5G Toolbox for Realizing Industrial Automation

Abdulrahman Alabbasi, Torsten Dudda, Zhenhua Zou, Jonas Kronander
Ericsson Research
(firstname.secondname@ericsson.com)

Abstract—Industry 4.0, a subset of the fourth industrial revolution, derives industrial automation for manufacturing sectors by setting targets and mechanisms for industrial data exchange. 5G system is considered as a communication platform that provides flexibility and wireless connectivity to support industrial automation. However, several issues are to be addressed in order for 5G NR to enable industrial automation, such as, efficient multiplexing between industrial flows i.e. maintaining QoS for critical traffic while enhancing efficiency for other traffic. In this paper, we present our view on how to solve several relevant issues, from medium access control (MAC) layer perspective. The main advantage of our solutions is that it requires very low complexity resulting in minimal changes to existing standardization specification.

Keywords— coexistence-of-services, factory automation, 5G

I. INTRODUCTION & RELATED WORK

5G is considered as a key candidate to support Industry 4.0, via enabling TSN (Time Sensitive Network, which is the envisaged communication standard based on IEEE to support critical communication) protocol over wireless connection [1], [2],[4]. This is motivated due to the fact that 5G can support multiple industrial use-cases under a unified network (5G system) umbrella. Such support requires two essential points: 1) Guarantee that 5G can support the targeted quality of services (QoS) for the heterogeneous classes of industrial traffic. Such different classes might contain both critical and non-critical traffic, hence require the support of ultra-reliable low latency communication (URLLC) on top of ordinary mobile broad band (MBB) traffic. 2) Guarantee that 5G is fully integrated with TSN standard.

Addressing URLLC requirements was one of the main challenges of NR (5G) standardization in 3GPP [9]. Several techniques in PHY, MAC, and upper layers have been introduced to support URLLC. For instance, new modulation and coding schemes (MCS) tables, puncturing (or pre-empting) of existing MBB transmission by newly arrived URLLC packets, higher numerologies (hence shorter slot duration), mini-slot based transmission, reduced processing time for hybrid automatic repeat request (HARQ) and scheduling request (SR) to allocated UL resource operation, configured grant transmission, and protocol data unit (PDU) duplication were some introduced features in the standard to meet the URLLC QoS of the ITU requirements [3][4].

Coexistence between critical and non-critical (e.g. MBB) traffic is an interesting topic for academia, it is also perceived as a practical case that often occurs in industrial automation use-cases. For instance, authors of [11] have proposed a joint link adaptation and resource allocation policy that adapt block error probability of URLLC to meet 99.999% reliability for a latency of 1 – 1.3 ms while maintaining less than 10% degradation in MBB throughput performance. Authors of [12] have investigated dynamic multiplexing between both services using queuing analysis and showed significant improvement to the resource efficiency of the wireless system. Authors of [13] investigated useful mechanisms for URLLC traffic to puncture MBB. They proposed a recovery mechanism of the of punctured MBB transmission. A null-space-based spatial preemptive scheduler for joint URLLC and MBB traffic is proposed and analyzed for dense 5G networks in [15]. A cross-objective optimization problem is formulated to guarantee URLLC QoS while maximize MBB’s ergodic capacity.

We now introduce common terminology related to MAC that will be used in the paper. Most of such terms and functionalities can be found from 3GPP specification document [5]. MAC layer has several functionalities, we describe the related ones here: 1) Mapping between LCHs and transport channels. 2) Multiplexing (de-) of MAC SDU (service data unit). 3) LCH prioritization.

Multiplexing of MAC SDUs and other MAC control elements (CEs) to form the MAC PDU depends on several sub-process in MAC, among which are the LCH prioritization (LCP), LCP restriction process, and the resource allocation process.

Each LCH has a priority, configured by radio resource control layer (RRC), that indicate the order of the LCHs to be selected to fill in the MAC PDUs. Such that SDUs from LCH with higher priority fill into the PDU before lower priority ones (depending on the PDU’s size there might not be space left for low priority LCH’s SDUs).

LCP restriction process decides whether the associated LCH matches a certain characteristic of the UL grant. Such target characteristic is set by the restrictions. For instance, the “maxPUSCH-Duration” restriction sets the maximum grant duration allowed for transmission of the associated LCH.

In this work, we focus on supporting UL mixed-services scenario. We address this issue from 3GPP RAN2 perspective, while aiming at low complexity solutions that do not require much change to the standard. Low complexity and small changes to exiting standard are important features when solving wireless problems, because it guarantees backward compatibility and higher adoption by network operation to such industrial automation feature. To achieve such objective, it is essential to enable the network to allocate overlapping UL resources and enable UEs to handle such collision in the best way. NR (new radio access technology) Rel.15 does not handle such overlapping resources in the desired mechanism. Therefore, we propose, in Rel.16, to enable the network to allocate overlapping UL resources and propose near optimal UE behavior (once receiving such overlapping) to enhance spectral efficiency while maintaining critical requirement based on which data has just arrived at the MAC buffer.

II. USE-CASES AND TRAFFIC ASSUMPTIONS

As described above, cyber-physical production systems are based on wireless connectivity and computation. From Industry 4.0 perspective several infrastructures that interconnect people, products, and machines in a flexible, secure, and consistent manner. There are several applications that describe such area, such as, Process automation, HMIs
and production IT, Logistics and warehousing, and Monitoring and maintenance [6]. However, in this paper we focus on Factory Automation, which addresses the automated control, monitoring and optimization of processes and workflows within a factory.

Based on the above application areas several traffic types, based on traffic periodicity, survival time, reliability and latency targets, are recognized as follows [6]:

**Periodic deterministic traffic** is periodic with stringent requirements on timeliness and availability of the communication service. A transmission occurs every transfer interval \( P \), which identifies the periodicity of traffic. An extreme example of this traffic type is the motion control [6], which targets a service availability up to 99.9999%, end-to-end latency of less than 500 µs, transfer interval (periodicity) is about 500 µs.

**Aperiodic deterministic traffic** is known for non-preset transfer time, it has stringent reliability and latency targets. **Non-deterministic traffic** consists of all other traffic types other than deterministic traffic. Also, it is not characterized by periodical transfer interval, MBB such as video streaming can be of this traffic type.

**Mixed traffic** can be a mixture of any or all the above traffic types originated from/to a single UE (called Intra-UE) or form/to multiple UEs (called Inter-UE).

**TSN flow** can also be considered as a mixed traffic flow due to the fact that TSN streams carries deterministic or non-deterministic, periodic or aperiodic service flows.

### III. Proposed Scheduling Tools for Industrial Automation

In this section, we illustrate our proposed solution to enable Intra-UE grant prioritization that support industrial automation. We begin by explaining the overall proposed solution. It is followed by defining related sub-problems and the associated solution that complete the picture of how 5G could support all aspects of industrial traffic.

#### A. General Proposed Solution

One of the main solutions proposed for enabling higher spectral efficiency while meeting industrial traffic critical requirements is to enable overlapping grants. In this section, we present our proposed algorithm that describes the solution for intra-UE grant multiplexing/prioritization for mixed traffic scenario, in Figure 1. We assume that gNB send overlapping grants to UE to efficiently handle the mixed traffic case. For each reception of a grant, the UE begins by checking whether **overlapping of grants exist or not**. We then decide upon **availability of restricted data** (means that the data after applying LCP restriction is available), in each or set of LCHs associated with each grant. If restricted data is available for a single grant, no overlapping handling is needed, hence we continue with conventional MAC procedures. Otherwise, MAC checks if the case is about overlapping between two dynamic grants. MAC layer in the UE decides to build two MAC PDUs and sends both MAC PDUs to PHY layer with the understanding that the later grant has a higher transmission profile index (TPI). TPI of a grant is used to indicate which LCH IDs (group of LCHs) or LCH priority is allowed (based on LCP procedures in [7] clause 5.4.3.1) to be sent on such grant’s resources with this transmission profile index, more on TPI can be found in [8]. PHY layer can perform cancelation if any of the transmissions has not started (subject to other processing timing limitation) or preemption if one transmission has started or cannot be stopped. The term ‘Later grant’ indicates the timing of DCI reception. This reflects that the gNB issuing the dynamic grants is aware of the impact of sending both grants and that sending the later grant in all cases as intentionally done by the gNB, i.e. the selection of the later grant is intended and should not be conditional. If the prioritization is among grants, with at least a single configured grant, the following is considered. If the MAC PDU of a previous grant has not been assembled by MAC, the UE can select to construct a MAC PDU for any of the previous or newer grants. Such MAC PDU assembly is according to the selected grant. The grant shall be selected based on which has the highest priority LCH that is mapped to the grant’s TPI and have available data for transmission (filtered according to LCH restrictions). In case of similar grants’ priorities, 1) the grant with the highest transport block size (TBS) or 2) the later grant sent by gNB can be selected. Otherwise, if the MAC PDU of a previous grant had been assembled by MAC already, the UE should start **Pre-emption procedure (involving configured grant)**. That is, when the MAC PDU of a previous grant has been assembled by MAC, UE may or may not override parts of the resources used for the MAC PDU / TB according to the original grant, with a newly constructed MAC PDU / TB according to the new/selected grant. The **UE may decide to prepare new MAC PDU** base on the later grant, and send it to PHY to pre-empt existing grant, in which case the UE might multiplex the pre-empted data and MAC CE in the HARQ buffer of the pre-empted grant into a new grant. Or, the **UE may ignore the later grant and does not prepare new MAC PDU**. The decision is based on the comparison of the TPI and the available restricted data in mapped LCH to each TPI of the pre-empted and possibility pre-empting grants.

After the above procedures, the selected (or preempts) grant is considered in the HARQ entity, which obtain the

![Figure 1. General algorithm that addresses UL overlapping grant issues for mixed services.](image)
MAC PDU from the Assembly and Multiplexing Entity. This entity constructs the MAC PDU based on the selected (decided on) grant’s TPI, conventional LCP and restrictions rules, and the reliability restriction. After obtaining the MAC PDU, the HARQ entity pass it to the HARQ process.

B. Enhancing TSN scheduling

In this section, we introduce some detailed sub-problems and the associated solutions to enhance TSN scheduling in cooperation with the above general solution to complete the full picture of 5G support for industrial automation. Figure 2. Industrial deterministic flows with different pattern, periodicity, latency & reliability requirements

In order to enable the UE to support multiple TSN flows with different characteristics, multiple pre-configured transmission occasions with different settings, e.g. periodicity, time offset, frequency resources, MCS index, etc., per single UE should be supported. For instance, illustrated in Figure 2, the red flow can be transmitted on configuration 1, the dotted flow can be transmitted on configuration 2, and the yellow flow can be transmitted on configuration 3. Each of configurations 1, 2, and 3, has different offset from the beginning of the slot/frame. Also, the time duration, and allocated resources, can be different. Additionally, other parameters, such as periodicity, MCS indices, and repetitions, could be different to fit the requirement of the industrial (TSN) flow. Therefore, with such multiple configurations the network can satisfy the QoS requirements (that can be translated into RAN parameter, i.e., periodicity, time duration, frequency resources, MCS index, and repetition) for all industrial flows per TSN node (UE). Hence, enabling multiple configurations of configured grant and semi-persistent scheduling is essential to support multiple TSN flows. However, what if a TSN stream arrives earlier or later than expected? Then the data of that stream will be allocated fully or partly in a non-intended configuration. In order to avoid such event, and to guarantee consistent scheduling, we propose to have a new LCP restriction that links a single or a set of configurations to a single or a set of LCHs.

Another issue is the delay induced by the arrival mis-alignment of data at the transmitter side. This occurs if data at application layer was generated earlier/later than expected (or other reasons like network congestions), hence TSN data arrival at MAC/PHY will be mis-aligned with the granted resources. Therefore, gNB could allocate multiple configurations CG each with shifted starting time compared to the other configuration, as illustrated in Figure 3. Such that if the arrived data missed the time required for preparation to configuration 1, it can be sent over configuration 2 or 3, i.e., the closest possible configuration to its arrival. Hence, UE can choose the configuration that suite the data arrival, such solution is already covered by the general solution described in section III.A.

Adjusting the configured grants temporal characteristics (e.g., starting offset and periodicity) allows a closer matching of transmission occasions and TSN arrival patterns. However, assuming that the network meets the delay budget of each stream, it might over-provision resources and over-reduce the transmission latency. This might result in an early arrival of the data at the receiver side. Therefore, we propose to introduce a de-jitter buffer at the egress of the 5G system that reduces the delivery jitter at the next TSN switch. The information of the TSN flows at both the ingress and the egress points of the 5G system might be beneficial to meet the TSN requirements, i.e., providing a bounded low latency, depending on the implementation. The egress information is needed for the de-jittering function and should be provided to UPF for UL traffic. This can be obtained via the integration unit between the TSN controller and the 5G core unit. With the de-jittering function, there is no requirement on RAN to provide the same deterministic latency for every packet, but instead only that packets are scheduled efficiently i.e. only delivered within their delay budget, while jitter is allowed.

C. Tools for Enabling Fast Access to UL Channel

In here, we introduce some detailed sub-problems and its solutions, aiming at enhancing the UL channel access delay, in cooperation with the general solution (in section IIIA).

1) As illustrated in Figure 4, a BSR miss-trigger might occur when multiple critical TSN flows originated from the same industrial UE (section 6.5, [10]). For instance, assume a UE already received a sufficiently large UL grant to serve a critical flow (labeled CF1). Then, another critical traffic (labeled CF2) of the same priority arrived, it can be aperiodic or other type of traffic. In such case, UE will not trigger BSR based on following rule [7]:

- CF2 belong to an LCH with equal or lower priority than any LCH which contain available data. And,
- any of the logical channels which belong to same logical channel group (LCG), as the LCG of CF2 contains any available UL data.

Such behaviour causes the CF2 data to wait in the corresponding LCH buffer without the ability to trigger scheduling request (SR) or inform the gNB about the volume of data waiting in UE’s LCH buffer (via Regular BSR), as illustrated in Figure 4. One possible counter measure to such problem is that gNB allocate a periodic BSR reporting mechanism (for all UEs). However, we don’t consider such way because it exhausts system resources, since MAC control element (CE) BSR will always be in the MAC PDU (even if there is no data, where UE will use Padding).

To solve this issue, we propose to enable independent BSR triggering for sets of, or for all LCHs not only logical channel groups. In this case, if critical data arrived on LCH that is of lower priority than the critical LCH, which already have a grant, UE is still able to trigger BSR and initiate faster and more efficient access to the wireless channel.

2) Issues impacting the delay of informing gNB about available critical data for transmission:

a) One case occurs if a critical data (called CF2) arrived after the construction of the MAC PDU of the earliest
PUSCH, done by MAC’s Multiplexing and Assembly Entity, assuming that this MAC PDU contain a Traffic (T1). If the control channel overlaps with the grant (G1 for data channel) intended for the constructed MAC PDU, which contains T1, the specification forces the UE to not transmit the SR even if it was triggered by critical traffic [7]. If G1’s PUSCH has a long duration, then informing the gNB about waiting data (CF2) in LCH will be delayed. Figure 5 illustrates this case.

![Figure 4. Delayed BSR triggering in Rel. 15.](image)

In a follow-up case, the transmission of G1’s PUSCH might be corrupted due to high spectral efficiency with low reliability that might be intended for the NCF1. Hence, the MAC CE BSR intended for CF2 will be lost. Also, due to current MAC specification, the UE will cancel the CF2’s associated SR because BSR assumed to be triggered and transmitted successfully.

![Figure 5. Delayed informing gNB about critical data, Rel. 15.](image)

In order to solve the above issues, 2a) and 2b), we propose to allow sending SR on an overlapping UL data resource, under the condition: where the priority of the LCH that triggers the SR is strictly larger than the highest priority of the LCH(s) to be transmitted or is under transmission on the UL data resource. On top of this, we enable not cancelling the SR if the BSR MAC CE is included in the less reliable data channel that does not meet the LCP mapping restrictions that is associated the LCH which triggered the BSR/SR.

3) Another issue that might arise is the MAC SDU segmentation issue. Consider new critical data (labeled CF2) arrived while available configured grant (CG1), with limited TBS, that barely fit the intended critical data (CF1) and the associated MAC CE. If CF2 is of similar or lower priority than existing CF1, MAC will not trigger a regular BSR to inform gNB about the waiting CF2. Hence, gNB will not know about the CF2 data volume. Therefore, gNB might not be efficient in allocating the next dynamic grant to accommodate CF2. This also will result in segmenting CF2 data. MAC SDU segmentation for critical packets is not desirable because it results in delaying the reception of a complete packet. The obvious solution to such problem is that gNB always over-provision resources such that if new data arrives it always fits in the TB, without being segmented. This over-provision consumes undesirably large bandwidth.

We look at the problem from another angle, where critical SDUs are known to have small size. Hence, we argue that filling the MAC PDU with SDUs from high priority LCH instead of MAC CE (which is in the order of critical SDU size) is a preferred solution to this issue. Therefore, we propose to flexibly increase the critical LCH priority compared to MAC CE, such that such small size critical PDUs are not segmented. This solution goes inline with the independent triggering of BSR that we proposed earlier in this section.

IV. CONCLUSION

In this paper, we present our view on how the 5G system can enable industrial automation use-cases, especially enable the co-existence between MBB and critical TSN (or URLLC) types of traffic in such scenario. Enabling co-existence means that we maintain latency and reliability targets, while aiming at high spectral efficiency performance. We allow the gNB to allocate a UE with (at least) two overlapping grants, one to accommodate critical traffic and the other to accommodate MBB traffic efficiently. Upon arrival of data UE selects among the provided grants the one that satisfies the traffic requirement. The proposed mechanism enables higher spectral efficiency while maintaining URLLC requirements. Such advantage of the proposed mechanism is achieved with low complexity and minimum changes to the existing network and UE.

V. REFERENCES

[2] “Industry 4.0: the fourth industrial revolution – guide to Industrie 4.0”, url = https://www.i-scoop.eu/industry-4-0/
[9] 3GPP TR 38.824, “Study on physical layer enhancements for NR ultra-reliable and low latency case (URLLC)”, Release 16
[10] 3GPP TS 38.825, “Study on NR Industrial Internet of Things (IoT)”, Release 16