# Semi-distributed Contention-based Resource Allocation for Ultra Reliable Low Latency Communications

Patrick Brown, Salah Eddine Elayoubi

# Semi-distributed Contention-based Resource Allocation for Ultra Reliable Low Latency Communications

Patrick Brown
Orange Labs
905 Rue Albert Einstein
06560 Valbonne, France

Salah Eddine Elayoubi
L2S, UMR CNRS 8506, CentraleSupélec
3 rue Joliot Curie
91190 Gif-Sur-Yvette, France

***Abstract*—Ultra-Reliable Low Latency Communications (URLLC) and especially those related to the Industrial Internet of Things (IIoT) are characterized by a large number of users transmitting sporadically information to a central controller. We consider in this paper scenarios where transmitted packets have to be conveyed within a very short time so that it is not possible to make per-packet resource reservation, i.e. contention-based access is needed. Moreover, in case of loss, there is no room for waiting for acknowledgement before retransmissions so that blind replication is needed for reaching the ultra high reliability targets. Knowing the limited, but large, number of potential users in the system, we propose a semi-centralized resource allocations scheme where each user is pre-allocated positions for its replicas in case he has a packet to convey. We show, using coding theory, how to design sequences for users so that the number of collisions is minimized. We further exploit our pre-allocation scheme to develop a successive interference cancellation method where the base station tries to decode a packet based on the knowledge of the already decoded colliding packet. We show that the proposed schemes succeed to attain very low loss rates with low resource reservation.**

## I. INTRODUCTION

Achieving ultra reliability and low latency simultaneously is in general a very challenging task in wireless networks. Indeed, the classical way for recovering from errors is through retransmission after the reception of a Negative Acknowledgement (NACK), which inexorably increases delays. Meeting the stringent requirements of Industrial Internet of Things (IIoT) use cases (1 ms target covering the end-to-end path between the sensors/actuators and a central controller with a reliability of $99.99999\,\%$ [1]) calls then for new wireless communication schemes. First, quick access to radio resources is primordial and *grant-free* scheduling, instead of the classical grant-based scheduling approach of Long-Term Evolution (LTE), is needed as waiting for the grant penalizes the latency [2]. And second, blind packet replication is a most as there is no room for successive retransmissions as in classical Hybrid Automatic Repeat Request (HARQ). We differentiate here, as 3GPP does [3], between retransmission that happens after a NACK and blind replication that repeats a packet without waiting for receiver feedback.

We now turn to the design of the grant-free resource access. Two flavours exist: a reservation scheme where several resources are reserved for each user at each TTI, and a contention-based scheme where a pool of resources is reserved for URLLC users and they contend to it. While the former scheme is suitable for deterministic traffic patterns (e.g. 1 packet per user per Transmission Time Interval (TTI)), it is unfeasible for a large number of users with sporadic packet arrivals. Our focus in this paper is on IIoT use cases with a very large number of users with sporadic packet arrivals and a very stringent reliability requirement, so that contention-based replication scheme is adopted. Such an approach will result in collisions between some of the replicated packets, which may impact the reliability level if the system is not well designed. Hence, it is important to determine the resource allocation policy an active user will follow to send the replicas of each of its packets. The problem corresponds to the choice of radio/frequency resources where replicas are placed, and this is equivalent to finding sequences of 0's and 1's for each user. The easiest and most distributed contention-based scheme with replicas corresponds to a random choice by the user of the placement of his 1's, as advocated in [4] and [5], but this has two drawbacks. First, it is not very unlikely (relative to the very high reliability requirements) that two users may choose exactly the same sequences, leading to a possible loss even when only two users are active. And, second, the base station does not know the placements of replicas and cannot use advanced decoding schemes for resolving collisions. Another limitation of these schemes is that they consider use cases where the latency constraint allows that replicas can be sent over multiple TTIs, which is not suitable for many IIoT applications. We consider replication only in the frequency dimension, so that latency does not exceed 2 TTIs. We propose in this paper a scheme that overcomes the previous limitations and is suitable for use cases with very stringent delay and reliability requirements as follows:

- We propose a semi-distributed scheme where the positions where each user places its replicas are predetermined by a central entity and sent to users. Note that this scheme does not impact the 'user place' latency, as the codes are sent only once, and not upon the generation of every packet. The 'control-plane' latency, i.e. the initial latency when the user is attached to the network, may however be impacted, which is not in general an issue (an as long as 10 ms initial control plane latency is acceptable [1]). This central allocation allows overcoming the random allocation drawbacks (different codes per user can be allocated and the base station knows where replicas can be found).

- In order to further enhance the reliability, we explore the usage of new constant-weight codes with minimum pairwise Hamming code distance. A packet loss can occur only if the number of active users is larger than a value that depends on the weight of the code and the minimum pairwise distance. These constant-weight codes have been studied in the literature for several applications, such as frequency reuse in cellular networks [6], and we propose in the Appendix simple algorithms that allow deriving such codes for different code lengths and weights.
- Once these sequences are allocated to users, we further apply signal processing at the base station for enhancing the reliability, by successively cancelling interference for decoding the packets, exploiting the knowledge of already decoded packets for the decoding of packets belonging to colliding users.

The rest of the paper is structured as follows. In Section II , we derive the loss probability under the contention-based approach when the positions of replicas are chosen randomly by the network and pre-allocated to users and illustrate the system performance. Section III considers the problem of resource allocation as the design of codes with minimum pairwise Hamming distance. It also presents closed form expression for reliability performance that fits very well with simulation results. Section IV proposes a packet decoding scheme that, combined with the proposed contention-based access, achieves the reliability targets of URLLC with low resource consumption. We draw conclusions in Section V. Finally, the appendix provides a simple algorithm for generating sequences intersecting at most in one slot.

## II. Contention-based allocation with pre-allocated sequences: Basic scheme

### A. Resource allocation

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 the frequency domain, the total bandwidth is divided in sub-channels and 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 each TTI, packet arrivals are sporadic and reserving resources for each user is clearly under 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. In contrast with other works [4][5] where multiple replicas (of the same packet) are sent in different TTIs, we consider scenarios where there is no room for using several TTIs without breaking the latency target. The resource pool is thus of size equal to the number of packets that can be served jointly in one TTI. Let $K$ be the number of such "resource units"[1].

### B. Pre-determined sequence allocation

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. We will show in section IV how these errors can be included in the developed models.

Let the time/frequency resources at each TTI 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 $\beta$ places where he places replicas 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 network allocates positions to the UE. This scheme has three advantages over the random one where the UE selects new positions at each cycle. First, it can be ensured that users have distinct sequences, i.e. there is a need for more than one other active user so that there is complete collision. Second, the sequences can be chosen carefully so that the probability of collisions is reduced, as will be seen in the next section. Third, as the base station knows the positions of replicas, it can use Successive Interference Cancellation (SIC) in case of collisions.

### C. Loss probability for pairwise distinct sequences

We start by investigating the first advantage of centrally designed sequences, that is the possibility of allocating distinct sequences to all users.

**Proposition 1.** *The loss probability for pairwise distinct random sequences is computed by:*

$$\begin{aligned}
&e_d(N,K,\beta,p) = 1- \\
&\sum_{l=1}^{\beta}(-1)^{l+1}C_\beta^l\left((1-p)+p\frac{C_{K-l}^\beta}{C_K^\beta})\right)^{N-1} - \\
&(1-(1-p\frac{1}{C_K^\beta})^{N-1}) \qquad (1)
\end{aligned}$$

[1] $K$ can be calculated as follows. For a size of an application packet of $b$ bits, a spectral efficiency of the used Modulation and Coding Scheme (MCS) of $\eta$ *bit/s/Hz*, a bandwidth per RB of $\omega$ and a TTI $\tau$, the number of physical RBs, $R$, for transmitting an application packet is: $R=\lceil\frac{b}{\eta\tau\omega}\rceil$. $K$ is obtained by dividing the amount of available spectrum $W$ by the available amount of spectral resources per unit $K=\lfloor\frac{W}{R\omega}\rfloor$

*Proof.* We start the proof by deriving the probability of loss when users choose their resources at random, without any coordination, and then use this expression to derive the distinct case. 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 $\beta$ 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 $\beta$ resources there are $C_\beta^l$ subsets of size $l$. All $l$ resources will be collision-free if all other users are either not transmitting or non of their $\beta$ replicas fall in the $l$ resources. For a given user, this happens with probability

$$1 - p + p\frac{C_{K-l}^\beta}{C_K^\beta}, \tag{2}$$

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:

$$\mathbb{P}\{\mathcal{A}_1 \cap \ldots \cap \mathcal{A}_l\} = \left(1 - p + p\frac{C_{K-l}^\beta}{C_K^\beta}\right)^{N-1}. \tag{3}$$

Using the above, we conclude that

$$\mathbb{P}\{\mathcal{A}_1 \cup \ldots \cup \mathcal{A}_\beta\} = \sum_{l=1}^{\beta}(-1)^{l+1}C_\beta^l\,\mathbb{P}\{\mathcal{A}_1 \cap \ldots \cap \mathcal{A}_l\} =$$
$$\sum_{l=1}^{\beta}(-1)^{l+1}C_\beta^l\left(1 - p + p\frac{C_{K-l}^\beta}{C_K^\beta}\right)^{N-1}. \tag{4}$$

Leading to the loss probability in the random case:

$$e_r(N, K, \beta, p) = 1 - \sum_{l=1}^{\beta}(-1)^{l+1}C_\beta^l\left((1-p) + p\frac{C_{K-l}^\beta}{C_K^\beta})\right)^{N-1} \tag{5}$$

Consider now the distinct sequences case. A given user transmits in an interval and the probability of at least one other user being active and having the same random slot sequence is the complement of the event that non of the other users are both active and have the same sequence, which is equal to:

$$(1 - p\frac{1}{C_K^\beta})^{N-1},$$

where $1/C_K^\beta$ is the probability of having been assigned the same sequence as the given user. Subtracting the probability of another user being active with same slot sequence from the probability of collision with random slot assignments we obtain (1).

□

Figure 1 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 $N = 50$, $K = 24$ and $\beta = 3$. 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 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 (1) to numerical simulations. For this aim, we developed a simulator that pre-assigns a sequence to each of the $N$ users. At each iteration, active users are selected at random following the activation probability $p$ and each of them generates replicas at the pre-determined $\beta$ positions. A packet is lost if all of its replicas collide. A perfect fit is observed between the simulation results and the theoretical analysis.

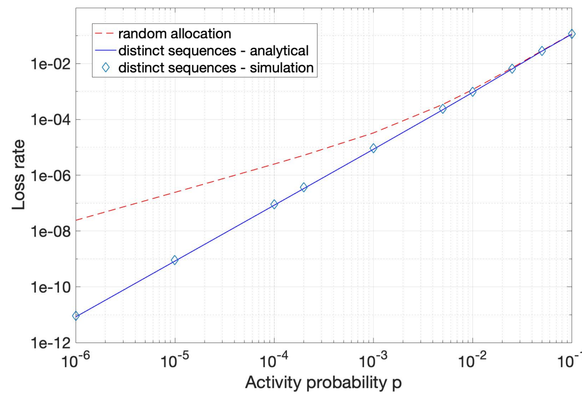


Figure 1: Probability of collisions in the case when sequences are distinct.

## III. Optimized Sequence Design

### A. General considerations on sequence design

First let us present some general expressions concerning the possible sequences, $v_i$, assigned to users. Let us assume there are $N$ users. Let $\sigma_k$ be the number of users, slot $k$ is assigned to. Note that we must have:

$$\Sigma_{k=1}^K \sigma_k = N\beta.$$

So the mean slot occupancy $\sigma$ depends only on the population size, the number of resources and the number of copies per transmission:

$$\sigma = \frac{\Sigma_{k=1}^K \sigma_k}{K} = \frac{N\beta}{K}, \tag{6}$$

Let $M$ be the matrix composed of the $N$ column vectors $v_j$. Then $1_K^t M = (\beta, .., \beta)_N$ and $M 1_N = (\sigma_1, ..., \sigma_K)^t$.

Note that $v_k^t v_j$ is the number of slots used both by users $j$ and $k$. In the case $j = k$ we have a fixed value $v_j^t v_j = \beta$. Then element $m_{kj}$ of the $N \times N$ matrix $M^t M$ gives the number of common slots used by $j$ and $k$. Note that the sum of all elements of $M^t M$ gives the sum of common slot positions between all couples of users (including identical users). It is given by:

$$1_N^t M^t M 1_N = (\sigma_1, ..., \sigma_K)(\sigma_1, ..., \sigma_K)^t = \Sigma_{k=1}^K \sigma_k^2.$$

Note that this sum represent a significant characteristic of the system. Define $\sigma_u$ as the average slot occupancy as seen by users in contrast to the average system slot occupancy taken over all slots, $\sigma$.

**Proposition 2.** *The average slot occupancy, $\sigma_u$, seen by users is:*

$$\sigma_u = \frac{\Sigma_{k=1}^{K} \sigma_k^2 - N\beta}{N\beta}. \quad (7)$$

*Proof.* A user sees $\sigma_k - 1$ other users for each slot $k$ among the $\beta$ slots assigned to him. Let $i_1, .., i_\beta$ denote the $\beta$ slots used by user $i$. Taking the average over all users:

$$\sigma_u = \frac{1}{N} \sum_{\text{users } i} \frac{(\sigma_{i_1} - 1) + ... + (\sigma_{i_\beta} - 1)}{\beta}.$$

Summing over the slots in the system instead of over the users in the expression of the average:

$$\sigma_u = \frac{1}{N\beta} \sum_{k=1}^{K} \sigma_k(\sigma_k - 1) = \frac{1}{N\beta} \left( \sum_{k=1}^{K} \sigma_k^2 - \sum_{k=1}^{K} \sigma_k \right),$$

as a slot with occupancy $\sigma_k$ is assigned to $\sigma_k$ different users. □

We derive the following property of slot assignments.

**Proposition 3.** *The average slot occupancy; $\sigma_u$, seen by users is minimized when all slots have the same occupancy $\sigma_k = \sigma = \frac{N\beta}{K}$ assuming this is an integer value. In which case:*

$$\sigma_u = \sigma - 1. \quad (8)$$

*Proof.* If all slots have the same occupancy (6), the average slot occupancy seen by users is $\sigma - 1$. This expresses the minimum attainable average user slot occupancy, $\sigma_u$, since it minimizes the slot occupancy variance: $\Sigma_{k=1}^{K} \sigma_k^2/K - (\Sigma_{k=1}^{K} \sigma_k/K)^2$. □

However this minimum alone is not necessarily an indication of a good assignment. Notice that it may be attained when many users are assigned the same set of slots.

As an example consider a system with $\beta = 2$ copies, $K = 6$ slots and $N = 9$ users such that three users occupy the first two slots, three users occupy the next two slots and the last three users occupy the last two slots. As any user shares its slots with only two other users, its collision probability is in this case:

$$1 - (1 - p)^2.$$

However if each user is assigned one of the following sequences of slots: $\{1, 2\}$, $\{2, 3\}$, $\{3, 4\}$, $\{4, 5\}$, $\{5, 6\}$, $\{1, 6\}$, $\{1, 3\}$, $\{2, 5\}$, $\{4, 6\}$, then the slot occupancy is still 3, but two other users must active in order to occupy the slots of any given user. The resulting collision probability is:

$$(1 - (1 - p)^2)^2 < 1 - (1 - p)^2.$$

To measure the quality of a set of predefined sequences it is thus necessary to derive the collision probabilities related to it which we proceed to do in the next section.

### B. Loss probability for carefully designed sequences

To minimize the collision probability we focus on sequences such that sequences assigned to two different users have at most one slot in common. However this constraint limits the number of possible sequences for a given number of slots $K$. So it may be interesting to allow sequences to have more than one slot in common. We derive next the collision probabilities for both cases.

We will be considering the collision rate for a given user. Without loss of generality we may assume slots are numbered such that the given user is assigned slots 1 to $\beta$.

*1) Sequences intersecting at most in one slot:* We start by the interesting case of sequences intersecting at most in one slot. This corresponds to generating codes of length $K$, weight $\beta$ and minimal pairwise Hamming distance of $2\beta - 1$. In addition to the good properties of these codes in terms of low collision rates, as will be shown next, there is a rich literature on how to generate them and on the maximal number of such sequences for given $K$ and $\beta$, as in [7] and [6].

In this case users occupying two slots assigned to our given user say $j$, are different users. Their probability of arrival are independent. As a result the probabilities of different slots, assigned to $j$, being occupied by other users are independent. There are $\sigma_i - 1$ other users sharing a slot $i$ with our given user. Thus the probability this slot is used by other users is:

$$1 - (1 - p)^{\sigma_i - 1}.$$

The loss probability is the probability all $\beta$ slots are used by other users. Since these probabilities are independent, this probability is given by the following proposition.

**Proposition 4.** *The loss probability of a user assigned slots with occupancy $\sigma_1$,...,$\sigma_\beta$, with a sequence intersecting only in one slot with other user sequences, is given by:*

$$(1 - (1 - p)^{\sigma_1 - 1})...(1 - (1 - p)^{\sigma_\beta - 1}). \quad (9)$$

The performance objective is that the loss probability (9), for each user, should be smaller or equal the target loss probability $\Theta$. As a consequence of this expression, the asymptotic behavior when $p$ tends to zero for sequence assignments with only one slot in common is:

$$(\sigma_1 - 1)...(\sigma_\beta - 1)p^\beta, \quad (10)$$

which drops to zero as $p$ to the power of $\beta$. Thus setting the number of replicas $\beta$ is a means of reaching very low loss rates in the context of small arrival rates with many UEs.

Let $S(N, K, \beta)$ designate a set of $N$ $\beta$-length slot sequences for a system with $K$ available slots. When designing a set of slot sequences, $S(N, K, \beta)$, we would like to know if it is preferable to have balanced slot occupancies, $\sigma_i$, around their average value (6), or if it is preferable to have more dispersed occupancy values.

From the expression (7) of the average occupancy seen by users, $\sigma_u$, we conclude that increasing the variance of the system slot occupancy, increases the average occupancy experienced by users, which has a negative effect on their collision rate.

However from the following proposition we observe that a UE allowed to choose between two sequences, one with constant occupancy, the second with dispersed values, should choose the second to minimize its collision rate.

**Proposition 5.** *Consider two slot sequences with occupancies $s = (\sigma_1, ..., \sigma_\beta)$, such that $\sigma_1 = ... = \sigma_\beta$, and $s' = (\sigma'_1, ..., \sigma'_\beta)$ such that both have identical average occupancies, $\sigma$. Then the collision rate of $s'$ will be inferior to that of $s$:*

$$(1-(1-p)^{\sigma_1 - 1})...(1-(1-p)^{\sigma_\beta - 1}) \geq (1-(1-p)^{\sigma'_1 - 1})...(1-(1-p)^{\sigma'_\beta - 1}). \qquad (11)$$

*Proof.* These inequalities result from the application of Jensen's inequality to the function $f(x) = log(1-(1-p)^x)$ which can be shown to be concave, as its second derivative is negative. This implies:

$$f(\sigma_1) + ... + f(\sigma_\beta) = \beta f(\frac{\sigma_1 + ... + \sigma_\beta}{\beta}) = \beta f(\frac{\sigma'_1 + ... + \sigma'_\beta}{\beta}) \geq f(\sigma'_1) + ... + f(\sigma'_\beta),$$

where the first equality results from the fact the occupancies in $s$ are constant, the second equality results from the fact both average occupancies are equal, and the last inequality results from Jensen's inequality for concave functions. Finally as the $log$ function is increasing, we conclude the same inequalities apply to the collision probabilities (9). □

We now consider the theoretical maximum offered load, $q_{max} = Np/K$, the resources may sustain for a target collision rate of $\Theta$ in the case of pre-assigned sequences intersecting at most in one slot. We denote $x = N/K$ the ratio of over-allocation. We first derive the maximum offered load when all slot occupancies are equal to the mean slot occupancy $\sigma = N\beta/K = x\beta$. We will next compare with the maximum allowable load in a specific case of unbalanced occupancies. In the case of equal slot occupancies we must solve for $q_{max}$ in the following equation:

$$(1-(1-q_{max}/x)^{x\beta-1})^\beta = \Theta, \qquad (12)$$

which derives from (9). The maximum offered load to the system in this case is a function of the over-allocation ratio, the number of copies and the maximum allowed collision rate.

The values of $q_{max}$ plotted on Figure 2, in the case $\Theta = 10^{-6}$ as certain URLLC services require, show that high over-allocation ratios seem attainable while still having acceptable offered loads.

We compare these values with the maximum achievable offered load in the unbalanced case, when occupancies depart from the average value, $\sigma$. We consider the case where occupancies can take two values. The first $k$ system slot occupancies are equal to a *large* value, $\gamma > \sigma$ while the last

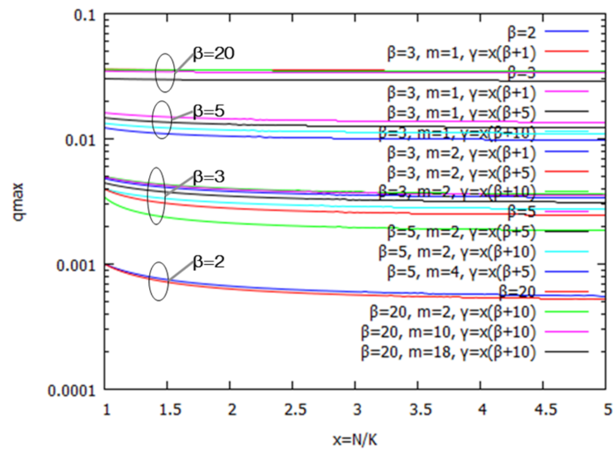


Figure 2: Maximum offered load, $q_{max} = Np/K$, for collision rate $\Theta = 10^{-6}$, for pre-assigned sequences intersecting at most in one slot, when slots are evenly ($\beta$) or unevenly occupied ($\beta, m, \gamma$), as a function of the slot over-allocation ratio, $x = N/K$.

$K - k$ system slot occupancies are all equal to a *small* value $\alpha < \gamma$:

$$\sigma_1 + ... + \sigma_K = N * \beta,$$
$$k\gamma + (K-k) * \alpha = N * \beta.$$

On the user side we assume each user is assigned $m$ slots with large occupancies and $\beta - m$ slots with small occupancies. Equating the sum of occupancies on slots and on users:

$$k\gamma = mN,$$
$$(K-k) * \alpha = N * (\beta - m).$$

Possible values for $m$ are $m = 1, ..., \beta$, while $\gamma$ must take values larger than the system average $x\beta$. The collision probability for this system is:

$$(1-(1-q/x)^{\gamma-1})^m (1-(1-q/x)^{\alpha-1})^{\beta-m}.$$

The values of $q_{max}$ for the dispersed case are compared to average occupancy case in Figure 2 for a collision rate $\Theta = 10^{-6}$. We express $\gamma$ as $\gamma = x(\beta + \delta) = \sigma + x\delta$. So we are plotting the maximum allowed offered traffic $q_{max}$, as a function of parameters $\beta$ (number of replicas), $x$ (slot over-allocation ratio), $m$ (less utilized slots), and $\delta$ (slot dispersion parameter).

We see from Figure 2 that, among the configurations analyzed, the lowest collision rates are obtained by choosing occupancies equal to the average value which is in accordance with the fact unbalanced slot occupancies increase the average occupancy seen by users as shown in Proposition 3. Note also that, unless the occupancies are very unbalanced, dispersion has little effect on the performance.

*2) General case:* We now turn to the more general case where sequences do not intersect necessarily in only one slot. We consider without loss of generality, a given user with $\beta$ slots numbered 1 to $\beta$. Define $S_\beta$ to be the set $\{1, .., \beta\}$ and $S^k_\beta$ to be the subsets of $S_\beta$ of size $k$. For $\{i_1, ..., i_k\} \in S^k_\beta$, let $\#(i_1...i_k)$ be the number of other users which have been assigned any slot in $\{i_1, ..., i_k\}$. We have the following result.

**Proposition 6.** *The loss probability in the general case, for a user with slots* $1$ *to* $\beta$ *is calculated by:*

$$1 - \sum_{k=1}^{\beta}(-1)^{k+1} \sum_{i_1...i_k \in S_\beta^k} (1-p)^{\#(i_1...i_k)}. \quad (13)$$

*Proof.* Recall $A_i$ is the event: slot $i$ is not occupied by another user. The probability "at least one of these slots is not occupied by another user" is:

$$Pr(A_1 \bigcup ... \bigcup A_\beta) = \sum_{k=1}^{\beta}(-1)^{k+1} \sum_{i_1...i_k \in S_\beta^k} Pr(A_{i_1} \bigcap ... \bigcap A_{i_k}).$$

The event all slots $i_1, ..., i_k$ are not occupied is equal to the event that other users are either not active or that these slots are not assigned to them:

$$\begin{aligned} Pr(A_{i_1} \bigcap ... \bigcap A_{i_k}) &= \prod_{j=1}^{N-1}(1 - p * \mathbf{1}_{\{j \text{ uses } i_1 \text{ or } ... \text{ or } i_k\}}) \\ &= \prod_{j=1}^{N-1}(1-p)^{\mathbf{1}_{\{j \text{ uses } i_1 \text{ or } ... \text{ or } i_k\}}} \\ &= (1-p)^{\#(i_1...i_k)} \end{aligned}$$

□

Note that equation (13) reduces to (9) when the given user has only one slot in common with other users. Without any further simplification (13) requires to calculate the number of users for the $2^\beta$ subsets of $S_\beta$.

However a probabilistic argument shows that (13) will behave asymptotically as $p^d$, when $p$ tends to zero, where $d$ is the minimum number of other users required to cover all slots $S_\beta$. The probability of collision due to the collision of all slots with a higher number of users tends to zero with a higher power of $p$ and is thus negligible compared to the term $p^d$. This gives an indication on how to better design sets of sequences, where sequences intersect on more than one position. As an example consider a user having two slots in common with two other users respectively in positions $\{i_1, i_2\}$ and $\{j_1, j_2\}$. Let $c$ be the number of different slots in $\{i_1, i_2\} \bigcup \{j_1, j_2\}$ and $c'$ be how many of these two users are required to cover $\{i_1, i_2\} \bigcup \{j_1, j_2\}$. Then $d = \beta - c + c'$. Three cases may occur:

1) $\{i_1, i_2\} = \{j_1, j_2\}$ then $d = \beta - 1$,
2) $\{i_1, i_2\}$ and $\{j_1, j_2\}$ intersect in one position: then $d = \beta - 1$,
3) $\{i_1, i_2\}$ and $\{j_1, j_2\}$ don't intersect: then $d = \beta - 2$,

the last case giving the worst performance.

*3) Achieved performance:* We plot in Figure 3 the loss rate for the low collision sequence design. An additional gain, above the one obtained from using distinct sequences, is obtained. For a target reliability of $10^{-6}$, the maximum $p$ is now equal to $2 * 10^{-3}$, i.e. a load 5 times larger than the case of pairwise distinct (but random) sequences.

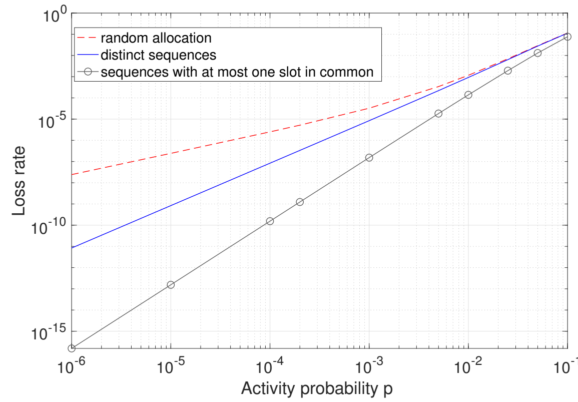

Figure 3: Low collision sequences and their impact on the loss rate ($N = 50$, $\beta = 3$, $K = 24$).

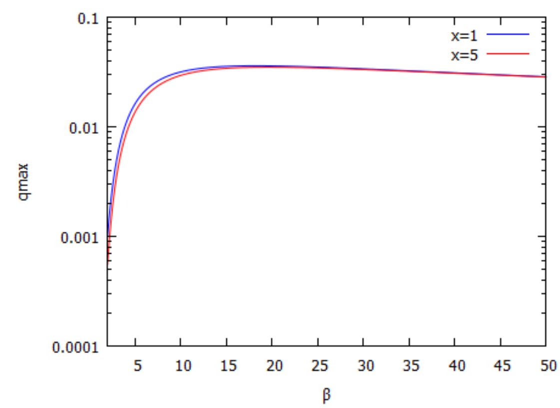

Figure 4: Impact of the number of replicas.

To attain these performance levels we must be able to produce sequences intersecting in one slot for a large range of values of $\beta$, $K$ and $N$. We propose a systematic method to attain this objective which we detail in the appendix.

We now turn to the choice of the number of replicas $\beta$. We are interest in the possible gain obtained from multiple transmissions and in how effectively we can use the radio resources while still offering the stringent performance required by the URLLC services. To this end we plot (12), the theoretical collision probability for pre-allocated slot sequences when occupancies are equal to the system average, $\sigma$, in the case of two over-allocation factors, $x = N/K = 1$ and $x = 5$ for $\Theta = 10^{-6}$. Figure 4 shows that the maximal supported offered traffic increases when the number of replicas increases, reaches a maximum and then decreases. Note that $q_{max}$ is not very sensitive to the exact value of $\beta$ around the optimum value. This maximum is reached around $\beta = 9$, but the majority of the gain is captured with $\beta = 5$, achieving a gain of a factor 10 in maximum offered traffic with respect to the transmission of a single copy, $\beta = 1$.

## IV. SUCESSIVE INTERFERENCE CANCELLATION FOR PRE-DEFINED SEQUENCES

In the previous sections, we computed the probability of packet loss as the probability that all the replicas of this packet have collided with replicas belonging to other users. However, iterative schemes may resolve collisions and reduce the loss rate. There are several such schemes where the packets with largest Signal to Interference and Noise (SINR) ratio are decoded first, and then subtracted from other signals in order to decode them [8]. Note that such schemes need that the base station knows where the replicas of all users are situated, i.e. that the sequences of all users be pre-determined, as we advocate in this paper.

Recently, an iterative decoding scheme that is specific to URLLC has been developed in [9], where two replicas are sent for each user, one on a dedicated resource and the other on a shared one. Our scheme goes beyond this specific scheme to cover sequences with general design. On the other hand, authors in [10] considered a system similar to ours, where users are equipped with a single antenna and a base station with $M > 1$ antennas and derived the power allocation that ensures a high reliability with a low energy consumption while transmitting on the same time-frequency resources. Our proposed scheme can build on top of this physical layer scheme and does not need necessarily to rely on iterative decoding when resolving collisions.

### A. Integrating losses due to radio fluctuations

As we aim in this section at exploiting physical layer techniques for enhanced performance, we cannot neglect the presence of wireless imperfections that may lead to packet losses even without collisions. We then show first how to take into account the radio imperfections in the loss probability. Note that, in order to satisfy reliability targets for URLLC, users are generally assigned a robust MCS that ensures a low Block Error Rate (BLER). However, some packets will be lost with a packet error rate that depends on the chosen MCS. Let $\delta$ 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 7.** *The loss probability integrating wireless errors can be expressed by:*

$$e_c(N,K,\beta,p,\delta) = 1- \sum_{l=1}^{\beta}(-1)^{l+1}C_\beta^l\left((1-p)+p\frac{C_{K-l}^{\beta}}{C_K^{\beta}})\right)^{N-1}(1-\delta)^l. \quad (14)$$

*for the random allocation case, by:*

$$1-\sum_{k=1}^{\beta}(-1)^k(1-\delta)^k\sum_{i_1...i_k}(1-p)^{\#(i_1...i_k)}. \quad (15)$$

*(where $\#(i_1...i_k)$ is the number of other users sharing slots $i_1...i_k$) for general pre-allocated sequences, and by:*

$$(1-(1-\delta)(1-p)^{\sigma_1-1})...(1-(1-\delta)(1-p)^{\sigma_\beta-1}), \quad (16)$$

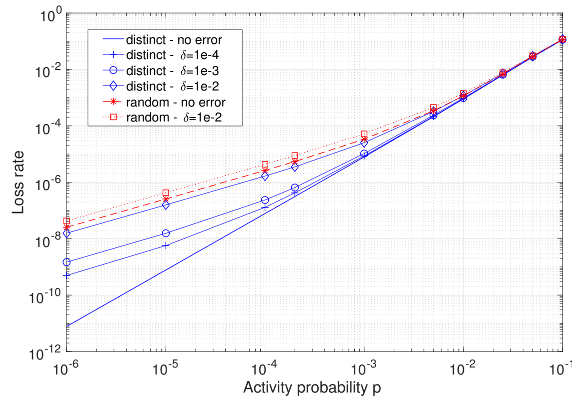


Figure 5: Impact of radio errors on the loss rates.

*for pre-allocated sequences intersecting in one slot.*

*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 $\beta$ resources 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:

$$\mathbb{P}\{\mathcal{A}_1\cap\ldots\cap\mathcal{A}_l\} = \left(1-p+p\frac{C_{K-l}^{\beta}}{C_K^{\beta}}\right)^{N-1}(1-\delta)^l. \quad (17)$$

Which leads to the expression (14). The same reasoning leads to the expression (15) in the pre-allocated case.

For (16) each factor is the probability correct decoding could not be achieved for the corresponding slot. This is the complementary event to a correct decoding, which takes place if the slot was not occupied by other users and if the slot was correctly decoded. The probability of this event is $(1-p)^{\sigma_k-1}(1-\delta)$. ☐

We now illustrate how the radio errors impact the loss rate for both random and pre-determined sequences. Figure 5 shows the loss performance when introducing radio errors, always with the same configuration (50 users, 24 reserved resources and 3 replicas per packet). We can observe that, for very low activity radios, wireless errors have a significant impact on the performance for distinct sequences, while they doe not have much influence on the random scheme. This is because wireless errors re-introduce losses when only one other user is active even if its chosen sequence is different from the active user. However, for the usual target reliability of $10^{-6}$, there is no significant impact, as long as the used MCS is sufficiently robust for ensuring an error rate around $\delta = 10^{-3}$. This calls for the usage of robust MCS, as advocated for URLLC.

Before moving to the iterative decoding scheme, we note that the presence of radio errors largely disadvantages the exclusive reservation scheme (not shown on the figure) where a resource is individually reserved for each user, unless the MCS is very robust (i.e. achieves the target reliability of $10^{-6}$ with one replica, which is very costly in terms of spectral efficiency). Otherwise, at least two resources have to be reserved for each packet ($K = 2N$).

### B. Successive interference cancellation performance

We do not aim at developing sophisticated decoding schemes (this may be the subject of future research), but to illustrate how simple decoding schemes, which take advantage of multiple transmissions, can further enhance the performance. Although sophisticated SIC schemes may decode packets even when replicas of all users are subject to collisions, we limit ourselves to SIC that is performed only when a replica collides with a packet that has been already decoded. Consider for instance a packet belonging to user $i$ that has been transmitted on resource $l$ without any collision. The same user $i$ has also transmitted a replica on resource $k$, that collides with a packet belonging to user $j$. Once the packet belonging to user $i$ is decoded on resource $l$, it can be subtracted from the signal received at resource $k$ in order to decode the packet of user $j$. An example of such iterative scheme exploiting the collision-free packets has been proposed in [9], where a Minimum Mean Square Error SIC (MMSE-SIC) decoder has been proposed, exploiting the analogy with a Multiple-Input-Multiple-Output (MIMO) where each User Equipment (UE) corresponds to a single transmit antenna, and each time-frequency slot is served by a different virtual receive antenna. Note that there is always a residual interference so that the collision resolution probability $q$ (defined as the probability that a packet in collision with a replica of an already decoded packet can be successfully decoded) is less than 1. We determine this probability using link level simulations.

We show in Figure 6 the probability of loss with SIC, taking as reference the low-collision design case without SIC (the best scheme in Figure 3). If the decoding scheme is perfect (probability of resolving a collision with an already decoded packet $q = 1$), the maximal allowed $p$ for a target reliability of $\Theta = 10^{-6}$ is equal to $10^{-2}$, i.e. a factor of 5 compared to a case without SIC. However, this gain reduces when the imperfection of the decoding is taken into account. For instance, our link level simulations indicate, for an MCS of 4QAM with a Polar Alamouti 2*2 MIMO scheme and a SINR of 15 dB a radio error probability of $\delta = 10^{-3}$, and a probability of successful collision resolution of $q = 0.9$. This leads to a maximum allowed activity probability of $p = 8 * 10^{-3}$ for the target reliability, i.e. a gain of 4. Further research studies on decoding schemes could achieve larger gains, but these are out of the scope of the current paper.

## V. Conclusion

In this paper, we developed a framework for radio resource allocation for URLLC traffic in 5G networks. We designed a contention-based scheme, where several replicas of each packet are systematically placed at different positions of the resource grid in order to increase the probability of success, despite possible collisions. The positions of replicas are a priori determined by a central entity, ensuring that users do not choose exactly the same positions of all their replicas and allowing a low-collision design. We model the problem as the design of sequences of zeroes and ones with large pairwise Hamming distance. We derived analytical expressions for the loss rate with sequences with low collisions and showed that large capacity gains can be achieved. We finally proposed an iterative decoding scheme that exploits the knowledge of the positions of replicas of different users in order to resolve collisions.

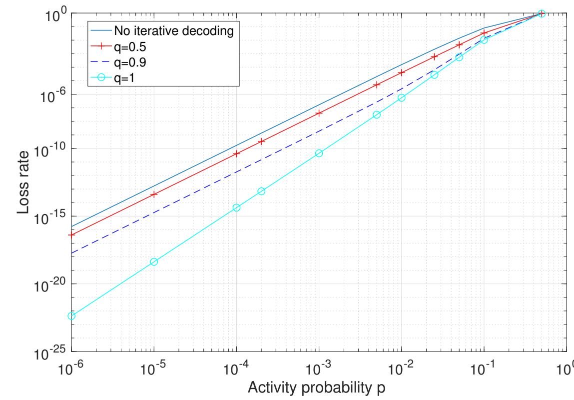


Figure 6: Impact of iterative decoding.

## Appendix

The contention based scheme we propose assumes we can define sets of sequences intersecting only in one slot for arbitrary size resources $K$ and that the system can scale for arbitrary size populations $N$.

We show in the following a systematic way to obtain a set of such sequences. The set of sequences obtained also shows that by scaling the resources by a factor $r$, we can scale the population by a factor $r^2$. It is thus possible to increase the over allocation factor $N/K$ and we have seen in Figures 2 and 4 this results in acceptable performance (when keeping equivalent offered load). Thus offering a solution for services to large populations with small traffic.

Denote $a\%b$, the value $a$ modulo $b$ for two integers $a$ and $b$.

**Proposition 8.** *Let $r \geq \beta - 1$ be a prime number. It is then possible to define $r^2$ sequences of $\beta$ slots on a set of $r\beta$ slots, such that the $(xr + y)$th sequence, where $x = 0, ..., r - 1$ and $y = 0, ..., r - 1$, is positioned on slots $(s_1, ..., s_\beta)$ such that:*

- $s_1 = x + 1$,
- $s_2 = y + r + 1$,
- $s_k = (x + y(k - 2))\%r + (k - 1)r + 1$, *for* $3 \leq k \leq \beta$.

*In addition the resulting average slot occupancy is $\sigma = r$.*

*Proof.* If such a set of sequences exists, its average slot occupancy is indeed $\sigma = r^2\beta/r\beta = r$, from (6).

This defines $r^2$ sequences (we show next to be distinct) as $(xr + y)$ takes $r^2$ different values.

Slot positions effectively range from 1 to $r\beta$ as can be seen from the range of values of $x$ and $y$.

We will be using the following property of the modulo operator:

(a) If $ac\%r = bc\%r$, $1 < c < r$, and $0 \leq a, b < r$, then $a = b$.

(We have $c\%r \neq 0\%r$ so $a\%r = b\%r$. Since $0 \leq a, b < r$ we must have $a = b$.)

Let us show that if two sequences $s$ and $s'$, respectively indexed by $xr + y$ and $x'r + y'$, have at least two slots $k \neq l$ in common then necessarily $x = x'$ and $y = y'$, so they are identical.

First note that $s_k = s'_l$ is possible only if $k = l$, since $s_k$ can only be placed on the subset of slots $(k-1)\beta + 1$ to $k\beta$. As a consequence, if two slots of $s$ and $s'$ are identical it must be that $s_k = s'_k$ and $s_l = s'_l$. Assume this is the case. Both sequences being interchangeable, there are four cases to be considered:

- $k = 1$ and $l = 2$: evidently $x = x'$ and $y = y'$,
- $k = 1$ and $l \geq 3$: then $x = x'$ and $(y - y')(l-2)\%r = 0$. So $y = y'$ due to (a),
- $k = 2$ and $l \geq 3$: then $y = y'$ so $y(l-2) = y'(l-2)$. Then we must have $x = x'$,
- $l > k \geq 3$: then $s_l - s_k = y(k-l)\%r = s'l - s'k = y'(k-l)\%r$ so $y = y'$ due to (a). Then $s_k = s'_k$ implies $x = x'$.

□

We next show how denser sets of sequences may be defined on a larger number of slots by combining smaller sets of sequences.

**Proposition 9.** *Let $\overline{S}$ be a set of $N_1$ sequences of $\beta$ slots intersecting only in one slot defined on a set of $K_1$ slots. It is then possible, for any prime number $r \geq \beta - 1$, to generate a set, $\overline{\Sigma}$, of $r^2 N_1$ sequences of $\beta$ slots over the set of $rK_1$ slots.*

*The $z(xr + y)$th sequence, $\Sigma$, of $\overline{\Sigma}$, is obtained from $z$th sequence, $S$, of $\overline{S}$ for $z = 1, ..., N_1$, and from the $(xr + y)$th sequence $s$ from $\overline{s}$ (obtained from Proposition 8) for $x = 0, ..., r-1$, and $y = 0, ..., r-1$, and has its $k$th slot positioned on:*

$$\Sigma_k = (S_k - 1)r + (s_k - (k-1)r - 1) + 1.$$

*Proof.* Since $S_k - 1$ takes all values between 0 and $K_1$, and since $(s_k - (k-1)r - 1)$ takes all values between 0 and $r - 1$, we conclude that $\Sigma_k$ takes all values between 1 and $rK_1$. So the resulting sequence covers the $rK_1$ slots.

Let us prove that two resulting sequences $\Sigma$ and $\Sigma'$ cannot intersect on two different slots without being identical. Suppose $\Sigma$ is the $z(xr + y)$th sequence of $\overline{\Sigma}$ and $\Sigma'$ is the $z'(x'r + y')$th sequence of $\overline{\Sigma}$.

Assume $\Sigma_k = \Sigma'_k$ and $\Sigma_l = \Sigma'_l$. If $z = z'$, then $s_k = s'_k$ and $s_l = s'_l$, which implies $x = x'$ and $y = y'$ from Proposition 8. So in this case both sequences $\Sigma$ and $\Sigma'$ correspond to the same $z(xr + y)$th sequence. If $z \neq z'$ then $S$ and $S'$ intersect only on one slot, say $l$. Then we cannot have $\Sigma_k = \Sigma'_k$ since they are positioned respectively on slots $((S_k - 1)r + 1), ..., S_k r$ and $((S'_k - 1)r + 1), ..., S'_k r$ which do not intersect since $S_k \neq S'_k$.

□

The set $\overline{S}$ may be obtained by hand on small values of $K_1$ or generated from a prime number greater or equal to $\beta - 1$ as in Proposition 8.

## References


[1] 3GPP, *Service requirements for next generation new services and markets*, 3 2018. V16.3.0.

[2] S. Sesia, I. Toufik, and M. Baker, *LTE, The UMTS Long Term Evolution: From Theory to Practice, 2nd Edition*. Wiley Publishing, 2011.

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

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

[5] M. Deghel, P. Brown, S. E. Elayoubi, and A. Galindo-Serrano, "Uplink contention-based transmission schemes for urllc services," in *Proceedings of the 12th EAI International Conference on Performance Evaluation Methodologies and Tools*, pp. 87–94, ACM, 2019.

[6] D. H. Smith, L. A. Hughes, and S. Perkins, "A new table of constant weight codes of length greater than 28," *the electronic journal of combinatorics*, vol. 13, no. 1, p. 2, 2006.

[7] L. B. S. Brouwer, Andries E. and N. I. A. Sloane, "A new table of constant weight codes," *IEEE Trans Inform Theory*, 1990.

[8] N. I. Miridakis and D. D. Vergados, "A survey on the successive interference cancellation performance for single-antenna and multiple-antenna ofdm systems," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 1, pp. 312–335, 2012.

[9] R. Kotaba, C. N. Manchón, T. Balercia, and P. Popovski, "Uplink transmissions in URLLC systems with shared diversity resources,"

[10] O. L. A. López, H. Alves, and M. Latva-aho, "Joint power control and rate allocation enabling ultra-reliability and energy efficiency in simo wireless networks," *IEEE Transactions on Communications*, 2019.