

Received March 28, 2022, accepted April 11, 2022, date of publication April 18, 2022, date of current version April 28, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3168136

# Worst-Case Latency Analysis for AVB Traffic Under Overlapping-Based Time-Triggered Windows in Time-Sensitive Networks

**KHALED M. SHALGHUM**<sup>1,2,3</sup>, (Student Member, IEEE),

**NOR K. NOORDIN**<sup>1,3</sup>, (Senior Member, IEEE),

**ADUWATI SALI**<sup>1,3</sup>, (Senior Member, IEEE), AND

**FAZIRULHISYAM HASHIM**<sup>1,3</sup>, (Member, IEEE)

<sup>1</sup>Department of Computer and Communication Engineering, Universiti Putra Malaysia, Serdang 43400, Malaysia

<sup>2</sup>Department of Electrical and Electronic Engineering, Azzaytuna University, Tarhuna, Libya

<sup>3</sup>Wireless and Photonic Networks Research Centre of Excellence (WiPNet), Faculty of Engineering, Universiti Putra Malaysia (UPM), Serdang 43400, Malaysia

Corresponding authors: Nor K. Noordin (nknordin@upm.edu.my) and Aduwati Sali (aduwati@upm.edu.my)

This work was supported in part by the BIDANET: Parametric Big Data Analytics over Wireless Networks under Grant UPM.RMC.800-3/3/1/GPB/2021/9696300, in part by the IGNITE-Interference Modeling for 5G and Fixed Satellite Services (FSS) Coexistence at mmWave with Climate Change Considerations in the Tropical Region under Grant FRGS/1/2021/TK0/UPM/01/1, and in part by the Universiti Putra Malaysia (UPM) Putra Berimpak under Fundamental Research Grant 9584300.

**ABSTRACT** Deterministic and low end-to-end latency communication is an urgent demand for many safety-critical applications such as autonomous vehicles and automated industries. The time-sensitive network (TSN) is introduced as Ethernet-based amendments in IEEE 802.1 TSN standards to support time-triggered (TT) traffic in these applications. In the presence of TT flows, TSN is designed to integrate Audio/Video Bridging (AVB) and Best Effort (BE) traffic types. Although AVB traffic has a lower priority than TT, it still requires low and deterministic latency performance, which may not be guaranteed under strict predefined TT scheduling constraints. For this reason, a window-overlapping scheduling algorithm is recently proposed in different works as analytical forms for TT latency under overlapping-windows based. But worst-case AVB latency evaluation under overlapped TT windows is also essential for critical optimizations and tradeoffs. In this paper, a worst-case end-to-end delay (*WCD*) for AVB traffic under overlapping-based TT windows (AVB-OBTTW) algorithm is proposed. Separate analytical models are derived using the network calculus (NC) approach for AVB-OBTTW with both non-preemption and preemption mechanisms. Using an actual vehicular use case, the proposed models are evaluated with back-to-back and porosity configurations under light and heavy loading scenarios. For specific AVB credit bounds, a clear *WCD* reduction has been achieved by increasing the overlapping ratio (*OR*), especially under back-to-back configuration. Preemption and non-preemption modes are compared under different loading conditions, resulting in lower *WCDs* using preemption mode than non-preemption, especially with porosity style. Compared to the latest related works, AVB-OBTTW reduces *WCD* bounds and increases unscheduled bandwidth, leading to the highest enhancements with the maximum allowable *OR*.

**INDEX TERMS** Safety-critical real-time systems, time-sensitive network (TSN), network calculus, worst-case latency analysis, AVB traffic, credit-based shaping (CBS).

## I. INTRODUCTION

Deterministic and low latency is a critical and significant design requirement to support urgent real-time applications, such as autonomous vehicles and automation industries.

The associate editor coordinating the review of this manuscript and approving it for publication was Fung Po Tso<sup>1</sup>.

Missing this need may lead to dangerous situations for humans or society. For this purpose, several technologies have been introduced to support related applications. For example, the Ethernet network was proposed recently to be an appropriate communication environment as it has sufficient bandwidth and low cost for such applications. Although multiple protocols have been previously introduced based

on Ethernet, such as Audio/Video Bridging (AVB) Ethernet and time-triggered (TT) Ethernet, none of which can provide safety-critical transmission requirements.

After adding several extensions to TT-Ethernet, the time-sensitive network (TSN) framework is presented in IEEE 802.1 TSN standards to manage and integrate safety-critical applications. These extensions include network management, synchronization, traffic scheduling and reliability aspects to guarantee no congestion loss, extremely low jitter, and deterministic end-to-end latency for time-triggered traffic [1]. With TT traffic, the TSN framework serves AVB (as soft real-time traffic), and Best Effort (BE) traffic without QoS guarantees. These features attract many relevant experts and companies to adopt TSN technology.

Many TSN standards have been presented to manage and control mixed-criticality traffic environments. For the scheduling aspect, the time-aware shaping (TAS) technique is defined in IEEE 802.1Qbv [2] to control TT forwarding according to a time-gating system defined in the gate control list (GCL) schedule in each network node. The predefined schedules (GCLs) are globally synchronized to specify the open/close time intervals for each priority queue at the egress port. The synchronization constraints for end-to-end connections are introduced in IEEE 802.1AS [3]. Based on the GCL pattern, the credit-based shaping (CBS) mechanism is used to share available bandwidth between AVB and BE flows, as defined in IEEE 802.1Qav [4].

Implementing an effective traffic schedule (GCL) for all selected nodes in the transmission path is a critical and complicated problem. The complexity induces not only from how to meet time-triggered requirements but also unscheduled critical time traffic (AVB) must be considered. AVB flows require enough bandwidth to ensure low and deterministic overall latency. Accordingly, several GCL implementations have been presented to guarantee TT requirements with considering AVB traffic. One of the attractive design ideas is to allow TT windows to overlap in each node leading to more available bandwidth for unscheduled traffic, as suggested in [5], [6]. The TT overlaps have been optimized in [7], [8] to ensure worst-case TT latency requirements. Together with the TT QoS needs, the worst-case performance evaluations for AVB traffic are essential under considering TT impacts. These evaluations assist the interested TSN designers to implement appropriate GCLs for each targeted use case. All the previous worst-case AVB latency evaluations have been implemented based on complete isolation between TT windows, as in [9]–[11]. Thus, a comprehensive view of the worst-case AVB latency performance under overlapping-based TT windows is essential to make critical tradeoffs with TT evaluations, resulting in an appropriate GCL design for each use case.

The trusted and safe worst-case latency representations are analytically based as all corner cases can be covered through the design stage. One of the analytical methods in real-time systems is the Network Calculus (NC) [12], which is preferred over other approaches as it produces less pessimistic

latency bounds [13], [14]. For this reason, the NC approach is adopted in this article to formulate the presented model.

In this paper, the worst-case end-to-end latency forms are implemented for AVB- $X$  ( $X \in \{A, B\}$ ) traffic under overlapping-based TT windows, with non-preemption and preemption modes. All worst cases for AVB- $X$  transmissions are assumed to build related  $WCD$  forms as a function of the overlapping ratio ( $OR$ ) between TT windows. The overlapping may occur from one or two sides for the TT window. Accordingly, the presented forms are evaluated under porosity (one-sided overlapping) and back-to-back (two-sides overlapping) configurations. These evaluations (AVB latencies under variable  $OR$ ) with those in [8] (TT latencies under variable  $OR$ ) can be considered as a complete helpful guide for TSN designers to implement more appropriate and trusted GCL schedules that guarantee worst-case latency deadlines for hard/soft real-time traffic. Thus, the main contributions of this paper can be summarized as:

- A worst-case AVB latency under overlapping-based TT windows (AVB-OBTTW) algorithm is proposed. The GCL schedules are mathematically represented in each selected node under adjustable  $OR$  between different priority TT windows in the hyper-period.
- Using Network Calculus (NC) approach, the worst-case end-to-end latency bounds for AVB- $X$  traffic are formulated with non-preemption and preemption modes. A realistic vehicular use case is used to evaluate the AVB-OBTTW algorithm under back-to-back and porosity configurations with light and heavy loading conditions.
- A comparison between non-preemption and preemption modes is provided under both loading scenarios.
- The AVB-OBTTW model reduces AVB- $WCD$  bounds and increases the unscheduled bandwidth compared to the previous works. The lowest  $WCD$ s are obtained with the maximum allowable  $OR$  that raises unscheduled bandwidth to the highest without missing TT deadlines.

The remainder of the article is structured as follows. Related works are discussed in Section II. Section III introduces a relevant background on the TAS mechanism, preemption modes and the CBS technique. The AVB-OBTTW system model with related design decisions is described in Section IV. Section V presents the impact of non-overlapped and overlapped TT open windows. The worst-case end-to-end latency analysis for AVB- $X$  traffic using AVB-OBTTW algorithm and related performance evaluations are presented with critical discussions in Sections VI and VII, respectively. Finally, the paper is concluded in Section VIII.

## II. RELATED WORKS

In this section, relevant scheduling research studies are discussed with related considerations regarding unscheduled critical time traffic, addressing some critical ideas that have been proposed to support unscheduled flows. Further, several analytical models that evaluate AVB performance are discussed here, considering their limitations and objectives.

Finally, this section focuses on the necessity to evaluate AVB performance under more flexible scheduling algorithms targeting more suitable GCL patterns for each use case.

Craciunas *et al.* [15] implemented main scheduling constraints that guarantee deterministic performance for scheduled traffic using pre-defined GCLs under complete timing isolation between scheduled and unscheduled transmissions. Based on [15], several scheduling amendments have been proposed. For example, an enhanced TAS (eTAS) was recently presented in [16] to support nonperiodic or unexpected critical time flows, such as alarms or emergency events. Although the OMNeT++ simulation results guaranteed emergency traffic (ET) deadlines with less impact on scheduled traffic (ST) performance, unscheduled critical time traffic was not considered. Moreover, several researchers argued that implementing GCLs with strict timing constraints to protect TT flows unrestrainedly may result in missing AVB requirements. The related timing constraints include the TT window duration, guard band, offsets, and isolation from other TT windows. Thus, more flexible GCL designs are recommended while ensuring TT demands.

Several scheduling solutions have been proposed to support unscheduled critical time flows. For example, Nasrallah *et al.* [17] proposed adjustable window durations for scheduled and unscheduled transmissions based on related latency deadlines and the network loading conditions. The simulation results proved that queuing delays for the associated traffic type decrease with increasing corresponding window lengths and vice versa. However, the presented algorithm considered only two priority queues at the node output port without differentiating unscheduled queues, which should be served under different transmission constraints to meet reasonable latency deadlines. Gavrilut *et al.* [18], [19] proposed to include AVB in the scheduling algorithm with TT traffic using the greedy randomized adaptive search procedure (GRASP). The results ensure feasible scheduling for AVB flows under low-scale network topology. However, AVB schedulability was not guaranteed under more complicated networking and loading scenarios. In [20], simulation-based end-to-end latency reduction has been obtained by using data compression methods for all TSN traffic types. Nevertheless, all mentioned algorithms were implemented using simulation methods, which fail to cover all corner cases and lead to untrusted performance evaluations.

The TAS mechanism requires a guard band to isolate scheduled and unscheduled transmissions, resulting in waste bandwidth [21]. Different guard band limitations are considered in TSN depending on the preemption mode used, as described in Section III-B. The preemption mechanism is well-defined in IEEE 802.1Qbu [22], allowing some higher priority traffic (express type) to interrupt the transmission of lower priority traffic (preemptable type). However, IEEE 802.1Qbu does not specify which priority queues are set as express and preemptable. Most TSN researchers consider TT queues as express and the others as preemptable.

But Ashjaei *et al.* [23] proposed different assumptions for express and preemptable assignments to evaluate the frame response time. As expected, the response time was improved for express-based queues as long as other lower priority queues were still preemptable. In [24], Guo *et al.* confirmed that the preemption technique could reduce latency and enhance network resource utilization. Recently, Li *et al.* [25] discussed the importance of bandwidth allocation in CBS under TT impact by deriving the shaping and service curves to calculate delay boundaries and backlogs of traffic, leading to an optimized bandwidth allocation. The authors confirmed that reserving more bandwidth for stream reservation traffic may not reduce latency performance, especially under heavy load conditions. Accordingly, bandwidth assignment in CBS should be granted to ensure determinism for traffic load.

Several researchers derived analytical forms to evaluate end-to-end AVB latency based on the AVB-Ethernet protocol, such as [26], [27]. However, the TSN amendments were not considered, i.e., the analysis was done without considering TT effects. In [9]–[11], other analytical models for AVB latency have been proposed with considering TT impact in the TSN system. Zhao *et al.* [9] formulated the worst-case AVB end-to-end latency with both allocation modes; non-preemption and preemption. The authors confirmed a non-overflow condition for the AVB credit. However, more analytical corrections were proposed in [10], [11] based on [9], resulting in less pessimistic AVB boundaries. In [10], Ren *et al.* obtained a slight reduction in AVB latencies under a more precise TT arrival curve compared to [9]. Moreover, Zhao *et al.* [9] implemented the AVB arrival curve as a sum of individual AVB-X arrivals that egress from the previous node without considering data speed and CBS shapers [11]. The same class flows can arrive simultaneously only at the first node (source). At non-first nodes, also the data link speed and the CBS shapers will control AVB-X arrivals. In [11], a more precise AVB arrival curve is considered to implement the worst-case latency for multiple AVB queues. The proposed analysis was implemented based on frozen and non-frozen AVB credits during guard band intervals. The evaluation results in [11] achieved a dramatic latency improvement for all AVB classes over that were obtained in [9].

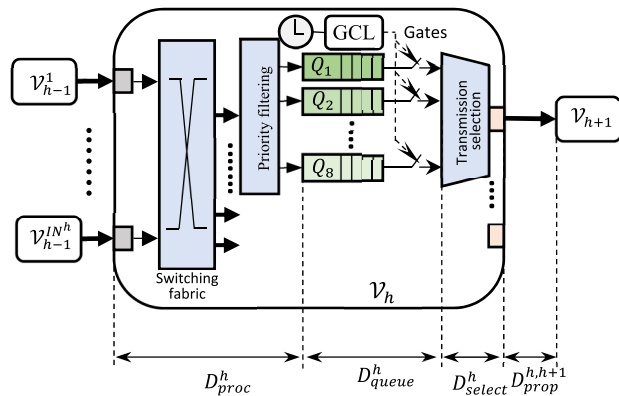
Nevertheless, all the previous analytical models considering AVB latency performance have been implemented based on complete isolation between TT windows. Although allowing TT windows to overlap will influence the overall time-triggered latency performance, complete TT isolations result in considerable waste of bandwidth in guard bands and may lead to missing QoS requirements for unscheduled critical time traffic (AVB). Accordingly, allowing TT to overlap will undoubtedly improve the unscheduled traffic performance [6]. But strict constraints must be applied to bound the overlapping ratio that guarantees TT requirements, as discussed in [8], [7]. Shalghum *et al.* [7] proposed an analytical model for the worst-case end-to-end latency of TT traffic based on flexible window-overlapping scheduling (FWOS) algorithm, which was used to determine the

maximum allowable overlapping ratio (*OR*) that guarantees TT latency deadlines. More latency-based optimizations for FWOS model were introduced in [8] as an optimized FWOS (OFWOS) algorithm, resulting in a complete view for the overlapping effect on the TT latency performance. As the TT overlapping was proposed to support soft real-time flows (AVB), a complete view for the worst-case AVB latency under these overlaps is essential to make critical tradeoffs with TT evaluations and implement the most appropriate GCL designs.

### III. GENERAL BACKGROUND

#### A. TIME-AWARE SHAPING (TAS) MECHANISM

The TSN structure consists of nodes (*V*) connected by physical links. The nodes include end systems (ESs) (information sources and destinations) and switches (SW). Appropriate connections among targeted ESs are managed and established by the stream reservation protocol (SRP) introduced in IEEE 802.1Qcc [28]. The time-aware shaping (TAS) mechanism selects data between TSN nodes, as defined in IEEE 802.1Qbv protocol [2]. The incoming flows are differentiated into eight priority-based queues in each node, as shown in Fig. 1. One or more are assigned for TT flows, two for AVB-A and AVB-B flows, and the remainder for BE flows. Forwarding frames from these queues to the associated egress port is controlled by the predefined GCLs, which specifies the opening and closing events for each queue-gate with guaranteed synchronization between nodes. Only one priority queue is allowed to forward frames when its gate is open by the first-in-first-out (FIFO) technique.

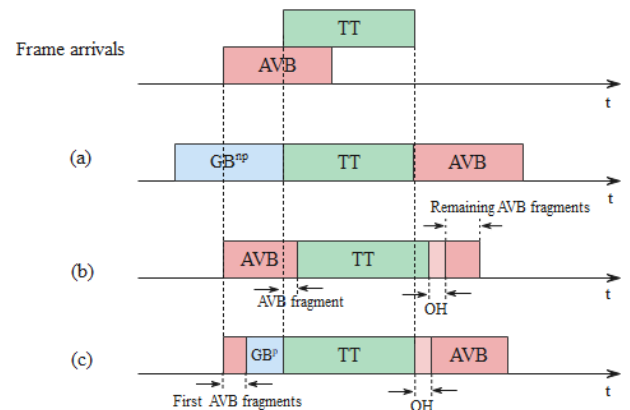


**FIGURE 1.** An IEEE 802.1Qbv capable switch with multiple input and output ports showing delay components between node-to-node connection.

As depicted in Fig. 1, the TSN-aware switch provides the processes of switching fabric, filtering, and data selection for incoming data from  $IN^h$  ingress ports, where  $h$  is the node order. Accordingly, between any nodes,  $h$  and  $h + 1$ , the associated delay includes the processing delay ( $D_{proc}^h$ ), queuing delay ( $D_{queue}^h$ ), selection delay ( $D_{select}^h$ ), and the propagation delay ( $D_{prop}^{h,h+1}$ ).

#### B. PREEMPTION MODES

The TSN framework supports non-preemption and preemption modes with different degrees of protection for critical time traffic, as depicted in Fig. 2.



**FIGURE 2.** Preemption mechanisms in TSN; (a) non-preemption mode, (b) preemption without Hold/Release, and (c) preemption with Hold/Release.

##### 1) NON-PREEMPTION MECHANISM

In this mode, the unscheduled traffic (AVB and BE) cannot be interrupted if it is already in the transmission status. A guard band interval must be allocated in front of each TT window to protect associated transmissions, with a length equal to the maximum transmission unit (MTU) of 1500 bytes in the Ethernet protocol, as depicted in Figure 2(a). The associated gates for all preemptable queues (AVB and BE) are closed during these guard bands. Accordingly, the preemptable traffic will experience a dramatic delay, and the bandwidth consumption will be degraded.

##### 2) PREEMPTION MECHANISM

The IEEE 802.1Qbu standard [12] introduces the preemptable frame to be interrupted when a TT window opens. The TSN standard defines two slightly different versions of preemptions; with HOLD/RELEASE and without HOLD/RELEASE.

- Preemption without HOLD and RELEASE: In this version, the preemptable frames can be transmitted at any time, i.e., there is no close gate for the preemptable queues. When a fragment of a preemptable frame starts to be transmitted before the opening edge of any TT window, the TT frame waits until finishing this fragment, and the remaining preemptable fragments will be resumed once the TT transmission has completed, as shown in Fig. 2(b). Thus, no guard bands are required before TT windows. However, an overhead (OH) is required to isolate and reassemble frames at the destination, with 24 bytes length as given by the standard. The TSN standard specifies fragments with up to 123 bytes that cannot be preempted by TT transmissions [22]



(Annex R-3). Thus, these fragments will increase the experienced TT delay.

- Preemption with HOLD and RELEASE: In this version, all preemptable gates are open if at least a frame is in the associated queue. The Hold and Release feature allows the designer to implement a guard band before the TT window with a relatively more minor length than in the non-preemption mode, as depicted in Fig. 2(c). This guard band is used to preempt fragments, not the entire frame, with sizes larger than 124 bytes [22] (Annex R-3). Thus, this function ensures complete protection for TT frames compared to that without Hold and Release while simultaneously reducing their impact on the bandwidth available for preemptable traffic compared to non-preemption mode.

### C. CREDIT-BASED SHAPING (CBS) MECHANISM

According to path-GCLs, AVB and BE flows share the available bandwidth using the credit-based shaping (CBS) technique in each selected node, as defined in IEEE 802.1Qav [4]. The CBS mechanism controls the AVB traffic forwarding to avoid starvation conditions for lower priority flows. The starvation means that even if an AVB frame has a higher priority than other AVB or BE frames, it cannot dominate the available bandwidth, and all of them will share it according to CBS limitations. The AVB gate selects the frame only if: the gate is open, the frame is allowed to be transmitted by the CBS constraints, and no higher priority AVB frame is being transmitted. The CBS parameters for AVB-X queue include the minimum credit limit required for transmission ( $cr_X^{min}$ ), maximum credit limit ( $cr_X^{max}$ ), sending slope ( $sdSl_X$ ), and idle slope ( $idSl_X$ ). As defined in the IEEE 802.1Qcc [28] and IEEE 802.1Qbv [2] standards, a frame can be forwarded only if the credit is higher than or equal to zero. If the node is selected in a transmission path, the AVB-X credit will be initiated to zero and then decreased with  $sdSl_X$  during AVB-X transmissions, increased with  $idSl_X$  when AVB-X is waiting to be transmitted and frozen during TT and guard band intervals. Furthermore, when the AVB-X credit is positive for an empty queue, the credit is updated as zero. But when the associated credit is negative for an empty queue, it is increased with  $idSl_X$  until zero. The minimum and maximum credit limits for AVB queues can be mathematically derived as in Appendix A.

### IV. SYSTEM MODEL AND DESIGN DECISIONS

There are two methods to combine scheduled and unscheduled transmissions in the TSN system, back-to-back, and porosity configurations, as shown in Fig. 3(a) and (b).

It is noticed that more available bandwidth will be saved for unscheduled transmissions when TT windows overlap, especially with a back-to-back style where the overlapping can be implemented from both window sides. Only one guard band is required during the hyper-period. In [7] and [8], all overlapping situations are considered to express and evaluate the worst-case TT latency behavior with both configuration

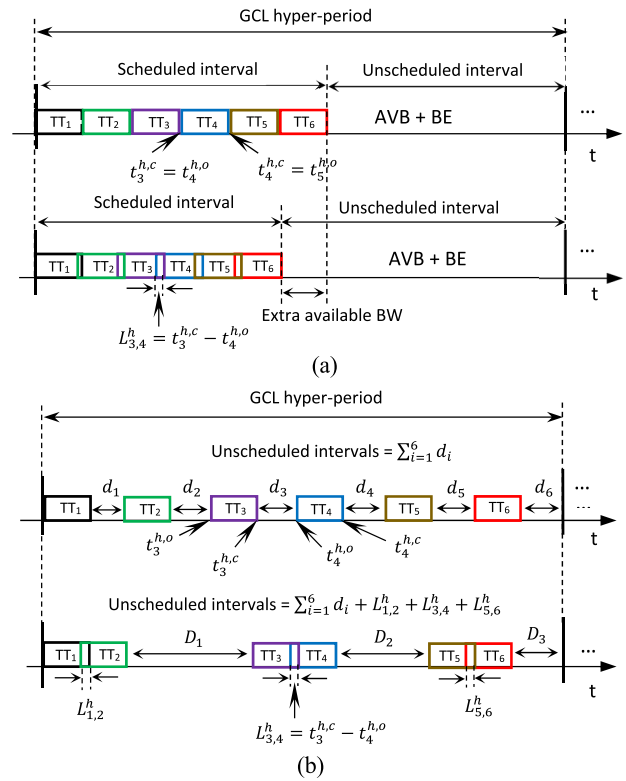
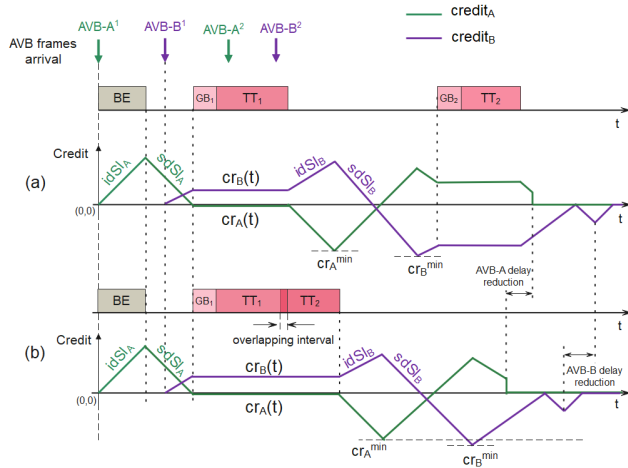


FIGURE 3. Non-overlapping and overlapping-based TT window offsets through the hyper-period with; (a) back-to-back configuration. (b) porosity configuration.

methods. Nevertheless, the worst-case end-to-end latency for AVB traffic under those overlapping situations is critically essential to provide critical optimizations and tradeoffs.

As the AVB transmissions are based on CBS, saving more intervals by minimizing the number of guard bands and overlapping TT windows will reduce the waiting time for the traffic in the associated queue. Fig. 4 shows an example for improving the credit variation for AVB queues under the overlapped TT window. Note that Fig. 4 is drawn assuming both AVB queues are empty when  $t < 0$ . Accordingly, it can be noticed that delay reduction will be obtained for AVB queues under overlapping-based TT windows compared to non-overlapping conditions.

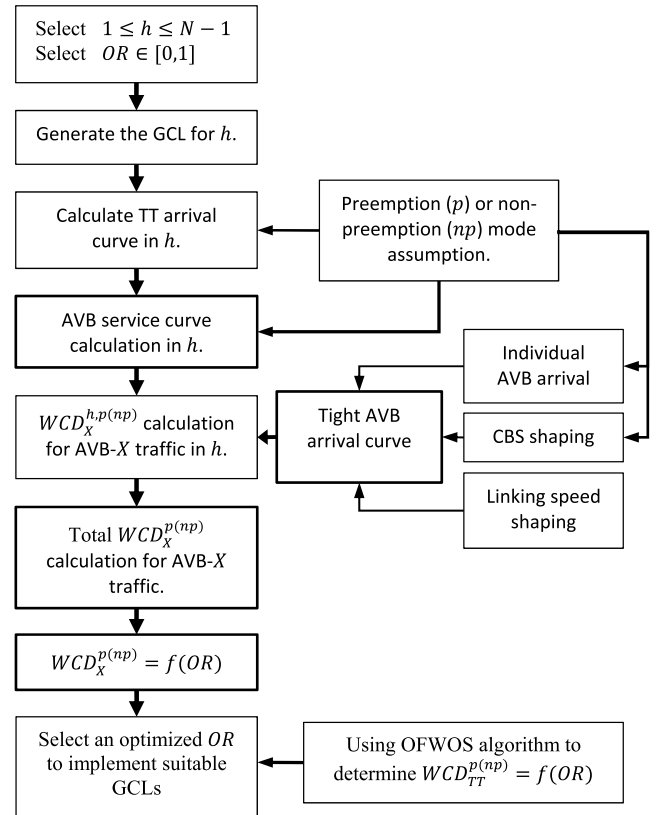
This article formulates the worst-case end-to-end latency for AVB traffic under the overlapping-based TT windows. The overlapping between any two adjacent TT windows  $k$  and  $k + 1$  is denoted as  $L_{k,k+1}^h = t_k^{h,c} - t_{k+1}^{h,o}, \forall 1 \leq k \leq NW_{GCL}^h - 1$ , where  $t_k^{h,c}$  is the closing time of the  $k$ -th TT window,  $t_{k+1}^{h,o}$  is the opening time of the  $(k+1)$ -th TT window, and  $NW_{GCL}^h$  represents the number of TT windows in the hyper-period ( $T_{GCL}^h$ ). For performance evaluation, we consider the overlapping ratio ( $OR$ ) as a design parameter represented by the ratio between the overlapping interval ( $L_{k-1,k}^h$ ) and the overall duration of the left overlapped window ( $W_{TT_{k-1}}^h$ ), as  $OR_{k-1,k}^h = L_{k-1,k}^h / W_{TT_k}^h$ .



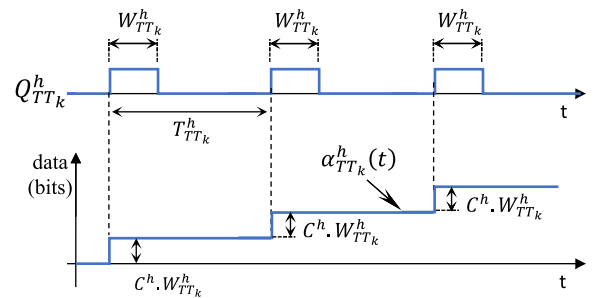
**FIGURE 4.** An example of AVB credit representation under; (a) non-overlapping-based TT windows and (b) overlapping-based TT windows.

The main design stages of the worst-case AVB latency under overlapping-based TT windows (AVB-OBTTW) algorithm are illustrated in Fig. 5. In the first, the GCL is implemented for all selected nodes by assuming variable  $OR$  between TT windows. The implemented GCLs are used to calculate the TT arrival curves with related guard bands effect, and then, the lower bound AVB- $X$  service curve is determined. The resulted service curve is used with the tight AVB- $X$  arrival curve to compute the worst-case end-to-end latency ( $WCD_X^{p(np)}$ ) for AVB- $X$  traffic as a function of  $OR$ . All overlapping situation is considered to evaluate  $WCD$  performance. Thus, gathering the evaluation results in this work with those that consider worst-case latency performance for TT traffic under the same overlapping conditions in [8] (OFWOS algorithm) will introduce a complete view for the TSN designer to select an optimized  $OR$  for appropriate GCL implementations. The aforementioned design stages to build AVB-OBTTW are considered according to the following overall assumptions and limitations.

As the proposed model is implemented based on the TAS mechanism, all network elements are assumed to be fully synchronized without considering any synchronization errors between them. Further, the GCLs for end-to-end selected nodes are offline-based implemented without considering scheduling dynamism. Another significant TSN feature is the frame replication and elimination for reliability (FRER) to ensure zero congestion loss for time-triggered frames, as defined in IEEE 802.1CB. However, this function is not provided in this work. Moreover, the preemption with Hold and Release is assumed here to compare with the non-preemption mode under related guard band limitations, as described in Section III-B. During these guard bands, the credit for AVB queues is assumed to be frozen in the whole system analysis. Note that all the notations used in this article are summarized in Table 1.



**FIGURE 5.** Main design stages of the worst-case AVB latency under overlapping-based TT windows (AVB-OBTTW) algorithm.



**FIGURE 6.** The arrival curve from an individual TT queue.

## V. TT IMPACT CONSIDERATIONS

### A. TT IMPACT BASED ON NON-OVERLAPPING WINDOWS

This section expresses the aggregate arrival curve in a node  $h$  for all TT traffic incoming from other nodes that are multiplexed in  $h$ . For an individual  $k$ -th queue ( $Q_{TT_k}^h$ ) as shown in Fig. 6, the number of associated periodic windows to transmit  $\tau_{TT_k}$  frames from an arbitrary node  $h$  to the transmission link during any arbitrary interval  $\Delta t = [s, t]$ , where  $\forall s, t \in \mathcal{R}^+$ ,  $t \geq s$  and  $\mathcal{R}^+$  is the set of positive real numbers, is  $M_{TT_k}^h(s, t)$ , which can be represented as [29]

$$M_{TT_k}^h(s, t) = \left\lceil \frac{t - s}{T_{TT_k}^h} \right\rceil, \quad (1)$$

TABLE 1. Summary of used notations.

Notation	Description
$N$	Number of nodes in the end-to-end path.
$N_{TT}^h$	Number of TT queues in node $h$ .
$T_{GCL}^h$	The GCL hyper-period
$M_{TT_k}^h$	Number of $k$ -th open windows in the hyper-period.
$NW_{GCL}^h$	Number of TT windows in the hyper-period
$C^h$	Egressing rate from node $h$ .
$W_{TT_k}^h, T_{TT_k}^h$	Duration and period of the $m$ -th window in the hyper-period at node $h$ .
$Q_{TT_k}^h$	The $k$ -th queue and its gate at node $h$ .
$\tau_{TT_k}$	The $k$ -th TT flow.
$\tau_{x_m}$	The $m$ -th AVB- $X$ flow.
$f_{AVB}^{h,max}$	Largest size of AVB frames at node $h$ .
$idSl_X$ & $sdSl_X$	Idle and sending slopes of AVB- $X$ queue.
$cr_X(t)$	Credit of the AVB- $X$ queue.
$cr_X^{min}$ & $cr_X^{max}$	Minimum and maximum credit bounds of AVB- $X$ queue.
$t_k^{h,o}, t_k^{h,c}$	Opening and closing times of the $k$ -th TT window in the hyper-period.
$L_{k,k+1}^h$	Length of overlapping interval between $k$ -th and $(k + 1)$ -th windows at node $h$ .
$OR_{k,k+1}^h$	Overlapping ratio between $k$ -th and $(k + 1)$ -th windows in the hyper-period at node $h$ .
$OD_{j,i}^h$	Offset difference between the $j$ -th and $i$ -th TT windows at node $h$ .
$OH_j^h$	Overhead interval after $j$ -th TT window at node $h$ .
$GAP_{j,j+1}^h$	Expected unoccupied gap between $j$ -th and $(j + 1)$ -th TT windows at node $h$ .
$GB_j^{h,p}$ & $GB_j^{h,np}$	Length of preemption and non-preemption guard bands for the $j$ -th TT window at node $h$ .
$R_{TT_k}^h(t)$	Arrival process of the $k$ -th TT traffic.
$\alpha_{TT_k}^h(t)$	Arrival curve of the $k$ -th TT traffic.
$\alpha_{OH}^{h,i}(t)$	Arrival curve during overhead intervals in the $i$ -the cycle.
$\alpha_{GAP+TT}^{h,i}(t)$	Aggregate arrival curve from TT windows and related preemption protection intervals in the $i$ -the cycle.
$\alpha_{GAP+TT+GAP}^{h,i}(t)$	Aggregate arrival curve from TT windows, related preemption protection intervals and expected unoccupied gaps in the $i$ -the cycle.
$\alpha_{Gnp+TT}^{h,i}(t)$	Aggregate arrival curve from TT windows and related non-preemption protection intervals in the $i$ -the cycle.
$R_X^h(t), R_X^{h*}(t)$	Arrival and departure processes for AVB- $X$ traffic at node $h$ .
$\Delta t_X^+, \Delta t_X^-, \Delta t_X^0$	Credit increasing, decreasing, and frozen intervals for AVB- $X$ queue.
$\sigma_{x_m}^h$ & $\rho_{x_m}^h$	Framing size and long-term rate of $\tau_{x_m}$ at node $h$ .
$\alpha_{X,sum}^{[h-1,h],p(np)}(t)$	Sum of individual AVB- $X$ arrival curves transferred from $h - 1$ to $h$ .
$\alpha_{X,link}^{[h-1,h]}(t)$	Arrival curve of AVB- $X$ traffic with link speed shaper.
$\alpha_{X,CBS}^{[h-1,h],p(np)}(t)$	Arrival curve of AVB- $X$ traffic with CBS constraints.
$f_X^{[h-1,h],max}$	Largest AVB- $X$ frame sent from $h - 1$ to $h$ .
$\alpha_X^{[h-1,h],p(np)}(t)$	Aggregate arrival curve for AVB- $X$ traffic sent from $h - 1$ to $h$ .
$D_{X,select}^h$	Selection delay of the AVB- $X$ frame from node $h$ .

TABLE 1. (Continued.) Summary of used notations.

$D_{prop}^{[h-1,h]}$	Propagation delay between $h - 1$ and $h$ .
$D_{proc}^h$	Processing delay at node $h$ .
$D_{X,queue}^{h,p(np)}$	Queuing delay experienced by the AVB- $X$ frame at node $h$ with preemption (p) or non-preemption (np) mode.
$\beta_X^{h,p(np)}(t)$	Lower bound service curve for the AVB- $X$ frame at node $h$ with preemption (p) or non-preemption (np) mode.
$\alpha_X^{h,p(np)}(t)$	Upper bound arrival curve for the AVB- $X$ frame at node $h$ with preemption (p) or non-preemption (np) mode.
$WCD_X^{h,p(np)}$	Worst-case delay experienced by the AVB- $X$ frame at node $h$ with preemption (p) or non-preemption (np) mode.
$WCD_X^{p(np)}$	Worst-case end-to-end delay experienced by the AVB- $X$ frame with preemption (p) or non-preemption (np) mode.

where  $T_{TT_k}^h$  is the period of  $Q_{TT_k}^h$  windows. By assuming  $R_{TT_k}^h(t)$  is the arrival process of  $Q_{TT_k}^h$  frames in node  $h$ , we have

$$R_{TT_k}^h(t) - R_{TT_k}^h(s) \leq W_{TT_k}^h \cdot C^h \cdot M_{TT_k}^h(s, t) \leq W_{TT_k}^h \cdot C^h \cdot \left\lceil \frac{t-s}{T_{TT_k}^h} \right\rceil, \quad (2)$$

where  $W_{TT_k}^h$  is the duration of  $Q_{TT_k}^h$  window, and  $C^h$  represents the egressing bit rate from node  $h$ . Therefore, the arrival curve for  $Q_{TT_k}^h$  is

$$\alpha_{TT_k}^h(t) = \begin{cases} W_{TT_k}^h \cdot C^h \cdot \left\lceil \frac{t-s}{T_{TT_k}^h} \right\rceil, & t > s \\ 0, & t \leq s. \end{cases} \quad (3)$$

If the number of TT queues in node  $h$  is  $N_{TT}^h$ , the hyper-period is equal to the least common multiple of periods for all  $N_{TT}^h$  queues ( $T_{GCL}^h = \text{LCM}_{1 \leq k \leq N_{TT}^h} (T_{TT_k}^h)$ ).

In each hyper-period, there is  $M_{TT_k}^h$  open windows to transmit  $Q_{TT_k}^h$  frames, where  $M_k^h = T_{GCL}^h / T_{TT_k}^h$ . For a given scheduling timetable (GCL) for an egress port at node  $h$ , the number of all TT windows in  $T_{GCL}^h$  is

$$NW_{GCL}^h = \sum_{k=1}^{N_{TT}^h} M_{TT_k}^h = \sum_{k=1}^{N_{TT}^h} \frac{T_{GCL}^h}{T_{TT_k}^h}. \quad (4)$$

By taking the  $i$ -th TT window ( $i \in NW_{GCL}^h$ ) as a reference with assuming  $s = 0$ , the aggregate TT arrival curve can be represented as

$$\alpha_{TT}^{h,i}(t) = \sum_{j=i}^{i+NW_{GCL}^h-1} W_{TT_j}^h \cdot C^h \cdot \left\lceil \frac{t - OD_{j,i}^h}{T_{GCL}^h} \right\rceil, \quad (5)$$

where  $OD_{j,i}^h$  represents the offset difference between the  $i$ -th and  $j$ -th TT windows in  $T_{GCL}^h$ , i.e.,  $OD_{j,i}^h = t_j^{h,o} - t_i^{h,o}$ .

It can be noticed that the above expression of the aggregate arrival curve is formulated based on complete isolation between TT windows. This article considers the overlapping-based GCLs to implement end-to-end latency for AVB queues, including both the preemption and non-preemption modes.

**B. TT IMPACT BASED ON OVERLAPPING WINDOWS**

This subsection starts with formulating GCL during the hyper-period by defining the opening and closing times for all TT windows with flexible overlapping between them. Then, the aggregate TT arrival curve is represented based on these GCLs for preemption and preemption modes.

**1) GCL FORMULATION BASED ON OVERLAPPING TT WINDOWS**

Each TSN-aware node has a GCL that specifies all queue gates' opening and closing events. In our model, the GCLs are represented as mathematical expressions for each node without considering the priority difference between TT queues, as they have the same impact on the AVB performance. Accordingly, we refer to each TT window by its order in the hyper-period, not by priority. By considering the  $i$ -th TT window as a reference, the associated opening and closing times can be given as,

$$\forall 1 \leq i \leq NW_{GCL}^h$$

$$t_i^{h,o} = OD_{i,0}^h \& t_i^{h,c} = t_i^{h,o} + W_{TT_i}^h, \quad (6)$$

where  $OD_{i,0}^h$  equals the interval from zero to the  $i$ -th opening time. Accordingly, the opening and closing times for other TT windows in the hyper-period can be determined as,

$$\forall i < k \leq i + NW_{GCL}^h - 1$$

$$t_k^{h,o} = t_{k-1}^{h,o} + OD_{k-1,k}^h \& t_k^{h,c} = t_k^{h,o} + W_{TT_k}^h, \quad (7)$$

where  $OD_{k-1,k}^h = t_k^{h,o} - t_{k-1}^{h,o}$ , and based on the offset difference, we can bound the overlapping ratio between  $k - 1$  and  $k$  windows as,

$$OR_{k-1,k}^h = \begin{cases} \frac{t_{k-1}^{h,c} - t_k^{h,o}}{W_{TT_{k-1}}^h}, & OD_{k-1,k}^h > t_{k-1}^{h,c} - OD_{i,0}^h \\ 0, & OD_{k-1,k}^h < t_{k-1}^{h,c} - OD_{i,0}^h. \end{cases} \quad (8)$$

Based on the above expressions, the end-to-end GCLs are implemented, and then, the aggregate TT arrival curve is determined to calculate the worst-case AVB latencies.

**2) AGGREGATE TT ARRIVAL CURVE WITH NON-PREEMPTION MODE**

On a non-preemption basis, a guard band interval is established before each TT window to avoid any expected interference with unscheduled traffic transmissions. During these guard bands, the AVB transmissions are not allowed. The length of the non-preemption guard band for the  $j$ -th TT window ( $GB_j^{h,np}$ , where  $j \in NW_{GCL}^h$ ) equals the time interval required to transmit the largest size of AVB frames that

compete to egress from the associated node, i.e.,  $GB_j^{h,np} = f_{AVB}^{h,max} / C^h$ . It is expected that the time interval between the adjacent windows is less than  $GB_j^{h,np}$ . Accordingly, the  $j$ -th TT window must be protected by

$$G_j^{h,np} = \min \left\{ GB_j^{h,np}, \left[ t_j^{h,o} - t_{j-1}^{h,c} \right]^+ \right\}, \quad (9)$$

where  $[x]^+ = x$  when  $x \geq 0$ , and zero otherwise. Hence, we assume that the credit of both AVB queues is frozen during all  $G_j^{h,np}$  and  $W_{TT_j}^h$  intervals. When the  $i$ -th TT window is considered as a reference, the aggregate arrival curve can be formulated as a sum of arrival curves produced from all TT windows and related guard bands in  $T_{GCL}^h$ , as follows,

$$\alpha_{G^{np+TT}}^{h,i}(t) = \sum_{j=i}^{i+NW_{GCL}^h-1} \left\{ C^h \cdot \left( W_{TT_j}^h + G_j^{h,np} - L_{j,j+1}^h \right) \cdot \left[ \frac{t - OD_{j,i}^h + G_j^{h,np} - G_i^{h,np}}{T_{GCL}^h} \right] \right\}, \quad (10)$$

where  $L_{j-1,j}^h$  is the overlapping interval between  $j$ -th and  $(j + 1)$ -th TT windows, and it equals

$$L_{j,j+1}^h = \left[ t_j^{h,c} - t_{j+1}^{h,o} \right]^+. \quad (11)$$

Hence, to clarify the effect of TT overlapping,  $L_{j,j+1}^h$  is expressed in terms of  $OR$  between each pair of adjacent TT windows, as follows,

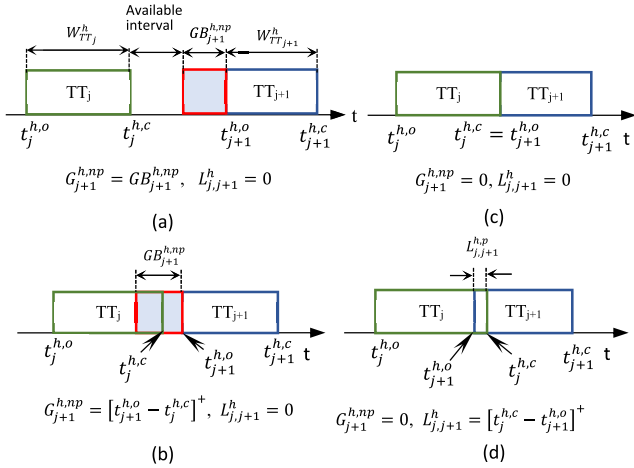
$$L_{j,j+1}^h = OR_{j,j+1}^h \cdot W_{TT_j}^h, \quad (12)$$

where  $OR_{j,j+1}^h$  represents the percent of the overlapping interval between  $j$ -th and  $(j + 1)$ -th TT windows concerning the whole duration of the  $j$ -th TT window. Moreover, as the guard band duration may vary with time depending on the length of AVB/BE frames that compete on the current GCL cycle, the distance between frozen intervals is adjusted by the term  $G_j^{h,np} - G_i^{h,np}$ . The expected overlapping interval is discarded from the frozen interval in each TT window using the term  $L_{j,j+1}^h$ , which is calculated between the current TT window and the other allocated on the right, i.e., the closing-edge overlapping. The opening-edge overlapping interval is calculated as a closing-edge overlapping interval with the previous TT window from the left. Thus, all expected overlaps are considered in all  $NW_{GCL}^h$  TT windows. As examples, different offset situations for adjacent TT windows are presented in Fig. 7 with critical intervals' determinations under non-preemption mode.

**3) AGGREGATE TT ARRIVAL CURVE WITH PREEMPTION MODE**

As classified in Section II-B, the TSN standard supports preemption with Hold/Release and without Hold/Release functions. In this article, the preemption with Hold/Release is considered to avoid degradation in the TT performance. Even a guard band must be used in the Hold/Release feature; it is





**FIGURE 7.** Possible offset situations for adjacent TT windows with critical intervals' determinations under non-preemption mode.

a smaller length than the non-preemption mode. Thus, the protection interval for the  $j$ -th TT window will be limited as

$$G_j^{h,p} = \min \left\{ GB_j^{h,p}, [t_j^{h,o} - t_{j-1}^{h,c}]^+ \right\}, \quad (13)$$

where  $GB_j^{h,p}$  equals the time required to transmit at most 124 bytes of AVB fragment. Under the worst transmission situation, there are no AVB transmissions during protection intervals, and the credit of both AVB queues will be frozen. Similar to the non-preemption case, the aggregate arrival curve can be formulated as a sum of arrival curves produced from all TT windows and related preemption guard bands in  $T_{GCL}^h$  with considering the  $i$ -th TT window as a reference, as follows

$$\alpha_{Gp+TT}^{h,i}(t) = \sum_{j=i}^{i+NW_{GCL}^h-1} \left\{ C^h \cdot \left( W_{TT_j}^h + G_j^{h,p} - L_{j,j+1}^h \right) \cdot \left[ \frac{t - OD_{j,i}^h + G_j^{h,p} - G_i^{h,p}}{T_{GCL}^h} \right] \right\}. \quad (14)$$

With the preemption mode, an overhead interval ( $OH$ ) will be used in front of the remaining AVB fragments after each TT window. It is expected that the gap between the  $j$ -th and  $(j+1)$ -th TT windows ( $GAP_{j,j+1}^h$ ) is not larger than  $OH + GB_{j+1}^{h,p}$ , i.e.,  $GAP_{j,j+1}^h \leq OH + GB_{j+1}^{h,p}$ . In this case, the  $OH$  interval will not be allocated between TT windows, and we assume frozen AVB credits during these gaps. Accordingly, the  $OH$  interval can be determined for the  $j$ -th TT window as follows,

$$OH_j^h = \begin{cases} OH, & t_{j+1}^{h,o} - t_j^{h,c} > GB_j^{h,p} + OH \\ 0, & t_{j+1}^{h,o} - t_j^{h,c} \leq GB_j^{h,p} + OH. \end{cases} \quad (15)$$

Accordingly, the expected gap between  $j$ -th and  $(j+1)$ -th TT windows that cannot be used for AVB transmissions is

determined as

$$GAP_{j,j+1}^h = \begin{cases} t_{j+1}^{h,o} - t_j^{h,c} - GB_j^{h,p}, & OH_j^h = 0 \\ 0, & OH_j^h \neq 0. \end{cases} \quad (16)$$

Fig. 8 shows different situations for the adjacent TT window offsets, determining related intervals under preemption mode. As shown in Fig. 8(a) and (b), it can be noticed that only one of  $OH_j^h$  and  $GAP_{j,j+1}^h$  is larger than zero for each TT window, and both are located after the associated TT window. This means that  $\forall j \in NW_{GCL}^h$  when  $OH_j^h > 0$ ,  $GAP_{j,j+1}^h = 0$ , and vice versa. Hence, the AVB credits during  $OH_j^h$  and  $GAP_{j,j+1}^h$  will be different. In  $OH_j^h$ , the credit increases for the associated AVB queue and decreases for the other. In  $GAP_{j,j+1}^h$ , both AVB credits are frozen as no one can be served. Thus, after considering the  $i$ -th TT window as a reference, the arrival curve during  $OH_j^h$  interval can be given as

$$\alpha_{OH}^{h,i}(t) = \sum_{j=i}^{i+NW_{GCL}^h-1} C^h \cdot OH_j^h \cdot \left[ \frac{t - OD_{j,i}^h - W_j^h}{T_{GCL}^h} \right]. \quad (17)$$

But the arrival curve during  $GAP_{j,j+1}^h$  intervals are combined with TT and protection intervals, as follows

$$\alpha_{Gp+TT+GAP}^{h,i}(t) = \sum_{j=i}^{i+NW_{GCL}^h-1} \left\{ C^h \cdot \left( W_{TT_j}^h + G_j^{h,p} - L_{j,j+1}^h + GAP_{j,j+1}^h \right) \cdot \left[ \frac{t - OD_{j,i}^h + G_j^{h,p} - G_i^{h,p}}{T_{GCL}^h} \right] \right\}. \quad (18)$$

## VI. WORST-CASE LATENCY ANALYSIS FOR AVB TRAFFIC

In this section, the worst-case AVB-X end-to-end latency is formulated for both allocation modes, i.e., non-preemption and preemption. Firstly, the AVB-X service curve is expressed based on the aggregate TT arrival curves determined in the previous section. Then, the AVB-X arrival curve is determined by considering three different shapers. Finally, the worst-case AVB-X latency is found for nodes separately and then for the end-to-end path.

### A. DETERMINATION OF LOWER-BOUND AVB-X SERVICE CURVE ( $\beta_X^{h,p(np)}(t)$ )

#### 1) WITH NON-PREEMPTION MODE

Assume that  $R_X^h(t)$  and  $R_X^{h*}(t)$  are the arrival and departure processes for class AVB-X traffic at node  $h$ . The arrival and departure events will be considered during any arbitrary interval  $\Delta t = [s, t]$ , where  $\forall s, t \in \mathcal{R}^+, t \geq s$ . Assume that all priority queues at the egress port of node  $h$  are empty at  $s$ , i.e.,  $R_X^{h*}(s) = R_X^h(s)$ ,  $R_{TT}^{h*}(s) = R_{TT}^h(s)$  and  $R_{Gnp}^{h*}(s) = R_{Gnp}^h(s)$ . Thus, the AVB credits are zero at  $s$  ( $cr_X^h(s) = 0$ ).  $\Delta t$  can be expressed as  $\Delta t = \Delta t_X^+ + \Delta t_X^- + \Delta t^0$  concerning the status of AVB credit.  $\Delta t^+$  represents the time intervals when the AVB credit is increasing,  $\Delta t^-$  is when the AVB

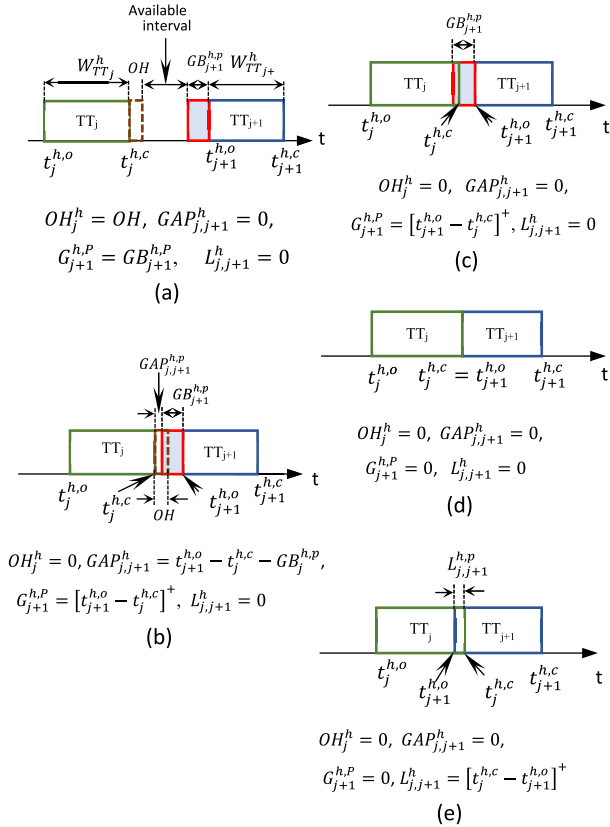


FIGURE 8. Possible offset situations for adjacent TT windows with critical intervals' determinations under preemption mode.

credit is decreasing, and  $\Delta t^0$  is when the AVB credit is frozen. With non-preemption mode,  $\Delta t^0$  represents all TT windows ( $W_{TT_j}^h$ ) and protection intervals ( $G_j^{h,np}$ ), taking into account discarding the overlapping interval between adjacent windows from one of them. For each TT window, we discard the closing-edge overlapping interval with the adjacent TT window ( $L_{j,j+1}^h$ ), which will be encountered with  $W_{TT_{j+1}}^h$ , i.e.,

$$\begin{aligned} \Delta t^0 &= \Delta t_{TT}^0 + \Delta t_{Gnp}^0 \\ &= \sum_{\Delta t} (W_{TT_j}^h - L_{j,j+1}^h) + \sum_{\Delta t} G_j^{h,np}. \end{aligned} \quad (19)$$

The credit variation for  $Q_X^h$  during  $\Delta t$  can be given as,

$$\begin{aligned} \Delta cr_X^h(\Delta t) &= cr_X^h(t) - cr_X^h(s) = cr_X^h(t) \\ &= \Delta t_X^+ .idSL_X + \Delta t_X^- .sdSL_X. \end{aligned} \quad (20)$$

By considering  $\Delta t_X^+ = \Delta t - \Delta t^0 - \Delta t_X^-$  in the above equation, we can obtain

$$\Delta t_X^- = \frac{(\Delta t - \Delta t^0) .idSL_X - cr_X^h(t)}{idSL_X - sdSL_X}. \quad (21)$$

Under the worst-case, the egressed TT frames during  $\Delta t$  can be expressed as

$$\begin{aligned} R_{TT}^{h*}(t) - R_{TT}^{h*}(s) &= R_{TT}^{h*}(t) - R_{TT}^h(s) = C^h . \Delta t_{TT}^0 \\ &= C^h . \sum_{\Delta t} (W_{TT_j}^h - L_{j,j+1}^h), \end{aligned} \quad (22)$$

and the protection intervals during will result in wasted service, as follows

$$\begin{aligned} R_{Gnp}^{h*}(t) - R_{Gnp}^{h*}(s) &= R_{Gnp}^{h*}(t) - R_{Gnp}^h(s) \\ &= C^h . \Delta t_{Gnp}^0 = C^h . \sum_{\Delta t} G_j^{h,np}. \end{aligned} \quad (23)$$

As the service during  $\Delta t_{TT}^0$  and  $\Delta t_{Gnp}^0$  intervals after any time  $t$  is less than or equal the number of ingress frames during these intervals, i.e.,

$$R_{TT}^{h*}(t) + R_{Gnp}^{h*}(t) \leq R_{TT}^h(t) + R_{Gnp}^h(t). \quad (24)$$

Thus, the sum of frozen-credit intervals during  $\Delta t$  can be bounded as

$$\begin{aligned} \Delta t^0 &\leq \frac{1}{C^h} \left\{ \left( R_{TT}^h(t) + R_{Gnp}^h(t) - \left( R_{TT}^h(s) + R_{Gnp}^h(s) \right) \right) \right\} \\ &\leq \frac{1}{C^h} \alpha_{Gnp+TT}^{h,i}(t), \end{aligned} \quad (25)$$

where  $\alpha_{Gnp+TT}^{h,i}(t)$  is the aggregate arrival curve by the TT windows and the related non-preemption guard bands when the  $i$ -th TT window is taken as a reference, as expressed in (10). After using (21) and (25), the departure process of AVB frames during  $\Delta t$  can be bounded by

$$\begin{aligned} R_X^{h*}(t) - R_X^{h*}(z) &= C^h . \Delta t_X^- \\ &\geq C^h . \left\{ \frac{(\Delta t - \alpha_{Gnp+TT}^{h,i}(t) / C^h) .idSL_X - cr_X^{max}}{idSL_X - sdSL_X} \right\} \\ &\geq (\Delta t - \alpha_{Gnp+TT}^{h,i}(t) / C^h) .idSL_X - cr_X^{max}. \end{aligned} \quad (26)$$

Due to  $R_X^{h*}(t)$  is a wide-sense increasing function, it can be written as

$$\begin{aligned} R_X^{h*}(t) - R_X^{h*}(s) &\geq \sup_{0 \leq u \leq \Delta t} \left\{ \left( u - \frac{\alpha_{Gnp+TT}^{h,i}(u)}{C^h} \right) .idSL_X - cr_X^{max} \right\} \\ &\geq idSL_X \left[ \sup_{0 \leq u \leq \Delta t} \left\{ u - \frac{\alpha_{Gnp+TT}^{h,i}(u)}{C^h} - \frac{cr_X^{max}}{idSL_X} \right\} \right]^+, \end{aligned} \quad (27)$$

where  $sup$  means supremum (lowest upper bound). As the term  $R_X^{h*}(t) - R_X^{h*}(s)$  represents the service curve for AVB-X traffic during  $\Delta t$ , the lowest service curve for AVB-X traffic can be given under non-preemption mode as

$$\beta_X^{h,np,i}(\Delta t) = idSL_X \left[ \sup_{0 \leq u \leq \Delta t} \left\{ u - \frac{\alpha_{Gnp+TT}^{h,i}(u)}{C^h} - \frac{cr_X^{max}}{idSL_X} \right\} \right]^+. \quad (28)$$

Then,

$$\begin{aligned} R_X^{h*}(t) &\geq R_X^{h*}(s) + \beta_X^{h,np}(\Delta t) \\ &\geq \inf_{0 \leq u \leq t} \left\{ R_X^h(s) + \beta_X^{h,np}(t-s) \right\} = (R_X^h \otimes \beta_X^{h,np})(t), \end{aligned} \quad (29)$$

where  $inf$  means infimum (largest lower bound), and the notation  $\otimes$  represents the convolution of the min-plus operation.

## 2) WITH PREEMPTION MODE

Similar to the non-preemption mode,  $\Delta t$  can be represented as a summation of three different intervals depending on the AVB credit situation: frozen, increasing, decreasing credit intervals. The frozen credit intervals can be given as

$$\begin{aligned}\Delta t^0 &= \Delta t_{TT}^0 + \Delta t_{Gp}^0 + \Delta t_{GAP}^0 \\ &= \sum_{\Delta t} (W_j^h - L_{j,j+1}^h) + \sum_{\Delta t} G_j^{h,p} + \sum_{\Delta t} GAP_{j,j+1}^h, \quad (30)\end{aligned}$$

where  $\Delta t_{Gp}^0$  represents the protection intervals in the preemption case, as limited in (13), and  $\Delta t_{GAP}^0$  represents possible unused gabs between TT windows, as bounded in (16). The AVB-X credit decreases during  $\Delta t_X^- = \Delta t_{X,OH}^- + \Delta t_{X,send}^-$ , where  $\Delta t_{X,OH}^-$  represents the overhead intervals as defined in (15) and  $\Delta t_{X,send}^-$  represents the AVB-X transmission intervals, which can be given as

$$\begin{aligned}\Delta t_{X,send}^- &= \frac{(\Delta t - \Delta t_X^0) idSL_X - \Delta t_{X,OH}^- (idSL_X - sdSL_X) - cr_X^h(t)}{idSL_X - sdSL_X}. \quad (31)\end{aligned}$$

By assuming worst transmission cases as in the non-preemption case, we can obtain the following

$$\begin{aligned}R_G^{h*}(t) - R_G^{h*}(s) &= R_G^{h*}(t) - R_G^h(s) \\ &= C^h \cdot \Delta t_G^{0,p} = C^h \cdot \sum_{\Delta t} G_j^{h,p}, \quad (32)\end{aligned}$$

$$\begin{aligned}R_{X,OH}^{h*}(t) - R_{X,OH}^{h*}(s) &= R_{X,OH}^{h*}(t) - R_{X,OH}^h(s) \\ &= C^h \cdot \Delta t_{X,OH}^- = C^h \cdot \sum_{\Delta t} OH_j^h, \quad (33)\end{aligned}$$

$$\begin{aligned}R_{GAP}^{h*}(t) - R_{GAP}^{h*}(s) &= R_{GAP}^{h*}(t) - R_{GAP}^h(s) \\ &= C^h \cdot \Delta t_{GAP}^0 = C^h \cdot \sum_{\Delta t} GAP_{j,j+1}^h. \quad (34)\end{aligned}$$

After considering the  $i$ -th TT window as a reference, the frozen-credit and overhead intervals can be bounded by

$$\begin{aligned}\Delta t^0 &= \Delta t_{TT}^0 + \Delta t_{Gp}^0 + \Delta t_{GAP}^0 \\ &\leq \frac{1}{C^h} \alpha_{Gp+TT+GAP}^{h,i}(\Delta t), \quad (35)\end{aligned}$$

$$\Delta t_{X,OH}^- \leq \frac{1}{C^h} \alpha_{OH}^{h,i}(\Delta t). \quad (36)$$

Thus, by considering (35), (36), and (31), the egressed frames from node  $h$  can be expressed as

$$\begin{aligned}R_X^{h*}(t) - R_X^{h*}(s) &= C^h \cdot \Delta t_{X,send}^- \\ &\geq \left( \Delta t - \alpha_{Gp+TT+GAP}^{h,i}(t) / C^h \right) \cdot idSL_X - \alpha_{OH}^{h,i}(\Delta t) - cr_X^{max} \\ &\geq idSL_X \cdot \left\{ \Delta t - \frac{\alpha_{Gp+TT+GAP}^{h,i}(t)}{C^h} - \frac{\alpha_{OH}^{h,i}(\Delta t)}{idSL_X} - \frac{cr_X^{max}}{idSL_X} \right\}. \quad (37)\end{aligned}$$

Finally, using the similar way as in the non-preemption case, the service curve for AVB-X traffic can be formulated during  $\Delta t$  with the preemption mode, by considering the  $i$ -th TT window as a reference, to be as

$$\beta_X^{h,p,i}(\Delta t) = idSL_X \left[ \substack{\text{sub} \\ 0 \leq u \leq \Delta t} \left\{ u - \frac{\alpha_{Gp+TT+GAP}^{h,i}(u)}{C^h} \right\}^+ \right]. \quad (38)$$

It can be noticed that the service curve can be changed according to the reference TT window. Thus, to ensure worst-case evaluations for both preemption modes, the lowest service curve must be simultaneously obtained with considering all possible service curves during the hyper-period using the following

$$\beta_{X,l}^{h,p(np)}(t) = \min_{1 \leq i \leq NW_{GCL}^h} \left\{ \beta_X^{h,p(np),i}(t) \right\}, \quad (39)$$

where  $\beta_X^{h,np,i}(t)$  and  $\beta_X^{h,p,i}(t)$  are obtained using (28) and (38), respectively.

B. DETERMINATION OF UPPER-BOUND AVB-X ARRIVAL CURVE ( $\alpha_X^{h,p(np)}(t)$ )

At the first node (source), the arrival curve for the  $m$ -th AVB-X flow ( $\tau_{X_m}$ ) can be given as

$$\alpha_{X_m}^{FN}(t) = \sigma_{X_m}^{FN} + \rho_{X_m}^{FN} \cdot t, \quad (40)$$

where  $\sigma_{X_m}^{FN}$  is the burst of  $\tau_{X_m}$  ( $\sigma_{X_m}^{FN} = f_{X_m}$ , where  $f_{X_m}$  represents the framing size of  $\tau_{X_m}$ ), and  $\rho_{X_m}^{FN}$  is the long-term rate of  $\tau_{X_m}$  ( $\rho_{X_m}^{FN} = f_{X_m} / T_{X_m}$ , where  $T_{X_m}$  represents the period of  $\tau_{X_m}$ ). If there are many AVB-X flows compete on transmission at the source, the aggregate arrival curve at the associated egress port will be formed as

$$\alpha_X^{FN}(t) = \sum_{\tau_{X_m} \in FN} \sigma_{X_m}^{FN} + \sum_{\tau_{X_m} \in FN} \rho_{X_m}^{FN} \cdot t. \quad (41)$$

For the non-first node, three shapers limit the upper-bound arrival curve for AVB-X traffic from node  $h-1$  to node  $h$ . The first shaper represents the sum of individual AVB-X arrival curves that egressed from  $h-1$  and sent to  $h$  [9], [10], i.e.,

$$\alpha_{X,sum}^{[h-1,h],p(np)}(t) = \sum_{\tau_{X_m} \in [h-1,h]} \alpha_{X_m}^{h-1,p(np)}(t). \quad (42)$$

As the above arrival shaper depends on the previous node servicing, it is defined based on the preemption ( $p$ ) and non-preemption ( $np$ ) modes. The second shaper represents the constraint of egressing bit rate from  $h-1$ . This means that the AVB-X group in  $h-1$  will arrive in  $h$  by a speed  $C^{h-1}$ . Thus, the arrival curve limited by the link speed can be given as [30], [11]

$$\begin{aligned}\alpha_{X,link}^{[h-1,h]}(t) &= \sigma_{link}(t) + f_X^{[h-1,h],max} \\ &= C^{h-1} \cdot t + f_X^{[h-1,h],max}, \quad (43)\end{aligned}$$

where  $f_X^{[h-1,h],max}$  is the largest AVB-X frame transmitted from  $h - 1$  to  $h$ . The third shaper represents the CBS limitation, as discussed in [11]. The arrival curve under this constraint is formed as [11]

$$\alpha_{X,CBS}^{[h-1,h],p(np)}(t) = \sigma_X^{h-1,p(np)}(t) + f_X^{[h-1,h],max} \\ = idSl_X \cdot \left[ t - \frac{\beta_{TT}^{h-1}(t)}{C^{h-1}} \right]^+ + cr_X^{max} - cr_X^{min}. \quad (44)$$

Then, the aggregate arrival curve for AVB-X traffic that transfer from  $h - 1$  to  $h$  will be bounded as

$$\alpha_X^{[h-1,h],p(np)}(t) = \min \left\{ \begin{array}{l} \alpha_{X,sum}^{[h-1,h],p(np)}(t), \\ \alpha_{X,link}^{[h-1,h]}(t), \\ \alpha_{X,CBS}^{[h-1,h],p(np)}(t) \end{array} \right\}. \quad (45)$$

Subsequently, the upper-bound arrival curve for  $\tau_X$  in each non-first node  $h$  can be found using the input arrival curve at the previous node  $h - 1$ , as follows

$$\alpha_X^{h,p(np)}(t) = \alpha_X^{[h-1,h],p(np)}\left(t + WCD_X^{h-1,p(np)}\right), \quad (46)$$

where  $WCD_X^{h-1,p(np)}$  represents the worst-case latency that  $\tau_X$  experiences to transfer from the  $X$ -th queue in  $h - 1$  ( $Q_X^{h-1}$ ) to the  $X$ -th queue in  $h$  ( $Q_X^h$ ), i.e.,

$$WCD_X^{h-1,p(np)} \\ = \left( D_{X,queue}^{h-1,p(np)} + D_{X,select}^{h-1} + D_{prop}^{h-1,h} + D_{proc}^h \right), \quad (47)$$

where  $D_{X,queue}^{h-1,p(np)}$  is the queuing delay in node  $h - 1$ ,  $D_{X,select}^{h-1} = f_X^{h-1,max} / C^{h-1}$  represents the selection delay from node  $h - 1$  for the entire AVB-X frame,  $D_{prop}^{h-1,h}$  represents the propagation delay from  $h - 1$  to  $h$ , and  $D_{proc}^h$  represents the processing delay that the frame experiences from the frame ingress time at node  $h$  until reaching the associated queue. Hence,  $D_{prop}^{h-1,h}$  and  $D_{proc}^h$  are assumed here to be constant, while  $D_{X,queue}^{h-1,p(np)}$  represents the worst-case that the  $Q_X^{h-1}$  frame experiences from its arrival at the queue until the start of selection time from the egress port, as calculated using (48). Hence, the queuing delay represents the critical part of the overall delay and mainly depends on the GCL implementation, applied preemption mechanism, related CBS parameters, and traffic intensity in the associated node.

### C. WORST-CASE END-TO-END LATENCY FOR AVB-X TRAFFIC

In the network calculus approach, the worst-case queuing delay in node  $h$  for AVB-X traffic equals the maximum horizontal distance between  $\alpha_X^{h-1,p(np)}(t)$  and  $\beta_X^{h-1,p(np)}(t)$  as

$$D_{X,queue}^{h-1,p(np)} = H \left( \alpha_X^{h-1,p(np)}(t), \beta_X^{h-1,p(np)}(t) \right). \quad (48)$$

Then, the worst-case end-to-end latency for AVB-X traffic can be calculated as a sum of the worst-case latencies experienced in the selected path, as

$$WCD_X^{total,p(np)} = \sum_{h=1}^{N-1} WCD_X^{h,p(np)}, \quad (49)$$

where  $N$  represents the number of nodes in the selected path that the AVB-X traffic passed through.

## VII. PERFORMANCE EVALUATION

### A. CASE STUDY AND EXPERIMENTAL SETUP

As the network calculus approach is used to formulate our model, the related GCLs are built using Java API of the Real-Time Calculus (RTC) toolbox [31], which is mainly implemented by min-plus and max-plus operators to represent real-time systems. Our evaluation programs are run using an Intel Core i7-3770 CPU computer at 3.40GHz and RAM 12 GB.

For performance evaluations, a realistic vehicular use case is used as one of the TSN targeted applications, as shown in Fig. 9(a). The connected vehicle can be represented by a simple networking topology, as shown in Fig. 9(b) [5], [8]. The associated end systems could be sensors, actuators, or cameras distributed in the vehicle to gather relevant driving information from the surrounding area.

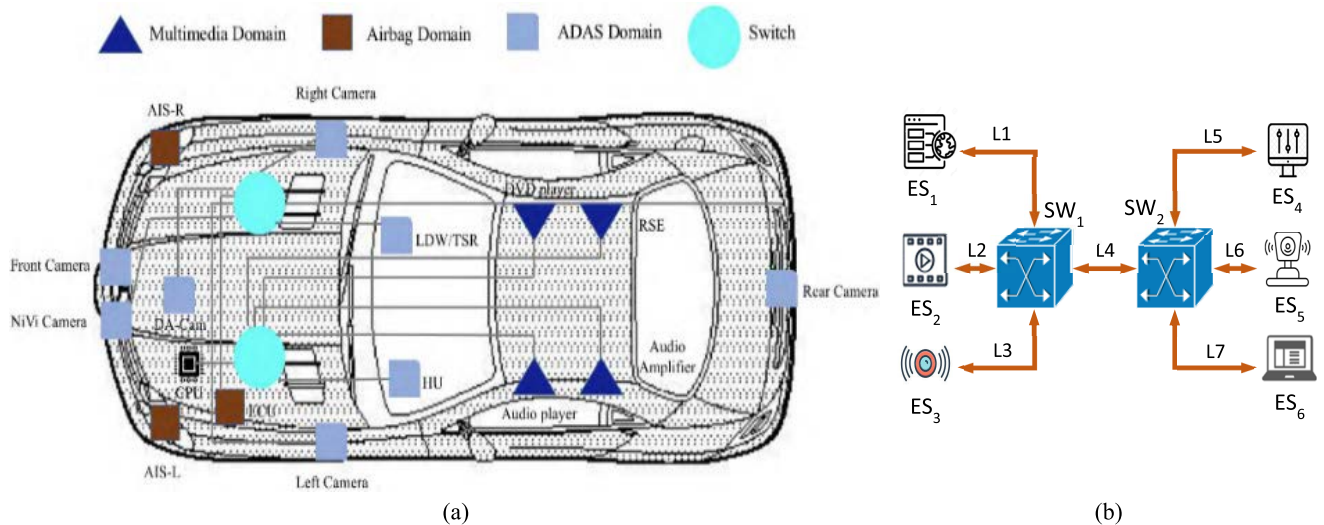
It is assumed that the data rate equals 1 Gbps for all physical links in all experiments. Further, incoming TT flows are differentiated into eight different priority queues in each node. The predefined timetable controls open and close events for each queue-gate (GCL), implemented as described in Section IV-B. The opening-edge times for all TT windows in the hyper-period in node  $h$  ( $t_k^{h,o}$ ) are tabulated in Table 2. The closing-edge times in the GCL implementations are updated using (6) and V-B2 accordingly. It is assumed that the duration of the open window is  $20\mu s$  with a  $500\mu s$  period for all TT queues in each node.

TABLE 2. GCL implementation in the first node for each path.

TT- priority class	Opening-edge time in node $h$ ( $t_k^{h,o}$ ) ( $\mu s$ )	
	Back-to-back configuration	Porosity configuration
$k = 1$	20	20
$k = 2$	$40 - W_{TT_1}^h \cdot OR_{1,2}^h$	$40 - W_{TT_1}^h \cdot OR_{1,2}^h$
$k = 3$	$60 - W_{TT_2}^h \cdot OR_{2,3}^h$	90
$k = 4$	$80 - W_{TT_3}^h \cdot OR_{3,4}^h$	$110 - W_{TT_3}^h \cdot OR_{3,4}^h$
$k = 5$	$100 - W_{TT_4}^h \cdot OR_{4,5}^h$	160
$k = 6$	$120 - W_{TT_5}^h \cdot OR_{5,6}^h$	$180 - W_{TT_5}^h \cdot OR_{5,6}^h$
$k = 7$	$140 - W_{TT_6}^h \cdot OR_{6,7}^h$	230
$k = 8$	$160 - W_{TT_7}^h \cdot OR_{7,8}^h$	$250 - W_{TT_7}^h \cdot OR_{7,8}^h$

Thus, from Table 2, we can observe that all GCL implementations for each networking path are based on variable





**FIGURE 9.** A graphical model for the applied use case: (a) a realistic vehicle use case with related networking components (b) representative model for the vehicular networking scenario.

OR between TT windows. Note that the symbol  $k$  represents both the TT priority class or the TT window order in the hyper-period as window durations and periods for all TT queues are the same. The WCD for AVB traffic is evaluated under one-sided and two-sided overlapping scenarios (porosity and back-to-back configurations) as illustrated in Table 2, using two different AVB loading cases (light and heavy loads) as described in Table 3.

**B. NUMERICAL RESULTS AND DISCUSSION**

This subsection presents our results under preemption mode in detail, and then, a brief comparison between the preemption and non-preemption modes is provided. Finally, our findings are compared with the previous related works.

**1) AVB LATENCY EVALUATIONS UNDER LIGHT LOAD**

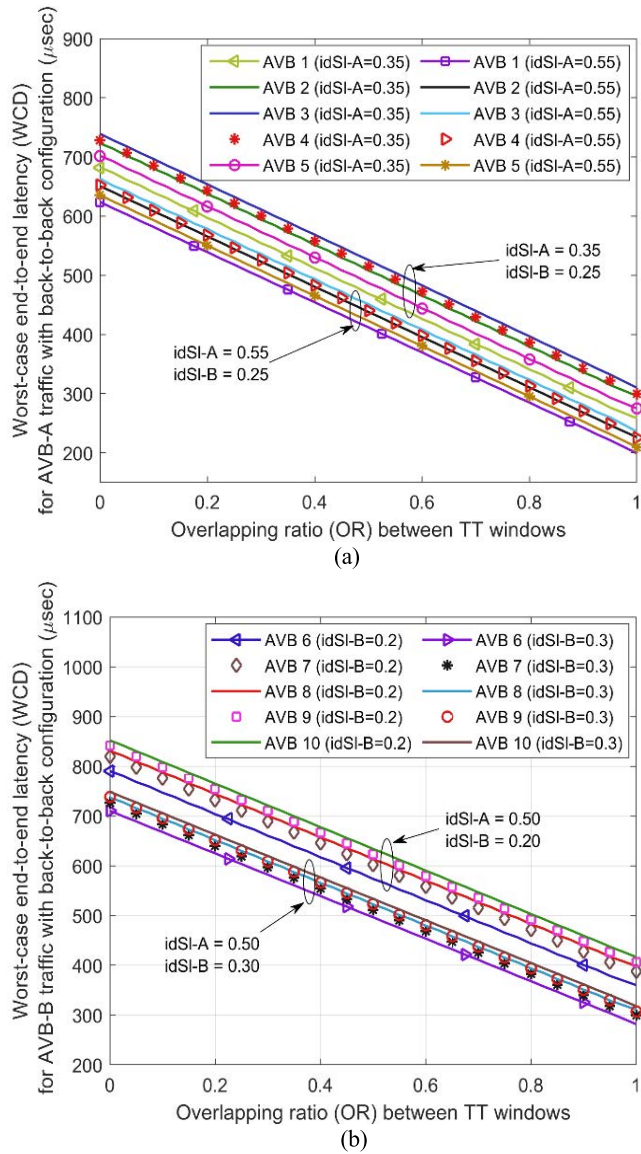
For this case, 10 AVB flows (5 AVB-A and 5 AVB-B flows) are considered to select the routes in Fig. 9(b), as specified and distributed in Table 3.

Fig. 10(a) and (b) show the effect of TT overlapping on the WCD of AVB-A and B traffic, respectively, with back-to-back configuration under multiple CBS settings. It is noticed that WCD for all AVB flows decreases dramatically as much as OR increases. In Fig. 10(a), the WCD performance for AVB-A flows is assessed under two different AVB-A idle slopes ( $idSl_A = 0.35, 0.55$ ) and fixed AVB-B idle slope ( $idSl_B = 0.25$ ). As expected, less WCDs has obtained with  $idSl_A = 0.55$  than with 0.35, as higher values of  $idSl_A$  means offering more bandwidth for AVB-A transmissions. For each CBS setting, AVB-A flows experience slightly different WCDs depending on their selected paths and their sizes. For example, AVB 1 and 5 experience lower WCDs than others as they do not have competitions with other AVB-A flows on Link 6 and Link 3, respectively. However, AVB 1 experiences a slightly lower WCD than AVB 5 as

**TABLE 3.** AVB loading scenarios on related physical links.

End-to-end path	Flow		Type	Size (Bytes)	Period ( $\mu$ s)
	Light load case	Heavy load case			
ES1-SW1-SW2-ES5	AVB 1	AVB 1,11,21	A	500	2000
ES1-SW1-SW2-ES6	AVB 2	AVB 2,12,22	A	1050	2000
ES2-SW1-SW2-ES4	AVB 3	AVB 3,13,23	A	1200	2000
ES2-SW1-SW2-ES4	AVB 4	AVB 4,14,24	A	750	2000
ES3-SW1-SW2-ES6	AVB 5	AVB 5,15,25	A	800	2000
ES1-SW1-SW2-ES6	AVB 6	AVB 6,16,26	B	800	2000
ES2-SW1-SW2-ES4	AVB 7	AVB 7,17,27	B	400	2000
ES2-SW1-SW2-ES4	AVB 8	AVB 8,18,28	B	850	2000
ES3-SW1-SW2-ES5	AVB 9	AVB 9,19,29	B	550	2000
ES3-SW1-SW2-ES5	AVB 10	AVB 10,20,30	B	1000	2000

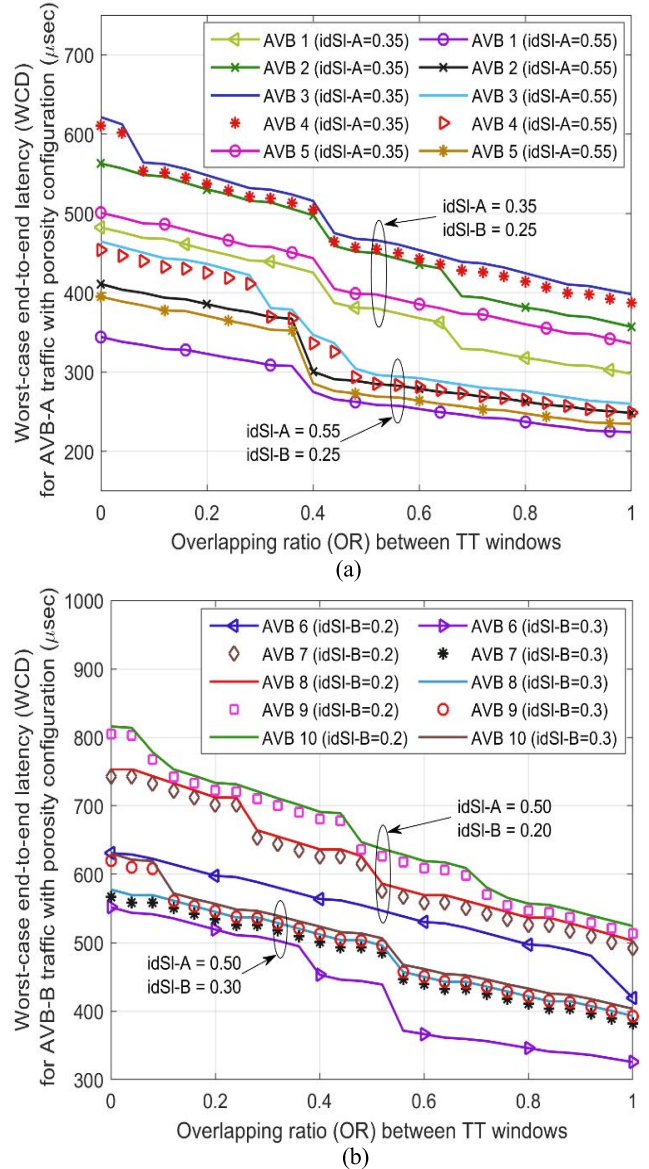
it has a smaller framing size, i.e.,  $f_{AVB1}^{h,max} = 500$  bytes and  $f_{AVB5}^{h,max} = 800$  bytes. Further, AVB 3 and AVB 4 experience the highest WCDs as they compete with each other on the same path (ES2-SW1-SW2-ES4). As AVB 3 has a bigger size (1200 bytes) than AVB 4 (750 bytes), it experiences higher selection delays in the associated nodes, leading to higher end-to-end latency. In Fig. 10(b), the WCD behavior for AVB-B flows is examined under two different AVB-B idle slopes ( $idSl_B = 0.20, 0.30$ ) and fixed AVB-A idle slope ( $idSl_A = 0.50$ ). As observed from the figure, lower latencies have been obtained using  $idSl_B = 0.30$  than 0.20. Moreover, frames with different sizes experience unequal WCDs even if they share the same path. For example, AVB 7 and 8 share the path of ES2-SW1-SW2-ES4, but their WCDs are slightly



**FIGURE 10.** The relationship between WCDs for AVB traffic and OR between TT windows with back-to-back configuration and only 10 AVB flows compete on the links; (a) AVB-A latencies under different values of AVB-A idle slope ( $idSI_A = 0.35, 0.55$ ) and fixed AVB-B idle slope ( $idSI_B = 0.25$ ). (b) AVB-B latencies under different values of AVB-B idle slope ( $idSI_B = 0.20, 0.30$ ) and fixed AVB-A idle slope ( $idSI_A = 0.50$ ).

different. The reason is that their sizes are different (400 and 850 bytes), resulting in various selection delays, as formulated in VI-C. Further, AVB 6 has the lowest WCD as there is no other AVB-B flows compete on Link 1 and Link 7. The highest WCD has been experienced by AVB 9 and 10 as they compete on the same path and their sizes together are 1550 bytes, which is larger than all AVB-B loads on other links. AVB 10 has a larger size (1000 bytes) than AVB 9 (550 bytes), it experiences higher end-to-end latency.

With porosity configuration, the WCD performance with respect to OR is evaluated for AVB-A and AVB-B flows with multiple CBS settings in Fig. 11(a) and (b), respectively.



**FIGURE 11.** The relationship between WCDs for AVB traffic and OR between TT windows with porosity configuration and only 10 AVB flows compete on the links; (a) AVB-A latencies under different values of AVB-A idle slope ( $idSI_A = 0.35, 0.55$ ) and fixed AVB-B idle slope ( $idSI_B = 0.25$ ). (b) AVB-B latencies under different values of AVB-B idle slope ( $idSI_B = 0.20, 0.30$ ) and fixed AVB-A idle slope ( $idSI_A = 0.50$ ).

Compared to the back-to-back configuration, WCD for all AVB flows with the porosity style decreases by lower percentages with the increase in OR. Furthermore, the WCD behavior differs from one flow to another as the AVB transmissions are interrupted by multiple TT windows, differing from that in the back-to-back style where the AVB transmissions are allocated at one blank interval in the hyper-period. It can be noticed that the WCD performance with the porosity configuration for all AVB flows is better than that with back-to-back during lower OR percentages under light loading case (10 AVB flows). However, the back-to-back design gives lower latencies than the porosity style under a higher degree



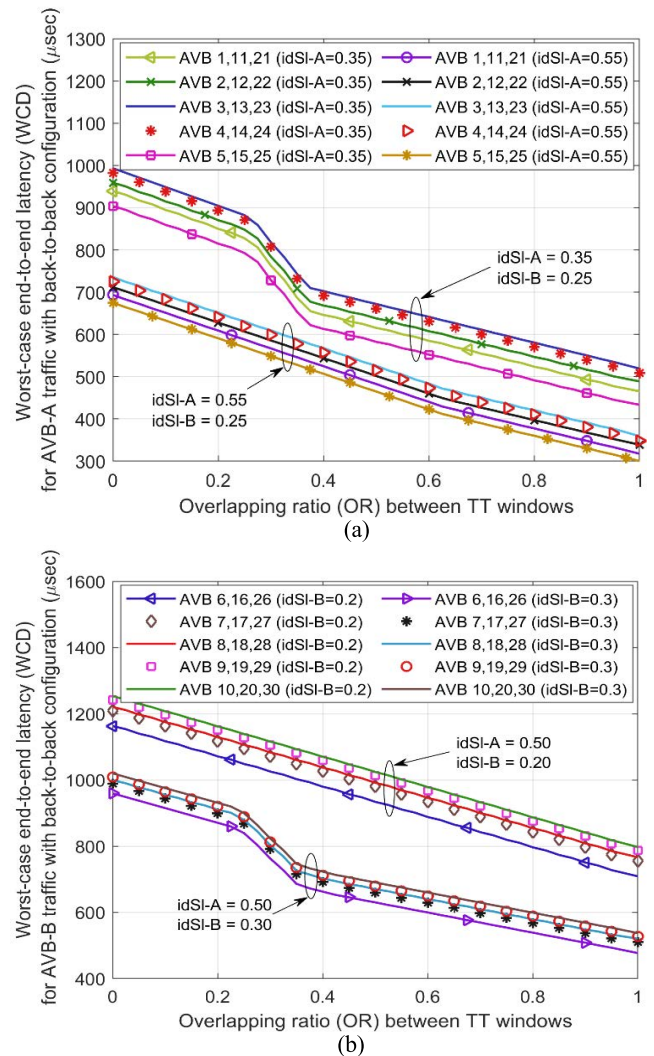
of overlapping, as shown in Fig. 10 and 11. As a result, the TT overlapping with back-to-back configuration gives more *WCD* reduction than that with porosity. For example, under 10% and 30% of TT overlapping with the back-to-back, the *WCD* for AVB-A is reduced by 6.57% and 19.72% on average, and by 5.59% and 16.77% on average for AVB-B traffic. However, with porosity style, the AVB-A *WCD* is reduced by 2.50% and 11.43% on average, and by 2.40% and 12.52% for AVB-B traffic under 10% and 30%, respectively. Compared to the AVB-A flows, AVB 1 experiences the lowest *WCD* and AVB 3 has the highest. Moreover, AVB 6 still obtains the lowest *WCD* and AVB 10 is the highest related to the AVB-B flows for the same reasons as justified in the back-to-back case.

## 2) AVB LATENCY EVALUATIONS UNDER HEAVY LOAD

For this case, 30 AVB flows (15 AVB-A and 15 AVB-B) are assumed to share the links in Fig. 9(b), as specified in Table 3. Fig. 12(a) and (b) show the impact of *OR* on *WCD* with the back-to-back style for AVB-A and AVB-B flows, respectively, and Fig. 13(a) and (b) with the porosity configuration. For these evaluations, we use the same idle slopes for AVB-A and AVB-B as those in the light loading case (10 AVB flows).

Fig. 12(a) and (b) prove that the effect of the heavy load situation on the latency performance is higher if compared with the light load case in Fig. 10(a) and (b). However, the *WCD* behavior is still almost the same as that under light load conditions, i.e., the *OR* increase reduces *WCD*s for all AVB flows by the same percentages with the associated idle slopes. Moreover, there are some AVB flows experience the same *WCD* as they share the same end-to-end networking path and have equal sizes of frames, such as AVB 1, 11, and 21. A slight change in *WCD*s for some AVB-A flows compared with each other. For example, AVB 1 experiences lower *WCD*s than AVB 5 in the light load case, as shown in Fig. 10(a), but under heavy loading scenarios, AVB 5 experiences the lowest, as shown in Fig. 12(a). This change is expected as these flows cross different paths with different loads and have unequal framing sizes.

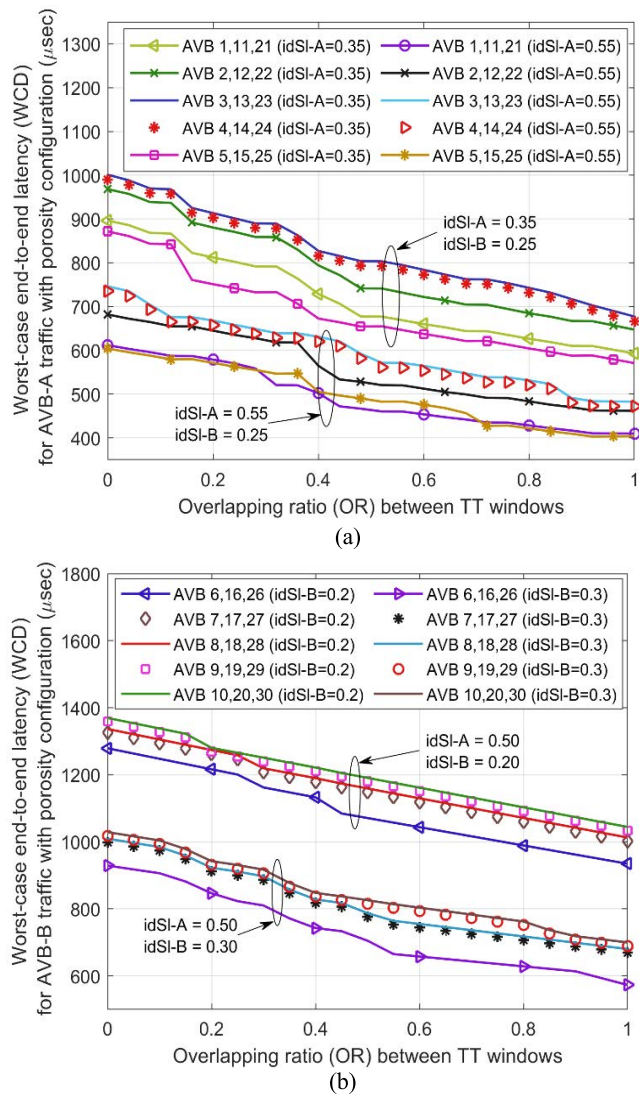
Fig. 13(a) and (b) show that as observed under the light load case, the *WCD* reduction with the *OR* increase in porosity style is lower than back-to-back. For example, under 10% *OR* and 30% *OR* with back-to-back, the *WCD* for AVB-A is reduced by 5.93% and 17.79% on average, and by 4.14% and 12.40% for AVB-B traffic. However, with porosity style, the AVB-A *WCD* is reduced by 5.15% and 11.48% on average, and by 3.70% and 11.54% for AVB-B traffic under 10% *OR* and 30% *OR*, respectively. Furthermore, after comparing the results in Fig. 13(a) and (b) with those in Fig. 12(a) and (b), it can be noticed that the *WCD* performance with the back-to-back is better than porosity in most overlapping situations. This happens because there is larger waste bandwidth in the porosity style, as more guard bands and overhead intervals are required. The effect of these unoccupied intervals appears



**FIGURE 12.** The relationship between *WCD*s for AVB traffic and *OR* between TT windows with back-to-back configuration and 30 AVB flows compete on the links; (a) AVB-A latencies under different values of AVB-A idle slope ( $idSI_A = 0.35, 0.55$ ) and fixed AVB-B idle slope ( $idSI_B = 0.25$ ). (b) AVB-B latencies under different values of AVB-B idle slope ( $idSI_B = 0.20, 0.30$ ) and fixed AVB-A idle slope ( $idSI_A = 0.50$ ).

in heavy load situations because most blank intervals will be used for AVB transmissions.

A slight fluctuation can happen between any *WCD* lines close to each other, depending on the related frame sizes,  $idSI_A$ ,  $idSI_B$ , and selected path. For example, the *WCD* behavior for AVB 1,11,21 and AVB 5,15,25 frames fluctuate with *OR*. Although the size of AVB 1,11,21 frames (500 bytes) is smaller than AVB 5,15,25 (800 bytes) and both are class A, they experience a slightly higher *WCD* when  $30\% \leq OR \leq 70\%$ , as shown in Fig. 13(a). This fluctuation happens because the frames share different paths, as listed in Table 3. AVB 1,11,21 frames share *ES1-SWI-SW2-ES5* and AVB 5,15,25 share *ES3-SWI-SW2-ES6*. As formulated in VII), total *WCD* is calculated by aggregating all *WCD*s experienced through the selected links, and the upper-bound arrivals in each node depend on the *WCD* experienced in the



**FIGURE 13.** The relationship between WCDs for AVB traffic and OR between TT windows with porosity configuration and 30 AVB flows compete on the links; (a) AVB-A latencies under different values of AVB-A idle slope ( $idSI_A = 0.35, 0.55$ ) and fixed AVB-B idle slope ( $idSI_B = 0.25$ ). (b) AVB-B latencies under different values of AVB-B idle slope ( $idSI_B = 0.20, 0.30$ ) and fixed AVB-A idle slope ( $idSI_A = 0.50$ ).

previous node. Varying *OR* with the same value in all nodes will affect all experienced *WCDs* through the selected links by different percentages, leading to fluctuated *WCDs*.

### 3) COMPARISON BETWEEN PREEMPTION AND NON-PREEMPTION MODES WITH AVB LATENCY

In this subsection, a critical comparison between preemption and non-preemption modes is presented concerning *WCD* for AVB traffic. For a meaningful comparison, the presented results for preemption mode in the previous subsection are summarized with the non-preemption for the back-to-back and porosity configurations in Fig. 14(a) and (b), respectively. In both figures, the non-preemption results are presented by solid lines and the preemption by markers, keeping the same color for both modes under the same configuration

and loading condition. With assuming larger bandwidth for AVB-A transmissions than AVB-B in related idle slopes, the average *WCD* for AVB-A traffic is always lower than AVB-B traffic.

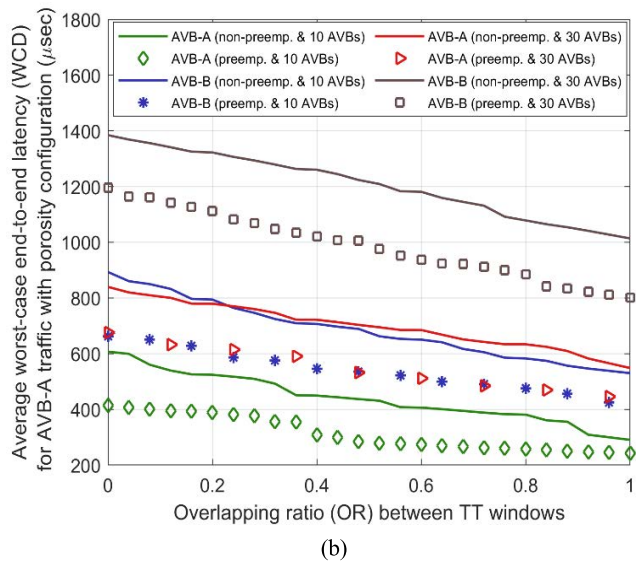
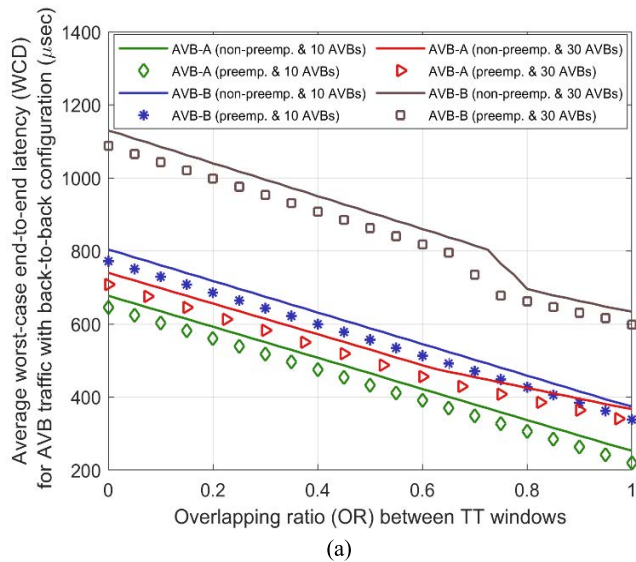
As expected, the preemption mode gives less *WCDs* on average compared to the non-preemption for all configurations, loading conditions, and overlapping ratios (*ORs*). Moreover, the porosity style results in lower *WCDs* on average than back-to-back, as shown in Fig. 14(a) and (b). For example, with 20% *OR* and using back-to-back configuration, the preemption mode minimizes the *WCD* for AVB-A traffic by 5.48% and 4.91% on average, and by 4.35% and 4.04% for AVB-B traffic under light and heavy loading cases, respectively, compared to the non-preemption. However, using the same *OR* with the porosity style, the preemption mode is superior to non-preemption by 25.91% and 19.94% *WCD* reduction on average for AVB-A traffic, and 23.76% and 15.89% for AVB-B traffic under light and heavy loading cases, respectively. The observed advantage in *WCD* performance using preemption mode returns to saving more bandwidth with smaller sizes of guard bands compared to non-preemption. As described in Section III-B, the preemption guard band is only 123 bytes, but the non-preemption requires 1500 bytes to protect TT windows from unscheduled transmissions. As the proposed model adopts the preemption with the Hold and Release feature, there is no negative impact will happen on the TT latency performance.

### 4) AVB LATENCY COMPARISON WITH THE PREVIOUS WORKS

This subsection compares our results with the previous related works in [9] (AVB-TSN18) and [11] (AVB-TSN21). The comparison is provided using preemption mode for both configurations (back-to-back and porosity) under assumptions of light and heavy load (10 AVB and 30 AVB flows). Fig. 15(a) and (b) present *WCD* bounds that experienced individually by all AVB flows under light loading condition with back-to-back and porosity configurations, respectively, and Fig. 16(a) and (b) under heavy loading case.

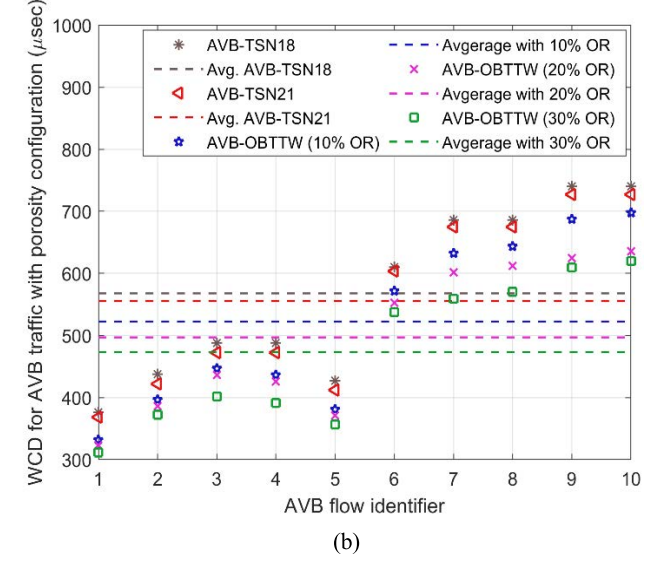
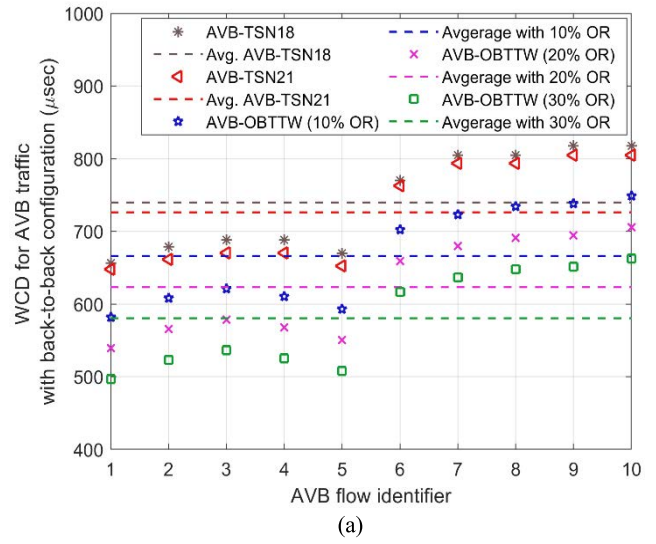
The proposed model in [11] improved the associated analysis by implementing more tight arrival curves for AVB traffic than that presented in [9]. In [9], the authors formulated the AVB arrival curve based on the aggregate arrivals from all individual nodes connected to the associated node's ingress port. However, Zhao et al. [11] added two other arrival shapers that control the AVB arrivals, i.e., link speed and CBS shapers. In AVB-OBTTW, all three arrival shapers are considered, as presented in Section V-B. We select three overlapping ratios to compare with the previous works, i.e., 10%, 20%, and 30%. It is noticed that AVB-OBTTW reduces *WCD* for all AVB flows compared to AVB-TSN18 and AVB-TSN21. The *WCD* reduction with back-to-back style is rather than that with porosity. For light load case as shown in Fig. 15(a) and (b), AVB-OBTTW reduces *WCD* with back-to-back by 8.28%, 14.17%, and 20.07% compared to AVB-TSN21, and by 5.94%, 10.55%, and 14.85% with porosity





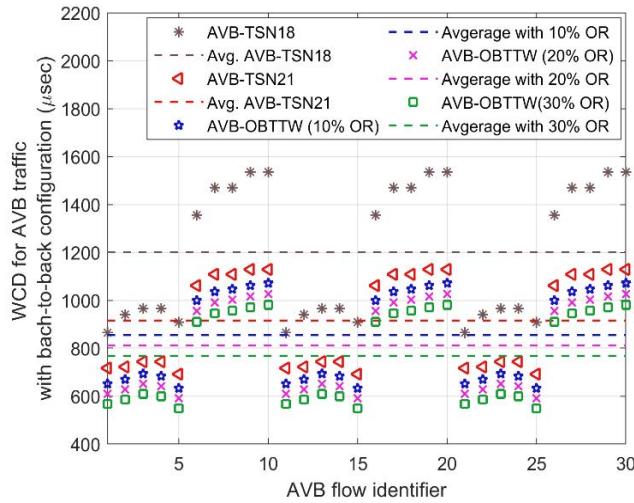
**FIGURE 14.** Comparison between preemption and non-preemption modes concerning average WCDs for AVB traffic by assuming 10 and 30 AVB flows compete on the links, separately with; (a) back-to-back configuration. (b) porosity configuration.

under 10% *OR*, 20% *OR*, and 30% *OR*, respectively. Furthermore, by considering the selection delay in our *WCD* calculations, as presented in VI-C, it is worthy to know that AVB-OBTTW obtains more accurate results than other models, which gave equally *WCD*s for AVB flows that share the same path. For example, the AVB 3 and 4 experience the same *WCD* in AVB-TSN18 and AVB-TSN21 models, while they have different sizes of frames. In contrast, AVB-OBTTW obtains *WCD*s proportional to the sizes of frames, resulting in different *WCD*s for flows that share the same path and have unequal frame sizes, such as AVB 3 and 4, AVB 7 and 8, and AVB 9 and 10. For example, as shown in Fig. 15(a) and (b), AVB 3 (with the size of 1200 bytes) experiences a larger *WCD* than AVB 4 (with the size of 750 bytes).

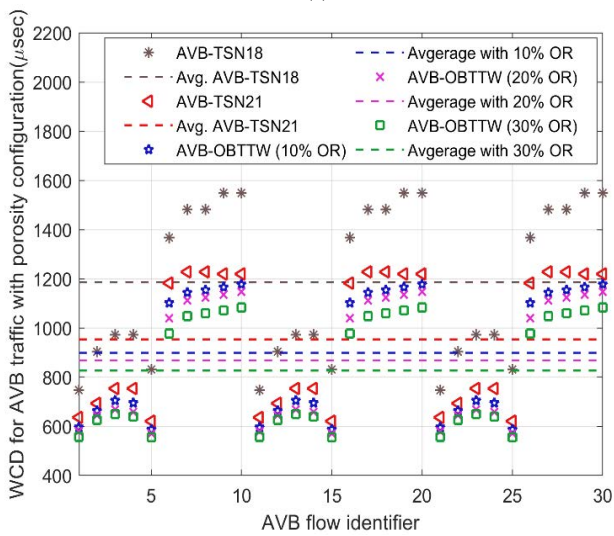


**FIGURE 15.** Comparison between *WCD* performances using preemption mode for three different models; our model (AVB-OBTTW), AVB-TSN18[9], and AVB-TSN21 [11] by assuming only 10 AVB flows compete on the links with; (a) back-to-back configuration. (b) porosity configuration.

For the heavy load case as shown in Fig. 16(a) and (b), it can be observed that using the AVB-TSN18 model produces more pessimistic *WCD*s than AVB-TSN21 as it is based on the aggregate arrivals for individual AVB flows. Accordingly, other arrival shapers, i.e., link speed and CBS shapers, have a considerable impact on *WCD* boundaries, as applied on AVB-TSN21 and AVB-OBTTW models. Further improvements have been achieved using AVB-OBTTW over AVB-TSN21 under different *OR* percentages. For example, AVB-OBTTW reduces *WCD* with back-to-back by 6.64%, 11.39%, and 16.15% compared to AVB-TSN21, and by 5.73%, 8.97%, and 13.28% with porosity under 10% *OR*, 20% *OR*, and 30% *OR*, respectively. Thus, *WCD* boundaries for AVB flows can be reduced by a percentage depending on the maximum allowable *OR* between TT windows, which can be calculated



(a)



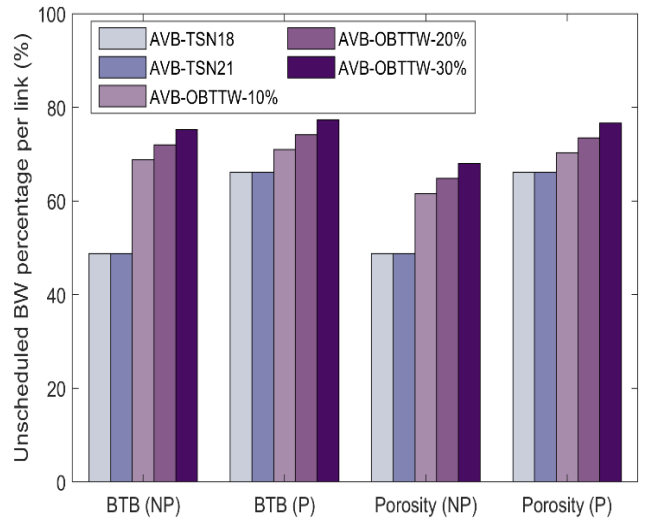
(b)

**FIGURE 16.** Comparison between WCD performances using preemption mode for three different models; our model (AVB-OBTWW), AVB-TSN18[9], and AVB-TSN21 [11] by assuming 30 AVB flows compete on the links with; (a) back-to-back configuration. (b) porosity configuration.

using the OFWOS model in [8]. The maximum allowable *OR* is the highest overlapping ratio between TT windows that guarantees end-to-end latency deadlines for the associated TT queue.

5) BW COMPARISON WITH THE PREVIOUS WORKS

Allowing TT windows to overlap contributes to minimizing the bandwidth waste in each link. When the overlapping occurs, there is no need to assign a guard band between the overlapped windows against unscheduled transmissions. Accordingly, the bandwidth waste will be reduced, especially with the back-to-back configuration. The unallocated guard bands and the overlapping intervals lead to saving more BW for unscheduled traffic. The highest unscheduled BW that can be saved is obtained with the maximum allowable *OR*.



**FIGURE 17.** Comparison between unscheduled BW percentages using AVB-OBTWW, AVB-TSN18 [9], and AVB-TSN21 [11] with both preemption modes, under back-to-back and porosity configurations.

Fig. 17 shows the available unscheduled BW using AVB-TSN18, AVB-TSN21, and AVB-OBTWW models under 10%, 20%, and 30% *OR*s. The comparison is achieved with the preemption and non-preemption modes under back-to-back and porosity styles. As AVB-TSN18 and AVB-TSN21 did not consider the TT overlapping, just AVB-OBTWW is assessed under multiple *OR*s. Further, the AVB-TSN18 and AVB-TSN21 give the same BW assignment with no difference between their GCL implementations. For the same configuration, the preemption mode offers a higher unscheduled BW than non-preemption in all models, as it requires a smaller size guard band. It can be noticed that AVB-OBTWW saves more unscheduled BW than others for all cases. The BW improvement increases when *OR* increases. For example, under porosity and preemption mode, 10%, 20%, and 30% *OR*s in AVB-OBTWW offer unscheduled BW 70.26%, 73.46%, and 76.66%, respectively, where only 66.12% is available using AVB-TSN18 and AVB-TSN21. At the same *OR*, the highest BW can be obtained is under back-to-back with preemption. Compared to the previous models, the highest BW enhancement is obtained with the non-preemption mode. For example, the non-preemption mode with 10% *OR* increases the unscheduled BW by 20% and 12.8% under back-to-back and porosity modes, respectively. But for preemption mode with the same *OR*, the BW percentage is enhanced by 4.84% and 4.14% under back-to-back and porosity, respectively.

VIII. CONCLUSION

The TSN framework is introduced as an appropriate environment for mixed-criticality applications, such as automotive and automated industries. In TSN, a time-aware shaping mechanism schedules TT streams through a window-based transmission system. Complete isolation between TT windows minimizes the available bandwidth for unscheduled traffic (AVB and BE). Although several scheduling

algorithms have been implemented based on overlapped TT windows, none evaluated the worst-case AVB performance under these overlaps.

In this article, a worst-case AVB latency under overlapping-based TT windows (AVB-OBTTW) algorithm is proposed. Complete WCD forms are presented separately with preemption and non-preemption modes. These models are evaluated using a realistic vehicular scenario by assuming back-to-back and porosity configurations in the hyper-period under light and heavy loading conditions. The numerical results confirm WCD reduction with increasing OR for all experimental settings. As higher bandwidth is given for AVB-A traffic than AVB-B, AVB-A flows experience less WCDs than AVB-B traffic that share the same path. The overlapping effect in the back-to-back style is higher than that in porosity. On average, the porosity style gives less WCDs under a light loading case, but as much as the load increases, the back-to-back becomes better than the porosity. Furthermore, the preemption mode gives less WCDs than non-preemption under all experimental settings, especially using the porosity style. For example, with 20% OR, the preemption mode reduces WCD compared to non-preemption by 4.92% and 4.48% on average using back-to-back type, and by 24.84% and 17.92% using porosity under light and heavy loading scenarios, respectively. Compared to the latest related work, our model obtains less WCDs, especially under large ORs. For example, with 10% and 30% ORs, AVB-OBTTW minimizes WCD using back-to-back style by 8.28% and 20.07% on average and using porosity by 5.94% and 14.85%, respectively. Thus, combining our findings with that for TT traffic will be a helpful guide for vehicular designers to implement suitable GCL patterns that can effectively serve incoming data from interconnected devices, such as car sensors, actuators, and cameras.

Another significant TSN feature is the frame replication and elimination for reliability (FRER) to ensure zero congestion loss for time-triggered frames, as defined in IEEE 802.1CB. This function is not provided for the selected nodes in this work. The FRER feature will increase the ingress frames in each node, leading to higher WCDs for all critical time traffic. An exciting research point for future work is to apply other real-time networking topologies considering this feature in all selected nodes.

## APPENDIX A MINIMUM AND MAXIMUM CREDIT LIMITS FOR AVB QUEUES

### A. AVB-A CREDIT BOUNDING

According to the transmission situation, the AVB-X credit will increase by a slope  $idSl_X$  or decrease by a slope  $sdSl_X$  between the minimum and maximum bounds, as

$$cr_X^{min} \leq cr_X(t) \leq cr_X^{max} \quad (50)$$

AVB-A frame can be transmitted if: it is in the associated queue, its credit is greater than zero ( $cr_A(t) > 0$ ), and there

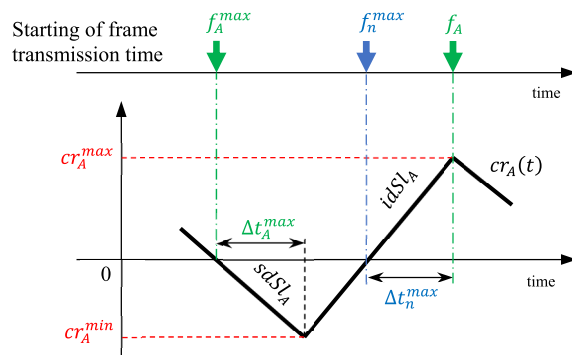


FIGURE 18. Maximum and minimum credits for the AVB-A queue.

is no lower priority frame being in transmission. As shown in Fig. 18, The minimum bound of AVB-A credit ( $cr_A^{min}$ ) can be reached after transmitting the largest size of AVB-A frame when the associated credit at the beginning of transmission is greater than zero by a minimal positive value ( $cr_A(t) = \epsilon^+$ , where  $\epsilon^+ \rightarrow 0^+$ ), as given by [32]

$$cr_A^{min} = sdSl_A \cdot \Delta t_A^{max} = sdSl_A \frac{f_A^{max}}{C}. \quad (51)$$

The maximum bound of AVB-A credit ( $cr_A^{max}$ ) is reached when the associated frame encounters the worst transmission situation, as shown in Fig. 18. This case can happen when the associated credit is less than zero by a very small value ( $cr_A(t) = \epsilon^-$ , where  $\epsilon^- \rightarrow 0^-$ ), and there is a lower priority frame (AVB-B or BE) that starts the transmission. AVB-A frame will wait  $\Delta t_n^{max}$  ( $\Delta t_n^{max} = f_n^{max}/C$ ), where  $f_n^{max} = \max(f_B^{max}, f_{BE}^{max})$ . Therefore, the AVB-A credit will be bounded as [32]

$$sdSl_A \frac{f_A^{max}}{C} \leq cr_A(t) \leq idSl_A \frac{f_n^{max}}{C}. \quad (52)$$

### B. AVB-B CREDIT BOUNDING

In similar with AVB-A case, the minimum bound of AVB-B credit can be given as [32]

$$cr_B^{min} = sdSl_B \cdot \Delta t_B^{max} = sdSl_B \frac{f_B^{max}}{C}. \quad (53)$$

As shown in Fig. 19, the maximum AVB-B credit is reached if; the associated frame arrives in the queue and its credit is less than zero; there is a BE frame that starts the transmission when  $cr_B(t) = \epsilon^-$  (where  $\epsilon^- \rightarrow 0^-$ ), and before completing BE transmission the largest AVB-A frame arrives the associated queue with the maximum credit bound. Thus, from the moment when  $cr_B(t) = \epsilon^-$ , the AVB-B frame will wait  $\Delta t_{BE}^{max} + \Delta t_A^{max}$ , where  $\Delta t_{BE}^{max} = f_{BE}^{max}/C$  and  $\Delta t_A^{max}$  is the time interval required to change the AVB-A credit from  $cr_A^{max}$  to  $cr_A^{min}$ , as

$$\Delta t_A^{max} = \frac{cr_A^{min} - cr_A^{max}}{sdSl_A} = \frac{f_A^{max}}{C} - \frac{f_n^{max}}{C} \cdot \frac{idSl_A}{sdSl_A}. \quad (54)$$



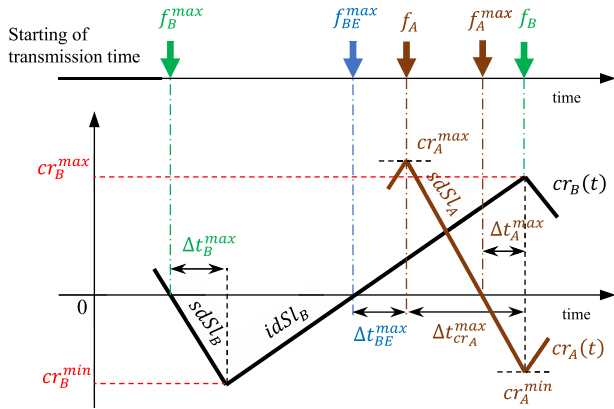


FIGURE 19. Maximum and minimum credits for the AVB-B queue.

Then, the maximum bound of AVB-B credit can be represented as [32]

$$\begin{aligned} cr_B^{max} &= idSL_B \cdot (\Delta t_{BE}^{max} + \Delta t_A^{max}) \\ &= \frac{idSL_B}{C} \left( f_{BE}^{max} + f_A^{max} - f_n^{max} \cdot \frac{idSL_A}{sdSL_A} \right). \end{aligned} \quad (55)$$

### APPENDIX B NETWORK CALCULUS BASIS FOR WORST-CASE LATENCY CALCULATION

The network calculus (NC) approach is commonly used in networking communications to formulate critical time flow transmissions and evaluate related QoS parameters, such as network utilization and worst-case latency boundaries. This article uses NC to assess worst-case latency for AVB traffic in TSN. To fulfill that, the arrival and service curves must be determined to express traffic intensity and servicing availability in each selected node. These curves are mathematically formulated using min-plus operations. The arrival curve  $\alpha(t)$  can be found by specifying the flow arrival process  $(R(t))$ , which counts the bits enter a node until  $t$ , as [33]

$$\forall \lambda < t$$

$$R(t) \leq \inf_{0 \leq \lambda \leq t} \{R(\lambda) + \alpha(t - \lambda)\} = (R \otimes \alpha)(t), \quad (56)$$

where  $\inf$  denotes infimum (highest lower bound) and  $\otimes$  is the min-plus convolution. The typical arrival curve is the leaky bucket model, which is given by [26]

$$\alpha_{\sigma, \rho}(t) = \begin{cases} \rho t + \sigma, & t \geq 0 \\ 0, & t < 0, \end{cases} \quad (57)$$

where  $\sigma$  is the largest flow burst and  $\rho$  is the maximum limit of the flow's long-term average rate. The service curve  $\beta(t)$  can be found by specifying the data departure process  $(R^*(t))$ , which counts the bits egress from a node until  $t$ , as [33]

$$\forall \lambda < t$$

$$R^*(t) \leq \inf_{0 \leq \lambda \leq t} \{R(\lambda) + \beta(t - \lambda)\} = (R \otimes \beta)(t). \quad (58)$$

The typical service curve example is the rate-latency service curve, which is given by [26]

$$\beta_{C, T}(t) = C[t - T]^+, \quad (59)$$

where  $C$  is the egress port servicing rate,  $T$  is the servicing latency, and  $[x]^+$  is equal to  $x$  when  $x \geq 0$  and 0 otherwise. As shown in Fig. 20, the worst-case latency bound can be calculated in each node as a maximum horizontal distance ( $D_{max}$ ) between  $\alpha(t)$  and  $\beta(t)$ , as [33]

$$\begin{aligned} D_{max} &\stackrel{def}{=} Del(\alpha, \beta) \\ &= \sup_{\lambda \geq 0} \{ \inf_{\tau \geq 0} \{ \alpha(\lambda) \leq \beta(\lambda + \tau) \} \}. \end{aligned} \quad (60)$$

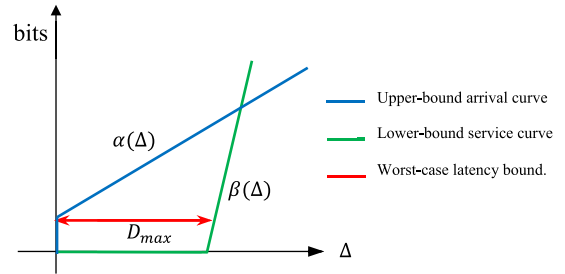


FIGURE 20. The worst-case latency bound using the Network Calculus approach.

### APPENDIX C CONSTRAINTS

Similarly, as presented in [20], the arrival curve is expressed here under the CBS shaper. First, the AVB-X credit deviation during  $\Delta t$  vary according to waiting and sending intervals and related slopes, as

$$\begin{aligned} \Delta cr_X^h &= cr_X^h(t + \Delta t) - cr_X^h(t) = \Delta t_X^+ \cdot idSL_X + \Delta t_X^- \cdot sdSL_X \end{aligned} \quad (61)$$

As  $\Delta t_X^+ = \Delta t - \Delta t^0 - \Delta t_X^-$ , we have

$$\begin{aligned} \Delta cr_X^h &= (\Delta t - \Delta t^0) \cdot idSL_X - \Delta t_X^- \cdot (idSL_X - sdSL_X) \\ &= (\Delta t - \Delta t^0) \cdot idSL_X - \Delta t_X^- \cdot C^h. \end{aligned} \quad (62)$$

Thus,  $\Delta cr_X^h$  will be bounded by the maximum and minimum credit variations, i.e.,  $\Delta cr_X^{h, max}$  and  $\Delta cr_X^{h, min}$ , as

$$\begin{aligned} \Delta cr_X^{h, min} &\leq \Delta cr_X^h \leq \Delta cr_X^{h, max}, \\ cr_X^{h, min} - cr_X^{h, max} &\leq \Delta cr_X^h \leq cr_X^{h, max} - cr_X^{h, min}, \\ \Delta cr_X^h &= (\Delta t - \Delta t^0) \cdot idSL_X - \Delta t_X^- \cdot C^h \\ &\geq cr_X^{h, min} - cr_X^{h, max}, \\ \Delta t_X^- \cdot C^h &\leq (\Delta t - \Delta t^0) \cdot idSL_X + cr_X^{h, max} \\ &\quad - cr_X^{h, min}. \end{aligned} \quad (63)$$

Hence, the highest arrival rate for AVB-X traffic must be considered to ensure safe latency calculations. Although with



non-preemption mode,  $\Delta t^0 = \Delta t_{TT}^0 + \Delta t_{GAP}^0$ , we assume all guard bands are used to send the largest-size AVB-X frames that start their transmissions. Thus,  $\Delta t^0 = \Delta t_{TT}^0$ , and

$$\begin{aligned} R_{TT}^{h*,np}(t + \Delta t) - R_{TT}^{h*,np}(t) \\ = C^h \cdot \Delta t_{TT}^0 = C^h \cdot \sum_{\Delta t} \left( W_{TT_j}^h - L_{j,j+1}^h \right) \geq \beta_{TT}^{h,np}(\Delta t), \end{aligned} \quad (64)$$

where  $\beta_{TT}^{h,np}(\Delta t)$  is the lower-bound service curve for TT traffic with non-preemption mode. Since  $R_X^{h*,np}(t)$  cannot be a decreasing function and using (64), (63) can be formed as

$$\begin{aligned} R_X^{h*,np}(t + \Delta t) - R_X^{h*,np}(t) \\ = \Delta t_X^- \cdot C^h \\ \leq \left[ \Delta t - \beta_{TT}^{h,np}(t) / C^h \right]^+ \cdot idSL_X + cr_X^{h,max} - cr_X^{h,min}. \end{aligned} \quad (65)$$

Hence, the upper-bound AVB-X arrivals happen when the TT service is assumed to be the lowest, which may differ depending on the referenced TT window. Thus, the lowest TT service curve can be found as

$$\beta_{TT}^{h,np}(\Delta t) = \min_{i \in \Delta t} \left\{ \beta_{TT}^{h,np,i}(\Delta t) \right\}, \quad (66)$$

where  $i$  refers to the reference TT window.  $\beta_{TT}^{h,np,i}(\Delta t)$  can be determined by aggregating the services produced from all TT windows with the  $i$ -th reference window, as

$$\beta_{TT}^{h,np,i}(t) = \sum_{j=i}^{i+N W_{GCL}^h - 1} \beta_{TT_j}^{h,np,i}(t), \quad (67)$$

and

$$\begin{aligned} \beta_{TT_j}^{h,np,i}(t) \\ = \beta_{T_{GCL}^h, W_{TT_j}^h}^{h,np} \left( t + T_{GCL}^h - W_{TT_j}^h + L_{j,j+1}^h - OD_{0,i}^h - OD_{j,i}^h \right), \end{aligned} \quad (68)$$

where  $L_{j,j+1}^h = \left[ t_j^{h,c} - t_{j+1}^{h,o} \right]^+$ ,  $OD_{0,i}^h = t_i^{h,o} - t_{i-1}^{h,c}$ , and  $OD_{j,i}^h = t_j^{h,o} - t_i^{h,c}$ . It can be noticed that the closing-edge overlapping ( $L_{j,j+1}^h$ ) of the  $j$ -th TT window is reduced from the overall window duration ( $W_{TT_j}^h$ ) as it will be considered to serve the next TT queue ( $Q_{TT_{j+1}}^h$ ) during its opening-edge ( $W_{TT_{j+1}}^h$ ). Although  $L_{j,j+1}^h$  cannot be guaranteed to serve  $Q_{TT_{j+1}}^h$  traffic, we consider here the aggregate effect from all TT queues on the AVB-X traffic, i.e., if  $L_{j,j+1}^h$  does not serve  $Q_{TT_{j+1}}^h$ , the  $Q_{TT_j}^h$  traffic will be served. Thus, it is assumed that the overlapping (contention-based [8]) interval is guaranteed to serve traffic from one overlapped TT window. Note that  $\beta_{T,L}^h(t)$  is expressed for the TDMA protocol [8] as

$$\beta_{T,L}^h(t) = C^h \cdot \max \left\{ \left\lfloor \frac{t}{T} \right\rfloor L, t - \left\lfloor \frac{t}{T} \right\rfloor (T - L) \right\}. \quad (69)$$

However, for preemption mode, unused gaps ( $GAP_{j,j+1}^h$ ), as bounded in (16), must be considered as frozen intervals and all guard bands are fully occupied by AVB-X transmissions, i.e., (65) is modified as

$$\begin{aligned} R_X^{h*,p}(t + \Delta t) - R_X^{h*,p}(t) \\ \leq \left[ \Delta t - \beta_{TT+GAP}^{h,p}(\Delta t) / C^h \right]^+ \cdot idSL_X + cr_X^{h,max} - cr_X^{h,min}, \end{aligned} \quad (70)$$

where

$$\begin{aligned} \beta_{TT+GAP}^{h,p}(\Delta t) \\ \leq C^h \cdot \left( \Delta t_{TT}^0 + \Delta t_{GAP}^0 \right) \\ = C^h \cdot \sum_{\Delta t} \left( W_j^h - L_{j,j+1}^h \right) + C^h \cdot \sum_{\Delta t} GAP_{j,j+1}^h. \end{aligned} \quad (71)$$

Thus, the lower-bound TT service curve with preemption mode ( $\beta_{TT+GAP}^{h,p}(\Delta t)$ ) can be found in the same manner as in (66) and (67). Hence, the service during the  $j$ -th TT window and the expected posterior empty gap can be formed as

$$\begin{aligned} \beta_{TT_j+GAP_{j,j+1}}^{h,p,i}(t) \\ = \beta_{T_{GCL}^h, W_{TT_j}^h}^{h,p} \left( t + T_{GCL}^h - W_{TT_j}^h + L_{j,j+1}^h \right. \\ \left. - GAP_{j,j+1}^h - OD_{0,i}^h - OD_{j,i}^h \right). \end{aligned} \quad (72)$$

By the same formulations as in (66)(67) and (69) with (72), the aggregate service curve for TT and GAP intervals can be determined with preemption mode.

## REFERENCES

- [1] N. Finn, "Introduction to time-sensitive networking," *IEEE Commun. Standards Mag.*, vol. 2, no. 2, pp. 22–28, Jun. 2018.
- [2] *IEEE Standard for Local and Metropolitan Area Networks Bridges and Bridged Networks Amendment 25: Enhancements for Scheduled Traffic*, Standard 802.1QbvTM-2015, Dec. 2015, pp. 1–57.
- [3] *IEEE Standard for Local and Metropolitan Area Networks Timing and Synchronization for Time-Sensitive Applications in Bridged Local Area Networks*, Standard 802.1AS-2020, Jan. 2020, pp. 1–421.
- [4] *Virtual Bridged Local Area Networks Amendment 12: Forwarding and Queuing Enhancements for Time Sensitive Streams*, Standard 802.1QavTM-2009, Dec. 2009, pp. 1–87.
- [5] 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.
- [6] L. Zhao, P. Pop, and S. S. Craciunas, "Worst-case latency analysis for IEEE 802.1Qbv time sensitive networks using network calculus," *IEEE Access*, vol. 6, pp. 41803–41815, 2018.
- [7] K. M. Shalghum, N. K. Noordin, A. Sali, and F. Hashim, "Network calculus-based latency for time-triggered traffic under flexible window-overlapping scheduling (FWOS) in a time-sensitive network (TSN)," *Appl. Sci.*, vol. 11, no. 9, p. 3896, Apr. 2021.
- [8] K. M. Shalghum, N. K. Noordin, A. Sali, and F. Hashim, "Critical offset optimizations for overlapping-based time-triggered Windows in time-sensitive network," *IEEE Access*, vol. 9, pp. 130484–130501, 2021.
- [9] L. 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. (RTAS)*, Apr. 2018, pp. 25–36.
- [10] J. Ren, D. Yang, E. Cui, and K. Gong, "An analytical latency model for AVB traffic in TSN considering time-triggered traffic," in *Proc. Int. Conf. Commun. Technol. (ICCT)*, Oct. 2020, pp. 938–943.

- [11] L. Zhao, P. Pop, Z. Zheng, H. Daigmore, and M. Boyer, "Latency analysis of multiple classes of AVB traffic in TSN with standard credit behavior using network calculus," *IEEE Trans. Ind. Electron.*, vol. 68, no. 10, pp. 10291–10302, Oct. 2021.
- [12] J.-Y. Le Boudec and P. Thiran, *Network Calculus: A Theory of Deterministic Queuing Systems for the Internet*, vol. 2050. Berlin, Germany: Springer-Verlag, 2004, p. 274. [Online]. Available: <http://portal.acm.org/citation.cfm?id=1755809>
- [13] L. Maile, K.-S. Hielscher, and R. German, "Network calculus results for TSN: An introduction," in *Proc. Inf. Commun. Technol. Conf. (ICTC)*, May 2020, pp. 131–140.
- [14] A. Finzi, A. Mifdaoui, F. Frances, and E. Lochin, "Network calculus-based timing analysis of AFDX networks with strict priority and TSN/BLS shapers," in *Proc. IEEE 13th Int. Symp. Ind. Embedded Syst. (SIES)*, Jun. 2018, pp. 1–10.
- [15] S. S. Craciunas, R. S. Oliver, M. Chmelfk, and W. Steiner, "Scheduling real-time communication in IEEE 802.1Qbv time sensitive networks," in *Proc. 24th Int. Conf. Real-Time Netw. Syst. (RTNS)*, Oct. 2016, pp. 183–192.
- [16] M. Kim, D. Hyeon, and J. Paek, "eTAS: Enhanced time-aware shaper for supporting non-isochronous emergency traffic in time-sensitive networks," *IEEE Internet Things J.*, early access, Nov. 13, 2021, doi: [10.1109/JIOT.2021.3124508](https://doi.org/10.1109/JIOT.2021.3124508).
- [17] A. Nasrallah, A. S. Thyagaturu, Z. Alharbi, C. Wang, X. Shao, M. Reisslein, and H. Elbakoury, "Performance comparison of IEEE 802.1 TSN time aware shaper (TAS) and asynchronous traffic shaper (ATS)," *IEEE Access*, vol. 7, pp. 44165–44181, 2019.
- [18] V. Gavrilut, L. Zhao, M. L. Raagaard, and P. Pop, "AVB-aware routing and scheduling of time-triggered traffic for TSN," *IEEE Access*, vol. 6, pp. 75229–75243, 2018.
- [19] V. Gavrilut and P. Pop, "Scheduling in time sensitive networks (TSN) for mixed-criticality industrial applications," in *Proc. 14th IEEE Int. Workshop Factory Commun. Syst. (WFCS)*, Jun. 2018, pp. 1–4.
- [20] M. M. Hasan, S. Khan, H. Feng, Q. Li, S. M. Masum, and M. T. Hasan, "Improved end to end delay in CBS using data compression for time sensitive network," in *Proc. 2nd Inf. Commun. Technol. Conf. (ICTC)*, May 2021, pp. 172–177.
- [21] A. Arestova, K.-S. J. Hielscher, and R. German, "Simulative evaluation of the TSN mechanisms time-aware shaper and frame preemption and their suitability for industrial use cases," in *Proc. IFIP Netw. Conf. (IFIP Netw.)*, Jun. 2021, pp. 1–6.
- [22] *IEEE Standard for Local and Metropolitan Area Networks Bridges and Bridged Networks—Amendment 26: Frame Preemption*, Standard 802.1Qbu-2016, Jun. 2016, pp. 1–52.
- [23] M. Ashjaei, L. Murselovic, and S. Mubeen, "Implications of various preemption configurations in TSN networks," *IEEE Embedded Syst. Lett.*, vol. 14, no. 1, pp. 39–42, Mar. 2022.
- [24] W. Guo, Y. Huang, J. Shi, Z. Hou, and Y. Yang, "A formal method for evaluating the performance of TSN traffic shapers using Uppaal," in *Proc. IEEE 46th Conf. Local Comput. Netw. (LCN)*, Oct. 2021, pp. 241–248.
- [25] E. Li, F. He, Q. Li, and H. Xiong, "Bandwidth allocation of stream-reservation traffic in TSN," *IEEE Trans. Netw. Service Manage.*, vol. 19, no. 1, pp. 741–755, Mar. 2022.
- [26] L. Zhao, F. He, and J. Lu, "Comparison of AFDX and audio video bridging forwarding methods using network calculus approach," in *Proc. IEEE/AIAA 36th Digit. Avionics Syst. Conf. (DASC)*, Sep. 2017, pp. 1–7.
- [27] X. Li and L. George, "Deterministic delay analysis of AVB switched Ethernet networks using an extended trajectory approach," *Real-Time Syst.*, vol. 53, no. 1, pp. 121–186, Jan. 2017.
- [28] *IEEE Standard for Local and Metropolitan Area Networks—Bridges and Bridged Networks—Amendment 31: Stream Reservation Protocol (SRP) Enhancements and Performance Improvements*, Standard 802.1Qcc-2018, Jun. 2018, pp. 1–208.
- [29] L. Zhao, P. Pop, Q. Li, J. Chen, and H. Xiong, "Timing analysis of rate-constrained traffic in TTEthernet using network calculus," *Real-Time Syst.*, vol. 53, no. 2, pp. 254–287, 2017.
- [30] 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. (ITC)*, Sep. 2018, pp. 1–6.
- [31] E. Wandeler and L. Thiele. (2006). *Real-Time Calculus (RTC) Toolbox*. [Online]. Available: <http://www.mpa.ethz.ch/Rtctoolbox>
- [32] J. A. R. De Azua and M. Boyer, "Complete modelling of AVB in network calculus framework," in *Proc. 22nd Int. Conf. Real-Time Netw. Syst. (RTNS)*, 2014, pp. 55–64.
- [33] E. Wandeler, L. Thiele, M. Verhoef, and P. Lieverse, "System architecture evaluation using modular performance analysis: A case study," *Int. J. Softw. Tools Technol. Transf.*, vol. 8, no. 6, pp. 649–667, 2006.



**KHALED M. SHALGHUM** (Student Member, IEEE) received the B.Sc. and M.Sc. degrees in communication engineering from the Department of Electrical and Electronic Engineering, University of Tripoli, Tripoli, Libya, in 2000 and 2008, respectively. He is currently pursuing the Ph.D. degree with University Putra Malaysia (UPM). From 2002 to 2008, he worked as an Engineer and, after that, as a Lecturer at the Faculty of Technical Engineering, Libya. From 2010 to 2018, he worked as a Lecturer at the Department of Electrical and Electronic Engineering, Azzaytuna University, Tarhuna, Libya. In 2018, he was offered a scholarship to complete his graduate study from the Libyan Ministry of Education, UPM. His research interests include wireless communications, broadband wireless access networks, WiMAX technology, time-sensitive networks, and latency performance evaluation.



**NOR K. NOORDIN** (Senior Member, IEEE) received the B.Sc. degree in electrical engineering from The University of Alabama, Tuscaloosa, AL, USA, in 1987, the M.Eng. degree from Universiti Teknologi Malaysia, and the Ph.D. degree from Universiti Putra Malaysia. She is currently working as a Professor with the Department of Computer and Communication System Engineering, Universiti Putra Malaysia. She has published more than 300 journals, book chapters, and conference papers. She has led many research projects. Her research interests include wireless communications and network systems.



**ADUWATI SALI** (Senior Member, IEEE) received the B.Eng. degree in electrical electronics engineering from The University of Edinburgh, U.K., in 1999, the M.Sc. degree in communications and network engineering from Universiti Putra Malaysia, in 2002, and the Ph.D. degree in mobile and satellite communications from the University of Surrey, U.K., in 2009. She was the Deputy Director of the Research Management Centre (RMC), Universiti Putra Malaysia, from 2016 to 2019. She is currently a Professor with the Department of Computer and Communication System Engineering and a Researcher with the Wireless and Photonic Network Research Center of Excellence (WiPNET), Universiti Putra Malaysia. Her research interests include mobile and satellite communication systems.



**FAZIRULHISYAM HASHIM** (Member, IEEE) received the M.Sc. degree from Universiti Sains Malaysia and the Ph.D. degree in telecommunication engineering from the University of Sydney, Australia. He is currently an Associate Professor with the Department of Computer and Communication Systems Engineering, Universiti Putra Malaysia. His research interests include heterogeneous wireless communication systems and network security.

• • •