Network-calculus service curves of the interleaved regulator

Ludovic Thomas, Jean-Yves Le Boudec

Network-Calculus Service Curves of the Interleaved Regulator

Ludovic Thomas[a,*], Jean-Yves Le Boudec[b]

[a] *CNRS, LORIA, Campus scientifique, Vandoeuvre-les-Nancy, 54506, Grand-Est, France*
[b] *EPFL, Route Cantonale, Lausanne, 1015, Vaud, Switzerland*

Abstract

The interleaved regulator (implemented by IEEE TSN Asynchronous Traffic Shaping) is used in time-sensitive networks for reshaping the flows with per-flow contracts. When applied to an aggregate of flows that come from a FIFO system, an interleaved regulator that reshapes the flows with their initial contracts does not increase the worst-case delay of the aggregate. This shaping-for-free property supports the computation of end-to-end latency bounds and the validation of the network's timing requirements. A common method to establish the properties of a network element is to obtain a network-calculus service-curve model. The existence of such a model for the interleaved regulator remains an open question. If a service-curve model were found for the interleaved regulator, then the analysis of this mechanism would no longer be limited to the situations where the shaping-for-free holds, which would widen its use in time-sensitive networks. In this paper, we investigate if network-calculus service curves can capture the behavior of the interleaved regulator. For an interleaved regulator that is placed outside of the shaping-for-free requirements (after a non-FIFO system), we develop Spring, an adversarial traffic generation that yields unbounded latencies. Consequently, we prove that no network-calculus service curve exists to explain the interleaved regulator's behavior. It is still possible to find non-trivial service curves for the interleaved regulator. However, their long-term rate cannot be large enough to provide any guarantee. Specifically, we prove that for the regulators that process at least four flows with the same contract, the long-term rate of any service curve is upper bounded by three times the rate of the per-flow contract.

Keywords: Network Calculus, Service Curve, Interleaved Regulator (IR), Time-Sensitive Networking (TSN), Asynchronous Traffic Shaping (ATS)

*Corresponding author
Email addresses: ludovic.thomas@cnrs.fr (Ludovic Thomas), jean-yves.leboudec@epfl.ch (Jean-Yves Le Boudec)

2024-09-05

1. Introduction

Time-sensitive networks, as specified by the time-sensitive networking (TSN) task group of the Institute of Electrical and Electronics Engineers (IEEE), support safety-critical applications in the aerospace, automation, and automotive domains [1]. To do so, time-sensitive networks provide a deterministic service with guaranteed bounded latencies.

These guarantees must be validated through a deterministic worst-case timing analysis that can be performed with network calculus. This mathematical framework obtains worst-case performance bounds by modeling the flows with the concept of arrival curves and the network elements with the concept of service curves. Service curves constrain the minimum amount of service that network elements provide to a flow or aggregate of flows.

Time-sensitive networks can also rely on a set of mechanisms that improve the traditional forwarding process of an output port. The traffic regulators are such hardware elements that support higher scalability and efficiency of time-sensitive networks. Placed after a multiplexing stage, they reshape the flows with per-flow shaping curves (by delaying packets if required) and remove the increase of the flows' burstiness due to their interference with other flows.

Traffic regulators come in two flavors: per-flow regulators (PFRs) and interleaved regulators (IRs). A PFR processes a unique flow. It stores the packets of the flow in a first in, first out (FIFO) queue and releases the head-of-line packet as soon as doing so does not violate the configured shaping curve for the flow. In contrast, the IR processes an aggregate of flows with a unique FIFO queue. Each flow has its own configured shaping curve, but the IR analyses only the head-of-line packet and releases it as soon as doing so does not violate the shaping curve of the associated flow. The packets in the IR queue are blocked by the head-of-line even if they belong to other flows. This second flavor is implemented within IEEE TSN under the name asynchronous traffic shaping (ATS) [2].

In time-sensitive networks that contain traffic regulators, end-to-end latency bounds are obtained from the knowledge of the shaping curves enforced by the regulators and from the essential "shaping-for-free" property. It states that the traffic regulators do not increase the worst-case latency of the flow (or of the flow aggregate) under certain conditions that depend on the type of the regulator (PFR or IR). Most analyses of traffic regulators rely on this property.

For the PFR, the shaping-for-free property is well understood because a PFR with a concave shaping curve can be modeled with a context-agnostic service curve, *i.e.,* a service curve that only depends on the configuration of the PFR but not on the context in which the PFR is placed. This service curve proves the shaping-for-free property when the PFR is placed in the appropriate context. When the PFR's context deviates from the shaping-for-free requirements, the context-agnostic service curve still provides performance bounds for the PFR, and slight deviations of the context lead to bounded delay penalties. On the contrary, the only context-agnostic service curve known for the IR is the trivial function $t \mapsto 0$. The non-trivial service-curve models published in the litera-

ture [3, 4] are context-dependent and always assume that the shaping-for-free property holds. Without a context-agnostic service-curve model, performance bounds cannot be obtained for an IR placed outside the shaping-for-free requirements, which restrains its use in time-sensitive networks.

In this paper, we investigate if the behavior of the IR can be modeled by a context-agnostic network-calculus service curve. Our contributions are:

- As opposed to the shaping-for-free property when the IR is placed after a FIFO system, we prove that the IR can yield unbounded latencies when placed after a non-FIFO system, even if the latter is FIFO-per-flow and lossless. To obtain this result, we develop Spring, an adversarial packet sequence that generates unbounded latencies in an IR placed after a non-FIFO system.
- Based also on Spring, we prove that the shaping-for-free property of the IR cannot be explained by any network-calculus service-curve model.
- For any IR that processes at least four flows, we prove that any context-agnostic service curve for an individual flow is upper bounded by a constant.
- We exhibit a strict service curve of the IR and a function that upper bounds any other context-agnostic strict service curve.
- For any IR that processes at least four flows, we show that the long-term rate of any context-agnostic service curve for the aggregate is upper bounded by three times the rate of the per-flow contract.
- We analyze IEEE TSN's implementation of the IR, *asynchronous traffic shaping* [2] and discuss possible changes in the ATS specifications for removing the IR's head-of-line blocking phenomenon exploited by the Spring adversary and solving the issues raised in this paper.

The paper is organized as follows. We provide the background on network-calculus service curves and regulators in Section 2. We discuss the related work in Section 3 and provide the system model in Section 4. We then analyze the role of the FIFO assumption in the shaping-for-free property of the IR and provide an intuition of the Spring adversary in Section 5. Afterward, we discuss the context-agnostic service curves of the IR in Section 6. In Section 7, we discuss the consequences for TSN ATS and describe possible changes in the specifications to solve the issue. In Section 8, we provide the formal description of the Spring adversary. We provide our conclusive remarks in Section 9.

With respect to our previous work [**?**], this paper contains the novel Section 7 that is useful for the TSN community as it discusses *asynchronous traffic shaping*, the TSN's implementation of the IR and the possible changes in the specifications for resolving the issues raised in this paper. Spring's trajectory is fundamental in analyzing the IR's behavior and proving the inexistence of useful service curves. In [**?**], we only provided an intuition of Spring. In this paper, we keep the intuition in Section 5 but additionally provide the formal description of the adversary in Section 8 (by using a marked-point-process description). This description ensures that the properties of the adversary and its consequences for the IR can be formally verified. Last, we provide the novel Appendix A that contains all formal proofs of our results.

2. Background

In time-sensitive networks, performance metrics such as flows' end-to-end latencies have to be bounded in the worst case, not in average. The network-calculus framework [5, 6, 7] can provide such performance bounds. It describes the data traffic with cumulative functions, such as $R^{\mathtt{A}}$, where, for $t \in \mathbb{R}^+$, $R^{\mathtt{A}}(t)$ is the amount of data that cross the observation point $\mathtt{A}$ between an arbitrary time reference 0 and t. Cumulative functions belong to $\mathfrak{F}_0 = \{\mathfrak{f} : \mathbb{R}^+ \to \mathbb{R}^+ | \mathfrak{f}(0) = 0\}$.

2.1. Network-Calculus Service Curves

A causal[1] network system S offers a service curve β if (a) β is wide-sense increasing and (b) for any input cumulative function $R^{\mathtt{A}}(t)$, the resulting output traffic $R^{\mathtt{B}}(t)$ verifies

$$\forall t \geq 0, \quad R^{\mathtt{B}}(t) \geq (R^{\mathtt{A}} \otimes \beta)(t) \tag{1}$$

where $\otimes$ describes the min-plus convolution (Table 1). Common service curves are of the form *rate latency* $\beta_{R,T} : t \mapsto R \cdot [t - T]^+$ with rate R and latency T, and $[\cdot]^+ = \max(\cdot, 0)$.

Some network systems provide stronger guarantees by offering a *strict* service curve. A causal network system S offers a strict service curve β^{strict} if (a) β^{strict} is wide-sense increasing and (b) during any time interval $(s, t]$ in which the system is never empty (a so-called *backlogged period*), the output $R^{\mathtt{B}}$ verifies

$$\forall t \geq s \geq 0, \quad R^{\mathtt{B}}(t) - R^{\mathtt{B}}(s) \geq \beta^{\text{strict}}(t - s) \tag{2}$$

Such a system then also offers β^{strict} as a service curve: β^{strict} also verifies Inequation (1) [5, Prop. 1.3.5].

To distinguish the service curves (strict or not) that model the intrinsic behavior of a system from those that rely on assumptions on the external environment of the system, we define in this paper the concept of *context-agnostic* service curves:

Definition 1 (Context-agnostic service curves)**.** We say that a network system offers a *context-agnostic* service curve [resp., strict service curve] if (1) [resp., (2)] holds for any input $R^{\mathtt{A}}(t)$, without any other assumption on the upstream systems.

Reciprocally, network calculus models traffic flows by arrival curves: a wide-sense increasing function $\alpha \in \mathfrak{F}_0$ is an arrival curve for the traffic at $\mathtt{A}$ if

$$\forall\ t \geq s \geq 0, \quad R^{\mathtt{A}}(t) - R^{\mathtt{A}}(s) \leq \alpha(t - s) \tag{3}$$

Inequation 3 states that α is an arrival curve for the traffic at $\mathtt{A}$ if, for any u, the amount of data that cross the observation point $\mathtt{A}$ during any time interval

[1] In this paper, a network system is causal if it does not produce any data internally, if it does not duplicate and does not expand the data.

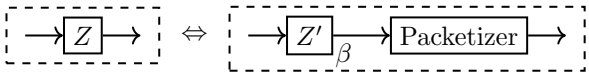


Figure 1: Notations of Definition 2. Z offers the *fluid service curve* β if it can be modelled as the concatenation of Z' followed by a packetizer, where Z' offers the service curve β.

of duration u is less than $\alpha(u)$. We also denote $R^{\mathtt{A}} \sim \alpha$ and say that the traffic is α-constrained. Common arrival curves are of the form *leaky bucket* $\gamma_{r,b}$ with a rate r and a burst b: $\forall t > 0, \gamma_{r,b}(t) = rt + b$.

Given some arrival-curve and service-curve constraints, network-calculus results provide delay and backlog bounds at a network element. A common approach for computing end-to-end performance bounds in time-sensitive networks consists in obtaining an arrival-curve model for each flow and a service-curve model for each network element. Service-curve models for most IEEE TSN mechanisms can be found in [8, 9].

2.2. The Packetizer and Fluid Service Curves

In packet-switching time-sensitive networks, the stream of data at an observation point $\mathtt{A}$ can either be fluid (*e.g.,* on the transmission links) or packetized (packet-by-packet, *e.g.,* within the switches). A packetizer transforms a fluid stream into a packetized stream by releasing the packet's bits only when the last bit is received. The packetizer does not increase the end-to-end latency bounds [5, Thm. 1.7.5].

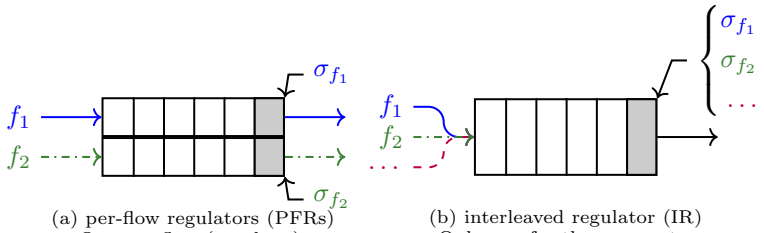
When a system Z with packetized input and output can be split into a fluid service-curve element followed by a packetizer, we say that Z offers a *fluid* service curve.

Definition 2 (Fluid service curve)**.** Consider a function $\beta \in \mathfrak{F}_0$ and a system Z with packetized input and output. We say that Z offers β as a *fluid service curve* if there exists a system Z' that offers the service curve β such that Z can be realized as the concatenation of Z' followed by a packetizer (Figure 1).

Some network systems (*e.g.,* the traffic regulators) can only work when placed within the switches, where the flow of data is packetized: The algorithms that define their behavior assume that the input is packetized [2, §8.6.11.3]. For the service curves [resp., strict service curves] of these systems, we continue to use the adjective *context-agnostic* (Definition 1) when Inequations (1) [resp., (2)] hold for any *packetized* input $R^A(t)$. Indeed, restraining (1) and (2) to packetized inputs is more an assumption on the type of the system's border as it is an assumption on the environment in which the system is placed.

2.3. Individual Service Curve for a Flow

In time-sensitive networks, the service modeled by the service curves of Sections 2.1 and 2.2 is shared between the flows of the aggregate $\mathcal{F}$. In (1), the cumulative arrival function $R^{\mathtt{A}}$ of the aggregate at $\mathtt{A}$ is the sum of the individual



(a) per-flow regulators (PFRs)
One per flow (two here).

(b) interleaved regulator (IR)
Only one for the aggregate.

Figure 2: Two flavors of traffic regulators. With PFRs, we need one PFR per flow. In contrast, the IR uses a single FIFO queue to shape several flows. Best with colors (not required).

arrival functions $\{R_f^{\mathtt{A}}\}_{f\in\mathcal{F}}$ for each flow f in the aggregate $\mathcal{F}$: $\forall t \in \mathbb{R}^+, R^{\mathtt{A}}(t) = \sum_{f\in\mathcal{F}} R_f^{\mathtt{A}}(t)$.

We say that a system S offers to flow g the *individual* service curve β_g if (a) β_g is wide-sense increasing, and (b) for any cumulative function $R_g^{\mathtt{A}}$ of the flow g at the input $\mathtt{A}$ of S, the cumulative function $R_g^{\mathtt{B}}$ of g at its output $\mathtt{B}$ verifies

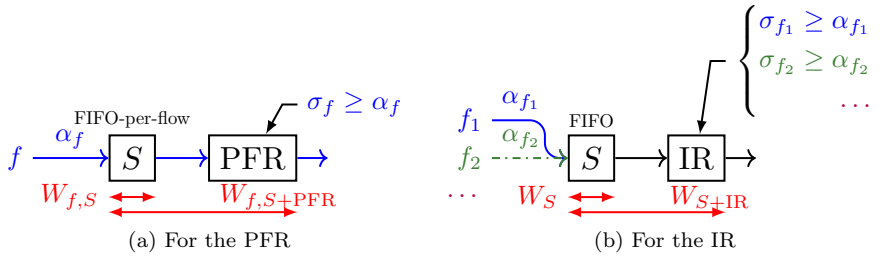
$$\forall t \geq 0, \quad R_g^{\mathtt{B}}(t) \geq (R_g^{\mathtt{A}} \otimes \beta_g)(t) \tag{4}$$

2.4. Traffic Regulators and their Shaping-For-Free Properties

Traffic regulators are hardware elements placed before a multiplexing stage to remove the increased burstiness due to interference with other flows in previous hops. They enable the computation of guaranteed latency bounds in networks with cyclic dependencies [3, 10, 11]. They come in two flavors.

A per-flow regulator (PFR) is a causal, lossless, FIFO system configured for a unique flow f with a shaping curve σ_f (Figure 2a). It stores the packets of f in order of arrival and releases the head-of-line (HOL) packet at the earliest time such that the resulting output has σ_f as an arrival curve. In a network with multiple flows, there is one PFR per flow.

The interleaved regulator (IR) is a causal, lossless, and FIFO system that processes an aggregate $\mathcal{F} = \{f_1, f_2, \dots\}$ of several flows, each one with its own shaping curve $(\sigma_{f_1}, \sigma_{f_2}, \dots,$ see Figure 2b). It stores all the packets of the aggregate $\mathcal{F}$ in order of arrival into a single FIFO queue and only looks at the head-of-line (HOL) packet. The HOL packet p is released as soon as doing so does not violate the configured shaping curve for the associated flow f_i: Packet p can either be immediately released (if the resulting traffic for f_i at the IR's output is σ_{f_i}-constrained) or delayed to the earliest date that ensures that f_i is σ_{f_i}-constrained at the IR's output. This delay depends on the shaping curve σ_{f_i} for the associated flow f_i and the history of departure dates for previous packets of the same flow. During this delay, any other packet p' in the queue is blocked by the HOL packet p, even if p' belongs to another flow f_j and even if p' could be immediately released without violating the shaping curve σ_{f_j} for its flow f_j.



(a) For the PFR (b) For the IR

Figure 3: Shaping-for-free properties of the traffic regulators. For the PFR, the system S only needs to be FIFO-per-flow. For the IR, S must be FIFO for the aggregate. Best with colors (not required).

Traffic regulators can delay individual packets, but there exist specific conditions in which they do not increase the worst-case latency bounds of the flows. This fundamental *shaping-for-free* property is central in the analysis of time-sensitive networks with traffic regulators. It slightly differs for the two flavors.

Theorem 1 (Shaping-for-free property of the PFR [12, Thm. 3])**.** Consider a flow f with input arrival curve α_f that crosses in sequence a causal system S followed by a PFR (Figure 3a). If the PFR is configured with $\sigma_f \geq \alpha_f$ and if S is FIFO for f, then the worst-case delay $W_{f,S+\mathrm{PFR}}$ of f through the concatenation is equal to the worst-case delay $W_{f,S}$ of the flow through the previous system S only.

Theorem 2 (Shaping-for-free property of the IR [12, Thm. 4])**.** Consider an aggregate $\mathcal{F} = \{f_1, f_2, \dots\}$ with input arrival curves $\{\alpha_f\}_{f\in\mathcal{F}}$ that crosses in sequence a causal system S followed by an IR (Figure 3b). If the IR is configured with $\forall i, \sigma_{f_i} \geq \alpha_{f_i}$ and if S is FIFO for the aggregate, then the worst-case delay $W_{S+\mathrm{IR}}$ of the aggregate $\mathcal{F}$ through the concatenation is equal to the worst-case delay W_S of the aggregate through the previous system S only.

Theorems 1 and 2 exhibit two fundamental differences. First, Theorem 2 only ensures that the worst-case delay of the aggregate is not increased, whereas Theorem 1 guarantees that the worst-case delay of the individual flow is preserved. Within an aggregate, the first bound can be larger than the latter, *e.g.,* when the flows have different packet sizes. Second, Theorem 1 only requires the previous system S to be FIFO for each flow individually (FIFO-per-flow), whereas the same system is required to be globally FIFO for Theorem 2.

3. Related Work on the Modeling of Traffic Regulators

In time-sensitive networks with traffic regulators, end-to-end latency bounds for the flows are obtained by combining the shaping-for-free property for traffic regulators with service-curve-based network-calculus results for other systems.

This differentiated treatment of network elements (traffic regulators vs. other systems with service-curve models) restrains the choice of the end-to-end analysis method. Methods based on total-flow analysis (TFA) [13, §3.2] can be adapted to networks with traffic regulators [3, 11]. Other approaches, such as single-flow analysis (SFA) [13, §3.3], pay multiplexing only once (PMOO) [14] and flow prolongation [15] provide tighter end-to-end latency bounds than TFA in several types of networks [16], but they heavily rely on service-curve models. In addition, service-curve models provide continuity and differentiability properties, which allows for synthesizing network designs from the performance requirements, as shown by Geyer and Bondorf in [17]. Hence, a need exists for obtaining service-curve models for all elements of time-sensitive networks, including traffic regulators such as PFRs and IRs.

The per-flow regulator (PFR) was introduced under the name *packetized greedy shaper* in [5, §1.7.4]. Le Boudec and Thiran proved in [5, §1.7.4] that if σ_f is concave and such that $\lim_{t\to 0^+} \sigma_f(t)$ is larger than the maximum packet size of f, then the PFR offers σ_f as a fluid service curve [5, Thm. 1.7.3]. This model proves Theorem 1. In this paper, we say that the curve σ_f *explains the shaping-for-free property* of the PFR and we formally define this notion in Section 5.2. Due to its network-calculus service-curve model, the behavior of the PFR can also be studied in situations where Theorem 1 does not apply. In [18], the consequence of redundancy mechanisms – that can affect the FIFO property – is studied, and end-to-end latency bounds are obtained for flows in networks with redundancy mechanisms and PFRs. In [12, §IV.A], Le Boudec also provides an input-output characterization of the PFR. This type of model does not rely on the concept of service curve but describes the PFR's output packet sequence as a function of the input packet sequence.

The interleaved regulator (IR) was introduced by Specht and Samii under the name *urgency-based scheduler* [10]. As opposed to the PFR, its shaping-for-free property was proved without the concept of service curves: with a trajectorial approach in [10] and with an input-output characterization in [12, §V]. The equivalence between the theoretical model of the IR and the TSN implementation (*asynchronous traffic shaping*, [2]) was proved by Boyer in [19], who also provides a second input-output characterization [19, §5.3]. The only useful service curves that are known for the IR are only valid when the IR is placed in a context that meets the conditions of Theorem 2. A first context-dependent service curve is provided in [3, §IV.A.1] and then slightly improved in [4, §III.B.1]. Because of their dependency on the context, none of these curves can be used to analyze the behavior of the IR in the situations where the shaping-for-free does not hold as in [18]. In contrast, the only context-agnostic service curve known for the IR is the trivial function $t \mapsto 0$. In [20], Hamscher mentions the first conjecture on a non-trivial context-agnostic service curve for the IR and uses a linear-programming approach for hardening their conjecture pending formal proof. The conjecture was not shared (and, to our knowledge, has not been published at the time of this writing). However, the presentation triggered discussions on whether the IR's behavior could be captured by context-agnostic service curves. These discussions motivated this paper.

Table 1: Notations

Common Operators		
$a \vee b$	$= \max(a, b)$	Maximum of a and b.
$a \wedge b$	$= \min(a, b)$	Minimum of a and b.
$[c]^+$	$= \max(0, c)$	
$\lfloor x \rfloor$	$= \max\{n \in \mathbb{Z} \mid n \leq x\}$	Floor function.
$\mathfrak{f} \otimes \mathfrak{g}$	$t \mapsto \inf_{s \leq t} \mathfrak{f}(s) + \mathfrak{g}(t-s)$	Min-plus convolution
$\mathfrak{f} \mathbin{\overline{\otimes}} \mathfrak{g}$	$t \mapsto \sup_{s \leq t} \mathfrak{f}(s) + \mathfrak{g}(t-s)$	Max-plus convolution
$\mathfrak{f} \mathbin{\overline{\oslash}} \mathfrak{g}$	$t \mapsto \inf_{u \geq 0} \mathfrak{f}(t+u) + \mathfrak{g}(u)$	Max-plus deconvolution
$\mathfrak{F}_0$	$= \{\mathfrak{f} : R^+ \to R^+ \mid \mathfrak{f}(0) = 0\}$	Set of curves
Common Curves		
$\gamma_{r,b}$	$t \mapsto \begin{cases} 0 & \text{if } t = 0 \\ rt + b & \text{if } t > 0 \end{cases}$	Leaky-bucket curve.
$\beta_{R,T}$	$t \mapsto R[t-T]^+$	Rate-latency curve.
Flows		
$f \in \mathcal{F}$	A flow f in the set of flows $\mathcal{F}$	
$\{\sigma_f\}_{f \in \mathcal{F}}$	A set of shaping curves for the flows $\mathcal{F}$	
$L_f^{\min}$, $L_f^{\max}$	Minimum [resp., maximum] packet size of flow f	
Trajectory Description		
x	A trajectory: Description of all the events in the network	
$\mathtt{M}$	An observation point	
$\mathscr{M}^x$	Packet sequence at $\mathtt{M}$ in trajectory x	
$R^{x,\mathtt{M}}$	Cumulative function of the aggregate...	
[resp., $R_f^{x,\mathtt{M}}$]	... [resp., of f] at $\mathtt{M}$ in Trajectory x.	
$R^{x,\mathtt{M}} \sim \alpha$	$R^{x,\mathtt{M}}$ is constrained by α, Equation (3)	
Parameters of the Spring adversary (Section 5.1)		
I	Expected spacing for same-flow packets after the IR	
d	Maximum delay in the Spring-controlled system S_1	
ϵ	Margin (minimum packet spacing after S_1)	
τ	Period of the six-packet-long profile	

Hence, two questions remain open: Beyond the function $\beta : t \mapsto 0$, what other context-agnostic service curves does the IR provide? Do any of them explain Theorem 2? We address these two questions in this paper.

4. System Model and Notations

We consider an asynchronous packet-switching time-sensitive network that contains traffic regulators. We focus on a particular traffic regulator within this network. It can either be a per-flow regulator (PFR) that processes a single flow $\mathcal{F} = \{f\}$ with shaping curve σ_f or an interleaved regulator (IR) that processes an aggregate $\mathcal{F}$ with leaky-bucket shaping curves $\{\sigma_f\}_{f \in \mathcal{F}} = \{\gamma_{r_f, b_f}\}_{f \in \mathcal{F}}$. We focus on the subset of flows $\mathcal{F}$ that cross the regulator. We model any other

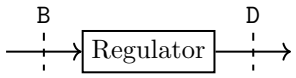


Figure 4: Input [resp., output] observation point B [resp., D] for a regulator.

network elements (queues, schedulers, switching fabrics, transmission links, ...) or sequence of network elements crossed by the flows $\mathcal{F}$ between their sources, the regulator, and their destinations as black-box systems. Each system has a traffic input and a traffic output and is only assumed to be causal and lossless: it neither produces nor loses any data internally. Data is produced at the flows' sources and consumed at the flows' destinations.

A *trajectory* x is a description of all the events in the network (packet arrival, packet departure). It is *acceptable* if all known constraints are satisfied. For an observation point M, we denote by $R^{x,\mathtt{M}}$ [resp., $R_f^{x,\mathtt{M}}$] the cumulative function of the aggregate [resp., of the flow $f \in \mathcal{F}$] at observation point M in trajectory x. If the stream is packetized at M, we call $\mathscr{M}^x$ the packet sequence that describes the packets' arrival date, size, and associated flow at M in trajectory x.

For an input packet sequence $\mathscr{B}^x$ at the input B of the regulator (Figure 4), we use the equivalent input-output characterizations of traffic regulators from [12] and [19] to obtain the output packet sequence $\mathscr{D}^x$ at the output D of the regulator.

We list the notations in Table 1.

5. Limits of the Shaping-For-Free Property for the Interleaved Regulator

The shaping-for-free property is a strong attribute of the interleaved regulator (IR). However, it is context dependent: It makes assumptions on the context in which the IR is placed. In this section, we investigate the limits of these assumptions.

First, we observe that Theorem 2 requires the upstream system to be FIFO. In Section 5.1, we prove that removing this assumption makes the IR unstable: it can yield unbounded latencies. We then prove in Section 5.2 that there exists no service-curve model of the IR that can explain Theorem 2.

5.1. Instability of the IR when Placed after a Non-FIFO System

In this subsection, we discuss the role of the FIFO assumption in Theorem 2. When removed, we prove that the IR can yield unbounded latencies. Specifically, we prove

Theorem 3 (Instability of the IR after a non-FIFO system)**.** Consider an IR that processes three or more flows with the same leaky-bucket shaping curve for the first three flows: $\forall f_i \in \{f_1, f_2, f_3\}, \sigma_{f_i} = \gamma_{r,b}$ with $r > 0$ and b greater than the maximum packet size of f_1, f_2, f_3. For any $W > 0$, there exists a system S_1 and a source ϕ (Figure 5), such that:

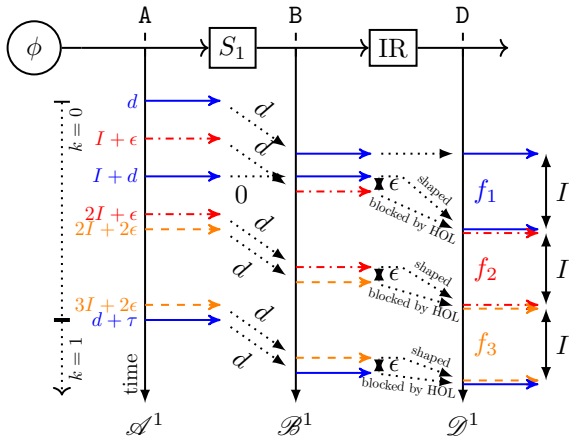


Figure 5: The network $\mathcal{N}_1$ and the Spring-generated Trajectory 1 that yields unbounded latencies in the IR when S_1 is not assumed FIFO. Best with colors (not required).

1/ each flow f_i is σ_{f_i}-constrained at the source ϕ,
2/ S_1 is causal, lossless and FIFO-per-flow (but globally non-FIFO),
3/ when the system S_1 is placed after the source as in Figure 5, then the delay of each flow within S_1 is upper-bounded by W,
4/ when the IR is placed after S_1 as in Figure 5, then the delay of any flow within the IR is not bounded.

The proof of Theorem 3 relies on an adversarial traffic generation that we call "Spring". Spring is an adversary that knows the values of b, r and W in Theorem 3 and controls the source ϕ and the system S_1 of Figure 5 such that Properties 1 to 4 of Theorem 3 hold. It defines the constants I, d, ϵ and τ as follows

$$I \triangleq \frac{b}{r} \;;\; 0 < d < \min(I, W) \;;\; 0 < \epsilon < \min(I - d, \frac{d}{3}) \;;\; \tau \triangleq 3I + 3\epsilon - d \quad (5)$$

The formal description of Spring and the formal proof of Theorem 3 are available in Section 8. They require formal notations for the packet sequences that we detail in Section 8. In the following, we provide an intuition of the Spring adversary without the formal notations.

Intuition. Trajectory 1 generated by Spring is illustrated in Figure 5. All packets have the size b. The far-left timeline shows the packet sequence $\mathscr{A}^1$ for the three flows at the output of the Spring-controlled source. A sequence of six packets is repeated with period τ. Only the period $k = 0$ is shown.

The dotted arrows that lead to the second timeline highlight each packet's delay in the Spring-controlled system S_1 and the resulting packet sequence $\mathscr{B}^1$. The main property of Trajectory 1 is that the first packet of the dash-dotted red flow f_2 and the second packet of the solid blue flow f_1 have exchanged their

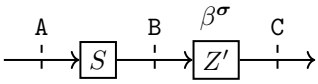


Figure 6: Notations of Definition 3. β^{σ} explains the shaping-for-free property if any system Z' that offers β^{σ} as a service curve does not increase the worst-case delay of the flows when placed after any FIFO system S.

order at B compared to their order at A. This is because the former suffers a delay d through S_1, but the latter does not suffer any delay. Note that the Spring-controlled system S_1 is not FIFO but remains causal, lossless, and FIFO-per-flow with a delay bound $d < W$.

The dotted arrows that link the second to the third timeline describe the behavior of the IR (not controlled by Spring) when provided with the input sequence $\mathscr{B}^1$. For example, the first packet of f_1 is immediately released by the IR because the network was previously empty. However, the second solid blue packet of f_1 is shaped (delayed) by the IR because releasing it would violate the $\gamma_{r,b}$ shaping constraint for f_1 at the output of the IR. This packet is released as soon as doing so does not violate the $\gamma_{r,b}$ constraint, *i.e.,* I seconds after the previous packet. Because of this, the first dash-dotted red packet of the flow f_2 is blocked by the head-of-line (HOL). And the second packet of f_2 is shaped and delayed to ensure a distance of I from the previous packet of f_2.

As a result, it takes $3I$ seconds for the IR to output the six packets of the first period, whereas they entered the IR within τ seconds. As $\tau < 3I$, we can generate a constant build-up of delay and backlog in the IR by repeating the six-packet-long profile every τ seconds.

5.2. The Shaping-for-Free Property of the Interleaved Regulator Cannot be Explained by a Service Curve

Theorem 3 shows that the IR does not provide any context-agnostic delay guarantees as a stand-alone network element. In contrast, if a system Z offers a context-agnostic service curve β that explains a context-dependent property (*e.g.,* shaping-for-free), then β continues to hold when Z is placed in a context that differs from the assumptions of the context-dependent property. β can be used to compute the consequences of the deviation from the assumptions and their resulting penalties on performance bounds. For such a system, slight deviations from the assumptions should lead to small delay penalties.

This is the case for the PFR, for which we can find a service-curve model that explains its shaping-for-free property. We formally define this notion as follows.

Definition 3 (A curve explains the shaping-for-free property)**.** Consider a set of flows $\mathcal{F}$ and a set of shaping curves $\boldsymbol{\sigma} = \{\sigma_f\}_{f\in\mathcal{F}}$. We say that $\beta^{\boldsymbol{\sigma}} \in \mathfrak{F}_0$ explains the shaping-for-free property if and only if: For any causal, lossless and FIFO systems Z' and S, if Z' offers $\beta^{\boldsymbol{\sigma}}$ as a service curve, then the worst-case delay of the aggregate between A and B (Figure 6) over all the trajectories

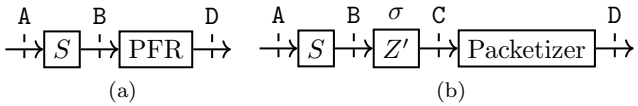


Figure 7: Application of Proposition 1 to prove Theorem 1. (a) A PFR placed in the conditions of Theorem 1. (b) The equivalent model as per Proposition 1.

$X = \{x | \forall f \in \mathcal{F}, R_f^{\mathtt{A},x} \sim \sigma_f\}$ equals the worst-case delay between A and C over the same set X.

As per this definition, a function $\beta^{\boldsymbol{\sigma}}$ explains the shaping-for-free property if any system Z' that is only assumed to offer $\beta^{\boldsymbol{\sigma}}$ as a service curve does not increase the worst-case delay of the aggregate $\mathcal{F}$ when placed in a context that meets the assumptions of Theorems 1 and 2 (*i.e.,* placed after a FIFO system S with flows that are initially constrained by their shaping curves). For the PFR, we have a positive result:

Proposition 1. Consider a PFR that shapes a single flow $\mathcal{F} = \{f\}$ with a concave shaping curve σ_f such that $\lim_{t \to 0} \sigma_f(t) \geq L_f^{\max}$. Then the PFR offers the fluid service curve $\beta^{\boldsymbol{\sigma}} = \sigma_f$ that explains the shaping-for-free property.

This result directly derives from [5, Thm. 1.7.3]. We provide its formal proof in Appendix A.1. Note that Proposition 1 contains two statements: (1) σ_f is a fluid service curve of the PFR. (2) σ_f explains the shaping-for-free property (Definition 3).

Let us discuss why the two statements of Proposition 1 prove Theorem 1. Consider a causal, lossless, and FIFO system S and a PFR configured with σ_f placed after S (Figure 7a). By combining the first statement of Proposition 1 with Definition 2, the PFR can be realized as the concatenation of Z' followed by a packetizer (Figure 7b), where Z' is a causal, lossless, and FIFO system that offers $\beta^{\boldsymbol{\sigma}} = \sigma_f$ as a service curve. We then combine the second statement of Proposition 1 with Definition 3. We obtain that if σ_f is an arrival curve for f at the input of S, then the worst-case delay of the flow f through S equals the worst-case delay of the flow through the concatenation of S and Z'. Finally, the packetizer does not increase the worst-case latency bounds [12, Thm. 1.7.1], which proves Theorem 1.

As opposed to the PFR, we now prove that no fluid service curve can explain the shaping-for-free property of the IR.

Theorem 4. An IR that processes at least three flows with the same leaky-bucket shaping curve does not have any fluid service curve that explains its shaping-for-free property.

To prove Theorem 4, we rely on the following Lemma 1, which we prove in Appendix A.2.

Lemma 1. If $\beta^{\boldsymbol{\sigma}}$ explains the shaping-for-free (Definition 3), then $\beta^{\boldsymbol{\sigma}} \geq \sum_{f \in \mathcal{F}} \sigma_f$

To prove Theorem 4, we exhibit a variant of the trajectory from Figure 5 that shows that a function larger than $\sum_{f \in \mathcal{F}} \sigma_f$ cannot be a fluid service curve of the IR. This variant of the Spring trajectory is formally described in Section 8.3 together with the formal proof of Theorem 4.

6. Service Curves of The Interleaved Regulator

In the previous section, we use the sensitivity of the interleaved regulator (IR) to the FIFO assumption in Theorem 2 to prove that the IR has no fluid service curve that can explain its shaping-for-free property.

In this section, we show that the IR still offers a family of non-trivial context-agnostic service curves. In particular, we exhibit a non-bounded strict service curve for the IR (Theorem 5). This strict-service-curve model is of interest for understanding the behavior of the IR in situations that differ from the shaping-for-free property. It can be used to model the IR in service-curve-oriented analysis like SFA and PMOO.

However, any service curves of the IR are also fluid service curves of the IR, and we know from Theorem 4 that they must be weak because they cannot explain its shaping-for-free property. Indeed, their long-term rate is upper bounded: For an IR that processes more than four flows, the long-term rate of any of its service curves is upper bounded by three times the rate enforced for a single flow, as we show in Theorem 7.

For an IR that processes at least four flows, we also prove that any individual service curve β_g offered to a single flow g is upper bounded by its minimum packet size $L_g^{\min}$ (Theorem 6).

The section is organized as follows. First, we obtain a strict service curve of the IR (Theorem 5) by using the input/output models of [12, 19]. Then, we use the Spring trajectory from Section 5 to obtain upper bounds on the individual service curve of the IR for a single flow (Theorem 6). Last, we use Theorem 6 to upper-bound the long-term rate of any context-agnostic service curve of the IR for the aggregate (Theorem 7).

6.1. A Strict Service Curve for the Aggregate

Even though Spring's Trajectory from Section 5.1 generates a constant build-up of delay in the IR, we can observe in Figure 5 that the IR continuously outputs two packets every I seconds. In fact, we can find a minimum output rate whenever the IR is non-empty. This shows that the IR offers a strict service curve as defined in Section 2.1:

Theorem 5 (IR strict service curve)**.** Consider an IR that processes an aggregate of flows $\mathcal{F}$ with leaky-bucket shaping curves: $\forall f \in \mathcal{F}, \sigma_f = \gamma_{r_f,b_f}$ with $r_f > 0$ and $b_f \geq L_f^{\max}$, where $L_f^{\min}$ [resp., $L_f^{\max}$] is the minimum [resp., maximum] packet size of f. Define

$$L^{\min} = \min_{f \in \mathcal{F}} L_f^{\min} \quad I^{\max} = \max_{f \in \mathcal{F}} \frac{L_f^{\max}}{r_f} \tag{6}$$

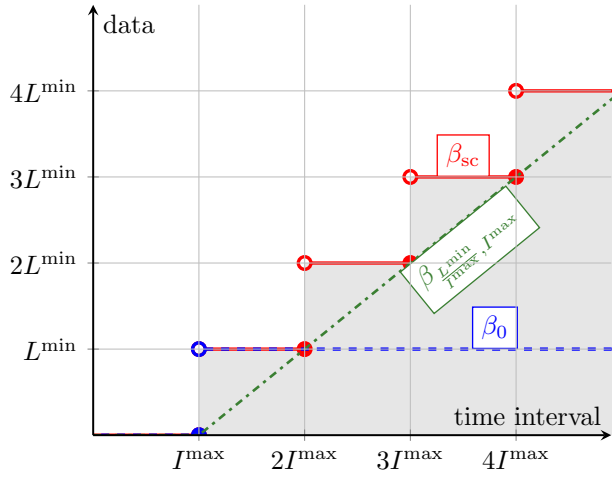


Figure 8: Three different strict service curves of the IR (Theorem 5). The step dashed blue function β_0 is the first curve obtained in the theorem's proof. Its supper-additive closure, the solid red function β_{sc} is then also a strict service curve, as well as the dash-dotted green rate-latency curve $\beta_{\frac{L^{\min}}{I^{\max}},I^{\max}}$. Best with colors (not required).

Then the staircase curve $\beta_{\text{sc}} : t \mapsto \lfloor t/I^{\max} \rfloor \cdot L^{\min}$ (with $\lfloor \cdot \rfloor$ the floor function, Table 1) and the rate-latency curve $\beta_{R,T}$ with $T = I^{\max}$ and $R = L^{\min}/I^{\max}$ (Figure 7) are context-agnostic strict service curves of the IR for the aggregate.

To prove this result, we consider a non-empty IR. The output time of the head-of-line packet is given by the input/output characterizations of [12, 19]. It can be upper-bounded. From this we obtain that the dashed-blue curve β_0 in Figure 8 is a strict service curve of the IR, and so is its supper-additive closure [7, Prop 5.6], defined in [7, §2.4] as the function

$$\beta_0 \vee (\beta_0 \overline{\otimes} \beta_0) \vee ((\beta_0 \overline{\otimes} \beta_0) \overline{\otimes} \beta_0) \vee \dots \tag{7}$$

where $\vee$ is the maximum and $\overline{\otimes}$ is the max-plus convolution (Table 1). The computation of (7) gives β_{sc}, shown in solid red in Figure 8. Any wide-sense increasing curve that remains below the β_{sc} service curve is also a strict service curve of the IR [7, Prop 5.6]. This is the case for the rate-latency service curve $\beta_{L^{\min}/I^{\max},I^{\max}}$ shown with a dash-dotted green line in Figure 8. We provide the formal proof of Theorem 5 in Appendix A.3.

Application to the Situation of Section 5.1. Figure 9 shows the cumulative arrival function R^{B} at the input of the IR as well as the cumulative departure function R^{D} at the output of the IR, in Spring's trajectory[2] described in Section 5.1 and Figure 5. We also provide the arrival curve α^{B} of the aggregate at the input of the IR.

[2]With the notations of (5), the parameters used in Figure 9 are: $W = 0.86I, d = 0.85I$, $\epsilon = 0.05I$.

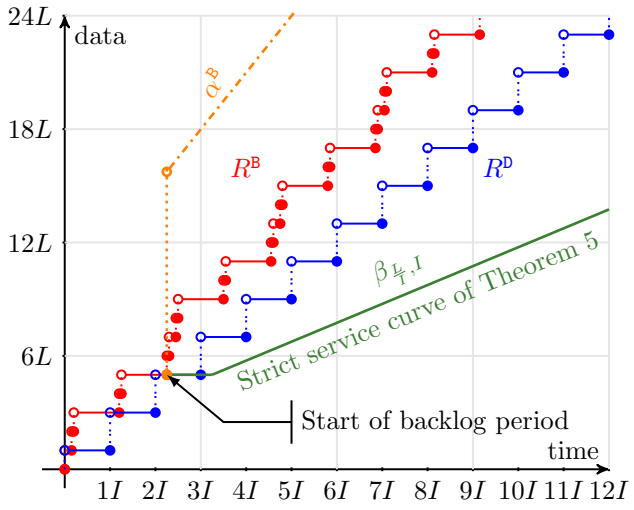


Figure 9: Application of Theorem 5 to the Spring trajectory of Section 5.1. The dotted red [resp., blue] curve is the cumulative function at the input [resp., output] of the IR. The dash-dotted orange leaky-bucket curve is an arrival curve of the aggregate at the input of the IR. The solid green rate-latency curve is a strict service curve of the IR, as proved by Theorem 5. This figure should be printed with colors.

In Section 5.1, all packets of the aggregate have the same size L and all three flows have the same leaky-bucket shaping curve $\sigma_{f_1} = \sigma_{f_2} = \sigma_{f_3} = \gamma_{r,b}$ with $b = L$. The application of Theorem 5 gives that $\beta_{\frac{L}{I},I} = \beta_{r,\frac{b}{r}}$ is a context-agnostic strict service curve of the IR for the aggregate. In Figure 9, we place this curve in green at the beginning of the backlog period. We confirm that when the IR is non-empty, the cumulative output R^{B} is larger than the strict service curve. Also, it is clear that the horizontal deviation between the arrival curve α^{B} and the service curve $\beta_{\frac{L}{I},I}$ is not bounded. This means that the network-calculus theory cannot provide a latency bound with this service-curve model, which is consistent with Theorem 3.

In Figure 9, the rate $\frac{L}{I} = r$ of the green service curve $\beta_{\frac{L}{I},I}$ does not follow the long-term rate $\frac{2L}{I} = 2r$ of the output cumulative function R^{D}. However, no better rate for the rate-latency context-agnostic strict service curve can be achieved:

Proposition 2 (Upper bound on the strict service curve)**.** Consider an IR that processes an aggregate $\mathcal{F}$ with leaky-bucket shaping curves $\{\sigma_f\}_{f\in\mathcal{F}} = \{\gamma_{r_f,b_f}\}_{f\in\mathcal{F}}$. Consider a curve β^{strict} and assume that β^{strict} is a context-agnostic strict service curve of the IR. Then

$$\forall t \geq 0, \quad \beta^{\text{strict}}(t) \leq \min_{f\in\mathcal{F}} \sigma_f(t) \tag{8}$$

In particular, if $\beta^{\text{strict}} = \beta_{R,T}$ is a rate-latency curve, then $R \leq \min_{f\in\mathcal{F}} r_f$.

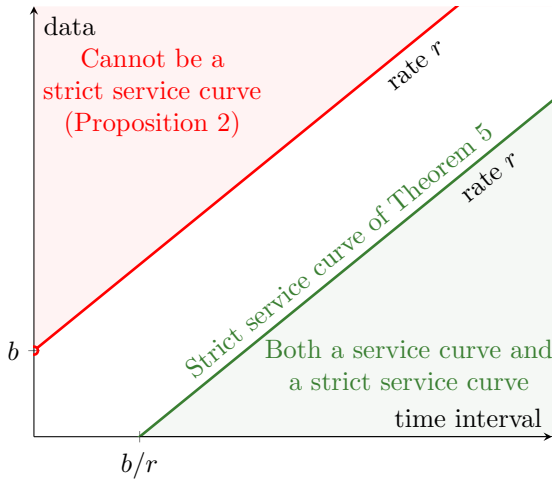


Figure 10: Areas of proven, possible, and impossible strict service curves for an IR that processes the flows with the same shaping curve $\gamma_{r,b}$ and if all packets have size b. This figure should be printed with colors.

To prove this result, we consider a set of trajectories $\{x_f\}_{f\in\mathcal{F}}$. For each flow $f \in \mathcal{F}$, the trajectory x_f is obtained by having only flow f send packets of size b_f to the IR at twice the frequency allowed by its shaping curve. Then the backlog of the IR quickly becomes non-empty, but the cumulative output of the IR, $R^{x_f,\mathsf{D}} = R_f^{x_f,\mathsf{D}}$ is constrained by the shaping curve σ_f: $\forall 0 \leq s \leq t, R^{x_f,\mathsf{D}}(t) - R^{x_f,\mathsf{D}}(s) \leq \sigma_f(t-s)$. Combined with the definition of a strict service curve (2), this gives $\beta^{\text{strict}} \leq \sigma_f$. This is valid for all trajectories $\{x_f\}_{f\in\mathcal{F}}$, hence the result. We provide the formal proof of Proposition 2 in Appendix A.4.

Figure 10 shows the areas of proven, possible, and impossible strict service curves of an IR that processes the flows with the same leaky-bucket shaping curve $\gamma_{r,b}$, assuming that all the packets have the size b. Any wide-sense increasing function that remains in the green area is a strict service curve of the IR, as proven by Theorem 5. In contrast, any function that enters the red area cannot be a strict service curve of the IR, as proven by Proposition 2.

Incidentally, any wide-sense increasing function that remains within the green area is also a service curve of the IR. So far, as opposed to the strict-service-curve property, we have not obtained any limit for the service curve (1). To obtain this limit, we first need to consider the *individual* service curve offered to any flow by the IR, which we analyze in the following subsection.

6.2. Upper-Bound on the Individual Service Curve

Consider an IR that processes an aggregate $\mathcal{F}$ and take a flow $g \in \mathcal{F}$. In this section, we are interested in the service the IR guarantees to g, *i.e.,* in an *individual* service curve of the IR for g.

If each flow $f \in \mathcal{F}\backslash\{g\}$ enters the IR with an arrival curve α_f^{B} that is equal or smaller than its shaping curve σ_f, then none of the packets of the flows $\mathcal{F}\backslash\{g\}$

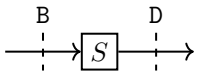


Figure 11: Input [resp., output] observation point B [resp., D] for a system S.

is ever delayed by the IR. In this case, the IR acts as a PFR for g and provides an individual fluid service curve σ_g to g.

In a more likely setting, though, the IR reshapes flows that exhibit an input arrival curve (strictly) larger than their configured shaping curve. In such a case, and if the IR processes more than four flows, no useful individual service curve for g can be obtained:

Theorem 6 (Upper-bound on the individual service curve)**.** Consider an IR that processes an aggregate $\mathcal{F}$ of at least four flows and with the same leaky-bucket shaping curve for at least three of them: $\forall f_i \in \{f_1, f_2, f_3\}, \sigma_{f_i} = \gamma_{r,b}$. Consider a flow $g \in \mathcal{F} \backslash \{f_1, f_2, f_3\}$ and assume that for each $f_i \in \{f_1, f_2, f_3\}, \gamma_{r,b_i}$ is an arrival curve for f_i at the input B of the IR (Figure 4), with $b_1 > b$, $b_2 \geq b$, $b_3 \geq b$ (permutating the indexes if required). Last, consider a curve $\beta_g \in \mathfrak{F}_0$ that can depend on the knowledge of $\{\sigma_f\}_{f \in \mathcal{F}}$ and $\{\alpha_h^{\mathtt{B}}\}_{h \in \mathcal{F} \backslash \{g\}}$.

If β_g is an individual service curve of the IR for the flow g, then β_g is upper-bounded by g's minimum packet size $L_g^{\min}$.

Theorem 6 shows that any individual-service-curve model of the IR for a flow g can only guarantee that one single packet of g will ever cross the IR over the entire network's lifetime. Hence, no useful context-agnostic service curve exists to model the service offered to a single flow by the IR (each flow is likely to send many packets).

To prove Theorem 6, we reuse the Spring trajectory of Theorem 3 for the three flows f_1, f_2, f_3. From Theorem 3, we know that the delay of each individual flow $f_i \in \{f_1, f_2, f_3\}$ through the IR is not bounded.

Remark: We could combine the result of Theorem 3 with the contrapositive of [5, Thm. 1.4.2]. We would obtain that the horizontal distance between the individual arrival curve α_{f_i} and the individual service curve β_{f_i} is not bounded. However, this approach only provides a result on the long-term rate of the individual service curve β_{f_i}. To obtain a stronger result (an upper bound on the individual service curve), we need a result different from [5, Thm. 1.4.2], and whose contrapositive provides an upper bound on the individual service curve. This result is the following proposition.

Proposition 3 (Delay bound of the first packet in a system that offers an individual service curve)**.** Consider a causal, lossless and FIFO system S with a packetized input B and a packetized output D (Figure 11). Consider the aggregate of flows $\mathcal{F}$ that cross S. Take a flow $g \in \mathcal{F}$ and assume that S offers to g the individual service curve β_g. Denote by $L_g^{\min}$ the minimum packet size of g. If there exists $u \in \mathbb{R}^+$ such that $\beta_g(u) > L_g^{\min}$, and if the first packet of g is of size $L_g^{\min}$, then the delay of this first packet in S is upper-bounded by u.

The proof of Proposition 3 is in Appendix A.5. We observe that we cannot trigger the contrapositive of Proposition 3 with the flows f_1, f_2, f_3 of Theorem 3. Indeed, with Spring, the delay of the first packets of f_1, f_2, f_3 is bounded. However, the trajectory used by Spring creates a constant build-up of delay and backlog inside the IR. Hence, if we consider a fourth flow, g, then the first packet of g can suffer a delay as large as desired within the Spring trajectory: We send it when the accumulated delay in the IR is large enough. The combination of this observation with the contrapositive of Proposition 3 provides the result of Theorem 6. The formal proof of Theorem 6 is in Appendix A.6.

6.3. Limits on the Aggregate Service Curve

Let us go back to the analysis of the service offered to the aggregate, by focusing on (non-strict) service-curve models. One consequence of Theorem 6 is that the long-term rate of the service curve for the aggregate is upper-bounded by three times the rate of a single contract.

Theorem 7 (Maximum long-term rate of any service curve)**.** Consider an IR that processes an aggregate $\mathcal{F}$ of at least four flows and with the same leaky-bucket shaping curve for at least three of them: $\forall f_i \in \{f_1, f_2, f_3\}, \sigma_{f_i} = \gamma_{r,b}$. Consider a curve $\beta \in \mathfrak{F}_0$ that can depend on $\{\sigma_f\}_{f\in\mathcal{F}}$. If the IR offers β as a context-agnostic service curve, then

$$\liminf_{t\to+\infty} \frac{\beta(t)}{t} \leq 3r \tag{9}$$

To prove Theorem 7, we pick one flow $g \in \mathcal{F}\backslash\{f_1, f_2, f_3\}$ and a traffic arrival for $h \in \mathcal{F}\backslash\{g\}$ that meets the requirements of Theorem 6. The IR is a FIFO system. Hence, if β is a service curve of the IR, then for any $\theta \in \mathbb{R}^+$, the curve β_g^θ defined by

$$\beta_g^\theta : t \mapsto \left[\beta(t) - \sum_{f\in\mathcal{F}\backslash\{g\}} \alpha_f(t-\theta)\right]^+ \cdot \mathbb{1}_{\{t>\theta\}} \tag{10}$$

verifies Inequation (4) [5, Prop. 6.4.1]. In (10), $\mathbb{1}_{\{\texttt{cond}\}}$ equals 1 when `cond` is true, 0 otherwise. Note that β_g^θ may not be wide-sense increasing, thus not an individual service curve for g. Inspired by [7, §5.2.1], we resolve this by considering the curve $\beta_g^\theta \overline{\oslash} \mathbf{0} : t \mapsto \inf_{s\geq t} \beta_g^\theta(s)$, where $\overline{\oslash}$ is the max-plus deconvolution (Table 1) and $\mathbf{0}$ is the zero function $t \mapsto 0$. The function $\beta_g^\theta \overline{\oslash} \mathbf{0}$ is wide-sense increasing, smaller than β_g^θ, thus an individual service curve for g. From Theorem 6, we obtain $\forall\theta \geq 0, \forall t \geq 0$, $(\beta_g^\theta \overline{\oslash} \mathbf{0})(t) \leq L_g^{\min}$. The left-hand side of this inequation is an infimum, but the result is valid for any $\theta \geq 0, t \geq 0$. Hence, we can derive a bound on the long-term rate of β. We provide the formal proof of Theorem 7 in Appendix A.7.

With Theorem 7, we can conclude that for an IR that processes more than four flows, no useful context-agnostic service curve exists to model the IR. Indeed, the aggregate of the four flows can exhibit a sustained rate of four times

Table 2: Context-Agnostic Service Curves for an IR that Processes at least Four Flows with the Same Leaky-Bucket Shaping Curve $\gamma_{r,b}$, Assuming All Packets Have Size b.

Service-curve type	Curve exhibited in this paper	Limit exhibited in this paper
Service curve β	$\beta_{r,\frac{b}{r}}$	$\liminf_{t\to+\infty} \frac{\beta(t)}{t} \leq 3r$
Strict service curve β^{strict}	$\beta_{r,\frac{b}{r}}$	$\forall t \geq 0,\ \beta^{\text{strict}}(t) \leq \gamma_{r,b}(t)$
Individual service curve $\beta_g,\ \forall g \in \mathcal{F}$	None $(t \mapsto 0)$	$\forall t \geq 0,\ \beta_g(t) \leq L_g^{\min}$

the rate of a single-flow contract, whereas a context-agnostic service curve can only guarantee a long-term service rate of three times this value. Hence, Theorem 7 concludes our search of context-agnostic service-curve models for the IR and the results of Section 6 are summarized in Table 2.

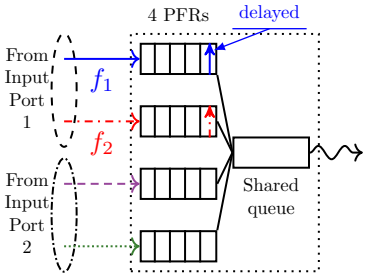
7. Relation with TSN Asynchronous Traffic Shaping (ATS)

The results in this paper are based on the theoretical model of an interleaved regulator (IR), as described in [10] and formally modeled in [5]. Based on the equivalence result of Boyer [19], we argue in this section that the TSN's implementation of the IR, called *asynchronous traffic shaping* (ATS), suffers from the same issues. By mapping the theoretical model of the IR with the ATS standard [2], we then identify which specifications of the ATS behavior create the head-of-line blocking phenomenon exploited by Spring to generate unbounded latencies in Section 5.1. We suggest a minor modification to the ATS standard to remove this head-of-line blocking phenomenon.

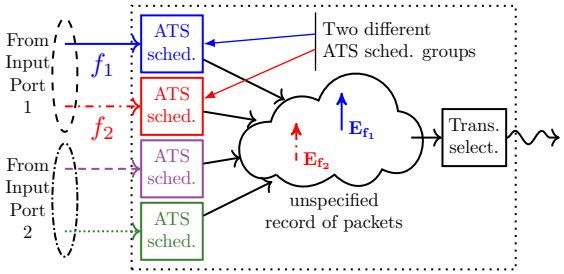
7.1. Consequence for the Analysis of Networks that Use TSN ATS

Figure 12 contains four models of the same output port (in dotted boxes), with a unique traffic class in which four different flows that come from two different input ports compete to access the transmission link (wavy arrow on the right). The figure compares the theoretical models of the output port that are used to analyze the network's performance with network calculus (subfigures on the left) and the specified implementation in IEEE TSN *asynchronous traffic shaping* (ATS, subfigures on the right)[2]. It also compares whether the output port uses per-flow regulators (PFRs) (top subfigures) or interleaved regulators (IRs) (bottom subfigures) to reshape the flows. The analysis in this paper is focused on the aggregate $\mathcal{F}$ of the flows that come from Input Port 1. In this figure, $\mathcal{F}$ contains only two flows: the blue solid flow f_1 and the red dash-dotted flow f_2.

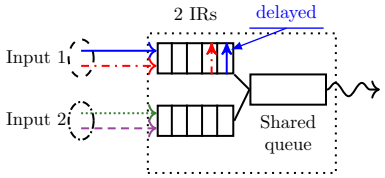
Figure 12a depicts the theoretical model of the output port equipped with per-flow regulators (PFRs) to reshape the flows after an upstream not-shown multiplexing stage. As discussed in Section 2.4, each flow is assigned to its



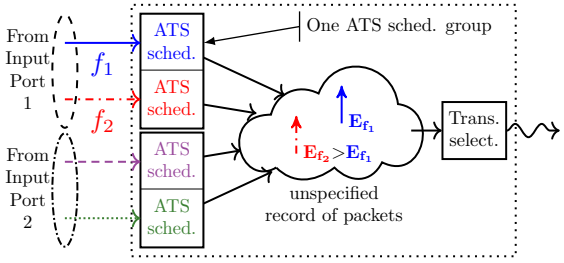
(a) Theoretical model for an output port with per-flow regulators (one PFR per flow, 4 here)



(b) ATS implementation that could behave as 4 PFRs with little modification to the specification



(c) Theoretical model for an output port with interleaved regulators (one IR per input port, 2 here)



(d) Specified ATS implementation that behaves as 2IRs in an IEEE TSN output port

Figure 12: Comparison of the theoretical models of traffic regulators (a, c) and their corresponding ATS configurations (b, d). The configuration (b) is not supported in the current ATS specification [2]. Removing only one sentence from [2] allows this configuration and removes any head-of-line blocking phenomenon. This figure should be printed with colors.

own PFR. The flow's packets are stored in a FIFO queue. When the head-of-line packet is released according to the shaping curve of the PFR, it joins a queue shared by all flows where it further competes with other packets to be transmitted. This competition depends on some subsequent scheduling rule. Two PFRs are necessary to process our aggregate of interest $\mathcal{F} = \{f_1, f_2\}$, and a total of four PFRs is required in this output port to also accommodate for the other two flows that come from the other input port. As each PFR relies on a FIFO queue, this solution requires four FIFO queues in this output port.

Figure 12c depicts the theoretical model of the output port equipped with interleaved regulators (IRs), as proposed in the original paper of Specht and Samii [10]. The flows f_1, f_2 of our aggregate of interest $\mathcal{F}$ share the same FIFO queue. As f_1, f_2 enter from the same input port, they also compete in the same upstream output port. If all previous network elements are globally FIFO for the aggregate $\mathcal{F} = \{f_1, f_2\}$ (upstream output port, input port, switching fabric), then the shaping-for-free property of the IR holds and this IR does not increase the worst-case delay bound of the aggregate $\mathcal{F}$. Hence, Figure 12c illustrates the main property that motivated the invention of the IR: if all network elements are globally FIFO, only one IR per input port is required to keep the shaping-for-free property. This design requires a total of only two FIFO queues instead of four for the design with PFRs.

We can note that the reduction of the implementation complexity brought by the IR relies on the assumption that FIFO queues are the default storage units in hardware components and that they can be combined in the two-step-queuing format (regulators $\rightarrow$ shared queue) shown in Figures 12a and 12c. Hence, the fewer FIFOs queues are required by the design, the simpler the implementation is.

But, in IEEE TSN, the output-port specification contains a unique *queuing function* per traffic class. Moreover, in TSN, a "*queue [...] is not necessarily a single FIFO data structure. A queue is a record of all [packets] of a given traffic class awaiting transmission [...]*" [21, §8.6.6]. Figure 12d depicts the specification of asynchronous traffic shaping (ATS), the TSN's implementation of the IR. In TSN ATS, each flow is processed by an ATS scheduler (ATS sched., one per flow) that computes for each packet an eligibility time based on the algorithm described in [21, §8.6.11.3]. The packets and their eligibility times are then stored in the record of packets (represented by a cloud in Figure 12d). At any time instant, only the packets whose eligibility times are in the past can be selected for transmission by the transmission selection function (trans. select.).

In [19], Boyer proves that the specified ATS implementation (Figure 12d) is equivalent to the theoretical output-port model with one IR per input port (Figure 12c): Packets become eligible for transmission in Figure 12d at the same time instant at which they would have left the IR to reach the shared queue in Figure 12c. Combined with the results of our paper, this means that there exists no useful network-calculus service curve for modeling TSN ATS (Theorem 7). The networks that use the TSN ATS feature can only be validated with a worst-case delay analysis if the conditions of the shaping-for-free strictly

hold (Theorem 2). If the globally-FIFO property of any network element is lost (due, for example, to parallel switching fabrics or redundancy mechanisms), and if at least three flows enter from the same input port, then the latencies can be unbounded inside the output port that uses ATS (Theorem 3).

7.2. Discussion on the Head-of-Line Blocking in TSN ATS

The main property of the IR that the Spring adversary exploits to generate these unbounded latencies is the head-of-line blocking phenomenon in the IR. In Trajectory 1 in Figure 5, the Spring-controlled system S_1 inverts the first packet of the dash-dotted red flow f_2 and the second packet of the solid blue flow f_1. The consequence of this for the theoretical model of the IR is shown in Figure 12c: The packet of the solid blue flow f_1 is delayed due to shaping, and the packet of the dash-dotted red flow f_2 is blocked by f_1 in the IR's FIFO queue during this delay.

Interestingly, the same head-of-line blocking phenomenon is standardized in the ATS specification, even though TSN's output port *queuing function* is not necessarily FIFO [21, §8.6.6]. Indeed, "*ATS schedulers are organized into ATS scheduler groups*," and "*there is one ATS scheduler group per [input port]*" [21, §8.6.5.6]. The ATS schedulers of one group communicate with each other to ensure that they assign non-decreasing eligibility times within the group.

The consequence of this normative behavior for our example is illustrated in Figure 12c. When the solid blue packet of f_1 arrives at the blue ATS scheduler, it must be delayed because of f_1's shaping curve. It is assigned an eligibility time E_{f_1} in the future. Here, the ATS schedulers for f_1 and f_2 belong to the same ATS scheduler group. Hence, when the dash-dotted red packet of f_2 arrives, it is assigned by the red ATS scheduler an eligibility time E_{f_2} larger than E_{f_1}, even if it could be immediately released according to f_2's shaping curve (*i.e.,* E_{f_2} set to the current time). In the current specification of ATS, the packet of f_2 is blocked by the head-of-line packet of f_1.

This behavior seems to be standardized so that hardware implementations of the TSN queuing function can make assumptions on the monotonicity of eligibility times: "*The organization of ATS schedulers into groups [. . .] permits frames of one group to be queued in FIFO order*" [21, §8.6.5.6, Note 2]. However, recent proposals of hardware implementations, such as *push-in first-out* queues [22] and *rotated gate-control-queues* [23], aim at organizing eligibility-based packet queues efficiently with little to no assumption on the monotonicity of the eligibility times. If these designs prove to be mature, then a TSN switch could use one of them to implement the TSN queuing function in the output port.

This situation is depicted in Figure 12b, where there is no longer any need for the eligibility time E_{f_2} to be larger than E_{f_1}. Figure 12b is not possible with the current ATS specifications, but removing the normative sentence "*There is one ATS scheduler group per [input port]*" [21, §8.6.5.6] and leaving the network engineer to decide the best allocation of groups is sufficient: In Figure 12b, each ATS scheduler is alone in its own group. The best selection of groups then

depends on the hardware implementation of the TSN output port and the FIFO assumptions for upstream systems.

With the same principle as the equivalence proof in [19], the ATS configuration of Figure 12b is equivalent to the theoretical model of Figure 12a with per-flow regulators. Hence, for this configuration, context-agnostic network-calculus service curves are available to perform the worst-case analysis of a TSN network with switches that are configured as in Figure 12b (Proposition 1). As opposed to the configuration in Figure 12d, there exists no head-of-line blocking situation for the Spring adversary to generate unbounded latencies.

8. Formal Description of Spring and Proof of Instability

In this section, we formally define the Spring adversary at the core of most of the negative results in our paper. For this, we first introduce the formal notations of packet sequences $\mathscr{M}^x = (M^x, L^{x,\mathtt{M}}, F^{x,\mathtt{M}})$ using marked-point processes, and we recall the formal model of traffic regulators of [12]. Then, we provide the formal proof of Theorem 3 (*Instability of the IR placed after a non-FIFO system*) by relying on a first Spring-generated trajectory. Last, we provide the formal proof of Theorem 4 (*The IR has no fluid service curve that explains its shaping-for-free property*) that relies on a few variants of this trajectory.

8.1. Formal Notations for Describing Spring

The Spring adversary relies on the formal model of the traffic regulators as introduced in [5]. Hence, to describe Spring, we use the marked-point-process notation of [12, §II.A] as a formal description of the packet sequences $\mathscr{M}^x$ introduced in Section 4. Table 3 lists the extended formal notations used in this section and completes Table 1.

8.1.1. Notation with Marked-Point Processes

For a trajectory x and for a packetized observation point $\mathtt{M}$ in the network $\mathcal{N}$, we define the packet sequence $\mathscr{M}^x = (M^x, L^{x,\mathtt{M}}, F^{x,\mathtt{M}})$ as in [12, §II.A]: $M^x = (M_1^x, M_2^x, \dots)$ is the sequence of packet arrival time instants at $\mathtt{M}$, in chronological order. $L^{x,\mathtt{M}} = (L_1^{x,\mathtt{M}}, L_2^{x,\mathtt{M}}, \dots)$ is the sequence of packet lengths, in chronological order. $F^{x,\mathtt{M}} = (F_1^{x,\mathtt{M}}, F_2^{x,\mathtt{M}}, \dots)$ is the sequence of flow indexes: $F_n^{x,\mathtt{M}} = f$ if and only if the n-th packet that crosses $\mathtt{M}$ in trajectory x belongs to f. For another observation point $\mathtt{A}$, we note $\mathscr{A}^x = (A^x, L^{x,\mathtt{A}}, F^{x,\mathtt{A}})$ the packet sequence at $\mathtt{A}$.

For two infinite sequence of real numbers $A = (A_n)_{n\in\mathbb{N}^*}$, $B = (B_n)_{n\in\mathbb{N}^*}$, we note $B \geq A$ if $\forall n \in \mathbb{N}^*$, $B_n \geq A_n$.

8.1.2. Mapping the Notation to Cumulative Arrival Functions

For a packetized observation point $\mathtt{M}$ and its packet sequence $\mathscr{M}^x$ defined in Section 8.1.1, we denote by $\mathscr{R}(\mathscr{M}^x)$ the cumulative arrival function:

$$\mathscr{R}(\mathscr{M}^x) : t \mapsto \sum_{n\in\mathbb{N}^*} L_n^{x,\mathtt{M}} \mathbb{1}_{\{M_n^x < t\}} \tag{11}$$

Table 3: Extended Formal Notations for the Description of Spring

	Common Operators	
$\mathfrak{f}^{\downarrow}$	$w \mapsto \sup\{s \geq 0 \vert \mathfrak{f}(s) \geq w\}$	Pseudo-inverse
$h(\mathfrak{f}, \mathfrak{g})$	$\sup_{t>0}\{\inf\{d \geq 0 \vert \mathfrak{f}(t) \leq \mathfrak{g}(t+d)\}\}$	Horizontal deviation
	Common Curves	
$\gamma_{r,b}$	$t \mapsto \begin{cases} 0 & \text{if } t = 0 \\ rt + b & \text{if } t > 0 \end{cases}$	Leaky-bucket curve.
$\beta_{R,T}$	$t \mapsto R[t-T]^+$	Rate-latency curve.
δ_W	$t \mapsto \begin{cases} 0 & \text{if } t \leq W \\ +\infty & \text{otherwise} \end{cases}$	Bounded-delay curve.
	Notation with Marked-Point Processes	
$\mathtt{M}$	An observation point	
M^x	Sequence of packet arrival dates...	
$L^{x,\mathtt{M}}$	Sequence of packet sizes...	
$F^{x,\mathtt{M}}$	Sequence of flow indexes...	
$\mathscr{M}^x$	$\triangleq (M^x, L^{x,\mathtt{M}}, F^{x,\mathtt{M}})$	Packet sequence...
	...at the packetized observation point $\mathtt{M}$ in trajectory x	
	Mapping of the Notations	
$\mathscr{R}$	$\mathscr{M}^x \mapsto R^{x,\mathtt{M}}$	Maps the packet sequence...
	... $\mathcal{M}^x$ at observation point $\mathtt{M}$ to the cumulative function $R^{x,\mathtt{M}}$	
	at the same observation point $\mathtt{M}$ in the same trajectory x.	
	Traffic-Regulator Model [12]	
$\Pi^{\gamma_{r,b}}$	Pi-operator (13) associated with shaping curve $\gamma_{r,b}$	
$\text{index}(n, f)$	Index of the n-th packet in the subsequence that only contains packets of the flow f.	

where $\mathbb{1}_{\{\texttt{cond}\}}$ equals 1 when `cond` is true, 0 otherwise. From the definition of the packet sequence, we have $\forall x, \forall \texttt{M}$, $\mathscr{R}(\mathscr{M}^x) = R^{x,\texttt{M}}$ is the cumulative arrival function of the aggregate at `M` in trajectory x, as defined in Section 4.

8.1.3. Formal Model For Traffic Regulators

We briefly recall the input-output characterization of traffic regulators from Le Boudec [12], that we restrain to leaky-bucket shaping curves.

The PFR for the single flow f, configured with a leaky-bucket shaping curve $\sigma_f = \gamma_{r_f,b_f}$, is a causal, lossless and FIFO system (Figure 4) that transforms, for any acceptable trajectory x, the input packet sequence $\mathscr{B}^x = (B^x, L^{x,\texttt{B}}, F^{x,\texttt{B}})$ with $F^{x,\texttt{B}} = \{f, f \dots\}$ into the output packet sequence $\mathscr{D}^x = (D^x, L^{x,\texttt{D}}, F^{x,\texttt{D}})$ with $L^{x,\texttt{D}} = L^{x,\texttt{B}}$, $F^{x,\texttt{D}} = F^{x,\texttt{B}}$ and $\forall n \in \mathbb{N}^*$

$$D_n^x = \max\left(B_n^x, D_{n-1}^x, \Pi^{\gamma_{r_f,b_f}}(D^x, L^{x,\texttt{D}})_n\right) \tag{12}$$

By convention, $\forall x, D_0^x = 0$. The Π operator $\Pi^{\gamma_{r_f,b_f}}$ associated with the shaping curve γ_{r_f,b_f} is defined in [12, §III.A] for two sequences D, L by, $\forall n \in \mathbb{N}^*$

$$\left(\Pi^{\gamma_{r_f,b_f}}(D,L)\right)_n = \max_{1\le m\le n-1}\left\{D_m + \gamma^{\downarrow}_{r_f,b_f}\left(\sum_{j=m}^{n} L_j\right)\right\} \tag{13}$$

$\gamma^{\downarrow}_{r_f,b_f} : w \mapsto [w-b]^+ \cdot \frac{1}{r}$ is the pseudo-inverse of γ_{r_f,b_f} (see Table 3).

The IR for the aggregate $\mathcal{F}$, configured with the leaky-bucket shaping curves $\{\gamma_{r_f,b_f}\}_{f\in\mathcal{F}}$ is a causal, lossless and FIFO system (Figure 4) that transforms the input packet sequence $\mathscr{B}^x = (B^x, L^{x,\texttt{B}}, F^{x,\texttt{B}})$ into the output sequence $\mathscr{D}^x = (D^x, L^{x,\texttt{D}}, F^{x,\texttt{D}})$ with $L^{x,\texttt{D}} = L^{x,\texttt{B}}$, $F^{x,\texttt{D}} = F^{x,\texttt{B}}$ and $\forall n \in \mathbb{N}^*$

$$D_n^x = \max\left(B_n^x, D_{n-1}^x, \Pi^{\gamma_{r_f,b_f}}([D^x]^f, [L^{x,\texttt{D}}]^f)_{\mathrm{index}(n,f)}\right) \tag{14}$$

where $f = F_n^{x,\texttt{D}}$ is the flow $f \in \mathcal{F}$ that owns the n-th packet that crosses the IR and $\mathrm{index}(n,f)$ is the index of the n-th packet in the sequence that contains only the packets of f. By convention $\forall x, D_0^x = 0$. Equation (14) is similar to (12), except that the flow $f \in \mathcal{F}$ that defines the applied shaping curve γ_{r_f,b_f} changes at every new n and the Π operator is applied only to the subsequences $[D^x]^f$, $[L^{x,\texttt{D}}]^f$, obtained from D^x, $L^{x,\texttt{D}}$ by keeping only the packets that belong to f, *i.e.,* to the same flow as the current packet n.

8.2. Description of Spring for Proving Theorem 3

Proof of Theorem 3. Consider an IR that processes at least three flows with the same leaky-bucket shaping curve for three of them: $\forall f_i \in \{f_1, f_2, f_3\}$, $\sigma_{f_i} = \gamma_{r,b}$ with $r > 0$ and b greater than the maximum packet size of f_1, f_2, f_3.

Let the adversary Spring define the constants I, d, ϵ and τ as in Equation (5):

$$I \triangleq \frac{b}{r} \;;\; 0 < d < \min(I, W) \;;\; 0 < \epsilon < \min(I-d, \frac{d}{3}) \;;\; \tau \triangleq 3I + 3\epsilon - d$$

Table 4: Packet Sequences Used by Spring

	$(k-1)^{\text{th}}$ period		k^{th} period						$(k+1)^{\text{th}}$ period	
$n:$	$\ldots,$	$6(k-1)+6$	$6k+1$	$6k+2$	$6k+3$	$6k+4$	$6k+5$	$6k+6$	$6(k+1)+1$	$\ldots$
$A^1 = A^2 =$	$(\ldots,$	$3I+2\epsilon+(k-1)\tau,$	$d+k\tau,$	$I+\epsilon+k\tau,$	$I+d+k\tau,$	$2I+\epsilon+k\tau,$	$2I+2\epsilon+k\tau,$	$3I+2\epsilon+k\tau,$	$d+(k+1)\tau,$	$\ldots)$
$F^{1,\mathtt{A}} = F^{2,\mathtt{A}} =$	$(\ldots,$	$f_3,$	$f_1,$	$f_2,$	$f_1,$	$f_2,$	$f_3,$	$f_3,$	$f_1,$	$\ldots)$
$B^1 = B^2 = B^3 =$	$(\ldots,$	$3I+2\epsilon+(k-1)\tau,$	$2d+k\tau,$	$I+d+k\tau,$	$I+\epsilon+d+k\tau,$	$2I+\epsilon+d+k\tau,$	$2I+2\epsilon+d+k\tau,$	$3I+2\epsilon+d+k\tau,$	$2d+(k+1)\tau,$	$\ldots)$
$F^{1,\mathtt{B}} = F^{3,\mathtt{B}} =$	$(\ldots,$	$f_3,$	$f_1,$	$f_1,$	$f_2,$	$f_2,$	$f_3,$	$f_3,$	$f_1,$	$\ldots)$
$F^{2,\mathtt{B}} =$	$(\ldots,$	$f_3,$	$f_1,$	$f_2,$	$f_1,$	$f_2,$	$f_3,$	$f_3,$	$f_1,$	$\ldots)$
$D^1 = D^3 \geq$	$(\ldots,$	$B_1^1+3(k-1)I+3I,$	$B_1^1+3kI,$	$B_1^1+3kI+I,$	$B_1^1+3kI+I,$	$B_1^1+3kI+2I,$	$B_1^1+3kI+2I,$	$B_1^1+3kI+3I,$	$B_1^1+3(k+1)I,$	$\ldots)$

We consider the network $\mathcal{N}_1$ of Figure 5, obtained by concatenating the Spring-controlled source ϕ, the Spring-controlled system S_1 and the IR (not controlled by Spring). We denote by $x = 1$ the trajectory on the network $\mathcal{N}_1$ that results from the choices of Spring, and we formally describe it by using the packet sequences of the form $\mathscr{M}^1 = (M^1, L^{1,\mathtt{M}}, F^{1,\mathtt{M}})$ for an observation point $\mathtt{M}$.

The Spring-controlled source ϕ generates at observation point $\mathtt{A}$ in Figure 5 the packet sequence $\mathscr{A}^1 = (A^1, L^{1,\mathtt{A}}, F^{1,\mathtt{A}})$ defined in Table 4. It comprises a sub-sequence of six packets that repeats every τ seconds. All packets have the same size[3] b, thus $L^{1,\mathtt{A}} = (b, b, b, \dots)$.

By using the definitions of I, d, ϵ and τ, one can verify that the sequence A^1 in Table 4 is increasing. For example,

$$A^1_{6k+2} - A^1_{6k+1} = I + \epsilon + k\tau - d - k\tau = I + \epsilon - d \geq \epsilon > 0$$

because $d < I$ by (5). We can also compute the minimum distance between any two packets of the same flow. For f_1:

$$A^1_{6k+3} - A^1_{6k+1} = I = \frac{b}{r}$$

and

$$A^1_{6(k+1)+1} - A^1_{6k+3} = \tau - I = 2I + 3\epsilon - d \geq I + 3\epsilon \geq \frac{b}{r} \quad \text{because } d < I \quad (15)$$

We obtain similar results for f_2 and f_3. This proves that the minimum distance between any two packets of the same flow is $\frac{b}{r}$. Hence, each flow f_i is $\gamma_{r,b}$-constrained at its source ϕ and Property 1 of Theorem 3 holds.

Remark: Inequation (15) explains why three flows are necessary for Spring. Indeed, with three flows, the duration of the six-packet-long profile that is repeated by Spring is $\tau = 3I + 3\epsilon - d$ (Equation (5)). In this case, the duration $A^1_{6k+3} - A^1_{6k+1} = \tau - I$ is strictly larger than $I = \frac{b}{r}$. This ensures that Property 1 of Theorem 3 holds. With only two flows, Spring would generate a four-packet-long profile of duration $\tau' = 2I + 2\epsilon - d$. We observe that $\tau' - I$ is not larger than I: Property 1 of Theorem 3 does not hold with Spring and only two flows.

The Spring-controlled system S_1 outputs the packet sequence $\mathscr{B}^1$ defined by $\mathscr{B}^1 = (B^1, L^{1,\mathtt{B}}, F^{1,\mathtt{B}})$ with $L^{1,\mathtt{B}} = (b, b, b, \dots)$ and B^1, $F^{1,\mathtt{B}}$ defined in Table 4. For any $k \geq 0$, the $(6k + 2)$th packet in sequence $\mathscr{A}^1$ is now the $(6k + 3)$th packet in $\mathscr{B}^1$ and *vice versa*. This is reflected by the sequence $F^{1,\mathtt{B}}$ that differs from $F^{1,\mathtt{A}}$.

By comparing $F^{1,\mathtt{A}}$ and $F^{1,\mathtt{B}}$, we observe that the system S_1 is causal, lossless and FIFO-per-flow, *i.e.,* Property 2 of Theorem 3 holds. One can also verify from Table 4 that the delay of every packet through S_1 is lower-bounded by 0 and upper-bounded by d. This is clear for most packets. We do it for those

[3] If b is strictly greater than the maximum packet size, then we send two or more packets at the same time and such that their lengths sum up to b.

that exchange their order: $B^1_{6k+3} - A^1_{6k+2} = d$ and $B^1_{6k+2} - A^1_{6k+3} = 0$. Hence, Property 3 holds.

The sequence of packets $\mathscr{B}^1$ is now the input of the IR in Figure 5. The output $\mathscr{D}^1$ is obtained from $\mathscr{B}^1$ by applying the model of the IR from Section 8.1.3. Note that Spring does not control the IR. Let $\mathscr{D}^1 = (D^1, L^{1,\mathtt{D}}, F^{1,\mathtt{D}})$ be the packet sequence at the output of the IR. As the IR is causal, lossless and FIFO, we have $L^{1,\mathtt{D}} = L^{1,\mathtt{B}}$ and $F^{1,\mathtt{D}} = F^{1,\mathtt{B}}$. Furthermore,

Lemma 2. For any $h \in \mathbb{N}$, $D^1_{2h+1} \geq B^1_1 + hI$, and
$D^1_{2h+2} \geq B^1_1 + hI + I$

Proof of Lemma 2. We prove this by induction.

Base case $h = 0$. We have $F^{1,\mathtt{B}}_1 = F^{1,\mathtt{B}}_2 = f_1$. From Equation (14),

$$\begin{aligned} D^1_1 &= \max\left(B^1_1, D^1_0, \Pi^{\gamma_{r,b}}([D^1]^{f_1}, [L^{1,\mathtt{D}}]^{f_1})_{\mathrm{index}(1,f_1)}\right) \\ &\geq B^1_1 \\ D^1_1 &\geq B^1_1 + 0 \cdot I \end{aligned} \tag{16}$$

In addition,

$$\begin{aligned} D^1_2 &= \max\left(B^1_2, D^1_1, \Pi^{\gamma_{r,b}}([D^1]^{f_1}, [L^{1,\mathtt{D}}]^{f_1})_{\mathrm{index}(2,f_1)}\right) \\ &\geq \Pi^{\gamma_{r,b}}([D^1]^{f_1}, [L^{1,\mathtt{D}}]^{f_1})_{\mathrm{index}(2,f_1)} \\ &\geq D^1_1 + \gamma^{\downarrow}_{r,b}(L^{1,\mathtt{D}}_1 + L^{1,\mathtt{D}}_2) && \triangleright\ (13) \text{ with } m = 1 \\ &= D^1_1 + [2b - b]^+ \cdot \frac{1}{r} \\ &= D^1_1 + I \\ &\geq B^1_1 + I && \triangleright\ (16) \\ D^1_2 &\geq B^1_1 + 0 \cdot I + I \end{aligned} \tag{17}$$

Induction step. Consider $h \in \mathbb{N}$, assume that

$$D^1_{2h+1} \geq B^1_1 + hI \tag{18}$$

$$\text{and} \quad D^1_{2h+2} \geq B^1_1 + hI + I \tag{19}$$

We have

$$\begin{aligned} D^1_{2(h+1)+1} &= D^1_{2h+3} \\ &\geq D^1_{2h+2} && \triangleright\ (14) \\ &\geq B^1_1 + (h+1)I && \triangleright\ (19) \end{aligned} \tag{20}$$

Now denote $f \triangleq F^{1,\mathtt{B}}_{2h+4}$. The index $2h+4$ is even, hence from the description of the Spring trajectory (Table 4), we have that $F^{1,\mathtt{B}}_{2h+3}$ also equals f. Hence,

$$\begin{aligned}
D^1_{2h+4} &\geq \Pi^{\gamma_{r,b}}\left([D^1]^f, [L^{1,\mathtt{D}}]^f\right)_{\text{index}(2h+4,f)} && \triangleright\ (14)\\
&\geq D^1_{2h+3} + \gamma^{\downarrow}_{r,b}(L^{1,\mathtt{D}}_{2h+3} + L^{1,\mathtt{D}}_{2h+4}) && \triangleright\ (13) \text{ with } m = 2h+3\\
&= D^1_{2h+3} + [2b-b]^+ \cdot \frac{1}{r}\\
D^1_{2(h+1)+2} &\geq B^1_1 + (h+1)I + I && \triangleright\ (20) \qquad (21)
\end{aligned}$$

Equation (20) and (21) conclude the induction step of the proof for Lemma 2. □

Now, for $k \in \mathbb{N}$,

$$\begin{aligned}
D^1_{6k+1} - B^1_{6k+1} &\geq B^1_1 + 3kI - B^1_{6k+1} && \triangleright\ \text{Lemma 2}\\
&= B^1_1 + 3kI - 2d - k\tau && \triangleright\ \text{Table 4}\\
&= B^1_1 + 3kI - 2d - 3kI - 3k\epsilon + kd && \triangleright\ (5)\\
&= k(d-3\epsilon) + B^1_1 - 2d\\
D^1_{6k+1} - B^1_{6k+1} &\geq k(d-3\epsilon) + B^1_1 - 2d
\end{aligned}$$

As $\epsilon < \frac{d}{3}$, we obtain $\sup_{k\in\mathbb{N}} D^1_{6k+1} - B^1_{6k+1} = +\infty$. We can do the same for the other indices $6k+2, 6k+3, \ldots, 6k+6$. Finally, this gives $\sup_{n\in\mathbb{N}^*} D^1_n - B^1_n = +\infty$ and Property 4 of Theorem 3 holds.

This proves that the Spring adversary described in Table 4 meets the constraints of Theorem 3, does not control the IR of Figure 5, but generates unbounded latencies within the interleaved regulator.

□

8.3. Description of the Spring Variants used in Theorem 4

Proof of Theorem 4. Consider an IR that shapes at least three flows $\mathcal{F}$ with the same leaky-bucket shaping curve for three of them $\forall f \in \{f_1, f_2, f_3\} \subset \mathcal{F}$, $\sigma_f = \gamma_{r,b}$ with $b \geq \max_{f\in\{f_1,f_2,f_3\}} L^{\max}_f$. Assume that there exists a wide-sense increasing curve $\beta^{\boldsymbol{\sigma}}$ such that (a) $\beta^{\boldsymbol{\sigma}}$ is a fluid service-curve of the IR and (b) $\beta^{\boldsymbol{\sigma}}$ explains the shaping-for-free. Select an arbitrary $W > 0$, and consider a causal, lossless, FIFO system S_2 that offers the service curve δ_W. Consider now the network $\mathcal{N}_2$ shown in Figure 13. Assume that the flows f_1, f_2, f_3 are $\gamma_{r,b}$-constrained at $\mathtt{A}$.

We define the Spring constants I, d, ϵ, τ as in (5). We define Trajectory 2 with $\mathscr{A}^2 \triangleq \mathscr{A}^1$, where the supperscript 1 denotes Trajectory 1 described in Table 4, and $\mathscr{B}^2 \triangleq (B^2, L^{2,\mathtt{B}}, F^{2,\mathtt{B}})$ with $B^2 \triangleq B^1$, $L^{2,\mathtt{B}} \triangleq L^{1,\mathtt{B}}$. However, $F^{2,\mathtt{B}}$ differs from $F^{1,\mathtt{B}}$ by exchanging two packets in every period, as shown in Table 4. We define the packet sequence $\mathscr{D}^2$ as the result of the input-output characterization of the IR (Section 8.1.3) when it processes $\mathscr{B}^2$ as an input.

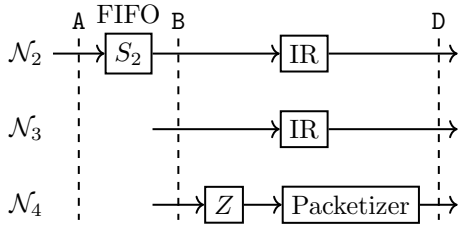


Figure 13: Several networks and their observation points used in the Proof of Theorem 4.

One can check that Trajectory 2 is acceptable for the given constraints. For example, $\mathscr{A}^2 = \mathscr{A}^1$, hence we know from Section 8.2 that each flow is $\gamma_{r,b}$-constrained at A in Trajectory 2.

In addition, $F^{2,\mathtt{B}}$ differs only slighty from $F^{1,\mathtt{B}}$, so we only need to recompute the delay through S_2 for the two packets with a different order: Packet $6k+2$ has a delay $B^2_{6k+2} - A^2_{6k+2} = d - \epsilon < d$ through S_2 and Packet $6k+3$ has a delay $B^2_{6k+3} - A^2_{6k+3} = \epsilon < d$ (Table 4). Consequently, the delay through S_2 in Trajectory 2 is upper-bounded by d. By (5), $d < W$, so Trajectory 2 is acceptable with respect to S_2's service-curve constraint δ_W.

Last, it is clear from Table 4 that the order of the packets is preserved between A and B in Trajectory 2, so Trajectory 2 is acceptable with respect to S_2's FIFO property. Also, $\mathscr{R}(\mathscr{B}^2)$ is constrained by the arrival curve $3\gamma_{r,b} \oslash \delta_W$ [5, Thm. 1.4.3].

Now we consider the network $\mathcal{N}_3$ shown in Figure 13, and we define Trajectory 3 by $\mathscr{B}^3 \triangleq \mathscr{B}^1$. In particular, we have $B^3 = B^2$.

The operator $\mathscr{R} : (M^x, L^{x,\mathtt{M}}, F^{x,\mathtt{M}}) \mapsto R^{x,\mathtt{M}}$ defined in Section 8.1.2 that returns the cumulative function $R^{x,\mathtt{M}}$ of the traffic as a function of the packet sequence $(M^x, L^{x,\mathtt{M}}, F^{x,\mathtt{M}})$, depends only on M^x and $L^{x,\mathtt{M}}$, thus

$$\mathscr{R}(\mathscr{B}^3) = \mathscr{R}(\mathscr{B}^2) \sim 3\gamma_{r,b} \oslash \delta_W \tag{22}$$

We now use the fact that $\beta^{\boldsymbol{\sigma}}$ is a fluid service curve of the IR. By Definition 2, the network $\mathcal{N}_3$ can be replaced by the network $\mathcal{N}_4$ and for the same input packet sequence $\mathscr{B}^4 \triangleq \mathscr{B}^3$, we have the same output packet sequence $\mathscr{D}^4 = \mathscr{D}^3$. In particular, for any $n \in \mathbb{N}^*$,

$$D^3_n - B^3_n = D^4_n - B^4_n \tag{23}$$

In $\mathcal{N}_4$, the packetizer does not increase the per-packet delay and Z offers the service-curve $\beta^{\boldsymbol{\sigma}}$. In addition, $\mathscr{B}^4 = \mathscr{B}^3$ so by Equation (22), $3\gamma_{r,b} \oslash \delta_W$ is an arrival curve for the aggregate at B in $\mathcal{N}_4$. By Definition 2, Z is also causal, lossless and FIFO, hence from [5, Thm 1.4.2] we obtain, $\forall n \in \mathbb{N}^*$,

$$\begin{aligned} D^3_n - B^3_n &= D^4_n - B^4_n && (24)\\ &\le h\left(3\gamma_{r,b} \oslash \delta_W, \beta^{\boldsymbol{\sigma}}\right) && (25)\\ &= h\left(3\gamma_{r,b}, \delta_W \otimes \beta^{\boldsymbol{\sigma}}\right) \qquad \triangleright\ [5, \text{Thm } 3.1.12] && (26) \end{aligned}$$

where $h(\cdot, \cdot)$ denotes the horizontal deviation between two curves (Table 3).

In addition, $\beta^{\boldsymbol{\sigma}}$ explains the shaping-for-free. By Lemma 1, we have $\beta^{\boldsymbol{\sigma}} \geq \sum_{f \in \mathcal{F}} \sigma_f \geq 3\gamma_{r,b}$. Combined with [5, Lemma 1.5.2], we obtain

$$D_n^3 - B_n^3 \leq h\left(3\gamma_{r,b}, \delta_W\right) \tag{27}$$

$$\leq W \tag{28}$$

But the packet sequence $\mathscr{B}^3$ in $\mathcal{N}_3$ is equal to the packet sequence $\mathscr{B}^1$ in the network $\mathcal{N}_1$ of Section 8.2 and Table 4. Hence, $\mathscr{D}^1$ is equal to $\mathscr{D}^3$. We have a contradiction because we showed in Section 8.2 that $\sup_n D_n^1 - B_n^1 = +\infty$. □

9. Conclusion

Network calculus is a framework for obtaining worst-case performance bounds of time-sensitive networks, as required for their validation. Most of the mechanisms standardized by the time-sensitive networking (TSN) task group of the IEEE enjoy a network-calculus service-curve model published in the literature. The interleaved regulator (IR), standardized as *asynchronous traffic shaping* (ATS) in TSN, is an exception. Its shaping-for-free property is instrumental in designing and analyzing time-sensitive networks but was proved without network-calculus service curves. The existence of a service-curve model that explains the IR's behavior and its shaping-for-free property remained an open question. If such a model existed, network engineers could use the IR outside of the shaping-for-free requirements and still compute end-to-end performance bounds with service-curve-oriented tools.

In this paper, we settled the question: Network-calculus service curves cannot explain the behavior of the IR. We show that the IR still offers non-trivial functions as (strict) service curves, but (a) none of them can explain the shaping-for-free property of the IR and (b) these curves are too weak to be helpful and cannot offer any delay guarantee in most cases. Consequently, performance bounds cannot be obtained with service-curve-oriented approaches when the IR is used in a context that differs from the shaping-for-free requirements, *e.g.,* after a non-FIFO system. We prove that these bounds do not even exist: Our Spring adversary can yield unbounded latencies in an IR placed after a non-FIFO system. This instability also affects TSN ATS, but removing a single line from the IEEE TSN normative requirements tackles the issues raised in this paper.

References

[1] J. Farkas, TSN Basic Concepts (Nov. 2018).

[2] IEEE Standard for Local and Metropolitan Area Networks–Bridges and Bridged Networks - Amendment 34:Asynchronous Traffic Shaping, IEEE Std 802.1Qcr-2020 (2020) 1–151doi:10.1109/IEEESTD.2020.9253013.

[3] E. Mohammadpour, E. Stai, M. Mohiuddin, J.-Y. Le Boudec, Latency and Backlog Bounds in Time-Sensitive Networking with Credit Based Shapers and Asynchronous Traffic Shaping, in: 2018 30th International Teletraffic Congress (ITC 30), Vol. 02, 2018, pp. 1–6. doi:10.1109/ITC30.2018.10053.

[4] L. Zhao, P. Pop, S. Steinhorst, Quantitative Performance Comparison of Various Traffic Shapers in Time-Sensitive Networking, IEEE Transactions on Network and Service Management 19 (3) (2022) 2899–2928. doi:10.1109/TNSM.2022.3180160.

[5] J.-Y. Le Boudec, P. Thiran, Network Calculus, Vol. 2050 of Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2001. doi:10.1007/3-540-45318-0.

[6] C.-S. Chang, Performance Guarantees in Communication Networks, Springer London, 2000. doi:10.1007/978-1-4471-0459-9.

[7] A. Bouillard, M. Boyer, E. Le Corronc, Deterministic Network Calculus, John Wiley & Sons, Inc., 2018. doi:10.1002/9781119440284.

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

[9] L. Zhao, P. Pop, Z. Zheng, H. Daigmorte, M. Boyer, Latency Analysis of Multiple Classes of AVB Traffic in TSN with Standard Credit Behavior using Network Calculus (May 2020). arXiv:2005.08256, doi:10.48550/arXiv.2005.08256.

[10] J. Specht, S. Samii, Urgency-Based Scheduler for Time-Sensitive Switched Ethernet Networks, in: 2016 28th Euromicro Conference on Real-Time Systems (ECRTS), 2016, pp. 75–85. doi:10.1109/ECRTS.2016.27.

[11] L. Thomas, J.-Y. Le Boudec, A. Mifdaoui, On Cyclic Dependencies and Regulators in Time-Sensitive Networks, in: 2019 IEEE Real-Time Systems Symposium (RTSS), 2019, pp. 299–311. doi:10.1109/RTSS46320.2019.00035.

[12] J.-Y. Le Boudec, A Theory of Traffic Regulators for Deterministic Networks With Application to Interleaved Regulators, IEEE/ACM Transactions on Networking 26 (6) (2018) 2721–2733. doi:10.1109/TNET.2018.2875191.

[13] J. B. Schmitt, F. A. Zdarsky, The DISCO network calculator: A toolbox for worst case analysis, in: Proceedings of the 1st International Conference on Performance Evaluation Methodolgies and Tools, Valuetools '06, Association for Computing Machinery, New York, NY, USA, 2006, pp. 8–es. doi:10.1145/1190095.1190105.

[14] J. B. Schmitt, F. A. Zdarsky, I. Martinovic, Improving Performance Bounds in Feed-Forward Networks by Paying Multiplexing Only Once, in: 14th GI/ITG Conference - Measurement, Modelling and Evalutation of Computer and Communication Systems, 2008, pp. 1–15.

[15] S. Bondorf, Better bounds by worse assumptions — Improving network calculus accuracy by adding pessimism to the network model, in: 2017 IEEE International Conference on Communications (ICC), 2017, pp. 1–7. doi:10.1109/ICC.2017.7996996.

[16] S. Bondorf, P. Nikolaus, J. B. Schmitt, Quality and Cost of Deterministic Network Calculus: Design and Evaluation of an Accurate and Fast Analysis, Proceedings of the ACM on Measurement and Analysis of Computing Systems 1 (1) (2017) 16:1–16:34. doi:10.1145/3084453.

[17] F. Geyer, S. Bondorf, Network Synthesis under Delay Constraints: The Power of Network Calculus Differentiability, in: IEEE INFOCOM 2022 - IEEE Conference on Computer Communications, 2022, pp. 1539–1548. doi:10.1109/INFOCOM48880.2022.9796777.

[18] L. Thomas, A. Mifdaoui, J.-Y. Le Boudec, Worst-Case Delay Bounds in Time-Sensitive Networks With Packet Replication and Elimination, IEEE/ACM Transactions on Networking (2022) 1–15doi:10.1109/TNET.2022.3180763.

[19] M. Boyer, Equivalence between the Urgency Based Shaper and Asynchronous Traffic Shaping in Time Sensitive Networking, DROPS-IDN/v2/document/10.4230/LITES.9.1.1 (2024). doi:10.4230/LITES.9.1.1.

[20] A. Hamscher, Using Mathematical Programming to Harden Conjectures on Service Curves (Sep. 2022).

[21] IEEE Standard for Local and Metropolitan Area Networks–Bridges and Bridged Networks, IEEE Std 802.1Q-2022 (Revision of IEEE Std 802.1Q-2018) (2022) 1–2163doi:10.1109/IEEESTD.2022.10004498.

[22] A. Sivaraman, S. Subramanian, M. Alizadeh, S. Chole, S.-T. Chuang, A. Agrawal, H. Balakrishnan, T. Edsall, S. Katti, N. McKeown, Programmable Packet Scheduling at Line Rate, in: Proceedings of the 2016 ACM SIGCOMM Conference, SIGCOMM '16, Association for Computing Machinery, New York, NY, USA, 2016, pp. 44–57. doi:10.1145/2934872.2934899.

[23] S. Ren, B. Liu, R. Meng, C. Wang, J.-Y. Le Boudec, E. Mohammadpour, A. Elfawal, Method for sending data packet, and network device (May 2023).

Appendix A. Formal Proofs

This appendix contains the formal proofs of the results in the paper. The proofs use the notations from Table 1 and Table 3.

Appendix A.1. Proof of Proposition 1

Proof of Proposition 1. We first prove that the PFR offers the fluid service curve $\beta^{\boldsymbol{\sigma}} \triangleq \sigma_f$. A shaping curve σ_f that meets the conditions of Proposition 1 verifies [5, Eq. (1.24)] (see the discussion in [5, §1.7.3]). From [5, Thm. 1.7.3], a PFR configured with such a service curve can be realized as the concatenation of a greedy shaper G with shaping curve σ_f, followed by a packetizer. From [5, Thm 1.5.1], G offers σ_f as a service curve. Hence, the PFR offers the fluid service curve σ_f as defined in Definition 2.

We now prove that σ_f explains the shaping-for-free property. Consider two causal, lossless and FIFO systems S and Z' such that Z' offers the service curve $\beta^{\boldsymbol{\sigma}} = \sigma_f$ (Figure 6). Consider now the network made of S followed by Z' and the subset X of the trajectories on this network such that $\forall x \in X, R_f^{\mathtt{A},x} \sim \sigma_f$. Denote by $W_f^{\mathtt{A}\to\mathtt{B}}$ the worst-case delay over X of the flow f between $\mathtt{A}$ and $\mathtt{B}$ and $W_f^{\mathtt{A}\to\mathtt{C}}$ the worst-case delay over X of the flow f between $\mathtt{A}$ and $\mathtt{C}$.

If $W_f^{\mathtt{A}\to\mathtt{B}} = +\infty$, then by causality $W_f^{\mathtt{A}\to\mathtt{C}} = +\infty$ and the result holds.

Assume that $W_f^{\mathtt{A}\to\mathtt{B}}$ is finite. By causality,

$$W_f^{\mathtt{A}\to\mathtt{C}} \geq W_f^{\mathtt{A}\to\mathtt{B}} \tag{A.1}$$

Within the subset of trajectories X, S offers to f the individual service curve $\beta_{S,f} = \delta_{W_f^{\mathtt{A}\to\mathtt{B}}}$. Hence, the concatenation of S and Z' offers to f the individual service curve $\beta_{S+Z,f} = \beta_{S,f} \otimes \beta_{Z,f} = \delta_{W_f^{\mathtt{A}\to\mathtt{B}}} \otimes \sigma_f$ [5, Thm. 1.4.6]. Within X, σ_f is an arrival curve for f at $\mathtt{A}$. Hence

$$\begin{aligned}
W_f^{\mathtt{A}\to\mathtt{C}} &\leq h(\sigma_f, \delta_{W_f^{\mathtt{A}\to\mathtt{B}}} \otimes \sigma_f) && \triangleright \text{[5, Thm. 1.4.2]} \\
&= h(\sigma_f, \sigma_f \otimes \delta_{W_f^{\mathtt{A}\to\mathtt{B}}}) && \triangleright \otimes \text{ commutative} \\
&= h(\sigma_f, \delta_{W_f^{\mathtt{A}\to\mathtt{B}}}) && \triangleright \text{[5, Lem. 1.5.2]} \\
W_f^{\mathtt{A}\to\mathtt{C}} &\leq W_f^{\mathtt{A}\to\mathtt{B}}
\end{aligned}$$

Combined with (A.1), we obtain the result.

□

Appendix A.2. Proof of Lemma 1

Proof of Lemma 1. Consider a set of shaping curves $\boldsymbol{\sigma} = \{\sigma_f\}_{f\in\mathcal{F}}$. Consider a wide-sense increasing function $\beta^{\boldsymbol{\sigma}}$. Assume that $\beta^{\boldsymbol{\sigma}}$ explains the shaping-for-free property (Definition 3).

Consider the system S that has no delay. S is causal, lossless and FIFO. Consider a causal, lossless and FIFO system Z' defined by, for all trajectory x,

$$R^{x,\mathtt{C}} = \min(R^{x,\mathtt{B}}, \beta^{\boldsymbol{\sigma}}) \tag{A.2}$$

where $R^{x,\mathtt{B}}$ [resp., $R^{x,\mathtt{C}}$] is the cumulative function of the aggregate at the input $\mathtt{B}$ of Z' [resp., at the output $\mathtt{C}$ of Z'], see Figure 6. By Equation (A.2), Z' offers $\beta^{\boldsymbol{\sigma}}$ as a service curve.

Concatenate now S and Z' as in Figure 6 and consider the Trajectory x_0 with $\forall f \in \mathcal{F}, R_f^{x_0,\mathtt{A}} = \sigma_f$. x_0 belongs to the set X defined in Definition 3. In addition, S has no delay so

$$R^{x_0,\mathtt{B}} = R^{x_0,\mathtt{A}} = \sum_{f\in\mathcal{F}} \sigma_f \tag{A.3}$$

By Definition 3, the worst-case delay over the set of trajectories X between $\mathtt{A}$ and $\mathtt{C}$ in Figure 6 equals the worst-case delay over X between $\mathtt{A}$ and $\mathtt{B}$, *i.e.,* zero (by construction of S). By causality, the worst-case delay between $\mathtt{B}$ and $\mathtt{C}$ also equals 0 over X.

As x_0 belongs to X, we have $R^{x_0,\mathtt{B}} = R^{x_0,\mathtt{C}}$. Combined with Equation (A.2), we have $\beta^{\boldsymbol{\sigma}} \geq R^{x_0,\mathtt{B}}$. With Equation (A.3), we finally obtain

$$\beta^{\boldsymbol{\sigma}} \geq \sum_{f\in\mathcal{F}} \sigma_f \tag{A.4}$$

□

Appendix A.3. Proof of Theorem 5

The proof of Theorem 5 relies on the model of traffic regulators from Boyer [19] that applies for an IR with leaky-bucket shaping curves.

It relies on the content of token buckets, one per flow. For a flow $f \in \mathcal{F}$, $\Lambda^x_{f,n}$ is defined as the number of tokens inside the bucket for flow f just after packet n is released from the IR in Trajectory x. The head-of-line packet is released as soon as the bucket for the associated flow contains at least as many tokens as the packet's size. With this model, the release time of packet n (Equation (14)) is [19, §5.3]

$$D^x_n = \max\left(B^x_n, D^x_{n-1}, \frac{L^{x,\mathtt{B}}_n - \Lambda^x_{f,n\ominus 1}}{r_f} + D_{n\ominus 1}\right) \tag{A.5}$$

where $f = F^{x,\mathtt{B}}_n$ is the flow that owns packet n and $n \ominus 1$ is the index of the last packet of f just before n. By convention, for any trajectory x and any flow f, $n \ominus 1 = 0$ if n is the first packet of f, and $\Lambda^x_{f,0}$ equals b_f, the shaping-curve burst configured for flow f. Last, $D^x_0 = 0$.

Boyer's [19] and Le Boudec's [12] models are equivalent for leaky-bucket shaping curves. The following proof uses Boyer's model because the result is faster to obtain based on this model.

Proof of Theorem 5. Consider an IR that processes an aggregate $\mathcal{F}$ of flows with leaky-bucket shaping curves: $\forall f \in \mathcal{F}$, $\sigma_f = \gamma_{r_f,b_f}$ with $r_f > 0$ and $b_f \geq L^{\max}_f$.

Define $L^{\min}$ and $I^{\max}$ as in Theorem 5. Denote by $\mathtt{B}$ [resp., $\mathtt{D}$] an observation point located at the input [resp., at the output] of the IR (as in Figure 4).

Consider an acceptable trajectory x. Consider two time instants $s < t$ such that $(s, t]$ is a backlogged period for the IR and consider $\epsilon \in \mathbb{R}$ such that $0 < \epsilon < t - s$. From the definition of the backlogged period, the cumulative functions at the input and the output verify $R^{x,\mathtt{B}}(s+\epsilon) > R^{x,\mathtt{D}}(s+\epsilon)$.

Both the input and the output of the IR are packetized. Hence, consider now the packet sequences $\mathscr{B}^x = (B^x, L^{x,\mathtt{B}}, F^{x,\mathtt{B}})$ and $\mathscr{D}^x = (D^x, L^{x,\mathtt{D}}, F^{x,\mathtt{D}})$ in trajectory x, at observation point $\mathtt{B}$ and $\mathtt{D}$ (as defined in Section 8.1.1). The IR is causal, lossless and FIFO, so we have $L^{x,\mathtt{B}} = L^{x,\mathtt{D}}$ [resp., $F^{x,\mathtt{B}} = F^{x,\mathtt{D}}$]. Denote by L^x [resp., F^x] this sequence.

At $s+\epsilon$, we have $R^{x,\mathtt{B}}(s+\epsilon) > R^{x,\mathtt{D}}(s+\epsilon)$. The IR is lossless, so at $s+\epsilon$, it is non-empty. In addition, its input and output are packetized, so the IR contains at least one packet at $t+\epsilon$. The IR is also FIFO, so denote by n the index within the sequences $\mathscr{B}^x$, $\mathscr{D}^x$ of the head-of-line packet inside the IR at $s+\epsilon$ and denote $g \triangleq F^x_n$ the flow to which packet n belongs in trajectory x.

We now use Equation (A.5) to compute the release time instant of packet n in trajectory x. At time instant $s+\epsilon$, packet n is the head-of-line packet in the IR, so $B_n \leq s+\epsilon$, $D_{n-1} \leq s+\epsilon$ and $D_{n\ominus 1} \leq s+\epsilon$. Hence,

$$D^x_n \leq \max\left\{ s+\epsilon, \frac{L^x_n - \Lambda^x_{g,n\ominus 1}}{r_g} + s + \epsilon \right\} \tag{A.6}$$

$$\leq \max\left\{ s+\epsilon, \frac{L^x_n}{r_g} + s + \epsilon \right\} \qquad \triangleright\, \Lambda^x_{g,n\ominus 1} \geq 0 \tag{A.7}$$

$$\leq \frac{L^x_n}{r_g} + s + \epsilon \qquad \triangleright\, L^x_n \geq 0 \tag{A.8}$$

With the terminology of [19], Equation (A.8) states that packet n stays at the head of the line for a duration that is less than the time required for the associated token bucket to regain as many credits as the length of n. We obtain

$$D^x_n \leq \frac{L^{\max}_g}{r_g} + s + \epsilon \tag{A.9}$$

$$\leq I^{\max} + s + \epsilon \tag{A.10}$$

The size L^x_n of packet n is greater than $L^{\min}$, hence the cumulative function $R^{x,\mathtt{D}}$ at the output of the IR in trajectory x verifies:

$$R^{x,\mathtt{D}}(t) - R^{x,\mathtt{D}}(s) \geq \begin{cases} 0 & \text{if } t-s \leq I^{\max}+\epsilon \\ L^{\min} & \text{if } t-s > I^{\max}+\epsilon \end{cases} \tag{A.11}$$

$$R^{x,\mathtt{D}}(t) - R^{x,\mathtt{D}}(s) \geq L^{\min} \wedge \left(\delta_{I^{\max}+\epsilon}\right)(t-s) \tag{A.12}$$

where $\wedge$ is the minimum and δ is the bounded-delay curve (Table 3). Equation (A.12) is valid for any ϵ such that $0 < \epsilon < t-s$, hence

$$R^{x,\mathtt{D}}(t) - R^{x,\mathtt{D}}(s) \geq L^{\min} \wedge \left(\delta_{I^{\max}}(t-s)\right) \tag{A.13}$$

Equation (A.13) is valid for any acceptable trajectory x with no assumption on the context of the IR and for any backlogged period $(s,t]$. In addition, note that the function $\beta_0 : t \mapsto L^{\min} \wedge \delta_{I^{\max}}(t)$ is wide-sense increasing. By definition, the IR offers β_0 as a context-agnostic strict service curve. The curve β_0 is shown with a dashed-blue line in Figure 8.

Let us now define the curve β_{sc} as

$$\beta_{\text{sc}} : t \mapsto \left\lfloor \frac{t}{I^{\max}} \right\rfloor \cdot L^{\min} \tag{A.14}$$

where $\lfloor \cdot \rfloor$ denotes the floor function. We observe that β_{sc} is the supper-additive closure of β_0 (Equation (7) in Section 6.1). Consequently, the IR also offers β_{sc} as a strict service curve [7, Prop. 5.6, Item 2].

In addition, any curve β' such that $\forall t \geq 0$, $\beta'(t) \leq \beta_{\text{sc}}$ is also a strict service curve of the IR. This is the case, in particular, for the rate-latency service curve $\beta_{R,T}$ with $R = \frac{L^{\min}}{I^{\max}}$ and $T = I^{\max}$.

□

Appendix A.4. Proof of Proposition 2

Proof of Proposition 2. Consider an IR that processes an aggregate $\mathcal{F}$ with leaky-bucket shaping curves $\{\sigma_f\}_{f\in\mathcal{F}} = \{\gamma_{r_f,b_f}\}_{f\in\mathcal{F}}$. Take a curve $\beta^{\text{strict}} \in \mathfrak{F}_0$ and assume that β^{strict} is a context-agnostic strict service curve of the IR.

Take $f \in \mathcal{F}$, and consider trajectory x_f. Define $I_f \triangleq \frac{b_f}{r_f}$. Let the packet sequence $\mathscr{B}^{x_f}$ at the input $\mathtt{B}$ of the IR be $B^{x_f} = (0, \frac{I_f}{2}, I_f, \frac{3I_f}{2}, 2I_f, \dots)$, $L^{x_f,\mathtt{B}} = (b_f, b_f, b_f, b_f, b_f, \dots)$, $F^{x_f,\mathtt{B}} = (f, f, f, f, f, \dots)$. That is, only flow f sends packets of size b_f at twice the rate allowed by its shaping curve. There are no packets for flows $\mathcal{F}\backslash\{f\}$.

Based on the input-output characterisation of the IR from Equation (14), we directly obtain that the packet sequence $\mathscr{D}^x$ at the output $\mathtt{D}$ of the IR is $L^{x_f,\mathtt{D}} = L^{x_f,\mathtt{B}}$, $F^{x_f,\mathtt{D}} = F^{x_f,\mathtt{B}}$ and $D^{x_f} = (0, I_f, 2I_f, 3I_f, 4I_f, \dots)$.

Hence, σ_f is an arrival curve for the output traffic $R^{x_f,\mathtt{D}}$ and the IR is non-empty after $\frac{I_f}{2}$. By combining (2) and (3), this gives

$$\forall t \geq 0, \quad \beta^{\text{strict}}(t) \leq \sigma_f(t) \tag{A.15}$$

Inequation (A.15) is valid for any $f \in \mathcal{F}$, hence the result.

□

Appendix A.5. Proof of Proposition 3

Proof of Proposition 3. Denote by v the value of $\beta_g(u)$. We have $v > L_g^{\min}$. β_g is an individual service curve of S for g. By definition, it is wide-sense increasing. Hence,

$$\forall t \geq u, \quad \beta_g(t) \geq v > L_g^{\min} \tag{A.16}$$

The red-shaded area in Figure A.14 represents the possible values of β_g.

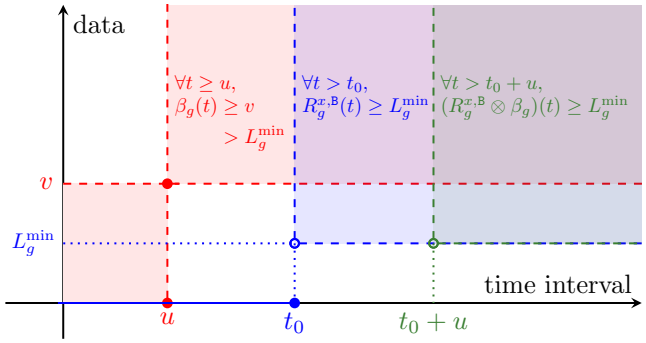


Figure A.14: If we assume that $\exists u \geq 0, \beta_g(u) = v > L_g^{\min}$ as in Proposition 3, then this figure depicts the envelopes of possible values for the *individual* service curve β_g for g (red area), for the cumulative function $R_g^{x,\mathtt{B}}$ of g at the input $\mathtt{B}$ of S with trajectory x (in blue), and for the resulting min-plus convolution $R_g^{x,\mathtt{B}} \otimes \beta_g$ (in green). This figure should be printed with colors.

The observation point $\mathtt{B}$ is packetized (Figure 11). Consider an acceptable trajectory x such that the first packet of g is of size $L_g^{\min}$. Denote by t_0 the arrival time instant of the first packet of g at observation point $\mathtt{B}$ in trajectory x. The cumulative arrival function $R_g^{x,\mathtt{B}}$ of g at $\mathtt{B}$ is wide-sense increasing and left-continuous, hence

$$\forall t > t_0, R_g^{x,\mathtt{B}}(t) \geq L_g^{\min} \tag{A.17}$$

The blue-shaded area in Figure A.14 represents the possible value of $R_g^{x,\mathtt{B}}$.

Consider now the min-plus convolution $R_g^{x,\mathtt{B}} \otimes \beta_g$. For $t > t_0 + u$, we have

$$\begin{aligned}(R_g^{x,\mathtt{B}} \otimes \beta_g)(t) &= \inf_{0 \leq s \leq t} \beta_g(s) + R_g^{x,\mathtt{B}}(t-s) \\ &= \min\left(\inf_{0 \leq s \leq u} \beta_g(s) + R_g^{x,\mathtt{B}}(t-s), \inf_{u < s \leq t} \beta_g(s) + R_g^{x,\mathtt{B}}(t-s)\right)\end{aligned}$$

For $s \in [0, u]$, $\beta_g(s) \geq 0$ and $t - s \geq t - u > t_0$, so $R_g^{x,\mathtt{B}}(t-s) \geq L_g^{\min}$ (A.17). For $s \in (u, t]$, $\beta_g(s) \geq L_g^{\min}$ (A.16) and $R_g^{x,\mathtt{B}}(t-s) \geq 0$. Hence, $\forall t > t_0 + u$,

$$(R_g^{x,\mathtt{B}} \otimes \beta_g)(t) \geq \min(0 + L_g^{\min}, L_g^{\min} + 0) = L_g^{\min} \tag{A.18}$$

By assumption, β_g is an individual service curve of S for g, so

$$\forall t > t_0 + u, \quad R_g^{x,\mathtt{D}}(t) \geq (R_g^{x,\mathtt{B}} \otimes \beta_g)(t) \geq L_g^{\min} \tag{A.19}$$

The green-shaded area in Figure A.14 represents the possible value of $R_g^{x,\mathtt{D}}$. The system S is causal, FIFO and lossless, and its output is packetized. Hence Inequation (A.19) implies that the first packet of g exits S before $t_0 + u$, *i.e.*, its delay through S is upper-bounded by u. □

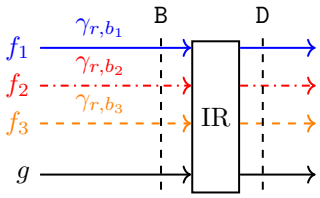


Figure A.15: Situation for the Proof of Theorem 6. The flows f_1, f_2, f_3 are processed by the IR with the same leaky-bucket shaping curve $\gamma_{r,b}$. They enter the IR with leaky-bucket arrival curves γ_{r_1,b_1},γ_{r_2,b_2} and γ_{r_3,b_3}. We are interested in an *individual* service curve that the IR offers to a fourth flow g. Best with colors (not required).

Appendix A.6. Proof of Theorem 6

Proof of Theorem 6. Consider an IR that processes an aggregate $\mathcal{F}$ of at least four flows with the same leaky-bucket shaping curve for three of them: $\forall f_i \in \{f_1, f_2, f_3\}$, $\sigma_{f_i} = \gamma_{r,b}$. Consider a flow $g \in \mathcal{F}\backslash\{f_1, f_2, f_3\}$. For $f_i \in \{f_1, f_2, f_3\}$, assume that the flow f_i is constrained at the input B of the IR by a leaky-bucket arrival curve γ_{r,b_i} with $b_1 > b$, $b_2 \geq b$ and $b_3 \geq b$ (Figure A.15). Assume that any other flow $h \in \mathcal{F}\backslash\{f_1, f_2, f_3, g\}$ exhibits an arrival curve $\alpha_h^{\mathtt{B}}$ at the input B of the interleaved regulator. Finally, assume that the IR offers to g the individual service curve β_g. This curve can depend on the shaping curves $\{\sigma_f\}_{f\in\mathcal{F}}$ and on the arrival curves of the other flows $\alpha_{f_1}^{\mathtt{B}} = \gamma_{r,b_1}$, $\alpha_{f_2}^{\mathtt{B}} = \gamma_{r,b_2}$, $\alpha_{f_3}^{\mathtt{B}} = \gamma_{r,b_3}$, $\{\alpha_h^{\mathtt{B}}\}_{h\in\mathcal{F}\backslash\{f_1,f_2,f_3,g\}}$. Consider X', the subset of trajectories such that for each $x \in X'$, the first packet of g that enters the IR is of size $L_g^{\min}$.

We obtain the proof of Theorem 6 through the following lemma.

Lemma 3. For any duration $M \in \mathbb{R}^+$, there exists an acceptable trajectory $x_M \in X'$ such that the delay of the first packet of g within the IR is greater than M.

Proof of Lemma 3 . For the three flows f_1, f_2, f_3, consider the Trajectory 3 of Table 4 with

$$W \triangleq \frac{b_1}{r} - \frac{b}{r} \tag{A.20}$$

Let the Spring adversary define its parameters I, d, ϵ, τ as in Equation (5). We have $W > 0$ because $b_1 > b$. Note that

$$W = I - \frac{2b}{r} + \frac{b_1}{r} \tag{A.21}$$

In Trajectory 3, the size of the packets is b and the duration between any two consecutive packets of f_2 [resp., f_3] is at least $I = b/r$ at B, so γ_{r,b_2} [resp., γ_{r,b_3}] is an arrival curve of f_2 [resp., f_3] at B in Trajectory 3 (because $b_2 \geq b$ and $b_3 \geq b$). Similarly, the duration between any two packets of f_1 of size b is

$I - d$ and

$$\begin{aligned} -W &< -d && \triangleright\,(5)\\ -I + \frac{2b - b_1}{r} &< -d && \triangleright\,(\text{A.21})\\ 2b &< r(I-d) + b_1 \end{aligned}$$

This shows that γ_{r,b_1} is an arrival curve for f_1 at B in Trajectory 3.

From Trajectory 3, we obtain trajectory x_M by inserting a packet of g, of size $L_g^{\min}$, between the $(6k+1)$th and $(6k+2)$th packets of Trajectory 3, with

$$k \triangleq \left\lceil \frac{M + d + I}{d - 3\epsilon} \right\rceil + 1 \tag{A.22}$$

where $\lceil \cdot \rceil$ is the ceiling function. This packet takes the index $6k+2$ in trajectory x_M, and we have

$$B^3_{6k+1} \leq B^{x_M}_{6k+2} \leq B^3_{6k+2} \tag{A.23}$$

No other flow $h \in \mathcal{F} \backslash \{f_1, f_2, f_3, g\}$ sends any packet in Trajectory x_M. For the three flows f_1, f_2, f_3, Trajectory x_M describes the same packet sequence in B as in Trajectory 3. Hence, $\gamma_{r,b_1}, \gamma_{r,b_2}, \gamma_{r,b_3}$ are also arrival curves of f_1, f_2, f_3 in Trajectory x_M. We define $\mathscr{D}^{x_M}$ as the result of the IR's input-output model (Section 8.1.3) on input $\mathscr{B}^{x_M}$. Hence, x_M is an acceptable trajectory, and, by definition $x_M \in X'$.

The packet sequences $\mathscr{B}^{x_M}$ and $\mathscr{B}^3$ are identical until index $6k+1$ included and the IR is causal. Hence, the packet sequences $\mathscr{D}^{x_M}$ and $\mathscr{D}^3$ are also identical until index $6k+1$ included. As the IR is FIFO, we obtain

$$D^{x_M}_{6k+2} \geq D^{x_M}_{6k+1} = D^3_{6k+1} \tag{A.24}$$

By combining (A.23) and (A.24), we obtain

$$\begin{aligned} D^{x_M}_{6k+2} - B^{x_M}_{6k+2} &\geq D^3_{6k+1} - B^3_{6k+2}\\ &= B^1_1 + 3kI - I - d - k\tau && \triangleright \text{ Table 4}\\ &\geq (-d - I) + k(3I - \tau) && \triangleright\, B^1_1 \geq 0\\ &= -(d+I) + k(d - 3\epsilon) && \triangleright\,(5)\\ &\geq -(d+I) + \frac{M+d+I}{d-3\epsilon}(d - 3\epsilon) && \triangleright\,(\text{A.22})\\ D^{x_M}_{6k+2} - B^{x_M}_{6k+2} &\geq M && (\text{A.25}) \end{aligned}$$

This gives the result of the Lemma. □

To prove Theorem 6, we now simply combine Lemma 3 with the contrapositive of Proposition 3.

Lemma 3 states that for any $u \in \mathbb{R}^+$, there exists a trajectory $x \in X'$ such that the delay of the first packet of g is not upper-bounded u (it is strictly larger than u). This means that the consequent of Proposition 3 is false. By contraposition, the antecedent is also false.

□

Appendix A.7. Proof of Theorem 7

Proof of Theorem 7. Consider an IR that processes an aggregate $\mathcal{F}$ of at least four flows with leaky-bucket shaping shaping curves $\{\sigma_h\}_{h\in\mathcal{F}}$. Assume that at least three flows $f_1, f_2, f_3 \in \mathcal{F}$ share the same leaky-bucket shaping curve: $\forall f_i \in \{f_1, f_2, f_3\}, \sigma_{f_i} = \gamma_{r,b}$. Consider now a function $\beta \in \mathfrak{F}_0$. β can depend on $\{\sigma_h\}_{h\in\mathcal{F}}$. Assume that the IR offers β as a context-agnostic service curve.

For any acceptable trajectory x with no assumption on the context of the IR, we have

$$\forall t \geq 0, \quad R^{x,\mathtt{D}}(t) \geq (R^{x,\mathtt{B}} \otimes \beta)(t) \tag{A.26}$$

where $R^{x,\mathtt{B}} : t \mapsto \sum_{h\in\mathcal{F}} R_h^{x,\mathtt{B}}(t)$ is the cumulative function of the aggregate at $\mathtt{B}$, the IR's input (Figure 4), in Trajectory x. And $R^{x,\mathtt{D}} : t \mapsto \sum_{h\in\mathcal{F}} R_h^{x,\mathtt{D}}(t)$ is the cumulative function at $\mathtt{D}$, the IR's output, in Trajectory x.

In $\mathcal{F}\backslash\{f_1, f_2, f_3\}$, choose arbitrarly one flow and call it g. For a given set of arrival curves $\{\alpha_h^{\mathtt{B}}\}_{h\in\mathcal{F}\backslash\{g\}}$, we define

$$\alpha_{\neg g} \triangleq \sum_{h\in\mathcal{F}\backslash\{g\}} \alpha_h^{\mathtt{B}} \tag{A.27}$$

where $\alpha_h^{\mathtt{B}}$ is an arrival curve for h at the input $\mathtt{B}$ of the IR. For each $\theta \in \mathbb{R}^+$, we also define β_θ as:

$$\beta_\theta : t \mapsto [\beta(t) - \alpha_{\neg g}(t-\theta)]^+ \, \mathbb{1}_{\{t>\theta\}} \tag{A.28}$$

where $\mathbb{1}_{t>\theta}$ equals 1 when $t > \theta$, 0 otherwise. We finally define $\overline{\beta}_\theta$ as:

$$\overline{\beta}_\theta = \beta_\theta \overline{\oslash} \mathbf{0} : t \mapsto \inf_{s\geq t}\{\beta_\theta(s)\} \tag{A.29}$$

where $\overline{\oslash}$ is the max-plus deconvolution (Table 1) and $\mathbf{0}$ is the zero function: $\forall t \geq 0,\ \mathbf{0}(t) = 0$.

Lemma 4. If, for each flow $h \in \mathcal{F}\backslash\{g\}$, $\alpha_h^{\mathtt{B}}$ is an arrival curve of h at the input of the IR, then for each $\theta \geq 0$, the IR offers $\overline{\beta}_\theta$ as an individual service curve for g.

Proof of Lemma 4. The IR is a FIFO system that offers the service curve β to the aggregate. By [5, Prop. 6.4.1], for each $\theta \in \mathbb{R}^+$, the function β_θ verifies Equation (4).

However, for $\theta \in \mathbb{R}^+$, β_θ may not be wide-sense increasing, thus not an individual service curve for g. But $\overline{\beta}_\theta$ is wide-sense increasing by construction and $\forall\theta \in \mathbb{R}^+, \forall t \geq 0, \overline{\beta}_\theta(t) \leq \beta_\theta(t)$. Hence, for $\theta \in \mathbb{R}^+$, $\overline{\beta}_\theta$ also verifies Equation (4), thus is an individual service curve of S for g. □

Take now $\epsilon > 0$ and denote by X_ϵ the subset of acceptable trajectories such that:

- $\gamma_{r,b+\epsilon}$ is an arrival curve for f_1 at $\mathtt{B}$, and
- $\gamma_{r,b}$ is an arrival curve for f_2 at $\mathtt{B}$, and

- $\gamma_{r,b}$ is an arrival curve for f_3 at $\mathtt{B}$, and
- $\forall h \in \mathcal{F}\backslash\{f_1, f_2, f_3, g\}$, $\forall t \geq 0$, $R_h^{x,\mathtt{B}} = R_h^{x,\mathtt{D}} = 0$

For all trajectories of X_ϵ, we have

$$\alpha_{\neg g} = \gamma_{3r,3b+\epsilon} \tag{A.30}$$

By applying Lemma 4, for each $\theta \in \mathbb{R}^+$, $\overline{\beta}_\theta$ is an individual service curve of the IR for g. In addition, $b + \epsilon > b$ so the subset of trajectories X_ϵ meets the assumptions of Theorem 6. By Theorem 6, we have, $\forall \theta \geq 0, \forall t \geq 0$,

$$\begin{aligned}
\overline{\beta}_\theta(t) &\leq L_g^{\min} \\
\inf_{s \geq t}\{\beta_\theta(s)\} &\leq L_g^{\min} \quad \triangleright \text{ (A.29)} \\
\inf_{s \geq t}\left\{[\beta(s) - \alpha_{\neg g}(s-\theta)]^+ \mathbb{1}_{\{s>\theta\}}\right\} &\leq L_g^{\min} \quad \triangleright \text{ (A.28)}
\end{aligned}$$

If we limit to $t > \theta$, we obtain, $\forall \theta \geq 0, \forall t > \theta$,

$$\inf_{s \geq t}\left\{[\beta(s) - \alpha_{\neg g}(s-\theta)]^+\right\} \leq L_g^{\min} \tag{A.31}$$

For any $x \in \mathbb{R}$, $x \leq [x]^+$, hence $\forall \theta \geq 0, \forall t > \theta$,

$$\inf_{s \geq t}\{\beta(s) - \alpha_{\neg g}(s-\theta)\} \leq \inf_{s \geq t}\left\{[\beta(s) - \alpha_{\neg g}(s-\theta)]^+\right\} \tag{A.32}$$

By combining Equations (A.31) and (A.32), we obtain $\forall \theta \geq 0, \forall t > \theta$,

$$\begin{aligned}
\inf_{s \geq t}\left\{\beta(s) - \alpha_{\neg g}(s-\theta)^+\right\} &\leq L_g^{\min} \\
\inf_{s \geq t}\{\beta(s) - \gamma_{3r,3b+\epsilon}(s-\theta)\} &\leq L_g^{\min} \quad \triangleright \text{ (A.30)} \\
\inf_{s \geq t}\{\beta(s) - 3rs\} &\leq L_g^{\min} - 3r\theta + 3b + \epsilon \qquad \text{(A.33)}
\end{aligned}$$

Equation (A.33) is valid for any $\epsilon > 0$. Hence,

$$\forall \theta \geq 0, \forall t > \theta, \quad \inf_{s \geq t}\{\beta(s) - 3rs\} \leq L_g^{\min} - 3r\theta + 3b$$

In particular,

$$\forall t > \frac{L_g^{\min}}{3r} + \frac{b}{r}, \quad \inf_{s \geq t}\{\beta(s) - 3rs\} \leq 0 \tag{A.34}$$

Denote $t_1 \triangleq \frac{L_g^{\min}}{3r} + \frac{b}{r}$. With the definition of the infimum, Equation (A.34)

gives

$$\forall t > t_1, \forall \epsilon' > 0, \exists s \geq t, \qquad \beta(s) - 3rs \leq \epsilon'$$

$$\forall t > t_1, \forall \epsilon' > 0, \exists s \geq t, \qquad \frac{\beta(s)}{s} \leq \frac{\epsilon'}{s} + 3r$$

$$\forall t > t_1, \qquad \inf_{s \geq t} \frac{\beta(s)}{s} \leq 3r$$

$$\text{This shows} \qquad \liminf_{t \to +\infty} \frac{\beta(t)}{t} \leq 3r \tag{A.35}$$

□