Design and Performance Evaluation of Contention-based Transmission Schemes for URLLC Services

Patrick Brown, Salah Eddine Elayoubi, Matha Deghel, Ana Galindo-Serrano

Design and Performance Evaluation of Contention-based Transmission Schemes for URLLC Services

Patrick Brown[1], Salah Eddine Elayoubi[2], Matha Deghel[1], Ana Galindo-Serrano[1]
[1]Orange Labs, France
[2] L2S, CentraleSupelec, France

Abstract

We investigate in this paper uplink multiple transmission schemes for 5G Ultra-Reliable Low Latency Communications (URLLC) traffic. The URLLC class of services has been defined for applications requiring extremely stringent latency and reliability. We show that, in systems with episodic traffic and many users compared with the number of transmission resources, randomly transmitting multiple copies of a packet allows to meet the URLLC requirements. We develop analytical models for the packet loss rate for two contention based multiple transmission schemes and show that one outperforms the other in the parameter range for which the URLLC requirements are met. We then propose an advanced replication scheme where positions for the different replicas are pre-allocated to users so that users have pairwise-distinct positions. We show that this advanced scheme achieves very high reliability with low resource consumption.

Keywords: 5G; URLLC; grant-free transmission; contention based

Design and Performance Evaluation of Contention-based Transmission Schemes for URLLC Services

Patrick Brown[1], Salah Eddine Elayoubi[2], Matha Deghel[1], Ana Galindo-Serrano[1]
[1]Orange Labs, France
[2] L2S, CentraleSupelec, France

1. Introduction

Ultra-Reliable Low-Latency Communication (URLLC) is a class of services targeted by 5G with very stringent latency and reliability requirements [1]. A general URLLC requirement is $99.999\,\%$ target reliability with 1 ms (two-way) user-plane latency [2]. These targets may be regarded as conflicting as achieving a high reliability requires retransmisisons of the lost packets, that comes at a latency cost. We study in this paper resource allocation for the most challenging URLLC use cases that involve sporadic uplink transmissions. Examples of such scenarios are those related to Industrial Internet of Things (IIoT) with cyclic and sporadic traffic generation.

At the radio level, as the main contributors to latency are the PHY and MAC layers [3], multiple solutions are being proposed to reduce the latency induced at these levels and help achieve the very low latency and high reliability required by URLLC. Decreasing the Transmission Time Interval (TTI) length is one efficient way to shorten the latency in the 5G system [4, 5]. The studies in [6, 7] show that shortening both TTI and the required time for retransmitting the packet, is essential. Besides shortening the TTI, flexibility in the numerology and in the usage of the duplexing mode makes the 5G radio frame suitable for a dynamic allocation of resources to URLLC users. Authors in [8] evaluated the performance at system level for the industrial scenario and showed that with the adequate 5G radio configuration, the required low-latency and high reliability for industrial applications can be achieved. However, the reliability and latency enablers differ depending on the centralized or distributed nature of the system.

In the downlink, where the resource allocation is centrally performed by the base station, preemption is considered as the most efficient resource allocation technique. The impact of the resource preemption technique on the user plane latency and its coexistence with eMBB have been evaluated in [9, 10]. In [11], a joint resource allocation and Modulation and Coding Scheme (MCS) selection was proposed, and the tradeoff between the resources reserved for the initial transmissions and the resources consumed by retransmissions was exploited. Starting from these considerations, a global downlink URLLC optimization framework has been proposed in [12], including a system dimensioning and an enhanced Hybrid Automatic Repeat Request (HARQ) scheme.

When considering the uplink, however, most of the approaches based on a centralized scheduler no more apply. For instance, preemption and automatic reservation of

resources for retransmissions become difficult, if not impossible. *Grant-free* scheduling has to be used instead of the classical grant-based scheduling approach, especially for scenarios with very sporadic traffic [13]. In this grant-free fast uplink access, neither issuing a scheduling request nor waiting for a scheduling grant are required [14]. On the other hand, retransmission is a key enabler for improving the reliability performance [15], but again, using classical HARQ retransmission procedures introduces additional latency [15] and other re-transmission schemes are needed for URLLC, such as blind retransmission of replicas of each packet without waiting for ACKs. The authors in [16] propose to send these replicas in consecutive TTIs, where the resources used by each replica are randomly selected from the set of available RBs in each TTI. A similar approach has been already adopted as a solution in the 3GPP standard [17]. In a previous work [18], we proposed a completely random choice of replicas' positions. Such an approach will result in collisions between some of the (re)transmitted packets, which will impact the reliability level that can be achieved. Hence, it is important to carefully design the contention-based scheme, which will determine the resource allocation policy an active user will follow to send the replicas of each of its packets.

In this paper, we further enhance the contention-based approach by exploiting the presence of a central entity in the network that may perform the long-term allocation of replicas' positions to users. Indeed, with a random scheme, two users may choose exactly the same places, leading to a possible loss even when only two users are active. We propose a semi-distributed scheme where the resources to be used by users are pre-determined by a central entity and sent to users. This central allocation allows overcoming the random allocation drawback as different resource sets per user can be allocated and the base station knows where replicas can be found. We derive the exact expression of collision probability, which represents a reliability measure. Using numerical results, we illustrate the performance improvement that our scheme can yield as compared to state of the art schemes [16] and to a completely random approach [18]. Note that the introducing of such a central entity increases the control plane latency once the first connection is established with the network, which is acceptable for URLLC as long as the user plane latency, i.e. the latency experienced by each packet, is not affected.

The rest of the paper is structured as follows. In Section 2, we derive the loss probability under the contention-based approach when the positions of replicas are chosen randomly by the UE at each packet generation and when sequences are pre-allocated by the base station. Section 3 illustrates the performance of the proposed schemes and compare them to the state of the art. We draw conclusions in Section 4.

2. Performance models for contention-based allocation with replicas

2.1. System model

We consider a system with N UEs. Radio resources are allocated into the time/frequency domain. In particular, in the time domain, they are allocated every TTI. In 4G, a TTI lasts for 1 ms, while different TTI sizes are being defined for 5G. In the frequency

domain, instead, the total bandwidth is divided in sub-channels[1]. A combination of a TTI and a subchannel is called Resource Block (RB) and corresponds to the smallest radio resource unit that can be assigned to a UE for data transmission.

In order to satisfy reliability targets for URLLC, users are assigned a robust Modulation and Coding Scheme (MCS) that ensures a low Block Error Rate (BLER). For a size of an applicative packet of b bits, a spectral efficiency of the used MCS of η *bit/s/Hz*, a bandwidth per RB of ω and a TTI τ, the number of physical RBs, R, for transmitting an applicative packet is:

$$R = \lceil \frac{b}{\eta \tau \omega} \rceil. \tag{1}$$

For the ease of reading and without any loss of generality, we define a "resource allocation unit" equal to R RBs, so that each packet occupies 1 unit. Let M be the amount of resource allocation units per TTI; it is obtained by dividing the amount of available spectrum W by the available amount of spectral resources per unit:

$$M = \lfloor \frac{W}{R\omega} \rfloor. \tag{2}$$

Time is divided into cycles, each comprising the same amount of time-frequency resources. In each cycle, packet arrivals are sporadic and reserving dedicated resources for each user is clearly sub-optimal, as the number of users, N, may be very large and the probability that a user generates a packet during a cycle, p, may be low. Our proposal is to deal with this traffic in a contention-based manner, i.e. to reserve a pool of resources where users who have packets to transmit contend. Packets are thus subject to collisions, in addition to the losses introduced by the wireless channel. In order to increase the probability of success, each packet may be sent $\beta \geq 1$ times. We call these replicas.

Let the amount of resource units in the resource pool be equal to K, the resources for first transmissions are spanned over a number of TTIs equal to $\lceil \frac{K}{M} \rceil$. Let the delay constraint of the service be equal to T, the amount of resources allocated to retransmissions, K, has to verify the following constraint:

$$\lceil \frac{K}{M} \rceil \leq \frac{T}{\tau}. \tag{3}$$

Depending on the service and system parameters (latency constraint, number of users, amount of available spectrum), this constraint may be feasible or not. We suppose that constraint given in (3) is feasible. We will derive in the numerical application section the optimal value K^* for satisfying the reliability constraint.

2.2. Loss probability under the random allocation scheme

We now provide the loss probability for the contention-based scheme, when each UE chooses a new set of positions for its replicas at each cycle. Note that when multiple

[1] 4G subchannels are of 180 kHz, each composed of 12 consecutive and equally spaced Orthogonal Frequency-Division Multiplexing (OFDM) subcarriers. Different subcarrier spacings are defined for 5G, but our model is sufficiently generic to cover the different cases.

copies of a packet are sent, a *collision* occurs if all these copies collide with other transmissions, in which case we consider a *loss* has occurred. The resulting collision rate is measured from a predefined-user perspective, given that this user has data to transmit.

Although imperfections of the radio channel (e.g. fast fading) may lead to a replica being lost even without a collision, we neglect the impact of such radio errors in this section and the next one and focus on losses due to collisions. The terms loss probability and collision probability will thus be used equivalently in these sections. While this is a reasonable assumption for sufficiently robust MCS, we will show afterwards how these errors can be included in the developed models.

Let $e_r(N, K, \beta, p)$ denote the loss probability under the random contention-based approach. Recall that each user is active with probability p and selects β resource units randomly from the set of K resources that are available in the contention cycle.

Proposition 1. *The loss probability under the contention-based approach with replicas can be expressed as follows*

$$e_r(N, K, \beta, p) = 1 - \sum_{l=1}^{\beta} (-1)^{l+1} C_\beta^l \left((1-p) + p \frac{C_{K-l}^\beta}{C_K^\beta}) \right)^{N-1} \tag{4}$$

where C_k^n denotes the binomial coefficient, or the number of k-combinations among n.

Proof. Define $\mathcal{A}_i$ to be the event that the i-th resource is free, i.e. no (other) active user chooses this resource for its packet transmission. We would like to express the probability that one of the β resources is free, i.e. $\mathbb{P}\{\mathcal{A}_1 \cup \ldots \cup \mathcal{A}_\beta\}$. To this end, we determine the probability that a subset of l resources is free. Note that in a set containing β resources there are C_β^l subsets of size l. All l resources will be collision-free if all other users are either not transmitting or non of their β replicas fall in the l resources. For a given user, this happens with probability

$$1 - p + p \frac{C_{K-l}^\beta}{C_K^\beta}, \tag{5}$$

where p represents the probability that a user is active. Since there are $N-1$ other users, the probability that all l slots of this subset are collision-free is:

$$\mathbb{P}\{\mathcal{A}_1 \cap \ldots \cap \mathcal{A}_l\} = \left(1 - p + p \frac{C_{K-l}^\beta}{C_K^\beta} \right)^{N-1}. \tag{6}$$

Using the above, we conclude that

$$\mathbb{P}\{\mathcal{A}_1 \cup \ldots \cup \mathcal{A}_\beta\} = \sum_{l=1}^{\beta} (-1)^{l+1} C_\beta^l \, \mathbb{P}\{\mathcal{A}_1 \cap \ldots \cap \mathcal{A}_l\} =$$

$$\sum_{l=1}^{\beta} (-1)^{l+1} C_\beta^l \left(1 - p + p \frac{C_{K-l}^\beta}{C_K^\beta} \right)^{N-1}. \tag{7}$$

Leading to the loss probability (4), which concludes the proof. □

For small values of p, which interest us as we aim for very small loss probabilities in the case of relatively large populations, the loss probability is essentially due to the presence of a single other user occupying the same slots.

Proposition 2. *For small probabilities p, the loss probability for random slot assignments is asymptotically equivalent to the probability of another user being active and having the same slot assignment:*

$$e_r(N, K, \beta, p) \sim (N-1)p\frac{1}{C_K^\beta}. \tag{8}$$

Proof. First note that the right hand term in (8) is indeed equivalent asymptotically to the probability of another user being active and having the same slot assignment. These probabilities are independent. The probability another user is active is equivalent to $(N-1)$ times the probability a single user is active, p, when p tends to zero. The probability another user having the same sequence of slots is one over the number of different sequences, C_K^β.

Let us now show that the left hand term in (8) is indeed equivalent to this product. Equation (4) can be written:

$$e_r(N, K, \beta, p) = \sum_{l=0}^{\beta}(-1)^l C_\beta^l \left((1-p) + p\frac{C_{K-l}^\beta}{C_K^\beta}\right)^{N-1}. \tag{9}$$

A first order expansion around $p = 0$ gives:

$$\begin{aligned} e_r(N, K, \beta, p) &\sim \sum_{l=0}^{\beta}(-1)^l C_\beta^l \left((1-(N-1)p) + (N-1)p\frac{C_{K-l}^\beta}{C_K^\beta}\right) \\ &\sim (N-1)p\left(\sum_{l=0}^{\beta}(-1)^l C_\beta^l \frac{C_{K-l}^\beta}{C_K^\beta}\right), \end{aligned} \tag{10}$$

since

$$\begin{aligned} \sum_{l=0}^{\beta}(-1)^l C_\beta^l (1-(N-1)p) &= (1-(N-1)p)\sum_{l=0}^{\beta}(-1)^l C_\beta^l \\ &= (1-(N-1)p)(1-1)^\beta = 0. \end{aligned}$$

From (9) we see that the last multiplicative factor in (10) is in fact equal to $e_r(2, K, \beta, 1)$, the probability that a single other user, surely present, i.e. $p = 1$, collides with our given user in a system where that are only two users, $N = 2$. Finally note that $e_r(2, K, \beta, 1)$ is equal to the probability of a single user choosing the slot assignment of our given user, which is $1/C_K^\beta$.

(This proves the combinatorial formulae: $\sum_{l=0}^{\beta}(-1)^l C_\beta^l C_{K-l}^\beta = 1$.) □

Note that, in [19] the authors derive a different, but equivalent, expression for the loss probability for the random sequence selection case. They then use it to compare

its performance in the case sequences are chosen from a list so that any two sequences overlap in only one position. As this limits the number of possibilities they propose to chose additional sequences at random whenever the number of users exceeds the list size. As we show in Proposition 2 the probability of the random sequence being exactly equal to one of the active sequences has a strong (and negative) impact on the loss rate. A possible improvement to the scheme they propose would be to use our result from Proposition 3 in the following section, where the additional sequences are assigned at random but are restricted to being different of all the already assigned ones.

2.3. Pre-determined sequence allocation

We now move to the case where the replicas occupy the same positions at each cycle. Let the time/frequency resources at each cycle be numbered from 1 to K. We define the "sequence" for UE i as a vector v_i of length K, composed of 1's in the β places where replicas are placed and 0's elsewhere.

As the choice of these positions is done once, we suppose that the base station is aware of this choice, e.g. the UE informs the network about its choice when it registers to the network or the network allocates the UE its positions. This scheme can then ensure that users have distinct sequences, in which case there needs to be more than one other active user for a complete collision to occur.

Proposition 3. *The loss probability for pairwise distinct random sequences is computed by:*

$$e_d(N, K, \beta, p) = \sum_{k=2}^{N-1} C_{N-1}^{k} p^k (1-p)^{N-1-k} \frac{b_k}{d_k} \tag{11}$$

where d_k is the number of combinations of k sequences of β distinct slots chosen among K, such that the sequences are distinct from each other and from a given user's:

$$d_k = (C_K^\beta - 1) \dots (C_K^\beta - k) \tag{12}$$

and b_k is the number of combinations of k such sequences which produce a loss for that given user.

The values of the b_k's are the solution of the following system of equations:

$$\begin{cases} a_2 &= b_2 \\ &\dots \\ a_k &= b_2 c_{2k} + \dots + b_{k-1} c_{k-1k} + b_k \end{cases} \tag{13}$$

where a_k is the number of combinations of k sequences of β slots chosen among K, which are distinct from a given user's (but not from each other) and which produce a loss for that given user:

$$a_k = \sum_{l=0}^{\beta} (-1)^l C_\beta^l \left(C_{K-l}^\beta \right)^k - ((C_K^\beta)^k - (C_K^\beta - 1)^k) \tag{14}$$

and $c_{lk} = l^{k-l}$ is the number of possible l draws chosen among $k-l$ elements.

Proof. Let us start with the expression of a_k. It is the number of k sequences producing a loss for a given user's sequence minus the number of such k sequences where one of them is equal to the given user's. The probability for k sequences to produce a loss for a given user is $e_d(k+1, K, \beta, 1)$ so that the number of sequences is:

$$C_K^\beta \times e_d(k+1, K, \beta, 1) = \sum_{l=0}^{\beta} (-1)^l C_\beta^l \left(C_{K-l}^\beta \right)^k .$$

The number of k sequences where at least one sequence is equal to the given user's is the number of k sequences minus the number of sequences where none is equal to the given user's:

$$(C_K^\beta)^k - (C_K^\beta - 1)^k.$$

Subtracting this last expression from the previous one we obtain a_k in (14).

The expression for b_k is difficult to express, and instead in (13) we express the a_k's as a function of the b_l's for l smaller than k from which we may iteratively solve for b_k. To this end we decompose a_k as a function of the number of distinct sequences l producing a loss for the given user. The remaining $k - l$ sequences must be chosen among the l distinct sequences. There are $c_{lk} = l^{k-l}$ such choices, while there are b_l ways to chose l distinct sequences to produce a loss for the given user. We obtain finally:

$$\begin{aligned} a_k = &\quad b_2 c_{2k} + \ldots + b_{k-1} c_{k-1k} + b_k c_{kk} \\ = &\quad b_2 c_{2k} + \ldots + b_{k-1} c_{k-1k} + b_k \end{aligned}$$

as $c_{kk} = 1$.

The probability for k random and distinct sequences from each other and from a given user's, to produce a loss, is the ratio of b_k to d_k, the number of combinations of k random and distinct sequences from each other and from a given user's. To obtain the expression for d_k consider that there are $C_K^\beta - 1$ ways to choose a first sequence distinct from the given user's. There are $C_K^\beta - 2$ ways to choose a second sequence distinct from the two previous ones, and so on.

To obtain the loss probability for pairwise distinct random sequences (11) one conditions on the probability for k users, other than the given user, to transmit, $C_{N-1}^k p^k (1-p)^{N-1-k}$. Conditioned on this event the probability of loss is then $\frac{b_k}{d_k}$. At least two distinct sequences are necessary to produce a loss and there are at most $N-1$ other users in case of a total of N users. □

Proposition 4. *The asymptotic decay rate of the loss probability for pairwise distinct random sequences when p tends to zero is:*

$$p^2 C_{N-1}^2 \times \left(e_r(3, K, \beta, 1) - (1 - (1 - \frac{1}{C_K^\beta})^2) \right) \left(\frac{(C_K^\beta - 1)(C_K^\beta - 2)}{(C_K^\beta)^2} \right)^{-1} . \quad (15)$$

Proof. From (11) the first non null term in the development of $e_d(N, K, \beta, p)$ in terms of p is:

$$C_{N-1}^2 p^2 \frac{b_2}{d_2} = C_{N-1}^2 p^2 \times \frac{b_2}{(C_K^\beta)^2} \left(\frac{d_2}{(C_K^\beta)^2} \right)^{-1}. \tag{16}$$

This term will determine the asymptotic decay rate for small values of p when p decreases. The left and right multiplicative terms in (15) and (16) coincide. It remains to prove that:

$$\frac{b_2}{(C_K^\beta)^2} = e_r(3, K, \beta, 1) - (1 - (1 - \frac{1}{C_K^\beta})^2). \tag{17}$$

From (13), $b_2 = a_2$, and from (14):

$$\begin{aligned}
\frac{b_2}{(C_K^\beta)^2} &= \frac{a_2}{(C_K^\beta)^2} \\
&= \frac{\sum_{l=0}^{\beta} (-1)^l C_\beta^l \left(C_{K-l}^\beta \right)^2 - ((C_K^\beta)^2 - (C_K^\beta - 1)^2)}{(C_K^\beta)^2} \\
&= \sum_{l=0}^{\beta} (-1)^l C_\beta^l \left(\frac{C_{K-l}^\beta}{C_K^\beta} \right)^2 - (1 - (1 - \frac{1}{C_K^\beta})^2) \\
&= e_r(3, K, \beta, 1) - (1 - (1 - \frac{1}{C_K^\beta})^2).
\end{aligned}$$

□

Note as the asymptotic decay rate for distinct random sequences is a constant times p^2 instead of p as for the case of random sequences, we expect the loss probability to be smaller in the first case than in the second for small transmission probabilities p. This is indeed what is observed in the numerical results of section 3.2.

3. Numerical applications

3.1. Random sequences

Before moving to the performance evaluation using the analytical formula (4), we first perform some comparisons of this formula with respect to discrete event simulations. These simulations generate, for each active user, a random sequence of β positions where its replicas are placed and verify for the target user (numbered 1) if there is a replica free of collision. Furthermore, to reduce simulation times while obtaining small variances for small activation probabilities p, we use importance sampling. In a first step, simulations are performed a large number of times with n active users, with n ranging from 1 to N. For each n we obtain an estimate of the collision probability, $\widehat{e}_n$. Then in a second step, the resulting conditional collision probabilities, $\widehat{e}_n$, are used to obtain the collision probabilities for different activation probabilities p, by multiplying

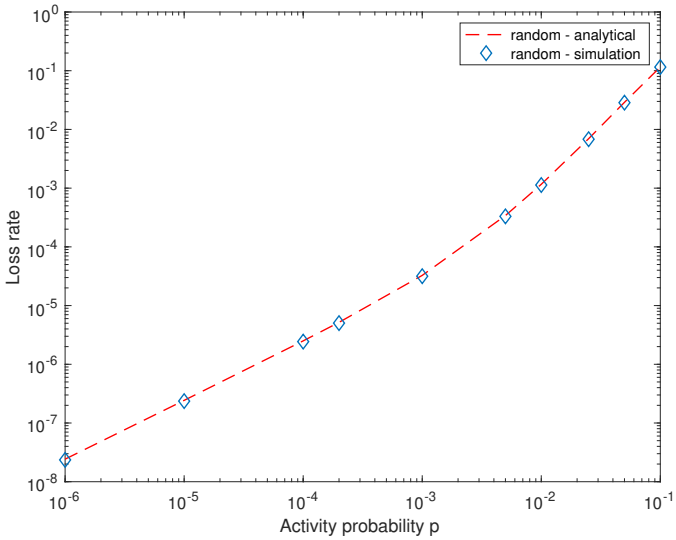

Figure 1: Comparison of analytical expression and simulation for the probability of collision in the random allocation case for $N = 50$, $K = 24$ and $\beta = 3$.

by the probability of having $n-1$ additional active users with respect to the one for which we are trying to obtain the collision rate:

$$\sum_{n=1}^{N} C_{N-1}^{n-1} p^{n-1} (1-p)^{N-1-n+1} \widehat{e}_n. \tag{18}$$

We plot in Figure 1 the packet loss probability, when varying the activity ratio p, using (4) and the discrete event simulator for $N = 50$, $K = 24$ and $\beta = 3$. As the analytical expression models without any approximation the system simulated, the figure confirms a perfect match for all values of p.

3.2. Distinct sequences

Figure 2 shows the loss probability when ensuring that sequences of users are distinct, compared to a baseline where the allocation is performed completely at random for the same parameters as in Figure 1. We observe as predicted that the decrease rate of the loss probability as p decreases, is much steeper than for completely random slot assignments. In order to illustrate the gain, if a target reliability corresponding to a loss probability of 10^{-6} is sought, the maximal activation probability p is equal to $4 * 10^{-5}$ and $4 * 10^{-4}$, for the random and pairwise distinct schemes, respectively (ten times larger load). The figure also compares the analytical expression (11) to discrete event simulations where, each time a user selects a sequence already assigned, its assignment is modified to another sequence. A perfect fit is observed.

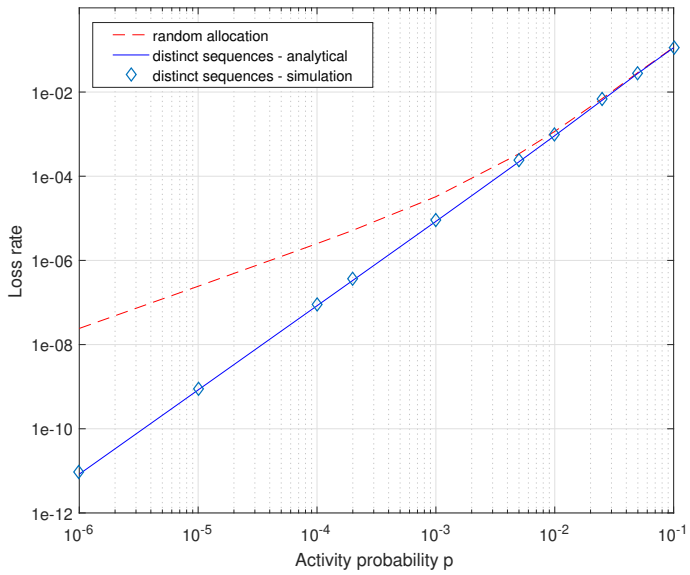

Figure 2: Probability of collisions in the case when sequences are distinct.

We then illustrate in figure 3 the impact of the varying number of users on the performance. The same trend is observed for all values of N, where the distinct sequence allocation clearly outperforms the random one for low activity rates, i.e. in the interesting URLLC region.

3.3. Comparison with the state of the art

For the sake of comparison, we call our random scheme RT for Random Transmissions of replicas. Define URT (for Unique Random Transmission) to be the baseline scheme where only one packet replica is sent. This scheme can then be seen as a special case of RT for $\beta = 1$. We also compare our scheme to the scheme called OT (for One Transmission per TTI) proposed in [16] that sends systematically one replica in each of the consecutive TTIs, where the resources used by each replica are randomly selected from the set of available RBs in each TTI. The performance model for the OT scheme can be found in [16] and an enhanced version for the OT model for low traffic loads can be found in [18].

In Figure 4, we illustrate the reliability performance achieved by URT, OT and RT, by plotting the variation of the collision probability with respect to p. We consider the case of $N = 30$ users and $K = 12$ resource units per cycle, with a cycle of 2 TTIs. In this case the OT scheme sends two replicas per cycle, as when $\beta = 2$ for RT. It can be seen that for $p < 10^{-2}$ and a number of packet replicas $\beta = 2$, 4 or 6, our proposed RT scheme is able to reach high reliability levels compared to the other schemes by producing lower collision probabilities. We can notice that the performance is initially enhanced when the number of replicas increases ($\beta = 2$, 4 or

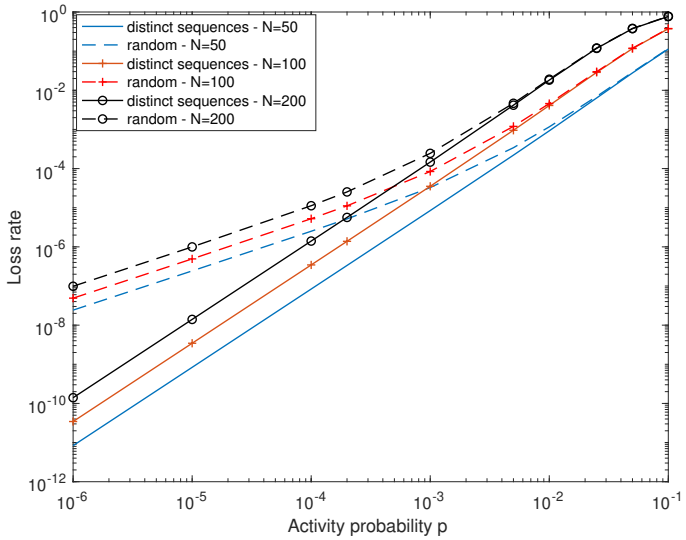

Figure 3: Impact of the number of users on the performance.

6) but degrades when the transmission probability, p, increases beyond a certain value which depends on β. There is then an inflexion point. However this degradation occurs for transmission probabilities, p, for which none of the schemes allow reaching the low collision probabilities required for URLLC services. Note that increasing the number of replicas leads eventually to a loss, as for the case where $\beta = K$, that corresponds to the case with only one resource and N contending users, with a loss rate of $1-(1-p)^{N-1}$

To close this paragraph, we compare the maximum load for which the target performance is achievable with the different schemes. If a collision probability of 10^{-5} is targeted, the baseline (URT) scheme can only support a activity probability of $p = 2\times10^{-6}$, OT supports a maximal activity of 7×10^{-6}, while RT can support an activity probability of almost 10^{-4} (50 times larger than the baseline) in the case $\beta = 4$.

3.4. Impact of radio errors

As mentioned earlier, URLLC users are generally assigned a robust MCS that ensures a low error rate. However, some packets will still be lost even without collisions due to radio imperfections. Let γ be the probability that a resource is subject to degraded radio condition so that a replica that is transmitted on it would be lost even without collision. We have the following result.

Proposition 5. *The loss probability integrating wireless errors can be expressed by:*

$$e_r(N,K,\beta,p,\gamma) = 1 - \sum_{l=1}^{\beta}(-1)^{l+1}C_\beta^l\left((1-p)+p\frac{C_{K-l}^{\beta}}{C_K^{\beta}}\right)^{N-1}(1-\gamma)^l. \qquad (19)$$

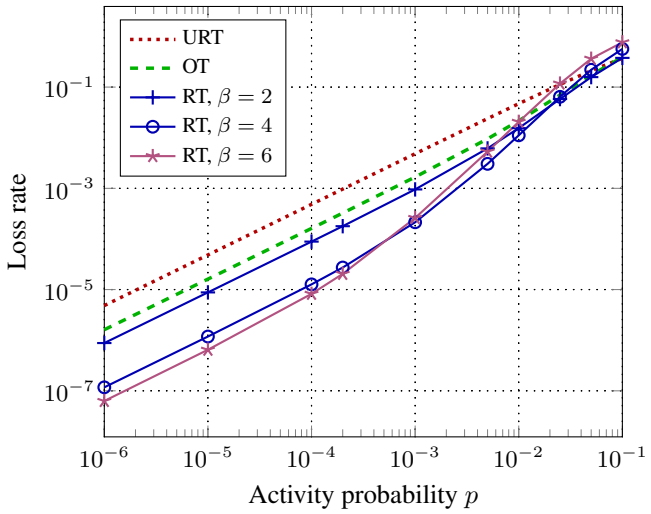

Figure 4: Collision probability vs p for URT, OT and RT, for $N = 30$ users sharing $K = 12$ resource units per cycle.

for the random scheme, and by:

$$e_d(N, K, \beta, p) = \sum_{k=2}^{N-1} C_{N-1}^{k} p^k (1-p)^{N-1-k} \frac{b'_k}{d_k} \tag{20}$$

for distinct sequence allocations, where the b_k's are the solution of the following system of equations:

$$\begin{cases} a'_2 &= b'_2 \\ &\dots \\ a'_k &= b'_2 c_{2k} + \ldots + b'_{k-1} c_{k-1k} + b'_k \end{cases} \tag{21}$$

where a'_k is the number of combinations of k sequences of β slots chosen among K, which are distinct from a given user's (but not from each other) and which produce a collision or a transmission loss for that given user:

$$a'_k = \sum_{l=0}^{\beta} (-1)^l C_\beta^l \left(C_{K-l}^{\beta} \right)^k (1-\gamma)^l - ((C_K^\beta)^k - (C_K^\beta - 1)^k) \tag{22}$$

and d_k and c_{lk} are defined as in Proposition 3.

Proof. We now define $\mathcal{A}_i$ to be the event that the i-th resource is free, i.e. no (other) active user chooses this resource for its packet transmissions and this resource is not subject to a radio error. These events (occupancy and error) are independent. As before, we determine the probability that a subset of l resources among the β resources

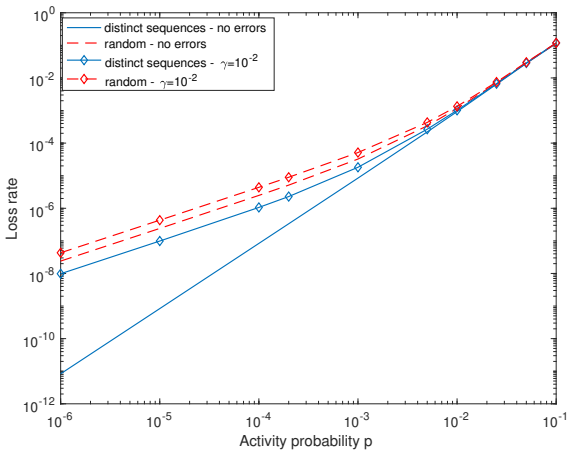

Figure 5: Impact of radio errors on the loss rates.

allocated to the target user is free. Since there are $N-1$ other users and errors are independent, the probability that all l slots of this subset are collision-free and error-free, in the random case, is:

$$\mathbb{P}\{\mathcal{A}_1 \cap \ldots \cap \mathcal{A}_l\} = \left((1-p) + p\frac{C_{K-l}^{\beta}}{C_K^{\beta}}\right)^{N-1}(1-\gamma)^l. \tag{23}$$

Which leads to the expression (19). We proceed similarly to calculate a'_k in (22) including collisions and radio losses. The expression for a'_k is obtained as in (23) where p is set to 1 and N to k. Then, proceeding as in Proposition 3, the b'_k are obtained from equations (21). □

We now illustrate how the radio errors impact the loss rate. Figure 5 shows the loss performance when introducing radio errors with probability $\gamma = 10^{-2}$, always with the same configuration (50 users, 24 reserved resources and 3 replicas per packet), for the distinct and random sequences case. The impact is large, especially for the scheme with distinct sequences as the radio errors reintroduce losses even when there is only one other active user. However, for the usual target loss probability of 10^{-5}, there is still a large advantage of the proposed scheme with respect to the random scheme, even if the radio error rate is as high as $\gamma = 10^{-2}$ (9 times larger acceptable activity factor). Interestingly, this example shows that multiple transmissions may allow reaching the target reliability even in mediocre radio conditions, without relying on the HARQ retransmissions mechanism, thus attaining the required reliability while still respecting the latency constraint.

3.5. Optimal retransmissions and resource allocation

Equation (4) gives the packet loss probability for a given number of replicas β and a given set of reserved resources K. These parameters, β and K, have to be chosen so

that the latency and reliability constraints are satisfied with the lowest possible resource reservation. Using the same notation as in section 2.1, the latency constraint can be expressed as:

$$\lceil \frac{K}{M} \rceil \leq \frac{T}{\tau} \tag{24}$$

where $M = \lfloor \frac{W}{R\omega} \rfloor$ is the amount of resource units per TTI. The number of replicas and the amount of resources are to be chosen so that K is minimized while satisfying latency constraint (24) and the reliability constraint:

$$e_r(N, K, \beta, p) \leq \Theta \tag{25}$$

Figure 6 shows the number of replicas needed for obtaining a reliability target of 10^{-6} loss rate, for different numbers of users N and different sizes of the reserved pool K, while keeping $p = 10^{-4}$. We can observe that the number of replicas increases when N increases, but also when resources become scarcer. Also, the reliability target becomes unfeasible when N is too large or K is too small.

We now move to the computation of the amount of resources to be reserved for ensuring both the target reliability and the target latency constraint of 1 ms (equation (24)). Figure 7 shows the minimal amount of resources to be reserved for achieving the targets, and the corresponding amount of spectrum to be reserved, for a system whose parameters are expressed in Table 1, and based on the number of replicas represented in Figure 6. It can be observed that the amount of required spectrum increases with the number of users, reaching up to 10 MHz for N=100. Note that, in comparison with a deterministic reservation schemes where dedicated resources are allocated to each user, as usually advocated for URLLC, the resource consumption is far less ($K = N$ in the deterministic case).

Table 1: System and service parameters

Applicative packet size (b)	100 bits
Modulation and Coding Scheme (MCS)	Polar Alamouti 2*2 4QAM
Subcarrier spacing	15 KHz
smallest time scheduling unit (τ)	0.125 ms
Acknowledgment time (t_a)	0.256 ms
Latency constraint	1 ms

4. Conclusions

In this paper we consider sporadic uplink transmissions for URLLC services. We combine grant-free contention-based transmissions with packet repetitions as a means to increase the reliability while respecting the latency budget. We explore contention-based schemes and develop an analytical model for the resulting collision probability. We validate this model through simulations and use it to design the transmissions strategies that allow meeting the URLLC requirements. In particular, we find that a strategy

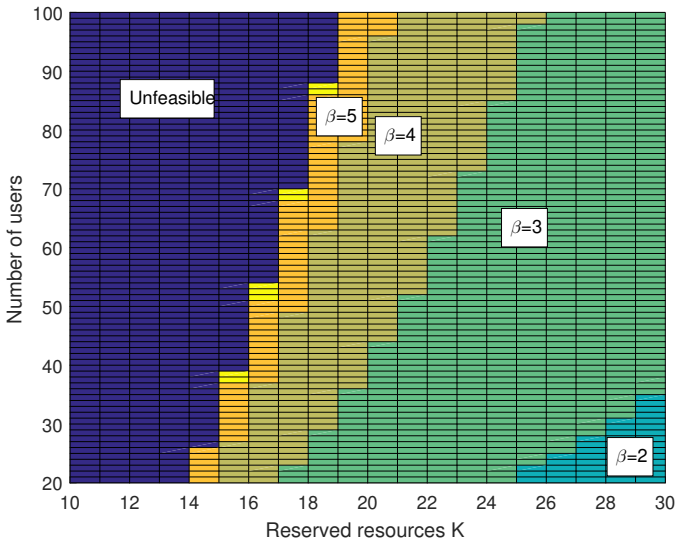

Figure 6: Number of needed replicas for the target reliability.

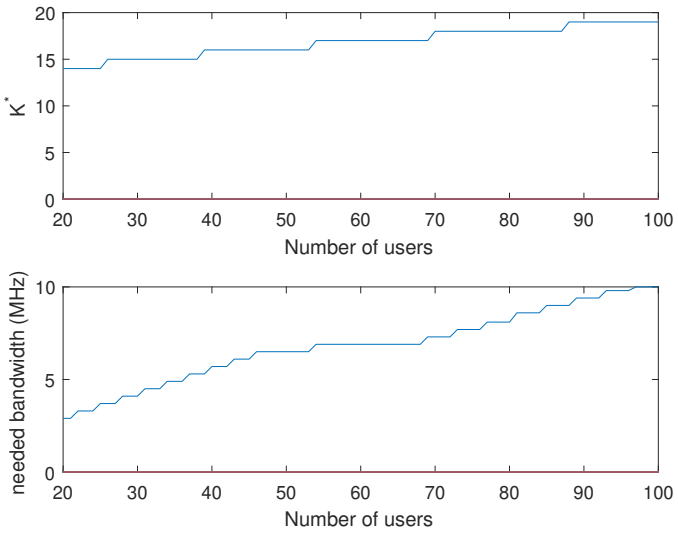

Figure 7: Reserved spectrum for the target reliability.

that allocates replicas randomly to available resources achieves better performance than a strategy that allocates one packet per TTI. We also showed that a centralized, offline, allocation of resources to users that ensures that a complete collision cannot happen when only two users are active enhances drastically the loss rate. An important result of this paper is to show how to achieve very stringent reliability targets with a proper dimensioning of the common resource pool, without having to perform hard resource reservation per user.

As of future work, we aim at studying the optimal pre-allocation of resources to users that minimizes the collision probability and to investigate the iterative decoding and collision resolution knowing this centralized allocation.

References

[1] P. Popovski. Ultra-reliable communication in 5G wireless systems. In *1st International Conference on 5G for Ubiquitous Connectivity*, pages 146–151, Nov 2014.

[2] 3GPP. Study on scenarios and requirements for next generation access technologies. 3GPP TR 38.913 v14.2.0, Tech. Rep., June 2017.

[3] Bernd Holfeld, Dennis Wieruch, Thomas Wirth, Lars Thiele, Shehzad Ali Ashraf, Jorg Huschke, Ismet Aktas, and Junaid Ansari. Wireless communication for factory automation: An opportunity for LTE and 5G systems. *IEEE Communications Magazine*, 54(6):36–43, 2016.

[4] 3GPP. Study on latency reduction techniques for LTE. 3GPP TR 36.881 v14.0.0, Tech. Rep., June 2016.

[5] K. I. Pedersen, S. R. Khosravirad, G. Berardinelli, and F. Frederiksen. Rethink hybrid automatic repeat request design for 5G: Five configurable enhancements. *IEEE Wireless Communications*, PP(99):2–8, 2017.

[6] Chih-Ping Li, Jing Jiang, Wanshi Chen, Tingfang Ji, and John Smee. 5G ultra-reliable and low-latency systems design. In *2017 European Conference on Networks and Communications (EuCNC)*, pages 1–5. IEEE, 2017.

[7] Guillermo Pocovi, Beatriz Soret, Klaus I Pedersen, and Preben Mogensen. MAC layer enhancements for ultra-reliable low-latency communications in cellular networks. In *2017 IEEE International Conference on Communications Workshops (ICC Workshops)*, pages 1005–1010, 2017.

[8] Shehzad Ashraf, Y-P Eric Wang, Sameh Eldessoki, Bernd Holfeld, Donald Parruca, Martin Serror, and James Gross. From radio design to system evaluations for ultra-reliable and low-latency communication. In *European Wireless 2017; 23th European Wireless Conference*, pages 1–8, 2017.

[9] Klaus I Pedersen, Guillermo Pocovi, and Jens Steiner. Preemptive scheduling of latency critical traffic and its impact on mobile broadband performance. In *2018 IEEE 87th Vehicular Technology Conference (VTC Spring)*, pages 1–6. IEEE, 2018.

[10] Ali A Esswie and Klaus I Pedersen. Opportunistic spatial preemptive scheduling for URLLC and eMBB coexistence in multi-user 5G networks. *IEEE Access*, 6:38451–38463, 2018.

[11] Yishu Han, Salah Eddine Elayoubi, Ana Galindo-Serrano, Vineeth S Varma, and Malek Messai. Periodic Radio Resource Allocation to Meet Latency and Reliability Requirements in 5G Networks. In *2018 IEEE 87th Vehicular Technology Conference (VTC Spring)*, pages 1–6, 2018.

[12] Arjun Anand and Gustavo de Veciana. Resource allocation and HARQ optimization for URLLC traffic in 5G wireless networks. *IEEE Journal on Selected Areas in Communications*, 36(11):2411–2421, 2018.

[13] Salah Eddine Elayoubi, Patrick Brown, Matha Deghel, and Ana Galindo-Serrano. Radio resource allocation and retransmission schemes for urllc over 5g networks. *IEEE Journal on Selected Areas in Communications*, 37(4):896–904, 2019.

[14] 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*, 55(2):70–78, February 2017.

[15] Stefania Sesia, Issam Toufik, and Matthew Baker. *LTE, The UMTS Long Term Evolution: From Theory to Practice, 2nd Edition*. Wiley Publishing, 2011.

[16] B. Singh, O. Tirkkonen, Z. Li, and M. A. Uusitalo. Contention-based access for ultra-reliable low latency uplink transmissions. *IEEE Wireless Communications Letters*, 7(2):182–185, April 2018.

[17] 3GPP. Physical layer procedures for data. 3GPP TR 38.214 v15.1.0, Tech. Rep., March 2018.

[18] Matha Deghel, Patrick Brown, Salah Eddine Elayoubi, and Ana Galindo-Serrano. Uplink contention-based transmission schemes for urllc services. In *Proceedings of the 12th EAI International Conference on Performance Evaluation Methodologies and Tools*, pages 87–94. ACM, 2019.

[19] Gino T. Peeters, R. Bocklandt, and Benny Van Houdt. Multiple access algorithms without feedback using combinatorial designs. *IEEE Transactions on Communications*, 57, 2009.