Improved Network-Calculus Nodal Delay-Bounds in Time-Sensitive Networks

Ehsan Mohammadpour, Eleni Stai, Member, IEEE, and Jean-Yves Le Boudec, Fellow, IEEE

Abstract—In time-sensitive networks, bounds on worst-case delays are typically obtained by using network calculus and assuming that flows are constrained by bit-level arrival curves. However, in IEEE TSN or IETF DetNet, source flows are constrained on the number of packets rather than bits. A common approach to obtain a delay bound is to derive a bit-level arrival curve from a packet-level arrival curve. However, such a method is not tight: we show that better bounds can be obtained by directly exploiting the arrival curves expressed at the packet level. Our analysis method also obtains better bounds when flows are constrained with g-regulation, such as the recently proposed Length-Rate Quotient rule. It can also be used to generalize some recently proposed network-calculus delay-bounds for a service curve element with known transmission rate.

Index Terms—Time-sensitive networks, delay bound, arrival curve, packet-level constraint, bit-level constraint, network calculus.

I. INTRODUCTION

TIME-SENSITIVE networks provide real-time guarantees for applications such as avionics, automobiles, industrial automation, etc [1], [2], [3], [4], [5], [6]. IETF Deterministic Networking (DetNet) [7] and IEEE Time-Sensitive Networking (TSN) [8] formalize the requirements and provide standardization for such networks. One of the main goals in time-sensitive networks is to provide guarantees on worst-case delay, and not average delay. In order to obtain such guarantees, flows are assumed to be regulated at the sources. A classical form of source regulation is the bit-level arrival curve constraint: a flow satisfies a bit-level arrival curve constraint \( \alpha() \) if the number of bits over any time interval of any duration \( t \) is upper bounded by \( \alpha(t) \) [9], [10]. The celebrated token bucket constraint with rate \( r \) and burst \( b \) is an example, with \( \alpha(t) = rt + b \) for \( t > 0 \). Formally proven delay bounds can be obtained by using network calculus, which combines arrival and service curves [9], [10]. A service curve is an abstraction that conveys information on the minimum service provided by a system (see Section III). Then the classical network-calculus delay-bound is obtained by taking the horizontal deviation of the arrival and service curves [9, Section 3] [10, Section 3.1.2]. This delay bound can be improved for some systems with known transmission rate [11].

However, in time-sensitive networks standards, flow regulation at sources is expressed in terms of number of packets rather than number of bits [12] [13, Section 35.2.2.8.4]; for example in TSN, the number of packets observed within any fixed class measurement interval (CMI) is upper-bounded by a constant value. To obtain delay bounds for such flows, network calculus is often used [14], for which a bit-level characterization is required. Hence, a common approach is to derive a bit-level arrival curve from the packet-level constraint and then apply network calculus [15], [16]. However, as we show in Table II and Section VIII, this can lead to delay bounds that can be improved. Indeed, our main result, in Theorem 1, is a new delay bound for flows regulated with packet-level constraints, which improves on the one that is obtained by deriving a bit-level arrival curve from the packet-level regulation constraint. We show that the obtained delay bound is tight at least for \( c \)-Lipschitz [17] service curves where \( c \) is the physical line-speed (Theorems 2 and 3).

Our method uses a novel modelling of packet-level constraint with g-regulation (Proposition 2). The concept of g-regulation was introduced by C.S. Chang [18] as an alternative to bit-level arrival curve and uses max-plus algebra, whereas results for bit-level arrival curves tend to use min-plus algebra. Our improvement in the delay bound is made possible by combining min-plus representation of service curves and max-plus representation of the input traffic to obtain a bound on queuing delay (Lemma 1).

As a by-product of our method of proof, we also obtain delay bounds for flows with g-regulation, such as Length-Rate Quotient (LRQ) [19], that share a FIFO system with known service curve (Theorem 4). This bound improves on the one obtained by deriving a bit-level arrival curve from g-regulation and then using the state-of-the-art network calculus bound. Moreover, it generalizes the results by [19] that was specifically applicable to flows with LRQ constraints and priority queues with constant rate servers. Last, the state-of-the-art network-calculus delay-bound for bit-level arrival curve can also be generalized to a wider family of service curves than rate-latency ones (Theorem 5).

The rest of the paper is organized as follows. Section II presents the state-of-the-art and related works. Section III includes the system model, notation, and the definition of the considered flow regulation constraints. Section IV presents the relations among different regulation constraints. In Section V,
we present the main contribution of this paper, namely, new delay bounds for flows with packet-level constraints, and we show that these bounds are tight. Section VI gives a generalization of the existing delay bounds for flows with g-regularity constraint and bit-level arrival curve, as a by-product of our method of proof in Section V. Section VII shows that the best delay bound for a given flow is obtained by directly applying the theorem corresponding to its initial constraint. Section VIII provides a numerical illustration of the theorems presented in this paper and Section IX concludes the paper. The Appendix gives proofs.

II. RELATED WORKS

The classical network-calculus proves delay bounds for a FIFO system with bit-level arrival and service curves [9], [10]. Specifically, for a flow with bit-level arrival curve \( \alpha \) that enters a FIFO system with service curve \( \beta \), the bound is obtained by taking the horizontal deviation between the two curves, i.e., \( h(\alpha, \beta) \), which is defined in [9, Section 3.1.11] and recalled in Section III. For example, when a flow has leaky bucket arrival curve \( \alpha(t) = rt + b \) and the FIFO system has rate-latency service curve \( \beta(t) = \max((R(t - T), 0)) \), we have \( h(\alpha, \beta) = T + \frac{b}{R} \) provided that \( r \leq R \). This bound is tight if only arrival and service curves are known. However, we often have more information, specially in the context of time-sensitive networks, e.g., TSN schedulers with Credit-Based Shapers [20] and Deficit Round Robin (DRR) schedulers [21]: when a packet starts its transmission, it is transmitted with full line rate. Recently, such information was exploited for rate-latency service curves to provide a delay bound that improves the classical network-calculus bound [11]; more precisely, the classical network-calculus bound is reduced by \( L_{\text{min}}(\frac{1}{R} - \frac{1}{r}) \), where \( c \) is the physical line-rate and \( L_{\text{min}} \) is the minimum packet length of the input traffic.

While rate-latency service curves are commonly used in the literature, they may not provide a perfect characterization of a system. A number of works provide more complex service curves that in turn leads to better delay bounds. In [21], the authors obtained a non rate-latency service curve for Deficit Round-Robin (DRR) schedulers that can incorporate the arrival curves of the interfering flows. Such non rate-latency service curves are obtained as well for Weighted Round-Robin (WRR) [10, Section 8.2.4] and Interleaved WRR [22]. Hence, the improvement of [11] does not apply to these service curves. In this paper, we generalize the results of [11] to improve delay bounds for non rate-latency service curves. Furthermore, we improve the bounds in [11] when the flows have packet-level constraints.

A number of other works focus on improving the arrival curves of the flows in time-sensitive networks by taking advantage of input-line shaping effect. For example, for a flow with leaky-bucket arrival curve \( \alpha(t) = rt + b \) that passes a physical line with rate \( c \), we can obtain a better arrival curve \( \alpha'(t) = \min(rt + b, ct) \). In [15] and [16], the authors study a TSN network and assume the input traffic has packet-level constraint [13, Section 35.2.2.8.4]. To obtain delay bounds, they first derive a bit-level arrival curve from the packet-level constraint; then they exploit input-line shaping to improve the obtained arrival curve. Finally, they use network calculus to obtain delay bounds. The improvement by the input-line shaping effect is complementary to the improvements in this paper.

Recently, LRQ was introduced in the context of interleaved regulators in [19] as a per-flow regulation constraint that is simpler to implement than token-bucket. LRQ is in fact a specific case of Shifted-Rate Regulator (when the delay term is set to zero) [18, Section 6.2.1], which belongs to the family of g-regulation. Reference [19] obtains delay bounds for flows with LRQ constraint within constant-rate servers and strict-priority queuing. This analysis was extended for guaranteed-rate servers [23]. Unlike the previous works, our results on g-regulated flows cover the whole family of g-regulation (i.e., are not limited to LRQ constraint) and apply to a broader class of network nodes (i.e., nodes with arbitrary service curves).

III. SYSTEM MODEL AND GENERAL PREREQUISITE

In this section, we first describe the systems to which our analysis applies. Second, we recall the definition of pseudo-inverse functions as they are used throughout the paper. Third, we give a description of the flow regulation constraints considered in this paper, i.e., bit-level arrival curve, g-regulation and packet-level arrival curve. Finally, we give a summary of necessary mathematical definitions required to follow the content of the paper. For the reader’s convenience, Table I gives the notation used throughout the paper.

A. System Model

We study a FIFO system with a queue and a transmission link, as in Fig. 1. Each queue is shared among a number of flows. Upon arrival, packets of different flows enter the queue and are stored in FIFO order. A scheduler decides when the packet at the head of the queue is selected for transmission. The scheduler typically arbitrates between this queue and other queues (not shown), therefore the packet at the head of the queue may have to wait even if there is no packet of this queue in transmission. When the packet at the head of this queue is selected for transmission, it is transmitted at a constant rate \( c \) until it is completely transmitted, i.e., there is no preemption.

Every flow is assumed to be constrained by either bit-level arrival curve (Section III-C), g-regulation (Section III-D) or packet-level arrival curve (Section III-E). For each flow \( j \), let \( L_j^{\text{min}} \) and \( L_j^{\text{max}} \) denote the minimum and maximum packet lengths, in bits. We also assume that the total flow of all incoming packets is packetized, i.e., we consider that all bits of any packet arrive at the same time instant.

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
Let $A_n$ be the arrival time of packet $n$ to the FIFO system, where the numbering of packets is by order of arrival and $Q_n$ be the time at which packet $n$ is selected for transmission. The FIFO assumption means that $Q_n \leq Q_{n+1}$. We call $Q_n - A_n$ the queuing delay of packet $n$ from the FIFO system. Let $l_n$ be the length in bits of packet $n$, so that $L_{\text{min}} \leq l_n \leq L_{\text{max}}$, then packet $n$ leaves the system at time $D_n = Q_n + \frac{l_n}{c}$. We call $D_n - A_n$ the response time of the FIFO system for packet $n$.

We assume that the scheduler is such that the FIFO system offers to the total flow of all incoming packets a service curve $\beta$, i.e., the total flow of all incoming packets a service curve $\beta$. Formally, this is defined as follows [9, Section 1.3]. Let $\mathcal{F}^0_{\text{inc}}$ be the set of wide-sense increasing functions $w : [0, +\infty) \to [0, +\infty]$ such that $w(0) = 0$. Let $A, D \in \mathcal{F}^0_{\text{inc}}$ be such that $A(t)$ [resp. $D(t)$] denotes the cumulative number of bits arrived in [resp. departed from] the system until time $t$ (excluded). A function $\beta \in \mathcal{F}^0_{\text{inc}}$ is a service curve for the system if for every time $t \geq 0$ there exists a time $s \leq t$ such that

$$D(t) \geq \beta(t-s) + A(s).$$  \hspace{1cm} (1)

The concept of service curve was initially introduced to specify the minimum amount of bits that departs from the system during any interval of time when the system backlog is non-zero [24], i.e., $D(t) - D(s) \geq \beta(t-s)$. Eq. (1) is an extended and more common definition introduced by [9] which enables obtaining of service curves for bounded-delay systems and for a concatenation of systems with known service curves. Service curve characterizations are available for many systems, see for example [25], [26], [27], [28], [29]. Rate-latency service curves are functions of the form $\beta(r) = R(t-T) + R_T$ the rate and latency terms. A system that offers a rate-latency service curve can be interpreted as behaving, for the flows of interest, as if it would be a server with rate $R$ and vacation $T$. The rate $R$ is the rate guaranteed to the flow and is typically less than the line rate $c$. Rate-latency service curves are often used because of their simplicity, but better delay bounds can also be obtained with more complex service curves [21].

### B. Pseudo-Inverse Functions

For $w \in \mathcal{F}^0_{\text{inc}}$, the lower and upper pseudo-inverses are respectively $w^\dagger$ and $w^\ddagger$, and are defined as:

$$w^\dagger(x) \overset{\text{def}}= \inf \{s \geq 0 \mid w(s) \geq x\},$$  \hspace{1cm} (2)

$$w^\ddagger(x) \overset{\text{def}}= \sup \{s \geq 0 \mid w(s) \leq x\}.$$  \hspace{1cm} (3)

Eq. (2) illustrates the pseudo-inverse functions and the differences between them. By [30, Section 10.1]:

- $w^\dagger$ is non-decreasing and left continuous.
- $w^\ddagger$ is non-decreasing and right continuous.

### C. Bit-Level Arrival Curve

The bit-level arrival curve constraint is the most widespread form of traffic regulation [9]. Consider a left-continuous function $A \in \mathcal{F}^0_{\text{inc}}$, where $A(t)$ denotes the cumulative number of bits observed on the flow of interest until time $t$ (excluded).
We say that the flow is constrained by the bit-level arrival curve \( \alpha \in \mathcal{F}_\text{inc}^0 \) if and only if
\[
A(t) - A(s) \leq \alpha(t - s) \quad \text{for all } t \geq 0 \text{ and } s \in [0, t]. \tag{4}
\]

In other words, \(4\) states that the total number of bits observed a packetized flow, it is shown in \([31]\) that this is equivalent to
\[
\sum_{k=m}^{n} l_k \leq \alpha^+(A_n - A_m) \quad \text{for all } m, n \in \mathbb{N}, m \leq n \tag{5}
\]
where \(\alpha^+\) is the right-limit of \(\alpha\), and this is also equivalent to
\[
A_n - A_m \geq \alpha^+(\sum_{k=m}^{n} l_k) \quad \text{for all } m, n \in \mathbb{N}, m \leq n. \tag{6}
\]

By \([9, \text{Lemma 1.2.1}]\), if \(\alpha\) is a bit-level arrival curve for a flow, then so is its left limit, i.e., \(\alpha^-\). Also, since input is packetized, \(\alpha^-\) is an upper bound on the size of any packet. Therefore, in the rest of this paper, we assume that \(\alpha\) is left-continuous and \(\alpha^+(t) \geq L_{\text{max}}\).

Frequently used bit-level arrival curves are \([9, \text{1.2.2}]\):

- the token-bucket (or leaky-bucket) arrival curve with rate \(r\) and burst \(b\), defined by \(\alpha(t) = rt+b; t > 0, \alpha(0) = 0\);
- the staircase function with burst \(b\) and interval \(\tau\) defined by \(\alpha(t) = b\lceil\frac{t}{\tau}\rceil\). It expresses the constraint that the flow has at most \(b\) bits within any interval of duration \(\tau\).

D. g-Regulation

The g-regulation constraint was introduced in \([18]\); it specifies that the time inter-spacing between packets is lower bounded by a left-continuous function \(g \in \mathcal{F}_\text{inc}^0\) \([18]\).\(^1\) More formally, a packetized input has g-regulation constraint if and only if for any packet indices \(m, n\), where \(m \leq n\):
\[
A_n - A_m \geq g \left( \sum_{k=m}^{n-1} l_k \right). \tag{7}
\]

Note the difference between the bit-level arrival curve in \(6\) and the g-regulation in \(7\): for the g-regulation the length of the last packet, i.e., \(l_n\), does not play a role in \(7\) while it is included in \(6\). Since the argument of the function \(g\) only takes values in a discrete set containing sums of packet sizes, we assume \(g\) is left continuous at each point in this set.

**Shifted-Rate Regulation** \([18, \text{Section 6.2.1}]\) is a type of g-regulation with \(g(x) = \frac{1}{r}[x - d]^+,\) where \(r\) is the regulation rate and \(d\) is the regulation delay, i.e., the constraint is
\[
A_n - A_m \geq \frac{1}{r} \left( \sum_{k=m}^{n-1} l_k - d \right)^+, \quad \forall m, n \in \mathbb{N}, m \leq n. \tag{8}
\]

**Length-Rate Quotient** (LRQ), introduced by \([19]\), is a traffic regulation which specifies the minimum interspacing between two consecutive packets as a function of a regulation rate \(r\) and the length of the earlier packet. i.e. the constraint is
\[
A_n - A_{n-1} \geq \frac{n-1}{r}, \quad \forall n \in \mathbb{N}, n \geq 2. \tag{9}
\]

By LRQ, the arrival of packet \(n\) is constrained by the regulation rate, the arrival time and the length of the previous packet; the fact that there is dependency on only the last packet, renders the implementation of LRQ very simple. It is easy to see that, when \(d = 0\), \(8\) is equivalent to \(9\), namely, LRQ is shifted-rate regulation with \(d = 0\); therefore, LRQ is a form of g-regulation with \(g(x) = \frac{x}{r}\).

E. Packet-Level Arrival Curve

The packet-level arrival curve is mainly used in the context of IEEE TSN and IETF DetNet and expresses traffic constraints at the packet level. More formally, consider a left-continuous wide-sense increasing function \(N : \mathbb{R}^+ \to \mathbb{N} \cup \{0\}\), where \(N(t)\) is the cumulative number of packets observed on a flow of interest until time \(t\) (excluded). Then, we say that the flow has a packet-level arrival curve \(\alpha_{\text{pkt}} : \mathbb{R}^+ \to \mathbb{N}\), \(\alpha_{\text{pkt}}(0) = 0\), if and only if
\[
N(t) - N(s) \leq \alpha_{\text{pkt}}(t - s), \quad \text{for all } t \geq 0 \text{ and } s \in [0, t]. \tag{10}
\]

Similarly to the bit-level arrival curve constraint, \(10\) is equivalent to:
\[
n - m + 1 \leq \alpha_{\text{pkt}}^+(A_n - A_m) \quad \text{for all } m, n \in \mathbb{N}, m \leq n. \tag{11}
\]

It is known that an arrival curve can always be assumed to be sub-additive (i.e. satisfy \(\alpha_{\text{pkt}}(s + t) \leq \alpha_{\text{pkt}}(s) + \alpha_{\text{pkt}}(t)\)), since otherwise it can be replaced by its sub-additive closure \([9]\). Also, following the same steps as the proof of \([9, \text{Lemma 1.2.1}]\), we can prove that if \(\alpha_{\text{pkt}}\) is a sub-additive packet-level arrival curve for a flow, then so is its left limit \(\alpha_{\text{pkt}}^-\). Also, since the input is packetized, \(\alpha_{\text{pkt}}^+(0)\) gives an upper bound on one packet. Therefore, in the rest of this paper, we assume that \(\alpha_{\text{pkt}}\) is sub-additive, left continuous and \(\alpha_{\text{pkt}}^+(0) \geq 1\).

In IEEE TSN the traffic specifications for a flow is defined as \([13, \text{Section 34.6.1}]\):

"...during each class measurement interval, it [a source] can place up to MaxIntervalFrames data frames, each no longer than MaxFrameSize into that stream’s queue."

Similarly, the traffic specification in IETF DetNet allows packet level constraint \([12, \text{Section 5.5}]\) by the attributes "Interval" as the period of time in which the traffic specification is specified and "MaxPacketsPerInterval" as the maximum number of packets that a source transmits in one Interval.

By the above specifications, the interval (i.e. class measurement interval in TSN or Interval in IETF DetNet) can be interpreted differently as seen in Fig. 3:

1) Sliding interval \((\tau, K)\): the number of packets is limited by \(K = \text{MaxIntervalFrames} (\text{or MaxPacketsPerInterval})\) at any sliding interval of duration \(\tau\).
2) Fixed interval \((\tau, K)\): the number of packets is limited by \(K = \text{MaxIntervalFrames} \) or \(\text{MaxPacketsPerInterval}\) at any reference interval of duration \(\tau\); the reference intervals are consecutive and non-overlapping.

The first interpretation, sliding interval \((\tau, K)\), is equivalent to

\[
N(t + \tau) - N(t) \leq K; \quad \forall t \geq 0,
\]

which is also equivalent to saying that the flow has a packet-level arrival curve given by

\[
\alpha_{\text{pkt}}(t) = K\left\lfloor \frac{t}{\tau} \right\rfloor.
\]

This is a staircase packet-level arrival curve with burst equal to \(K\) packets and period \(\tau\). It is easy to verify that it is sub-additive and left-continuous. Applying the right limit of (13) into (11), (12) is equivalent to the following for any packet indices \(m, n, m \leq n:\)

\[
n - m + 1 \leq K\left\lfloor \frac{A_n - A_m}{\tau} \right\rfloor + K.
\]

The second interpretation, fixed interval \((\tau, K)\), is equivalent to saying that there exists some offset \(\theta\) such that:

\[
N(\theta) = N(0) = 0,
\]

\[
N(\theta + (i + 1)\tau) - N(\theta + i\tau) \leq K; \quad \forall i \in \mathbb{N} \cup \{0\}.
\]

As shown in Lemma 5 in Appendix A, the second interpretation, fixed interval \((\tau, K)\), implies a packet-level arrival curve with,

\[
\alpha_{\text{pkt}}(0) = 0, \quad \alpha_{\text{pkt}}(t) = K\left\lfloor \frac{t}{\tau} \right\rfloor + K : t > 0,
\]

It is easy to verify that this function is sub-additive and left-continuous. However, the converse does not hold, i.e., it is not true that all flows that have the packet-level arrival curve in (16) satisfy fixed interval \((\tau, K)\) (because such an arrival curve allows \(2K\) packets in an interval of duration less than \(\tau\)). Nonetheless, we show in Section V that the delay bound obtained using the packet-level arrival curve in (16), is tight for flows with the fixed interval \((\tau, K)\) regulation constraint.

Another form of packet-level regulation found in the literature is the token-bucket packet-level constraint [31], which limits the number of packets within any time interval \(t\) to \(\rho t + B\), where \(\rho > 0\) and \(B \geq 1\) are respectively the packet rate and burst. This constraint enjoys the superposition property, i.e., for an aggregation of flows each with such constraint, the superposition is constrained by a token-bucket packet-level constraint with \(\rho\) and \(B\) equal to the sum of the packet rates and bursts of the flows respectively. Such a constraint is equivalent to the following packet-level arrival curve:

\[
\alpha_{\text{pkt}}(0) = 0, \quad \alpha_{\text{pkt}}(t) = \lceil \rho t + B - 1 \rceil : t > 0.
\]

It can be easily verified that this function is sub-additive and left-continuous. Now, by (11), for any packet indices \(m, n\), where \(m \leq n\), (17) is equivalent to:

\[
n - m + 1 \leq \lfloor \rho(A_n - A_m) + B \rfloor.
\]

The \((\lambda, \nu)\)-constraint introduced in [32], is equal to the token-bucket packet-level arrival curve with \(\rho = \lambda\) and \(B = \nu + 1\). The staircase packet-level arrival curve in (11) implies a token-bucket packet-level arrival curve with \(\rho = \frac{\tau}{K}, B = K\).

**F. Additional Definitions**

For two functions \(w, w' \in F_{\text{inc}}^0\), the horizontal deviation from \(w'\) to \(w\) is defined as [9]:

\[
h(w', w) \overset{\text{def}}{=} \sup_{t \geq 0} \{w(t) - w'(t)\}.
\]

See Fig. 4.

A function \(w \in F_{\text{inc}}^0\) is called \(c\)-Lipschitz (for \(c \geq 0\)) if \(\forall t_1, t_2 \in \mathbb{R}^+\) [17, Section 41.5]:

\[
|w(t_2) - w(t_1)| \leq c|t_2 - t_1|.
\]

A \(c\)-Lipschitz function is necessarily continuous and the slope of the function is within \([-c, c]\) at any point in time. Fig. 5 shows examples of a \(c\)-Lipschitz and non-\(c\)-Lipschitz functions.

For \(w \in F_{\text{inc}}^0\), \(w^+\) and \(w^-\) are the right and left limits, defined as:

\[
\forall x \in \mathbb{R}^+ : \quad w^+(x) = \lim_{\varepsilon \to 0^+} w(x + \varepsilon),
\]

\[
\forall x \in \mathbb{R}^+ : \quad w^-(x) = \lim_{\varepsilon \to 0^+} w(x - \varepsilon).
\]
IV. RELATIONS AMONG FLOW REGULATIONS

We have seen various families of traffic regulations in Section III. We can find regulation-specific toolbox for delay analysis, e.g., [9], [10] for bit-level arrival curves. The immediate question is whether there is a relation among these regulation types. Such relations can open up new opportunities to apply the existing toolbox of one regulation type to another, that may lead to delay improvements, as we will see in Section V. The goal of this section is to present such derivations between the regulation types.

In the next proposition, we present a relation between g-regularity and bit-level arrival curve. The derivation of bit-level arrival curve from g-regularity already exists in literature [18] and we put it here for completeness.

**Proposition 1:** Consider a flow \( f \).

1. If \( f \) has a bit-level arrival curve \( \alpha(t) \), it also conforms to a g-regularity constraint with \( g(x) = \alpha^{-1}(x+L_{\min}) \), \( x > 0; \alpha(0) = 0 \).
2. If \( f \) conforms to a g-regularity constraint, it also has a bit-level arrival curve \( \alpha(t) = g(\x(t)) + L_{\max} \).
3. The sequential application of items 1 and 2 or items 2 and 1 gives a regulation constraint that is weaker than the initial one, except if all packets have the same size.

The proof is in Appendix B-A. Proposition 1 shows that even though bit-level arrival curve and g-regularity are two different families of traffic constraints, they can be derived from each other in items (1) and (2). However, item (3) shows that when packets are of different sizes, such derivations lead to weaker constraints than the initial traffic constraints.

**Proposition 2:** Consider a flow with packet-level arrival curve \( \alpha_{\text{pkt}} \), then:

1. The flow conforms to a g-regularity constraint with:
   \[
   g(x) = \alpha_{\text{pkt}}^{-1}\left(\frac{x}{L_{\max}} + 1\right),
   \]
   and to a bit-level arrival curve constraint, \( \alpha \), with:
   \[
   \alpha(t) = L_{\max} \alpha_{\text{pkt}}(t).
   \]
2. The g-regularity constraint in (23) is stronger than the bit-level arrival constraint in (24); i.e., the sequential application of (23) and then item 2 of Proposition 1 gives (24) while the sequential application of (24) and then item 1 of Proposition 1 gives a looser constraint than (23), except if all packets have the same size.

The proof is in Appendix B-B. By item (1), we can use the bit-level arrival curve and g-regularity to compute delay bounds for flows with packet-level arrival curves. In fact, for TSN/DetNet traffic regulation, such bit-level arrival curve derivation already exists [15] and is used for delay-bound computation. While in literature only the bit-level arrival curve derivation is used to obtain delay bounds, item (2) shows that the g-regularity derivation is a stronger constraint, and it leads to delay improvements as we will see in Theorem 1.

V. DELAY BOUNDS FOR PACKET-LEVEL ARRIVAL CURVES

The focus of this section is to find delay bounds for flows with packet-level arrival curve that share a FIFO system described in Section III. To do so, we use a combination of g-regularity and bit-level arrival curves derived from packet-level arrival curves. We start with a technical lemma about queuing delay.

**Lemma 1:** Consider a FIFO system with service curve \( \beta \in \mathcal{F}_0 \). Assume that we know some wide-sense increasing function \( w : [0, +\infty) \rightarrow [0, +\infty] \) such that for any two packet indices \( m, n, m \leq n \), we have

\[
\sum_{k=m}^{n-1} l_k \leq w(A_n - A_m),
\]

where \( A_n, A_m \) are packet arrival times and \( l_k \) is the length of packet \( k \). Then, the queuing delay of the FIFO system is bounded by \( h(w, \beta) \), i.e., for any packet index \( n \):

\[
Q_n - A_n \leq h(w, \beta);
\]

where \( h \) is the horizontal deviation.

The proof is in Appendix B-C. It combines min-plus representation of the service curve and the max-plus representation of the input traffic in (25).

Using Lemma 1, we obtain our first result on delay bounds.

**Lemma 2:** Consider a FIFO system offering a service curve \( \beta \) to the aggregate of the flows sharing it. Assume that (i) flow \( 1 \) is the flow of interest and conforms to a g-regularity constraint, and (ii) flow 2 represents the aggregate of the remaining flows sharing the FIFO system and has bit-level arrival curve \( \alpha \). In addition, assume that as soon as a packet starts to be transmitted, it is transmitted with rate \( c \). Then,

(i) An upper bound on the response time of a packet with length \( l \) of flow \( 1 \) at this FIFO system is:

\[
\Delta_{\text{AG}}(l) = h(g_1^{l} + \alpha^+, \beta) + \frac{l}{c},
\]

(ii) An upper bound on the response time of any packet of flow 1 at this FIFO system is:

\[
\Delta_{\text{AG}} = h(g_1^{l} + \alpha^+, \beta) + \frac{L_{\max}}{c}.
\]

The proof is in Appendix B-D. We can now apply Lemma 2 to packet-level arrival curves and obtain the following theorem:

**Theorem 1:** Consider a FIFO system offering a service curve \( \beta \) to the aggregate of the flows sharing it. Assume any flow \( f \) has \( \alpha_{\text{pkt},f} \) as packet-level arrival curve at the entrance of the FIFO system and has maximum packet length \( L_{\max} \). In addition, assume that as soon as a packet starts to be transmitted, it is transmitted with rate \( c \). Then,

(i) An upper bound on the response time of a packet with length \( l \) of a flow \( f \) at this FIFO system is:

\[
\Delta_{\text{pkt}}(l) = h\left(\sum_{i=1}^{M} L_{i}^{\max} \alpha_{\text{pkt},i} + L_{f}^{\max} \beta\right) + \frac{l}{c}.
\]

(ii) An upper bound on the response time of any packet of flow \( f \) at this FIFO system is:

\[
\Delta_{\text{pkt}} = h\left(\sum_{i=1}^{M} L_{i}^{\max} \alpha_{\text{pkt},i} + L_{f}^{\max} \beta\right) + \frac{L_{\max}}{c}.
\]

Proof: Use item 1 of Proposition 2 to derive a g-regularity constraint for the flow of interest and a
Theorem 2: The bound of Theorem 1 is tight. More specifically, consider a c-Lipschitz service curve $\beta$, sub-additive left-continuous functions $\alpha_{\text{pkt}, f}$ as packet-level arrival curves for flow $f \in \{1, \ldots, M\}$, with maximum packet lengths $L_{f}^{\max}$, and a transmission rate $c$. There exists a simulation trace of a FIFO system, shared between the $M$ flows, where a packet of flow 1 reaches the bound in (29) of Theorem 1.

The proof is in Appendix B-E; it consists in building a trajectory with greedy sources, i.e., with sources for which the cumulative packet arrival function $N(t)$ satisfies $N(t) = \alpha_{\text{pkt}}(t-t_0)$ for some offset $t_0$.

Recall that there are two interpretations of the packet-level regulation constraint of TSN/DetNet (Section III-E). The former, sliding interval, is equivalent to a packet-level arrival curve constraint; the latter, fixed interval, implies a packet-level arrival curve constraint but is not equivalent to it. However, as we show in the next theorem, the delay bound obtained by using packet-level arrival curves and Theorem 1 is the best possible bound for flows with TSN/DetNet constraints.

Theorem 3: Consider a FIFO system offering a $c$-Lipschitz service curve $\beta$ to an aggregate of $M$ flows, where, as soon as a packet starts to be transmitted, it is transmitted with rate $c$. Assume that every flow $f \in \{1, \ldots, M\}$ conforms to either the Sliding interval $(T_f, K_f)$ or the Fixed interval $(\tau_f, K_f)$ constraint of TSN/DetNet (Section III-E) and has maximum packet size $L_{f}^{\max}$. For every flow $f$, we can derive a packet-level arrival curve, by using (13) or (16), and apply Theorem 1 to obtain a delay bound $\Delta_{f}^{\text{pkt}}$.

Then, for every flow $f$ and every $\varepsilon > 0$, there exists a simulation trace of this system where a packet of flow $f$ experiences a delay in the interval $[\Delta_{f}^{\text{pkt}} - \varepsilon, \Delta_{f}^{\text{pkt}}]$.

The proof is in Appendix B-F. The main issue here is that a flow that conforms to Fixed interval $(\tau, K)$ has packet-level arrival curve given by (16), but a greedy source for this arrival curve does not satisfy Fixed interval $(\tau, K)$ (as it generates $2K$ packets in one interval of duration $\tau$). We overcome this by considering, for every $\varepsilon > 0$, a greedy source for the packet-level arrival curve $K[\frac{t_{\min} - \varepsilon}{\tau}] + K$, which does conform to Fixed interval $(\tau, K)$.

VI. IMPROVEMENTS ON EXISTING NETWORK-CALCULUS DELAY BOUNDS

In this section, as a by-product of Lemmas 1 and 2, we derive novel delay bounds for flows with g-regularity constraints (Theorem 4) and bit-level arrival curves (Theorem 5). The obtained delay bounds for flows with g-regularity constraint generalize the results of [19], which are specific to LRQ and strict-priority schedulers. Similarly, for flows with bit-level arrival curve, our delay bounds improve on the results obtained by classical network-calculus and generalize the results of [11], which are specific to rate-latency service curves.

Theorem 4: Consider a FIFO system offering a service curve $\beta$ to the aggregate of $n$ flows sharing it. Assume any flow $f \in \{1, \ldots, M\}$ has $g_f$-regulation constraint at the entrance of the FIFO system and has maximum packet length $L_{f}^{\max}$. In addition, assume that as soon as a packet starts to be transmitted, it is transmitted with rate $c$. Then,
(i) An upper bound on the response time of a packet with length $l$ of a flow $f$ at this FIFO system is given as follows:

$$\Delta_G^c(l) = h\left(\sum_{i=1}^{M} q_i^{1} + \sum_{i=1}^{M} L_i^{\max} 1_{i \neq f} \right) + \frac{l}{c}. \quad (32)$$

(ii) An upper bound on the response time of any packet of flow $f$ at this FIFO system is given as follows:

$$\Delta_G = h\left(\sum_{i=1}^{M} q_i^{1} + \sum_{i=1}^{M} L_i^{\max} 1_{i \neq f} \right) + \frac{L_i^{\max}}{c}. \quad (33)$$

The proof consists in two steps; we first take the sum of the bit-level arrival curves of the competing flows derived using item 1 of Proposition 1; second, we apply Lemma 2.

For the most common form of g-regularity constraint, i.e., LRQ, the current available delay bound is by [19] which only applies to strict-priority schedulers; under the same assumption, Theorem 4 gives the same delay bounds as [19]. The delay bounds presented in [19] do not apply to a vast majority of schedulers, e.g., WRR, Interleaved WRR, and DRR. Theorem 4, however, can be used to obtain delay bounds for such scheduling mechanisms. We provide a numerical illustration of Theorem 4 for a DRR scheduler in Section VIII-B.

Next, we present the delay bounds for flows with bit-level arrival curves. It follows from the definition that the bit-level arrival curves can be superimposed: an aggregate of flows conforms to a bit-level arrival curve equal to the sum of the bit-level arrival curves of the flows. Hence, in the next theorem, we assume an aggregate of flows (that share a FIFO system with the flow of interest) is constrained by one bit-level arrival curve constraint.

**Theorem 5:** Consider a FIFO system offering a service curve $\beta$ to the aggregate of the flows sharing it and such that, as soon as a packet starts to be transmitted, it is transmitted with rate $c$. Flow 1 is the flow of interest and flow 2 represents the aggregate of the remaining flows sharing the FIFO system. Assume that that flows 1, 2 have bit-level arrival curves $\alpha$, $\alpha'$ respectively; then,

(i) an upper bound on the response time of a packet with length $l$ of flow 1 at this FIFO system is

$$\Delta^A(l) = h(\alpha^+ + \alpha'^+ - l, \beta) + \frac{l}{c}; \quad (34)$$

(ii) if $\beta$ is $c$-Lipschitz, an upper bound on the response time of any packet of flow 1 at this FIFO system is

$$\Delta^A = h(\alpha + \alpha' - L_1^{\min}, \beta) + \frac{L_1^{\min}}{c}. \quad (35)$$

The proof is in Appendix B-G. It consists in two steps. We first derive the queuing delay bound for the FIFO system using Lemma 1 and then we add the transmission time for the packet of interests. To obtain the per-flow delay-bound in item (ii), we take the supremum of per-packet delay-bound in item (i) for all the range of packet lengths, i.e., $[L_1^{\min}, L_1^{\max}]$. When the service curve is $c$-Lipschitz (formally defined in (20)), the supremum is achieved at $l = L_1^{\min}$ (in Appendix D in the Supplementary Material, we show that this does not always hold for a non $c$-Lipschitz service curve). The $c$-Lipschitz condition in Theorem 5 implies that the slope of the service curve is not more than the transmission rate $c$ at any point in time. In fact, the rate-latency service curves as well as the existing service curves for the common scheduling mechanisms in time-sensitive networks, e.g., TSN schedulers with credit-based shapers [20], [25], DRR [21], [26], Interleaved WRR [22], are $c$-Lipschitz; therefore, the bound in Theorem 5 can be applied to such systems.

Theorem 5 gives the same delay bound as [11] for rate-latency service curves. For a number of scheduling mechanisms, e.g., DRR and Interleaved WRR, the rate-latency service curves are not accurate; to this end, better non rate-latency service curves are obtained that provide more precise characterization for these mechanisms [21], [22]. For such non rate-latency service curves, the delay-bound improvement of [11] cannot be used; instead, the existing approach is to use classical network calculus to obtain a delay bound, which is $h(\alpha + \alpha', \beta)$. This bound is improved by Theorem 5. We provide a numerical illustration of Theorem 5 for a non rate-latency service curve in Section VIII-C.

**VII. COMPARISON OF THE DIFFERENT BOUNDS**

In sections V and VI we obtained delay bounds for flows with various regulation types. Moreover, we showed in Section IV that regulation types can be derived from each other. Combining the mentioned results, we can derive a regulation type from another and then apply the corresponding theorem, e.g., deriving bit-level arrival curve from g-regularity using item (1) of Proposition 1 and then apply Theorem 5. Now, the question is how such delay bounds are compared with the ones obtained by directly applying the theorem to initial flow constraint. The goal of this section is address this question.

We start with packet-level arrival curve. Our tight delay bound for packet-level arrival curves in Theorem 1 uses the g-regularity derivation of Proposition 2 for the flow of interests. The other approach is to use bit-level derivation of packet-level arrival curves in Proposition 2; this is indeed the state-of-the-art approach. We already showed in Table II that this leads to sub-optimal delay bounds for rate-latency service curves. The following proposition shows that this is not accidental and is true in general.

**Proposition 3:** Consider the assumptions in Theorem 1. Using Proposition 2 we can derive a bit-level arrival curve and, by applying Theorem 5, obtain a delay bound $\Delta^A_f$ for every flow $f$. Let $\Delta^A_{f,\text{pkt}}$ be the delay bound obtained by a direct application of Theorem 1. Then $\Delta^A_{f,\text{pkt}} \leq \Delta^A_f$. Furthermore, if packets of flow $f$ have a constant size $(L_{f,\text{min}} = L_{f,\text{max}})$ then $\Delta^A_{f,\text{pkt}} = \Delta^A_f$, else $\Delta^A_{f,\text{pkt}} < \Delta^A_f$ in general.

The proof is in Appendix C-A in the Supplementary Material. In the above, “in general” means that we can find schedulers for which the inequality is strict, for example when the service curve is rate-latency with rate $R < c$.

Similarly, we evaluate the delay bounds achieved by deriving g-regularity from bit-level arrival curve [resp. bit-level arrival curve from g-regularity] and then applying Theorem 4 [resp. Theorem 5]. Then, the question is whether we obtain the same delay bounds by applying theorems 4 and 5 respectively.
to the derived g-regularity constraint and bit-level arrival curve. The following proposition shows that the answer to this question is negative and the obtained delay bound is generally worse, except for flows with packets of constant size.

**Proposition 4:** Consider the FIFO system assume in Theorem 5.

1) Assume that the flows have bit-level arrival curves. Then, Theorem 5 gives a better response time bound than the sequential application of item 1 of Proposition 1 and Theorem 4. Specifically: \( \Delta_f^A \leq \Delta_f^G \); furthermore, if packets of flow \( f \) have a constant size \( \Delta_f^A = \Delta_f^G \) and else \( \Delta_f^A < \Delta_f^G \) in general.

2) Assume that the flows have g-regularity constraints. Then, Theorem 4 gives a better response time upper bound than the sequential application of item 2 of Proposition 1 and Theorem 5. Specifically: \( \Delta_f^A \leq \Delta_f^G \); furthermore, if packets of flow \( f \) have a constant size \( \Delta_f^A = \Delta_f^G \) and else \( \Delta_f^G < \Delta_f^G \) in general.

The proof is in Appendix C-B in the Supplementary Material.

Propositions 3 and 4 indicate that the best delay bound for a given flow is obtained by directly applying the theorem corresponding to its initial constraint, as delay bounds obtained from derived flow constraints are pessimistic.

**VIII. NUMERICAL ILLUSTRATION**

This section provides three example applications of the theorems presented in this paper to commonly used schedulers. To compute the delay bounds, we use the RealTime-at-Work (RTaW) tool [33] that has efficient implementation of network calculus operations.

**A. Flows With Packet-Level Arrival Curve**

Consider a FIFO system with TSN scheduler and Credit-based Shapers (CBSs) with per-class FIFO queuing [20]; from highest to lowest priority, the classes are CDT, A, and Best Effort (BE). The CBSs are used separately for classes A and B. The CBS parameters \( \text{ideslopes} \) are set to 50% and 25% of the link rate, \( c = 1 \) Gbps, respectively for classes A and B. The CDT traffic has a token-bucket arrival curve with rate 6.4 Kbps and burst 64 Bytes. The maximum packet length of class BE is 1.5 KB. Using the results in [20], a rate-latency service curve offered to class A has latency \( T_A = 12.5 \mu s \) and rate \( R_A = 499.92 \) Mbps, and the one offered to class B has \( T_B = 36.6 \mu s \) and rate \( R_B = 249.75 \) Mbps.

There are 5 periodic flows for each of classes A and B; each flow send 1 packet at each period. Table III shows the flow information. We want to compute delay bounds for flows 1 and 6 in classes A and B.

The state-of-the-art approach is to obtain bit-level arrival curve for all flows using Proposition 2; then by Theorem 5 we obtain a delay bound of \( 63.58 \mu s \) for flow 1 (class A) and \( 158.47 \mu s \) for flow 6 (class B). However, we can also directly apply Theorem 1 to obtain delay bounds; for flows 1 and 6 we obtain 50.46 \( \mu s \) and 126.32 \( \mu s \) respectively, which shows 20% improvement in both cases.

**B. Flows With g-Regulation**

Consider a FIFO system with aggregate queuing and DRR arbitration policy, with \( n = 8 \) queues sharing a link with rate \( c = 1 \) Gbps. Assume all flows have maximum packet length of \( L_{\text{max}} = 1.5 \) KB and the queues have same quantum value \( Q = L_{\text{max}} \). Then, by [21, Theorem 1], we obtain a service curve offered to any aggregate queue,

\[
\beta(t) = (\theta \circ \zeta)(t) + \min \left(\left[ct - 2(n-1)(2L_{\text{max}} - c)\right]^+, \epsilon\right),
\]

with,

\[
\theta(t) = \inf_{0 \leq s \leq t} \left\{ t - s + Q \left\lfloor \frac{s}{(n-1)Q} \right\rfloor \right\},
\]

\[
\zeta(t) = \left(\frac{ct - (n-1)(4Q - c) + \epsilon}{c}\right)^+,
\]

and we set \( \epsilon = 1 \) Byte. Now, assume a flow of interest, conforms to LRQ with rate \( r \), shares a queue with a number of other flows with LRQ regulation where the sum of their maximum packet lengths is 5 KB. Also assume that the minimum packet length of the flow is 100 Bytes.

We obtain a bit-level arrival curve constraint for the flows using Proposition 1. Then by Theorem 5 we obtain a delay bound of 632.75 \( \mu s \) for the flow of interest. Using the results of this paper, we can also directly apply Theorem 4 to obtain delay bounds that gives 552.74 \( \mu s \), which shows 12.6% improvement.

**C. Flows With Bit-Level Arrival Curve**

Consider a FIFO system with per-flow queuing and DRR arbitration policy, with \( n = 16 \) queues sharing a link with rate \( c = 1 \) Gbps. Assume all flows have maximum packet length of \( L_{\text{max}} = 1.5 \) KB and the queues have same quantum value \( Q = L_{\text{max}} \). Then, similarly to Section VIII-B, we obtain a service curve offered to any per-flow queue.

Consider a flow with \( h(t) = rt + 2L_{\text{max}} \) and minimum packet length of 0.5 KB. Since, the service curve is not rate-latency, we cannot apply the improvement in [11]. Hence, the state-of-the-art approach is to use the classical network-calculus bound, \( h(\alpha, \beta) \); it is equal to 915.88 \( \mu s \). Using our improved delay bound in Theorem 5, the delay bound is reduced to 743.88 \( \mu s \), which improves the classical network-calculus bound by 18.8%.

**IX. CONCLUSION**

We presented a theory to compute delay bounds for flows with packet-level arrival curve, which improves the
state-of-the-art bounds. The improvement is made possible by a novel modelling of packet-level arrival curve with g-regularity and bit-level arrival curve together with exploiting the information on the transmission rate. Our method of proof led to delay improvement for flows with g-regularity constraint and bit-level arrival curve. In time-sensitive networks, this result can open a discussion on the operation of flow re-shaping: even though in such networks the traffic specification at a source is at the packet level, flow re-shaping is performed at the bit level, based on a bit-level arrival curve derived from the flow constraints at the source. As a result of the analysis of this paper, using packet-level traffic re-shaping leads to better delay bounds at the intermediate routers and switches. As the operation of bit-level re-shaping mechanisms is mainly based on full reception of a packet, e.g., as in IEEE802.1 Qcr Asynchronous Traffic Shaping, an implementation of packet-level re-shaping appear to be feasible and even simpler.

**APPENDIX A**

**TECHNICAL PREREQUISITES**

Consider a function \( f \in \mathcal{F}_{\text{inc}} \), then by [30, Section 10.1] [31],

\[
(f^+_f)^\downarrow = f^+_f, \quad (f^+_f)^+ = f^+_f, \quad (f^-_f)^- = f^-_f, \quad (f^-_f)^+ = f^-_f, \quad \forall y \in \mathbb{R}^+ : f(x) \leq y \implies x \leq f^+_f(y), \quad (40)
\]

\[
\forall y \in \mathbb{R}^+ : f(x) \geq y \implies x \geq f^-_f(y), \quad (41)
\]

where \( f^+ \) is the right limit of the function \( f \). Furthermore, by [30, Section 10.1] [34], if \( f \) is right continuous:

\[
f = (f^+_f)^\downarrow, \quad \forall w \in \mathcal{F}_{\text{inc}} : (f \circ w)^\downarrow(x) = (w^\downarrow \circ f^+_f)(x), \quad (43)
\]

and if \( f \) is left continuous:

\[
f = (f^-_f)^\downarrow, \quad \forall w \in \mathcal{F}_{\text{inc}} : (f \circ w)^\downarrow(x) = (w^\downarrow \circ f^-_f)(x), \quad (45)
\]

Note that \( \circ \) is the composition operator, i.e., \( (f \circ w)(x) = f(w(x)) \).

**Lemma 3:** Consider a left-continuous function \( f \in \mathcal{F}_{\text{inc}} \). Then \( (f^+_f)^\downarrow \downarrow = f \).

**Lemma 4:** Let \( f \) be a wide-sense increasing and \( c \)-Lipschitz function. Then, for \( x' \geq x \), we have:

\[
f^+_f(x') - f^+_f(x) \geq \frac{x' - x}{c}, \quad (46)
\]

**Lemma 5:** A flow with fixed interval \((\tau, K)\) conforms to a packet-level arrival curve \( \alpha_{\text{pkt}} \) with \( \alpha_{\text{pkt}}(0) = 0 \), \( \alpha_{\text{pkt}}(t) = K \left[ \frac{t}{\tau} \right] + K : t \geq 0 \).

The proofs of the above lemmas are in Appendix C in the Supplementary Material.

\(^{2}\)Flow re-shaping refers to the process of recreating the arrival curve of a flow as its source.

**APPENDIX B**

**SELECTED PROOFS**

Due to the page limits, some proofs are not included in the main manuscript. They are available in the supplementary material under Appendix C in the Supplementary Material.

**A. Proof of Proposition 1**

1) Since the flow has an arrival curve \( \alpha \), by (5) for any packet index \( m, n \) such that \( m \leq n \), we have:

\[
\sum_{i=m}^{n} l_i \leq \alpha^+(A_n - A_m). \quad (48)
\]

By excluding the last packet from the left hand-side of the inequality, we have:

\[
\sum_{i=m}^{n-1} l_i \leq \alpha^+(A_n - A_m) - l_n \leq \alpha^+(A_n - A_m) - L^\min. \quad (49)
\]

We define \( h(t) = [t - L^\min]^+ \) and \( x = \sum_{i=m}^{n-1} l_i \). Then, (49) can be rewritten as:

\[
x \leq (h \circ \alpha^+)(A_n - A_m). \quad (50)
\]

Applying (41) to (50):

\[
A_n - A_m \geq (h \circ \alpha^+)^\downarrow(x). \quad (51)
\]

Since \( h \) is continuous, by (43) we have:

\[
A_n - A_m \geq ((\alpha^+)^\downarrow \circ h^\downarrow)(x). \quad (52)
\]

Applying (36), we have:

\[
A_n - A_m \geq (\alpha^+ \circ h^\downarrow)(x). \quad (53)
\]

Finally, by (2), we have \( h^\downarrow(x) = x + L^\min \), \( x > 0 \); \( h^\downarrow(0) = 0 \). Then, for \( x > 0 \):

\[
A_n - A_m \geq \alpha^+(x + L^\min) = g(x), \quad (54)
\]

and for \( x = 0 \):

\[
A_n - A_m \geq \alpha^+(0) = 0 = g(0). \quad (55)
\]

2) By [18], Lemma 6.2.8], \( g^\downarrow(t) + L^\max \) is an arrival curve for the flow. Since, the left limit of an arrival curve is also an arrival curve for the flow, we have:

\[
\alpha(t) = \lim_{\varepsilon \to 0} g^\downarrow(t - \varepsilon) + L^\max = (g^\downarrow)^\downarrow(t) + L^\max = g^\downarrow(t) + L^\max. \quad (56)
\]

The last equality is by (38).

3) Assume a flow with arrival curve \( \alpha \) and let us first apply item 1 and then item 2. By item 1, we can find a \( g \)-regularity constraint:

\[
g(x) = \alpha^+(x + L^\min) = (\alpha \circ h)(x), \quad (57)
\]

where \( h(x) = x + L^\min \). Now, we apply item 2 to the obtained \( g \)-regularity constraint in (57) and derive an arrival curve \( \alpha' \):

\[
\alpha'(t) = g^\downarrow(t) + L^\max = ((\alpha \circ h)^\downarrow)^\downarrow(t) + L^\max = ((\alpha \circ h)^\downarrow)(t) + L^\max, \quad (58)
\]
According to (5), the flow conforms to an arrival curve that
for all packet indices
has the same length). We have

\[
\alpha'(t) = \left( \left(h^+ \circ (\alpha^+)^\dagger \right)^\dagger \right)(t) + L_{\text{max}}
\]

where \( h(t) = t + L_{\text{max}} \). Now, we apply item 1 to the obtained arrival curve (61) and derive a \( g^\dagger \)-regularity constraint:

\[
g'(x) = \alpha^+ (x + L_{\text{min}}) = (h \circ g^\dagger)(x + L_{\text{min}})
\]

By (45) and since \( h^\dagger(x) = [x - L_{\text{max}}]^+ \), we have:

\[
g'(x) = \left( (g^\dagger)^+ \circ h^\dagger \right)(x + L_{\text{min}})
\]

If \( g \) is left continuous, by Lemma 3, we have:

\[
g'(x) = g\left((x - L_{\text{max}})^+ + L_{\text{min}}\right).
\]

The above equation shows that applying item 2 and then item 1 does not give the same \( g \)-regularity as the initial one, i.e., \( g \neq g' \), except from the case that \( L_{\text{max}} = L_{\text{min}} \) (when all packets have the same length).

**B. Proof of Proposition 2**

1) According to the min-plus representation of packet-level arrival curve in Eq. (11), for any packets \( m, n \) with \( m \leq n \), we have:

\[
n - m + 1 \leq \alpha_{\text{pkt}}(E_n - E_m).
\]

Now let us multiply both sides of the inequality by \( L_{\text{max}} \):

\[
(n - m + 1)L_{\text{max}} \leq L_{\text{max}}\alpha_{\text{pkt}}(E_n - E_m).
\]

For all packet indices \( i \), it holds that \( l_i \leq L_{\text{max}} \). Thus,

\[
\sum_{i=m}^{n} l_i \leq L_{\text{max}}\alpha_{\text{pkt}}(E_n - E_m).
\]

According to (5), the flow conforms to an arrival curve \( \alpha = L_{\text{max}}\alpha_{\text{pkt}} \).

To obtain \( g \)-regularity, we use Eq. (66) as well as the fact that \( l_i \leq L_{\text{max}} \) for all packets \( i \), and we have:

\[
\sum_{i=m}^{n-1} l_i + L_{\text{max}} \leq L_{\text{max}}\alpha_{\text{pkt}}(E_n - E_m).
\]

Now, we divide the both sides by \( L_{\text{max}} \) and set \( x := \sum_{i=m}^{n-1} l_i \). Then,

\[
\frac{x}{L_{\text{max}}} + 1 \leq \alpha_{\text{pkt}}(E_n - E_m).
\]

Using (41) and then (36):

\[
E_n - E_m \geq \alpha_{\text{pkt}}\left(\frac{x}{L_{\text{max}}} + 1\right) = g(x).
\]

2) First we show that we can derive (24) using (23). By item 1, the flow with packet-level arrival curve \( \alpha_{\text{pkt}} \) conforms to \( g \)-regularity with \( g = \alpha_{\text{pkt}} \circ f \), where \( f(x) = \max\left(\frac{x}{L_{\text{max}}} + 1\right) \). The flow also conforms to a bit-level arrival curve \( \alpha' \) that is derived by applying item 2 of Proposition 1 to (23), i.e., \( \alpha'(t) = g'(t) + L_{\text{max}} \); then,

\[
\alpha'(t) = \left[(\alpha_{\text{pkt}} \circ f)^+\right]^\dagger(t) + L_{\text{max}}
\]

where the last equality is obtained by (38). Now, using (45) and left continuity of \( \alpha_{\text{pkt}} \):

\[
\alpha'(t) = \left(f^\dagger \circ (\alpha_{\text{pkt}}^\dagger)^\dagger\right)^\dagger(t) + L_{\text{max}}
\]

Since \( f^\dagger(t) = [L_{\text{max}} - L_{\text{max}}]^+ \), we have:

\[
\alpha'(t) = \left[L_{\text{max}}(\alpha_{\text{pkt}}^\dagger)^\dagger\right]^\dagger(t) - L_{\text{max}} + L_{\text{max}}
\]

Since \( \alpha_{\text{pkt}} \) is left continuous, by Lemma 3:

\[
\alpha'(t) = L_{\text{max}}\alpha_{\text{pkt}}(t) - L_{\text{max}} + L_{\text{max}}.
\]

Note that \( \alpha_{\text{pkt}}^\dagger\geq 1 \). Thus \( \alpha'(t) = \alpha(t) \) given by (24).

Second we show that (24) does not give (23). By item 1, the flow with packet-level arrival curve \( \alpha_{\text{pkt}} \) conforms to bit-level arrival curve with \( \alpha = f' \circ \alpha_{\text{pkt}} \), where \( f'(x) = L_{\text{max}}^\dagger \).

In addition, the flow also conforms to a \( g' \)-regularity constraint that derives by applying item of Proposition 1 to (24):

\[
g'(x) = \alpha^+ (x + L_{\text{min}}) = (f' \circ \alpha_{\text{pkt}})^+ (x + L_{\text{min}})
\]

where the last equality is obtained by using (43) and continuity of \( f' \). Now, since \( f'(x) = \frac{x}{L_{\text{max}}} \), we have:

\[
g'(x) = \alpha_{\text{pkt}}^\dagger \left(\frac{x + L_{\text{min}}}{L_{\text{max}}} \right).
\]

When all packets are of the same size \( L_{\text{max}} = L_{\text{min}} \), \( g' = g \).

When \( L_{\text{min}} < L_{\text{max}} \), since \( \alpha_{\text{pkt}}^\dagger \) is a wide-sense increasing function, we have \( g' \geq g \) and \( g' \neq g \), i.e. \( g' \) is a weaker constraint than \( g \) given in (23).

**C. Proof of Lemma 1**

Let \( n \) be the index of the packet of interest with length \( l_n \).

Using Lemma 1 of [11], there exists an \( m \leq n \) such that:

\[
\beta(Q_n - A_m) \leq \sum_{k=m}^{n-1} l_k.
\]

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
where $Q_n$ is the beginning of transmission of packet $n$. Using (40), Eq. (77) gives:

$$Q_n - A_m \leq \beta^1 \left( \sum_{k=m}^{n-1} l_k \right).$$ (78)

Therefore $Q_n$ satisfies,

$$Q_n \leq \max_{m\leq n} \left\{ A_m + \beta^1 \left( \sum_{k=m}^{n-1} l_k \right) \right\}. (79)$$

Since $\sum_{k=m}^{n-1} l_k \leq w(A_n - A_m)$, we have:

$$Q_n \leq \max_{m\leq n} \left\{ A_m + \beta^1 \left( w(A_n - A_m) \right) \right\}. (80)$$

By defining $t \triangleq A_n - A_m \geq 0$, we further obtain,

$$Q_n - A_n \leq \sup_{t \geq 0} \{ -t + \beta^1 (w(t)) \}. (81)$$

Applying (37) to (81):

$$Q_n - A_n \leq \sup_{t \geq 0} \{ (\beta^1)^+ (w(t)) - t \}. (82)$$

Next, we use Lemma 2 of [11], which implies that for, $f \in F_{inc}^0$,

$$\sup_{t \geq 0} \{ f^+(t) - Rt \} = \sup_{t \geq 0} \{ f(t) - Rt \}.$$ (91)

We apply this result to (82) by setting $\beta^1 = f$, and therefore:

$$Q_n - A_n \leq \sup_{t \geq 0} \{ \beta^1 (w(t)) - t \} = h(w, \beta), (83)$$

which completes the proof.

**D. Proof of Lemma 2**

(i) Let us remind that $l_k$ and $A_k$ are the length and the arrival time of the $k^{th}$ packet with $k = 1, 2 \ldots$ (Section III). Let $n$ be the index of the packet of interest belonging to flow 1 with length $l_n$, $l_n = l$. The sum of all packets can be split in two parts, one with packets belonging to flow 1 and the one with packets belonging to flow 2. Let $F(k)$ be the flow id of $k^{th}$ packet, then for any $0 \leq m \leq n$, we have:

$$\sum_{k=m}^{n-1} l_k = \sum_{k=m}^{n-1} 1_{\{F(k) = 1\}} l_k + \sum_{k=m}^{n-1} 1_{\{F(k) = 2\}} l_k. (84)$$

For flow 1 with g-regularity constraint, by applying (40) to (7), we have:

$$\sum_{k=m}^{n-1} 1_{\{F(k) = 1\}} l_k \leq g^1 (A_n - A_m), (85)$$

For flow 2 with bit-level arrival curve, using (5) we have:

$$\sum_{k=m}^{n} 1_{\{F(k) = 2\}} l_k \leq \alpha^+ (A_n - A_m). (86)$$

Note that since packet $n$ belongs to flow 1, $\sum_{k=m}^{n} 1_{\{F(k) = 2\}} l_k = \sum_{k=m}^{n-1} 1_{\{F(k) = 2\}} l_k$. Therefore, together with (85) and (86):

$$\sum_{k=m}^{n-1} 1_{\{F(k) = 1\}} l_k + \sum_{k=m}^{n-1} 1_{\{F(k) = 2\}} l_k \leq g^1 (A_n - A_m) + \alpha^+ (A_n - A_m). (87)$$

Using (87) in (84):

$$\sum_{k=m}^{n-1} l_k \leq w(A_n - A_m), (88)$$

where the function $w : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ is $w = g^1 + \alpha^+$. Then by Lemma 1 for packet $n$, we have:

$$Q_n - A_n \leq h(w, \beta), (89)$$

where $Q_n$ is start of transmission of packet $n$. Since the transmission time for the packet of interest, $n$, is $D_n - Q_n = \tau$, the delay bound is:

$$D_n - A_n = D_n - Q_n + Q_n - A_n \leq h(w, \beta) + \frac{l}{c}, (90)$$

which concludes the proof for item 1. Now, since $l \leq L_1^{\text{max}}$, $h(w, \beta) + \frac{L_1^{\text{max}}}{c}$ is a delay bound for flow 1 which completes the proof.

**E. Proof of Theorem 2**

Let us first define the function $w$ as:

$$w(t) = \sum_{u=1}^{M} L_{u}^{\max} \alpha_{\text{pkt},u}(t), (91)$$

The proof is in two steps: first, we construct a simulation trace; second, we verify its properties.

**Step 1.** We start the construction of a simulation trace.

(a) We determine the smallest time instant $t'$ with the following property:

$$\beta^1 (w(t') - L_1^{\text{max}}) - t' = \sup_{t \geq 0} \{ \beta^1 (w(t) - L_1^{\text{max}}) - t \}$$

$$= h(w - L_1^{\text{max}}, \beta). (92)$$

In fact, the time $t'$ will be the time at which the packet of interest arrives at the system and experiences the worst-case delay.

(b) Now, we generate the packet sequence for the $M$ flows. Flow 1 has $n_1 = \alpha_{\text{pkt},1}(t')$ packets presented as a pair $(A_1^1, L_1^1)$ where $A_1^1 = (A_1^1, A_1^1, \ldots, A_1^1)$ is the sequence of packet arrival times and $L_1^1 = (l_1^1, l_2^1, \ldots, l_n^1)$ is the packet length sequence, defined for all $i \in \{1, \ldots, n_1\}$ by:

$$l_i^1 = L_1^{\text{max}},$$

$$A_1^1 = \gamma + \inf\{ s \geq 0 \mid \alpha_{\text{pkt},1}(s) \geq i \} = \gamma + \alpha_{\text{pkt},1}(i), (93)$$

where $\gamma = t' - \alpha_{\text{pkt},1}(n_1)$. Lemma 6 shows that $\alpha_{\text{pkt},1}(n_1) \leq t'$ and hence $\gamma \geq 0$. The aforementioned packet sequence indicates that the packets have maximum length and the packet arrival is greedy, starting at time $\gamma$, and the last packet (packet of interest) arrives at $A_{n_1}^1 = t'$. 


The delay of the packet with length $L_i^{\text{max}}$ that arrives at time $t'$ is $\Delta^{\text{pkt}}$. Any other flow $f$, $f \neq 1$, has $n_f = \alpha_{\text{pkt}, f} f'(t')$ packets presented as a pair $(A^f, \mathcal{L}^f)$ where $A^f = (A^f_1, A^f_2, \ldots, A^f_{n_f})$ is the packet arrival sequence and $\mathcal{L}^f = (l^f_1, l^f_2, \ldots, l^f_{n_f})$ is the packet length sequence, that are defined for all $j \in \{1, \ldots, n_f\}$ as:

$$
A^f_j = \inf\{s \geq 0 : |\alpha_{\text{pkt}, f}(s) \geq j\} = \alpha_{\text{pkt}, f}^{-1}(j).
$$

The aforementioned packet sequence indicate that the packets have maximum length and their arrival is greedy starting at time $A^f_1 = \alpha_{\text{pkt}, f}^{-1}(1) = 0$. With the above arrival construction, we have the cumulative input packet-count function as:

$$
N_u(t) = \begin{cases} 
\sum_{i=1}^{n_u} A^i_u < t & t \leq t', \\
0 & \forall u = 1, 2, \ldots, M.
\end{cases}
$$

Now let us merge the two packet sequences to express the total traffic. We define the pair $(A, \mathcal{L})$ with arrival sequence $A = (A_1, A_2, \ldots, A_n)$ and lengths sequence $\mathcal{L} = (l_1, l_2, \ldots, l_n)$ of total packets and $n = \sum_{k=1}^{M} n_k$, i.e., $(A, \mathcal{L}) = \bigcup_{k=1}^{M} (A^k, \mathcal{L}^k)$. Lemma 6 indicates that for any $f \neq 1$:

$$
A^f_{n_f} = t' = A^1_{n_1}.
$$

In the case $A^f_{n_f} = A^1_{n_1} = t'$, assume that the last packet of flow 1 is enqueued after the last packet of other flows; hence, $A^1_{n_1} = t'$ and $l_n = L^{\text{max}}$. The cumulative input function, $I(t)$ is shown as the green line in Fig. 6 that is obtained as:

$$
I(t) = \sum_{u=1}^{M} L_u^{\text{max}} N_u(t) = \begin{cases} 
\sum_{k=1}^{\infty} L_k^{\text{max}} I(A_k < t) & t \leq t', \\
\sum_{u=1}^{M} n_u L_u^{\text{max}} w^+(t) & t > t'.
\end{cases}
$$

(c) For the output, we first construct the fluid output curve $F(t)$ (orange dotted-line in Fig. 6) given by

$$
F(t) = \inf_{0 \leq s \leq t} \{I(s) + \beta(t - s)\},
$$

so that the service curve property would be automatically satisfied if we would let the output cumulative function be $O(t) = F(t)$. However, we cannot take $O(t) = F(t)$ because $F(t)$ does not satisfy the condition that packet transmission is at rate $c$. In order to obtain the output function $O(t)$, we first observe the start and end of transmission time of a packet $l_i$, i.e., $Q_i$ and $D'_i$ as:

$$
D'_i = \inf\{s \geq 0 : F(s) \geq I^+(A_i)\} = F^1 (I^+(A_i)),
$$

$$
Q_i = \inf\{s \geq 0 : F(s) \geq I^+(A_i) - l_i\} = F^1 (I^+(A_i) - l_i).
$$

In the above definitions, we have $D'_i = Q_{i+1}$ as $I^+(A_{i+1}) = I^+(A_i) + l_i$ by definition of $I$. For the output function $O(t)$, we keep the same time $Q_i$ for the start of transmission of packet $i$, but, we let the transmission finish at time $D_i = Q_i + \frac{l_i}{c}$. Observe that:

$$
D'_i - D_i = F^1 (I^+(A_i)) - \left( F^1 (I^+(A_i) - l_i) + \frac{l_i}{c} \right).
$$

By Lemma 7, $F(t)$ is $c$-Lipschitz. Then using Lemma 4,

$$
D'_i - D_i \geq \frac{l_i}{c} - \frac{l_i}{c} = 0.
$$

Then, more precisely precisely $\forall i = 1, \ldots, n_1 + n_2$ and $\forall t \in [Q_i, D'_i]$, $O(t)$ is:

$$
O(t) = \begin{cases} 
L(t - Q_i) + F(Q_i) & Q_i \leq t < D_i, \\
F(Q_i) + l_i & D_i \leq t \leq D'_i.
\end{cases}
$$

$O(t)$ is shown with red line in Fig. 6.

Step 2. We verify that all requirements in the theorem are satisfied. First we show that the service curve property holds; to do this, since $F$ satisfies the service curve property, it is sufficient to show that $O(t) \geq F(t)$ : $\forall t \in [Q_i, D_i]$ for any $i = 1, \ldots, n$. Using Lipschitz continuity of $F$ for $t \in [Q_i, D_i]$, $F(t) \leq F(Q_i) + c(t - Q_i) = O(t)$.

For $t \in [D_i, D'_i]$, by construction:

$$
F(t) \leq F(Q_i) + l_i = O(t).
$$

The above equations imply $O(t) \geq F(t), \forall t \geq 0$; therefore, the service curve property is satisfied when the output is $O(t)$. Moreover, by construction the system is FIFO, the input is packetized and packet transmission occurs at rate $c$. We need to prove that the input conforms to the packet-level arrival curve. For any flow $f$, $f \neq 1$, consider two time instants $s, t \geq 0, s \leq t$.

If $t = s = 0$, then $N_f(t) - N_f(s) = \alpha_{\text{pkt}, f}(t - s) = 0$.

If $0 < t \leq s$, there exist a packet index $m'$ where $t \in (A_{m'}, A_{m'+1})$. Then by Lemma 8 with $k = 0$, we have $m' = \alpha_{\text{pkt}, f}(t)$.

Now if $s = 0$, we have $N_f(t) - N_f(s) = N_f(A_{m'+1}) - N_f(0) = m' = \alpha_{\text{pkt}, f}(t)$; otherwise, there exists a packet index $m (m \leq m')$, where $s \in (A_m, A_{m+1})$. Then:

$$
N_f(t) - N_f(s) = N_f(A_{m+1}) - N_f(A_m) = m' - m.
$$

By Lemma 8 with $k = 0$, $m' = \alpha_{\text{pkt}, f}(t)$ and $m = \alpha_{\text{pkt}, f}(s)$.

Therefore:

$$
N_f(t) - N_f(s) = \alpha_{\text{pkt}, f}(t) - \alpha_{\text{pkt}, f}(s) \leq \alpha_{\text{pkt}, f}(t - s),
$$

where the last inequality is due to sub-additivity of $\alpha_{\text{pkt}, f}$. 


If \( t > t' \), by construction, \( N_f(t) = n_f = \alpha^+_{\text{pkt}, f}(t') \). Now, for \( s = 0 \), we have
\[
N_f(t) - N_f(s) = n_f = \alpha^+_{\text{pkt}, f}(t' + t') - \alpha^+_{\text{pkt}, f}(t'); \quad \forall t > t'.
\]
For \( 0 < s \leq t' \), there exists a packet index \( m (m \leq m') \), where \( s \in (A_m, A_{m+1}] \). Then by Lemma 8 with \( k = 0 \), we have \( m = \alpha_{\text{pkt}}(s) \). Therefore,
\[
N_f(t) - N_f(s) = n_f - m = \alpha_{\text{pkt}, f}^+(t') - \alpha_{\text{pkt}, f}(s)
\leq \alpha_{\text{pkt}, f}(t - s); \quad \forall t > t',
\]
where the last inequality is due to sub-additivity of \( \alpha_{\text{pkt}, f} \).
For \( t' < s \leq t \), \( N_f(s) = n_f \). Then,
\[
N_f(t) - N_f(s) = n_f - n_f = 0 \leq \alpha_{\text{pkt}, f}(t - s),
\]
which shows flow \( f \) conforms to the packet-level arrival curve.

For flow 1, similarly to the above computation and using Lemma 8 with \( k = \gamma \), we obtain
\[
N_1(t) - N_1(s) \leq \alpha_{\text{pkt}}(t - s); \quad \forall s \leq t, \quad \forall t \geq 0,
\]
which shows flow 1 conforms to packet-level arrival curve. Lastly, we show that packet \( n \) achieves the delay bound. We have for packet \( n \), \( A_n = t' \); then
\[
Q_n = F^1((t' - l_n) = F^1(w^+(t') - L_1^{\max}).
\]
Furthermore, \( D_n = Q_n + \frac{w^+}{c} = Q_n + \frac{L_1^{\max}}{c} \), therefore,
\[
D_n - A_n = F^1(w^+(t') - L_1^{\max}) - t' + \frac{L_1^{\max}}{c}.
\]
By definition of \( F \) in (96), we have \( F \leq \beta \). Then by [30, Lemma 10.1], \( F^1 \geq \beta^1 \). Hence:
\[
D_n - A_n \geq \beta^1(w^+(t') - L_1^{\max}) - t' + \frac{L_1^{\max}}{c} + \frac{L_1^{\max}}{c} = \Delta_{\text{pkt}}.
\]
By Theorem 1, we have \( D_n - A_n \leq \Delta_{\text{pkt}} \); then together with the above equation, we have \( D_n - A_n = \Delta_{\text{pkt}} \).

\textbf{Lemma 6: Consider a wide-sense increasing function} \( f : \mathbb{R}^+ \rightarrow \mathbb{Z}^+ \), a positive real value \( t_0 \), a positive integer \( n = f^+(t_0) \), and a value \( x = f^1(n) \); then we have \( x \leq t_0 \).

The proof is in Appendix C-F in the Supplementary Material.

\textbf{Lemma 7: Consider two functions} \( f, g \) where \( f \) is \( L \)-Lipschitz and \( g \geq 0 \). Then \( z(t) = \inf_{0 \leq s \leq t} \{g(s) + f(t - s)\} \) is \( L \)-Lipschitz.

The proof is in Appendix C-G in the Supplementary Material.

\textbf{Lemma 8: Consider a left-continuous function} \( \alpha : \mathbb{R}^+ \rightarrow \mathbb{N} \) and a constant \( k \geq 0 \). Let \( A = (A_1, A_2, \ldots) \) be a sequence where \( A_i = k + \alpha^i(i) \). Assume a positive integer \( m \) and a time instant \( s \), such that \( s \in (A_m, A_{m+1}] \); then we have \( m = \alpha(t - k) \).

The proof is in Appendix C-H in the Supplementary Material.

\textbf{F. Proof of Theorem 3}

As discussed in Section III-E, the flows can have sliding-interval and fixed-interval regulation constraints; let \( F \) and \( S \) be respectively the sets of flows with fixed interval and sliding interval. By (13), if a flow \( i \) has the sliding interval \((\tau_i, K_i)\) constraint, this is equivalent to having the following packet-level arrival curve:
\[
\alpha_{\text{pkt}, i}(0) = 0, \quad \alpha_{\text{pkt}, i}(t) = K_i \left\lfloor \frac{t}{\tau_i} \right\rfloor, \quad t > 0.
\]
By (16), if a flow \( i \) has the fixed interval \((\tau_i, K_i)\), the constraint implies the following packet-level arrival curve:
\[
\alpha_{\text{pkt}, i}(0) = 0, \quad \alpha_{\text{pkt}, i}(t) = K_i \left\lfloor \frac{t}{\tau_i} \right\rfloor + K_i, \quad t > 0.
\]
Then the delay bound obtain by Theorem 1 for flow 1 is:
\[
\Delta_{\text{pkt}}^{\max} = h \left( \sum_{u=1}^{N} L_u^{\max} K_u \left\lfloor \frac{t}{\tau_u} \right\rfloor + \sum_{u=1}^{N} L_u^{\max} K_u 1_{\{u \in F\}} - L_1^{\max}, \beta \right) + \frac{L_1^{\max}}{c}.
\]

We want to construct a simulation trace where a packet of flow 1 experiences a delay arbitrarily close to \( D \). To this end, for flows in \( S \), we use (104) to construct the input packet sequence as it is equal to the sliding interval interpretation. For any other flow \( j \) with fixed interval \((\tau_j, K_j)\), we use the following packet-level arrival curve:
\[
\alpha_{\text{pkt}, j}(0) = 0, \quad \alpha_{\text{pkt}, j}(t) = K_j \left\lfloor \frac{t-\varepsilon}{\tau_j} \right\rfloor + K_j, \quad t > 0, \quad \forall \varepsilon \in (0, \min_u \{\tau_u\}).
\]

Lemma 9 shows that a greedy packet-sequence of \( \alpha_{\text{pkt}}^j \), starting at \( t_0 = \max_u \{\tau_u\} - \varepsilon \), conforms to Fixed interval \((\tau_j, K_j)\).

Now, let:
\[
w_{\varepsilon}(t) = \sum_{u=1}^{N} L_u^{\max} \alpha_{\text{pkt}}^u(t) 1_{\{u \in S\}} + \sum_{u=1}^{N} L_u^{\max} \alpha_{\text{pkt}, u}(t) 1_{\{u \in F\}}
\]
\[
= \sum_{u=1}^{N} L_u^{\max} K_u \left\lfloor \frac{t}{\tau_u} \right\rfloor 1_{\{u \in S\}}
\]
\[
+ \sum_{u=1}^{N} L_u^{\max} K_u \left\lfloor \frac{t-\varepsilon}{\tau_u} \right\rfloor 1_{\{u \in F\}} + \sum_{u=1}^{N} K_u 1_{\{u \in F\}}.
\]

The tightness scenario follows the same as Theorem 2 by generating a greedy packet-sequence for every flow using the packet-level arrival curves in (107) for fixed-interval flows and (104) for sliding-interval flows. To create a feasible greedy packet-sequence, we shift the start of the simulation by \( t_0 \) (mentioned earlier, this guarantees the existence of greedy packet-sequence for the fixed-interval flows). Therefore, a packet of a flow of interest, i.e., flow 1, experiences a delay of \( d_{\varepsilon} = h(w_{\varepsilon} - L_1^{\max}, \beta) + \frac{L_1^{\max}}{c} \) as shown in (103). By definition of \( w_{\varepsilon} \) in (108), \( d_{\varepsilon} \) is dependent on \( \varepsilon \). We show next that \( d_{\varepsilon} \geq \Delta_{\text{pkt}} - \varepsilon \).
To this end, let us define the following auxiliary functions:

\[ f(t) = \sum_{u=1}^{N} L_u^{\max} K_u \left[ \frac{t}{t_u} \right] 1_{\{u \in F\}}, \quad t > 0; \quad f(t) = 0, \quad t \leq 0, \]

\[ g(t) = \sum_{u=1}^{N} L_u^{\max} K_u \left[ \frac{t}{t_u} \right] 1_{\{u \in S\}} + \sum_{u=1}^{N} K_u 1_{\{u \in F\}} - L_1^{\max}. \]

Then we have

\[ w_{\epsilon}(t) = f(t) + g(t) + L_1^{\max} \Delta_{\text{pkt}} = h(f + g, \beta) + \frac{L_1^{\max}}{c}. \]

By Lemma 10, \( h(w_{\epsilon} - L_1^{\max}, \beta) \geq h(f + g, \beta) - \epsilon \); therefore,

\[ d_{\epsilon} \geq h(f + g, \beta) + \frac{L_1^{\max}}{c} - \epsilon = \Delta_{\text{pkt}} - \epsilon. \]

Moreover, by Theorem 1, \( d_{\epsilon} \leq \Delta_{\text{pkt}} \); hence,

\[ \Delta_{\text{pkt}} - \epsilon \leq d_{\epsilon} \leq \Delta_{\text{pkt}}. \]

Finally, when \( \epsilon \to 0 \), \( d_{\epsilon} \) gets arbitrary close to \( \Delta_{\text{pkt}} \).

**Lemma 9:** Consider the following packet-level arrival curve:

\[ \alpha_{\text{pkt}}(0) = 0, \]

\[ \alpha_{\text{pkt}}(t) = K \left[ \frac{t - \epsilon}{\tau} \right] + K : t > 0, \quad \forall \epsilon \in (0, \tau). \]

Then, a greedy packet-sequence of \( \alpha_{\text{pkt}} \) at any time \( t_0 \geq \tau - \epsilon \), conforms to the fixed-interval \((\tau, K)\) in (15).

**Proof:** Consider the cumulative packet function \( N \) defined as:

\[ N(0) = N(t_0) = 0, \quad N(t) = \alpha_{\text{pkt}}(t - t_0), \quad t > t_0. \]

By definition, the above function is greedy at time \( t_0 \). Now, in (15), let \( \theta = t_0 - \tau + \epsilon \). Since \( t_0 \geq \tau - \epsilon \), we have \( 0 \leq \theta < t_0 \); therefore, \( N(\theta) = 0 \). Next, we show that the greedy packet-sequence conforms to the fixed-interval constraint for any time \( \geq \theta \). For \( i = 0 \) in (15):

\[ N(\theta + \tau) - N(\theta) = N(t_0 + \epsilon) - 0 = \alpha_{\text{pkt}}(\epsilon) = K, \]

and for \( i \geq 1 \):

\[
N(\theta + (i + 1)\tau) - N(\theta + i\tau)
= N(t_0 + \epsilon + i\tau)
- N(t_0 + \epsilon + (i - 1)\tau) = \alpha_{\text{pkt}}(\epsilon + i\tau)
- \alpha_{\text{pkt}}(\epsilon + (i - 1)\tau)
= (iK + K) - ((i - 1)K + K) = K. \]

Therefore, \( \forall i \in \mathbb{N} \):

\[ N(\theta) = 0, \quad N(\theta + (i + 1)\tau) - N(\theta + i\tau) \leq K, \]

which shows there exists \( \theta = t_0 - \tau + \epsilon \) where the cumulative function \( N \), as a greedy packet-sequence of \( \alpha_{\text{pkt}} \) at \( t_0 \), conforms to the fixed interval \((\tau, K)\) constraint.

---

**Lemma 10:** Consider the functions \( f, g, \beta \in \mathcal{F}_{\text{inc}} \). Let

\[ f_{\epsilon}(t) = \begin{cases} f(t - \epsilon) & t \geq \epsilon, \\ 0 & t < \epsilon. \end{cases} \]

Then \( h(f_{\epsilon} + g, \beta) \geq h(f + g, \beta) - \epsilon \).

The proof is in Appendix C-I in the Supplementary Material.

**G. Proof of Theorem 5**

(i) Let us remind that \( l_k \) and \( A_k \) are the length and the arrival time of the \( k^{th} \) packet with \( k = 1, 2, \ldots \) (Section III). Let \( n \) be the index of the packet of interest belonging to flow 1 with length \( l_n \), \( l_n = l \). The sum of all packets can be split in two parts, one with packets belonging to flow 1 and the one with packets belonging flow 2. Let \( \hat{F}(k) \) be the flow id of \( k^{th} \) packet, then:

\[ \sum_{k=m}^{n-1} l_k = \sum_{k=m}^{n-1} 1_{\{F(k)=1\}} l_k + \sum_{k=m}^{n-1} 1_{\{F(k)=2\}} l_k. \]

The flow of interest 1 has a bit-level arrival curve; using its min-plus representation (5), for any \( 0 \leq m \leq n \), we can write:

\[ \sum_{k=m}^{n} 1_{\{F(k)=1\}} l_k \leq \alpha_{+}(A_n - A_m). \]

By excluding the last packet from the left-hand-side of (119) (note that \( l_n = l \)), we obtain:

\[ \sum_{k=m}^{n-1} 1_{\{F(k)=1\}} l_k \leq \alpha_{+}(A_n - A_m) - l. \]

Similarly, flow 2 has bit-level arrival curve; thus, using (5) for any \( 0 \leq m \leq n \), we can write:

\[ \sum_{k=m}^{n-1} 1_{\{F(k)=2\}} l_k \leq \alpha_{+}(A_n - A_m). \]

Note that since packet \( n \) belongs to flow 1, the above equation conforms the min-plus representation of bit-level arrival curve in (5) for flow 2. Now, we sum up (120) and (121):

\[ \sum_{k=m}^{n-1} 1_{\{F(k)=1\}} l_k + \sum_{k=m}^{n-1} 1_{\{F(k)=2\}} l_k \]

\[ \leq \alpha_{+}(A_n - A_m) + \alpha_{+}(A_n - A_m) - l. \]

Using (122) in (118):

\[ \sum_{k=m}^{n-1} l_k \leq w(A_n - A_m), \]

where the function \( w : \mathbb{R}^+ \to \mathbb{R}^+ \) is \( w = \alpha_{+} + \alpha_{2^+} - l \). Then by Lemma 1 for packet \( n \), we have:

\[ Q_n - A_n \leq h(w, \beta), \]

where \( Q_n \) is start of transmission of packet \( n \). Since the transmission time for the packet of interest, \( n \), is \( D_n - Q_n = \frac{1}{c} \), the delay bound is:

\[ D_n - A_n = D_n - Q_n + Q_n - A_n \leq h(w, \beta) + \frac{1}{c}, \]

which completes the proof.
(ii) We want to compute
\[
\sup_{l \leq L^1_{\min}, L^1_{\max}} \{ \Delta^b(t) \} 
= \sup_l \left[ \beta^b \left( \alpha^b + \alpha^b - l - t \right) - \frac{l}{c} \right] 
= \sup_l \left[ \beta^b \left( \alpha^b + \alpha^b - l \right) - \frac{l}{c} \right]. \tag{126} 
\]

Since \( \beta \) is \( \epsilon \)-Lipschitz, by Lemma 4, it satisfies:
\[
\beta^b \left( \alpha^b + \alpha^b - l \right) - \beta^b \left( \alpha^b + \alpha^b - l \right) \geq \frac{1}{c} \left( l - L^1_{\min} \right). \tag{127} 
\]
This gives,
\[
\beta^b \left( \alpha^b + \alpha^b - l \right) \leq \beta^b \left( \alpha^b + \alpha^b - L^1_{\min} \right) - \frac{1}{c} \left( l - L^1_{\min} \right). \tag{128} 
\]
By using the last relation in (126), we obtain,
\[
\sup_{l \leq L^1_{\min}, L^1_{\max}} \{ \Delta^b(t) \} 
= \sup_l \left[ \beta^b \left( \alpha^b + \alpha^b - L^1_{\min} \right) - \frac{l}{c} \right] 
= h \left( \alpha^b + \alpha^b - L^1_{\min}, \beta \right) + \frac{L^1_{\min}}{c}. \tag{129} 
\]
Since \( \beta \) is \( \epsilon \)-Lipschitz continuous and \( \alpha, \alpha' \) are left continuous, by [10, Theorem 5.6], we have:
\[
h \left( \alpha^b + \alpha^b - L^1_{\min}, \beta \right) = h \left( \alpha^b + \alpha^b - L^1_{\min}, \beta \right), \tag{130} 
\]
which together with (129) completes the proof.

REFERENCES

[1] IEEE Standard for Local and Metropolitan Area Networks—Audio Video Bridging (AVB) Systems, IEEE Standard 802.1BA-2011. [Online]. Available: https://ieeexplore.ieee.org/document/6032690

[2] Time-Sensitive Networking Profile for Industrial Automation, Standard IEC/IEEE 60802, (D1.1), 2019. [Online]. Available: http://www.ieee802.org/1/files/private/60802-drafts/d1/60802-d1-1.pdf

[3] IEEE Standard for Local and Metropolitan Area Networks—Time Sensitive Networking Profile for Automotive In-Vehicle Ethernet Communications. IEEE Standard P802.1DG/D1.1, Oct. 2019. [Online]. Available: http://www.ieee802.org/1/files/private/dg-drafts/d1/802-1DG-d1-1.pdf

[4] T. Ferrandiz, F. Frances, and C. Fraboul, “Using network calculus to compute end-to-end delays in SpaceWire networks,” ACM SIGBED Rev., vol. 8, no. 3, pp. 44–47, Sep. 2011, doi: 10.1145/2038617.2038627.

[5] T. Wong, N. Finn, and X. Wang, TSN Profile for Service Provider Networks. Accessed: Apr. 2022. [Online]. Available: https://www.ieee802.org/1/files/public/docs2018/new-tsn-wangtt-TSN-profile-for-service-provider-network-0718.pdf

[6] Representative Use Cases and Key Network Requirements for Network 2030, Standard 3000, 2020. [Online]. Available: https://www.int.intrec/T-REC-Y.Sup67-202007-I

[7] Deterministic Networking (DetNet). Accessed: Apr. 2022. [Online]. Available: https://datatracker.ietf.org/wg/detnet/about/

[8] Time-Sensitive Networking (TSN) Task Group. Accessed: Apr. 2022. [Online]. Available: https://1.ieee802.org/tsn/

[9] J.-Y. Le Boudec and P. Thiran, Network Calculus: A Theory of Deterministic Queuing Systems for the Internet, vol. 2050. Berlin, Germany: Springer, 2001.

[10] A. Bouillard, M. Boyer, and E. Le Corronc, Deterministic Network Calculus: From Theory to Practical Implementation. Hoboken, NJ, USA: Wiley, 2018.

[11] E. Mohammadpour, E. Stai, and J.-Y. Le Boudec, “Improved delay bound for a service curve element with known transmission rate,” IEEE Netw. Lett., vol. 1, no. 4, pp. 156–159, Dec. 2019.

[12] B. Varga, J. Farkas, R. Cummings, Y. Jiang, and D. Fedyk, Flow and Service Information Model for Deterministic Networking (DetNet), Standard 9016, 2021. [Online]. Available: https://www.rfc-editor.org/rfc/rfc9016

[13] IEEE Standard for Local and Metropolitan Area Network—Bridges and Bridged Bridges. IEEE Standard 802.1Q-2018, (Revision of IEEE Standard 802.1Q-2014, pp. 1–1993, 2018.

[14] N. Finn, J.-Y. Le Boudec, E. Mohammadpour, J. Zhang, B. Varga, and J. Farkas. (2021). DetNet Bounded Latency Draft. Accessed: Dec. 15, 2021. [Online]. Available: https://www.ietf.org/archive/id/draft-ietf-detten bounded-latency-07.html

[15] H. Daignerote, M. Boyer, and L. Zhao. (Jun. 2018). Modelling in Network Calculus: A TSN Architecture Mixing Time-Triggered, Credit Based Shaper and Best-Effort Queues. [Online]. Available: https://hal.archives-ouvertes.fr/hal-01814211

[16] L. Maile, K. Hielischer, and R. German, “Network calculus results for TSN: An introduction,” in Proc. Inf. Commun. Technol. Conf. (ICTC), May 2020, pp. 131–140.

[17] K. Eriksson, D. Este, C. Johnson, and J. Hoffman, Applied Mathematics: Body and Soul. Berlin, Germany: Springer, 2004.

[18] C.-S. Chang, Performance Guarantees in Communication Networks. New York, NY, USA: Springer-Verlag, 2000.

[19] J. Specht and S. Samii, “Urgency-based scheduler for time-sensitive switched Ethernet networks,” in Proc. 28th Euromicro Conf. Real-Time Syst. (ECRTS), Jul. 2016, pp. 75–85.

[20] E. Mohammadpour, M. Boyer, and L. Zhao, “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.

[21] S. M. Tabatabaee and J.-Y.-L. Boudec, “Deficit Round-Robin: A second network calculus analysis,” in Proc. IEEE 27th Real-Time Embedded Technol. Appl. Symp. (RTAS), May 2021, pp. 171–183.

[22] S. M. Tabatabaee, J.-Y. Le Boudec, and M. Boyer, “Interleaved weighted round-robin: A network calculus analysis,” JIECE Trans. Commun., vol. 104, no. 12, pp. 1479–1493, 2021.

[23] Y. Jiang, “Some properties of length rate quotient shapers,” 2021, arXiv:2107.05021.

[24] R. L. Cruz, “Quality of service guarantees in virtual circuit switched networks,” IEEE J. Sel. Areas Commun., vol. 13, no. 6, pp. 1048–1056, Aug. 1995.

[25] 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.

[26] M. Boyer, G. Stea, and W. M. Sofack, “Deficit round Robin with network calculus,” in Proc. 6th Int. Conf. Perform. Eval. Methodologies Tools, 2012, pp. 138–147.

[27] A. Finzi, A. M fldauoi, 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.

[28] W. M. Sofack and M. Boyer, “Non preemptive static priority with network calculus: Enhancement,” in Measurement, Modelling, and Evaluation of Computer Systems and Dependability and Fault Tolerance. Berlin, Germany: Springer, 2012, pp. 258–272.

[29] A. Burchard and J. Liebeherr, “A general per-flow service curve for GPS,” in Proc. 30th Int. Teletraffic Congr. (ITC), vol. 2, Sep. 2018, pp. 31–36.

[30] J. Liebeherr, “Duality of the max-plus and min-plus network calculus,” Found. Trends Netw., vol. 11, nos. 3–4, pp. 139–282, 2017.

[31] J. Le Boudec, “A theory of traffic regulators for deterministic networks with application to interleaved regulators,” IEEE/ACM Trans. Netw., vol. 26, no. 6, pp. 2721–2733, Dec. 2018.

[32] Y. Jiang, “A basic result on the superposition of arrival processes in deterministic networks,” in Proc. IEEE Global Commun. Conf. (GLOBECOM), Dec. 2018, pp. 1–6.

[33] RealTime-at-Work Online Min-Plus Interpreter for Network Calculus. Accessed: Apr. 2022. [Online]. Available: https://www.realtimeatwork.com/minplus-playground, accessed: year-month-day

[34] M. Boyer and P. Roux. (May 2016). A Common Framework Embedding Network Calculus and Event Stream Theory. [Online]. Available: https://hal.archives-ouvertes.fr/hal-01311502

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.