Multi-tenancy and URLLC on unlicensed spectrum

Ayat Zaki-Hindi, Salah-Eddine Elayoubi, Tijani Chahed

Multi-tenancy and URLLC on unlicensed spectrum: performance and design

Ayat Zaki-Hindi[a,c], Salah-Eddine Elayoubi[b], Tijani Chahed[c]

[a] *Orange Labs, 44 Avenue de la Republique, 92320, Chatillon, France*
[b] *CentraleSupelec, 3 Rue Joliot Curie, 91190 Gif-sur-Yvette, France*
[c] *Institut Polytechnique de Paris; Telecom SudParis; UMR CNRS SAMOVAR, 19 Place Marguerite Perey, Palaiseau, France*

Abstract

We study in this paper the transport of Ultra-Reliable Low-Latency Communication (URLLC) in a scenario of Industry 4.0 where transmission in uplink is possible in both licensed and unlicensed spectrum. We propose a transmission policy where the packet first attempts transmission in unlicensed spectrum during a time budget smaller than the delay constraint, if the transmission fails then the packet uses the remaining time budget to attempt transmission in licensed spectrum. The goal of using this policy is to minimize the cost of licensed bandwidth required for URLLC. We first consider the case of one tenant managing the industrial area and quantify the needed licensed bandwidth to attain reliability target within the target delay. We optimize the system by choosing the transmission policy which minimizes this cost. We then study the case of multiple tenants present in the same area and sharing unlicensed spectrum. Each tenant tries to minimize its cost of licensed bandwidth by utilizing unlicensed resources, which may result in the tragedy of the commons like situation. We formulate the problem using a game-theoretic approach to model the non-cooperative multi-tenant scenario. We model the medium access of unlicensed system for this case and derive the strategies that minimize individual cost functions. We then prove the existence of pure Nash equilibria analytically and identify them numerically. Finally, we quantify the so-called price of anarchy, i.e., ratio of the utility yielded by the competitive setting to the outcome of a cooperative scenario.
Keywords: URLLC, unlicensed spectrum, medium access, game theory

1. Introduction

The use of unlicensed spectrum for mobile communications is possible since 3GPP release 13, when Licensed-Assisted Access (LAA) LTE was first proposed in the downlink only, to be followed by the enhanced LAA (eLAA) for uplink and downlink in release 14 and the de-facto Multefire standard [1]. Afterwards, 3GPP worked on the definition of 5G New Radio (NR) which includes several unlicensed bands, illustrating the importance of unlicensed spectrum for 5G [2].

However, a main drawback for unlicensed spectrum is that Quality of Service (QoS) is not guaranteed because of the existence of other technologies in the same bands. This limitation is especially critical for Ultra-Reliable Low Latency Communications (URLLC) which transport critical information with stringent latency and reliability requirements, on the order of 1 to 5 ms end-to-end and 99.999%, respectively [3]. Nonetheless, unlicensed spectrum is being discussed for some URLLC services [4], notably for some smart factory use cases that are to be deployed in industrial areas where the environment can be controlled, e.g., by reducing the sources of outside interference. This may be true in environments managed by a single operator, but not in multi-tenant environments where several verticals manage plants in direct proximity. These co-existing networks operating in unlicensed spectrum create interference that degrades the QoS and compromises the value-add of unlicensed spectrum for verticals.

In this paper, we study the use of unlicensed spectrum for URLLC services and advocate its combined usage along with licensed spectrum to ensure the stringent latency and reliability targets. In particular, a generated packet attempts first transmission in unlicensed spectrum during a time budget shorter than the delay constraint, if it does not succeed then it is redirected to the licensed 5G spectrum. We show that this scheme drastically decreases the need for licensed spectrum resources compared to a classical licensed-only system by choosing the optimal transmission policy. In a multi-tenant environment, each of the verticals will want to maximize its economic gain from using unlicensed spectrum, which will increase the overall system interference and decrease the gain of the unlicensed spectrum for all tenants, leading to a tragedy of the commons like situation. We model this situation as a non-cooperative game between tenants, where each tenant strategy consists in using the unlicensed resources more or less aggressively so as to minimize its need for the expensive licensed ones.

The literature on URLLC is becoming rich. There has been early works which study the transport of URLLC over LTE, for instance [5], but the majority of the works deal with URLLC on 5G's licensed NR, considering grant-free fast uplink access, where neither issuing a scheduling request nor waiting for a scheduling grant are required [6]. This approach is often associated with the blind replication of packets, where the packet is sent several times within the delay budget without waiting for negative acknowledgement (NACK) so as to increase reliability [7, 8]. However, only few papers consider the use of unlicensed spectrum, such as [4], mainly because of the existence of other technologies such as WiFi on the same unlicensed bands, which decreases the reliability of the system.

As of works on modelling the medium access of unlicensed spectrum under delay and reliability constraints, such as our present one, unlicensed medium access using Listen Before Talk (LBT) has been widely studied for Wi-Fi systems, with the most cited model for Wi-Fi medium access being the one proposed by Bianchi in [9] based on discrete-time Markov chains. However, most studies focused on average measures such as average delay and throughput. Other variations of the Bianchi model were proposed and evaluate different performance metrics, as in [10] where the authors derived the probability generating function of the chain to obtain the delay distribution. Other methods proposed other probabilistic formulations to calculate the reliability for a given delay budget as in [11]. In this work, we prefer using Markov chains because of their simplicity of analysis.

The remainder of this paper is organised as follows: section 2 describes the system in one-tenant case, models the medium access in unlicensed and licensed spectrum and quantifies the performance of the system in terms of reliability within a delay budget. Section 3 formulates the problem of multi-tenant environment as a non-cooperative game, analyzes the performance in this case and proves the existence of pure Nash equilibria. We validate and illustrate our analysis numerically in section 5. Section 6 eventually concludes the paper.

2. URLLC transport in one-tenant environment

We consider a confined smart-factory context managed by one tenant having access to both unlicensed and licensed spectrum. The latter is more expensive and scarce compared to the former. In the sequel, we denote by 5G-U and 5G-L the parts of the system which use unlicensed and licensed resources, respectively.

We propose a scenario where the tenant deploys 5G-U Access Point(s) (AP) in its premises connected to a central controller, and uses a 5G-L Base Station (BS) of a mobile network operator covering the factory as a relay for the packets back to the controller. The number of APs may vary depending on the area and density of machines in the factory. We denote the transmitting machine by station and the number of stations by N, and assume that all stations are equipped with both 5G transmission systems. URLLC packets share the same latency and reliability requirements, denoted by T and R, respectively. Figure 1 illustrates the considered scenario.

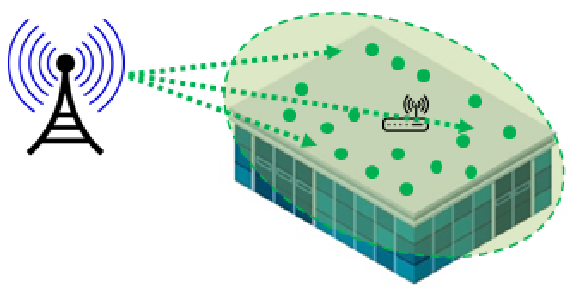

Figure 1: Scenario of industrial one-tenant environment

The proposed transmission mechanism is that when a packet is generated, it is first transmitted through 5G-U during a time budget $T_U < T$. If it is successfully transmitted then the process stops, if not, the packet is switched to 5G-L and is transmitted within the remaining time budget $T_L = T - T_U$. With this method, we decrease the load on 5G-L and hence the amount of licensed bandwidth BW that the tenant has to buy from the operator. Our aim is to study the effect of T_U on the performance of the system and choose the best value which minimizes the cost of licensed bandwidth.

2.1. Medium access model in unlicensed spectrum

In 5G-U, channel occupancy is managed by sensing the medium before transmission according to a random backoff procedure, called Listen-Before-Talk (LBT). When a packet is generated, it is associated to a backoff counter chosen randomly from the integer set $\{0, ..., W_0 - 1\}$ where W_0 is the maximum contention window size. Then the station senses the medium during one time slot (T_s), which is the smallest period required to sense the medium. If the medium is sensed idle then the backoff counter is decremented by one, else it is halted. This process is repeated until the counter reaches zero, in which case the packet is sent without sensing. A positive or negative acknowledgment ACK/NACK is expected within a given time, its absence is considered as a NACK. If the transmission is successful then the process ends here, else it is repeated for a number of attempts m, called stages. In our case, the number of stages is limited by the delay constraint T_U.

We denote the number of time slots needed for transmitting a URLLC packet by x, it comprises the time of packet transmission until the reception of ACK/NACK (or its absence). Since no collision avoidance mechanism is considered in our case, the duration of a collision is equal to a successful transmission. Assuming a perfect channel, stations sense x consecutive busy slots every time the medium is sensed busy.

We note that we deploy LBT cat3 with fixed contention window size in every stage instead of LBT cat4 deployed in most Wi-Fi-like systems which adapts the contention window according to collisions. This makes LBT cat3 more suitable to delay-constrained applications.

We note that following analysis can be easily extended to LBT cat4 or any scenario which uses LBT, e.g., LTE-LAA, by taking the changing contention window size into account in the model.

2.1.1. Stage-based modified Bianchi model model

We base our model on the one proposed by Bianchi [12] and modify it to suit our context, as illustrated in Figure 2, where we add three novel states: Start, Success and Failure; the latter refers to unsuccessful transmission after m attempts. This Markov chain is transient and describes the lifecycle of a packet from the moment it enters the system until it is either successfully transmitted or handed to the licensed band.

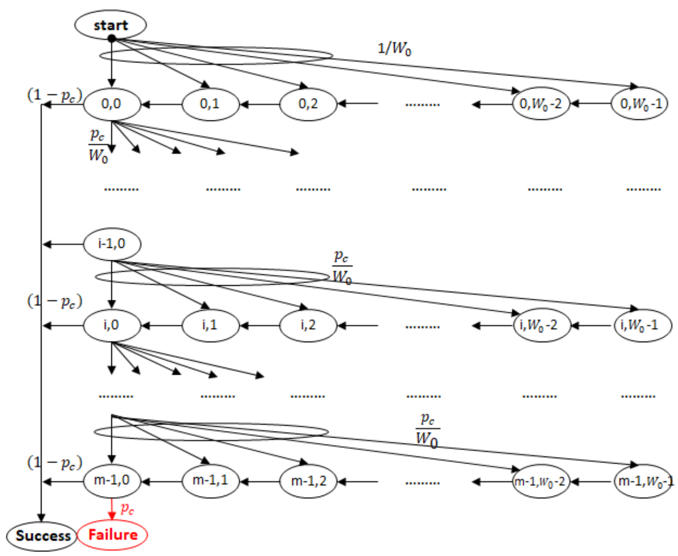

Figure 2: Modified Bianchi model for LBT cat3

Several assumptions are made to allow us use this model, such as perfect channel, i.e., packet-loss happens only when two or more transmissions coincide at the same time. We also consider a full-connected network, i.e., hidden-node problem does not exist. Moreover, N is finite and q is constant.

Each state of the 2-dimensional Markov chain is composed of two stochastic processes $\{s(t), b(t)\}$, representing the stage and the backoff counter at time t, respectively. The self-looping arrow represents the halting of the counter when the medium is sensed busy. We denote by p_c the probability of collision seen by the transmitting station in one slot, in other words p_c is the probability that at least one of the other $N-1$ stations is transmitting during the current time slot. p_c is considered constant and independent of time t and the backoff process.

The sporadic nature of URLLC implies that the stations are not saturated, we denote by q the probability of having a URLLC packet to transmit. The interval between two consecutive packet arrivals for the same station is assumed to be larger than T_U and so q is small enough to consider that the packets are not enqueued. We are first interested by the probability of packet loss, i.e., knowing that a packet has been generated, what is the probability that it reaches the "Failure" state. In order to compute this failure probability, we have to compute p_c which depends on the probability of users being active. For this sake, we have to compute the steady state probabilities of the states, or equivalently the proportion of time spent in each state. We then consider the non-transient Markov chain of figure 3 where the states "Start", "Failure" and "Success" are merged into one state where the user is inactive, which is adapted to the case of unsaturated sources. Each time slot, this slot loops on itself with probability $1-q$, and generates a packet with probability q.

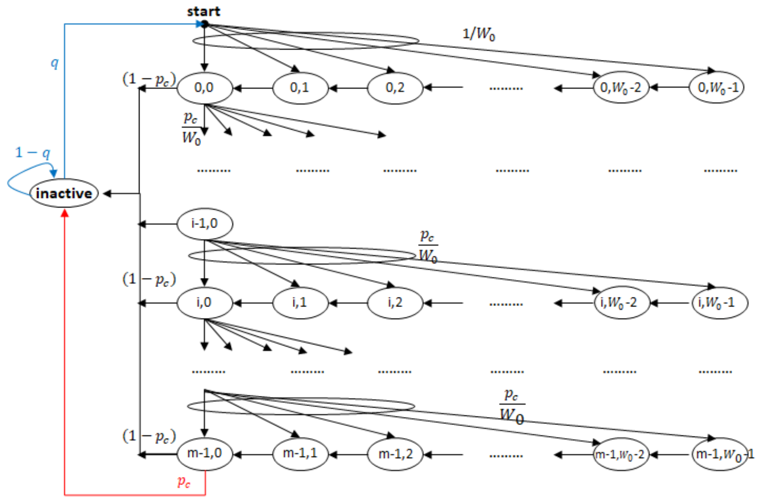

Figure 3: Complete model including the inactivity state

Denoting the probability of being in states $\{s(t), b(t)\} = \{i, j\}$ by $\Pi_{i,j}$ and in the inactive state by Π_{in}, we write the balance equations as follows:

$$\Pi_{0,W_0-1} = \frac{\Pi_{in} q}{W_0}$$

$$\Pi_{0,W_0-j} = jq\frac{\Pi_{in}}{W_0}, \quad 2 \leq j \leq W_0$$

$$\begin{aligned} \Pi_{i,W_0-j} &= j\frac{p_c}{W_0}\Pi_{i-1,0} \\ &= j\frac{p_c^i}{W_0}q\Pi_{in}, \qquad \begin{matrix} 1 \leq i \leq m-1 \\ 1 \leq j \leq W_0 \end{matrix} \end{aligned}$$

We can now compute Π_{in} by putting the sum of all state probabilities equal to 1 (the normalization condition), obtaining:

$$\Pi_{in} = [1 + q\frac{1-p_c^m}{1-p_c}\frac{W_0-1}{2})]^{-1}$$

Let τ be the probability of transmitting when $CW = 0$, computed by:

$$\tau = \sum_{i=0}^{m-1} \Pi_{i,0} = q\Pi_{in}\frac{1-p_c^m}{1-p_c} \tag{1}$$

We come to the computation of p_c, which corresponds to the probability of having at least one transmission from the other $N-1$ stations:

$$p_c = 1 - (1-\tau)^{N-1} \tag{2}$$

Equations (1) and (10) yield a fixed point of p_c, which we compute numerically. Having p_c, we can quantify the reliability of the system by evaluating the hitting probability of state Failure starting from state Start, which can be calculated directly from the balance equations by setting $q\Pi_{in} = 1$, where the reliability is $\Pi_{Success} = 1 - \Pi_{Failure}$ and:

$$\Pi_{Failure} = p_c\Pi_{m-1,0} = (p_c)^m$$

This probability expresses the probability of attempting to transmit m times without success, where we can only calculate the average delay at the end of the m stages and not the actual one.

We denote by $\overline{D}_{state}$ the average sojourn time in one state, it is given by $\overline{D}_{state} = [p_c x + (1-p_c)1]$ except for states with $b(t) = 0$ where their sojourn time is 0. After every stage when $b(t) = 0$, a transmission duration of $x+1$ is always sensed: x time slots for packet transmission plus one additional idle time slot that is necessarily sensed after every transmission if all stations are performing LBT and are in backoff during the transmission $CW \geq 1$. $b(t)$ is chosen from a uniform distribution with mean value $(W_0-1)/2$ excluding the state with 0 sojourn time, hence the average delay in one stage is: $\overline{D}_{stage} = (W_0-1)[p_c x + (1-p_c)]/2 + x + 1$, which yields the following average total delay:

$$\overline{D} = m \times \overline{D}_{stage} = m \times \left[\frac{(W_0-1) \times [p_c x + (1-p_c)]}{2} + x + 1\right] \tag{3}$$

which is the average delay corresponding to reaching the Failure state (after m attempts).

2.1.2. *Timer-based modified Bianchi model*

The above model gives average delay and not the actual one. Recall that reliability of URLLC in the unlicensed system is obtained when the packet is not successfully transmitted within T_U time units. This problem appears mainly because we do not distinguish between idle and busy slots which does not allow us to track the actual delay of the packet. To remedy to this, we propose the following model. We distinguish between the resulting states from the current one depending on the sensed medium: idle or busy, and track the delay of every state with a new stochastic process $d(t)$ which represents the delay at time t. Hence, we get a three-dimensional Markov chain which allows us to calculate the actual reliability for a given actual delay. This model is introduced in Figure 4(a) for the case of $T_U = 3x$ and $W_0 = 4$.

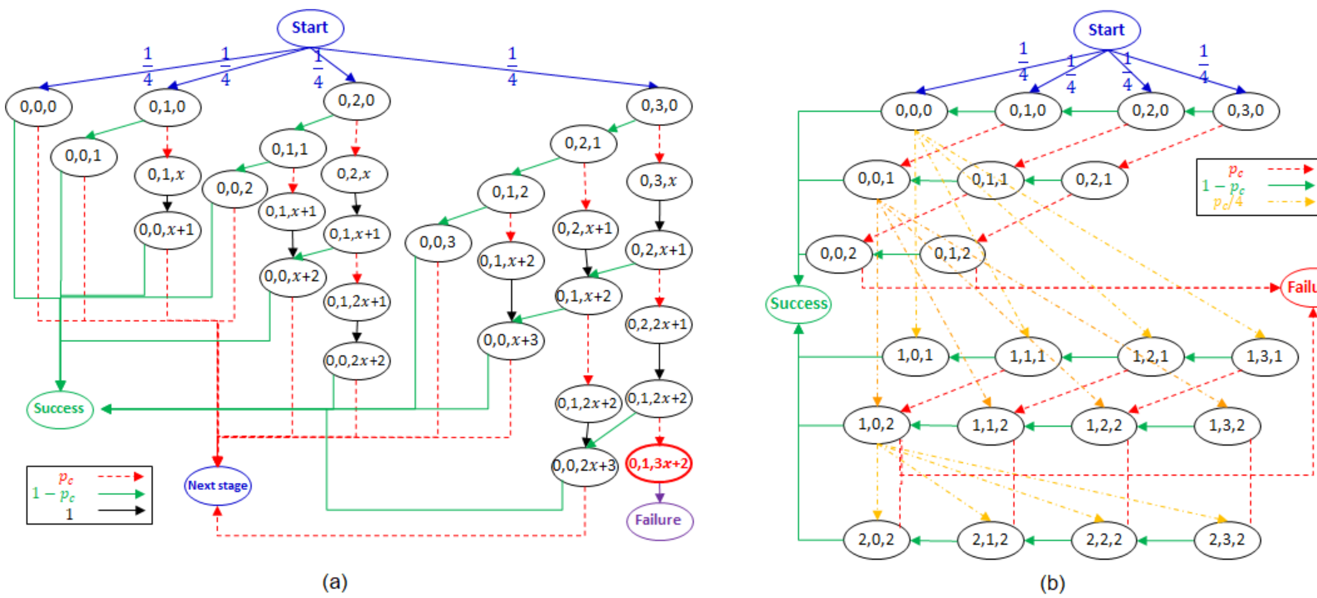

Figure 4: Example of 5G-U model with delay constraint (a) exact model (b) approximate model

Figure 4(a) illustrates the fact that every busy period is followed by at least one idle slot. The chain is built dynamically depending on the values of T_U and x, i.e., in every state we test if the constraint T_U is still respected and generate next states by adding 1 or x to the current delay, hence the number of stages is determined by the last possible state with $d(t) = T_U$.

Similar to our first model, the chain begins from stage 0 and a random backoff counter, generating the first row of states $\{0, b(0), 0\} : b(0) \in \{0, ..., W_0-1\}$, then every state has two possibilities to proceed depending on the sensed medium: idle or busy, except for the state $(0,0,0)$ where the packet is immediately transmitted without medium sensing. If the medium is sensed idle, then $b(t)$ is decremented by one and $d(t)$ is incremented by one, otherwise, the medium is busy for a duration of x consecutive time slots and $b(t)$ remains unchanged. The chain terminates when all its paths reach one of the absorbing states: Success or Failure. The state Success is reached when $b(t) = 0$ and the medium is idle having $d(t) < T_U$, while Failure is reached when $d(t) \geq T_U$ for any value of $s(t)$ and $b(t)$.

We assume in this example that $x > 3$ to get the illustrated chain, otherwise we would obtain another set of states. Due to the difficulty of displaying a large number of states, we illustrate in Figure 4(a) the states of the first stage only and gather the rest of the chain in one state called Next stage.

2.1.3. Approximate timer-based modified Bianchi model

The existence of two possible increments of $d(t)$ complicates the problem at hand because this generates a huge number of states for practical values of W_0, T_U and x, making the solution prohibitive. If we neglect the change in $d(t)$ after sensing an idle slot, then the states of different branches can be combined and the chain becomes more compact. This approximation affects the precision of calculating the probabilities of the chain, but the complexity reduction tips the balance in its favor. The new approximate model for the previous example is shown in Figure 4(b).

In Figure 4(b), we simplify the notation of $d(t)$ to express now the multiples of $x+1$ slots when the medium is sensed busy followed by an idle slot. This chain is no longer built dynamically state by state, but its probabilities can be deduced from a general formula. We can represent the balance equations of the non-transient form of the Markov chain (with state Inactive similar to that in Figure 3) in a three-dimensional matrix denoted by Π.

The dimensions of Π are $m \times W_0 \times m$, where m is the maximum number of stages that can be attained without violating the time constraint T_U assuming that $b(t)$ is always equal to 0:

$$m = \left\lfloor \frac{T_U}{x+1} \right\rfloor$$

where $\lfloor . \rfloor$ is the floor function.

We note that when $m > W_0$, T_U is not always attained in the first stages because we have W_0 busy periods at most, but to keep the homogeneity of the matrix dimensions in different stages we fill the additional states with zero probabilities. Starting by setting all probabilities to zero, we can describe the balance equations of the system in a recursive manner row by row starting from the columns with higher $b(t)$, as follows:

$$\Pi_{0,W_0-1,0} = q\frac{\Pi_{in}}{W_0}$$

$$\Pi_{0,W_0-j,0} = q\frac{\Pi_{in}}{W_0} + (1-p_c)\Pi_{0,W_0-j+1,0}, \quad 2 \leq j \leq W_0$$

$$\Pi_{0,W_0-j,k} = p_c\Pi_{0,W_0-j+1,k} + (1-p_c)\Pi_{0,W_0-j+1,k-1}, \quad \begin{array}{c} 1 \leq k \leq m-1 \\ k+1 \leq j \leq W_0 \end{array}$$

We note that the delay in one stage cannot be less than its number of stage, the first rows in the next stages remain zeros.

$$\Pi_{i,W_0-1,k} = \frac{p_c}{W_0}\Pi_{i-1,0,k-1}, \quad \begin{array}{c} 1 \leq i \leq m-1 \\ i \leq k \leq m-1 \end{array}$$

$$\Pi_{i,W_0-j,k} = \frac{p_c}{W_0}\Pi_{i-1,0,k-1} + (1-p_c)\Pi_{i,W_0-j+1,k} + p_c\Pi_{i,W_0-j+1,k-1}, \quad \begin{array}{c} 1 \leq i \leq m-1 \\ i \leq k \leq m-1 \\ 2 \leq j \leq W_0 \end{array}$$

We apply the normalization condition to calculate Π_{in}:

$$\Pi_{in} + \sum_{i=0}^{m-1}\sum_{j=0}^{W_0-1}\sum_{k=0}^{m-1}\Pi_{i,j,k} = 1$$

Since we now have several states with $b(t) = 0$ in every stage, the probability of transmission in one time slot is then:

$$\tau_i = \sum_{i=0}^{m-1}\sum_{k=0}^{m-1}\Pi_{i,0,k} \tag{4}$$

The expression of p_c in Equation (10) is still valid and is solved numerically with Equation (4) to converge to the fixed point of p_c. As a result, we have:

$$\Pi_{Failure} = p_c\sum_{i=0}^{m-1}\sum_{j=0}^{W_0-1}\Pi_{i,j,m-1} \tag{5}$$

In the sequel, we denote $\Pi_{Failure}$ by $P_{loss}^U(T_U)$, as the probability that a packet reaches delay T_U without being served in 5G-U.

2.2. Medium access model in licensed spectrum

Transmission in licensed spectrum is not subject to any regulations. Existing methods for uplink transmission are Grant-based (GB) scheduling and Grant-free (GF) on a common pool. GB scheduling is the one used in existing cellular systems, when a station wants to transmit, it sends a scheduling request to the base station through a random access channel, when the BS receives the request it allocates one or more resource blocks (RBs) to the station according to its demand and transmits back the positions of the allocated RBs in time and frequency to the station. This approach offers very high reliability and spectral efficiency since the resources are managed by one central unit. On the other hand, the resource reservation process is time consuming and unsuitable for delay constrained applications. In GF transmission, RBs are accessible without prescheduling similar to slotted Aloha protocol, which minimizes the delay on account of reliability degradation.

Time is slotted into small intervals called Transmission Time Intervals (TTIs) and available frequency bandwidth BW is also divided into K subcarriers of width w. The combination of one TTI-w is the RB, and we assume that one RB is sufficient to transmit one URLLC packet.

Assuming that stations are synchronized in time, when a station has a packet to transmit, it does so immediately at the beginning of the following TTI using a randomly chosen RB from the K available ones, as advocated by [7]. If the delay budget T_L is still respected, then more replicas in the following TTIs can be sent, without waiting for acknowledgment so as to reduce the delay. The number of allowed replicas is $\delta = \lfloor T_L/TTI \rfloor$. This mechanism is standardized and called TTI bundling for URLLC [13]. We denote by $\delta_{max} = \lfloor T/TTI \rfloor$ the maximum possible number of replicas.

We consider that URLLC transmission can be made robust to channel noise by increasing the transmission power or using spatial diversity. However, a packet can be damaged if other packets are sent over the same RB because of high interference, this will be considered as the only source of packet loss in our analysis. In the blind replication method, the packet is lost if and only if all its replicas collide with other transmissions.

From the point of view of the packet being transmitted, each replica out of the δ ones chooses one RB in every TTI with probability $1/K$. Accordingly, another active packet does not choose the same RB in the same TTI with probability $1 - 1/K$.

We can apply the chain rule to calculate the probability of loss of a packet, where this method is based on calculating the probability of loss given the number of active packets, then the total probability of loss.

Denoting by n the number of active packets other than the one being considered in one TTI; $n \in \{1, 2, ..., N-1\}$, we exclude $n = 0$ the case where there is no collision. A_n denotes the event of having n other packets in one TTI and C denotes the event of having a collision between the packet under study and at least one other packet from the n active ones in δ consecutive TTIs. We have:

$$\mathbb{P}(A_n) = \binom{N-1}{n} P_a^n (1 - P_a)^{N-1-n}$$

$$\mathbb{P}(C|A_n) = (1 - (1 - \frac{1}{K})^n)^\delta$$

where P_a is the probability of arrival to 5G-L.

By applying the chain rule, we obtain the probability of loss in 5G-L for K available RBs per TTI:

$$\begin{aligned} P_{loss}^L(\delta) &= \sum_{n=1}^{N-1} \mathbb{P}(C|A_n)\mathbb{P}(A_n) \\ &= \sum_{n=1}^{N-1} \binom{N-1}{n} P_a^n (1 - P_a)^{N-n-l} (1 - (1 - \frac{1}{K})^n)^\delta \end{aligned}$$

2.3. Combined unlicensed and licensed transmission

When combining 5G-U and 5G-L systems, we have to be careful about their time units. 5G-U operates in T_s unit which is considerably smaller than TTI used in 5G-L. To unify the units of the combined system, we assume that $TTI = zT_s$ where z is an integer. To adapt the aforementioned arrival probability q per T_s to TTI, and considering that a station generates one packet at most during δ TTIs, the packet arrival probability to the licensed system is given by $P_a = 1 - (1 - q)^{\delta z}$.

For the case of combined 5G-U then 5G-L transmission, packets arrive at 5G-L after failing the transmission on 5G-U with probability $P_{loss}^U(T_U)$, which can be written equivalently as $P_{loss}^U(\delta)$ since $T_U = T - \delta z$, resulting in:

$$P_a(\delta) = 1 - (1 - qP_{loss}^U(\delta))^{\delta z} \tag{6}$$

Since TTI bundling requires synchronization among stations, and since packets arrive randomly in time, if a packet reaches 5G-L amid the TTI, then it is postponed till the beginning of the next one, which may cause a maximum delay of the packet of $(z-1)T_s$ but its impact is small and is neglected in our analysis for simplification.

We derive now the formula of total probability of loss for the combined 5G-U and 5G-L system, denoted by $P_{loss}(\delta)$, which quantifies the reliability $R = 1 - P_{loss}(\delta)$ of the system under delay constraint $T = T_U + T_L$:

$$P_{loss}(\delta) = \sum_{n=1}^{N-1} \binom{N-1}{n} P_a(\delta)^n (1 - P_a(\delta))^{N-n-l} [1 - (1 - \frac{1}{K})^n]^\delta \tag{7}$$

Using this formula, we can determine numerically the reliability of the combined system for a given bandwidth K and a given policy δ, which is equivalent to finding the minimum K which satisfies the target reliability requirement for a given δ.

2.4. Optimal transmission policy

To take advantage of the combined 5G-U 5G-L system to the maximum, we need to determine the optimal time division between T_U and T_L which respects delay and reliability constraints while minimizing bandwidth cost in 5G-L. In other words, we opt to find the optimal pair $(\delta, K)^\star$ with minimum K subject to reliability and delay constraints.

From Equation (7), we notice that minimizing P_{loss} for a given K is equivalent to minimizing K for a target P_{loss}, hence $\delta^\star$ is the same for either cases.

For a given K, $P_{loss}(\delta)$ is the composition of two functions: $P_{loss}(\delta) = P^L_{loss}(P^U_{loss}(\delta))$ where P^U_{loss} is a decreasing function of δ and P^L_{loss} is an increasing one. Hence, it is hard to predict the behaviour of $P_{loss}(\delta)$ analytically, but we can predict that neither increasing nor decreasing δ would minimize the function. We illustrate later the previous study numerically.

3. URLLC transport in multi-tenant environment

In a real-life situation, an industrial area includes many automated factories in vicinity, creating non-negligible interference from one to another. If all factories are operated by one tenant then this interference can be managed easily, but usually it is not the case and factories are operated by different tenants. Assuming the existence of M tenants operating in proximity, 5G-U is no longer confined from the viewpoint of one tenant, and if every tenant tries to use it selfishly (without considering neighbouring interfering stations) then the overall interference could increase and the gain from unlicensed spectrum is reduced.

We denote a given tenant by v_i (for vertical) and we evaluate its system performance under interference from other tenants, $i \in \{1, ..., M\}$. v_1's coverage includes N stations transmitting in the uplink, N_i of which belong to v_i where $i \in \{1, ..., M\}$ and $N = N_1 + N_2 + ... + N_M$. N_i denotes the number of v_i stations inside the coverage area of v_1 only. This scenario is illustrated in Figure 5.

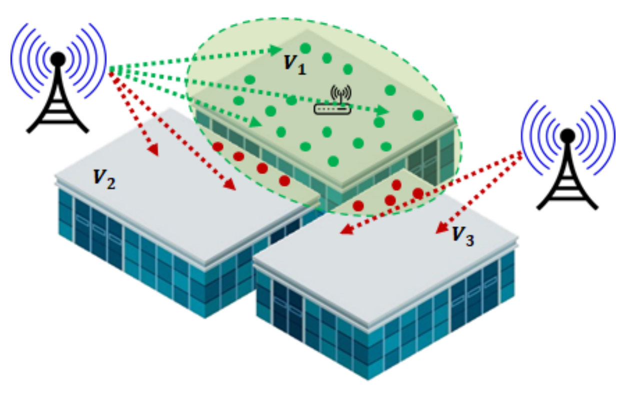

Figure 5: Scenario of industrial multi-tenant environment

Each vertical v_i deploys a URLLC transmission strategy δ_i; $i \in \{1, ..., M\}$, $\delta_i \in \{1, ..., \delta_{max} - 1\}$].

This situation can be represented by a non-cooperative game with triplet $G = (V, \{S_i\}_{i\in V}, \{u_i\}_{i\in V})$ where $V = \{v_1, v_2, \ldots, v_M\}$ is the finite set of players, S_i is the set of strategies of v_i represented by $\delta_i \in \{1, ..., \delta_{max} - 1\}$, and u_i is the utility function of v_i which is the inverse of its cost function represented by the required BW to satisfy the reliability and latency requirements, equivalent to $K_i(\delta_1, ..., \delta_M)$.

In non-cooperative games, each player aims to maximize its own utility over its strategy set, thus player i chooses the strategy s_i which maximizes its utility u_i for a given vector of strategies $\vec{s} = (s_1, ..., s_M)$. Thereafter, player i waits for others to change/keep their strategies, and then it changes/keeps its strategy accordingly. If there exists a vector of strategies $\vec{s}^\star = (s_1^\star, ..., s_M^\star)$ which satisfies $\forall i \in V, \forall s_i^{'} \in S_i, u_i(s_i^\star, \vec{s}^\star_{-i}) \geq u_i(s_i^{'}, \vec{s}^\star_{-i})$ where $\vec{s}^\star_{-i}$ refers to the set of strategies of all players except player i, then the game has Nash equilibria [14].

Our game can be considered as finite since it has finite sets of players and strategies. Nash showed in [15] that at least one equilibrium point exists in finite games. However, this proves the existence of mixed strategies only and not pure ones, so to determine whether our game has pure Nash equilibria or not, we have to either adopt a numerical approach or propose a new analysis methodology to obtain a closed-form formula of the utility function and determine if it is concave or quasi-concave, which leads to the proof of having pure-strategy Nash equilibria.

In the following, we model the system and illustrate the existence of pure Nash equilibria and identify the equilibrium point(s) in our proposed scenario.

3.1. Medium access model in unlicensed spectrum

3.1.1. Fixed point analysis

The proposed model in subsection 2.1.3 is still valid for the multi-tenant case, where the effect of other stations on the one being studied is present in p_c calculation, which we denote here by $p_{c,1}$ referring to the collision probability calculated by v_1.

Different policies δ_i suggest having different number of stages m_i, $i \in \{1, ..., M\}$, then the probability of transmission in Equation (4) becomes:

$$\tau_i(\delta_i) = \sum_{j=0}^{m_i-1} \sum_{k=0}^{m_i-1} \Pi_{j,0,k}, \quad i \in \{1, ..., M\} \tag{8}$$

Equation (10) is rewritten similarly for all tenants as:

$$p_{c,i}(\delta_1, ..., \delta_M) = 1 - (1 - q\tau_i(\delta_i))^{Ni-1} \times \sum_{\substack{j=0 \\ j \neq i}}^{M} (1 - q\tau_j(\delta_j))^{N_j} \tag{9}$$

Numerically, we can assess the impact of interfering stations from other tenants on v_1 by solving the set of fixed-point Equations (8) and (9), then plugging $p_{c,1}$ into Equation (5) to get the loss probability for v_1 denoted by $P_1(\delta_1, ..., \delta_M)$.

Solving the fixed point does not allow us to have a closed-form expression for the Nash equilibrium points of $P_1(\delta_1, ..., \delta_M)$. We propose next an approximate way to obtain such expression.

3.1.2. Closed-form analysis

For this purpose, we go back to the first model illustrated in Figure 2. Since the arrival of packets is random and the arrival rate q is assumed to be small, the random backoff process can be reduced to the arrival process only, and hence a time slot is busy if one or more packets arrive at the same time. $p_{c,1}$ is then expressed as follows:

$$\begin{aligned} p_{c,1} &= 1 - (1 - q)^{N-1} \\ &\approx 1 - [1 - (N - 1)q] \\ &= (N - 1)q \end{aligned} \tag{10}$$

$p_{c,1}$ depends also on the number of stages a packet goes through, because the actual number of packets in the system depends on their arrival and whether they were successfully transmitted or are still in backoff. Assuming that all tenants deploy the same W_0, we estimate the average number of stages a packet goes through $\overline{m}_i$ in a reverse manner of how we calculated the average delay of the packet in subsection 2.1. We use the formula in Equation (3) and adapt it to the multi-tenant scenario as:

$$\overline{D}_i = m_i \times \left[\frac{(W_0 - 1) \times [p_{c,i}x + (1 - p_{c,i})]}{2} + x + 1 \right], \quad i \in \{1, ..., M\}$$

Denoting the time budget of tenant v_i by $T_U^i = T - \delta_i z$, we calculate the average number of stages as:

$$\overline{m}_i = \frac{T_U^i}{\overline{D}_i}, \quad i \in \{1, ..., M\}$$

noting that $\overline{m}_i \in \mathbb{R}$.

The probability of going through $\overline{m}_i$ stages without success is $(p_{c,i})^{\overline{m}_i}$, and so staying in the backoff phase has a probability of $1 - (p_{c,i})^{\overline{m}_i}$. The actual number of packets that are still in backoff phase $\tilde{N}_i$ is approximated by:

$$\tilde{N}_i = N_i(1 - (p_{c,i})^{\overline{m}_i}), \quad i \in \{1, ..., M\}$$
$$\tilde{N} = \tilde{N}_1 + \tilde{N}_2 + ... + \tilde{N}_M$$

We inject $\tilde{N}$ into equation 10, so we can get a more accurate value of $p_{c,i}$ knowing that $\tilde{N}$ depends on $p_{c,i}$:

$$\begin{aligned} p'_{c,i} &= 1 - (1-q)^{\tilde{N}-1} \\ &\approx (\tilde{N} - 1)q \end{aligned} \tag{11}$$

The probability of failure of v_1 equals to the probability of going through $\overline{m}_1$ stages without success:

$$P_1^U(\delta_1, ..., \delta_M) = (p'_{c,1})^{\overline{m}_1}$$

And hence we obtain a closed-form formula for P_1^U.

In this approach, we are using the fact that q is infinitesimal to perform the approximations in equations (10) and (11). However, when q tends to grow, this approximation is no longer valid and we cannot use this approach anymore.

3.2. Combined unlicensed and licensed transmission

Transmission in 5G-L is the same as in the case of one-tenant, so does the analysis of the combined system by replacing the corresponding probabilities of v_1 by the probabilities of the one-tenant. We now perform the following analysis to demonstrate the existence of pure Nash equilibrium points analytically using the closed-form analysis.

Based on Equation (6), we simplify the probability of arriving to 5G-L after 5G-U by noting that $qP_1^U(\delta_1, ..., \delta_M) \ll 1$ as:

$$\begin{aligned} P_{a,1}(\delta_1, ..., \delta_M) &= 1 - (1 - qP_1^U(\delta_1, ..., \delta_M))^{\delta_1 z} \\ &\approx \delta_1 x q P_1^U(\delta_1, ..., \delta_M) \end{aligned}$$

The global probability of loss of the system is as indicated in Equation (7) but seen from the point of view of v_1 it is:

$$P_1(\delta_1, ..., \delta_M) = \sum_{n=1}^{N_1-1} \binom{N_1 - 1}{n} (P_{a,1})^n (1 - (P_{a,1}))^{N_1-n-l} [1 - (1 - \frac{1}{K_1})^n]^{\delta_1} \tag{12}$$

One way to a faster evaluation of expression (12) can be obtained by making further approximations. We first approximate $(1 - (1 - 1/K_1)^n)^\delta$ by $(n/K_1)^\delta$ for $n < K_1$, this is valid because even if $n \geq K_1$, $(P_{a,1})^n$ is in this case very small and can be taken as zero.

$$\begin{aligned} P_1 &\approx \sum_{n=1}^{N_1-1} \binom{N_1 - 1}{n} (P_{a,1})^n (1 - P_{a,1})^{N_1-n-l} (\frac{n}{K_1})_1^\delta \\ &= \frac{1}{K_1^\delta} \sum_{n=1}^{N_1-1} n_1^\delta \binom{N_1 - 1}{n} (P_{a,1})^n (1 - P_{a,1})^{N_1-n-l} \end{aligned}$$

This formula represents the δ_1th moment of a binomial distribution with probability $P_{a,1}$ and $N_1 - 1$ trials. The moment-generating function (MGF) is:

$$M_X(t) = (1 - P_{a,1} + P_{a,1}e^t)^{(N_1-1)}$$

To obtain the jth moment, $M_X(t)$ is differentiated j times then evaluated at $t = 0$. Evaluating the first moments we get:

$$M_1 = M_X(0)' = (N_1 - 1)P_{a,1}$$

$$M_2 = (N_1 - 1)P_{a,1} + (N_1 - 1)(N_1 - 2)(P_{a,1})^2$$

$$M_3 = (N_1 - 1)P_{a,1} + (N_1 - 1)(N_1 - 2)(P_{a,1})^2 + 2(N_1 - 1)(N_1 - 2)(P_{a,1})^2 + (N_1 - 1)(N_1 - 2)(N_1 - 3)(P_{a,1})^3$$

where M_j represents the jth moment of the distribution. Neglecting higher orders of $P_{a,1}$ because they become considerably small, and considering the first two orders only, we obtain a simple generalized formula of M_j as:

$$M_j = (N_1 - 1)P_{a,1} + (2^{j-1} - 1)(N_1 - 1)(N_1 - 2)(P_{a,1})^2$$

Thus:

$$P_1(\delta_1, ..., \delta_M) = \frac{1}{K_1^{\delta_1}}[(N_1 - 1)\delta_1 xqP_1^U + (2^{\delta_1 - 1} - 1)(N_1 - 1)(N_1 - 2)(\delta_1 xqP_1^U)^2] \quad (13)$$

We illustrate the existence of Nash equilibria by demonstrating the convexity of $P_1(\delta_1, ..., \delta_M)$ for all $\delta_1 : 1 \leq \delta_1 < \delta_{max}$. This function is differentiable and is verified to be convex because its second derivative is positive for the values that correspond to URLLC scenarios.

Since the set of possible values of δ_1 is limited, the simplest way to identify the equilibrium points is by brute force search of the points which minimize this function.

4. Numerical results

4.1. 5G-U model validation

In the numerical and simulation results presented next, we consider 5G-U system with parameters as defined in the latest IEEE 802.11 systems operating on the 5GHz unlicensed band, regarding the time slot duration T_s, the backoff periods $SIFS$, $DIFS$ and bit rate R_b. The transmission uses the whole available unlicensed band of spectrum, hence only one transmission can take place in a given time. The data packet size including all headers is denoted by L_{data} and the acknowledgment packet size by L_{ack}. The station receives the ACK/NACK after a duration of SIFS, then all stations backoff during a period of DIFS before starting to contend again for medium access. The transmission duration x is then calculated as: $x = \lceil \frac{(L_{data}+L_{ack})/R_a+SIFS+DIFS}{T_s} \rceil$, where $\lceil . \rceil$ is the ceiling function. The system deploys LBT-cat3 with a fixed contention window size of W_0. We consider that the station generates a packet every $10ms$ following a Poisson distribution, then we can estimate the probability of packet arrival q per T_s. The latency and reliability requirements are set to $T = 1ms$ and $R = 1 - 10^{-5}$, respectively. The numerical values are shown in Table 1.

Table 1: Numerical values of the system parameters

T_s	$9\mu s$	R_b	$100Mbps$	x	7
$SIFS$	$16\mu s$	L_{data}	$32Bytes$	W_0	16
$DIFS$	$34\mu s$	L_{ack}	$14Bytes$	q	≈ 0.001

We compare the results obtained from our analytical model with the results obtained from simulation. We build an event-driven simulator using MATLAB with time step of T_s. We designate a station to which we estimate p_c and P_{loss}^U. This station generates a packet as soon as its previous one quits the system, unlike the remaining $N-1$ stations which generate packets in a given T_s following a Poisson distribution of probability q. To conform with the hypothesis of our model that p_c is independent of time, we reduce the contention process of the $N-1$ stations to their arrival process only, i.e., in every T_s we generate a new vector of length $N-1$ that contains $N-1$ Poisson random variables with probability q representing the state of the medium: idle if the vector is empty and busy if not.

Packets of the designated station are tagged with a delay counter d initialized to 0 and a contention window CW chosen randomly between 0 and W_0. Every T_s, we check whether the designated packet is being transmitted when its $CW = 0$ or not. We also assess the state of the medium depending on the generated random vector. If the packet is being transmitted and the medium is busy then there is a collision, d is incremented by x and CW is chosen randomly again, if the medium is idle then it is a successful transmission and a new packet is generated in the next T_s. When the packet is not being transmitted, d is incremented by 1 or x and CW is decremented by 1 or remains unchanged depending whether the medium is sensed idle or busy. If d reaches the allowed delay budget T then the packet is lost and replaced with a new one.

To better simulate our model, we neglect the one time slot increment in delay when idle in simulation as in the model.

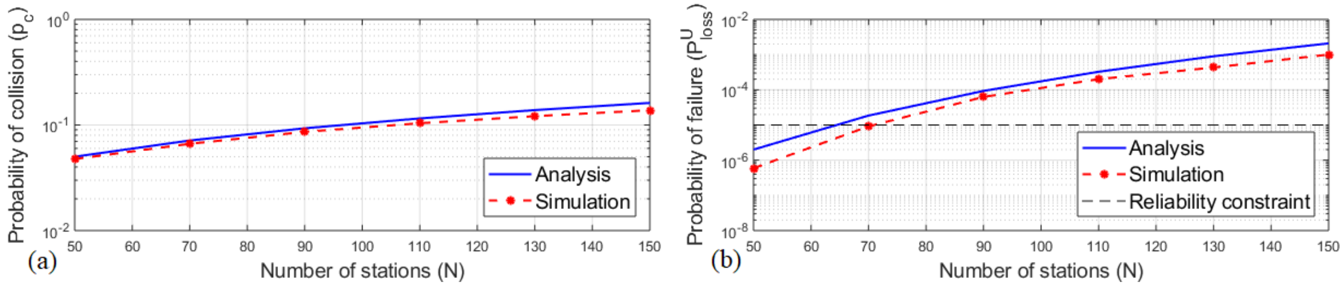

Figure 6: 5G-U model validation (a) probability of collision validation (b) probability of failure validation

We plot the curves obtained from analysis and simulation for p_c and P^U_{loss} under delay requirement, for different loads (N), and demonstrate the validity of our model in Figure 6.

Figure 6 shows indeed a good match for both p_c and P^U_{loss} curves. We infer the maximum number of stations before which 5G-U alone is capable of providing the required QoS from Figure 6(b), which corresponds to N which verifies $P^U_{loss} = 10^{-5}$. We denote this number by N_{max} which is approximately equal to 65 stations for our numerical application, based on the curve obtained from analysis.

4.2. 5G-L model validation

Transmission in licensed spectrum has become more flexible in 5G than in 4G. Depending on the application, it is now possible to choose TTI length from a range of values. For delay constrained applications, like URLLC, we choose the smallest length of TTI defined in the standards: $TTI = 0.125ms$, on account of the needed bandwidth for the same transmission.

We assume that a URLLC packet fits in one conventional LTE RB of $0.5ms$ duration and $180KHz$ bandwidth (12 subcarriers with carrier spacing of $15KHz$), having that $TTI = 0.125ms$, the bandwidth of our RB is then $w = 720KHz$.

Reliability and delay constraints are taken as $R = 1 - 10^{-5}$ and $T = 1ms$, respectively, which correspond to $\delta = 8$. To get the same characteristics as in 5G-U, we translate q into P_a as discussed in Section 3.2. We assume $K = 5$ RBs and we evaluate P^L_{loss}.

We compare the results obtained from the analytical model with a system simulation realised using MATLAB. The simulation output is calculated from the viewpoint of a designated station which is always active, the other $N - 1$ stations generate a packet every T time with probability $P_a \approx 0.1$.

We trace the curves obtained from analysis and simulation for different values of N in Figure 7(a) and observe a good match.

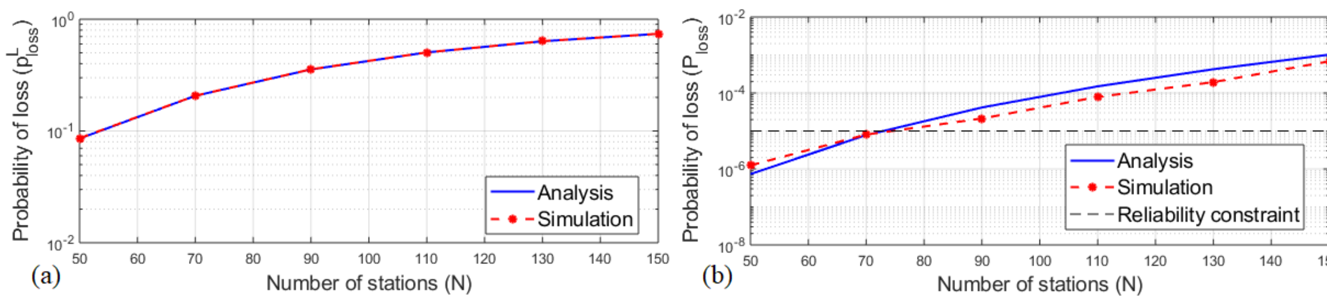

Figure 7: (a) 5G-L model validation (b) Combined 5G-U/5G-L model validation

We notice here that $K = 5$ is not enough to guarantee the reliability requirement when using 5G-L alone, where $\forall N : P^L_{loss} > 10^{-5}$.

4.3. Combined 5G-U/5G-L model validation

We validate our combined 5G-U/5G-L model against simulation for the case of $\delta = 3$ and $K = 5$, as illustrated in Figure 7(b). The curves show a good match. We notice that the reliability is enhanced compared to standalone 5G-U or 5G-L but it is still close to 5G-U performance since we are not considering a large number of licensed replicas ($\delta = 3$). One should note that it is necessary to adapt K to the number of stations to get the desired reliability; $K = 5$ is not sufficient for large N values.

4.4. Optimization problem illustration

As discussed in Section 2.4, there is a trade-off between time budget and reliability in the combined system that must be studied to determine the optimal transmission policy δ which minimizes the cost K. For this purpose, we use the analytical model to determine the minimum K which satisfies $P_{loss} \leq 10^{-5}$ for every possible δ. For a given N, we can determine $\delta^\star$ which yields to the smallest K: $K^\star$. This study is illustrated in Figure 8(a) for different values of $N \in \{100, 150, 200\}$, beyond the previously illustrated capacity of 5G-U, $N_{max} \approx 65$. The maximum available number of RBs is taken as 99, and reaching it means that reliability is not guaranteed yet.

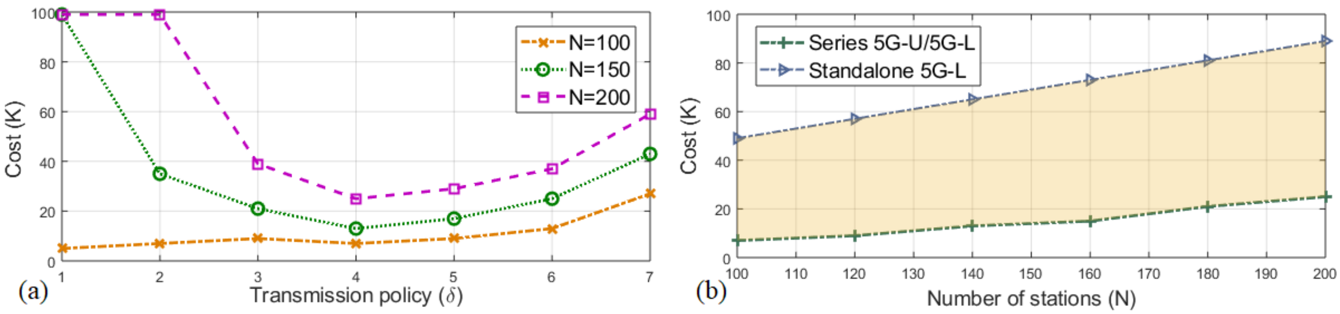

Figure 8: (a) Optimization problem illustration (b) Combined system gain illustration

We notice from Figure 8(a) that for $N = 100$ where N is close to N_{max}, 5G-U can handle most of the traffic, hence it is better to choose smaller δ. However, in the regime far from N_{max}, the trade-off becomes more obvious, and $\delta^\star$ tends to be in mid-range of possible δ values confirming the reasoning mentioned in Section 2.4.

For our numerical application, we obtain $\delta^\star = 1$ for $N = 100$ and $\delta^\star = 4$ for $N = 150$ and 200, noting that when having a more strict reliability constraint R for example, $\delta^\star$ tends to take larger values because 5G-L's role becomes more significant to cope with the demand.

4.5. System gain

To illustrate the importance of using unlicensed spectrum for URLLC transmission, we show in Figure 8(b) the gain in terms of number of RBs when using the combined system with optimal transmission $\delta^\star = 4$ compared to standalone 5G-L transmission over the whole time budget, , $T_U = 0$ and $T_L = T$. We notice a substantial gain, which corresponds to the added unlicensed bandwidth.

4.6. Other system parameters

If we explore other calibratable system parameters apart from δ, we observe having the contention window size W_0, which plays a big role in the waiting time of the packet in each stage, i.e., reducing W_0 reduces the average time spent in a stage which increases the total number of stages within T, increasing by that the reliability. However, decreasing W_0 is not always beneficial to the system performance, especially when a big number of full-buffer stations are contending to access the network, which increases the probability of collision. Hopefully, our scenario deals with sporadic traffic with low arrival rate, rendering the parameter calibration easier.

We illustrate in Figure 9 the effect of changing W_0 on the cost of the system, while fixing δ to 4.

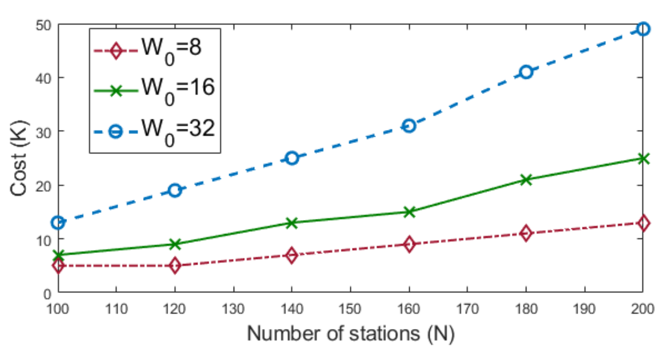

Figure 9: The effect of changing the contention window size W_0

We observe that decreasing W_0 helps reducing the cost largely. We can estimate from Figure 9 that cutting off W_0 by half discounts K as well by half.

We should also mention that different W_0 yield potentially to different optimal policy $\delta^\star$. To guarantee fairness among stations, we fix W_0 and calibrate δ only.

4.7. Multi-tenant model validation

In the case of multi-tenant existence, we validate the performance of the overall system model (combined 5G-U/5G-L), obtained from the two methods: fixed-point analysis and closed-form analysis, against simulation. We consider the existence of two tenants v_1 and v_2 with: $N_1 = 3N/4$ and $N_2 = N/4$. To illustrate the effect of different policies we choose $\delta_1 = 5$ and $\delta_2 = 3$. We fix the number of RBs available to v_1 to $K = 5$.

We modify the previous simulator used in one-tenant case which tracks every packet to distinguish the packets with different policies of different tenants. We compare in Figure 10(a) the curves obtained from the analytical models with the one obtained by simulation.

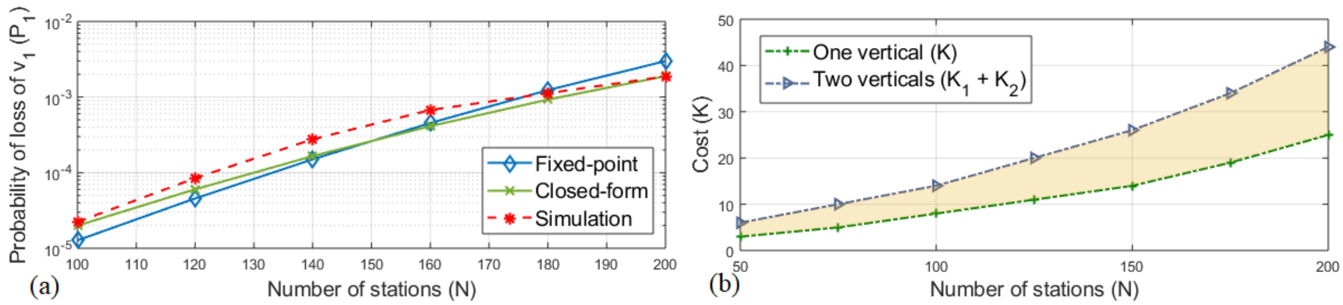

Figure 10: (a) Models validation of multi-vertical case (b) Price of anarchy

Figure 10(a) validates the two models, despite the approximations we did in the closed-form one. The small difference between simulation and analysis is due to p_c dependence on time in the simulation.

4.8. Nash equilibria illustration

For the case of two tenants, we illustrate Nash equilibria by evaluating the cost of all possible combinations of the pair (δ_1, δ_2). We first consider v_1 as the tenant of interest with $N = 180$: $N_1 = 135$ and $N_2 = 45$, then we consider v_2 as the tenant of interest with $N_1 = 45$ and $N_2 = 135$, which yields to a symmetrical scenario for both tenants. We illustrate in Table 2 the resulting cost matrix which gives the number of RBs that every tenant has to reserve for the different policies. Note that 99 RBs is the maximum offered cost, which also indicates that reliability target is not reached and further resources should be allocated to guarantee the required QoS.

Table 2: Nash equilibrium illustration

δ_2 \ δ_1	**1**	**2**	**3**	**4**	**5**	**6**	**7**
1	99, 99	56, 99	24, 99	18, 99	19, 99	23, 99	34, 99
2	99, 56	56, 56	24, 56	18, 55	19, 54	23, 51	34, 46
3	99, 24	56, 24	24, 24	18, 24	18, 23	23, 23	34, 21
4	99, 18	55, 18	24, 18	**18, 18**	**18, 17**	23, 17	34, 16
5	99, 19	54, 19	23, 18	**17, 18**	18, 18	23, 18	34, 17
6	99, 23	51, 23	23, 23	18, 23	18, 23	22, 22	32, 21
7	99, 34	46, 34	21, 34	16, 34	17, 34	21, 33	32, 32

We observe from Table 2 the existence of multiple equilibrium points, which correspond to the set of pure-strategy Nash equilibria: $(\delta_1, \delta_2) = (4, 4), (4, 5), (5, 4)$. If the game begins at any of these strategies, it is not in the interest of either player to change their strategy because it does not improve its payoff (minimizes its cost). Neighbouring values of δ_i can lead to the same cost due to the quantization granularity where multiple close values of P_1 lead to the same value of K. Table 2 illustrates the fact that decreasing the time budget in 5G-U for one player (increasing its δ_i) improves the performance for the other player. By this, we conclude that our game has pure-strategy Nash equilibria.

4.9. Price of anarchy

It is interesting to discuss the notion of *price of anarchy* in non-cooperative games, which measures the efficiency deterioration of the system in the presence of multiple non-cooperative players, compared to a cooperative system. We evaluate in Figure 10(b) the cost in terms of the number of RBs in the licensed band for the case of one player with N stations versus the case of two players with $N/2$ stations each; the cost in the second case is $K_1 + K_2$. Figure 10(b) confirms that a cooperative setting achieves higher gain than a non-cooperative one.

5. Conclusion

We studied in this work the transport of URLLC traffic in the uplink in a an industrial area with smart factories, where the transmission is considered on both unlicensed and licensed spectrum. We proposed a transmission policy where URLLC packets privilege the use of unlicensed spectrum to reduce the cost of transmission, but when the number of stations increases then we have to use additional resources of licensed spectrum to guarantee the stringent reliability and delay constraints. The station attempts to transmit its packet through unlicensed spectrum during a time budget smaller than the delay constraint, but if the packet fails to be successfully transmitted, it attempts in the remaining time to transmit the packet using licensed spectrum.

We first considered the case of one tenant managing the area and confined unlicensed spectrum, we modeled the medium access of the proposed system, validated the model against simulations and quantified its performance in terms of probability of loss under delay constraint and then optimized the transmission policy to minimize the cost (needed licensed resources).

We then considered the case of multi-tenant industrial environment where unlicensed spectrum is shared among stations in the vicinity belonging to different tenants, which degrades its value-add to the system. We modeled this scenario as a non-cooperative game where the tenants are the players deploying the above proposed transmission policy on both spectra, the cost function is the needed licensed resources to attain reliability constraint. We modeled the medium access in the system in two ways: fixed-point and closed-form analysis and showed their accuracy in comparison to system level simulations. We illustrated the existence of pure Nash equilibria of our game analytically using the closed-form formulation and verified it numerically. The existence of pure Nash equilibria shows that unlicensed spectrum is a viable option even in a multi-tenant scenario. However, as indicated by our analysis for the price of anarchy, a situation where a common operator manages the unlicensed spectrum access for all tenants would lead to a better overall utility.

Our setting considers an offline analysis of discussed problems for known number of stations, probability of transmission and channel occupancy time, hence, solving the set of fixed-point equations in one-tenant case, or using brute-force to find optimal policy in multi-tenant case is only performed once and has no impact on the processing time for URLLC packets. However, the model can be rendered more dynamic for non-stationary environments by implementing online-learning algorithms. For example, in one-tenant scenario, online-learning can be employed to directly assess external interference p_c from the channel, which can then be plugged into our analysis (no fixed-point is needed) to find optimal transmission policy $\delta^\star$. In this case, interference estimation is done periodically, and some processing time is added to the transmission one, related to the periodicity of learning. In a realistic case of our scenario, the number of stations does not vary a lot with time, hence learning frequency can be set low, leading to a negligible impact on packets delay.

References

[1] M. Alliance, "Multefire release 1.0 technical paper: A new way to wireless," *white paper, Jan*, 2017.

[2] 3GPP, "Physical layer procedures for data," 3GPP TR 38.214 v15.1.0, Tech. Rep., March 2018.

[3] M. Maternia, S. E. El Ayoubi, M. Fallgren, P. Spapis, Y. Qi, D. Martín-Sacristán, Ó. Carrasco, M. Fresia, M. Payaró, M. Schubert *et al.*, "5g ppp use cases and performance evaluation models," *5G PPP*, 2016.

[4] G. J. Sutton, J. Zeng, R. P. Liu, W. Ni, D. N. Nguyen, B. A. Jayawickrama, X. Huang, M. Abolhasan, and Z. Zhang, "Enabling ultra-reliable and low-latency communications through unlicensed spectrum," *IEEE Network*, vol. 32, no. 2, pp. 70–77, 2018.

[5] H. Chen, R. Abbas, P. Cheng, M. Shirvanimoghaddam, W. Hardjawana, W. Bao, Y. Li, and B. Vucetic, "Ultra-reliable low latency cellular networks: Use cases, challenges and approaches," *IEEE Communications Magazine*, vol. 56, no. 12, pp. 119–125, 2018.

[6] 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*, vol. 55, no. 2, pp. 70–78, February 2017.

[7] 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, no. 2, pp. 182–185, 2018.

[8] 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*, March 2019, pp. 87–94.

[9] G. Bianchi, "Performance analysis of the ieee 802.11 distributed coordination function," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 3, pp. 535–547, March 2000.

[10] H. Zhai, Y. Kwon, and Y. Fang, "Performance analysis of ieee 802.11 mac protocols in wireless lans," *Wireless communications and mobile computing*, vol. 4, no. 8, pp. 917–931, 2004.

[11] G. J. Sutton, R. P. Liu, and Y. J. Guo, "Delay and reliability of load-based listen-before-talk in laa," *IEEE Access*, vol. 6, 2018.

[12] G. Bianchi, "Performance analysis of the ieee 802.11 distributed coordination function," *IEEE Journal on selected areas in communications*, vol. 18, no. 3, pp. 535–547, 2000.

[13] 3GPP, "Study on latency reduction techniques for LTE," 3GPP TR 36.881 v14.0.0, Tech. Rep., June 2016.

[14] S. Lasaulce, M. Debbah, and E. Altman, "Methodologies for analyzing equilibria in wireless games," *IEEE Signal Processing Magazine*, vol. 26, no. 5, pp. 41–52, 2009.

[15] J. Nash, "Non-cooperative games," *Annals of mathematics*, pp. 286–295, 1951.