Multi-Base Station Coordination for Semi-Persistent Scheduling in Industrial 5G Wireless Time-Sensitive Networks

Javier Villares, Youssef El Kaisi and Olga Muñoz
Dept. of Signal Theory and Communications
Universitat Politècnica de Catalunya (UPC), Barcelona, Spain
E-mail: {javier.villares, olga.munoz}@upc.edu, youssef.elkaisi@estudiantat.upc.edu

***Abstract*—In this paper, we design a 5G-compliant semi-persistent scheduling algorithm that guarantees determinism (i.e., bounded delay and zero jitter) over the wireless link for delay-constrained traffic. This is especially important for industrial remote control applications handling time-sensitive traffic. To this end, we adapt the window reservation mechanism for wired networks introduced in IEEE 802.1Qbv to also ensure determinism over 5G wireless links. Based on the statistical characterization of the radio channel, we develop a rigorous method to determine the number of time-frequency resources that must be reserved in each 5G subframe to accommodate all time-sensitive traffic traversing the wireless link. Numerical simulations are conducted in representative industrial scenarios, showing that neighboring base stations have to coordinate their transmissions to avoid interference and mitigate channel fading. Numerical results demonstrate that window design for time-sensitive traffic benefits from coordination in two key ways: by extending coverage and significantly increasing the number of time-critical flows that a 5G network can support in industrial deployments.**

***Index Terms*—Time-sensitive networks (TSN), 5G, scheduling, industrial communications**

I. Introduction

As industrial automation and new latency-constrained applications become more sophisticated, determinism has emerged as a critical requirement for their underlying communication systems [1]. Determinism refers to the ability to guarantee that packets experience a stable end-to-end delay throughout the entire duration of the connection. Specifically, deterministic networks require that both the maximum delay and delay variability (jitter) are strictly bounded within predetermined limits. To support these requirements, the IEEE 802.1 Time-Sensitive Networking (TSN) Task Group has issued several standards [2]. Among them, two have attracted significant attention: IEEE 802.1as [3] and 802.1Qbv [4]. While IEEE 802.1as is aimed at synchronizing all network nodes with nanosecond precision, IEEE 802.1Qbv, also known as Time-Aware Shaping (TAS), focuses on defining precise transmission windows that allow forwarding packets between nodes without queuing delays. Although all these standards were initially conceived for Ethernet-based networks, considerable interest has recently arisen in incorporating wireless networks into TSN infrastructure [5][6].

In wireless environments, the time required for reliable packet transmission fluctuates throughout the connection due to various propagation impairments, including channel fading and interference. Accordingly, the duration of TAS windows must be calculated statistically to guarantee that the vast majority of packets can be transmitted within the limits of the designed windows. However, in critical applications requiring ultra-low packet loss rates, transmission windows must be long enough to transmit packets using the most robust modulation and coding scheme (MCS). While these conservative MCS configurations significantly reduce error probability, they also require longer transmission times due to increased redundancy and lower data rates. Consequently, the required frame duration, which is the sum of the duration of the TAS windows of all active flows, can be notably larger than in wired links, thus increasing the achievable control cycle in the case of industrial remote control applications.

In this paper, we propose a coordinated scheduling approach across multiple base stations to provide diversity for mobile terminals positioned within the overlapping coverage areas. By increasing diversity in wireless channels, we make performance more predictable, thereby enhancing network determinism. In the proposed approach, every TAS window in the shared radio frame is assigned to the base station with the best propagation channel. The other base stations cooperate with the serving BS in two different ways. They can remain silent to avoid interference or transmit the same packet to improve the quality of the received signal. The result is that windows can be shorter in time (reduced latency), and the aggregated coverage is extended compared to the uncoordinated case. While the proposed approach is broadly applicable across wireless technologies, we focus our analysis on 5G New Radio (NR) networks in industrial settings. Simulations have

Work supported by projects MAYTE (PID2022-136512OB-C21) and 6-SENSES (PID2022-138648OB-I00) funded by MCIN/AEI/10.13039/501100011033 and “ERDF A way of making Europe”, EU), PREDICT-6G (grant 101095890) by EU’s Horizon Europe, TIMING (TSI-063000-2021-145) by MINECO UNICO5G and, grants 2021 SGR 01033 and 2021 SGR 00772 by the Catalan Government (AGAUR).

revealed that coordination is required to support time-sensitive communications in the studied interference-limited scenarios.

II. Problem statement

We consider a system of M single-antenna base stations (gNBs in 5G NR terminology) operating with full frequency reuse across the same frequency band. Their coverage areas partially overlap to guarantee full coverage. A variable number N of single-antenna terminals is deployed randomly around the M gNBs. The channel impulse response between the mth gNB and the nth terminal at time t is denoted by $h_{m,n}(\tau, t)$ and accounts for pathloss, shadowing (slow fading), and multipath (fast fading). We consider a low-mobility scenario in which the coherence time of fading (T_{coh}) is much longer than the radio frame and, therefore, $h_{m,n}(\tau, t)$ varies enough slowly in time (t) to be estimated and tracked accurately.

Following the 5G NR standard [7], a resource block (RB) comprises 12 consecutive subcarriers over one OFDM symbol duration (T seconds), representing the minimum scheduling unit in our analysis. We denote by K the total number of RBs available for uplink (UL) and downlink (DL) transmissions. The OFDM symbol duration is given by $T = 1/\Delta f + T_{CP}$, where Δf is the subcarrier spacing and T_{CP} is the cyclic prefix duration, both determined by the adopted numerology parameter $\mu = 0, 1, 2, 3$. We assume hereinafter that the cyclic prefix exceeds the duration of the channel impulse response $h_{m,n}(\tau, t)$ for all gNB-user pairs (m, n) and time instances t. The duration of a subframe is 1 millisecond and the number of available RBs per subframe is specified in the standard as $2^{\mu}14K - S_{RB}$, where $2^{\mu}14$ represents the number of OFDM symbols per subframe and S_{RB} denotes the number of RBs for signaling.

Following the philosophy of the IEEE 802.1Qbv TSN standard, some RBs can be periodically reserved to transmit the periodic traffic of active time-sensitive (TS) flows. The integration of IEEE 802.1Qbv into 5G NR is viable because some semi-static scheduling mechanisms are already contemplated in 5G NR to reduce the overhead and latency of dynamic scheduling algorithms [7]. For simplicity, we assume that each terminal establishes a unique bidirectional periodic TS flow. Also, we assume that all TS flows consist of a sequence of packets (transport blocks) of constant size (T_{BS} bits) generated every T_{cycle} seconds, which is assumed to coincide with the NR subframe duration (1 ms) just for simplicity. In the 5G's context, T_{BS} corresponds to one of the admitted transport block sizes (TBS) in the standard [8].

Assuming that IEEE 802.1as is employed to synchronize all network nodes, including the data source and destination nodes, a TSN controller is responsible for programming a dedicated transmission window for each TS flow in all nodes along the route from the source to the destination. In particular, the gNB node is configured to reserve a dedicated transmission window for every active TS flow in every subframe that crosses the radio link. The programmed windows open precisely when the entire packet has arrived at the node, thereby minimizing the delay in queues.

The number of RBs that must be reserved in every window depends on the selected modulation and coding scheme (MCS). If R_i stands for the spectral efficiency associated with the ith MCS, expressed in bits per channel use, the number of required RBs depends on the adopted MCS index, i, as follows:

$$N_{RB}^{(i)} = \lceil T_{BS}/(12R_i) \rceil \tag{1}$$

For short packets and the fastest MCS values, it is possible to have multiple windows multiplexed in frequency within the same symbol, provided that $N_{RB}^{(i)}$ is lower than K. Otherwise, windows will span over multiple symbols.

The objective of this paper is to minimize the number of RBs required to multiplex packets from all active TS flows in each subframe, thereby maximizing the available RBs for non-TS flows. At the same time, windows should be as short as possible to minimize latency and as stable as possible to minimize jitter (delay variability). Note that jitter can always be reduced or even eliminated by voluntarily delaying packets at the receiving node to force a constant delay (also referred to as hold-and-forward in [9]). Consequently, a jitter-optimal strategy is to forward packets to the next node as fast as possible and let this node apply hold-and-forward, if required.

Unfortunately, the received signal varies in time due to fading and interference, and, consequently, the required number of resource blocks changes randomly in time. If channel statistics (i.e., fading and interference) are estimated accurately, then we can estimate the probability p_i of supporting the MCS index i. These probabilities will remain practically constant during the whole communication, only changing in the long term as the terminal moves and its pathloss changes. Accordingly, the number of reserved RBs for a given flow can be adjusted statistically in advance to $N_{RB}^{*} \equiv N_{RB}^{(i^*)}$ with i^* the greatest MCS index holding that

$$\sum_{i=0}^{i^*-1} p_i \leq p_{loss} \tag{2}$$

being p_{loss} a small value that fixes the probability of losing a packet because it does not fit in the reserved window. During transmission, if the MCS of the TS flows is adjusted dynamically to the channel state, the reserved window will not be fully occupied in those subframes where the index of the selected MCS is larger than i^*. In this case, unoccupied RBs can be used to allocate non-TS flows exchanged with the same user equipment, improving overall throughput.

III. Coordinated windows design

We consider that the M gNBs are coordinated to define a common radio frame with one gNB acting as the master. In each 1-millisecond subframe, periodic windows are programmed to transmit packets for the N active TS flows. While this work focuses on the downlink, the approach can be extended to uplink scenarios with appropriate adjustments. We consider two types of windows:

- **Uncoordinated Windows**: multiple gNBs are enabled to overlap their reserved windows provided that the links are sufficiently distant not to interfere with each other. Spatial reuse is leveraged through these windows, thereby increasing system throughput. In this case, gNBs must optimize their transmitted power to control interference and maximize the achieved rate.

- **Coordinated Windows**: multiple gNBs are coordinated to guarantee connectivity to terminals that are interference-limited. Two options are considered: selection and replication. In the first case, the terminal is assigned to the gNB offering the best propagation channel while switching off the other gNBs to avoid interference. In the second case, several gNBs transmit simultaneously the same message, generating additional multipath components that should be absorbed by the cyclic prefix. In both cases (selection and replication), the number of RBs required in a coordinated window is expected to be reduced because of the absence of interference and the use of spatial diversity.

A criterion is needed to determine the appropriate window type for each active TS flow. The adopted approach selects the option that maximizes the number of information bits that can be multiplexed per RB. Using this criterion, simulations conducted in reference industrial scenarios [10] have shown that interference is the limiting factor and some form of coordination is required. Simulations also reveal that selecting the best gNB for transmission is preferable to allowing all gNBs to replicate the same signal. Focusing then on this coordination strategy, the probabilities p_i associated with each active TS flow must be obtained before deciding the number of RBs reserved for this flow. The computation of p_i for the nth terminal follows the following steps:

- Compute the SNR received by the nth terminal from every gNB in its surroundings ($m = 1, ..., M$) and for all the subcarriers in use ($k \in \mathcal{K}$):

$$\mathrm{SNR}_{m,n}(k,j) = \frac{E_s}{N_0}|H_{m,n}(k,j)|^2 \quad (3)$$

  with E_s the transmitted (average) energy per symbol, N_0 the noise single-sided power spectral density and

$$H_{m,n}(k,j) = \sum_i h_{m,n}(iT_s, jT_{coh})e^{-j2\pi\frac{k}{N_{FFT}}i} \quad (4)$$

  the N_{FFT}-points DFT of $h_{m,n}(\tau, t)$ sampled in τ and t every $T_s = T/N_{FFT}$ and T_{coh} (coherence time) seconds, respectively. The DFT size (N_{FFT}) depends on the adopted bandwidth.
- Compute the effective SNR ($\mathrm{SNR}^{\mathrm{eff}}$). When channel coding is performed across multiple subcarriers, the effective SNR provides an accurate characterization of the performance of an OFDM-based physical layer under frequency-selective channels [11][12]. The calculation of $\mathrm{SNR}^{\mathrm{eff}}$ depends on the adopted modulation, which changes in NR form QPSK to 1024-QAM depending on the selected MCS. Considering the ith MCS, we obtain

$$\mathrm{SNR}^{\mathrm{eff}}_{m,n}(j) = f_i^{-1}\left(\frac{1}{|\mathcal{K}|}\sum_{k\in\mathcal{K}} f_i\left(\mathrm{SNR}_{m,n}(k,j)\right)\right) \quad (5)$$

  with $f_i(\mathrm{SNR})$ the received bit mutual information rate (RBIR) function associated with the modulation format used by the ith MCS. Note that, in general, $f_i(\mathrm{SNR})$ and its inverse must be evaluated numerically.
- Select the gNB offering the highest effective SNR. The other gNBs are switched off during this window to avoid interference.

$$\mathrm{SNR}^{\mathrm{eff}}_{n}(j) = \max_m \mathrm{SNR}^{\mathrm{eff}}_{m,n}(j) \quad (6)$$

- Start with the fastest MCS ($i = i_{max}$) and check that this MCS can be supported with the obtained $\mathrm{SNR}^{\mathrm{eff}}$. To do so, compare $\mathrm{SNR}^{\mathrm{eff}}_{n}(j)$ with the SNR that the ith MCS requires to achieve a given block error rate (BLER)[1]. If $\mathrm{SNR}^{\mathrm{eff}}_{n}(j)$ is not enough, reduce the MCS index ($i \to i-1$) and compute again $\mathrm{SNR}^{\mathrm{eff}}_{n}(j)$ for the new MCS. Continue this process repeatedly until you find the highest MCS index that satisfies the BLER requirement. Call this MCS, $\mathrm{MCS}_n(j)$.
- Repeat the entire procedure J times starting from statistically independent channel realization $h_{m,n}(\tau, jT_{coh})$. Then, collect the resulting J values into a vector:

$$\mathbf{MCS}_n = [\mathrm{MCS}_n(1), ..., \mathrm{MCS}_n(J)] \quad (7)$$

- Compute the histogram of $\mathrm{MCS}_n(j)$ from vector $\mathbf{MCS}_n$. Then, for J sufficiently large, the probabilities p_i can be estimated as follows:

$$p_i = \frac{1}{J}\sum_j \mathbb{I}_{\mathrm{MCS}_n(j)=i} \quad (8)$$

  with $\mathbb{I}_x$ equal to 1 if x is true and zero otherwise.
- Determine the number of RBs that must be reserved for the nth terminal $N^*_{RB}(n)$. Note that $N^*_{RB}(n)$ corresponds to the number of resource blocks when using the highest MCS index, i^*, that fulfills (2) for terminal n.
- Finally, compute the total number of RBs that must be reserved in every subframe to multiplex the N active TS flows:

$$N_{RB} = \sum_{n=1}^{N} N^*_{RB}(n) \quad (9)$$

IV. Simulation Results

In this section, the statistical window design procedure described in the last section is tested in a realistic industrial 5G deployment. We consider a factory building with dimensions 100 m (length) × 100 m (width) × 10 m (height), where four gNBs are mounted 8 m above the floor, positioned 25 m away from the walls, and spaced 50 m apart from each other. This deployment reproduces, in essence, the indoor factory scenario InF-SH reported in [10] (Sec. 7.8.4). As detailed in [10], InF-SH is a non-line of sight (NLOS) scenario with sparse clutter and high base station height, whose main characteristics are listed in Table I. The wireless channel is simulated in MATLAB

[1]The target SNR for each of the 26 MCSs defined in the standard (with LDPC encoding) [13] was obtained by the authors through Monte Carlo simulations using MATLAB (5G Toolbox) [14].

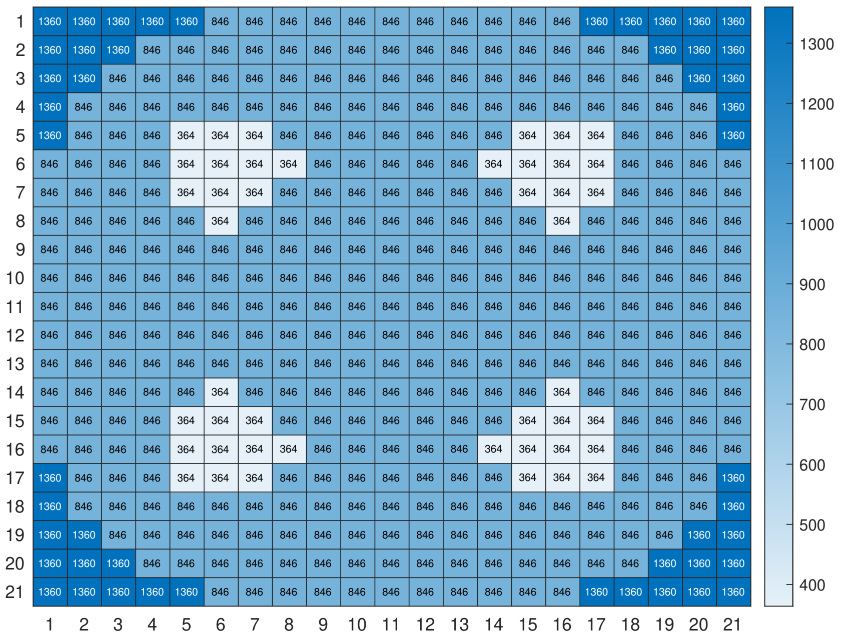

Fig. 1. NRB map when the transmitted E_s/N_0 is set to 87.7 dB.

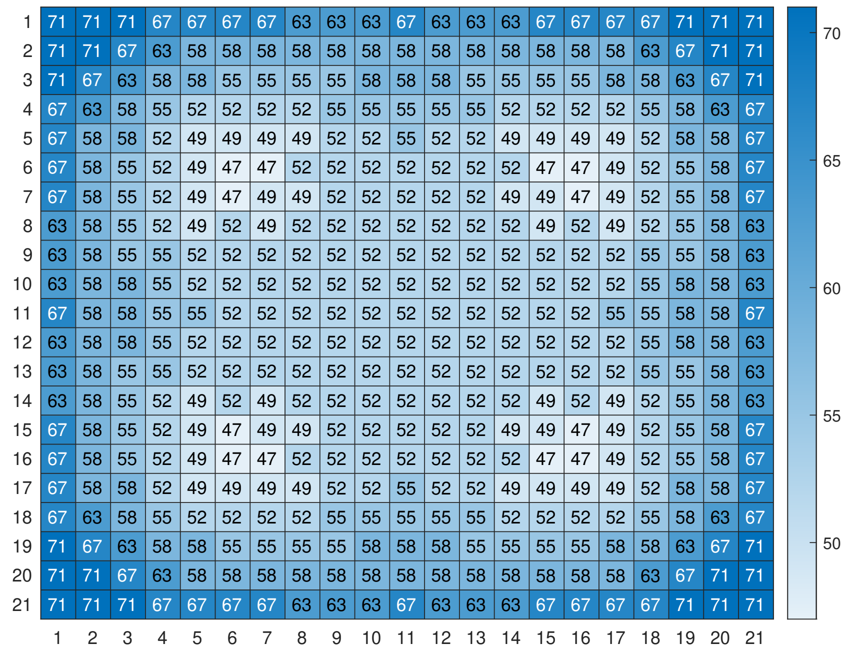

Fig. 2. NRB map when the transmitted E_s/N_0 is set to 112.4 dB.

(R2024a) using the `nrTDLChannel` object included in the 5G Toolbox (End-to-End Simulation section). The probability of overflowing the designed window is set to $p_{loss} = 5 \cdot 10^{-4}$. Regarding the 5G NR network, it is configured as indicated in Table II.

TABLE I
PROPAGATION CHANNEL PARAMETERS

Parameter	Value
Frequency (f)	3.5 GHz
Multipath model	TDL-B (Rayleigh taps)
Delay spread	57 ns
Shadowing (std)	5.9 dB (log-normal)
Pathloss model	$32.4 + 23\log_{10}(d) + 20\log_{10}(f)$
User terminal height	1.5 m

TABLE II
5G NR CONFIGURATION

Parameter	Value
Transmitted E_s/N_0 (dB)	87.7 - 130.9 dB
Bandwidth	50 MHz
Subframe duration	1 ms
Subcarrier spacing (Δf)	15 kHz ($\mu = 0$)
Symbol duration including CP (T)	71.4 μs
Transport block size (TBS)	3780 bits
Block Error Rate (BLER)	$5 \cdot 10^{-4}$
Number of RBs (NRB) per symbol (K)	270
Number of RBs (NRB) per subframe ($N_{RB,\max}$)	3600
Number of RBs (NRB) for signaling (S_{RB})	180 (approx. 5%)

The size of each user window, $N_{RB}^*(n)$, depends on the position of the terminal in the factory hall. Figs. 1 and 2 show a pair of maps illustrating the number of resource blocks required at different terminal positions across the hall. Terminal locations are discretized with a spatial resolution of 5 m. The first map in Fig. 1 corresponds to the case of transmitting the minimum power to provide coverage to all the locations on the map (i.e., MCS ≥ 0). In the second map (Fig. 2), the transmitted power is increased to boost throughput and, in this way, reduce the size of the windows. It can be shown that the window size is minimum for those terminals near the gNBs and maximum in the corners of the hall. The window size does not increase excessively at the center of the hall because the best gNB is always selected to exploit positively fading (spatial diversity) and maximize the received SNR. Although not plotted, the uncoordinated case was also simulated, showing that even with optimized transmit power, communication is only feasible in the immediate vicinity of the gNBs, and only at the lowest modulation and coding schemes (MCS 0 or 1).

To conclude, in Fig. 3, we have computed the total number of resource blocks N_{RB} (9) that are required to sustain N TS flows for different values of the transmit E_s/N_0 and assuming that terminals are deployed randomly on the map. Due to the random location of terminals, N_{RB} is random, and in Fig. 3, we have represented the average value of N_{RB} and, with dashed lines, $\pm$ 3 times its standard deviation. In Fig. 3, N_{RB} is compared with the maximum number of available RBs in one subframe ($N_{RB,\max}$), which is 3600 according to table II. It can be shown that the larger the number of TS flows, the higher the transmit power must be. In the limit, if the transmit E_s/N_0 is larger than 129.6 dB, the minimum window size ($N_{RB} = 35$) can be configured regardless of the terminal's location. Therefore, the maximum number of users that can be supported increases up to $N_{max} = \lfloor 3600/35 \rfloor = 102$. The above statements are clearly illustrated in Fig. 4, which shows the transmit E_s/N_0 required to support N TS flows for N ranging from 1 to $N_{max} = 102$.

A maximum delay of $D_{max}(n) = \lceil \frac{N_{RB}^*(n)}{270} \rceil T$ seconds is guaranteed to the nth TS flow as long as its position does not change significantly during the flow's lifespan. This maximum delay corresponds to its window duration. Depending on the terminal location and transmit power, $D_{max}(n)$ ranges from T to $6T$ symbols. Should the application impose a latency reduction, one can increase the transmit power or reduce T by increasing the bandwidth. On the other hand, all TS flows experience zero jitter, except when the reserved window needs to be reconfigured during the connection due to significant

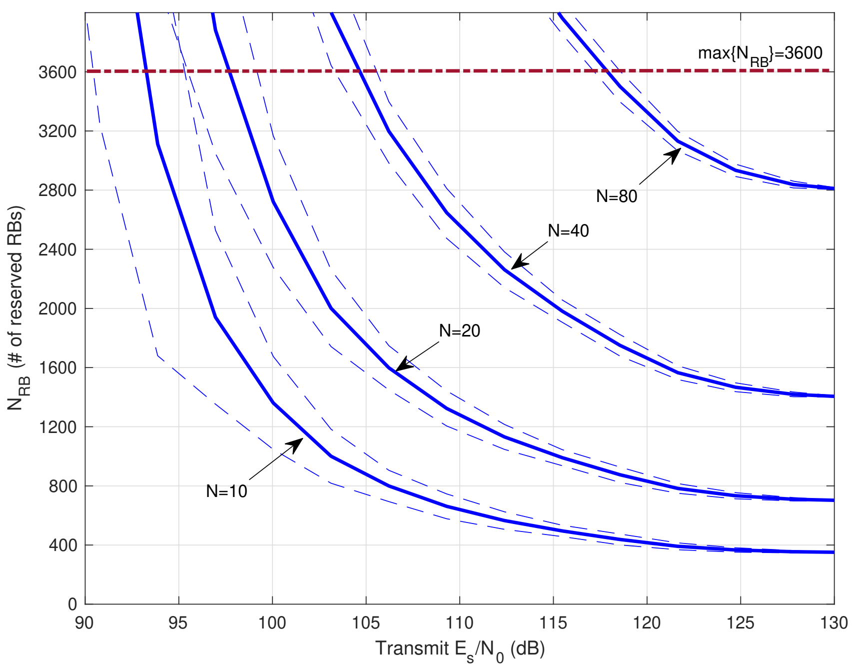

Fig. 3. Number of RBs per subframe as a function of the transmitted E_s/N_0 for $N = 10, 20, 40, 80$ user terminals.

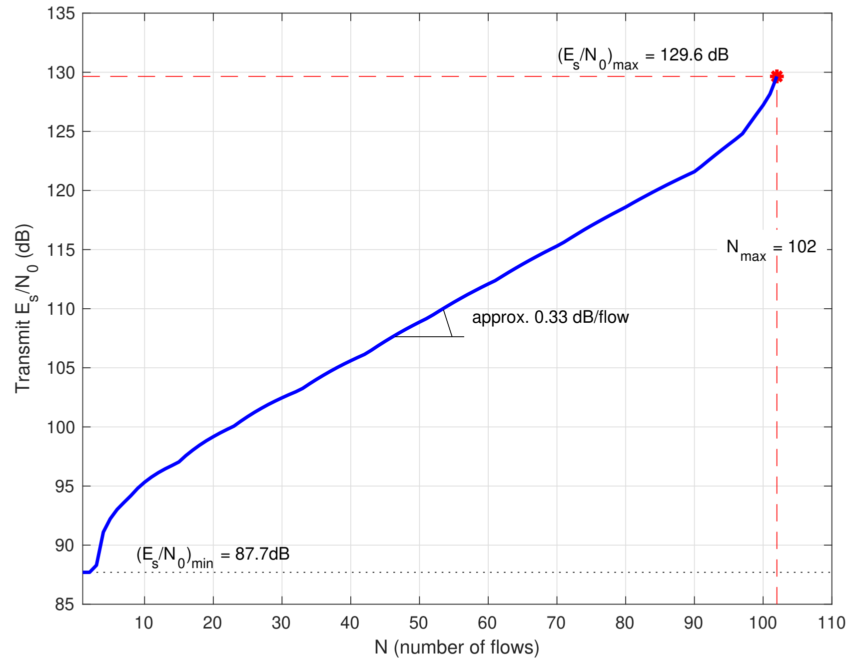

Fig. 4. Required E_s/N_0 (at transmission) as a function of the number of active TS flows. This curve is derived from Fig. 3 by identifying the point where the average N_{RB} plus three times its standard deviation intersects the number of available RBs ($N_{RB,\max} = 3600$).

terminal movement. Although reconfiguring windows in high-mobility scenarios is of great importance, it is not addressed in this paper and is left for future work.

V. Conclusions

Despite link adaptation and dynamic scheduling are generally beneficial to increase throughput, their intrinsic randomness have a negative impact on time-sensitive (TS) communications, which call for deterministic latency, with bounded delay and zero jitter. A successful TSN-based approach for stabilizing latency is to reserve periodic resources in each radio frame, enabling packets to be forwarded immediately upon arrival at the transmitter without being queued. In this paper we have developed a window reservation strategy for 5G networks that is compatible with the IEEE TSN 802.1Qbv standard. The number of time-frequency resources N_{RB} that are required to sustain a given number of TS data flows is determined concluding that: 1) interference constitutes the main limiting factor in industrial multi-gNB deployments, making coordination essential to cope with interference and exploit spatial diversity; 2) the number of resource blocks, N_{RB}, depends on the transmit power such that an increase in TS flows requires a corresponding increase in transmit power. By coordinating transmissions, we show that 5G NR can support a substantial number of TS connections, even under the baseline setup evaluated in this paper. Further improvements in the number of admissible TS flows and achievable latency are foreseen by reducing the transport block size or adopting higher numerologies with wider bandwidths, both at the expense of increased transmit power.

References

[1] "Time Sensitive Networks for Flexible Manufacturing Testbed Characterization and Mapping of Converged Traffic Types", Industrial Internet Consortium White Paper, Mar. 2019.

[2] Z. Maletic, M. Mladen and M. Ljubojevic, "A Survey on the Current State of Time-sensitive Networks Standardization," 2023 10th International Conference on Electrical, Electronic and Computing Engineering (IcETRAN), East Sarajevo, Bosnia and Herzegovina, 2023, pp. 1-6

[3] "IEEE Standard for Local and Metropolitan Area Networks–Timing and Synchronization for Time-Sensitive Applications," in IEEE Std 802.1AS-2020 (Revision of IEEE Std 802.1AS-2011) , vol., no., pp.1-421, 19 June 2020.

[4] "IEEE Standard for Local and Metropolitan Area Networks–Bridges and Bridged Networks," in IEEE Std 802.1Q-2022 (Revision of IEEE Std 802.1Q-2018) , vol., no., pp.1-2163, 22 Dec. 2022.

[5] "Integration of 5G with Time-Sensitive Networking for Industrial Communications", 5G-ACIA White Paper, Jan. 2021.

[6] Harmatos,J.; Maliosz,M. Architecture Integration of 5G Networks and Time-Sensitive Networking with Edge Computing for Smart Manufacturing. Electronics 2021,10,3085.

[7] E. Dahlman, S. Parkvall, and J. Sköld, 5G NR: The Next Generation Wireless Access Technology, 2nd ed. Academic Press, 2020.

[8] ETSI TS 38.214 V18.5.0, "5G; NR; Physical layer procedures for data" (3GPP TS 38.214 version 18.5.0 Release 18), Jan 2025.

[9] ETSI TS 123 501 V17.4.0 "5G; System architecture for the 5G System (5GS)" (3GPP TS 23.501 version 17.10.0 Release 17), Sep. 2023.

[10] 3GPP TR 38.901, "Study on channel model for frequencies from 0.5 to 100 GHz (Release 18)", V18.0.0, May 2024.

[11] "Physical Layer Abstraction for System-Level Simulation". WLAN Toolbox Documentation. Matlab R2021.

[12] Kim, H., Seol, K. and Kim, J. (2024), An optimal calibration factor for mutual information effective SINR mapping scheme in 5G NR. Electron. Lett., 60.

[13] ETSI TS 138 212 V18.2.0, "5G; NR; Multiplexing and channel coding" (3GPP TS 38.212 version 18.2.0 Release 18), May 2024.

[14] "5G LDPC Block Error Rate Simulation Using the Cloud or a Cluster". 5G Toolbox Documentation. Matlab R2024a.