to guarantee a target probability.

Simulation in Section III shows that the sizes of the reserved resources for both random access and group access approach are the same to achieve an equal target probability. However, group access approach reduces the decoding burden in the gNB. To decode a retransmitted packet, the gNB only needs to search a part of the reserved resource instead of the entire reserved resource as random access approach. Thereby, power consumption and processing time drop dramatically.

When two or more UEs of the same group having access to a part of the reserved resources have data coming at the same time and need to use the reserved resource for repetitions, they will compete for the same part of the reserved resources at the consecutive transmission occasions as show in Fig. 5. The UE 1 and 2 have 3 repetitions and data coming after the first two GF occasions so they compete two times at both the first and second reserved resources with the arrangement in the blue circle.

To avoid that problem, the UEs are assigned to different groups of the reserved resources for different transmission occasions as illustrated in the red circle. The UE 1 only has the competition in both two reserved resources when both the UE 2 and 3 need to use the part of the reserved resources assigned to the UE 1 so the probability of having a competition decreases from P_{data} to P_{data}^2 with P_{data} calculated from (1).

In each reserved resource, the collision probability is still the same as calculated from (9) but the overall reliability of a UE transmitting in the reserved resources at different transmission occasions will be improved with the hopping of the UEs to various groups in the different reserved resources.

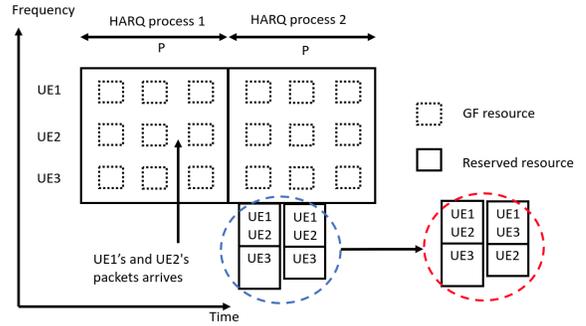


Fig. 5. Method to group the UEs in different transmission occasions.

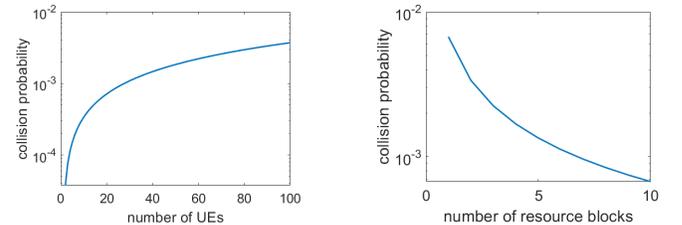
III. NUMERICAL RESULTS AND PERFORMANCE EVALUATION

A. Random access to the reserved resources

From (9), the number of the UEs that the system can support can be found. Besides, the number of resource blocks in each reserved resource are also calculated to sustain that system.

The simulation starts with the reserved resource in the first transmission occasion of a period. The set of parameters is $M_1 = 10$, $K = 4$, $\lambda = 1.25 \times 10^{-4}$.

Fig. 6(a) shows collision probability in the reserved resource in the first transmission occasion in terms of the number of the UEs sharing that resource. From the graph, we see that if the target collision probability (P_{c1}) is 10^{-3} , the system can support 28 UEs in the first reserved resources.



(a) Collision probability with respect to the number of the UEs. (b) Collision probability of the second reserved resource.

Fig. 6. Collision probability.

As mentioned in Section II-B and II-C, the GF resources in a frequency band can be shared by a group of the UEs. 28 UEs calculated above can be divided in to 4 groups with 7 UEs in each group. The collision probability in the GF resources calculated from (2) is 7.5×10^{-4} that is approximate to the collision probability of 10^{-3} in the reserved resource.

When all the parameters (λ , N , K and P_{ci}) are kept the same, the sizes of the reserved resources in the second and the third transmission occasions are calculated as shown in Table I.

TABLE I
SIZES OF THE RESERVED RESOURCES WITH $K = 4$ AND RANDOM ACCESS

Position of reserved resources	1	2	3
Number of blocks	10	7	3

The percentage of resources saved in comparison to using the same size of 10 resource blocks for all the resources in [11] and [12] is: $(1 - (10 + 7 + 3)/(10 \times 3)) \times 100\% = 33\%$.

Fig. 6(b) illustrates collision probability with respect to the reserved resource's size in the second transmission occasion.

Another scenario is considered with a bigger number of repetitions: $M_1 = 10, K = 8, \lambda = 1.25 \times 10^{-4}$.

The system can support 12 UEs in the first reserved resources to achieve Pc target of 10^{-3} . These UEs can be divided into 2 groups with 6 UEs in one group that each group uses the GF resources in one bandwidth part. The collision probability in GF regions of a group of 6 UEs as scheduled above is: 6.25×10^{-4} . With 12 UEs, the sizes of the reserved resources in the transmission occasions are shown in Table II.

TABLE II

SIZES OF THE RESERVED RESOURCES WITH $K = 8$ AND RANDOM ACCESS

Position of reserved resources	1	2	3	4	5	6	7
Number of blocks	10	8	7	6	4	3	2

The percentage of resources saved in comparison to using the same size of 10 resource blocks for all the resources in [11] and [12] is: $(1 - (10 + 8 + 7 + 6 + 4 + 3 + 2)/(10 \times 7)) \times 100\% = 42.86\%$

As can be seen from two scenarios, the use of reserved resources with optimization from (9) brings an efficiency of resource utilization, especially for high configured number of repetitions. The resources needed are less than the approach of choosing blindly the same sizes for all reserved resources in different transmission occasions while still guarantees the number of repetitions with a target reliability.

Moreover, in the conventional scheme in 3GPP Release 15, the UE needs to wait until the next period if it cannot do the configured number of repetitions in the current period to fulfill the reliability requirement. Therefore, with $K = 4$ and 4 transmission occasions in a period, in the worst case, the UE must wait 3 transmission occasions equal to 3 slots or 0.75ms with SCS 60kHz. This latency is close to the URLLC requirement of 1ms and causes the UEs not be able to make 4 repetitions as configured in the next period. In comparison, the proposed scheme allows the UE to start the transmission immediately and reach the configured number of repetitions in target latency of 1ms (4 repetitions consume 1ms with SCS 60kHz) as well as meet the reliability requirement.

B. Group access to the reserved resources

Group access is simulated with the parameters as the first scenario in random access: $N = 28, K = 4, \lambda = 1.25 \times 10^{-4}$.

28 UEs are divided into 4 groups of 7 UEs to access the reserved resources. Each group is only allowed to access the part of the reserved resources pre-configured to them.

To satisfy $P_c = 10^{-3}$, the sizes of the reserved resources are shown in Table III.

The resource consumption is equal for random access and group access to the reserved resources as shown in Table I

TABLE III

SIZES OF THE RESERVED RESOURCES WITH $K=4$ AND GROUP ACCESS

Position of reserved resources	1	2	3
Number of blocks per group of reserved resources	2	2	1
Total number of blocks of reserved resources	8	8	4

and Table III. However, with group access, the complexity and processing time of decoding in the gNB reduce substantially because in the reserved resource at the first transmission occasion, the gNB only needs to find data of a UE in 2 reserved blocks instead of finding blindly data of a UE in 10 reserved blocks as in random access approach.

IV. CONCLUSION

This paper presents an approach with reserved resources shared among the UEs that allows them to carry out repetitions and reach the configured number so that reliability and latency requirements can be satisfied. Each reserved resource has an optimized size to balance reliability and resource consumption.

ACKNOWLEDGMENTS

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

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] B. Singh, O. Tirkkonen, Z. Li and M. A. Uusitalo, "Contention-Based Access for Ultra-Reliable Low Latency Uplink Transmissions", IEEE Wireless Communications Letters, April 2018.