HAL open science

Latency Analysis of Multiple Classes of AVB Traffic in TSN With Standard Credit Behavior Using Network Calculus

Luxi Zhao, Paul Pop, Zhong Zheng, Hugo Daigmorte, Marc Boyer

▶ **To cite this version:**

Luxi Zhao, Paul Pop, Zhong Zheng, Hugo Daigmorte, Marc Boyer. Latency Analysis of Multiple Classes of AVB Traffic in TSN With Standard Credit Behavior Using Network Calculus. IEEE Transactions on Industrial Electronics, 2021, 68 (10), pp.10291-10302. 10.1109/TIE.2020.3021638 . hal-03577428

HAL Id: hal-03577428

https://hal.science/hal-03577428v1

Submitted on 16 Feb 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Latency Analysis of Multiple Classes of AVB Traffic in TSN with Standard Credit Behavior using Network Calculus

Luxi Zhao*[a], Paul Pop[a], Zhong Zheng[b], Hugo Daigmorte[c], and Marc Boyer[d]

[a]Department of Applied Mathematics and Computer Science, Technical University of Denmark, Denmark
[b]Department of Electronics and Information Engineering, Beihang University, Beijing, China
[c]RealTime-at-Work, Toulouse, France
[d]Department on Information Processing and Systems at ONERA, Toulouse, France

***Abstract*—Time-Sensitive Networking (TSN) is a set of amendments that extend Ethernet to support distributed safety-critical and real-time applications in the industrial automation, aerospace and automotive areas. TSN integrates multiple traffic types and supports interactions in several combinations. In this paper we consider the configuration supporting Scheduled Traffic (ST) traffic scheduled based on Gate-Control-Lists (GCLs), Audio-Video-Bridging (AVB) traffic according to IEEE 802.1BA that has bounded latencies, and Best-Effort (BE) traffic, for which no guarantees are provided. The paper extends the timing analysis method to multiple AVB classes and proofs the credit bounds for multiple classes of AVB traffic, respectively under frozen and non-frozen behaviors of credit during guard band (GB). They are prerequisites for non-overflow credits of Credit-Based Shaper (CBS) and preventing starvation of AVB traffic. Moreover, this paper proposes an improved timing analysis method reducing the pessimism for the worst-case end-to-end delays of AVB traffic by considering the limitations from the physical link rate and the output of CBS. Finally, we evaluate the improved analysis method on both synthetic and real-world test cases, showing the significant reduction of pessimism on latency bounds compared to related work, and presenting the correctness validation compared with simulation results. We also compare the AVB latency bounds in the case of frozen and non-frozen credit during GB. Additionally, we evaluate the scalability of our method with variation of the load of ST flows and of the bandwidth reservation for AVB traffic.**

I. Introduction

ETHERNET is a well-established network protocol that has excellent bandwidth, scalability, compatibility and cost properties [1]. However, it was not suitable for real-time and safety critical applications [2]. Distributed safety-critical applications, like those found in the aerospace, automotive and industrial automation domains, require certification evidence for the correct real-time behavior of critical communication. Therefore, several extensions to the Ethernet protocol have been proposed, such as, ARINC 664 Specification Part 7 [3], TTEthernet [4], and EtherCAT [5]. In 2012, the IEEE 802.1 Time-Sensitive Networking (TSN) Task Group [6] has been founded to define standard real-time and safety-critical enhancements for Ethernet.

E-mail: luxzha@dtu.dk; paupo@dtu.dk; zhengzhong@buaa.edu.cn; hugo.daigmorte@realtimeatwork.com; Marc.Boyer@onera.fr.

TSN integrates multiple traffic types and supports interactions in several combinations. In this paper we consider a TSN solution supporting Credit-Based Shaper (CBS), previously defined for Audio-Video Bridging (AVB) in 802.1BA [8], currently in [7, § 8.6.8.2], with enhancements for Scheduled Traffic (previously defined in 802.1Qbv [9], now in [7, § 8.6.8.4]) based on Gate-Control Lists (GCLs). We call this architecture TSN/GCL+CBS. We consider three traffic-types with varying criticalities: Scheduled Traffic (ST) which is also known as Time-Triggered (TT) traffic, AVB traffic and Best-Effort (BE) traffic. ST traffic supports hard real-time applications that require very low latency and jitter (e.g., industrial control requires 1–10 milliseconds latency and microseconds jitter). It is transmitted based on schedule tables called GCLs that rely on a global synchronized clock (802.1ASrev [10]). AVB traffic is intended for applications that require bounded end-to-end latencies, but has less time-critical compared with ST traffic. It uses the CBS to prevent the starvation of lower priority AVB traffic. The TSN Task Group has extended AVB architecture, allowing the definition of multiple CBS [7, § 5.4.1.5], so we consider any number of AVB classes. BE traffic is used for applications that do not require any timing guarantees and is mapped to the remaining queues.

The schedulability of the scheduled ST traffic can be guaranteed during design phase by considering individual scheduling of ST flows and synthesizing the GCLs [11]. For more generic window-based GCLs, Zhao et al. [12] have proposed a Network Calculus (NC)-based method to analyze the latency bounds for flexible window-based ST traffic. Researchers have extended the analysis to the preemption mode using Real-Time Calculus [13]. Moreover, AVB in TSN can be used for time-critical applications that require bounded latency even though it has lower time-critical than ST traffic. Hence it is very important to have a safe and guaranteed WCD analysis for AVB that takes into account the ST flows. Although latency analysis methods have been successfully applied to AVB traffic in AVB networks [14], [15], [16], they do not take account for the interference of ST traffic on the AVB traffic and only consider two or three AVB classes. Recently, Cao et al. [17]

have proposed an eligible interval based formal analysis on the calculation of worst-case performance bounds for multiple classes of AVB traffic. However, they do not consider the effect of ST traffic on AVB traffic, and focus only on the "relative delay" on a single hop. Network Calculus-based analyses have been proposed [18] to compute the WCDs of Rate-Constrained (RC) traffic with the consideration of the ST (also named Time-Triggered, or TT) frames in TTEthernet, but the techniques are not applicable for TSN: RC traffic has no CBS, the integration modes are different in TSN and TTEthernet and TSN schedules ST traffic differently from TTEthernet: individual ST frames are considered in TTEthernet, whereas TSN schedules ST windows which may include several ST frames.

The timing analysis of non-ST traffic types in TSN has been addressed in [19], considering closed-gate blocking, strict priorities and a "Peristaltic Traffic Shaper". However, their analysis is not applicable to AVB, which uses CBS. Researchers [20],[21] have proposed the latency bounds for AVB traffic affected by control-data-traffic (CDT) in TSN, but assuming CDT as leaky bucket (LB) or length rate quotient (LRQ), which does not fit for the ST traffic. The AVB Latency Math equation has been extended to consider the ST traffic in TSN [22]. However, it does not consider the actual situations for AVB flows in the network, but just assumes that maximum allowable bandwidth is occupied by the corresponding AVB traffic class, thus causing overly pessimistic, i.e., leading to overly large WCDs. In addition, [22] can only be used to determine the WCDs of AVB Class A traffic. Researchers have extended the Eligible Interval Analysis (EIA) to calculate the delay of two classes of AVB traffic in TSN [23]. However, it does not consider relative offsets between ST windows, but just assumes the ST windows always arranging back-to-back. The initial idea for the two classes of AVB analysis under the influence of ST traffic in TSN network based on the Network Calculus has been given by [24], and then credit bounds for CBS are improved in [25]. However, all these analyses above are based on a credit behavior deviating from the standard 802.1Q-2018 [7], i.e., assuming credit frozen during guard band before gate closing. Moreover, the latest standard 802.1Q supports any number of AVB classes, which are not supported in previous analysis work. Thus, in this paper, we are interested to propose a formal performance analysis method for arbitrary number of AVB classes in TSN networks, under both behaviors of frozen and non-frozen credit during guard band (GB, cf. Sect. III for details).

The main contributions of our paper are as follows:

- To the best of our knowledge, our analysis is the first one to prove the credit bounds of CBS in the cases of frozen and non-frozen credit during guard band. The analysis is able to handle with arbitrary number of AVB classes under the influence of ST traffic for the whole TSN network.
- We reduce the pessimism of the analysis and this provides tighter latency bounds for AVB, by introducing the limitations from the physical link rate and the output of CBS, which are denoted as the link and CBS shaping curves with the consideration of ST traffic.
- We evaluate the proposed approach on both synthetic and real-world test cases, comparing AVB latencies in cases of frozen and non-frozen credit during guard bands, and comparing with related work and simulation results to show the significant reduction of pessimism on latency bounds, and to validate the correctness and the scalability of our implementation.

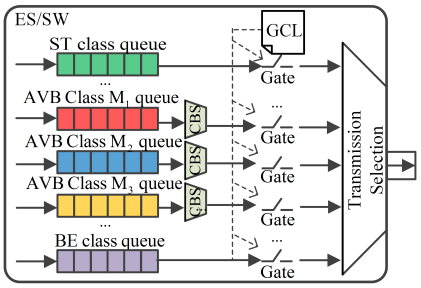

Fig. 1. TSN/GCL+CBS architecture for an output port in an ES/SW

The paper is organized as follows. Sect. II presents the system model. Sect. III introduces TSN. Sect. IV briefly introduces the Network Calculus concepts needed for the analysis. Sect. V gives the proof of credit bounds for multiple classes of AVB traffic respectively under frozen and non-frozen behaviors of credit during GB and presents tighter AVB latency bounds by introducing shaping curves. Sect. VI evaluates the proposed analysis and Sect. VII concludes the paper.

II. System Model

A TSN network is composed of a set of end systems (ES) and switches (SW) also called nodes, connected via physical links. In this paper, we assume, without loss of generality, that all physical links have the same rate C. The links are full duplex, allowing thus communication in both directions, and the networks can be multi-hop. The output port of a SW is connected to one ES or an input port of another SW.

The messages are sent from ESes via flows, which have a single source and may have multiple destinations. Each source ES is able to send multiple flows to the network. As mentioned, our TSN solution supports three traffic types: ST [7, § 8.6.8.4], AVB [7, § 8.6.8.2] and BE. We assume that the traffic type for each application has been decided by the designer. The subscripts M_i ($i \in [1,6]$) for AVB denote respectively the different AVB traffic classes. For ST traffic, we know the GCLs in each output port h of nodes, i.e., the opening and closing time of ST traffic (named as ST window) and GCL period (p^h_{GCL}). For an AVB flow $\tau_{M_i[k]} \in \tau_{M_i}$, we know its frame size $l_{M_i[k]}$, the minimum frame interval $p_{M_i[k]}$ in the source ES and the traffic class M_i it belongs to. The AVB Class M_i has higher priority than the AVB Class M_{i+1}. The flows assigned the same AVB traffic class M_i are served in FIFO order. Moreover, we know the maximum frame size l^{max}_{BE} of BE traffic.

III. TSN/GCL+CBS Output Port

The TSN Task group has defined a set of schedulers, which can interact in several combinations. In this paper, we assume a specific configuration, called TSN/GCL+CBS architecture to provide different real-time guarantees and will present how ST and AVB flows are transmitted in this case, as shown in Fig. 1. Each one has eight queues for storing frames that wait to be

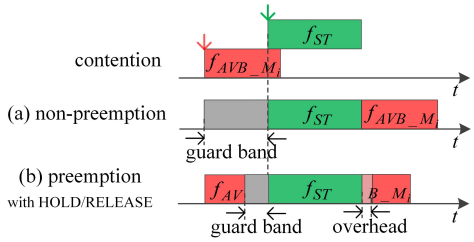

Fig. 2. Non-preemption integration modes

forwarded on the corresponding link, one or more for ST queues, two or more (n^h_{CBS}) for AVB queues (respectively for Class M_i) and the remaining queues are used for BE. Every queue has a gate with two states, open and closed. Frames waiting in the queue are eligible to be forwarded only if the associated gate is open.

The gates for each queue are controlled by GCLs, which are created offline and contain the times when the associated gates are open and closed [11]. In this study, we assume that when associated gate for ST traffic is open, the remaining gates for other traffic (AVB and BE) are closed, and vice versa (aka. exclusive gating). Therefore, AVB traffic is prevented from transmitting in the time windows reserved for ST frames.

In addition, two integration modes are introduced to solve the issue when an AVB frame is already in transmission at the beginning of time window for ST, i.e., non-preemption [7, Annex Q] and preemption [7, Annex S] modes. In this paper, we focus on the discussion of the *non-preemption* integration mode (see Fig. 2) and the analysis model is also suitable for the preemption integration mode combined with HOLD/RELEASE. The non-preemption mode uses a "guard band" before each ST window to prevent the AVB frame from initiating transmission if there is insufficient time available to transmit that entire frame before the open event of ST gate. The preemption mode combined with HOLD/RELEASE also uses a guard band, but with a rather smaller size [7, Annex S.4]. The "guard band" will lead to wasted bandwidth due to the guard band, but it ensures no delay for ST traffic.

An enqueued AVB frame is transmitted only if the associated gate is open and the Credit-Based Shaper (CBS) [7, § 8.6.8.2] permits it. Each AVB queue has a credit value, initialized to zero. When the AVB gate is closed, the associated credit is "frozen". When the associated AVB gate is open, the variation of credit is the same as the one in AVB networks [8], i.e., decreasing with a send slope ($sdSl_{M_i}$) during the transmission of an AVB frame and increasing with an idle slope ($idSl_{M_i}$) when AVB frames are waiting to be transmitted due to other higher priority AVB frames or the negative credit. This is illustrated for example in Fig. 3, where we have three AVB classes M_i ($i = 1,2,3$), and show the variation of credit for respective AVB class. Particularly, during guard bands, we assume two behaviors of the credit for AVB traffic. One is "non-frozen", obeying the behavior of credit in the standard. The other is "frozen", of which the behavior deviating from the standard, but is a model assumption for most existing work. Daigmorte et al. [26] have shown that the credit evolution rule with guard band in the standard is unfair, which is an argument for an update of the

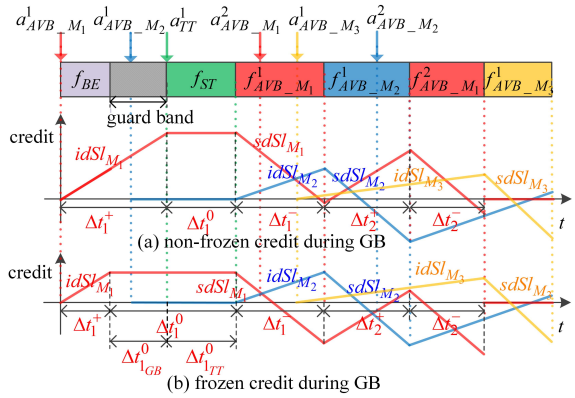

Fig. 3. CBS example with non-preemption mode

standard to consider the credit evolution rule to be frozen.

IV. Network Calculus Background

Network Calculus [27] is a mature theory proposed for deterministic performance analysis. It is used to construct arrival and service curve models for the investigated flows and network nodes. Network Calculus functions mainly belong to non-decreasing functions and null before 0: $\mathscr{F}_{\uparrow} = \{f : \mathbb{R}_+ \to \mathbb{R} | x_1 < x_2 \Rightarrow f(x_1) < f(x_2), x < 0 \Rightarrow f(x) = 0\}$. Two basic operators on $\mathscr{F}_{\uparrow}$ are the convolution $\otimes$,

$$(f \otimes g)(t) = \inf_{0 \le s \le t} \{f(t-s) + g(s)\}, \tag{1}$$

and deconvolution $\oslash$,

$$(f \oslash g)(t) = \sup_{s \ge 0} \{f(t+s) - g(s)\}, \tag{2}$$

where inf means infimum and sup means supremum.

An arrival curve $\alpha(t)$ is a model constraining the arrival process $R(t)$ of a flow, in which $R(t)$ represents the input cumulative function counting the total data bits of the flow that has arrived in the network node up to time t. We say that $R(t)$ is constrained by $\alpha(t)$ iff for all $s \le t$,

$$R(t) - R(s) \le \alpha(t-s). \tag{3}$$

A service curve $\beta(t)$ models the processing capability of the available resource. Assume that $R^*(t)$ is the departure process, which is the output cumulative function that counts the total data bits of the flow departure from the network node up to time t. There are several definitions for service curve. We say that the network node offers the min-plus minimum service curve $\beta(t)$ for the flow iff

$$R^*(t) \ge \inf_{0 \le s \le t} \{R(s) + \beta(t-s)\} = (R \otimes \beta)(t), \tag{4}$$

and offers the strict service curve $\beta(t)$ iff

$$R^*(t+\Delta t) - R^*(t) \ge \beta(\Delta t), \tag{5}$$

during any backlog period $(t, t+\Delta t]$. In addition, in order to evaluate service curves we will use the non-decreasing non-negative closure defined by

$$[f(t)]^+_{\uparrow} = \max_{0 \le s \le t} \{f(s), 0\}.$$

A shaping curve $\sigma(t)$ characterizes the maximum number of bits that are served during a period of time Δt, which means that the departure process $R^*(t)$ from the server always respects the shaping curve. A server offers a shaping curve $\sigma(t)$ iff $\sigma(t)$

could be an arrival curve for all output cumulative function $R^*(t)$ (Eq. 3). Note that the shaping curve $\sigma(t)$ from the output can also be said as an arrival curve for the subsequent nodes.

If a flow $R(t)$ of arrival curve $\alpha(t)$ crosses a server with the service curve $\beta(t)$, then the output flow $R^*(t)$ can be bounded by the arrival curve $\alpha'(t)$,

$$\alpha'(t) = \alpha \oslash \beta(t) = \sup_{s\geq 0}\{\alpha(t+s) - \beta(s)\}, \tag{6}$$

It can also be taken as an arrival curve of input flow for the next node.

Let us assume that the flow constrained by the arrival curve $\alpha(t)$ traverses the network node offering the service curve $\beta(t)$. Then, the latency experienced by the flow in the network node is bounded by the maximum horizontal deviation between the graphs of two curves $\alpha(t)$ and $\beta(t)$,

$$h(\alpha,\beta) = \sup_{s\geq 0}\{\inf\{\tau \geq 0 \mid \alpha(s) \leq \beta(s+\tau)\}\}. \tag{7}$$

The network end-to-end delay of a frame along its route is defined as the time duration between the instant it is transmitted by the source ES on the link and the instant it is fully received by the destination ES. The worst-case end-to-end delay of the flow can be bounded by the sum of latency bounds in each network nodes along its route.

V. Worst-case Analysis for AVB Traffic

A. Service Curve for AVB Traffic

In this section, we will extend the service curve in [24] to multiple AVB Class M_i ($i \in [1, n^h_{CBS}]$) with respectively (i) non-frozen and (ii) frozen credit during guard band. The service curve for AVB traffic depends on the working mechanism of CBS. As can be seen from Fig. 3, no matter the case (i) or (ii) for each AVB traffic Class M_i, any time interval Δt can be decomposed into

$$\Delta t = \Delta t^+ + \Delta t^- + \Delta t^0, \tag{8}$$

where $\Delta t^+ = \sum_i \Delta t_i^+$ (resp. $\Delta t^- = \sum_j \Delta t_j^-$) represents the accumulated length of all period where the credit increases (resp. decreases), and $\Delta t^0 = \sum_k \Delta t_k^0$ is the frozen time of credit. The service could only be supplied for AVB traffic during the descent time Δt^- of credit. Then, the service curve for multiple classes of AVB traffic is given by the following theorem.

Theorem 1: The min-plus minimum service curve for AVB Class M_i ($i \in [1, n^h_{CBS}]$) under non-preemption mode in an output port h is given by

$$\beta^h_{M_i(I)}(t) = idSl_{M_i}\left[t - \frac{\alpha^h_{H(I)}(t)}{C} - \frac{c^{max}_{M_i(I)}}{idSl_{M_i}}\right]^+_{\uparrow}, \tag{9}$$

where $I \in \{NF, F\}$ represents the behavior of credit during GB, and $\alpha^h_{H(I)}(t)$ is the arrival curve with regard to credit frozen. In case of $I = NF$, i.e., (i) non-frozen credit during GB, $\alpha^h_{H(I)}(t) = \alpha^h_{ST}(t)$ is only related to higher priority ST traffic, while in case of $I = F$, i.e., (ii) frozen credit during GB, $\alpha^h_{H(I)}(t) = \alpha^h_{GB+ST}(t)$ should additionally take guard band slots into account. $\alpha^h_{ST}(t)$ and $\alpha^h_{GB+ST}(t)$ are respectively given by Lemma 1 and Lemma 2 in Sect. V-C. $c^{max}_{M_i(I)}$ is the upper bound of credit of Class M_i under cases of (i) non-frozen or (ii) frozen credit during GB, and given by Eqs. (11) or (12). The proof of credit bound has to be extended to an arbitrary number of AVB classes and to consider different behaviors of credit during guard band, which is one of the challenges in this paper and will be discussed in Sect. V-B. The proof Theorem 1 is similar as the one given in [24], hence it will not be further discussed in this paper.

B. Bounding the Credit for AVB Traffic

In this section, we bound the credit for AVB traffic. Let us recall from Sect. III how AVB is transmitted.

Theorem 2: (Lower bound of Class M_i) Let $l^{max}_{M_i}$ be the maximal frame size of any flow crossing the AVB queue Q_{M_i}. Then, the credit $c_{M_i}(t)$ of Class M_i is lower bounded by [14], [24],

$$c_{M_i}(t) \geq \frac{l^{max}_{M_i}}{C} sdSl_{M_i} = c^{min}_{M_i}. \tag{10}$$

Before we start to prove the upper bound of credit, let us define some notations. Let $Q^{\leq i}_{AVB} = \bigcup_{1\leq j\leq i}\{Q_{M_j}\}$ denotes the same or higher priority AVB queues and $Q^{>i}_{AVB} = \bigcup_{i<j\leq n^h_{CBS}}\{Q_{M_j}\}$ denotes the lower priority AVB queues. Then for $\forall s,t \in \mathbb{R}^+, s \leq t$, the interval $[s,t]$ can be partitioned into several kinds of intervals: let $\Delta t_{M_i}(s,t)$ and $\Delta t^{<i}_{AVB}(s,t)$ respectively denote the duration of emissions of frames from the queue Q_{M_i} and of high priority frames from the queue $Q^{<i}_{AVB}$; $\Delta t_{ST}(s,t)$ is the time slot reserved for ST traffic, *i.e.*, the duration where the gates of AVB queues are closed; $\Delta t_{LP}(s,t)$ is the duration where a frame in $Q_{M_j} \in Q^{\leq i}_{AVB}$ waits due to a lower priority frame in queue $Q^{>i}_{AVB}$ or Q_{BE} that is being transmitted cannot be preempted; $\Delta t_{GB}(s,t)$ denotes the guard band duration where for all queues $Q_{M_j} \in Q^{\leq i}_{AVB}$ with $c_{M_j} > 0$ no frame can be sent, as there is no sufficient time available to transmit the entire frame before the next closing event; and $\Delta t_{Other}(s,t)$ is the interval except all the definitions above. Then we have $t - s = \Delta t_{M_i}(s,t) + \Delta t^{<i}_{AVB}(s,t) + \Delta t_{ST}(s,t) + \Delta t_{LP}(s,t) + \Delta t_{GB}(s,t) + \Delta t_{Other}(s,t)$.

Theorem 3: (Higher bound of Class M_i) Let l^{max}_{BE} be the maximal frame size of a BE flow. The credit $c_{M_i}(t)$ of Class M_i is upper bounded by,

(i) non-frozen credit during GB

$$c_{M_i}(t) \leq idSl_{M_i} \cdot \frac{\sum_{j=1}^{i-1} c^{min}_{M_j} - l^{max}_{>i} - \sigma^{M_i}_{GB}}{\rho^{M_i}_{GB} + \sum_{j=1}^{i-1} idSl_{M_j} - C} = c^{max}_{M_i(NF)}; \tag{11}$$

(ii) frozen credit during GB

$$c_{M_i}(t) \leq idSl_{M_i} \cdot \frac{\sum_{j=1}^{i-1} c^{min}_{M_j} - l^{max}_{>i}}{\sum_{j=1}^{i-1} idSl_{M_j} - C} = c^{max}_{M_i(F)}. \tag{12}$$

where $l^{max}_{>i} = max_{j\in[i+1,n^h_{CBS}]}\{l^{max}_{M_j}, l^{max}_{BE}\}$, $c^{min}_{M_j}$ is the lower bound of credit of Class M_j from Theorem 2, and $\sigma^{M_i}_{GB}$ and $\rho^{M_i}_{GB}$ are the parameters of upper envelope related to guard band duration and satisfy for $\forall s,t \in \mathbb{R}^+, s \leq t$,

$$C \cdot \Delta t_{GB}(s,t) \leq \sigma^{M_i}_{GB} + \rho^{M_i}_{GB} \cdot (t - s - \Delta t_{ST}(s,t)). \tag{13}$$

which will be discussed in the end of Sect.V-C.

Proof: Considering the evolution of the credit, the maximal value of credit will only happen at the instant satisfying the following definition. Let $t \in \mathbb{R}^+$ be a time point when AVB gates are in the open state, and $c_{M_i}(t) > 0$. Then let us define $s = \sup\left\{u \leq t \;\middle|\; \forall Q_{M_j} \in Q^{\leq i}_{AVB}, c_{M_j}(t) \leq 0\right\}$. It implies that $\forall u \in (s,t]$, $\exists Q_{M_j} \in Q^{\leq i}_{AVB}$, $c_{M_j}(u) > 0$, *i.e.*, there always

exists at least one queue in $Q_{AVB}^{\leq i}$ with some frame to send. Otherwise, we can always find another $s < s' \leq t$ that satisfies $\forall Q_{M_j} \in Q_{AVB}^{\leq i}, c_{M_j}(s') \leq 0$. As it always exists one queue in $Q_{AVB}^{\leq i}$ with some frame to send, either it succeeds (there is a frame of M_i emission, or a high-priority frame in $Q_{M_j} \in Q_{AVB}^{>i}$ emission) or is blocked (non-preemption or guard band or ST traffic transmission). Thus $\Delta t_{Other}(s,t) = 0$ here. In the following, as s,t are fixed, the $\Delta t_X(s,t)$ will be simplified to Δt_X.

(i) *For the case of non-frozen credit during GB*

Consider first the evolution of the credit value of M_i between s and t. The credit $c_{M_i}(t)$ increases at speed $idSl_{M_i}$ when the frame in the queue Q_{M_i} is waiting (during $\Delta t_{AVB}^{<i} + \Delta t_{LP} + \Delta t_{GB}$), decreases at speed $sdSl_{M_i} = idSl_{M_i} - C$ when the frame in the queue Q_{M_i} is transmitting (during Δt_{M_i}), is not modified when the gate of Q_{M_i} is closed during ST traffic transmission, and may be reduced from some positive value P to 0 due to resets. Thus the variation of $c_{M_i}(t)$ during $(s,t]$ is,

$$\begin{aligned} &c_{M_i}(t) - c_{M_i}(s) \\ &= \Delta t_{M_i} \cdot sdSl_{M_i} + \left(\Delta t_{AVB}^{<i} + \Delta t_{LP} + \Delta t_{GB}\right) \cdot idSl_{M_i} - P. \end{aligned} \tag{14}$$

Since $\Delta t_{AVB}^{<i} + \Delta t_{LP} + \Delta t_{GB} = s - t - \Delta t_{M_i} - \Delta t_{ST}$ and $P \geq 0$, Eq. (14) is modified into,

$$c_{M_i}(t) - c_{M_i}(s) \leq -\Delta t_{M_i} \cdot C + (t - s - \Delta t_{ST}) \cdot idSl_{M_i}. \tag{15}$$

Let $c_{<i}(t) = \sum_{j=1}^{i-1} c_{M_j}(t)$ denote the sum of credits of AVB traffic with the priority higher than M_i. Consider three cases: first, at any instant between s and t either a frame of M_i uses the link, or a low priority frame blocks the link or during GB, the credit of each non-empty queue $Q_{M_j} \in Q_{AVB}^{<i}$ increases with speed $idSl_{M_j}$. Then the sum of credits $c_{<i}(t)$ increases at most at speed $\sum_{j=1}^{i-1} idSl_{M_j}$. Or a frame from class with higher priority than M_i is being sent. In this case $c_{<i}(t)$ decrease at least at speed $\sum_{j=1}^{i-1} idSl_{M_j} - C$ (all the classes from $Q_{AVB}^{<i}$ gain credit expect one which loses credit). The last case is like for M_i, $c_{<i}(t)$ may decrease due to a set of resets. Then it is possible to upper bound the variation of $c_{<i}(t)$ between s and t,

$$\begin{aligned} &c_{<i}(t) - c_{<i}(s) \leq (\Delta t_{M_i} + \Delta t_{LP} + \Delta t_{GB}) \cdot \sum_{j=1}^{i-1} idSl_{M_j} \\ &+ (t - s - \Delta t_{ST} - \Delta t_{M_i} - \Delta t_{LP} - \Delta t_{GB}) \cdot \left(\sum_{j=1}^{i-1} idSl_{M_j} - C\right) \\ &= (\Delta t_{M_i} + \Delta t_{LP} + \Delta t_{GB}) \cdot C + (t - s - \Delta t_{ST}) \cdot \left(\sum_{j=1}^{i-1} idSl_{M_j} - C\right). \end{aligned} \tag{16}$$

By definition of s, it exists continuously some queue $Q_{M_j} \in Q_{AVB}^{\leq i}$ trying to send a frame. Hence there is at most one low priority frame that can get access to the link before s and go on transmission due to non-preemption. Thus

$$\Delta t_{LP} \cdot C \leq \max_{j \in [i+1, n_{CBS}^h]} \{l_{M_j}^{max}, l_{BE}^{max}\} = l_{>i}^{max}.$$

Moreover, we assume that there exists $\sigma_{GB}^{M_i}$ and $\rho_{GB}^{M_i}$ which will be discussed in Sect. V-C such that $\forall s,t \in \mathbb{R}^+, s \leq t$,

$$\Delta t_{GB} \cdot C \leq \sigma_{GB}^{M_i} + \rho_{GB}^{M_i} \cdot (t - s - \Delta t_{ST}).$$

Thus Eq. (16) is modified to obtain,

$$\begin{aligned} c_{<i}(t) - c_{<i}(s) \leq &\Delta t_{M_i} \cdot C + l_{>i}^{max} + \sigma_{GB}^{M_i} \\ &+ (t - s - \Delta t_{ST}) \cdot \left(\rho_{GB}^{M_i} + \sum_{j=1}^{i-1} idSl_{M_j} - C\right). \end{aligned} \tag{17}$$

Considering the system is assumed to be not overloaded, we have $\rho_{GB}^{M_i} + \sum_{j=1}^{i-1} idSl_{M_j} \leq C$. Thus, from Eq. (17)

$$t - s - \Delta t_{ST} \leq \frac{c_{<i}(t) - c_{<i}(s) - \Delta t_{M_i} \cdot C - l_{>i}^{max} - \sigma_{GB}^{M_i}}{\rho_{GB}^{M_i} + \sum_{j=1}^{i-1} idSl_{M_j} - C}. \tag{18}$$

Then from (18), (15) is modified to obtain,

$$c_{M_i}(t) - c_{M_i}(s) \leq idSl_{M_i} \frac{c_{<i}(t) - c_{<i}(s) - l_{>i}^{max} - \sigma_{GB}^{M_i}}{\rho_{GB}^{M_i} + \sum_{j=1}^{i-1} idSl_{M_j} - C}.$$

By definition of s, we have $c_{<i}(s) \leq 0$ and $c_{M_i}(s) \leq 0$ as well. Moreover, if $c_{M_j}^{min}$ is the lower bound of credit of Class M_j, then $c_{<i}(t) \geq \sum_{j=1}^{i-1} c_{M_j}^{min}$ so to conclude,

$$c_{M_i}(t) \leq idSl_{M_i} \cdot \frac{\sum_{j=1}^{i-1} c_{M_j}^{min} - l_{>i}^{max} - \sigma_{GB}^{M_i}}{\rho_{GB}^{M_i} + \sum_{j=1}^{i-1} idSl_{M_j} - C}.$$

(ii) *For the case of frozen credit during GB*

The proof is similar to the above process, except the difference of the evolution of credit of M_i during GB. The credit $c_{M_i}(t)$ increases at speed $idSl_{M_i}$ during $\Delta t_{AVB}^{<i} + \Delta t_{LP}$, decreases at speed $sdSl_{M_i} = idSl_{M_i} - C$ during Δt_{M_i}, and is not modified during guard band Δt_{GB} and ST traffic transmission Δt_{ST}. Then the variation of $c_{M_i}(t)$ during $(s,t]$ is upper bounded by,

$$c_{M_i}(t) - c_{M_i}(s) \leq -\Delta t_{M_i} \cdot C + (t - s - \Delta t_{GB} - \Delta t_{ST}) \cdot idSl_{M_i}. \tag{19}$$

Additionally, for the sum of credits $c_{<i}(t)$ of AVB traffic with the priority higher than M_i, it increases at most as speed $\sum_{j=1}^{i-1} idSl_{M_j}$ when the frame of M_i uses the link (during Δt_{M_i}) or a low priority frame blocks the link (during Δt_{LP}),

$$\begin{aligned} &c_{<i}(t) - c_{<i}(s) \leq (\Delta t_{M_i} + \Delta t_{LP}) \cdot \sum_{j=1}^{i-1} idSl_{M_j} \\ &+ (t - s - \Delta t_{GB} - \Delta t_{ST} - \Delta t_{M_i} - \Delta t_{LP}) \cdot \left(\sum_{j=1}^{i-1} idSl_{M_j} - C\right) \\ &\leq \Delta t_{M_i} \cdot C + l_{>i}^{max} + (t - s - \Delta t_{GB} - \Delta t_{ST}) \cdot \left(\sum_{j=1}^{i-1} idSl_{M_j} - C\right). \end{aligned} \tag{20}$$

Similarly, considering the system is assumed to be not overloaded, we have $\sum_{j=1}^{i-1} idSl_{M_j} \leq C$. Thus, from Eq. (20)

$$t - s - \Delta t_{GB} - \Delta t_{ST} \leq \frac{c_{<i}(t) - c_{<i}(x) - \Delta t_{M_i} \cdot C - l_{>i}^{max}}{\sum_{j=1}^{i-1} idSl_{M_j} - C}. \tag{21}$$

Then from (21), (19) is modified to obtain,

$$c_{M_i}(t) - c_{M_i}(s) \leq idSl_{M_i} \frac{c_{<i}(t) - c_{<i}(x) - l_{>i}^{max}}{\sum_{j=1}^{i-1} idSl_{M_j} - C}.$$

Thus, the upper bound of $c_{M_i}(t)$ is to conclude,

$$c_{M_i}(t) \leq idSl_{M_i} \cdot \frac{\sum_{j=1}^{i-1} c_{M_j}^{min} - l_{>i}^{max}}{\sum_{j=1}^{i-1} idSl_{M_j} - C}.$$

■

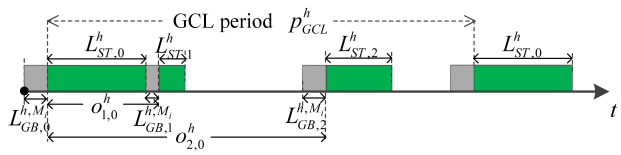

Fig. 4. Guard bands and ST windows

C. Arrival Curve of ST/GB duration

In this section, the arrival curve of ST/GB duration is constructed. It is used to derive the leftover service curve for AVB traffic of Class M_i, as AVB traffic cannot be transmitted when credit is frozen. As discussed above, credit will be frozen during ST time slots (windows) depending on the GCL, and whether it is frozen during GB intervals depends on the behavior selection.

ST traffic is scheduled within specific ST windows according to GCLs. The GCL for the output port h is repeated after the GCL period p^h_{GCL}, see the example in Fig. 4. Let N^h_{ST} be the number of ST windows in the GCL period p^h_{GCL}. It is assumed that the ith ST window in output port h starts at o^h_i and has duration $L^h_{ST,i}$, and the relative offset between the starting time of the ith and jth ST windows is $o^h_{j,i} = o^h_j - o^h_i$ For the case of credit non-frozen for AVB traffic during guard bands, the arrival curve for the credit frozen part is only related to ST windows, which is given by the following Lemma.

Lemma 1: The arrival curve of the credit frozen part due to ST traffic in an output port h is given by, for all $t \in \mathbb{R}^+$

$$\begin{aligned} \alpha^h_{ST}(t) &= \max_{0 \le i \le N^h_{ST}-1} \left\{ \alpha^h_{ST,i}(t) \right\} \\ \alpha^h_{ST,i}(t) &= \sum_{j=i}^{i+N^h_{ST}-1} L^h_{ST,j} C \cdot \left\lceil \frac{t - o^h_{j,i}}{p^h_{\text{GCL}}} \right\rceil, \end{aligned} \tag{22}$$

where is $\alpha^h_{ST,i}(t)$ is one possible arrival curve by selecting ith ($i \in \left[0, N^h_{ST,i}\right]$) ST window as the reference; $L^h_{ST,j} \cdot C$ represents the maximum number of bits the could be transmitted during the ST traffic window of length $L^h_{ST,j}$; each staircase function shows the upper bound of ST transmission in the periodic ST windows of length $L^h_{ST,j}$, and the relative offsets $o^h_{j,i}$ give the relationships between different ST windows within the GCL period. The proof of the lemma is similar to the proof for TT flows in TTEthernet [18].

Moreover, for the case of credit frozen for AVB traffic of Class M_i during GB, the arrival curve for the credit frozen part is not only related to ST windows, but also to GB intervals. A guard band may appear before a ST window, depending on the flow backlog. In the worst-case, the time duration of the guard band $L^{h,M_i}_{GB,j}$ before the jth ($j \in [0, N^h_{ST}-1]$) ST window equals to the minimum value of the maximum transmission time ($l^{max}_{\le i}/C$) of AVB frames of Class M_i competing in output port h and the idle time interval between two consecutive ST windows ($(j-1+N^h_{ST})\%N^h_{ST}$th and jth windows). We merge the guard band and ST window together to construct the arrival curve of the credit frozen part, the following Lemma is given,

Lemma 2: The arrival curve of the credit frozen part due to ST traffic and guard band in an output port h is given by, for all $t \in \mathbb{R}^+$

$$\alpha^{h,M_i}_{GB+ST}(t) = \max_{0 \le j \le N^h_{ST}-1} \left\{ \alpha^{h,M_i}_{GB+ST,j}(t) \right\} \tag{23}$$

$$\alpha^{h,M_i}_{GB+ST,j}(t) = \sum_{k=j}^{j+N^h_{ST}-1} (L^h_{ST,k} + L^{h,M_i}_{GB,k}) C \left\lceil \frac{t - o^h_{k,j} + L^{h,M_i}_{GB,k} - L^{h,M_i}_{GB,j}}{p^h_{\text{GCL}}} \right\rceil,$$

where is $\alpha^{h,M_i}_{GB+ST,j}(t)$ is one possible arrival curve by selecting the jth ($j \in \left[0, N^h_{ST}-1\right]$) ST window as the reference.

In the following, we are interested to derive $\sigma^{M_i}_{GB}$ and $\rho^{M_i}_{GB}$ defined in Eq. (13) which is used for deriving credit upper bound of AVB Class M_i. They are parameters (burst and long-term rate) related to the upper envelope of accumulated bits in guard band duration $\Delta t_{GB}(s,t)$ in any interval $[s,t]$. Additionally, note that $\sigma^{M_i}_{GB}$ and $\rho^{M_i}_{GB}$ are parameters defined based on $t - s - \Delta t_{ST}(s,t)$ in Eq. (13). The following Theorem gives the upper envelope of the guard band defined in Eq. (13) in the form of the staircase function.

Theorem 4: The staircase upper bound on the GB duration defined from Eq. (13) for AVB Class M_i traffic for all $t \in \mathbb{R}^+$ is as follows,

$$\alpha^{h,M_i}_{GB}(t) = \max_{0 \le j \le N^h_{ST}-1} \left\{ \alpha^{h,M_i}_{GB,j}(t - \Delta t_{ST}(t)) \right\}, \tag{24}$$

where

$$\alpha^{h,M_i}_{GB,j}(t - \Delta t_{ST}(t)) = \sum_{k=j}^{j+N^h_{ST}-1} L^{h,M_i}_{GB,k} C \left\lceil \frac{t - o^h_{k,j} + L^{h,M_i}_{GB,k} + \sum_{p=j}^{k} L^h_{ST,p}}{p^h_{\text{GCL}} - \sum_{p=j}^{j+N^h_{ST}-1} L^h_{ST,p}} \right\rceil. \tag{25}$$

Proof: We need to upper bound $\Delta t_{\text{GB}}(s, s+t)$ the number of $L^{h,M_i}_{\text{GB},j}$ in $[s, s+t]$. Introduce $\forall j : u_j = o_j - L^{h,M_i}_{\text{GB},j}$. First, consider some u_j and $t \ge 0$, let $k = \max\left\{p \mid u_p < u_j + t\right\}$

$$\Delta t_{\text{GB}}(u_j, u_j + t) \le L^{h,M_i}_{\text{GB},j} + \ldots + L^{h,M_i}_{\text{GB},k-1} + L^{h,M_i}_{\text{GB},k} \tag{26}$$

$$= \sum_{k=j}^{\infty} L^{h,M_i}_{\text{GB},k} \mathbb{1}_{\{t > u_k - u_j\}} \tag{27}$$

Noticing that $\Delta t_{\text{ST}}(u_j, u_k) = \sum_{p=j}^{k-1} L^h_{ST,p}$, Chasle's relation $\Delta t_{\text{ST}}(u_j, u_j + t) = \Delta t_{\text{ST}}(u_j, u_k) + \Delta t_{\text{ST}}(u_k, u_k + t)$ and $\Delta t_{\text{ST}}(u_k, u_j + t) \le L^h_{ST,k}$ leads to

$$\Delta t_{\text{GB}}(u_j, u_j + d) \le \sum_{k=j}^{\infty} L^{h,M_i}_{\text{GB},k} \mathbb{1}_{\left\{t - \Delta t_{\text{ST}}(u_j, u_j + d) > u_k - u_j - \sum_{p=j}^{k} L^h_{ST,p}\right\}} \tag{28}$$

But any infinite sum on any expression $E(k)$, can be decomposed as a double sum using $k = k' + nN^h_{ST}$ with $k' = k \ mod \ N^h_{ST}$

$$\sum_{k=j}^{\infty} E(k) = \sum_{k'=j}^{j+N^h_{ST}-1} \sum_{n=0}^{\infty} E(k' + nN^h_{ST}) \tag{29}$$

The system behavior is periodic, so $\forall k, u_{k+N^h_{ST}} = u_k + p^h_{\text{GCL}}$, $L^h_{ST,k+N^h_{ST}} = L^h_{ST,k}$, $L^{h,M_i}_{\text{GB},k+N^h_{ST}} = L^{h,M_i}_{\text{GB},k}$, leading to, $\forall t$

$$\begin{aligned} &\sum_{k=j}^{\infty} \mathbb{1}_{\left\{t > u_k - u_j - \sum_{p=j}^{k} L^h_{ST,p}\right\}} \\ &\quad = \sum_{k'=j}^{j+N^h_{ST}-1} \sum_{n=0}^{\infty} \mathbb{1}_{\left\{t > u_{k'} + np^h_{\text{GCL}} - u_j - \sum_{p=j}^{k'} L^h_{ST,p} - nL^h_{\text{ST,GCL}}\right\}} \end{aligned} \tag{30}$$

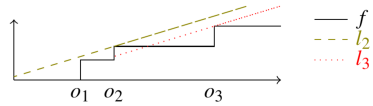

Fig. 5. Linear bound on staircase function

with $L^h_{\text{ST,GCL}} = \sum_{k=0}^{N^h_{ST}-1} L^h_{ST,k}$. Now, remark that $\forall t, J, P \in \mathbb{R}_+$, $P > 0$,

$$\left\lceil \frac{t-J}{P} \right\rceil^+ = \sum_{n=0}^{\infty} \mathbb{1}_{\{t > J+nP\}}. \tag{31}$$

Combining all leads to

$$\sum_{k=j}^{\infty} L^{h,M_i}_{\text{GB},k} \mathbb{1}_{\left\{t > u_k - u_j - \sum_{p=j}^{k} L^h_{ST,p}\right\}}$$
$$= \sum_{k=j}^{j+N^h_{ST}-1} L^{h,M_i}_{\text{GB},k} \left\lceil \frac{t - (u_k - u_j - \sum_{p=j}^{k} L^h_{ST,p})}{p^h_{\text{GCL}} - L^h_{\text{ST,GCL}}} \right\rceil^+ \stackrel{def}{=} F_j(t) \tag{32}$$

Let $s \geq 0$, it exists o_j such that $s \in [o_{j-1}, o_j)$. If $s \in [o_{j-1}, u_j]$, $\Delta t_{\text{GB}}(s, s+t) = \Delta t_{\text{GB}}(u_j, s+t)$. Using eq. (28), (32), $\Delta t_{\text{GB}}(s, s+t) \leq F_j(s+t-u_j-\Delta t_{\text{ST}}(u_j, s+t))$. From Chasles's relation and $\Delta t_{\text{ST}}(s, u_j) \leq u_j - s$ comes $s - u_j - \Delta t_{\text{ST}}(u_j, s+t) \leq -\Delta t_{\text{ST}}(s, s+t)$. And since F_j is non decreasing, it leads to

$$\Delta t_{\text{GB}}(s, s+t) \leq F_j(t - \Delta t_{\text{ST}}(s, s+t)) \tag{33}$$

If $s \in [u_j, o_j)$, $\Delta t_{\text{GB}}(s, s+t) = \Delta t_{\text{GB}}(u_j, s+t) - (s - u_j)$. But for any x, y, z, $z \leq y$ $\Delta t_{\text{GB}}(x, x+y) - z \leq \Delta t_{\text{GB}}(x, x+y-z)$, so $\Delta t_{\text{GB}}(s, s+t) \leq \Delta t_{\text{GB}}(u_j, u_j+t) \leq F_j(t - \Delta t_{\text{ST}}(u_j, u_j+t))$. Consider now Δt_{ST}: from $s \in [u_j, o_j]$, $\Delta t_{\text{ST}}(u_j, u_j+t) = \Delta t_{\text{ST}}(s, u_j+t)$, and from Chalses's relation $\Delta t_{\text{ST}}(s, s+t) = \Delta t_{\text{ST}}(s, u_j+t) + \Delta t_{\text{ST}}(u_j+t, s+t)$. Since $s \in [u_j, o_j]$, $s - u_j \leq o_j - u_j = L^{h,M_i}_{\text{GB},j}$, then $\Delta t_{\text{ST}}(u_j+t, s+t) \leq \Delta t_{\text{ST}}(u_j+t, o_j+t) \leq L^{h,M_i}_{\text{GB},j}$. So, $-\Delta t_{\text{ST}}(u_j, u_j+t) \leq -\Delta t_{\text{ST}}(s, s+t) + L^{h,M_i}_{\text{GB},j}$, and since F_j is non decreasing,

$$\Delta t_{\text{GB}}(s, s+t) \leq F_j(t + L^{h,M_i}_{\text{GB},j} - \Delta t_{\text{ST}}(s, s+t)). \tag{34}$$

Last, to get rid of the position of s wrt o_i and u_i, one gets the maximum

$$\Delta t_{\text{GB}}(s, s+t) \leq \max_{j \in \mathbb{N}} \left\{ F_j(t + L^{h,M_i}_{\text{GB},j} - \Delta t_{\text{ST}}(s, s+t)) \right\} \tag{35}$$
$$= \max_{0 \leq j \leq N^h_{ST}-1} \left\{ F_j(t + L^{h,M_i}_{\text{GB},j} - \Delta t_{\text{ST}}(s, s+t)) \right\} \tag{36}$$

Last, we are looking for a bound on $C \cdot \Delta t_{\text{ST}}(s, s+t)$ and $t + L^{h,M_i}_{\text{GB},j} - u_k + u_j = t + L^{h,M_i}_{\text{GB},j} - (o^h_k - L^{h,M_i}_{\text{GB},k}) + o^h_j - L^{h,M_i}_{\text{GB},j} = t - o^h_{k,j} + L^{h,M_i}_{\text{GB},k}$, leading to eq. (25).

■

Lemma 3: (Linear bound on staircase function) Let $P \in \mathbb{R}^+$, $\{o_1, ..., o_n\}$, $\{l_1, ..., l_n\} \in \mathbb{R}^+$ be two sets of non-negative values such that $o_1 \leq o_2 \leq ... \leq o_n \leq P$. Let $f(t) = \sum_{j=1}^{n} l_j \left\lceil \frac{t-o_j}{P} \right\rceil$. Then for all $t \in \mathbb{R}^+$

$$f(t) \leq \rho t + \max\{\sigma_1, ... \sigma_n\}, \tag{37}$$

with $\sigma_k = \sum_{j=1}^{k} l_j - \rho o_k$, $\rho = \sum_{j=1}^{n} l_j / P$.

Proof: For any k, let $l_k : t \mapsto \sigma_k + \rho t$. Since $f(t+P) = f(t) + \rho P$, if $\forall u \in [0, P]$, $f(u) \leq l_k(u)$, then $f \leq l_k$. Notice that $l_k(o_k) = f(o_k)$, and $\forall u \in [o_k, o_{k+1}]$ $f(u) \leq \max\{l_k, l_{k+1}\}$ (cf. Fig. 5). So, $\forall u \in [0, P]$, $f(u) \leq \max_{k \in [1,n]}\{l_k(u)\}$.

■

D. Tighter latency bounds by introducing shaping curves for arrival of AVB traffic

According to Network Calculus, the upper bound latency of a Class M_i flow $\tau_{M_i[k]}$ in output port h is given by the maximum horizontal deviation between the aggregate arrival curve $\alpha^h_{M_i}(t)$ of intersecting flows of AVB Class M_i and the service curve $\beta^h_{M_i(I)}(t)$ for AVB Class M_i in h,

$$D^h_{M_i[k]} = h(\alpha^h_{M_i}(t), \beta^h_{M_i(I)}(t)), \tag{38}$$

where the service curve $\beta^h_{M_i(I)}(t)$ is from Theorem 1.

The individual arrival curve for the flow $\tau_{M_i[k]}$ in the source ES (h_0) can be given by

$$\alpha^{h_0}_{M_i[k]}(t) = l_{M_i[k]} + \frac{l_{M_i[k]}}{p_{M_i[k]}} \cdot t \tag{39}$$

where $l_{M_i[k]}$ is the burst of the flow $\tau_{M_i[k]}$ in h_0, $l_{M_i[k]}/p_{M_i[k]}$ is the long-term rate of $\tau_{M_i[k]}$ sending from h_0. The individual arrival curve $\alpha^h_{M_i[k]}(t)$ for $\tau_{M_i[k]}$ in the intermediate node is calculated from the previous node port h' along the path of $\tau_{M_i[k]}$,

$$\alpha^h_{M_i[k]}(t) = \alpha^{h'}_{M_i[k]}(t) \oslash \delta_{D^{h'}_{M_i[k]}}(t), \tag{40}$$

where $D_{M_i[k]}$ is the maximum queuing latency of the flow $\tau_{M_i[k]}$ in output port of h', and $\delta_D(t)$ is the burst-delay function [27] which equals to 0 if $t \leq D$ and ∞ otherwise.

Moreover, we provide a solution that leads to tighter aggregate arrival curves for AVB traffic thus reducing pessimism of worst-case delays for AVB flows in intermediate nodes, by taking flows $\tau_{M_i[k]}$ of Class M_i from the same previous node port h' as a group. On one hand, such grouped flows are limited by the physical link speed, meaning that they cannot arrive simultaneously. The shaping curve of physical link is given by,

$$\sigma_{link}(t) = C \cdot t. \tag{41}$$

On the other hand, the grouped flows are constrained by CBS as well. The shaping curve of CBS $\sigma^{h'}_{M_i}(t)$ will be discussed in Sect. V-E with the consideration of ST effects.

Thus the aggregate arrival curve $\alpha^h_{M_i,h'}(t)$ of grouped flows in the input of h from the preceding node port h' is constrained by all three aspects, *i.e.*, sum of individual output arrival curves from h' (Eq. (40)), the shaping curve of physical link (Eq. (41)) and the shaping curve of CBS (Eq. (46)). Then it is given by,

$$\alpha^h_{M_i,h'}(t) = \left(\sum_{\tau_{M_i[k]} \in [h',h]} \alpha^h_{M_i[k]}(t) \right) \wedge \left(\sigma_{link}(t) + l^{h,max}_{M_i,h'} \right) \wedge \left(\sigma^{h'}_{M_i}(t) + l^{h,max}_{M_i,h'} \right), \tag{42}$$

where $x \wedge y = min\{x, y\}$, $l^{h,max}_{M_i,h'}$ is the maximum frame size of flows of Class M_i from h' to h.

Then the aggregate arrival curve $\alpha^h_{M_i}(t)$ for the output port h is the sum of all grouping arrival curves of Class M_i. Moreover, it is assumed that there is no limitation for arriving of flows in the source ES (h_0). Then they can simultaneously arrive and thus the aggregate arrival curve $\alpha^{h_0}_{M_i}(t)$ for the port h_0 is the sum of all individual arrival curves of Class M_i flows.

By disseminating the computation of latency bounds along the routing of $\tau_{M_i[k]}$, its WCD is obtained by the sum of delays from its source ES to its destination ES,

$$D_{M_i[k]} = \sum_{h \in dr_{M_i[k]}} D^h_{M_i[k]} + (|h| - 1) \cdot d_{tech}, \tag{43}$$

where $|h|-1$ represents the number of SWs the flow passing through and d_{tech} is the constant technical latency in a SW.

E. CBS Shaping Curve for AVB traffic

Lemma 4: With the non-preemption mode, the strict service curve for ST traffic in an output port h is given by,

$$\beta^h_{ST}(t) = min_{0\le i\le N^h_{ST}-1}\left\{\sum_{j=i}^{i+N^h_{ST}-1}\beta^h_{TDMA}(t+t_0, L^h_{ST,j})\right\}, \quad (44)$$

where

$$\beta^h_{TDMA}(t,L) = C\cdot max\left\{\left\lfloor\frac{t}{p^h_{\mathrm{GCL}}}\right\rfloor L, t-\left\lceil\frac{t}{p^h_{\mathrm{GCL}}}\right\rceil(p^h_{\mathrm{GCL}}-L)\right\},$$

and

$$t_0 = p^h_{\mathrm{GCL}} - L^h_{ST,j} - o^h_{0,i} - o^h_{j,i}.$$

Proof: As discussed in Sect. V-A, for the given GCL in an output port, there are N^h_{ST} number of ST windows in the GCL period p^h_{GCL}. Then the service for ST traffic can be taken as N^h_{ST} sets of periodic windows of known lengths and relative offsets.

If considering one set of periodic (p^h_{GCL}) ST windows of length L_{ST}, its service in the port h is similar to the TDMA communication [28]. In the worst-case, the service cannot be guaranteed for ST traffic during any time interval $0\le\Delta t < p^h_{\mathrm{GCL}} - L^h_{ST}$, but can be guaranteed of $C\cdot(\Delta t - p^h_{\mathrm{GCL}} - L^h_{ST})$ in any time interval $p^h_{\mathrm{GCL}} - L^h_{ST} \le \Delta t < p^h_{\mathrm{GCL}}$. Then the service curve for periodic ST windows during Δt can be given by,

$$\begin{aligned}&\beta^h_{TDMA}(\Delta t, L^h_{ST})\\ &= C\cdot max\left\{\left\lfloor\frac{\Delta t}{p^h_{\mathrm{GCL}}}\right\rfloor L^h_{ST}, \Delta t-\left\lceil\frac{\Delta t}{p^h_{\mathrm{GCL}}}\right\rceil(p^h_{\mathrm{GCL}}-L^h_{ST})\right\}.\end{aligned} \quad (45)$$

Then the service for all N^h_{ST} sets of periodic ST windows is derived from (45) as well as the relative offsets between two ST windows from different sets. Taking the ith ($i\in[0, N^h_{ST}-1]$) ST window as the reference, we have that the first served frame is from the ith ST window during the backlogged period Δt. Assume that $o^h_{0,i}$ is the maximum idle time interval before the opening time of the ith window, then the service guaranteed for the ith set of ST windows is the curve in (45) shifted to the left with the positive value $p^h_{\mathrm{GCL}} - L^h_{ST,i} - o^h_{0,i}$,

$$\beta^h_{ST,i,i}(\Delta t) = \beta^h_{TDMA}(\Delta t + p^h_{\mathrm{GCL}} - L^h_{ST,i} - o^h_{0,i}, L^h_{ST,i}).$$

Moreover, for the other set of the jth ($j\in[i+1, i+N^h_{ST}-1]$) traffic windows, with the known of the relative offset $o^h_{j,i}$ by considering the ith ST window as the benchmark, the service guaranteed for ST traffic in the jth set of windows can be given by shifting the curve in (45) to the left with the positive value $p^h_{\mathrm{GCL}} - L^h_{ST,j} - o^h_{0,i} - o^h_{j,i}$,

$$\beta^h_{ST,j,i}(\Delta t) = \beta^h_{TDMA}(\Delta t + p^h_{\mathrm{GCL}} - L^h_{ST,j} - o^h_{0,i} - o^h_{j,i}, L^h_{ST,j}).$$

Note that $\beta^h_{ST,j,i}(\Delta t)$ equals to $\beta^h_{ST,i,i}(\Delta t)$ if $j=i$.

Thus if the ith ST window is as benchmark, the strict service for ST traffic is as follows after considering all N^h_{ST} sets of periodic ST windows,

$$\beta^h_{ST,i}(\Delta t) = \sum_{j=i}^{i+N^h_{ST}-1}\beta^h_{ST,j,i}(\Delta t).$$

Then the strict service curve for ST traffic in an output port h is the lower envelope of $\beta^h_{ST,i}(\Delta t)$ by considering each ST windows in the GCL period as benchmark,

$$\beta^h_{ST}(\Delta t) = min_{0\le i\le N^h_{ST}-1}\{\beta^h_{ST,i}(\Delta t)\}.$$

■

Theorem 5: The CBS shaping curve of Class M_i for the non-preemption mode is given by,

$$\sigma^h_{M_i}(t) = \left[t - \frac{\beta^h_{ST}(t)}{C}\right]^+_{\uparrow}\cdot idSl_{M_i} + c^{max}_{M_i} - c^{min}_{M_i}, \quad (46)$$

where $\beta^h_{ST}(t)$ is given by the Lemma 4, and $c^{max}_{M_i}$ and $c^{min}_{M_i}$ are respectively the upper and lower bounds of credit of M_i given by Theorems 3 and 2.

Proof: For an arbitrary period of time Δt, the variation of credit during Δt satisfies

$$\begin{aligned}\Delta c_{M_i} &= c_{M_i}(t+\Delta t) - c_{M_i}(t) = \Delta t^+\cdot idSl_{M_i} + \Delta t^-\cdot sdSl_{M_i}\\ &= (\Delta t - \Delta t^0)\cdot idSl_{M_i} - \Delta t^-\cdot(idSl_{M_i} - sdSl_{M_i}).\end{aligned} \quad (47)$$

In the best-case, the frozen duration $\Delta t^0 = \Delta t^{ST}$ is only related to ST windows, and has nothing to do with guard bands, for the non-preemption integration mode. Considering the strict service curve of ST traffic expressed in the Lemma 4, Δt^{ST} is limited by,

$$R^{h*}_{ST}(t+\Delta t) - R^{h*}_{ST}(t) = \Delta t^{ST}\cdot C \ge \beta^h_{ST}(\Delta t). \quad (48)$$

Moreover, Δt^- can be expressed by,

$$R^{h*}_{M_i}(t+\Delta t) - R^{h*}_{M_i}(t) = \Delta t^-\cdot C, \quad (49)$$

and Δc_{M_i} satisfies the relationship as follows,

$$\Delta c_{M_i} \ge c^{min}_{M_i} - c^{max}_{M_i}. \quad (50)$$

Due to non-decreasing function $R^{h*}_{M_i}(t)$ and using the expressions (48), (49), (50) and $sdSl_{M_i} = idSl_{M_i} - C$, (47) is modified to obtain,

$$R^{h*}_{M_i}(t+\Delta t) - R^{h*}_{M_i}(t) \le \left[\Delta t - \frac{\beta^h_{ST}(\Delta t)}{C}\right]^+_{\uparrow}\cdot idSl_{M_i} + c^{max}_{M_i} - c^{min}_{M_i}.$$

■

VI. Experimental Results

We have evaluated our proposed improving Network Calculus-based WCD analysis for multiple AVB classes with consideration of both frozen and non-frozen credit during guard band (GB) in TSN (called impNC/TSN) as follows. impNC/TSN is implemented in C++ using the Java kernel of the RTC toolbox [29], running on a computer with Intel Core i7-8550U CPU at 1.80 GHz and 8 GB of RAM.

A. Synthetic Test Cases

In this section, our evaluation focuses on a topology of 6 ESes and 2 SWs, connected via physical links with rates of 100Mb/s. The test case (TC1) has 12 ST flows and 10 AVB flows with two classes (M_1 and M_2). Only two classes of AVB traffic are considered in TC1, as we are interested in comparing our improved method (impNC/TSN) with the other existing methods [23], [24] from related work, which only support the analysis for two classes of AVB traffic. The average load of links is about 14.4% and the maximum link load is about 33.7%. The idle slopes of Class M_1 and Class M_2 are respectively set to 40% and 20% [(1)].

[(1)] More detailed information for TC1 and TC2 can be downloaded from https://zenodo.org/record/3785915#.XrDsH2j7RPZ

TABLE I
COMPARED RESULTS BY DIFFERENT METHODS

Flow	EIA17 (μs)	NC/TSN18 (μs)	impNC/TSN (μs)	Sim (μs)
AVB1	3798	3449	2456	1448
AVB2	3778	3273	2379	1559
AVB3	4771	3663	2466	1529
AVB4	5230	5635	4442	2699
AVB5	4971	4612	3865	1721
AVB6	5151	5287	4421	2615
AVB7	4969	3744	2331	1306
AVB8	4642	4236	3490	1996
AVB9	4757	5635	4442	2494
AVB10	5005	5287	4421	2436

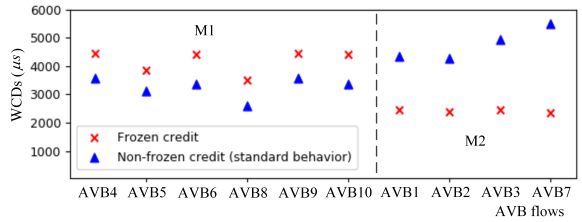

Fig. 6. Comparison of WCDs with frozen and non-frozen credit during GB

Firstly, we are interested to show the improvement of AVB WCDs in TSN network obtained with our proposed method by considering physical link and CBS shaping curves in comparison to existing work [23], [24]. The earliest method proposed for AVB analysis in TSN network is based on Eligible Interval Analysis [23], which we denote with EIA17. The NC-based method from Zhao [24] without shaping constraints is denoted with NC/TSN18. Additionally, we give simulation results (denoted as Sim) in order to show the correctness of the proposed method to some extent. Note that here we use the same assumption of the credit variation as related work, i.e., credit frozen during guard band.

The results for each AVB flow of TC1 are presented in Table I, calculated from different timing analysis methods. As we can see, compared with our impNC/TSN method, EIA17 suffers the larger WCDs, with 43.6% larger on average. This is expected, since EIA17 does not consider the relative locations of ST windows. The pessimism of EIA17 will be even larger with the increasing interval among ST windows. Compared with NC/TSN18, our method by introducing physical link shaping and CBS shaping curves is able to significantly reduce the WCDs compared to [24]. As we can see from the third and forth columns in Table I, impNC/TSN is able to reduce on average the WCDs obtained with [24] by 23.5% and as much as 37.7% in some cases. Additionally, we give simulation results in the last column. The simulation results denoted with Sim are useful for validating our approach, as our obtained WCDs should all be larger than the maximum delays obtained by simulation, which is the case in our experiments. However, as rare events can be missed, a simulation approach will not be able to determine the WCDs, hence it is not useful for safety critical applications that require safety guarantees.

For the second experiment, to show the impact of frozen and non-frozen credit during GB on the WCD of AVB traffic, we use the same TC1, but just assume different credit variations during guard band. We show the results in Fig. 6. The values

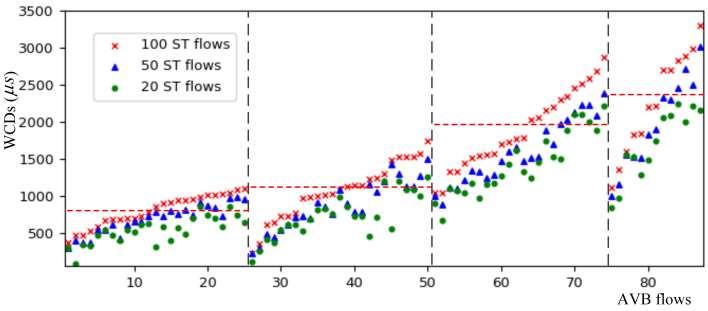

Fig. 7. Compared WCDs with different number of ST flows

on the x-axis give the AVB flows, from AVB1 to AVB10, and on the y-axis show the WCDs in microseconds. As it can seen in Fig. 6, WCDs of Class M_1 under frozen credit assumption are larger than the Class M_1 results under the non-frozen credit assumption, while WCDs of Class M_2 are just the opposite. This confirms our intuition, since if considering non-frozen credit, both credits for Class M_1 and Class M_2 will be increased during GB. Moreover, as Class M_2 has lower priority than Class M_1, the increased possibility of M_1 obtaining service means that the possibility of M_2 getting service is reduced.

B. Evaluation on a Larger Realistic Test Case

In this section, we use a larger real-world test case (TC2), adapted from the Orion Crew Exploration Vehicle (CEV) [30] by using the same topology and ST flows, and considering rate-constrained (RC) flows as AVB flows. We investigate the scalability of our method, show the extension of the analysis method on multiple AVB classes and investigate the influence of varied idle slopes on multiple classes of AVB traffic. Without loss of generality, the credit is assumed to be frozen during GB in this subsection. CEV has a topology consisting of 31 ESes, 15 SWs, 188 dataflow routes, connected by dataflow links transmitting at 1 Gbps, and running 100 ST flows, 25 AVB flows of Class M_1, 25 AVB flows of Class M_2, 24 AVB flows of Class M_3 and 13 AVB flows of Class M_4 (including multicast flows).

In such experiment, we would like to evaluate the scalability of our analysis impNC/TSN with the number of ST flows. We reduce the total number of ST flows from 100 to 50 and 20 respectively. The idle slope for Classes M_i $(i = 1,2,3,4)$ are respectively 40%, 20%, 10% and 5%. The obtained results are visually shown in Fig. 7 As shown in Fig. 7, the obtained results are grouped by the AVB Classes with vertical dotted lines and respectively sorted in increasing order by results from the case with 100 ST flows. As expected and can be seen from Fig. 7, the WCDs of the AVB flows grow with the increasing number of ST flows. In addition, the average WCD (horizontal dotted lines) of Class M_i is smaller than M_{i+1}, since the AVB class of higher priority has larger bandwidth guarantee, i.e., $idSl_{M_i} > idSl_{M_{i+1}}$.

VII. CONCLUSION

The TSN IEEE task group is defining new extensions of Ethernet, devoted to real-time and safety-critical application areas. These types of applications require a method to bound the worst-case delays of a given configuration. This paper has presented an improved Network Calculus-based approach by

considering the limitations from physical link rate and the output of CBS, which are modeled as shaping curves to compute tighter bounds. Moreover, the timing analysis method has be extended to arbitrary number of AVB traffic classes, considering both frozen and non-frozen of credit during guard band, in the case of a TSN/GCL+CBS architecture.

Our analysis is, to the best of our knowledge, the first one to handle with both credit frozen and non-frozen during guard bands and with any number of AVB classes under the influence of ST traffic for the whole TSN network. We have evaluated the proposed approach on both synthetic and realistic test cases. The comparison results to the existing approaches show that the Network Calculus approach is a viable approach for the analysis of TSN. Our approach provides safe upper bounds on WCDs, reduces the pessimism of the analysis (tighter WCD bounds), and is scalable to handle large problem sizes.

REFERENCES

[1] IEEE, "802.3 Standard for Ethernet," 2015.

[2] J. D. Decotignie, "Ethernet-based real-time and industrial communications," *Proceedings of the IEEE*, vol. 93, no. 6, pp. 1102-1117, 2005.

[3] ARINC 664, Aircraft Data Network, Part 7: Deterministic networks, 2003.

[4] SAE AS6802: Time-Triggered Ethernet, Technical report, 2011.

[5] D. Jansen, and B. Holger, "Real-time Ethernet: the EtherCAT solution," *Computing and Control Engineering*, vol. 15, no. 1, pp. 16-21, 2004.

[6] IEEE, "Time-Sensitive Networking Task Group," http://www.ieee802.org/1/pages/tsn.html, 2016.

[7] IEEE, "802.1Q—IEEE Standard for Local and Metropolitan Area Networks—Bridges and Bridged Networks," https://standards.ieee.org/standard/802_1Q-2018.html, 2018.

[8] IEEE, "802.1BA—Audio Video Bridging (AVB) Systems," http://www.ieee802.org/1/pages/802.1ba.html, 2011.

[9] IEEE, "802.1Qbv—Enhancements for Scheduled Traffic," http://www.ieee802.org/1/pages/802.1bv.html, 2015.

[10] IEEE, "802.1ASrev—Timing and Synchronization for Time-Sensitive Applications," http://www.ieee802.org/1/pages/802.1AS-rev.html, 2017.

[11] S. S. Craciunas, R. Serna Oliver, M. Chmelik, and W. Steiner, "Scheduling Real-Time Communication in IEEE 802.1Qbv Time Sensitive Networks," in *Proc. of the 24th Int. Conf. on Real-Time Networks and Systems*, 2016.

[12] L. Zhao, P. Pop, and S. S. Craciunas, "Worst-case latency analysis for IEEE 802.1 Qbv time sensitive networks using network calculus," *IEEE Access*, vol. 6, pp. 41803-41815, 2018.

[13] P. Zhang, Y. Liu, J. Shi, Y. Huang, and Y. Zhao, "A Feasibility Analysis Framework of Time-Sensitive Networking Using Real-Time Calculus," *IEEE Access*, vol. 7, pp. 90069-90081, 2019.

[14] J. A. R. Azua, and M. Boyer, "Complete modelling of AVB in network calculus framework," in *Proc. of the 22nd Int. Conf. on Real-Time Networks and Systems*, 2014.

[15] A. Philip, D. Thiele, R. Ernst, and J. Diemer, "Exploiting shaper context to improve performance bounds of ethernet avb networks," in *Proc. of the 51st Annual Design Automation Conference*, 2014.

[16] L. Zhao, F. He, and E. Li, "Improving worst-case delay analysis for traffic of additional stream reservation class in Ethernet-AVB Network," *Sensors*, vol 18, no. 11, 2018.

[17] J. Y. Cao, P. J. L. Cuijpers, R. J. Bril, and J. J. Lukkien, "Independent WCRT analysis for individual priority classes in Ethernet AVB. *Real-Time Systems*, vol. 54, no. 4, pp. 861-911, 2018.

[18] L. Zhao, P. Pop, Q. Li, J. Y. Chen, and H. G. Xiong, "Timing analysis of rate-constrained traffic in TTEthernet using network calculus," *Real-Time Systems*, vol. 53, no. 2, pp. 254-287, 2017.

[19] D. Thiele, R. Ernst, J. Diemer "Formal Worst-Case Timing Analysis of Ethernet TSN's Time-Aware and Peristaltic Shapers," in *Proc. of the IEEE Vehicular Networking Conference*, 2015.

[20] 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. of the 30th International Teletraffic Congress*, 2018.

[21] J. Zhang, L. Chen, T. Wang, and X. Wang, "Analysis of TSN for industrial automation based on network calculus," in *Proc. of 34th IEEE Int. Conf. on Emerging Technologies and Factory Automation*, 2019.

[22] S. M. Laursen, P. Pop, and W. Steiner, "Routing Optimization of AVB Streams in TSN Networks," *ACM Sigbed Review*, vol. 13, no. 4, pp. 43-48, 2016.

[23] D. Maxim, and Y. Q. Song, "Delay Analysis of AVB traffic in Time-Sensitive Networks (TSN)," in *Proc. of the 25th Int. Conf. on Real-Time Networks and Systems*, 2017.

[24] L. X. Zhao, P. Pop, Z. Zheng, and Q. Li, "Timing Analysis of AVB Traffic in TSN Networks using Network Calculus," in *Proc. of IEEE Real Time and Embedded Technology and Applications Symposium*, 2018.

[25] E. Mohammadpour, E. Stai, and J. Y. Le Boudec, "Improved Credit Bounds for the Credit-Based Shaper in Time-Sensitive Networking," *IEEE Networking Letters*, vol. 1, no. 3, 2019.

[26] H. Daigmorte, and M. Boyer, "Impact on credit freeze before gate closing in CBS and GCL integration into TSN," in *Proc. of International Conference on Real-Time Networks and Systems*, 2019.

[27] J. Y. Le Boudec, and P. Thiran, "Network calculus: A theory of deterministic queuing systems for the internet," *Springer-Verlag Lecture Notes on Computer Science*, 5th ed., New York, 2001.

[28] W. Ernesto, L. Thiele, "Optimal TDMA time slot and cycle length allocation for hard real-time systems," *Proc. of the 2006 Asia and South Pacific Design Automation Conference*, 2006.

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

[30] D. Tamas-Selicean, P. Pop, and W. Steiner, "Design optimization of TTEthernet-based distributed real-time systems," *Real-Time Systems*, vol. 51, no. 1, pp.1-35, 2015.