INVITED PAPER *Special Section on Recent Progress in Networking Science and Practice in Conjunction with Main Topics of ITC32*

Interleaved Weighted Round-Robin: A Network Calculus Analysis

Seyed Mohammadhossein TABATABAEE[†a)], **Jean-Yves LE BOUDEC**[†b)], ***and*** **Marc BOYER**[††c)], ***Nonmembers***

SUMMARY Weighted Round-Robin (WRR) is often used, due to its simplicity, for scheduling packets or tasks. With WRR, a number of packets equal to the weight allocated to a flow can be served consecutively, which leads to a bursty service. Interleaved Weighted Round-Robin (IWRR) is a variant that mitigates this effect. We are interested in finding bounds on worst-case delay obtained with IWRR. To this end, we use a network calculus approach and find a strict service curve for IWRR. The result is obtained using the pseudo-inverse of a function. We show that the strict service curve is the best obtainable one, and that delay bounds derived from it are tight (i.e., worst-case) for flows of packets of constant size. Furthermore, the IWRR strict service curve dominates the strict service curve for WRR that was previously published. We provide some numerical examples to illustrate the reduction in worst-case delays caused by IWRR compared to WRR.

key words: *weighted round-robin, delay bound, worst-case delay, network calculus, strict service curve*

1. Introduction

Weighted Round-Robin (WRR) is a scheduling algorithm that is often used for scheduling tasks, or packets, in real-time systems or communication networks. The capacity is shared among several clients or queues by giving each of them a weight, which is a positive integer, and by providing more service to those with larger weights. Specifically, queues are visited one after the other, and when a queue i with weight w_i has an emission opportunity, it sends w_i packets, or less if fewer packets are present. The advantage of WRR is that it is fair and simple. However, the service is bursty because up to w_i packets can be served consecutively for queue i, which can cause a large worst-case waiting time for other queues. Interleaved Weighted Round-Robin (IWRR) mitigates this effect [1]. With IWRR, a queue i with weight w_i has w_i emission opportunities per round and can send up to one packet at every emission opportunity. In contrast, with WRR, it has one emission opportunity per round and can send up to w_i packets at every emission opportunity. Hence, IWRR spreads out emission opportunities of each queue in a round, which is expected to result in a smoother service and lower worst-case delays.

Manuscript received December 16, 2020.
Manuscript revised April 16, 2021.
Manuscript publicized July 1, 2021.
†The authors are with EPFL, Switzerland.
††The author is with ONERA/DTIS, Université de Toulouse, France.
a) E-mail: hossein.tabatabaee@epfl.ch
b) E-mail: jean-yves.leboudec@epfl.ch
c) E-mail: marc.boyer@onera.fr
DOI: 10.1587/transcom.2021ITI0001

There exist several versions of IWRR; we focus on the simplest one, where queue i has emission opportunities in the first w_i cycles within a round (see Sect. 3 for a formal description of IWRR and Sect. 4 for WRR variants).

We are interested in delay bounds for the worst case, as is typical in the context of deterministic networking. To this end, a standard approach is network calculus. Specifically, with network calculus, the service offered to a flow of interest by a system is abstracted by means of a service curve. A bound on the worst-case delay is obtained by combining the service curve with an arrival curve for the flow of interest. An arrival curve is a constraint on the amount of data that the flow of interest can send; such a constraint is necessary to the existence of a finite delay bound. The exact definitions are recalled in Sect. 2.

The network calculus approach was applied to WRR in [2, Sec. 8.2.4], where a *strict* service curve is obtained. As explained in Sect. 2, a strict service curve is a special case of a service curve hence can be used to derive delay (and backlog) bounds. Our first contribution is to obtain a strict service curve for IWRR. Compared to WRR, the interleaving in IWRR makes the analysis more difficult, and the method of proof in [2] cannot easily be extended. To circumvent this difficulty, we rely heavily on the method of pseudo-inverse, recalled in Sect. 2. As expected, the IWRR strict service curve dominates that of WRR, hence the resulting delay bounds for IWRR are always less than or equal to those for WRR.

The strict service curve enables us to obtain delay bounds by using network calculus, but such bounds might not always be tight, i.e., they might not always be equal to worst-cases. This is because the strict service curve is an abstraction of the system. Our second contribution is to show that, for flows with packets of constant sizes, the strict service curve obtained for IWRR provides tight delay bounds. We show that the same result holds for the existing strict service curve of WRR. Extending such results to flows with packets of variable sizes is left for further study.

The strict service curve obtained for IWRR has some description complexity, see also Fig. 3. Therefore, we provide simplified lower bounds that can be used, at the expense of sub-optimality, when analytic, closed-form expressions are important.

After giving some necessary background on network calculus and the lower-pseudo inverse technique in Sect. 2, we describe our system model in Sect. 3. We describe the state of the art in Sect. 4. In Sect. 5, we present our strict service curve for IWRR. In Sect. 6, we show that both the IWRR and WRR strict service curves are the best possible and that they give tight delay bounds for a flow with constant packet sizes. We use numerical examples to illustrate the worst-case latency improvement of IWRR over WRR obtained with our method in Sect. 7. We present proofs of results in Sect. 8.

2. Background

We use the framework of network calculus [2]–[4]. A flow is represented by a cumulative arrival function $R \in \mathscr{F}$, where $\mathscr{F}$ denotes the set of wide-sense increasing functions $f : \mathbb{R}^+ \mapsto \mathbb{R}^+ \cup \{+\infty\}$ and $R(t)$ is the number of bits observed on the flow between times 0 and t. We say that a flow R has $\alpha \in \mathscr{F}$ as arrival curve if for all $s \le t$, $R(t) - R(s) \le \alpha(t-s)$. A frequently used arrival curve is $\alpha = \gamma_{r,b}$, defined by $\gamma_{r,b}(t) = rt + b$ for $t > 0$ and $\gamma_{r,b}(t) = 0$ for $t = 0$ (token bucket arrival curve, with rate r and burst b). An arrival curve α can always be assumed to be sub-additive, i.e., to satisfy $\alpha(s+t) \le \alpha(s) + \alpha(t)$ for all s, t.

For two functions f and g in $\mathscr{F}$, the min-plus convolution is defined by $(f \otimes g)(t) = \inf_{0 \le s \le t}\{f(t-s) + g(s)\}$. An example of min-plus convolution used in this paper is illustrated in Fig. 1.

Consider a system S and a flow through S with input and output functions R and R^* and let $\beta \in \mathscr{F}$. We say that the system S offers β as a service curve to the flow if $R^* \ge R \otimes \beta$, which often means that for every $t \ge 0$ there exists some $s \le t$ such that $R^*(t) \ge R(s) + \beta(t-s)$ [2, Sec. 3.2.2]. We say that system S offers a *strict* service curve $\beta \in \mathscr{F}$ to the flow if $R^*(t) - R^*(s) \ge \beta(t-s)$ whenever $(s,t]$ is a backlogged period (i.e., $R(\tau) > R^*(\tau)$ for all τ such that $s < \tau \le t$). If β is a strict service curve, then it is a service curve, but the converse is not always true [3, Sec. 1.3]. A frequently used service curve is the rate-latency function $\beta_{r,T}$ that is the function in $\mathscr{F}$ defined by $\beta_{r,T}(t) = r[t-T]^+$, where we use the notation $[x]^+ = \max\{x, 0\}$. Saying that a system offers a service curve $\beta_{r,T}$ to a flow expresses that the flow is guaranteed a service rate r, except for possible interruptions that might impact the delay by at most T. Saying that a system offers a *strict* service curve $\beta_{r,T}$ to a flow expresses that the flow is guaranteed a service rate r, except for possible interruptions that might not exceed T in total per backlogged period. A strict service curve β can always be assumed to be super-additive, i.e., to satisfy $\beta(s+t) \ge \beta(s) + \beta(t)$ for all s, t (otherwise, it can be replaced by its super-additive closure [2, Prop. 5.6]).

Assume that a flow, constrained by arrival curve α, traverses a system that offers a service curve β to the flow and that respects the ordering of the flow (FIFO per-flow). The delay of the flow is upper bounded by $h(\alpha, \beta)$ (horizontal deviation), defined by

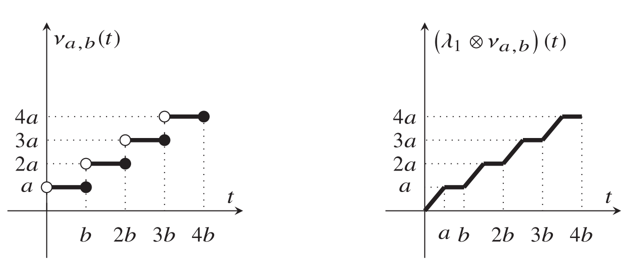
$$h(\alpha,\beta) = \sup_{t \ge 0}\{\inf\{d \ge 0 | \alpha(t) \le \beta(t+d)\}\} \tag{1}$$

Our technique of proof uses the lower pseudo-inverse. The lower pseudo-inverse $f^{\downarrow}$ of a function $f \in \mathscr{F}$ is defined by

$$f^{\downarrow}(y) = \inf\{x | f(x) \ge y\} = \sup\{x | f(x) < y\} \tag{2}$$

We use the following property from [5, Sec. 10.1]:

$$\forall x, y \in \mathbb{R}^+, y \le f(x) \Rightarrow x \ge f^{\downarrow}(y) \tag{3}$$



(a) $\nu_{a,b}$ (b) $(\lambda_1 \otimes \nu_{a,b})$

Fig. 1 Left: the stair function $\nu_{a,b} \in \mathscr{F}$ defined for $t \ge 0$ by $\nu_{a,b}(t) = a\left\lceil \frac{t}{b} \right\rceil$. Right: min-plus convolution of $\nu_{a,b}$ with the function $\lambda_1 \in \mathscr{F}$ defined by $\lambda_1(t) = t$ for $t \ge 0$. When $a \le b$, the discontinuities are smoothed, and replaced with a unit slope.

3. System Model

We consider a weighted round-robin subsystem that serves n input flows, has one queue per flow, and uses a weighted round-robin algorithm to arbitrate between flows. The arbitration algorithm assumed in this paper is IWRR, shown in Algorithm 1. When a packet of flow i enters the weighted round-robin subsystem, it is put into queue i. The weight of flow i is w_i. IWRR runs an infinite loop of *rounds*. In one round, each queue i has w_i emission opportunities; one packet can be sent during one emission opportunity. The inner loop defines a *cycle*, where each queue is visited but only those with a weight not smaller than the cycle number have an emission opportunity. The `send` instruction is assumed to be the only one with a non-null duration. Its actual duration depends on the packet size but also on the amount of service available to the entire weighted round-robin subsystem. See Fig. 2 for an illustration.

The weighted round-robin subsystem is itself placed in a larger system, and can compete with other queuing subsystems. For example, consider the case of a constant-rate server with several priority levels, without preemption, and where the weighted round-robin subsystem is at a priority level that is not the highest, as in [6, Sec. 8.6.8.3]. Assuming some arrival curve constraints for the higher priority traffic, the service received by the entire weighted round-robin subsystem can be modelled using a strict service curve [2, Sec. 8.3.2].

This motivates us to assume that the aggregate of all flows in the weighted round-robin subsystem receives a strict service curve, say $\beta \in \mathscr{F}$ that we call "aggregate strict service curve". If the weighted round-robin subsystem has exclusive access to a transmission line of rate c,

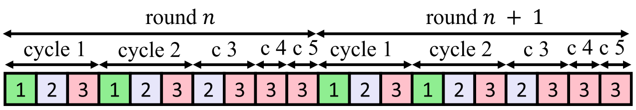


Fig. 2 Emission opportunities on two successive rounds for IWRR with three flows and $w_1 = 2$, $w_2 = 3$, $w_3 = 5$. Mind that this is not the temporal behaviour: each opportunity can lead to an empty interval if the queue is empty at this time. Furthermore, the duration of each non-empty interval depends on the packet size and the aggregate service available (we do not assume constant rate service).

Algorithm 1 Interleaved Weighted Round-Robin

```
Input: Integer weights w_1 ≤ w_2 ≤ .. ≤ w_n
w_max = max{w_1, .., w_n}
while True do                                   ▷ A round starts.
    for C ← 1 to w_max do                       ▷ A cycle starts.
        for i ← 1 to n do
            if C ≤ w_i then
                if (not empty(i)) then
                                                ▷ A service for queue i.
                    print(now,i);
                    send(head(i));
                    removeHead(i);
                end if
            end if
        end for
    end for                                     ▷ A cycle finishes.
end while                                       ▷ A round finishes.
```

then $\beta(t) = ct$ for $t \geq 0$. We assume that $\beta(t)$ is finite for every (finite) t and, without loss of generality, we assume β to be super-additive. Furthermore, we need an additional technical assumption, primarily for establishing the tightness result: we assume that β is Lipschitz-continuous, i.e., there exists a constant $K > 0$ such that $\frac{\beta(t)-\beta(s)}{t-s} \leq K$ for all $0 \leq s < t$; this does not appear to be a restriction as the rate at which data is served has a physical limit.

Here, we use the context of communication networks, but the results equally apply to real-time systems: Simply map flow to task, packet to job, packet size to job execution time and strict service curve to "delivery curve" [7], [8].

4. State of the Art

One of the first use of round-robin scheduling in the network context appeared in [9], with a fairness objective, i.e., a fair way to share the bandwidth among sessions. It is also mentioned in [10] as a way to implement "fair queueing".

The term "Weighed Round-Robin" was coined in [1] as a generalisation of round-robin to share the bandwidth "in proportion to prescripted weights" in the context of ATM (i.e., with constant-size packets). Two versions of the algorithm are presented in [1]. The former is presented in Algorithm 1: at cycle C (with C between 1 and $w_{\max}$), only flows with weight $w_i \geq C$ can emit one packet. We call this version IWRR. The latter version assumes that there exists for each flow i a bit-list of length $w_{\max}$, $o_i \in \{0,1\}^{w_{\max}}$, such that $w_i = \sum_{k=1}^{w_{\max}} o_i[k]$. A flow i can emit a packet at cycle C only if $o_i[C] = 1$. A strategy is given to build these vectors in [1] and is refined with fairness objectives in [11]. Call LIWRR (list-based IWRR) this version.

IWRR is modified into WRR/SB in [12] to enable some flow to send slightly more packets than permitted in a cycle, and to decrease accordingly at the next cycle.

As mentioned in Section 1, plain WRR (which we simply call "WRR") enables each flow i to send up to w_i packets every time it is selected [13]. A "Multiclass WRR" is also defined in [13]. Surprisingly, the authors of [13] were not aware of [1] and have re-invented LIWRR. Note that even if WRR was designed for packets of constant size, it has been applied in network of variable size packets such as Ethernet [6, Sec. 8.6, Sec. 8.6.8.3, Sec. 37], in request balancing in cloud infrastructures [14], in the LinuxVirtualServer scheduling [15], in network of chip [16], and so on. In fact, looking for expression "weighted round-robin" in the title or abstracts of papers index by Scopus returns more than 400 entries (March 2020), and Google references more than 4000 patents with this expression (March 2020). Unfortunately, when authors refer to WRR, they often do not explicit which version of WRR it is.

A WRR server is also a latency-rate server, with latency and rates given in [17] for packets of constant size. The latency result is generalised to LIWRR in [18]. Even if the notion of latency-rate server is very close to the one of a service curve $\beta_{r,T}$ in network calculus, both notions are slightly different, and results cannot be directly imported from one theory to the other [19]. In [16], the authors consider a Network on Chip (NoC), with WRR arbitration at the flit level. A flit is the elementary data unit of the NoC, one flit is sent per CPU/NoC cycle. Assuming that the weights are such that packets are never fragmented by the arbiter, a strict service curve β_{R_i,T_i} for flow i is found, with $R_i = \frac{w_i}{\sum_k w_k}$, $T_i = \sum_{j \neq i} w_j$.

WRR arbitration in an Ethernet switch is also considered in [20], with the assumption that all flows of an output ports have the same constant packet size. It then computes, in the network calculus framework, a residual service with service curve β_{R_i,T_i} with $R_i = \frac{w_i}{\sum_k w_k} C$, $T_i = \frac{\sum_{j \neq i} w_j}{C}$, where C is the link rate. We assume that the missing packet size in the T_i term was a typo. This network calculus result on conventional WRR arbitration in Ethernet is refined in [21], considering packets of variable size, leading to residual service with strict service curve β_{R_i,T_i} with $R_i = \frac{w_i l_i^{\min}}{w_i l_i^{\min} + \sum_{j \neq i} w_j l_j^{\max}} C$ and $T_i = \frac{\sum_{j \neq i} w_j l_j^{\max}}{C}$ (cf. Eq. (1) and (2) in [21]) where $l_i^{\min}$, $l_i^{\max}$ are, respectively, lower and upper bounds on the size of the packets in the flow i. It refines this result by subtracting the part of the bandwidth not used by interfering flows (considering their arrival curves).

Observe that computing a residual service with a $\beta_{R,T}$ curve is pessimistic as it assumes that, once the worst latency is paid, each packet is served with the long-term residual rate. Whereas, in reality, each packet, when it is selected for emission, is transmitted at full link speed up to completion. A residual service for the conventional WRR with a

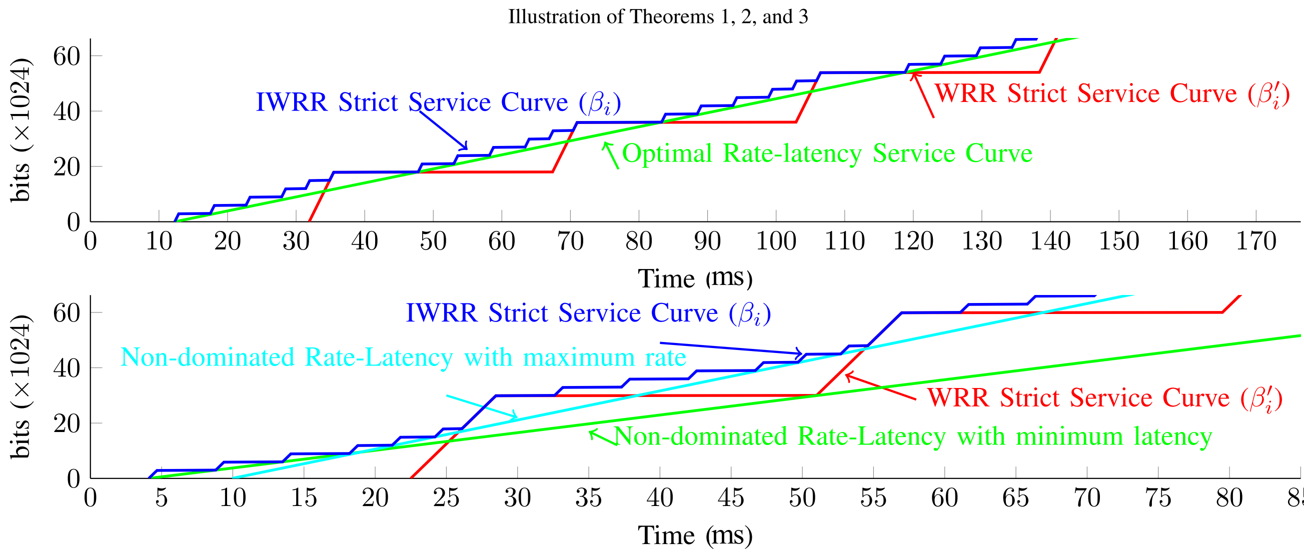


Fig. 3 Strict service curves obtained in Sect. 5 for an example with four input flows, weights = $\{4, 6, 7, 10\}$, $l^{\min} = \{4096, 3072, 4608, 3072\}$ bits, $l^{\max} = \{8704, 5632, 6656, 8192\}$ bits and $\beta(t) = ct$ with $c = 10$ Mb/s (i.e., the aggregate of all flows is served at a constant rate). The figure shows the IWRR service curve β_i and the WRR strict service curve β'_i for two of the flows; it also shows the non-dominated rate-latency strict service curves $\beta_{r^*_0,T^*_0}$ and $\beta_{r^*_{k^*},T^*_{k^*}}$ of Theorem 3 (in the top panel both are equal).

curve that is an alternation of full services and plateaus is given in [2, Sec. 8.2.4]. This effect of “full speed up to completion” can also be captured when computing the local delay of a server with $\beta_{R,T}$ service curve [22].

5. Strict Service Curves for IWRR

Our first result is a strict service curve for IWRR that, as we show in Sect. 6, is the best possible. We compare it to WRR and also give simpler, lower approximations.

Theorem 1 (Strict Service Curve of IWRR): Let S be a server shared by n flows that uses IWRR as explained in Sect. 3, with weight w_i for flow i. Recall that the server offers a strict service curve β to the aggregate of the n flows. For any flow i, $l_i^{\min}$ [resp. $l_i^{\max}$] is a lower [resp. upper] bound on the packet size.

Then, S offers to every flow i a strict service curve β_i given by $\beta_i(t) = \gamma_i(\beta(t))$ with

$$\gamma_i = \lambda_1 \otimes U_i \tag{4}$$

$$U_i(x) \overset{def}{=} \sum_{k=0}^{w_i-1} \nu_{l_i^{\min},L_{\text{tot}}}\left(\left[x - \psi_i(kl_i^{\min})\right]^+\right) \tag{5}$$

$$L_{\text{tot}} = w_i l_i^{\min} + \sum_{j,j\neq i} w_j l_j^{\max} \tag{6}$$

$$\psi_i(x) \overset{def}{=} x + \sum_{j,j\neq i} \phi_{i,j}\left(\left\lfloor \frac{x}{l_i^{\min}} \right\rfloor\right) l_j^{\max} \tag{7}$$

$$\phi_{i,j}(x) \overset{def}{=} \left\lfloor \frac{x}{w_i} \right\rfloor w_j + \left[w_j - w_i\right]^+ + \min(x \mod w_i + 1, w_j) \tag{8}$$

In the above, $\nu_{a,b}$ is the stair function, λ_1 is the unit rate function and $\otimes$ is the min-plus convolution, all are described in Fig. 1.

Furthermore, β_i is super-additive.

The proof is in Sect. 8.1. See Fig. 3 for some illustration of β_i. Observe that γ_i in Eq. (4) is the strict service curve obtained when the aggregate strict service curve is $\beta = \lambda_1$ (i.e., when the aggregate is served at a constant, unit rate). In the common case where β is equal to a rate-latency function, say $\beta_{c,T}$, we have $\beta_i(t) = \gamma_i(c(t-T))$ for $t \geq T$ and $\beta_i(t) = 0$ for $t \leq T$, namely, β_i is derived from γ_i by a rescaling of the x axis and a right-shift.

As mentioned in Sect. 2, any strict service curve that is not super-additive can be improved, by replacing it by its super-additive closure. The last statement in the theorem guarantees that it is not possible to improve the obtained service curve in this way.

We now compare to WRR. The best known service curve for (non-interleaved) WRR is given in [2, Sec. 8.2.4] and is

$$\beta'_i(t) = (\lambda_1 \otimes \nu_{q_i,L_{\text{tot}}})\left([\beta(t) - Q_i]^+\right) \tag{9}$$

with $q_i = w_i l_i^{\min}$ and $Q_i = \sum_{j,j\neq i} w_j l_j^{\max}$. In Sect. 6, we show that $\beta'_i(t)$ is indeed the best possible strict service curve for WRR. Furthermore, it is dominated by the strict service curve for IWRR:

Theorem 2: With the assumptions in Theorem 1 and in Eq. (9):

$$\beta'_i \leq \beta_i \tag{10}$$

The proof is in Sect. 8.2. Figure 3 illustrates how the strict service curve for IWRR improves on that for WRR, by providing a smoother, and generally larger, service.

The service curve found in Theorem 1 is the best possible one but has a complex expression. If there is interest in a simpler expression, any lower bounding function is a strict service curve; in particular, the strict service curve β'_i for WRR is also a valid, though suboptimal, strict service curve for IWRR. There is often interest in service curves that are rate-latency functions. Observe that, if the aggregate service curve β is a rate-latency function, then replacing γ_i by a rate-latency lower-bounding function also yields a rate-latency function for β_i, and vice-versa. Therefore, we are interested in rate-latency functions that lower bound γ_i.

Among all of these, there is not a single best one, as some have a smaller latency while others have a larger rate. We say that a rate-latency function $\beta_{r,T}$ that lower bounds γ_i is non-dominated if there is no other rate latency function $\beta_{r',T'}$ that lower bounds γ_i and dominates $\beta_{r,T}$, i.e., such that $r' \geq r$ and $T' \leq T$. The following result gives all such non-dominated rate-latency functions. Let $r^* = \frac{q_i}{L_{\text{tot}}} = \frac{w_i l_i^{\min}}{L_{\text{tot}}}$, $r_{w_i-1} = 1$, and

$$r_k = \frac{l_i^{\min}}{\psi_i((k+1)l_i^{\min}) - \psi_i(kl_i^{\min})},\ 0 \leq k < w_i - 1 \tag{11}$$

$$k^* = \min\{0 \leq k < w_i \mid r_k \geq r^*\} \tag{12}$$

$$r_k^* = \min(r_k, r^*),\ 0 \leq k \leq k^* \tag{13}$$

Theorem 3: With the assumptions in Theorem 1 and the definitions (11)–(13), a rate-latency function $\beta_{r,T}$ lower bounds γ_i and is non-dominated if and only if $r = r_{k^*}^*$ and $T = \psi_i(k^* l_i^{\min}) - \frac{k^* l_i^{\min}}{r}$, or $r_{k-1}^* \leq r < r_k^*$ and $T = \psi_i(kl_i^{\min}) - \frac{kl_i^{\min}}{r}$ for some integer k with $0 < k \leq k^*$. Among all such rate-latency functions, the one with lowest latency is $\beta_{r_0^*,T_0^*}$ and the one with largest rate is $\beta_{r_{k^*}^*,T_{k^*}^*}$.

The proof is in Sect. 8.3. Figure 3 illustrates $\beta_{r_0^*,T_0^*}$ and $\beta_{r_{k^*}^*,T_{k^*}^*}$ in some examples. Observe that $k \mapsto r_k^*$ is wide-sense increasing with k for $0 \leq k \leq k^*$, but the values of r_k^* are not necessarily all distinct. It can also occur that $k^* = 0$ (as in the top panel of Fig. 3); in which case, there is one optimal rate-latency service curve. In general, however, this does not occur, and a simple lower bounding approximation can be obtained with the supremum of all non-dominated rate-latency service curves, as given by the next theorem.

Theorem 4: With the assumptions in Theorem 3, the supremum of all non-dominated rate-latencies is equal to $\max\left(\beta_{r_0^*,T_0^*}, \ldots, \beta_{r_{k^*}^*,T_{k^*}^*}\right)$, and it is the largest convex function that lower bounds γ_i.

The proof is in Sect. 8.4. There is often interest in service curves that are piecewise-linear and convex. Specifically, convex piecewise-linear functions are stable under addition and maximum and the min-plus convolution can be computed in automatic tools very efficiently [2, Sec. 4.2]. The above theorem thus gives the best such strict service curve.

6. Tightness

We first show that the strict service curve we have obtained is the best possible. The proofs of all results in this section are in Sect. 8.

6.1 Tightness of Strict Service Curve

Theorem 5: (Tightness of the IWRR Service Curve) Consider a weighted round-robin subsystem that uses the IWRR scheduling algorithm, as defined in Sect. 3. Assume the following system parameters are fixed: the number of input flows, the weight w_j allocated to every flow j, the bounds on packet sizes $l_j^{\min}$ and $l_j^{\max}$ for every flow j, and the strict service curve β for the aggregate of all flows. Let i be the index of one of the flows.

Assume that $b_i \in \mathscr{F}$ is a strict service curve for flow i in any system that satisfies the specifications above. Then $b_i \leq \beta_i$ where β_i is given in Theorem 1.

Interestingly, we obtain a similar result for WRR. Recall that β'_i is the strict service curve for flow i, described in Eq. (9), which was obtained in [2, Sec. 8.2.4].

Theorem 6: (Tightness of the WRR Service Curve) Theorem 5 is also valid if we replace IWRR with WRR. Specifically, using WRR as a scheduling policy, β'_i is the largest possible strict service curve for flow i.

6.2 Tightness of Delay Bounds with Constant Packet Sizes

Having obtained the best-possible strict service curve does not guarantee that the delay bounds derived from it are tight, i.e., are worst-case delays. This is because a service curve is only an abstraction of the system; and we have obtained a strict service curve, and non-strict service curves might provide better results. However, we show that, for flows of packets of constant size, we do obtain tight delay bounds. We show that it holds for IWRR and for WRR.

Recall that a delay bound requires the knowledge of an arrival curve α_i for the flow of interest. If this flow generates only packets of length l, then α_i can be assumed to be a multiple of l and sub-additive. A delay bound for this flow is then equal to $h(\alpha_i, \beta_i)$ (see Eq. (1)).

Theorem 7: (Tightness of Delay Bound for IWRR with Constant Packet Size) Consider a system, as in Theorem 5, with the additional assumption that, for the flow of interest i, $l_i^{\min} = l_i^{\max} = l$.

Let $\alpha_i \in \mathscr{F}$ be a sub-additive function that is an integer multiple of l, and assume that flow i has α_i as arrival curve. The network calculus delay bound is tight, i.e, there exists a trajectory where the delay of one packet of flow i is equal to $h(\alpha_i, \beta_i)$.

Theorem 8: (Tightness of Delay Bound for WRR with

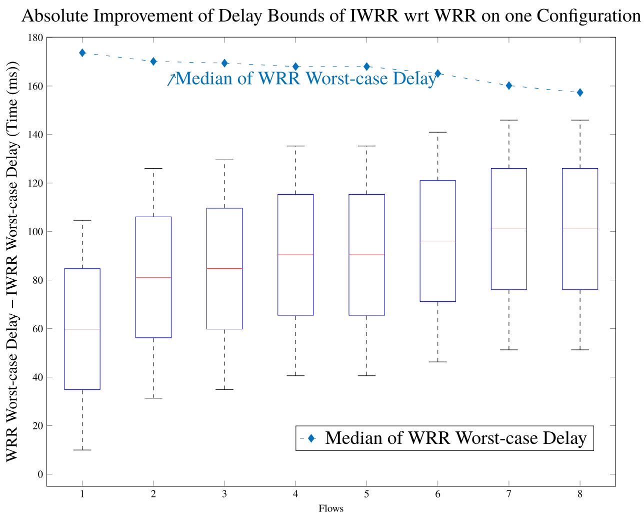


Fig. 4 Box-and-whisker plots of difference between WRR and IWRR delay bounds with weights {22, 27, 28, 30, 30, 34, 41, 45} and $l = 7119$ bit with random arrival curves. Median WRR delay bounds are also provided.

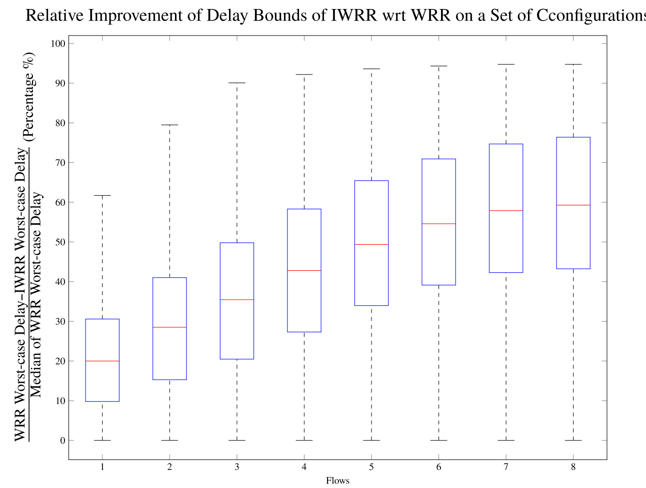


Fig. 5 Box-and-whisker plots of difference between WRR and IWRR delay bounds normalized to the median of WRR delay bounds, for several systems with weights picked uniformly at random in [10, 50], assigned to flow by increasing order, and a packet length picked uniformly at random in [64, 1522] bytes.

Constant Packet Size) Theorem 7 is also valid for the WRR policy.

7. Numerical Examples

To compare IWRR and WRR worst-case delays, we provide some numerical examples. First, we consider a system of 8 input flows $f_1, \dots, f_8$ with respective weights $\{22, 27, 28, 30, 30, 34, 41, 45\}$ and $l^{\min} = l^{\max} = l = 7119$ bit. Let the aggregate service, β, be a constant bit rate of 10 Mb/s. For every flow i, we compute the IWRR and WRR strict service curves β_i, β'_i. Then, for every i, we generate $N = 1000$ leaky-bucket arrival curves γ_{r,b_k}, $k = 1, \dots, N$, with rate $r = 0.5$ Mb/s and burst b_k picked uniformly at random in [1, 20] packets. Then, we use $\alpha_i^k = \lceil \frac{\gamma_{r,b_k}}{l} \rceil l$ to satisfy the conditions of Theorems 7 and 8 and to compute $d_i^k = h(\alpha_i^k, \beta_i)$ and $\dot{d}_i^k = h(\alpha_i^k, \beta'_i)$. Figure 4 gives the box-and-whisker plots of the $\dot{d}_i^k - d_i^k$ series. The median of WRR delay bounds $\dot{d}_i^k$ are also provided to illustrate the improvement.

Second, we repeated the same study for $M = 10000$ sets of system parameters. For each system, we choose the weights of 8 flows by picking them uniformly at random between 10 and 50, and we pick a packet length l uniformly at random between 64 to 1522 bytes. For each experiment, we call flow 1 the flow with the smallest weight, flow 2 with second smallest weight, and so on. As the scale of delay bounds depends on the choices of weights and the packet length, the $\dot{d}_i^k - d_i^k$ series are divided by $\dot{d}_i^{\bar{m}}$, the median of WRR delay bounds for flow i. Figure 5 gives the box-and-whisker plots of the $\frac{\dot{d}_i^k - d_i^k}{\dot{d}_i^{\bar{m}}}$ series. Using IWRR improves worst-case delays, as expected, and the improvement is larger for flows with larger weights.

8. Proofs

8.1 Proof of Theorem 1

The idea of proof is as follows. We consider a backlogged period $(s, t]$ of flow of interest i, and we let p be the number of packets of flow i that are entirely served during this period. For every other flow j, the number of packets that are entirely served is upper bounded by a function of p, given in Lemma 3. Also, p is upper bounded by a function of the amount of service received by flow i in Lemma 5. Combining these two results gives an implicit inequality for the total amount of service in Eq. (26). By using the technique of pseudo-inverse, this inequality is inverted and provides a lower bound for the amount of service received by the flow of interest.

8.1.1 Key Variables and Basic Properties

Let $(s, t]$ be a backlogged period of flow i. Let (τ_k, f_k) be couples of (instant, flow), printed at line 8 of Algorithm 1. Note that $\tau_k < \tau_{k+1}$ as the `send` instruction has a non-null duration (because the aggregate service curve β is Lipschitz continuous). Let $\sigma(0), \sigma(1), \dots$ be the sequence of service opportunities for flow i at or after s, i.e., $\sigma(0) = \min\{m \mid \tau_m \geq s, f_m = i\}$ and $\sigma(k) = \min\{m \mid \tau_m > \tau_{\sigma(k-1)}, f_m = i\}$. The kth service opportunity for flow i occurs at time $\tau_{\sigma(k-1)}$; we say that it is "complete" if $\tau_{\sigma(k-1)+1} \leq t$, i.e., the interval taken by this service is entirely in $[s, t]$. Let $p \geq 0$ be the number of complete service opportunities. Observe that it is possible that $p = 0$, and it might happen that $\tau_{\sigma(p)} < t$ or $\tau_{\sigma(p)} \geq t$ (see Fig. 6).

In each service of flow i, during a backlogged period, it sends one packet with a length $\geq l_i^{\min}$, thus, for all $k =$

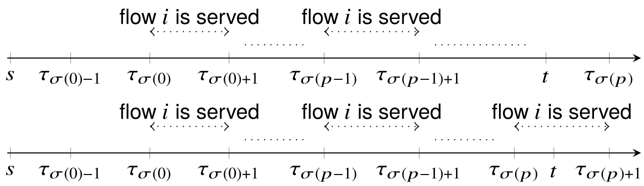


Fig. 6 Illustration of two possible cases of $\tau_{\sigma(p)} \geq t$ and $\tau_{\sigma(p)} < t$.

$0 \dots (p-1)$, we have $R_i^*(\tau_{\sigma(k+1)}) - R_i^*(\tau_{\sigma(k)}) \geq l_i^{\min}$, therefore

$$R_i^*(\tau_{\sigma(p)}) - R_i^*(\tau_{\sigma(0)}) \geq p l_i^{\min} \tag{14}$$

8.1.2 Amount of Service to Other Flows

In order to upper bound the number of emission opportunities for another flow j, we first find an expression, in Lemma 1, for the number of emission opportunities for flow j between two consecutive emission opportunities for flow i. Lemma 2 then finds an upper bound on the number of emission opportunities for flow j in $(s, \tau_{\sigma(p)})$, as a function of the cycle number (variable C in Algorithm 1) at $\tau_{\sigma(0)}$. Lastly, Lemma 3 maximizes the previous upper bound over all values of C.

Lemma 1: The number of emission opportunities for flow $j \neq i$ between two consecutive emission opportunities for flow i, given that the latter emission opportunity for flow i occurs at cycle C, is equal to

$$q_{g,f}(C) = \begin{cases} 0 & \text{if } 1 < C \leq w_i \text{ and } w_j < C \\ 1 & \text{if } 1 < C \leq w_i \text{ and } w_j \geq C \\ \left[w_j - w_i\right]^+ + 1 & \text{if } C = 1 \end{cases} \tag{15}$$

Proof: According to Algorithm 1, flow i has emission opportunities only in the first w_i cycles of each round. Both emission opportunities are either in the same round (Case 1) or in two consecutive rounds (Case 2). As C is the cycle number for the second emission opportunity for flow i, Case 1 can occur only when $1 < C \leq w_i$, and Case 2 can occur when $C = 1$. For Case 1, we further differentiate between $w_j < C$ and $w_j \geq C$.

Case 1a: $1 < C \leq w_i$ and $w_j < C$: Queue j does not have an emission opportunity in cycle C because $w_j < C$. Also, we must have $w_j < w_i$, thus queue j does not have any emission opportunity after i in cycle $C-1$. Hence, $q_{i,j}(C) = 0$.

Case 1b: $1 < C \leq w_i$ and $w_j \geq C$: If $w_j > w_i$, then queue j has an emission opportunity after queue i in cycle $C-1$. If $w_j = w_i$, then queue j has an emission opportunity before i in cycle C, or after i in cycle $C-1$. Else, $C \leq w_j < w_i$ and queue j has an emission opportunity in cycle C, before i. In all cases, $q_{i,j}(C) = 1$.

Case 2: $C = 1$: The first emission opportunity for i is in the last cycle of a round that includes i (cycle w_i). If $w_j > w_i$, then queue j has an emission opportunity in the rest of cycle w_i and also has emission opportunities during the next $(w_j - w_i)$ cycles of the last round. In this case, $q_{i,j}(C) = w_j - w_i + 1$, which is also the value in the last line of Eq. (15). Else if $w_j = w_i$, queue j has an emission opportunity before i in this cycle or after i in cycle w_i of the first round, thus $q_{i,j}(C) = 1$, which is also the value in the last line of Eq. (15). Else, $w_j < w_i$ and queue j has an emission opportunity before i in this cycle. Here too, $q_{i,j}(C) = 1$, the value in the last line of Eq. (15). □

Lemma 2: The number of emission opportunities for flow $j \neq i$ in $(s, \tau_{\sigma(p)})$, for any backlogged period $(s, t]$ of flow i with p complete services, given that the first service starts at cycle number C (cycle number at time $\tau_{\sigma(0)}$) is upper bounded by

$$q'_{i,j}(C, p) \stackrel{def}{=} \sum_{k=0}^{p} q_{i,j}\left((C + k - 1) \mod w_i + 1\right) \tag{16}$$

Also, let $C'(p)$ be the cycle number at $\tau_{\sigma(p)}$. Then,

$$C'(p) = (C + p - 1) \mod w_i + 1 \tag{17}$$

Proof: By induction on p.

Base Case: $p = 0$

In this case, $q'_{i,j}(C, 0)$ is the number of emission opportunities for flow j between two consecutive emission opportunities for flow i that by Lemma 1, is equal to $q_{i,j}(C)$. As $1 \leq C \leq w_i$, $(C-1) \bmod w_i + 1 = C$ thus $q_{i,j}(C) = q_{i,j}((C-1) \bmod w_i + 1)$. This shows Eq. (16). Also, by definition, $C'(0) = C$; using again $(C-1) \bmod w_i + 1 = C$ shows that Eq. (17) holds.

Induction step:

We assume that Eq. (16) and Eq. (17) hold for $p-1$, and we want to show that they also hold for p.

First, let us prove Eq. (17). There are two possible cases: (a) if $0 \leq C'(p-1) < w_i$, then both $(p-1)$st and pth emission opportunities occur in the same round, thus $C'(p) = C'(p-1) + 1$. By the induction hypothesis, $(C + p - 2) \bmod w_i + 1 < w_i$, i.e., $(C + p - 2) \bmod w_i < w_i - 1$. Note that, for any integer x

$$(x{+}1) \mod w = \begin{cases} (x \mod w) + 1 \text{ if } (x \mod w) < w - 1 \\ 0 \text{ otherwise} \end{cases} \tag{18}$$

By using Eq. (18), we obtain that $C'(p)$ is given by Eq. (17) as required. (b) In the second case, $C'(p-1) = w_i$ then the next emission opportunity occurs in the first cycle of the next round, thus $C'(p) = 1$. Here too, applying Eq. (18) shows that $C'(p)$ is given by Eq. (17) as required.

Then, we prove Eq. (16). Let N be the number of emission opportunities for flow j in $[s, \tau_{\sigma(p)})$. N is the sum of N_1, the number of emission opportunities in $[s, \tau_{\sigma(p-1)})$, and N_2, the number of emission opportunities in $(\tau_{\sigma(p-1)}, \tau_{\sigma(p)})$. By the induction hypothesis, $N_1 \leq q'_{i,j}(C, p-1)$. Also, by

Lemma 1, we have $N_2 \le q_{i,j}(C'(p))$. Thus, by using Eq. (17) which was just shown to also hold for p, we obtain

$$N \le \sum_{k=0}^{p-1} q_{i,j}\left((C+k-1) \mod w_i + 1\right) + q_{i,j}\left((C+p-1) \mod w_i + 1\right) \tag{19}$$

where the right-hand side is equal to $q'_{i,j}(C,p)$ as required. □

Lemma 3: For any backlogged period $(s,t]$ of flow i with p complete services, the number of emission opportunities for flow $j \neq i$ in $(s, \tau_{\sigma(p)})$ is upper bounded by $\phi_{i,j}(p)$, defined in Eq. (8).

Proof: Lemma 2 gives the number of emission opportunities for flow $j \neq i$ in $(s, \tau_{\sigma(p)})$, for any backlogged period $(s,t]$ of flow i with p complete services, when the first service starts at cycle number C (cycle number at time $\tau_{\sigma(0)}$). To obtain the lemma, we maximize this result over C. We show the following properties.

(P1) For any integer $C \in [1, w_i]$,

$$\sum_{k=0}^{w_i-1} q_{i,j}\left((C+k-1) \mod w_i + 1\right) = w_j \tag{20}$$

The mapping $k \mapsto (C+k-1) \bmod w_i + 1$ is one-to-one from $\{0, ..., w_i - 1\}$ onto $\{1, ..., w_i\}$, thus the left-hand side of Eq. (20) is equal to $\sum_{k=1}^{w_i} q_{i,j}(k)$ that as we show now, is equal to w_j. First, we have $q_{i,j}(1) = \left[w_j - w_i\right]^+ + 1$. Also, $q_{i,j}(k) = 1$ when $k > 1$ and $w_j \ge k+1$. Thus, $\sum_{k=2}^{w_i} q_{i,j}(k) = \min(w_i - 1, w_j - 1)$ and finally the left-hand side is equal to $\left[w_j - w_i\right]^+ + \min(w_i - 1, w_j - 1) + 1$, which is equal to w_j.

(P2) For any integers $C \in [1, w_i]$ and $p \ge 0$,

$$q'_{i,j}(C,p) = \left\lfloor \frac{p}{w_i} \right\rfloor w_j + \sum_{k=0}^{p \bmod w_i} q_{i,j}\left((C+k-1) \mod w_i + 1\right) \tag{21}$$

$q_{i,j}$ is a periodic function with period w_i. By (P1), the sum over one complete period is w_j. Also, we can write $p = \left\lfloor \frac{p}{w_i} \right\rfloor w_i + p \bmod w_i$. Thus, we have $\left\lfloor \frac{p}{w_i} \right\rfloor$ complete rounds, and the sum in Eq. (21) is the remainder.

(P3) $q_{i,j}$ is a wide-sense decreasing function. This means that for any integer $k \in [1, w_i)$, $q_{i,j}(k+1) \le q_{i,j}(k)$. If $k = 1$, this follows from $q_{i,j}(1) \ge 1$ and $q_{i,j}(2) \le 1$. Else if $k \le w_j < k+1$, then $q_{i,j}(k+1) = 0$ and $q_{i,j}(k) = 1$. Else, they are equal. Hence, in all cases the property holds.

(P4) For any integer $C \in [1, w_i]$ and $p \ge 0$,

$$q'_{i,j}(C,p) \le q'_{i,j}(1,p) \tag{22}$$

By using (P2), we should show that

$$\sum_{k=0}^{p \bmod w_i} q_{i,j}\left((C+k-1) \mod w_i + 1\right)$$

is upper bounded by $\sum_{k=0}^{p \bmod w_i} q_{i,j}(k \bmod w_i + 1)$. Note that here we have $k \bmod w_i = k$. Both sides are the sum of $a \overset{def}{=} p \bmod w_i + 1$ unique elements of the set $\{q_{i,j}(k)\}_{k \in [1, w_i]}$. By (P3), the right-hand side is the maximum sum of a unique elements of this set.

(P5) For any integer $p \ge 0$,

$$q'_{i,j}(1,p) = \phi_{i,j}(p) \tag{23}$$

We apply (P2) with $C = 1$ to compute $q'_{i,j}(1,p)$. Then, the sum in the right-hand side of Eq. (21) is equal to $\sum_{k=0}^{p \bmod w_i} q_{i,j}(k+1)$, as $k \bmod w_i = k$. Then, by using the same argument after Eq. (20), it is equal to $\left[w_j - w_i\right]^+ + 1 + \min(p \bmod w_i, w_j - 1)$, which, by Eq. (8), is precisely $\phi_{i,j}(p)$.

The lemma then follows directly from (P4) and (P5). □

Lemma 4: For every flow $j \neq i$,

$$R_j^*(t) \le R_j^*(\tau_{\sigma(p)}) \tag{24}$$

Proof: If $t \le \tau_{\sigma(p)}$, the result follows from R_j^* being wide-sense increasing. Else, we have $t > \tau_{\sigma(p)}$; this implies that flow i is served during $[\tau_{\sigma(p)}, t]$; thus for any other flow j, $R_j^*(t) = R_j^*(\tau_{\sigma(p)})$. □

8.1.3 Amount of Service to Flow of Interest

Lemma 5: The number of complete services, p, of flow of interest, i, in $(s,t]$ is upper bounded by

$$p \le \left\lfloor \frac{R_i^*(t) - R_i^*(s)}{l_i^{\min}} \right\rfloor \tag{25}$$

Proof: First, $R_i^*(s) \le R_i^*(\tau_{\sigma(0)})$, as $s \le \tau_{\sigma(0)}$ and R_i^* is wide-sense increasing. Second, consider the two cases in 8.1.1. If $t \ge \tau_{\sigma(p)}$, the property holds. Else, the scheduler in not serving flow i in $[\tau_{\sigma(p-1)+1}, \tau_{\sigma(p)})$, thus, $R_i^*(t) = R_i^*(\tau_{\sigma(p)})$. Hence, in both cases $R_i^*(t) \ge R_i^*(\tau_{\sigma(p)})$. By Eq. (14), $R_i^*(t) - R_i^*(s) \ge p l_i^{\min}$. Then, observe that p is integer. □

8.1.4 Total Amount of Service

Lemma 6: For any backlogged period $(s,t]$ of the flow of interest i,

$$\beta(t-s) \le \psi_i\left(R_i^*(t) - R_i^*(s)\right) \tag{26}$$

where ψ_i is defined in Eq. (7).

Proof: As the interval $(s,t]$ is a backlogged period, by the definition of the strict service curve for the aggregate of flows, $\beta(t-s) \le \sum_j R_j^*(t) - R_j^*(s)$. We upper bound $R_j^*(t)$ for all $j \neq i$ by applying Lemma 4,

$$\beta(t-s) \le (R_i^*(t) - R_i^*(s)) + \sum_{j, j \neq i} R_j^*(\tau_{\sigma(p)}) - R_j^*(s) \tag{27}$$

Each flow j has at most $\phi_{i,j}(p)$ emission opportunities during

$\left(s, \tau_{\sigma(p)}\right)$ (Lemma 3) and can send at most one packet of maximum size in each. Thus,

$$\beta(t-s) \leq (R_i^*(t) - R_i^*(s)) + \sum_{j,j\neq i} \phi_{i,j}(p) l_j^{\max} \tag{28}$$

Also, Lemma 5 finds an upper bound on p. Thereby,

$$\begin{aligned} \beta(t-s) \leq &(R_i^*(t) - R_i^*(s)) \\ &+ \sum_{j,j\neq i} \phi_{i,j}\left(\left\lceil \frac{R_i^*(t) - R_i^*(s)}{l_i^{\min}} \right\rceil\right) l_j^{\max} \end{aligned} \tag{29}$$

where the right-hand side is equal to $\psi_i(R_i^*(t) - R_i^*(s))$. □

8.1.5 Lower Pseudo-Inverse of ψ_i

Our next step is to invert Eq. (26) by computing the lower-pseudo inverse of ψ_i. As the calculus of pseudo inverses applies to wide-sense increasing functions, we first show:

Lemma 7: ψ_i, defined in Eq. (7), is wide-sense increasing.

Proof: It is sufficient to show that $\phi_{i,j}$, defined in Eq. (8), is a wide-sense increasing function. For any non-negative integers x and y such that $y \leq x$, we can write $x = kw_i + (x \bmod w_i)$ and $y = k'w_i + (y \bmod w_i)$, where k and k' are non-negative integers. We must have $k \leq k'$. If $k = k'$, we know that ($y \bmod w_i \leq x \bmod w_i$) and $\left\lfloor \frac{x}{w_i} \right\rfloor = \left\lfloor \frac{y}{w_i} \right\rfloor$. Hence, $\phi_{i,j}(y) \leq \phi_{i,j}(x)$. Else, $k > k'$ and $\left\lfloor \frac{x}{w_i} \right\rfloor > \left\lfloor \frac{y}{w_i} \right\rfloor$. Thereby, $\phi_{i,j}(x)$ is at least one w_j larger than $\phi_{i,j}(y)$. Hence, $\phi_{i,j}(y) < \phi_{i,j}(x)$. □

Lemma 8: Let $g_0, g_1, \dots, g_k, \dots$ be a non-negative sequence such that $g_{k+1} - g_k \geq 1$. The sequence can be extended to a function in $\mathscr{F}$ by $g(x) = g_{\lfloor x \rfloor}$ and let $g^\downarrow$ be its lower pseudo-inverse, so that $g^\downarrow(y) = k+1 \in \mathbb{N} \Leftrightarrow g_k < y \leq g_{k+1}$. Define $f \in \mathscr{F}$ by $f(x) = g_{\lfloor x \rfloor} + x \bmod 1$. Then, $f^\downarrow = \lambda_1 \otimes g^\downarrow$.

Proof: Observe that convolving $g^\downarrow$ with λ_1 consists in smoothing the unit steps with a slope of 1 (Fig. 1). Thus $(\lambda_1 \otimes g^\downarrow)(y) = k + y - g_k$ whenever $g_k \leq y \leq g_k + 1$ and $(\lambda_1 \otimes g^\downarrow)(y) = k + 1$ whenever $g_k + 1 \leq y \leq g_{k+1}$.

Also, f is piecewise linear and can be inverted in closed form on every interval where it is linear. A direct calculation gives $f^\downarrow(y) = k + y - g_k$ whenever $g_k \leq y \leq g_k + 1$ and $f^\downarrow(y) = k + 1$ whenever $g_k + 1 \leq y \leq g_{k+1}$. □

Lemma 9: Let $f \in \mathscr{F}$ and $l, m > 0$. Define $h \in \mathscr{F}$ by $h(x) = mf\left(\frac{x}{l}\right)$. Then, for all $y \geq 0$, $h^\downarrow(y) = lf^\downarrow\left(\frac{y}{m}\right)$.

Proof: Let $B(f, y) \overset{def}{=} \{x \geq 0, h(x) \geq y\}$ so that $f^\downarrow(y) = \inf B(y, f)$. Observe that $x \in B(h, y) \Leftrightarrow \frac{x}{l} \in B\left(f, \frac{y}{m}\right)$. □

Lemma 10: Let $a \in \mathscr{F}$ and $l > 0$. Define $b \in \mathscr{F}$ by $b(x) = lf\left(\frac{x}{l}\right)$. Then, for all $x \geq 0$, $(\lambda_1 \otimes b)(x) = l(\lambda_1 \otimes a)\left(\frac{x}{l}\right)$.

Do the change of variable $u = lv$ in the expansion $(\lambda_1 \otimes b)(x) = \inf_{0 \leq u \leq x}(u + b(x-u))$ and obtain $(\lambda_1 \otimes b)(x) = \inf_{0 \leq v \leq \frac{x}{l}}\left(lv + a\left(\frac{x}{l} - v\right)\right) = l(\lambda_1 \otimes a)\left(\frac{x}{l}\right)$. □

We can now compute the lower-pseudo inverse of ψ_i. First, define the sequence g by $g_k = \frac{1}{l_i^{\min}} \psi_i\left(k l_i^{\min}\right)$. As in Lemma 8, g can be extended to a piecewise constant function whose lower-pseudo inverse, $g^\downarrow$, can be directly computed:

$$g^\downarrow(x) = \frac{1}{l_i^{\min}} \sum_{k=0}^{w_i - 1} \nu_{l_i^{\min}, L_{\text{tot}}}\left(l_i^{\min}[x - g_k]^+\right) \tag{30}$$

Second, observe that for all $x \geq 0$, $\psi_i(x) = \psi_i(\lfloor \frac{x}{l_i^{\min}} \rfloor l_i^{\min}) + x \bmod l_i^{\min}$. Define f and h from g as in Lemmas 8 and 9 with $l = m = l_i^{\min}$, so that $h = \psi_i$. Apply Lemmas 8 and 9 and obtain $\psi_i^\downarrow(x) = l_i^{\min}\left(\lambda_1 \otimes g^\downarrow\right)\left(\frac{x}{l_i^{\min}}\right)$. Now apply Lemma 10 with $a = g^\downarrow$, $l = l_i^{\min}$, and $b = U_i$ to obtain

$$\psi_i^\downarrow = \lambda_1 \otimes U_i \tag{31}$$

Proof of Theorem 1: Lemma 6 gives, in Eq. (26), an upper bound on the total amount of service as a function of the service received by the flow of interest. We invert Eq. (26) by the lower-pseudo inverse technique in Eq. (3) and obtain $R_i^*(t) - R_i^*(s) \geq \psi_i^\downarrow(\beta(t-s))$. The lower-pseudo inverse of ψ_i is given by Eq. (31), thus

$$R_i^*(t) - R_i^*(s) \geq (\lambda_1 \otimes U_i)(\beta(t-s)) = \beta_i(t-s) \tag{32}$$

Lastly, we need to prove that β_i is super-additive. This follows from the tightness result in Theorem 5 (the proof of which is independent of rest of this proof). Indeed, the super-additive closure $\bar{\beta}_i$ of β_i is also a strict service curve, and $\bar{\beta}_i(t) \geq \beta_i(t)$ for all t [2, Prop. 5.6]). By Theorem 5, we also have $\bar{\beta}_i(t) \leq \beta_i(t)$ for all t, hence $\bar{\beta}_i = \beta_i$. □

8.2 Proof of Theorem 2

Proof: The WRR strict service curve [2, Sec. 8.2.4] is defined by $\beta_i'(t) = \gamma_i'(\beta(t))$ with

$$\gamma_i' = (\lambda_1 \otimes \nu_{q_i, L_{\text{tot}}})([t - Q_i]^+) \tag{33}$$

$$\psi_i'(x) \overset{def}{=} x + \sum_{j,j\neq i} \phi_{i,j}'\left(\left\lceil \frac{x}{l_i^{\min}} \right\rceil\right) l_j^{\max} \tag{34}$$

$$\phi_{i,j}'(x) \overset{def}{=} \left(1 + \left\lfloor \frac{x}{w_i} \right\rfloor\right) w_j \tag{35}$$

where γ_i' is the lower-pseudo inverse of ψ_i'. We know that for IWRR, γ_i is also the lower-pseudo inverse of ψ_i (defined in Eq. (7)). We first show that $\psi_i \leq \psi_i'$.

It is sufficient to prove that for all $j \neq i$ and for all $k \in \mathbb{N}$, $\phi_{i,j}(k) \leq \phi_{i,j}'(k)$. From the definition of $\phi_{i,j}$ and as $\min(x \bmod w_i + 1, w_j) \leq \min(w_i, w_j)$,

$$\phi_{i,j}(x) \leq \left\lfloor \frac{x}{w_i} \right\rfloor w_j + \left[w_j - w_i\right]^+ + \min(w_i, w_j) \tag{36}$$

Observe that $\left[w_j - w_i\right]^+ + \min(w_i, w_j) = w_j$. Hence, the right-hand side is $\phi_{i,j}'(x)$. This shows that

$$\psi_i \le \psi_i' \tag{37}$$

In [5, Sec. 10.1], it is shown that

$$\forall f, g \in \mathscr{F}, f \ge g \Rightarrow f^{\downarrow} \le g^{\downarrow} \tag{38}$$

Apply Eq. (38) to Eq. (37) to conclude the proof. □

8.3 Proof of Theorem 3

Lemma 11: Consider some integers $w \ge 1$ and $0 \le k \le w-1$, a finite sequence $g_0, g_1, \dots, g_{w-1}$, and a number $a \in \mathbb{R}$ that satisfy

1. $\forall \ell \in \mathbb{N}$ if $0 \le \ell \le w-2$ then $g_{\ell+1} - g_\ell \ge 1$
2. $\forall \ell \in \mathbb{N}$ if $0 \le \ell \le w-3$ then $g_{\ell+2} - g_{\ell+1} \le g_{\ell+1} - g_\ell$
3. if $k \le w-2$ then $a \ge g_{k+1} - g_k$ else $a \ge 1$
4. if $k \ge 1$ then $a \le g_k - g_{k-1}$

Define f: $[0, w) \to \mathbb{R}$ by $f(x) = g_{\lfloor x \rfloor} + x \bmod 1$ and h: $[0, w) \to \mathbb{R}$ by $h(x) = a(x-k) + g_k$. Then $h \ge f$.

Proof: First we show that

$$\forall \ell \in \{0, \dots, w-1\}, g_k - g_\ell \ge a(k-\ell) \tag{39}$$

Case 1: $\ell < k$. Then $g_k - g_\ell = \sum_{k'=\ell}^{k-1}(g_{k'+1} - g_{k'})$. By 2) every term in the sum is $\ge g_k - g_{k-1}$, by 4) is also $\ge a$ and there are $(k-\ell)$ terms, which shows Eq. (39).
Case 2: $\ell = k$. Then Eq. (39) is obvious.
Case 3: $\ell > k$. Then $g_\ell - g_k = \sum_{k'=k}^{\ell-1}(g_{k'+1} - g_{k'})$. By 2) every term in the sum is $\le g_{k+1} - g_k$; note that we must have $k \le w-2$ thus by 3), every term in the sum is also $\le a$; also, there are $\ell - k$ terms. Thus $g_\ell - g_k \le a(\ell - k)$, which shows Eq. (39) in this case.

We now proceed with the proof of the lemma. Consider some arbitrary $x \in [0, w)$ and let $\ell = \lfloor x \rfloor$. Then

$$f(x) = x - \ell + g_\ell \tag{40}$$
$$h(x) = a(x-\ell) + a(\ell - k) + g_k \tag{41}$$
$$h(x) - f(x) = \underbrace{(a-1)(x-\ell)}_{A} + \underbrace{g_k - g_\ell - a(k-\ell)}_{B} \tag{42}$$

Observe that we must have $a \ge 1$: if $k = w-1$ this follows from 3), and if $k \le w-2$ it follows from 3) and 1); thus $A \ge 0$. Also $B \ge 0$ by Eq. (39). □

Lemma 12: Let $T > 0$ and P a bounded, wide-sense increasing function $[0, T) \to \mathbb{R}$. Extend P to a function $\bar{P} \in \mathscr{F}$ by $\forall x \ge 0, \bar{P}(x) = \left\lfloor \frac{x}{T} \right\rfloor P(T^-) + P(x \bmod T)$ where $P(T^-) \stackrel{def}{=} \sup_{0 \le t < T} P(t)$.

Also, consider an affine function L, defined by $L(x) = ax + b$ for some $a \ge \frac{P(T^-)}{T}$ and some $b \in \mathbb{R}$.

If $L(x) \ge P(x)$ for all x in $[0, T)$ then $L \ge \bar{P}$.

Proof: Observe that, for $x \ge 0$, $L(x) = a\left\lfloor \frac{x}{T} \right\rfloor T + L(x \bmod T)$. Now $L(x \bmod T) \ge P(x \bmod T)$ by hypothesis. Thus

$$L(x) \ge a\left\lfloor \frac{x}{T} \right\rfloor T + P(x \bmod T) \tag{43}$$

$$\ge \frac{P(T^-)}{T}\left\lfloor \frac{x}{T} \right\rfloor T + P(x \bmod T) = \bar{P}(x) \tag{44}$$

□

Lemma 13: Let $f \in \mathscr{F}$ and a rate-latency function $\beta_{r,T}$ such that $r > 0$, $T > 0$, and $\beta_{r,T} \le f$. Assume that $\beta_{r,T}(x_1) = f(x_1)$ for $x_1 > T$.

Then there is no other rate-latency function $\beta_{r',T'}$ (i.e., with $(r', T') \ne (r, T)$) such that $\beta_{r,T} \le \beta_{r',T'} \le f$.

Proof: Assume that $\beta_{r,T} \le \beta_{r',T'} \le f$. The proof consists in showing that $(r, T) = (r', T')$.

First, we know that $\beta_{r,T}(x_1) = f(x_1)$ and $x_1 > T$; thus $r(x_1 - T) = f(x_1)$ and

$$T = x_1 - \frac{f(x_1)}{r} \tag{45}$$

Second, observe that we must have $T' \le T$, since otherwise $\beta_{r,T}(T') > 0 = \beta_{r',T'}(T')$.

Third, observe that $f(x_1) = \beta_{r,T}(x_1) \le \beta_{r',T'}(x_1) \le f(x_1)$ thus $\beta_{r',T'}(x_1) = f(x_1)$ and

$$T' = x_1 - \frac{f(x_1)}{r'} \tag{46}$$

Combining the last three paragraphs, it follows $x_1 - \frac{f(x_1)}{r'} \le x_1 - \frac{f(x_1)}{r}$, i.e., $r' \le r$. Also, we must have $r' \ge r$, since otherwise $\forall x > x_0$, $\beta_{r,T}(x) > \beta_{r',T'}(x)$ with $x_0 = \frac{rT - r'T'}{r - r'}$. Thus, $r' = r$, and it follows from Eq. (45) and Eq. (46) that $T' = T$. □

Now we proceed with the proof of Theorem 3.

1) We first show that $r_k \le r_{k+1}$ for $k = 0, \dots, w_i - 2$. Define sequence g by $g_k = \frac{1}{l_i^{\min}} \psi_i\left(k l_i^{\min}\right)$ for $k = 0, \dots, w_i - 1$. By definition, we have $g_{k+1} - g_k =$

$$1 + \frac{1}{l_i^{\min}} \sum_{j, j \ne i} \left(\min(k+2, w_j) - \min(k+1, w_j)\right) l_j^{\max} \tag{47}$$

Observe that $\left(\min(k+2, w_j) - \min(k+1, w_j)\right)$ is equal to 1 if $k+1 < w_j$, and equal to 0 otherwise. Thus, $g_{k+2} - g_{k+1} \le g_{k+1} - g_k$ for $0 \le k < w_i - 2$, which shows that $r_k \le r_{k+1}$ for $k = 0, \dots, w_i - 3$. Also, observe that $g_{k+1} - g_k \ge 1$, i.e., $r_k \le 1$, for $0 \le k \le w_i - 2$. Hence, $r_{w_i - 2} \le r_{w_i - 1}$.

2) Let $r \in [r_0^*, r_{k^*}^*]$ and let $T(r)$ be the value of T defined in the Theorem, namely, $T(r) \stackrel{def}{=} \psi_i(k l_i^{\min}) - \frac{k l_i^{\min}}{r}$, where k is defined by $r_{k-1}^* \le r < r_k^*$ if $r \in [r_0^*, r_{k^*}^*)$ and $k = k^*$ if $r = r_{k^*}^*$. We now show that $\beta_{r,T(r)} \le \gamma_i$.

We consider two cases: $r_0^* \le r < r_{k^*}^*$ or $r = r_{k^*}^*$. For the former case, for any r, apply Lemma 11 with $w = w_i$, g as defined in 1), k as defined in the paragraph above, and $a = \frac{1}{r}$. As by construction $\frac{1}{r_k} < a \le \frac{1}{r_{k-1}}$ and $\frac{1}{r_{k-1}} = g_k - g_{k-1}$, 3) and 4) are satisfied. For the latter case, apply again Lemma 11 with the same g and $w = w_i$ but now with $k = k^*$ and $a = \frac{1}{r} = \frac{1}{r_{k^*}^*}$. By construction, we have $\frac{1}{r_{k^*}^*} \ge \frac{1}{r_{k^*}} = g_{k^*+1} - g_{k^*}$ and $\frac{1}{r_{k^*}^*} \le \frac{1}{r_{k^*-1}} = g_{k^*} - g_{k^*-1}$. Thus, conditions 3) and 4) of Lemma 11 are satisfied. Let f be the corresponding function f in Lemma 11, i.e., $f(x) = g_{\lfloor x \rfloor} + x \bmod 1$ for $0 \le x < w_i$.

Note that for both cases f is the same. Also, let f_r be the corresponding function h in Lemma 11, i.e., $f_r(x) = \frac{1}{r}(x - k) + g_k$ for $0 \le x < w_i$. By Lemma 11, $f_r \ge f$.

Observe that $f(w_i^-) = \frac{1}{l_i^{\min}}\left(\psi_i((w_i-1)l_i^{\min}) + 1\right) = \frac{1}{l_i^{\min}}\left(w_i l_i^{\min} + \sum_{j,j\neq i} w_j l_j^{\max}\right) = \frac{L_{\text{tot}}}{l_i^{\min}} = \frac{w_i}{r^*}$. Then, as $f_r(x) \ge f(x)$ for $0 \le x < w_i$ and $\frac{1}{r} \ge \frac{1}{r^*} = \frac{f(w_i^-)}{w_i}$, we can apply Lemma 12 with $P = f$ and $L = f_r$. It gives us $\bar{f}$ defined by $\bar{f}(x) = \lfloor \frac{x}{w_i} \rfloor \frac{L_{\text{tot}}}{l_i^{\min}} + f(x \bmod w_i)$ such that $f_r \ge \bar{f}$.

Then, by using Eq. (38), $f_r^{\downarrow} \le \bar{f}^{\downarrow}$. Also, as $\bar{f}^{\downarrow} \ge 0$, we have $\left[f_r^{\downarrow}\right]^+ \le \bar{f}^{\downarrow}$. Note that for an increasing, linear function L, defined by $\forall x \ge 0, L(x) = ax + b$ with some $a > 0$ and $b > 0$, we have $\left[L^{\downarrow}\right]^+ = \beta_{\frac{1}{a},b}$; and observe that $f_r(x) = \frac{x}{r} + g_k - \frac{k}{r} = \frac{x}{r} + \frac{T(r)}{l_i^{\min}}$. Hence, $\left[f_r^{\downarrow}\right]^+ = \beta_{r,\frac{T(r)}{l_i^{\min}}}$.

Until now, we have shown that $\beta_{r,\frac{T(r)}{l_i^{\min}}} \le \bar{f}^{\downarrow}$. Lastly, we show that $l_i^{\min}\bar{f}^{\downarrow}(\frac{x}{l_i^{\min}}) = \gamma_i(x)$ and $l_i^{\min}\beta_{r,\frac{T(r)}{l_i^{\min}}}(\frac{x}{l_i^{\min}}) = \beta_{r,T(r)}(x)$. Observe that $l_i^{\min}\bar{f}(\frac{x}{l_i^{\min}}) = \lfloor \frac{x}{w_i l_i^{\min}} \rfloor L_{\text{tot}} + \psi_i((\frac{x}{l_i^{\min}} \bmod w_i) l_i^{\min})$. Also, $\psi_i(x) = \lfloor \frac{x}{w_i l_i^{\min}} \rfloor L_{\text{tot}} + \psi_i(x \bmod w_i l_i^{\min})$. Hence, we have $\psi_i(x) = l_i^{\min}\bar{f}(\frac{x}{l_i^{\min}})$. By using Lemma 9 with $l = m = l_i^{\min}$, $l_i^{\min}\bar{f}^{\downarrow}(\frac{x}{l_i^{\min}}) = \psi_i^{\downarrow}(x) = \gamma_i(x)$. Also, observe that $l_i^{\min}\beta_{r,\frac{T(r)}{l_i^{\min}}}(\frac{x}{l_i^{\min}}) = \beta_{r,T(r)}(x)$.

Combine the last paragraphs to conclude that $\beta_{r,T(r)} \le \gamma_i$ for all r in $[r_0^*, r_{k^*}^*]$.

3) We now show that for any $r \in [r_0^*, r_{k^*}^*]$, $\beta_{r,T(r)}$ is a non-dominated lower-bound of γ_i. Let $r' \ge 0, T' \ge 0$ such that $\beta_{r,T(r)} \le \beta_{r',T'} \le \gamma_i$. We have to show that $r' = r$ and $T' = T(r)$.

First, if r in $[r_0^*, r_{k^*}^*)$, observe that $\beta_{r,T(r)}(x) = \gamma_i(x)$ for $x = \psi_i(kl_i^{\min}) > \psi_i(kl_i^{\min}) - \frac{kl_i^{\min}}{r} = T(r)$. Then, apply Lemma 13 with $\beta_{r,T} = \beta_{r,T(r)}$ and $f = \gamma_i$ to conclude that $r' = r$ and $T' = T(r)$.

Second, if $r = r_{k^*}^*$, observe that $\beta_{r,T(r)}(x) = \gamma_i(x)$ for $x = \psi_i(k^* l_i^{\min}) + L_{\text{tot}} > T(r)$. Again, apply Lemma 13 with $\beta_{r,T} = \beta_{r,T(r)}$ and $f = \gamma_i$ to conclude that $r' = r$ and $T' = T(r)$.

4) We now show that there is no other non-dominated rate-latency function, $\beta_{r',T'}$, that is upper bounded by γ_i.

First, we must have $T' \ge T(r_0^*)$. This is because $\gamma_i(x) = 0$ for $x \le \psi_i(0) = T(r_0^*)$.

Second, we must have $r' \ge r_0^*$. Otherwise, we have $r' < r_0^*$ and we previously showed $T' \ge T(r_0^*)$. Thus, $\beta_{r',T'} \le \beta_{r_0^*,T(r_0^*)} \le \gamma_i$, which is in contradiction with $\beta_{r',T'}$ being non-dominated.

Third, we must have $r' \le r_{k^*}^*$. We proceed to prove this by contradiction. If $T' \ge T(r_{k^*}^*)$ and $r' > r_{k^*}^*$, observe that $\beta_{r',T'}(x_0) = \beta_{r_{k^*}^*,T(r_{k^*}^*)}(x_0)$ with $x_0 = \frac{r'T' + r_{k^*}^* T(r_{k^*}^*)}{r' - r_{k^*}^*}$ and $\forall x, x > x_0 \Rightarrow \beta_{r',T'}(x) > \beta_{r_{k^*}^*,T(r_{k^*}^*)}(x)$; for any arbitrary, non-negative integer k, let x_k be defined by $x_k = \psi_i(k^* l_i^{\min}) + kL_{\text{tot}}$. Then observe that $\beta_{r_{k^*}^*,T(r_{k^*}^*)}(x_k) = \gamma_i(x_k)$. Choose some k large enough such that $x_k > x_0$; then, $\beta_{r',T'}(x_k) > \beta_{r_{k^*}^*,T(r_{k^*}^*)}(x_k) = \gamma_i(x_k)$, which is in contradiction with $\beta_{r',T'} \le \gamma_i$. Also, if $T' < T(r_{k^*}^*)$ and $r' > r_{k^*}^*$, we have $\forall x, x > T' \Rightarrow \beta_{r',T'}(x) > \beta_{r_{k^*}^*,T(r_{k^*}^*)}(x)$. Choose some k large enough such that $x_k > T'$; then, $\beta_{r',T'}(x_k) > \beta_{r_{k^*}^*,T(r_{k^*}^*)}(x_k) = \gamma_i(x_k)$, which is in contradiction with $\beta_{r',T'} \le \gamma_i$. Therefore, $r' > r_{k^*}^*$ is in contradiction with $\beta_{r',T'} \le \gamma_i$.

Therefore, we must have r' in $[r_0^*, r_{k^*}^*]$. We now show that $T' = T(r')$. Because otherwise, if $T' < T(r')$, we have $\beta_{r',T(r')} \le \beta_{r',T'} \le \gamma_i$, which is in contradiction with $\beta_{r',T(r')}$ being a non-dominated rate latency function. Also, if $T' > T(r')$, we have $\beta_{r',T'} \le \beta_{r',T(r')} \le \gamma_i$, which is in contradiction with $\beta_{r',T'}$ being non-dominated. □

8.4 Proof of Theorem 4

Let us call the supremum of all non-dominated rate-latency functions B. We want to show that $B = \max\left(\beta_{r_0^*,T_0^*}, \ldots, \beta_{r_{k^*}^*,T_{k^*}^*}\right)$. The proof consists on three steps.

1) $B(x) = 0$ for all x in $[0, l_i^{\min} g_0]$.

2) $B(x) = \beta_{r_{k-1}^*,T_{k-1}^*}(x)$ for all x in $[l_i^{\min} g_{k-1}, l_i^{\min} g_k]$ and $k = 1 \ldots k^*$.

3) $B(x) = \beta_{r_{k^*}^*,T_{k^*}^*}(x)$ for all $x \ge l_i^{\min} g_{k^*}$.

To prove 1), as every non-dominated rate-latency function is equal to zero before $l_i^{\min} g_0$, we have $B(x) = 0$ for all x in $[0, l_i^{\min} g_0]$.

To prove 2), we consider two cases for any other non-dominated rate-latency $\beta_{r',T(r')}$: First, $r_0^* \le r' < r_{k-1}^*$. Second, $r_{k-1}^* < r' \le r_{k^*}^*$.

For the former case, we show that

$$\beta_{r',T(r')}\left(l_i^{\min} g_{k-1}\right) \le \beta_{r_{k-1}^*,T_{k-1}^*}\left(l_i^{\min} g_{k-1}\right) \tag{48}$$

Then, as $r' < r_{k-1}^*$, it follows $\beta_{r',T(r')}(x) \le \beta_{r_{k-1}^*,T_{k-1}^*}(x)$ for all x in $[l_i^{\min} g_{k-1}, l_i^{\min} g_k]$.

Let k' defined by $r' \in [r_{k'-1}^*, r_{k'}^*)$. Then, by definition $\beta_{r',T(r')}\left(l_i^{\min} g_{k-1}\right) =$

$$r'\left(l_i^{\min} g_{k-1} - l_i^{\min} g_{k'}\right) + k' l_i^{\min} \tag{49}$$

$$= r'\left(\sum_{e=k'}^{k-2} g_{e+1} - g_e\right) l_i^{\min} + k' l_i^{\min} \tag{50}$$

$$\le r_{k'}^*\left(\sum_{e=k'}^{k-2} g_{e+1} - g_e\right) l_i^{\min} + k' l_i^{\min} \tag{51}$$

$$= \frac{\sum_{e=k'}^{k-2} g_{e+1} - g_e}{g_{k'+1} - g_{k'}} l_i^{\min} + k' l_i^{\min} \tag{52}$$

Then, as $g_{e+1} - g_e$ is decreasing, we have $\sum_{e=k'}^{k-2} g_{e+1} - g_e \le (k - 1 - k')(g_{k'+1} - g_{k'})$. Combine it with Eq. (52) to conclude that $\beta_{r',T(r')}\left(l_i^{\min} g_{k-1}\right) \le (k-1)\, l_i^{\min}$; lastly, observe that $\beta_{r_{k-1}^*,T_{k-1}^*}\left(l_i^{\min} g_{k-1}\right) = (k-1)\, l_i^{\min}$. Therefore, Eq. (48) is proven.

For the latter case, we show that

$$\beta_{r',T(r')}\left(l_i^{\min} g_k\right) \le \beta_{r_{k-1}^*,T_{k-1}^*}\left(l_i^{\min} g_k\right) \tag{53}$$

Then, as $r' > r_{k-1}^*$, it follows $\beta_{r',T(r')}(x) \le \beta_{r_{k-1}^*,T_{k-1}^*}(x)$ for all

x in $[l_i^{\min} g_{k-1}, l_i^{\min} g_k]$.

Let k' defined by $r' \in [r^*_{k'-1}, r^*_{k'})$. Then, by definition $\beta_{r',T(r')}\left(l_i^{\min} g_k\right) =$

$$r'\left(l_i^{\min} g_k - l_i^{\min} g_{k'}\right) + k' l_i^{\min} \tag{54}$$

$$= r'\left(-\sum_{e=k}^{k'-1} g_{e+1} - g_e\right) l_i^{\min} + k' l_i^{\min} \tag{55}$$

$$\leq r^*_{k'}\left(-\sum_{e=k}^{k'-1} g_{e+1} - g_e\right) l_i^{\min} + k' l_i^{\min} \tag{56}$$

$$\leq \frac{-\sum_{e=k}^{k'-1} g_{e+1} - g_e}{g_{k'+1} - g_{k'}} l_i^{\min} + k' l_i^{\min} \tag{57}$$

Then, as $g_{e+1} - g_e$ is decreasing, we have $-\sum_{e=k}^{k'-1} g_{e+1} - g_e \leq (k - k')(g_{k'+1} - g_{k'})$. Combine it with Eq. (57) to conclude that $\beta_{r',T(r')}\left(l_i^{\min} g_k\right) \leq k l_i^{\min}$; lastly, observe that $\beta_{r^*_{k-1},T^*_{k-1}}\left(l_i^{\min} g_k\right) = k l_i^{\min}$. Therefore, Eq. (53) is proven.

Combining these two cases, 2) is proven.

To prove 3), for any other non-dominated rate-latency $\beta_{r',T(r')}$, we show that

$$\beta_{r',T(r')}\left(l_i^{\min} g_{k^*}\right) \leq \beta_{r^*_{k^*},T^*_{k^*}}\left(l_i^{\min} g_{k^*}\right) \tag{58}$$

Then, as $r' < r^*_{k^*}$, it follows $\beta_{r',T(r')}(x) \leq \beta_{r^*_{k^*},T^*_{k^*}}(x)$ for all $x \geq l_i^{\min} g_{k^*}$.

Let k' defined by $r' \in [r^*_{k'-1}, r^*_{k'})$. Then, by definition $\beta_{r',T(r')}\left(l_i^{\min} g_{k^*}\right) =$

$$r'\left(l_i^{\min} g_{k^*} - l_i^{\min} g_{k'}\right) + k' l_i^{\min} \tag{59}$$

$$= r'\left(\sum_{e=k'}^{k^*-1} g_{e+1} - g_e\right) l_i^{\min} + k' l_i^{\min} \tag{60}$$

$$\leq r^*_{k'}\left(\sum_{e=k'}^{k^*-1} g_{e+1} - g_e\right) l_i^{\min} + k' l_i^{\min} \tag{61}$$

$$\leq \frac{\sum_{e=k'}^{k^*-1} g_{e+1} - g_e}{g_{k'+1} - g_{k'}} l_i^{\min} + k' l_i^{\min} \tag{62}$$

Then, as $g_{e+1} - g_e$ is decreasing, we have $\sum_{e=k'}^{k^*-1} g_{e+1} - g_e \leq (k^* - k')(g_{k'+1} - g_{k'})$. Combine it with Eq. (62) to conclude that $\beta_{r',T(r')}\left(l_i^{\min} g_{k^*}\right) \leq k^* l_i^{\min}$; lastly, observe that $\beta_{r^*_{k^*},T^*_{k^*}}\left(l_i^{\min} g_{k^*}\right) = k^* l_i^{\min}$. Therefore, Eq. (58) is proven.

Until now, we have shown that **1)**, **2)**, and **3)**. Let $A = \max\left(\beta_{r^*_0,T^*_0}, \ldots, \beta_{r^*_{k^*},T^*_{k^*}}\right)$. Observe that first, by **1)**, it follows $A = 0$ for all x in $[0, l_i^{\min} g_0]$; second, by **2**, it follows $A(x) = \beta_{r^*_{k-1},T^*_{k-1}}(x)$ for all x in $[l_i^{\min} g_{k-1}, l_i^{\min} g_k]$ and $k = 1 \ldots k^*$; lastly, by **3)**, $A(x) = \beta_{r^*_{k^*},T^*_{k^*}}(x)$ for all $x \geq l_i^{\min} g_{k^*}$. Therefore, $A = B$, i.e., $B = \max\left(\beta_{r^*_0,T^*_0}, \ldots, \beta_{r^*_{k^*},T^*_{k^*}}\right)$.

We now want to show that B is the largest convex function upper bounded by γ_i, i.e., if f is a convex function and is upper bounded by γ_i, then $f \leq B$.

Pick an arbitrary $x \geq 0$. Let G_x be a subgradient of f at x. Note that a subgradient exists because f is convex [23, Sec. 5.4]. By definition of subgradient [23, Sec. 5.4], for L, defined by $\forall x' \geq 0, L(x') = G_x(x'-x) + f(x)$, we have $L \leq f$; then, as $f \leq \gamma_i$, we have $L \leq \gamma_i$. We now consider two cases for G_x and proceed the proof to show that $f(x) \leq B(x)$ in both cases.

Case 1: $G_x \leq 0$

As $L(0) \leq \gamma_i(0) = 0$, we have $L \leq 0$; also, observe that $B \geq 0$. Hence, $L \leq B$. It follows $L(x) = f(x) \leq B(x)$.

Case 1: $G_x > 0$

Define $\beta_{r,T}$ with $r = G_x$ and $T = x - \frac{f(x)}{G_x}$. Observe that $r \geq 0$ and as $L(0) \leq \gamma_i(0) = 0$, it follows $T \geq 0$. We now proceed to show that $\beta_{r,T} \leq \gamma_i$. As $L \leq \gamma_i$ and $\gamma_i \geq 0$, we have $[L]^+ \leq \gamma_i$; also, observe that $[L]^+ = \beta_{r,T}$. Therefore, $\beta_{r,T} \leq \gamma_i$.

Then, as $\beta_{r,T}$ is a rate-latency function upper bounded by γ_i, it is dominated by one of non-dominated rate-latencies or is equal to one of them. It follows $\beta_{r,T} \leq B$; also, observe that $\beta_{r,T}(x) = f(x)$. Thus, $f(x) \leq B(x)$.

Lastly, the above result applies to any $x \geq 0$, thus $\forall x \geq 0, f(x) \leq B(x)$, i.e., $f \leq B$. □

8.5 Proof of Theorem 5

We use the following lemma about the lower pseudo-inverse technique.

Lemma 14: For a right-continuous function f in $\mathscr{F}$ and x, y in $\mathbb{R}^+$, $f^{\downarrow}(y) = x$ if and only if $f(x) \geq y$ and there exists some $\varepsilon > 0$ such that $\forall x' \in (x - \varepsilon, x), f(x') < y$.

Proof:

$\Rightarrow$:

Let $S = \{x', f(x') \geq y\}$ so that $x = \inf S$ (see Eq. (2)). From the definition of an inf, there exists a sequence x_n such that $x_n \in S$ for all n, $x_n \geq x$, and $\lim_{n\to\infty} x_n = x$. Since f is right-continuous, $\lim_{n\to\infty} f(x_n) = f(x)$, which shows that $f(x) \geq y$. Also, again by definition of an inf, any $x' < x$ does not belong to S, i.e. $\forall x' < x, f(x') < y$.

$\Leftarrow$:

By the first part of the hypothesis, $x \in S$ therefore $x \geq \inf S = f^{\downarrow}(y)$. Let also $S' = \{x', f(x') < y\}$ so that $f^{\downarrow}(y) = \sup S'$ (see Eq. (2)). By the second part of the hypothesis, S' contains the interval $(x-\varepsilon, x)$ hence $\sup S' \geq x$, which shows that $f^{\downarrow}(y) \geq x$. Combining the two shows that $f^{\downarrow}(y) = x$. □

Proof of Theorem 5

We prove that, for any value of the system parameters, for any $\tau > 0$, and for any flow i, there exists one trajectory of a system such that

$$\begin{aligned} &\exists s \geq 0,\ (s, s+\tau] \text{ is backlogged for flow } i \\ &\text{and } R_i^*(s+\tau) - R_i^*(s) = \beta_i(\tau) \end{aligned} \tag{63}$$

Step 1: Constructing the Trajectory

1) Flows are labeled in order of weights, i.e., $w_j \leq w_{j+1}$.

2) At time 0, the input of every queue $j \neq i$ is a burst of size $\left\lceil \frac{\beta(\tau)}{l_j^{\max}} \right\rceil l_j^{\max} + w_j l_j^{\max}$.

3) Every flow, $j \neq i$, is packetized according to its maximum packet size, $l_j^{\max}$.

4) The output of the system is at rate K (the Lipschitz constant of β) from time 0 to times s, which is defined as the time at which queue i is visited at cycle w_i in the first round, namely

$$s = \frac{1}{K}\sum_{j,j\neq i} \min\left(w_i - 1, w_j\right) l_j^{\max} \tag{64}$$

It follows that

$$\forall t \in [0, s], R^*(t) = Kt \tag{65}$$

5) The input of queue i starts just after time s, with a burst of size $\left\lceil \frac{\beta(\tau)}{l_i^{\min}} \right\rceil l_i^{\min}$.

6) Flow i is packetized according to its minimum packet size, $l_i^{\min}$.

7) After time s, the output of the system is equal to the guaranteed service; by 2) and 5), the busy period lasts for at least τ, i.e.,

$$\forall t \in [s, s+\tau], R^*(t) = R^*(s) + \beta(t-s) \tag{66}$$

In particular,

$$R^*(s+\tau) - R^*(s) = \beta(\tau) \tag{67}$$

If we apply $\psi_i^{\downarrow}$ to both sides of Eq. (67), the right-hand side is equal to $\beta_i(\tau)$. Thereby, we should prove

$$\psi_i^{\downarrow}\left(R^*(s+\tau) - R^*(s)\right) = R_i^*(s+\tau) - R_i^*(s) \tag{68}$$

Let $y = R^*(s+\tau) - R^*(s)$ and $x = R_i^*(s+\tau) - R_i^*(s)$. Our goal is now to prove that

$$\psi_i^{\downarrow}(y) = x \tag{69}$$

From 5), we know that the first packet of flow i is served at the first cycle of a round ($C = 1$ in Algorithm 1). Thus, applying Lemma 2 and (P5) in Lemma 3, the number of services to each flow j is equal to $\phi_{i,j}(p)$. From 2), flow j sends packets with the maximum length. Thus

$$\sum_{j,j\neq i} R_j^*(s+\tau_{\sigma(p)}) - R_j^*(s) = \sum_{j,j\neq i} \phi_{i,j}(p) l_j^{\max} \tag{70}$$

Now there are two cases for $s+\tau$ (8.1.1).

Case 1: $s+\tau < \tau_{\sigma(p)}$ In this case the scheduler is not serving flow i in $[\tau_{\sigma(p)}, s+\tau]$ and $x = p l_i^{\min}$. Thus $R_i^*(s+\tau) = R_i^*(\tau_{\sigma(p)})$. It follows that

$$\begin{aligned} \psi_i(x) &= x + \underbrace{\sum_{j,j\neq i} \phi_{i,j}(\lfloor \frac{x}{l_i^{\min}} \rfloor) l_j^{\max}}_{\sum_{j,j\neq i} R_j^*(\tau_{\sigma(p)}) - R_j^*(s)} \\ y &= x + \sum_{j,j\neq i} R_j^*(s+\tau) - R_j^*(s) \end{aligned} \tag{71}$$

and thus

$$\psi_i(x) \geq y \tag{72}$$

Let $x - l_i^{\min} < x' < x$; flow i's output becomes equal to x' during the emission of packet $p-1$ thus

$$\psi_i(x') = x' + \sum_{j,j\neq i} R_j^*(\tau_{\sigma(p-1)}) - R_j^*(s) \tag{73}$$

Hence

$$\forall x' \in (x - l_i^{\min}, x), \psi_i(x') < y \tag{74}$$

Combining Eq. (72) and Eq. (74) with Lemma 14 shows Eq. (69).

Case 2: $s+\tau \geq \tau_{\sigma(p)}$ In this case the scheduler is serving flow i in $[\tau_{\sigma(p)}, s+\tau]$. For every other flow j, we have $R_j^*(s+\tau) = R_j^*(\tau_{\sigma(p)})$. Hence,

$$\psi_i(x) = R_i^*(s+\tau) - R_i^*(s) + \sum_{j,j\neq i} \phi_{i,j}(p) l_j^{\max} = y \tag{75}$$

As with case 1, for any $x' \in ((p-1)l_i^{\min}, x)$, we have $\psi_i(x) < y$, which shows Eq. (69).

This shows that Eq. (63) holds. It remains to show that the system constraints are satisfied.

Step 2: Verifying the Trajectory

We need to verify that the service offered to the aggregate satisfies the strict service curve constraint. Our trajectory has one busy period, starting at time 0 and ending at some time $T_{\max} \geq \tau$. We need to verify that

$$\forall t_1, t_2 \in [0, T_{\max}] \text{ with } t_1 < t_2, R^*(t_2) - R^*(t_1) \geq \beta(t_2 - t_1) \tag{76}$$

Case 1: $t_2 < s$

Then $R^*(t_2) - R^*(t_1) = K(t_2 - t_1)$. Observe that, by the Lipschitz continuity condition on β, for all $t \geq 0$, $\beta(t) = \beta(t) - \beta(0) = \beta(t) \leq Kt$ thus $K(t_2 - t_1) \geq \beta(t_2 - t_1)$.

Case 2: $t_1 < s \leq t_2$

Then $R^*(t_2) - R^*(t_1) = \beta(t_2 - s) + K(s - t_1)$. By the Lipschitz continuity condition:

$$\beta(t_2 - t_1) - \beta(t_2 - s) \leq K(s - t_1) \tag{77}$$

thus $R^*(t_2) - R^*(t_1) \geq \beta(t_2 - t_1)$.

Case 3: $s \leq t_1 < t_2$

Then $R^*(t_2) - R^*(t_1) = \beta(t_2) - \beta(t_1) \geq \beta(t_2 - t_1)$ because β is super-additive. □

8.6 Proof of Theorem 6

Proof: The proof is very similar to the proof of Theorem 5. The necessary changes in the proof are the following:

1) s is the time of the first visit to flow i.

2) Instead of functions ψ_i and $\phi_{i,j}$, use functions ψ'_i and $\phi'_{i,j}$, defined in Eq. (34) and Eq. (35). □

8.7 Proof of Theorem 7

Proof: The proof contains the following steps:

1) Consider the same trajectory as in the proof of Theorem 5, yet with one difference: the input of flow i is $R_i(t) = \alpha_i(t-s)$ for $t \geq s$ and zero before s. Observer that as α_i is sub-additive, $\forall t_1, t_2$: $t_2 \geq t_1 \geq s \Rightarrow R_i(t_2) - R_i(t_1) = \alpha_i(t_2) - \alpha_i(t_1) \leq \alpha_i(t_2 - t_1)$.

2) Define $s' = \inf\{u > 0 | \alpha_i(u) \leq \beta_i(u)\}$. This is the first time after zero that the service curve meets the arrival curve. Note that s' can be infinite as well.

3) Then, it is guaranteed that flow i is backlogged in $(s, s+s']$. Therefore, using Eq. (63), we have $R_i^*(t) = \beta_i(t-s)$ for $t \geq s$ and zero before s.

4) Combining 1 and 3, the horizontal deviation of R_i and R_i^* in $(s, s+s']$ is equal to the horizontal deviation of α_i and β_i in $[0, s']$.

4) Using [2, Sec. 5.3.3], the horizontal deviation of α_i and β_i can be restricted to $[0, s']$.

Thereby, we find a valid trajectory (verified in the proof of Theorem 5) where the delay bound is achieved. □

8.8 Proof of Theorem 8

Proof: The same proof of Theorem 7 works here as well. However, we use the trajectory defined in the proof of Theorem 6. □

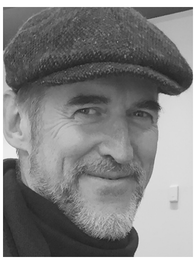
9. Conclusion

IWRR is a variant of WRR with the same long-term rate and the same complexity. We have provided a residual strict service curve for IWRR and have shown that it is the best possible one under general assumptions. For flows with packets of constant size, we have shown that the delay bounds derived from it are worst-case. We have proved that IWRR worst-case delay is not greater than WRR and shown on experiments that the gain is significant (20%–60%) in practice, which speaks in favour of using IWRR as a replacement to WRR. Our result assumes that the aggregate of all IWRR queues receives a strict service curve guarantee, and we find a strict service curve guarantee for every IWRR queue. Therefore, our results apply to hierarchical schedulers. In future research, we plan to improve the results with supplementary hypotheses on flows, considering arrival curves and packet size distribution, with "packet curves" [24].

References

[1] M. Katevenis, S. Sidiropoulos, and C. Courcoubetis, "Weighted round-robin cell multiplexing in a general-purpose ATM switch chip," IEEE J. Sel. Areas Commun., vol.9, no.8, pp.1265–1279, 1991.
[2] A. Bouillard, M. Boyer, and E. Le Corronc, Deterministic Network Calculus: From Theory to Practical Implementation, Wiley-ISTE, 2018.
[3] J.Y. Le Boudec and P. Thiran, Network Calculus: A Theory of Deterministic Queuing Systems for the Internet, Springer Science & Business Media, 2001.
[4] C.S. Chang, Performance Guarantees in Communication Networks, Springer-Verlag, New York, 2000.
[5] J. Liebeherr, "Duality of the max-plus and min-plus network calculus," Foundations and Trends in Networking, vol.11, no.3-4, pp.139–282, 2017.
[6] "IEEE standard for local and metropolitan area networks — Bridges and bridged networks," IEEE Standard 802.1Q, IEEE, 2018.
[7] D.B. Chokshi and P. Bhaduri, "Modeling fixed priority non-preemptive scheduling with real-time calculus," 2008 14th IEEE International Conference on Embedded and Real-Time Computing Systems and Applications, pp.387–392, Aug. 2008.
[8] L. Thiele, S. Chakraborty, and M. Naedele, "Real-time calculus for scheduling hard real-time systems," 2000 IEEE International Symposium on Circuits and Systems (ISCAS), vol.4, pp.101–104, May 2000.
[9] E.L. Hahne and R.G. Gallager, "Round robin scheduling for fair flow control in data communication networks," Proc. IEEE Int. Conf. on Communications (ICC 86), June 1986.
[10] J. Nagle, "On packet switches with infinite storage," IEEE Trans. Commun., vol.35, no.4, pp.435–438, April 1987.
[11] Y.T. Wang, T.P. Lin, and K.C. Gan, "An improved scheduling algorithm for weighted round-robin cell multiplexing in an ATM switch," Proc. International Conference on Communications (SUPERCOMM'94), vol.2, pp.1032–1037, May 1994.
[12] H. Shimonishi, M. Yoshida, R. Fan, and H. Suzuki, "An improvement of weighted round robin cell scheduling in ATM networks," Proc. IEEE Global Telecommunications Conference (GLOBECOM 97) vol.2, pp.1119–1123, Nov. 1997.
[13] H.M. Chaskar and U. Madhow, "Fair scheduling with tunable latency: A round robin approach," Proc. IEEE Global Telecommunications Conference (GLOBECOM'99), pp.1328–1333, IEEE, 1999.
[14] L.D. Dhinesh Babu and P.V. Krishna, "Honey bee behavior inspired load balancing of tasks in cloud computing environments," Applied Soft Computing, vol.13, no.5, pp.2292–2303, 2013.
[15] Wensong, "Weighted round-robin scheduling, documentation of the linuxvirtualserver knowledge base," 2005.
[16] Y. Qian, Z. Lu, and W. Dou, "Analysis of worst-case delay bounds for best-effort communication in wormhole networks on chip," Proc. 3rd ACM/IEEE International Symposium on Networks-on-Chip (NoCS 2009), pp.44–53, IEEE, 2009.
[17] D. Stiliadis and A. Varma, "Latency-rate servers: A general model for analysis of traffic scheduling algorithms," IEEE/ACM Trans. Netw., vol.6, no.5, pp.611–624, Oct. 1998.
[18] S. Nananukul, "Latency of weighted round-robin scheduler," Electron. Lett., vol.39, no.2, pp.256–257, 2003.
[19] Y. Jiang, "Relationship between guaranteed rate server and latency rate server," Computer Networks, vol.43, no.3, pp.307–315, 2003.
[20] J.P. Georges, T. Divoux, and E. Rondeau, "Network calculus: Application to switched real-time networking," Proc. 5th Int. ICST Conf. on Performance Evaluation Methodologies and Tools, VALUETOOLS'11, ICST, Brussels, Belgium, Belgium, pp.399–407, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2011.
[21] A. Soni, X. Li, J.L. Scharbarg, and C. Fraboul, "WCTT analysis of avionics switched ethernet network with WRR scheduling," Proc. 26th International Conference on Real-Time Networks and Systems (RTNS), pp.213–222, ACM, 2018.
[22] E. Mohammadpour, E. Stai, and J.Y. Le Boudec, "Improved delay bound for a service curve element with known transmission rate," IEEE Netw. Lett., vol.1, no.4, pp.156–159, 2019.
[23] D. Bertsekas, Convex Optimization Theory, Athena Scientific Optimization and Computation Series, Athena Scientific, 2009.
[24] A. Bouillard, N. Farhi, and B. Gaujal, "Packetization and packet curves in network calculus," Performance Evaluation Methodologies and Tools (VALUETOOLS), 2012 6th International Conference on, pp.136–137, IEEE, 2012.

Seyed Mohammadhossein Tabatabaee received his bachelor's degree in electrical engineering with a major in Telecommunication from Sharif University of Technology, Tehran, Iran in 2019. He is currently a PhD student in Computer and Communication Sciences at École Polytechnique Fédérale de Lausanne (EPFL) in Lausanne, Switzerland. His research interest is the analysis of deterministic networking to obtain performance bounds (e.g., delay and backlog) for worst cases.



Jean-Yves Le Boudec is a professor at EPFL and fellow of the IEEE. He graduated from Ecole Normale Supérieure de Saint-Cloud, Paris, where he obtained the Agrégation in Mathematics in 1980 and received his doctorate in 1984 from the University of Rennes, France. From 1984 to 1987 he was with INSA/IRISA, Rennes. In 1987 he joined Bell Northern Research, Ottawa, Canada, as a member of scientific staff in the Network and Product Traffic Design Department. In 1988, he joined the IBM Zurich Research Laboratory where he was manager of the Customer Premises Network Department. In 1994 he became associate professor at EPFL. His interests are in the performance and architecture of communication systems and smart grids. He co-authored a book on network calculus, which serves as a foundation for deterministic networking, an introductory textbook on Information Sciences, and is the author of the book "Performance Evaluation."

Marc Boyer received it's engineer diploma in computer science from ENSEEIHT (Toulouse, France) in 1996, and it's PhD in computer science from University Paul Sabatier (Toulouse, France) in 2001. From 2002 to 2008, he was assistant professor at ENSEEIHT, in the networking department. From 2008, he is researcher at ONERA, working on embedded networks. He is mainly interested in formal methods to compute upper bounds on network delay, especially the Network Calculus method.