Delay Analysis of TSN Based Industrial Networks With Preemptive Traffic Using Network Calculus

Mohamed Seliem
School of Computer Science and IT
University College Cork
Cork, Ireland
m.seliem@cs.ucc.ie

Ahmed Zahran
School of Computer Science and IT
University College Cork
Cork, Ireland
a.zahran@cs.ucc.ie

Dirk Pesch
School of Computer Science and IT
University College Cork
Cork, Ireland
d.pesch@cs.ucc.ie

***Abstract*—Time-Sensitive Networking (TSN) extends traditional Ethernet to support data traffic with ultra-reliability and time-critical requirements for a range of applications in industrial automation, automotive, and aerospace. The TSN standards present guidelines to integrate different types of data traffic over a single converged network. Therefore, it is becoming an enabling technology towards the Industry 4.0 vision of integrating information and operational technologies within future Industrial Internet of Things networks. In this paper, we develop a network calculus based framework to analyse TSN based industrial networks supporting a range of data traffic classes. We apply the framework to study and analyse a well-known industrial use case, Quality Checks After Production (QCAP), with four data traffic types with different requirements in terms of reliability and end-to-end latency. In our evaluation, we validate our framework with a computer simulation model and compare the tightness of the calculated delay bounds to a state-of-the-art approach. We then use our model to analyse the upper bounds on the worst-case delay of the different QCAP traffic types and assess the factors that impact end-to-end delay, e.g. flow offset and critical links. Finally, we compare various credit accumulation rates and their impact on the traffic delay bounds.**

***Index Terms*—Time-Sensitive Networking (TSN), Industrial networks, Network Calculus (NC), IIoT, Deterministic Latency, Quality control.**

I. Introduction

The evolving Industrial Internet of Things (IIoT) paradigm envisions a diverse set of applications [1] that feature a variety of data traffic classes with distinct characteristics and quality of service (QoS) requirements [2]. Among them, control and safety message traffic flows must meet tight deadlines to ensure safe operation of automation systems equipment. Ensuring reliable and timely delivery of these traffic flows is challenging in future converged local area networks, where a single network will carry both information technology and operational technology traffic. Hence, Time Sensitive Networking (TSN) [3] has evolved as a group of standards that enable the deployment of IIoT applications over converged local area networks.

This publication has emanated from research conducted with the financial support of Science Foundation Ireland under Grant number 16/RC/3918.

TSN standards define various mechanisms to provide delivery guarantees for time critical data flows while ensuring fair delivery of lower priority data flows, e.g. non-critical and best effort. Beyond node synchronization (IEEE 802.1AS [4]), traffic shaping and scheduling (e.g., IEEE 802.1Qbv and IEEE 802.3br [5]) plays an essential role in realizing TSN's goals. Specifically, TSN standards consider time-based traffic shaping and prioritized, preemptive frame scheduling to ensure reliable and timely delivery of critical traffic flows. Additionally, it employs rate-based traffic shaping to ensure deterministic behaviour of bursty traffic and to avoid starvation of low priority traffic. In order to address the needs to accommodate industrial use cases within the TSN standard, TSN profiles for different industrial applications have recently been defined in the IEC/IEEE 60802 standard [6]. This warrants the development of an analytic framework for analysing the performance bounds of these profiles to determine if TSN can meet their quality of service requirements.

Network Calculus (NC) [7] is a mathematical framework for analyzing delay and back-log (queue size) bounds of networked systems. Network calculus has been widely used in the analysis of TSN traffic shaping techniques [8]. Bondorf et al. [9] compare a number of NC-based delay analysis frameworks in regard to their bound tightness and complexity. In [10], Mohammadpour et al. propose latency bounds for coexisting control data and audio-video bridging traffic. J. Zhang et al. [11] made use of [10] to design a model based on real industrial automation use cases. They use network calculus to calculate the upper bound on latency for data traffic susceptible to Credit-Based Shaping (CBS) with Strict Priority (SP). However, the arrival curve of the scheduled traffic is modeled as a leaky bucket, which does not accurately model the scheduled control data traffic of industrial applications. Therefore, further efforts were needed to analyze diverse traffic management mechanisms defined in the TSN standards as highlighted in [12]. Zhao et al. [13] were the first to analyse two classes of audio-video bridging traffic that are impacted by scheduled traffic. Upper bounds on the credit accumulation were improved in [14] for CBS traffic. Moreover in [15], the authors analysed the

978-3-903176-57-7 © 2023 IFIP

CBS credit accumulation dynamics considering the limitations resulting from the physical link rate and the CBS output to identify a tighter latency bound for TSN. The goal of this paper is to perform timing analysis and to determine the Worst Case Delay (WCD) upper bounds for preemptive traffic in industrial networks using the most recent, enhanced NC approaches, while taking various TSN shaping strategies into account.

In this paper, we develop an NC based framework to analyse the worst-case end-to-end latency in a four traffic class TSN with frame prioritisation and preemption. The four classes represent some of the industrial applications described in [6]. To the best of our knowledge, this paper is the first to analyse the combination of frame prioritisation and preemption in the presence of time-based and rate-based shaping mechanisms. We apply our framework to analyse a Quality Control After Production (QCAP) [16] industrial use case. We validate our framework through a comparison with a computer simulation model and a state-of-the-art approach, and demonstrate that our framework produces a tighter delay bound for all traffic flows compared to the state of the art. We further use our framework to investigate the impact of link bandwidth and traffic load on delay performance in QCAP. Our evaluation shows that using an accurate arrival model for scheduled traffic, as well as considering the TSN preemption feature and the relationship between the best effort traffic and the credit accumulation, results in tighter delay bounds for all classes of traffic. Moreover, the results show a 91% reduction in the worst case end-to-end delay for scheduled traffic, when such traffic flows have different offset duration. However, this results in utilizing larger portion of the critical links available bandwidth. Finally, our analysis presents the impact of varying Audio Video Bridging (AVB) idle slopes on the WCD bound. As expected, increasing idle slopes will not affect the highest priority traffic, but impacts the other traffic classes.

The rest of the paper is organized as follows: section II presents an overview of the industrial profile of TSN and the basic concepts of network calculus. An overview of a QCAP cell components and a classification for the four types of QCAP traffic are presented in section III. In section IV, we describe the bahavior of a TSN interface and derive the worst case delay for each traffic class. In section V, we validate our framework and analyse the performance of the TSN-based QCAP use case. Finally, we conclude our work and discuss future directions in section VI.

II. Preliminaries

A. IEEE 802.1 Time Sensitive Networking

The IEEE 802.1 Working Group (WG) formed the Time Sensitive Networking (TSN) Task Group (TG) to inherit and extend the work of the IEEE 802.1 AVB TG. The TSN-TG has published a set of standards to create a guide for a TSN operating network with the following components:

Time synchronization ensures that all network nodes have the same time reference, which enables time aware data transmission and deterministic delivery. IEEE 802.1AS-2020 [4] defines the following procedures: 1) distributing synchronized time information, 2) selecting the best master clock source, and 3) identifying timing impairments events. Amended work in [17] specifies hot standby without using the Best Master Clock Algorithm (BCMA).

Bounded low latency through data flow classification based on the flow priority. Each flow is allocated a dedicated transmission slot according to a pre-determined schedule. Moreover, TSN shapes non-critical and bursty data traffic using token bucket based techniques, permitting less priority traffic to transmit while maintaining the required quality of service. The different TSN traffic scheduling and shaping techniques are defined in IEEE 802.1Q-2022 [18], which includes frame preemption, credit based shaping, enhancement for scheduled traffic, cyclic queuing and forwarding, and asynchronous traffic shaping. Moreover, [19] is an on-going amendment to recommend shaper parameter settings for bursty traffic requiring bounded latency.

High availability or ultra-reliability achieved through transmitting multiple copies of a critical frame using different paths. Upon successfully receiving a critical frame, duplicates are eliminated either at the final destination or at the final TSN switch directly before the final destination. Frame Replication and Elimination [20], Path Control and Reservation [21], Per-Stream Filtering and Policing [22] are the standards that cover the TSN reliability aspects.

Resources management uses Application Programming Interfaces (APIs) to configure the network resources to support time-critical applications according to the following information models: fully distributed user/network interface, distributed user/centralized network interface, and fully centralized user/network interface. This category includes the following [23] published standards: Stream Reservation Protocol, Link-local Registration Protocol, TSN Configuration, Foundational Bridge YANG, and YANG for CFM.

In addition, TSN-TG has ongoing projects to create and maintain different TSN profiles, which are:

- Front-haul, IEEE 802.1CM/de.
- Industrial Automation, IEC/IEEE 60802.
- Automotive In-Vehicle, IEEE 802.1DG.
- Service Provider, IEEE 802.1DF.
- Aerospace Onboard, IEEE 802.1DP/ SAE AS6675.

In this paper, we focus on the industrial automation profile, which is briefly described in the following section.

B. IEC/IEEE 60802 Industrial profile of TSN

The TSN industrial profile incorporates a hierarchical network architecture that allows merging Information Technology (IT) and industrial Operational Technology (OT) traffic, as foreseen in the Industry 4.0 vision [24]. Figure 1 represents OT functions that include devices/sensors, basic control/cells, and supervisory control/production line. Figure 2 shows an

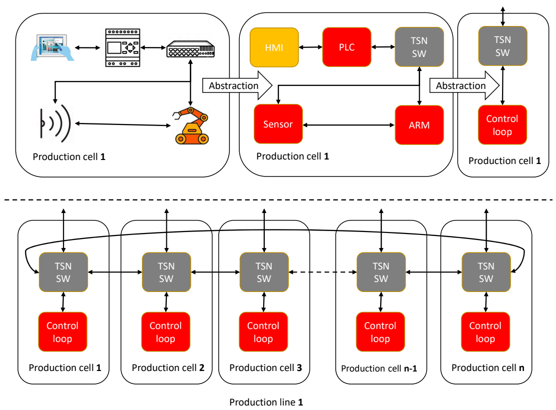

Fig. 1. Operation Technology (OT) hierarchical structure.

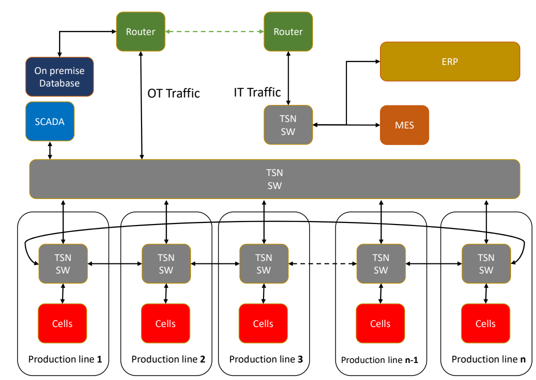

Fig. 2. Full architecture of TSN-based Industrial automation.

extension that includes the site operations/connected production lines as well as IT functions, i.e. site enterprise networks and business applications. In an industrial network based on TSN, the participating nodes abide by the rules outlined in the TSN standards. As a result, the intended industrial application could obtain TSN features and functionalities. Furthermore, the TSN industrial profile identifies the behavior and highlights the requirements for the different data flows that may exist in an industrial automation environment (see Table I).

C. Basic concepts of network calculus:

Network calculus is a theoretical framework for deterministic network system performance analysis. It is used to derive the worst-case/upper bound of data traffic delays traversing through network node(s), e.g. single/multiple hop(s). The process of data traffic arrival $R(t)$, which is defined as the input cumulative function of the accumulated bits of the data traffic, at a network node, up to time t. NC defines the arrival curve $\alpha(t)$ that constraints $R(t)$ given that for all $s \leq t$:

$$R(t) - R(s) \leq \alpha(t-s). \quad (1)$$

TABLE I
INDUSTRIAL AUTOMATION DATA TRAFFIC TYPES.

Type	Periodicity	Guarantees	Priority	Frame Size
Isochronous	Periodic (<2 ms)	Deadline	High	Fixed $30-100$B
Cyclic Real−time	Periodic ($2-20$ ms)	Latency	High	Fixed $50-100$B
Audio	Sporadic	Latency Bandwidth	High	Variable $100-1500$B
Video	Sporadic	Latency Bandwidth	Medium	Variable $100-1500$B
Network Control	Periodic (50 ms-1 s)	Bandwidth	Medium	Variable $50-500$B
Events	Sporadic	Latency	High	Variable $100-1500$B
Config & Diagnostics	Sporadic	Bandwidth	Medium	Variable $500-1500$B
Best Effort	Sporadic	None	Low	Variable $500-1500$B

The process of data traffic departure $R^*(t)$ is defined as the output cumulative function of the departed bits of the data traffic from a network node up to time t. NC defines the service curve $\beta(t)$ that models the processing service offered by this node, subject to:

$$R^*(t) \geq \inf\{R(s) + \beta(t-s)\}. \quad (2)$$

A strict service curve $\beta_{strict}(t)$ is offered upon satisfying the following expression during any backlog period $(t, t+\Delta t]$:

$$R^*(t+\Delta t) - R^*(t) \geq \beta_{strict}(\Delta t). \quad (3)$$

Another feature of using NC is that a given data traffic $R(t)$ is constrained by an arrival curve $\alpha(t)$ and while traversing a network node, is subject to a service curve $\beta(t)$. Then, the following parameters can be calculated:

- Per-hop delay $D(\alpha, \beta)$, the latency experienced by the data flow, which is the maximum horizontal deviation between $\alpha(t)$ and $\beta(t)$.

$$D(\alpha, \beta) = \sup_{s \geq 0}\{\inf\{\tau \geq 0 | \alpha(s) \leq \beta(s+\tau)\}\}. \quad (4)$$

- Per-hop buffer size $B(s)$, the maximum buffer size, which is the maximum vertical deviation between $\alpha(t)$ and $\beta(t)$.

$$B(s) = \max_{0 \leq s \leq t}\{\alpha(s) - \beta(s)\}. \quad (5)$$

- Upper bound, $\alpha^{`}(t)$, that constraints $R^*(t)$, which is the resulting arrival curve at the next hop node.

$$\alpha^{`}(t) = \sup_{s \geq 0}\{\alpha(t+s) - \beta(s)\}. \quad (6)$$

Obtaining these parameters provides insights, in terms of worst-case delay and queue size, on a deterministic service offered by the network.

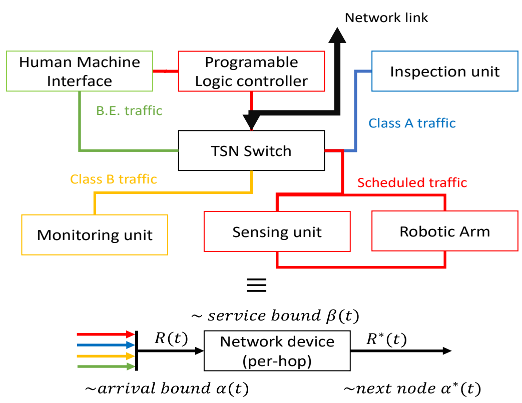

Fig. 3. Single hop model of QCAP TSN node.

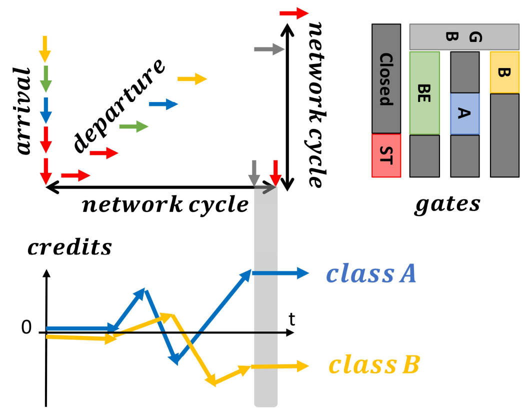

Fig. 4. TSN interface with data from different flows arriving simultaneously.

III. Quality Control After Production (QCAP)

A. QCAP Overview

QCAP [16] represents a typical quality-check industrial use case to ensure fault-free production. QCAP involves three main processes: 1) accurate detection of new product arrival, 2) collecting inspection data on the product, and 3) product classification after quality checking. Figure 3 depicts a QCAP cell including typical components and connections. These components include sensors for product detection, a camera used for product inspection, a floor monitoring unit, and a robotic arm to act upon quality inspection decisions. Moreover, a QCAP node may include a human-to-machine interface for diagnostics, device monitoring, and configuration. Hence, a QCAP cell could be considered a full industrial automation system that includes quality management as well as other relevant functions.

B. QCAP Traffic

We classify the different types of data traffic generated in a QCAP cell into four categories following the IEC/IEEE 60802 profiles. Those traffic types are:

Control data / Scheduled traffic (CDT/ST): Critical traffic flows with a hard deadline. Data frames for those flows are transmitted periodically, e.g., in cyclic real time (see Table I). For example, sensor readings have to be delivered to a computing unit within a deadline to provide updated information on product arrival, for example, per each application cycle. Control data frames are assigned the highest priority and are allocated exclusive transmission opportunities to ensure deterministic delivery behaviour across the entire network.

AVB traffic (class A): Non-critical traffic flows that nevertheless require bounded latency and bandwidth guarantees. Data frames from those flows are sporadic, large in size, and bursty in nature. The visual inspection of a detected product is an example of this type of data traffic in the industrial network. After enough bandwidth is allocated for critical traffic flows, the remaining bandwidth is shared by non-critical traffic flows, including class A traffic. In addition, network nodes assign class A traffic to a high-priority queue to ensure an upper bound on end-to-end latency.

AVB traffic (class B): Non-critical traffic flows with strict bandwidth requirements. Data frames in class B traffic have large sizes and persistence, which require the consistent allocation of network resources to ensure the quality of service requirements. Video traffic streamed by the QCAP monitoring unit is a typical source of class B frames. Although network nodes assign class B traffic a lower priority compared to class A traffic, token-based rate shaping techniques applied to class A create transmission opportunities for class B. Hence, class A QoS requirements can be met without starving lower-priority traffic.

Best effort traffic (B.E.): Non-critical traffic flows with no strict QoS requirements, which could be a result of an Ethernet device or human-to-machine interface connected to a QCAP node. Network nodes allocate the non-utilized bandwidth to carry best-effort frames. In this paper, we model B.E. traffic either as web browsing or file transfer. However, an analysis focusing on the impact of TCP/IP-based data frames on the critical data flows transmissions is an interesting future direction of our work.

IV. Delay Bound for TSN with Preemptive Scheduled Traffic

In this section, we first present the system model for TSN networks that consider preemptive scheduled traffic. We then derive the delay bounds for this system.

A. System overview

We consider a TSN-enabled network where each interface has four queues, one each for scheduled traffic, class A, class B, and best effort traffic.

Scheduled traffic frames have the highest priority. To ensure bounded delay for scheduled traffic, a time-aware shaping scheduler allocates an exclusive transmission window for the highest priority traffic. Hence, the scheduled traffic queue control gate is open upon frame arrival and control gates of the other queues are closed. Additionally, we consider the TSN Hold/Release preemption integration mode [28], i.e. a guard band exists before the scheduled traffic transmission window. This guard band is defined as the minimum preemptable portion of a lower priority frame. Each scheduled traffic frame is treated as a single unit of transmission, i.e. bits belonging to the same frame arrive at the same time instant.

Class A and class B traffic frames arrive as a bit stream with a burst. Therefore, credit-based shaping is used to regulate the traffic and limit the burst length before transmissions. Credit accumulation happens when class A/B traffic is idle and only outside both the scheduled traffic transmission window and the guard band period. The control gates of class A and class B queues open to allow frames transmission according to a defined schedule. Both classes share the transmission window that is left after transmitting the scheduled traffic. However, both traffic classes should be allocated different transmission opportunities, i.e. class A and class B gates should not open at the same time. Yet, the best effort traffic control gate could be simultaneously opened with any of the aforementioned AVB traffic classes to improve bandwidth utilization. If the frame transmission is preempted, the remaining part of the frame has to wait for the next available window for transmission. Finally, priority based transmission selection occurs when different traffic frames are queued for transmission.

Best effort traffic is a bit stream that arrives in bursts. Best effort frames typically share the transmission window with class A and class B traffic. A frame is preempted at a guard band instance, and has to wait for highest priority frame transmission if they arrived simultaneously. However, frame transmission will continue if it started before a preemptable frame with higher priority, e.g. class A/class B.

B. Delay Analysis

In this subsection, we derive the worst-case delay bounds for each considered traffic class of the TSN industrial profile. Based on the aforementioned assumptions, the arrival curve of the total scheduled traffic is defined by

$$\alpha_{ST}(t) = \sum_i L_i \times \frac{t}{T_i}, \qquad (7)$$

where L_i represents the maximum frame length and T_i represents the arrival period of the i-th scheduled traffic flow, respectively.

The arrival curves of class A and class B traffic are constrained by a token bucket, defined by

$$\alpha_{class}(t) = r_{class} \times t + \sigma_{class}, \qquad (8)$$

where r_{class} represents the sum of the traffic class data rate and σ_{class} represents the sum of the traffic class burst.

The service curve of the scheduled traffic has a rate-latency form that can be defined by

$$\beta_{ST}(t) = c \times (t - T_{ST}), \qquad (9)$$

where c represents the link speed and T_{ST} represents the maximum encountered delay for a scheduled traffic frame. T_{ST} can be expressed as

$$T_{ST} = \frac{L_{ST,k} - L_{minST,k}}{c}, \qquad (10)$$

where $L_{ST,k}$ represents the total length of the allocated window for all scheduled traffic frames at hop k and $L_{minST,k}$ represents the minimum frame length of the scheduled traffic at hop k.

The worst case scenario occurs when all frames of different scheduled traffic flows arrive simultaneously at hop k and the smallest scheduled traffic frame has to wait for all other scheduled frames to be transmitted.

Similar rate-latency equations apply to classes A and B traffic. For class A traffic, the service curve is given by

$$\beta_A(t) = R_A \times (t - T_A), \qquad (11)$$

where R_A represents the effective rate of class A traffic and T_A represents the maximum delay encountered by a bit arriving at an empty queue. R_A can be expressed as

$$R_A = \frac{I_A \times [c - r_{ST,k}]}{c} - \frac{I_A \times r_{GB}}{c - I_A}, \qquad (12)$$

where I_A represents the idle credit accumulation slope (bits/s) for AVB Class A traffic, $r_{ST,k}$ represents the sum of all scheduled traffic rates going through hop k, i.e., $r_{ST,k} = \sum_i (\frac{L_i}{T_i})$ per k, and r_{GB} represents the guard band rate that is calculated as $r_{GB} = \frac{L_{minPkt}}{T_{minPkt}}$, where L_{minPkt} represents the maximum length of the minimum non-preemptable portion of a low priority frame and T_{minPkt} represents the guard band duration. The first term in eq. (12) captures the transmission slope/rate of class A traffic outside the scheduled traffic transmission window, while the second term represents the guard band duration where class A transmissions are not allowed. T_A can be expressed as

$$T_A = \frac{L_{ST,k}}{r_{ST,k}} + \frac{L_{minPkt,k} + L_{BE}}{c - r_{ST,k}}, \qquad (13)$$

where L_{BE} represents the maximum frame length of the best effort traffic. In eq. (13) the first term represents the worst case which is when a guard band occurs and class A has negative credit. In this case, a class A frame has to wait for the guard band duration, the transmissions of all scheduled traffic, and the queued best effort traffic until it has accumulated enough positive credit to be allowed to be transmitted.

Similarly, the service curve of class B traffic can be defined as

$$\beta_B(t) = R_B \times (t - T_B), \qquad (14)$$

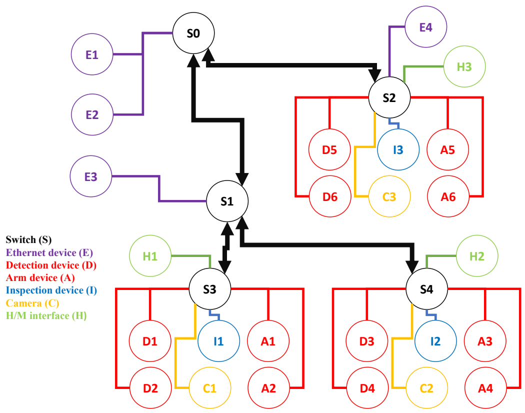

Fig. 5. Representative QCAP network topology.

where R_B represents the effective rate of class A traffic and T_B represents the maximum delay encountered by a bit arriving at an empty class B queue. R_B and T_B can be expressed as

$$R_B = \frac{I_B \times [c - r_{ST}]}{c} - \frac{I_B \times r_{GB}}{c - I_B}, \tag{15}$$

$$T_B = \frac{L_{ST,k}}{r_{ST,k}} + \frac{L_{minPkt,k} + L_A + L_{BE}}{c - r_{ST,k}}, \tag{16}$$

where I_B represents the idle slope of AVB Class B traffic and L_A represents the maximum frame length of class A traffic. Eq. (15) is analogues to eq. (12). Eq. (16) corresponds to the worst case when a guard band occurs and class B has negative credit. Therefore, class B frame must wait for the guard band duration, transmissions of all scheduled traffic, the class A transmission window, and queued best effort traffic given that class B still needs to accumulate enough positive credit. Finally, the BE service curve is defined by

$$\beta_{BE}(t) = c \times (t - T_{BE}) \tag{17}$$

$$T_{BE} = \sigma_{BE} + \frac{L_{ST,k}}{r_{ST,k}} + \frac{L_{minPkt} + L_A + L_B}{c - r_{ST}} \tag{18}$$

where σ_{BE} is the burst period of the BE traffic. In accordance with the network calculus framework, the delay bounds of a flow $f \epsilon Z_+^* = \{1, 2, ...\}$, which belongs to traffic class $cls \epsilon \{ST, A, B, BE\}$, and traversing a single hop k can be defined by

$$D_{f,k} = T_{cls} + \frac{\sigma_{cls} - L_{min,f}}{R_{cls}} + \frac{L_{min,f}}{c}, \tag{19}$$

For each traffic flow, the worst case end-to-end delay can be obtained by adding the per-hop delays along a path. In addition, the queue sizes of a node can be obtained by analyzing the number of flows traversing a node at time t. Our analysis here provides a guide to test the feasibility of network configurations against the network capacity to accommodate critical and non-critical traffic flows for industrial use cases such as the QCAP example.

TABLE II
QCAP DATA FLOW CHARACTERIZATIONS

Flow: source, sink	Type	Priority (PCP)[1]	Deadline
$f_1 ... f_6 : D_i, E_1$ $f_7 ... f_{12} : E_1, A_i$	CDT/ST	7	$100\mu s$
$f_{13} ... f_{15} : I_i, E_1$ $f_{16} ... f_{18} : E_1, I_i$	class A	5	$60ms$
$f_{19} ... f_{21} : C_i, E_2$ $f_{22} ... f_{24} : E_2, C_i$	class B	3	$80ms$
$f_{25} ... f_{27} : H_i, E_2$ $f_{28} ... f_{30} : E_2, H_i$ $f_{31} : E_3, E_4$ $f_{32} : E_4, E_3$	B.E.	0	-

V. Framework Validation and Performance Evaluation

In this section, we compare the performance of our industrial TSN network calculus (TSN/NC) framework in terms of calculating the worst case end-to-end delay upper bounds, with a computer simulation model (OMNET++) [25] and with a state of the art NC-based approach for industrial automation networks [11]. First, we describe the network topology and the data traffic flows for the QCAP use case. Afterwards, we present the simulation parameter setup and requirements. Finally, we analyse the performance of a TSN network implementing the QCAP use case.

A. Network Model

We consider a two-level tree topology of the single hop model of a QCAP node, depicted in Figure 3. The QCAP topology consists of three QCAP nodes, five switches, and four Ethernet devices as represented in Figure 5. The breakdown of QCAP node components is:

- Detection device (D): a sensing unit to detect the arrival of a new product for quality checking.
- Inspection device (I): a visual inspection unit to capture images of a detected product and submit those to make a decision on the quality of the product.
- Arm device (A): robotic arm used to remove a product not meeting the quality check from the production line.
- Camera (C): visual unit to real time monitor the floor area.
- H/M interface (H): Human machine interface to monitor and configure the operating devices of a QCAP node.
- Ethernet device (E): An Ethernet connected computing device that works as controller, computing unit, or monitoring interface for the industrial application.

The characterization of the data traffic flows are summarized in Table II. Generally, data exchanged as part of the control loop, e.g. direct communication between sensors, controllers and actuators, holds the highest priority, i.e. scheduled data traffic. The inspection unit produces class A traffic and the cameras produce class B traffic as detailed in Section III-B.

[1]Priority Code Point: 3-bits field, which is part of the VLAN header of an Ethernet frame to identify the data flow priority.

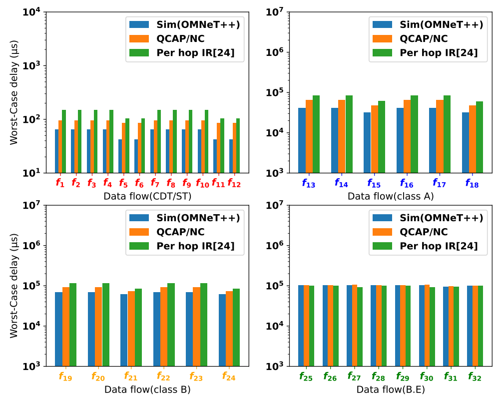

Fig. 6. Comparison between worst-case delay bounds for QCAP data flows.

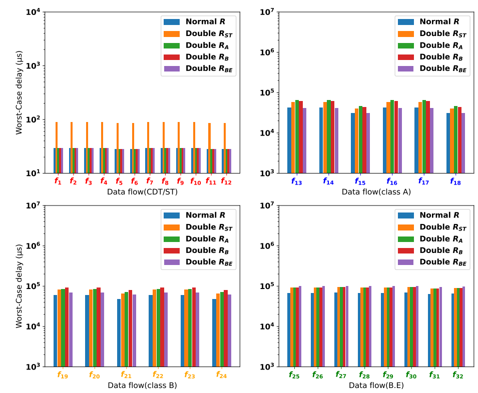

Fig. 7. Comparison between worst-case delay bounds for different flow rates.

B. Model implementation

For our evaluation, we have implemented our models using two approaches, OMNeT++ based discrete event simulation and Real-Time Calculus (RTC) toolbox [26] based implementation of the developed network calculus models. First, we obtain the worst-case end-to-end delay (WCD) bounds using the following implementations:

1) OMNeT++ based simulation framework for QCAP use case. [16].
2) RTC [26] based implementation of the presented industrial TSN/NC based WCD analysis, Section IV-B.
3) RTC based implementation of CBS+SP for industrial automation using NC [11].

Second, we use our industrial TSN/NC-based implementation to evaluate whether the QCAP network performance meets the different data flow QoS requirements. For this evaluation, we use typical network configurations and data traffic parameters for industrial applications, detailed in [27], which can be summarized as follows:

- Network link capacity (C): 100 Mbps or 1 Gbps.
- Maximum processing delay of a switch: 1 μs
- Maximum buffer size of a switch: 1000 frames.
- Propagation delay: $1\mu s/100m$
- Maximum frame size: Scheduled (L_i) = 128 B, class A (L_A) = 1500 B, class B (L_B) = 1500 B.
- Minimum cycle time/duration between frames: T_{ST} = 500 μs, T_A = 1 ms, T_B =500μs.
- Maximum data rate:r_{ST}: 2.05 Mbps, r_A = 12 Mbps, r_B = 24 Mbps.
- Idle slope: I_A = 40%, I_B = 20% of the link capacity.
- Maximum burst duration: σ_A = 2 sec, σ_B = 4 sec.

All of our experiments were run on a work station with Intel(R) Xeon(R) E-2174G CPU at 3.80GHz and 64 GB of RAM.

C. Evaluation and results

In our evaluation, we obtain and compare the maximum value (e.g. the worst case) of the end-to end delay for each flow using the three approaches described above. Firstly, we simulate the behaviour of the QCAP network model using OMNeT++. The model includes 32 flows, as highlighted in Table II. Next, we implement both our developed industrial TSN/NC model and the per hop IR NC-based model [11].

Figure 6 compares the delay bounds obtained by the three aforementioned methods. Figure 6 has four sub-plots with classified data flows, one plot for each data traffic class. Our NC model outperforms the other NC model by providing tighter delay bounds. This is expected, considering that the per hop IR model does not consider the preemption feature of the scheduled traffic. Additionally, it does not consider the relation between the best effort traffic and the credit accumulation of both class A and class B traffic. While our industrial TSN/NC framework exceeds the worst-case delay results when compared to the simulation model, its run-time is 10x faster than the network simulator with statistically safe results. Therefore, our TSN/NC framework can be used as a first stage assessment for the design of TSN-based industrial automation applications such as the QCAP use case.

Next, we focus on measuring the QCAP network performance for different configurations using our TSN/NC framework. For that, we start with relaxing the network configurations by choosing larger bandwidth (1 Gbps) links between switches. In addition, we separately double the transmission rates for each set of flows and then observe the impact of this increase on the Worst Case Delay (WCD) for all data flows. Figure 7 depicts the WCD bounds for the four different sets of flows of this experiment, where we observe the following:

- For scheduled flows (CDT/ST traffic), the WCD bound remains almost the same when doubling the other data flow rates. However, it only slightly changes when doubling the

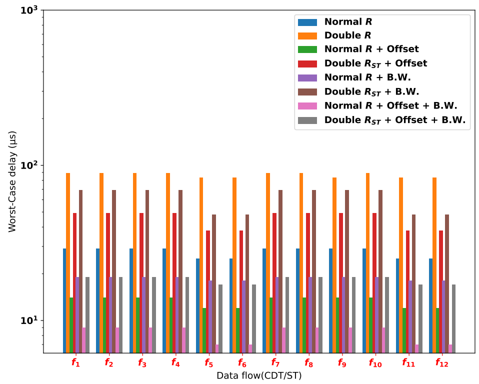

Fig. 8. The effect of transmission offset and increasing the bandwidth of critical interfaces on the WCD of scheduled traffic.

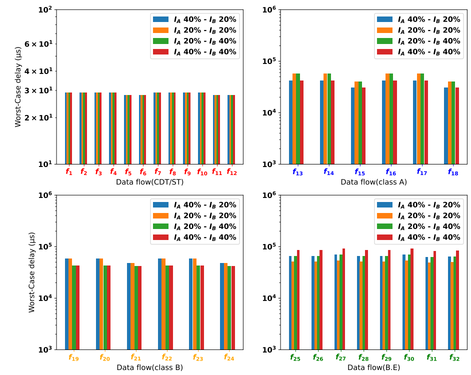

Fig. 9. Comparison between different idle slopes on the WCD of different traffic types.

rates of the CDT/ST traffic itself. This is due to the high priority of flows belonging to this type of traffic, as well as the preemption mode, which allows these data flows to be processed faster when they collide with lower priority traffic.

- For AVB traffic (class A and class B), the WCD bound increases progressively when data rates of the same type of traffic or rates of traffic with higher priority increase/double. Furthermore, the WCD bound becomes progressively less when low priority data rates increase. This is due to the credit based and strict priority shaping functionality offered by TSN switches.
- For best effort traffic, increasing the ST traffic rate leaves less bandwidth to be shared among the other three types of traffic, which increases the delay bound on the lowest priority traffic, e.g. BE class. However, increasing the data rates of AVB traffic creates more transmission opportunities for BE traffic while AVB traffic accumulates credit.

Based on the previous experiments, we can identify the WCD bounds for all traffic flows traversing the QCAP network. Based on this analysis, we demonstrate that the successful operation for each process of the QCAP use case can be guaranteed. Moreover, we show that our TSN/NC framework demonstrates lower WCD bounds on critical data traffic, allowing to accommodate more data flows if necessary. In Figure 8, we study two properties that impact the delay bounds of the critical traffic:

1) adding an offset period to the start of ST flow transmissions. This ensures that frames belonging to this traffic will arrive at different time instances, resulting in less queuing time.
2) identifying critical links to be configured with larger bandwidth, hence reducing the transmission and propagation delay. Therefore, the overall WCD for ST flows is reduced.

In this context, critical links do not imply links between switches only, where large number of flows are expected, but also interfaces where multiple ST flows are waiting for transmission. For example, in the QCAP topology, the link between TSN switch one (S1) and TSN switch zero (S0) is a critical link as many flows are traversing this link. However, the link that connects S0 to the Computing unit/Ethernet device 1 (E1) is the interface where all ST flows are transmitted through as part of the detection and inspection QCAP processes. In summary, setting a different offset for the ST critical flows and increasing the bandwidth of critical links can result in a 91% reduction in the worst case end-to-end delay of these flows, as depicted in Figure 8.

Our final experiment targets the influence of AVB traffic classes idle slopes on the WCD bounds of the different data flows. We expect that the larger the idle slope of an AVB class, the lower the WCD bound for this class of traffic. Moreover, the higher the WCD bound for the classes sharing bandwidth with this class of traffic. Figure 9 depicts the effect of varying idle slopes for class A and class B traffic. Starting with the normal case where $I_A = 40\%$ and $I_B = 20\%$, we increase the idle slope of class A to 20%. As expected, the ST flows' WCD is not affected. However, class A flows' WCD bound increases by a factor of 2 and the same applies to class B flows due to the dependency between the two types of AVB flows. Moreover, reducing the idle rates leads to reducing the WCD bound on the best effort data flows, which can be expected as more BE frames can be transmitted while AVB flows are accumulating credit to transmit. On the contrary, increasing idle slopes of AVB classes to $I_A = 40\%$ and $I_B = 40\%$ leads to faster credit accumulation for both class A and B, which reflects on the reduced WCD bound of the flows that belong to these classes. In addition, increasing idle slopes of AVB classes utilizes most of the available bandwidth, which leaves no room for best effort transmissions and hence increases its end-to-end-delay.

VI. Conclusion and Future Work

TSN is evolving to support industrial applications by including advanced traffic management techniques to ensure strict delay requirements for critical data streams. In this paper, we leverage network calculus to develop and validate a performance analysis framework to compute the worst-case delay bounds for four different types of data streams. The evaluation results show that our industrial TSN/NC framework provides tighter delay upper bounds compared to existing frameworks. In addition, it provides safer delay upper bounds when compared to a simulation based assessment for the same industrial application (QCAP), while running 10x faster than simulations. Additionally, we performed an extensive analysis of delay bounds for the QCAP use case to investigate the impact of progressively increasing data rates for all data flows on the end-to-end delay upper bounds. Increasing the data rates of traffic flows, or alternatively increasing the number of flows, will not only impact the flows of the same type, but also other types of traffic depending on the priority of those increased flow rates. Moreover, the impact of flow offset, the criticality of a physical link, and bandwidth availability for AVB traffic have been investigated. Hence, our developed model provides a tool that helps industrial network architects design optimised configurations to ensure the safe operation of the designated industrial application. In our future work, we plan to analyse a more general TSN setup for the eight classes of industrial traffic defined in IEC/IEEE 60802, including other TSN shaping and queuing features.

References

[1] Axelsson, Björn, and Geoff Easton. Industrial Networks (Routledge Revivals): A New View of Reality. Routledge, 2016.

[2] Sisinni, Emiliano, Abusayeed Saifullah, Song Han, Ulf Jennehag, and Mikael Gidlund. "Industrial internet of things: Challenges, opportunities, and directions." IEEE transactions on industrial informatics 14, no. 11 (2018): 4724-4734.

[3] Time-Sensitive Networking Task Group, [online] Available: https://www.ieee802.org/1/pages/tsn.html.

[4] "IEEE Standard for Local and Metropolitan Area Networks−Timing and Synchronization for Time-Sensitive Applications", IEEE 802.1AS−2020. [Online]. Available: https://standards.ieee.org/ieee/802.1AS/7121/

[5] "IEEE Standard for Local and Metropolitan Area Networks− Bridges and Bridged Networks, IEEE 802.1Q−2022". [Online]. Available: https://standards.ieee.org/ieee/802.1Q/10323/

[6] IEC/IEEE P60802, Use Cases IEC/IEEE 60802, [online] Available: https://1.ieee802.org/tsn/iec-ieee-60802/ /1/files/public/docs2018/60802-industrial-use-cases-0918-v13.pdf.

[7] J-Y. Le Boudec and P. Thiran, "Network calculus: a theory of deterministic queuing systems for the internet", Vol. 2050, Springer Science I& Business Media, 2001.

[8] IEEE Standard for Local and Metropolitan Area Networks—Bridges and Bridged Networks, IEEE Standard 802.1Q, Annex S.4, 2018. [Online]. Available: https://standards.ieee.org/standard/802_1Q-2018.html

[9] S. Bondorf, P. Nikolaus, J. B. Schmitt, "Quality and Cost of Deterministic Network Calculus Design and Evaluation of an Accurate and Fast Analysis" , In Proceedings of the ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS 2017), 2017.

[10] E. Mohammadpour, E. Stai, M. Mohiuddin, and J.-Y. Le Boudec, "End-to-end latency and backlog bounds in time-sensitive networking with credit based shapers and asynchronous traffic shaping," in Proc. 30th Int. Teletraffic Congr., 2018, pp. 1–6.

[11] J. Zhang, L. Chen, T. Wang and X. Wang, "Analysis of TSN for Industrial Automation based on Network Calculus," 2019 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), Zaragoza, Spain, 2019, pp. 240-247, doi: 10.1109/ETFA.2019.8869053.

[12] L. Maile, K. -S. Hielscher and R. German, "Network Calculus Results for TSN: An Introduction," 2020 Information Communication Technologies Conference (ICTC), Nanjing, China, 2020, pp. 131-140, doi: 10.1109/ICTC49638.2020.9123308.

[13] L. X. Zhao, P. Pop, Z. Zheng, and Q. Li, "Timing analysis of AVB traffic in TSN networks using network calculus," in Proc. IEEE Real Time Embedded Technol. Appl. Symp., 2018, pp. 25–36.

[14] E. Mohammadpour, E. Stai, and J. Y. Le Boudec, "Improved credit bounds for the credit-based shaper in time-sensitive networking," IEEE Netw. Lett., vol. 1, no. 3, pp. 136–139, Sep. 2019.

[15] L. Zhao, P. Pop, Z. Zheng, H. Daigmorte and M. Boyer, "Latency Analysis of Multiple Classes of AVB Traffic in TSN With Standard Credit Behavior Using Network Calculus," in IEEE Transactions on Industrial Electronics, vol. 68, no. 10, pp. 10291-10302, Oct. 2021, doi: 10.1109/TIE.2020.3021638.

[16] M. Seliem, A. Zahran and D. Pesch, "Quality Checks After Production: TSN-based Industrial Network Performance Evaluation," 2022 4th International Conference on Electrical, Control and Instrumentation Engineering (ICECIE), Kuala Lumpur, Malaysia, 2022, pp. 1-7, doi: 10.1109/ICECIE55199.2022.10000278.

[17] "IEEE Standard for Local and Metropolitan Area Networks−Timing and Synchronization for Time−Sensitive Applications−Amendment: Hot Standby", IEEE P802.1ASdm. [Online]. Available: https://1.ieee802.org/tsn/802−1asdm/

[18] "IEEE Standard for Local and Metropolitan Area Networks− Bridges and Bridged Networks, IEEE 802.1Q−2022". [Online]. Available: https://standards.ieee.org/ieee/802.1Q/10323/

[19] Shaper Parameter Settings for Bursty Traffic Requiring Bounded Latency, IEEE P802.1Qdq. [Online]. Available: https://1.ieee802.org/tsn/802-1qdq/

[20] "IEEE Standard for Local and metropolitan area networks–Frame Replication and Elimination for Reliability," in IEEE Std 802.1CB-2017 , vol., no., pp.1-102, 27 Oct. 2017, doi: 10.1109/IEEESTD.2017.8091139.

[21] "IEEE Standard for Local and metropolitan area networks— Bridges and Bridged Networks - Amendment 24: Path Control and Reservation," in IEEE Std 802.1Qca-2015 (Amendment to IEEE Std 802.1Q-2014 as amended by IEEE Std 802.1Qcd-2015 and IEEE Std 802.1Q-2014/Cor 1-2015) , vol., no., pp.1-120, 11 March 2016, doi: 10.1109/IEEESTD.2016.7434544.

[22] "IEEE Standard for Local and metropolitan area networks–Bridges and Bridged Networks–Amendment 28: Per-Stream Filtering and Policing," in IEEE Std 802.1Qci-2017 (Amendment to IEEE Std 802.1Q-2014 as amended by IEEE Std 802.1Qca-2015, IEEE Std 802.1Qcd-2015, IEEE Std 802.1Q-2014/Cor 1-2015, IEEE Std 802.1Qbv-2015, IEEE Std 802.1Qbu-2016, and IEEE Std 802.1Qbz-2016) , vol., no., pp.1-65, 28 Sept. 2017, doi: 10.1109/IEEESTD.2017.8064221.

[23] Time-Sensitive Networking (TSN) Task Group, [Online]. Available: https://1.ieee802.org/tsn/

[24] Lasi, Heiner, Peter Fettke, Hans-Georg Kemper, Thomas Feld, and Michael Hoffmann. "Industry 4.0." Business & information systems engineering 6 (2014): 239-242.

[25] OMNeT++ Discrete Event Simulator. [Online]. Available: https://www.omnetpp.org/. Accessed on: Jul. 1, 2022.

[26] E. Wanderler and L. Thiele, "Real-Time Calculus (RTC) Toolbox," 2006. [Online]. Available: http://www.mpa.ethz.ch/Rtctoolbox

[27] Seliem, M., Zahran, A. and Pesch, D. (2022) 'TSN-based industrial network performance analysis', 2022 8th World Forum on Internet of Things (WF-IoT), Yokohama, Japan, 26 October - 11 November. Available: http://hdl.handle.net/10468/14280

[28] "IEEE Standard for Local and metropolitan area networks – Bridges and Bridged Networks – Amendment 26: Frame Preemption," in IEEE Std 802.1Qbu-2016 (Amendment to IEEE Std 802.1Q-2014) , vol., no., pp.1-52, 30 Aug. 2016, doi: 10.1109/IEEESTD.2016.7553415.