

Flow vs. Time Sampling for Throughput Performance Evaluation

Thomas Bonald, Minh-Anh Tran

► **To cite this version:**

Thomas Bonald, Minh-Anh Tran. Flow vs. Time Sampling for Throughput Performance Evaluation. Performance Evaluation, Elsevier, 2007, 10.1016/j.peva.2007.06.010 . hal-01286338

HAL Id: hal-01286338

<https://hal.archives-ouvertes.fr/hal-01286338>

Submitted on 10 Mar 2016

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

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

Flow vs. Time Sampling for Throughput Performance Evaluation

T. Bonald^{a,*} M-A. Tran^b

^a*Orange Labs, 38-40 rue du général Leclerc, 92794 Issy-les-Moulineaux Cedex 9, France.*

^b*École Normale Supérieure, 45 rue d'Ulm, 75230 Paris Cedex 5, France.*

Abstract

We introduce two throughput metrics referred to as *flow-* and *time-sampled* throughputs. The former gives the throughput statistics of an arbitrary flow while the latter weights these throughput statistics by the flow durations. Under fair sharing assumptions, the latter is shown to coincide with the steady-state instantaneous throughput *weighted* by the number of flows, which provides a useful means to measure and estimate it. We give some generic properties satisfied by both metrics and illustrate their difference on a few examples.

Key words: Traffic modeling, Throughput performance metric

1 Introduction

Throughput is a key performance indicator for Internet access. Users perceive the quality of their Internet access through the time necessary to download and upload various files. Providers need throughput measurements to control quality of service and to take appropriate decisions about equipment upgrades. They also need simple analytical results that relate user throughput to the degree of network congestion in order to determine target traffic loads and to assess the efficiency of various control schemes like routing, scheduling and buffer management.

* Corresponding author.

Email addresses: thomas.bonald@orange-ftgroup.com (T. Bonald),
minh-anh.tran@polytechnique.org (M-A. Tran).

Although throughput is a very common performance metric, it is actually not immediately obvious how to measure and estimate it. One of the difficulties comes from the random nature of traffic. Throughput is most often the result of the statistical multiplexing of data flows that share the same network resources. These flows arrive at random and leave the network at the end of the corresponding document transfer. They adapt elastically to congestion, typically under the control of TCP [18]. Thus throughput is inherently random, determined by the complex interaction of a dynamic set of elastic flows through the congestion control algorithms.

Another difficulty is related to the well-known *observer's paradox*. The simplest way to measure the throughput achieved by users during a file download is to actually download this file, which changes the network state. Similarly, calculating user throughput by means of a traffic model requires the study of the transient behaviour of the system, given the presence of the considered file transfer. In both cases, it would be much more convenient to derive throughput estimates from the steady state of the network, either measured through passive monitoring tools or calculated using an appropriate traffic model.

In this paper, we introduce two throughput metrics referred to as *flow-* and *time-sampled* throughputs. The former gives the throughput statistics of an arbitrary flow. It turns out to be hard to evaluate analytically even for the simplest traffic models. This is why it is most often approximated by the steady-state instantaneous throughput [4,13,16]. We shall see that this approximation is in fact far from accurate. The latter weights the throughput statistics by the flow durations. A key result of the paper is that, under fair sharing assumptions, this throughput metric coincides with the steady-state instantaneous throughput *weighted* by the number of flows. This yields explicit, insensitive expressions for a large class of traffic models [2,5,9,12,15].

We highlight the fact that both throughput metrics are random variables that represent the whole throughput distribution. This is particularly relevant for adaptive streaming flows that typically require a minimum throughput. Though we mainly focus on elastic traffic, the results may be used as a conservative approximation for a more realistic scenario where both elastic and adaptive streaming flows are multiplexed [6]. It is also worth noting that many elastic flows in fact correspond to streamed audio and video files with long playback delays (a few seconds, say). It is clear that the mean throughput is not a good metric for assessing the quality of such flows. Asymptotic results on flow duration and throughput, see e.g. [3,7,11], are also not sufficient for this purpose. What is really needed is either the throughput distribution during the flow, as considered in the present paper, or some bounds or approximations on the flow duration conditionally to the flow size, as described in [8] for instance.

We describe the traffic model in the next section. Sections 3 and 4 are devoted to the flow- and time-sampled throughputs, respectively. These throughput metrics are compared with the steady-state instantaneous throughput in Section 5. The results are illustrated by a few examples in Section 6. Section 7 concludes the paper.

2 Traffic model

Consider a tagged flow of potentially infinite size starting at time 0. We view this flow as a fluid stream and denote by $\varphi(t)$ its instantaneous throughput at time t . This stochastic process is typically not stationary due to the impact of the tagged flow on the system state, which may be viewed as a manifestation of the *observer's paradox*. We assume it reaches a steady state characterized by some random variable φ^∞ equal to the weak limit of $\varphi(t)$ when t tends to infinity. We also assume that the process $\varphi(t)$ takes positive values in some discrete, bounded set \mathcal{X} , is regular, right-continuous with left-hand limits, and ergodic in the sense that:

$$\forall x \in \mathcal{X}, \quad \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \mathbb{I}_{\{\varphi(t)=x\}} dt = \Pr(\varphi^\infty = x) \quad \text{a.s.} \quad (1)$$

We denote by $\varphi^0 = \varphi(0)$ the throughput of a virtual flow of null size.

If the tagged flow is of size s , its duration $D(s)$ is the random variable defined by the equality:

$$s = \int_0^{D(s)} \varphi(t) dt.$$

We have the following simple result.

Proposition 1 *The throughputs of flows of infinitely small and infinitely large sizes are respectively given by:*

$$\lim_{s \rightarrow 0} \frac{s}{D(s)} = \varphi^0 \quad \text{a.s.} \quad (2)$$

and

$$\lim_{s \rightarrow \infty} \frac{s}{D(s)} = \mathbb{E}[\varphi^\infty] \quad \text{a.s.} \quad (3)$$

Proof. Let θ be the minimum time $t > 0$ such that $\varphi(t) \neq \varphi(0)$. We have $\varphi(t) = \varphi^0$ for all $t < \theta$ so that

$$D(s) = \frac{s}{\varphi^0} \quad \text{if } s < \varphi^0\theta.$$

We deduce (2) from the fact that $\varphi^0\theta > 0$ a.s.

Now, since \mathcal{X} is bounded, we have:

$$\lim_{s \rightarrow \infty} D(s) = \infty \quad \text{a.s.} \tag{4}$$

so that (3) follows by ergodicity. \square

In the following, we denote by S a random flow size and by $D = D(S)$ the corresponding flow duration:

$$S = \int_0^D \varphi(t) dt.$$

3 Flow-sampled throughput

3.1 Definition

Assume the tagged flow has random size S . For all $x \in \mathcal{X}$, consider the mean fraction of time the tagged flow has throughput x :

$$\mathbb{E} \left[\frac{1}{D} \int_0^D \mathbb{I}_{\{\varphi(t)=x\}} dt \right]. \tag{5}$$

This defines a distribution on \mathcal{X} . We refer to the corresponding random variable X^{flow} as the *flow-sampled* throughput.

The definition extends to any given flow size $s > 0$. The corresponding random variable $X^{\text{flow}}(s)$ satisfies for all $x \in \mathcal{X}$:

$$\Pr(X^{\text{flow}}(s) = x) = \mathbb{E} \left[\frac{1}{D(s)} \int_0^{D(s)} \mathbb{I}_{\{\varphi(t)=x\}} dt \right]. \tag{6}$$

3.2 Properties

The mean flow-sampled throughput, denoted by γ^{flow} , is given by:

$$\begin{aligned}
\gamma^{\text{flow}} &= \mathbb{E}[X^{\text{flow}}], \\
&= \sum_{x \in \mathcal{X}} x \mathbb{E} \left[\frac{1}{D} \int_0^D \mathbb{I}_{\{\varphi(t)=x\}} dt \right], \\
&= \mathbb{E} \left[\frac{1}{D} \int_0^D \left(\sum_{x \in \mathcal{X}} x \mathbb{I}_{\{\varphi(t)=x\}} \right) dt \right], \\
&= \mathbb{E} \left[\frac{1}{D} \int_0^D \varphi(t) dt \right], \\
&= \mathbb{E} \left[\frac{S}{D} \right]. \tag{7}
\end{aligned}$$

Similarly, we get the mean flow-sampled throughput of a flow of size s :

$$\gamma^{\text{flow}}(s) = \mathbb{E}[X^{\text{flow}}(s)] = \mathbb{E} \left[\frac{s}{D(s)} \right]. \tag{8}$$

We have the following result for the limiting cases of flows of infinitely small and infinitely large sizes:

Proposition 2 *The flow-sampled throughput $X^{\text{flow}}(s)$ tends in distribution to φ^0 when s tends to 0 and to φ^∞ when s tends to ∞ .*

Proof. Recall that:

$$\Pr(X^{\text{flow}}(s) = x) = \mathbb{E} \left[\frac{1}{D(s)} \int_0^{D(s)} \mathbb{I}_{\{\varphi(t)=x\}} dt \right].$$

As in the proof of Proposition 1, we denote by θ be the minimum time $t > 0$ such that $\varphi(t) \neq \varphi(0)$. We have:

$$\frac{1}{D(s)} \int_0^{D(s)} \mathbb{I}_{\{\varphi(t)=x\}} dt = \mathbb{I}_{\{\varphi^0=x\}} \quad \text{if } s < \varphi^0 \theta.$$

Since $\varphi^0 \theta > 0$ a.s., we deduce:

$$\lim_{s \rightarrow 0} \frac{1}{D(s)} \int_0^{D(s)} \mathbb{I}_{\{\varphi(t)=x\}} dt = \mathbb{I}_{\{\varphi^0=x\}} \quad \text{a.s.}$$

and

$$\lim_{s \rightarrow 0} \Pr(X^{\text{flow}}(s) = x) = \Pr(\varphi^0 = x).$$

The other limit follows by ergodicity, using (1) and (4):

$$\lim_{s \rightarrow \infty} \frac{1}{D(s)} \int_0^{D(s)} \mathbb{I}_{\{\varphi(t)=x\}} dt = \Pr(\varphi^\infty = x) \quad \text{a.s.}$$

□

4 Time-sampled throughput

4.1 Definition

Assume the tagged flow has random size S . For all $x \in \mathcal{X}$, consider the ratio of the mean duration at throughput x to the mean flow duration:

$$\frac{\mathbb{E} \left[\int_0^D \mathbb{I}_{\{\varphi(t)=x\}} dt \right]}{\mathbb{E}[D]}.$$
 (9)

This defines a distribution on \mathcal{X} provided the mean flow duration is finite. We refer to the corresponding random variable X^{time} as the *time-sampled* throughput.

The definition extends to any given flow size $s > 0$. The corresponding random variable $X^{\text{time}}(s)$ satisfies for all $x \in \mathcal{X}$:

$$\Pr(X^{\text{time}}(s) = x) = \frac{\mathbb{E} \left[\int_0^{D(s)} \mathbb{I}_{\{\varphi(t)=x\}} dt \right]}{\mathbb{E}[D(s)]}.$$
 (10)

To illustrate the difference with the other throughput metric, consider K flows of size s , respective durations d_1, \dots, d_K and fractions of time at throughput x f_1, \dots, f_K . For the flow-sampled throughput metric, the probability that a flow of size s has throughput x is simply estimated as the empirical mean:

$$\frac{1}{K} \sum_{k=1}^K f_k.$$

For the time-sampled throughput metric, it is estimated as the empirical mean *weighted* by the flow durations:

$$\frac{\sum_{k=1}^K d_k f_k}{\sum_{k=1}^K d_k}.$$

This corresponds to the fraction of *time* flows have throughput x . Note that, provided the K samples are i.i.d., the previous two expressions tend to (6) and (10), respectively, as K grows to infinity.

4.2 Properties

The mean time-sampled throughput, denoted by γ^{time} , is given by:

$$\begin{aligned} \gamma^{\text{time}} &= \mathbb{E}[X^{\text{time}}], \\ &= \sum_{x \in \mathcal{X}} x \frac{1}{\mathbb{E}[D]} \mathbb{E} \left[\int_0^D \mathbb{I}_{\{\varphi(t)=x\}} dt \right], \\ &= \frac{1}{\mathbb{E}[D]} \mathbb{E} \left[\int_0^D \left(\sum_{x \in \mathcal{X}} x \mathbb{I}_{\{\varphi(t)=x\}} \right) dt \right], \\ &= \frac{1}{\mathbb{E}[D]} \mathbb{E} \left[\int_0^D \varphi(t) dt \right], \\ &= \frac{\mathbb{E}[S]}{\mathbb{E}[D]}. \end{aligned} \tag{11}$$

Similarly, we get the mean time-sampled throughput of a flow of size s :

$$\gamma^{\text{time}}(s) = \mathbb{E}[X^{\text{time}}(s)] = \frac{s}{\mathbb{E}[D(s)]}. \tag{12}$$

Again, to illustrate the difference with the other throughput metric, consider K flows of size s and respective durations d_1, \dots, d_K . The mean flow-sampled throughput is simply estimated as the empirical mean:

$$\frac{1}{K} \sum_{k=1}^K \frac{s}{d_k},$$

while the mean time-sampled throughput is estimated as the empirical mean *weighted* by the flow durations:

$$\frac{\sum_{k=1}^K d_k \frac{s}{d_k}}{\sum_{k=1}^K d_k} = \frac{s}{\frac{1}{K} \sum_{k=1}^K d_k}.$$

Provided the K samples are i.i.d., these expressions tend to (8) and (12), respectively, as K tends to infinity. They correspond to the *arithmetic* and *harmonic* empirical means, respectively.

We have the following result:

Proposition 3 *For all flow sizes, the mean time-sampled throughput is less than the mean flow-sampled throughput:*

$$\forall s > 0, \quad \gamma^{\text{time}}(s) \leq \gamma^{\text{flow}}(s).$$

Proof. The proof follows from the convex inequality:

$$E[D(s)]E\left[\frac{1}{D(s)}\right] \geq 1.$$

□

For the limiting cases of flows of infinitely small and infinitely large sizes, we get:

Proposition 4 *The time-sampled throughput $X^{\text{time}}(s)$ tends in distribution to φ^∞ when s tends to ∞ . Moreover,*

$$\lim_{s \rightarrow 0} \Pr(X^{\text{time}}(s) = x) = \frac{\Pr(\varphi^0 = x)}{xE[1/\varphi^0]}.$$

Proof. Recall that:

$$\Pr(X^{\text{time}}(s) = x) = \frac{E\left[\int_0^{D(s)} \mathbb{I}_{\{\varphi(t)=x\}} dt\right]}{E[D(s)]}.$$

By ergodicity, it follows from (1), (3) and (4) that:

$$\lim_{s \rightarrow \infty} \frac{1}{s} \int_0^{D(s)} \mathbb{I}_{\{\varphi(t)=x\}} dt = \frac{\Pr(\varphi^\infty = x)}{E[\varphi^\infty]} \quad \text{a.s.}$$

and

$$\lim_{s \rightarrow \infty} \frac{\mathbb{E}[D(s)]}{s} = \frac{1}{\mathbb{E}[\varphi^\infty]},$$

so that

$$\lim_{s \rightarrow \infty} \frac{1}{\mathbb{E}[D(s)]} \int_0^{D(s)} \mathbb{I}_{\{\varphi(t)=x\}} dt = \Pr(\varphi^\infty = x) \quad \text{a.s.}$$

For the other limit, we use again the fact that $\varphi(t) = \varphi^0$ for all $t < \theta$ so that

$$\frac{1}{D(s)} \int_0^{D(s)} \mathbb{I}_{\{\varphi(t)=x\}} dt = \mathbb{I}_{\{\varphi^0=x\}} \quad \text{if } s < \varphi^0 \theta.$$

Using (2), we deduce:

$$\lim_{s \rightarrow 0} \frac{1}{s} \int_0^{D(s)} \mathbb{I}_{\{\varphi(t)=x\}} dt = \frac{1}{x} \mathbb{I}_{\{\varphi^0=x\}} \quad \text{a.s.}$$

and

$$\lim_{s \rightarrow 0} \frac{D(s)}{s} = \frac{1}{\varphi^0} \quad \text{a.s.}$$

so that

$$\lim_{s \rightarrow 0} \frac{\mathbb{E} \left[\int_0^{D(s)} \mathbb{I}_{\{\varphi(t)=x\}} dt \right]}{\mathbb{E}[D(s)]} = \frac{\Pr(\varphi^0 = x)}{x \mathbb{E}[1/\varphi^0]}.$$

□

5 Steady-state instantaneous throughput

The stochastic process $\varphi(t)$ involves the transient behaviour of the system and is therefore very hard to characterize even for the simplest traffic models. It is tempting to derive flow- and time-sampled throughput statistics from the steady-state behaviour of the system, as observed in practice by means of passive monitoring.

We here consider a stream of tagged flows arriving at rate λ . As above, we denote by S their random size and by D their random duration. Let n be the number of active tagged flows in steady state. We assume a perfectly fair sharing among these flows and denote by φ their common instantaneous throughput, defined when $n \geq 1$.

5.1 Conditional distribution

The steady-state instantaneous throughput, denoted by Y , is commonly used as an approximation for the flow-sampled throughput [13,16]. Since it is defined if and only if $n \geq 1$, its distribution is the *conditional* distribution:

$$\forall x \in \mathcal{X}, \quad \Pr(Y = x) = \Pr(\varphi = x | n \geq 1).$$

This throughput metric turns out to be biased: it depends on the proportion of tagged flows in the overall traffic stream. As shown in the example of §6.2 below, it is a loose approximation for the flow-sampled throughput when this proportion is equal to 1 and it coincides with the time-sampled throughput when this proportion tends to 0, which is in fact a simple consequence of the following more general result.

5.2 Weighted distribution

It turns out that the time-sampled throughput has the same distribution as the steady-state instantaneous throughput φ *weighted* by the number n of tagged flows. This weighting reflects the fact that congestion events impact the throughput metric in proportion to the number of active flows.

Proposition 5 *The time-sampled throughput satisfies:*

$$\forall x \in \mathcal{X}, \quad \Pr(X^{\text{time}} = x) = \frac{\mathbb{E}[n \mathbb{I}_{\{\varphi=x\}}]}{\mathbb{E}[n]}.$$

Proof. The result is a simple consequence of Little's law [1]. We have:

$$\mathbb{E}[n] = \lambda \mathbb{E}[D].$$

Moreover, considering those time periods where $\varphi = x$, we get

$$\mathbb{E}[n | \varphi = x] = \frac{\lambda}{\Pr(\varphi = x)} \mathbb{E} \left[\int_0^D \mathbb{I}_{\{\varphi(t)=x\}} dt \right],$$

so that:

$$\mathbb{E}[n \mathbb{I}_{\{\varphi=x\}}] = \lambda \mathbb{E} \left[\int_0^D \mathbb{I}_{\{\varphi(t)=x\}} dt \right]. \quad (13)$$

□

In view of Proposition 5, we have:

$$\gamma^{\text{time}} = \mathbb{E}[X^{\text{time}}] = \frac{\mathbb{E}[n\varphi]}{\mathbb{E}[n]}. \quad (14)$$

Using (13), we get:

$$\begin{aligned} \mathbb{E}[n\varphi] &= \mathbb{E}\left[n \sum_{x \in \mathcal{X}} x \mathbb{I}_{\{\varphi=x\}}\right], \\ &= \sum_{x \in \mathcal{X}} x \lambda \mathbb{E}\left[\int_0^D \mathbb{I}_{\{\varphi(t)=x\}} dt\right], \\ &= \lambda \mathbb{E}\left[\int_0^D \varphi(t) dt\right], \\ &= \lambda \mathbb{E}[S], \end{aligned}$$

which corresponds to the traffic intensity generated by tagged flows (in bit/s). We deduce the following expression for the mean time-sampled throughput:

$$\gamma^{\text{time}} = \frac{\lambda \mathbb{E}[S]}{\mathbb{E}[n]}. \quad (15)$$

Note that this expression in fact directly follows from (11) and Little's law.

6 Examples

We now apply previous results to a few key examples. In the following, we simply refer to flows of infinitely small and infinitely large sizes as *small* and *large* flows, respectively.

6.1 Fair sharing

Consider a unit capacity link. Flows arrive as a Poisson process of intensity λ and have i.i.d. sizes characterized by the random variable S . Sharing is assumed to be perfectly fair so that the instantaneous throughput φ of each flow is equal to $1/n$ in the presence of n flows, with $n \geq 1$. In particular, the throughput takes values in the discrete set $\mathcal{X} = \{1, 1/2, 1/3, \dots\}$.

The corresponding queueing system is the processor-sharing queue. Provided the traffic intensity $\rho = \lambda E[S]$ is less than 1, the queue is stable and has the stationary distribution:

$$\pi(n) = \rho^n(1 - \rho),$$

independently of the flow size distribution beyond the mean [17]. In the presence of a permanent flow, the stationary distribution of the number n of other flows becomes:

$$\pi'(n) = (n + 1)\rho^n(1 - \rho)^2.$$

The result is again insensitive to the flow size distribution beyond the mean. We deduce the following expressions for the throughput of a virtual flow of null size and for the throughput of a permanent flow of infinite size, where $x = 1/m$, $m \geq 1$, denotes any element of \mathcal{X} :

$$\Pr(\varphi^0 = x) = \pi(m - 1) = \rho^{m-1}(1 - \rho),$$

$$\Pr(\varphi^\infty = x) = \pi'(m - 1) = m\rho^{m-1}(1 - \rho)^2.$$

In particular, we have:

$$E[\varphi^0] = -\frac{\ln(1 - \rho)}{\rho}(1 - \rho) \quad \text{and} \quad E[\varphi^\infty] = E[1/\varphi^0]^{-1} = 1 - \rho. \quad (16)$$

Mean throughput

It turns out that the mean flow-sampled throughput has no simple expression, even in the simplest case of an exponential flow size distribution. We only have explicit expressions in the limiting cases of small flows and large flows. In view of Proposition 2 and (16),

$$\lim_{s \rightarrow 0} \gamma^{\text{flow}}(s) = -\frac{\ln(1 - \rho)}{\rho}(1 - \rho) \quad \text{and} \quad \lim_{s \rightarrow \infty} \gamma^{\text{flow}}(s) = 1 - \rho. \quad (17)$$

The mean time-sampled throughput, on the other hand, follows from (15):

$$\gamma^{\text{time}} = \frac{\rho}{E[n]} = 1 - \rho. \quad (18)$$

In view of Proposition 4 and (16), this is also the mean time-sampled throughput for small flows and large flows. The expression is in fact valid for all flow sizes. This follows from (12) and the fact that the mean sojourn time of a customer in a processor-sharing queue is proportional to its service requirement [14].

Figure 1 below compares both throughput metrics for an exponential flow size distribution of unit mean with $\rho = 0.8$. The mean flow-sampled throughput

and the mean time-sampled throughput are estimated using 10,000,000 simulation points (only 10,000 of them are shown on the graph). We observe that, whereas the time-sampled throughput is equal to 0.2 independently of the flow size, as expected, the mean flow-sampled throughput decreases from around 0.4 to 0.2, as predicted by (17), when the flow size grows from 0 to ∞ .

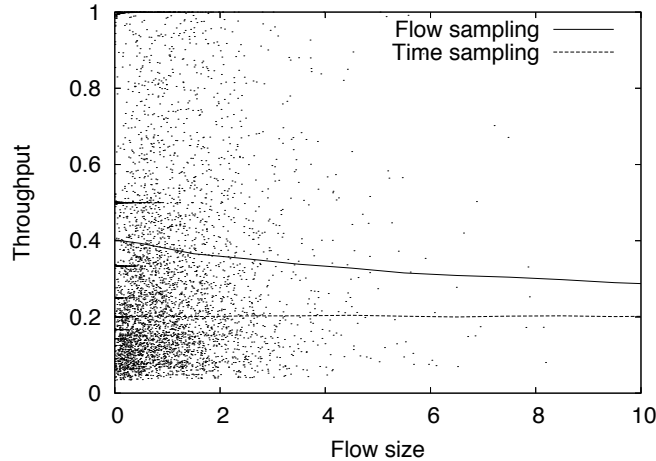


Fig. 1. Mean flow- and time-sampled throughputs (load $\rho = 0.8$).

Throughput distribution

There is again no explicit expression for the flow-sampled throughput distribution except in the limiting cases of small flows and large flows, cf. Proposition 2. For the time-sampled throughput distribution, it follows from Proposition 5 that for all $x = 1/m$, $m \geq 1$:

$$\Pr(X^{\text{time}} = x) = \frac{m\pi(m)}{E[n]} = m\rho^{m-1}(1-\rho)^2.$$

This is also the time-sampled throughput distribution for small flows and large flows, cf. Proposition 4. Again, this expression is in fact valid for all flow sizes, as explained in §6.2 below.

The throughput distribution of small flows is shown in Figure 2 for the same scenario as that of Figure 1. Note that the difference between both throughput metrics is significant. In particular, the probability that the throughput is maximum is equal to $1 - \rho$ (here 0.2) for the flow-sampled throughput and only $(1 - \rho)^2$ (here 0.04) for the time-sampled throughput.

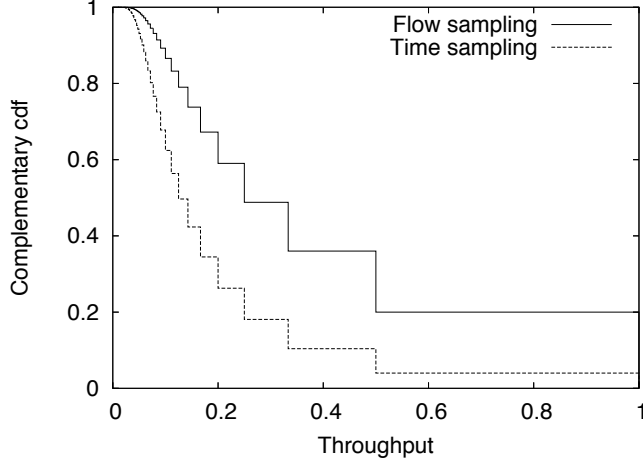


Fig. 2. Flow- and time-sampled throughput distributions of small flows (load $\rho = 0.8$).

6.2 Sampling bias

We now illustrate the sampling bias associated with the steady-state instantaneous throughput Y defined in §5.1. We consider the model of §6.1 with a proportion p of flows tagged at random. We refer to those flows as class-1 flows and to the other flows as class-2 flows. Denoting by $\rho_1 = p\rho$ and $\rho_2 = (1-p)\rho$ the corresponding traffic intensities, the stationary distribution of the number of flows of each class is given by:

$$\pi(n_1, n_2) = \binom{n_1 + n_2}{n_1} \rho_1^{n_1} \rho_2^{n_2} (1 - \rho).$$

We deduce the mean steady-state instantaneous throughput as evaluated through class 1:

$$\gamma_1 = \mathbb{E}[\varphi | n_1 \geq 1] = \frac{\sum_{n_1 \geq 1, n_2} \frac{\pi(n_1, n_2)}{n_1 + n_2}}{\sum_{n_1 \geq 1, n_2} \pi(n_1, n_2)} = -\frac{\ln\left(1 - \frac{\rho_1}{1 - \rho_2}\right)}{\frac{\rho_1}{1 - \rho_2}} (1 - \rho).$$

In view of (17) and (18), this coincides with the mean flow-sampled throughput of small flows for $p = 1$ and with the mean time-sampled throughput for $p \rightarrow 0$.

The bias vanishes when the probability measure is weighted by the number of flows. In view of (14), the mean time-sampled class-1 throughput is given by:

$$\gamma_1^{\text{time}} = \frac{\mathbb{E}[n_1 \varphi]}{\mathbb{E}[n_1]} = \frac{\sum_{n_1 \geq 1, n_2} \frac{n_1 \pi(n_1, n_2)}{n_1 + n_2}}{\sum_{n_1 \geq 1, n_2} n_1 \pi(n_1, n_2)} = 1 - \rho.$$

Similarly, for all $x = 1/m$, $m \geq 1$:

$$\Pr(X_1^{\text{time}} = x) = \frac{\mathbb{E}[n_1 \mathbb{I}_{\{\varphi=x\}}]}{\mathbb{E}[n_1]} = \frac{\sum_{n_1+n_2=m} n_1 \pi(n_1, n_2)}{\sum_{n_1 \geq 1, n_2} n_1 \pi(n_1, n_2)} = m \rho^{m-1} (1 - \rho)^2.$$

This expression is independent of the choice of class 1. Tagging all flows whose size is in the interval $(s - \delta, s + \delta)$, for all $s > 0$ and $\delta > 0$, we conclude by letting δ tend to 0 that the time-sampled throughput of a flow of size s is independent of s .

6.3 Rate limits

In many practical cases, flows do not have full access to the link capacity but are imposed a rate limit. In the simple case of a common rate limit c , the instantaneous throughput φ of each flow is equal to $\min(c, 1/n)$ in the presence of n flows, with $n \geq 1$. The stationary distribution becomes:

$$\pi(n) = \frac{\rho}{nc} \pi(n-1) \quad \text{if } nc \leq 1, \quad \pi(n) = \rho \pi(n-1) \quad \text{otherwise.}$$

In the presence of a permanent flow, we obtain:

$$\pi'(n) = \frac{\rho}{nc} \pi'(n-1) \quad \text{if } (n+1)c \leq 1, \quad \pi'(n) = \frac{n+1}{n} \rho \pi'(n-1) \quad \text{otherwise.}$$

We deduce the distributions of φ^0 and φ^∞ as above. We verify that the time-sampled throughput distribution is again the same for all flow sizes. Figures 3 and 4 below are the analogues of Figures 1 and 2 for a rate limit $c = 0.1$. Though the difference between both throughput metrics is much lower than without rate limit, it is still significant for short flows. For instance, the probability that short flows get the maximum throughput is equal to 0.59 under flow sampling and 0.49 under time sampling. As a comparison, the probability that the steady-state instantaneous throughput Y is maximum is equal to 0.67.

These results extend to more general models with different access rates and several resources shared according to balanced fairness [5]. The underlying queuing system is a Kelly-Whittle network. The same properties then hold: the time-sampled throughput distribution is insensitive to the flow size distribution beyond the mean and independent of the considered flow size.

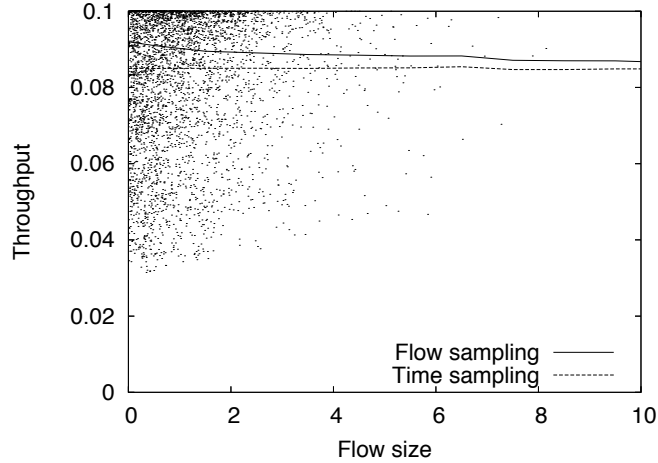


Fig. 3. Mean flow- and time-sampled throughputs (load $\rho = 0.8$, rate limit $c = 0.1$).

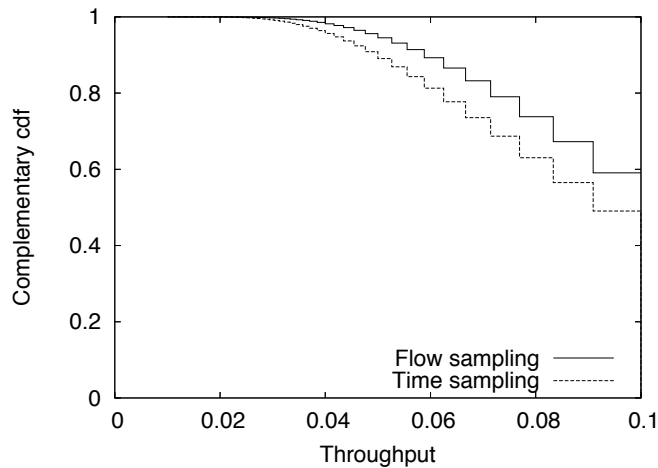


Fig. 4. Flow- and time-sampled throughput distributions of small flows (load $\rho = 0.8$, rate limit $c = 0.1$).

6.4 Unfair sharing

Finally, consider a unit capacity link under unfair sharing. Specifically, we consider two flow classes such that the throughput of a class-1 flow is equal to w_1/w_2 times that of a concurrent class-2 flow, for some weights w_1, w_2 . The corresponding model is the discriminatory processor-sharing queue [10].

There is no simple expression for the steady state distribution, even for an exponential flow size distribution. Figures 5 and 6 below are the analogues of Figures 1 and 2 for a weight ratio $w_1/w_2 = 10$. Note that both the flow-sampled throughput and the time-sampled throughput now depend on the considered flow size. The difference between both throughput metrics is significant.

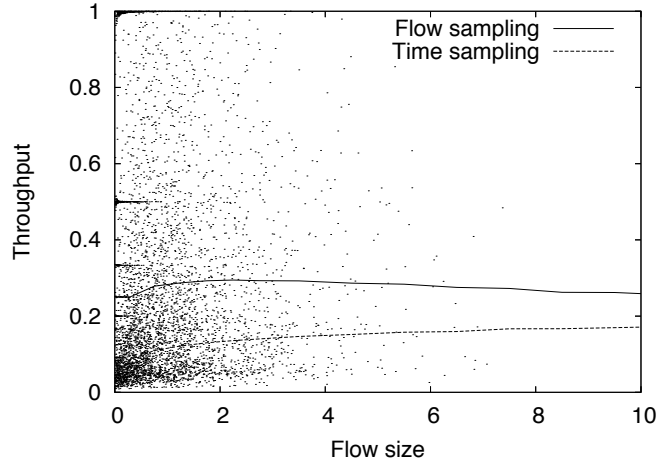


Fig. 5. Mean flow- and time-sampled throughputs of low priority flows (load $\rho = 0.8$).

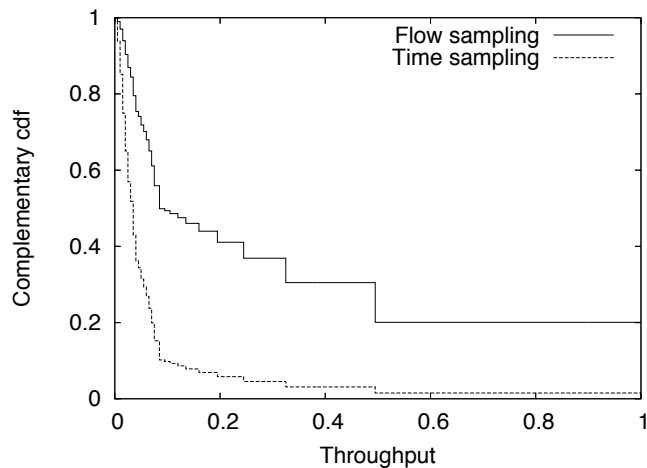


Fig. 6. Flow- and time-sampled throughput distributions of small, low priority flows (load $\rho = 0.8$).

7 Conclusion

We have introduced two metrics for throughput performance evaluation, we refer to as *flow-* and *time-sampled* throughputs. The former gives the throughput statistics of an arbitrary flow while the latter weights these throughput statistics by the flow durations. We have seen that, under fair sharing assumptions, the time-sampled throughput may be simply interpreted as the steady-state instantaneous throughput *weighted* by the number of flows. The weighting reflects the fact that congestion events impact the throughput metric in proportion to the number of active flows. In the absence of weighting, the steady-state instantaneous throughput metric is biased and provides only a loose approximation of the flow-sampled throughput.

References

- [1] F. Baccelli, P. Brémaud, *Elements of Queueing Theory*, Applications of Mathematics Series, Springer Verlag, 2002, second edition.
- [2] S. Ben Fredj, T. Bonald, A. Proutiere, G. Régnié and J.W. Roberts, Statistical bandwidth sharing: a study of congestion at flow level, Proc. of SIGCOMM 2001.
- [3] R. Bekker, S.C. Borst, and R. Nunez Queija, Performance of TCP-friendly streaming sessions in the presence of heavy-tailed elastic flows, Performance Evaluation 61 (2005) 143–162.
- [4] A.W. Berger, Y. Kogan, Dimensioning bandwidth for elastic traffic in high-speed data networks, IEEE/ACM Trans. on Networking 8(5) (2000) 643–654.
- [5] T. Bonald, L. Massoulié, A. Proutière, J. Virtamo, A queueing analysis of max-min fairness, proportional fairness and balanced fairness, Queueing Systems (2006).
- [6] T. Bonald, A. Proutière, On performance bounds for the integration of elastic and adaptive streaming traffic, Proc. of SIGMETRICS / Performance 2004.
- [7] S. Borst, D. van Ooteghem, B. Zwart, Tail asymptotics for discriminatory processor-sharing queues with heavy-tailed service requirements, Performance Evaluation 61, 2-3 (2005) 281–298.
- [8] S.K. Cheung, J.L. van den Berg, R.J. Boucherie, Insensitive bounds for the moments of the sojourn time distribution in the M/G/1 processor-sharing queue, Queueing systems 53, 1-2 (2006) 7–18.
- [9] F. Delcoigne, A. Proutière, G. Regnié, Modeling integration of streaming and data traffic, Performance Evaluation 55, 3-4 (2004) 185–209.
- [10] G. Fayolle, I. Mitrani, R. Iasnogorodski, Sharing a processor among many classes, Journal of the ACM 27 (1980) 519–532.
- [11] F. Guillemin, P. Robert, B. Zwart, Tail asymptotics for processor-sharing queues, Adv. in Appl. Probab. 36, 2 (2004) 525–543.
- [12] D.P. Heyman, T.V. Lakshman, A.L. Neidhardt, A new method for analysing feedback-based protocols with applications to engineering Web traffic over the Internet, Proc. of SIGMETRICS 1997.
- [13] A.A. Kherani, A. Kumar, Stochastic models for throughput analysis of randomly arriving elastic flows in the Internet, Proc. of INFOCOM 2002.
- [14] L. Kleinrock, *Queueing Systems, Vol. 2: Computer Applications*, Wiley, 1976.
- [15] R. Litjens, R. Boucherie, Elastic calls in an integrated services network: the greater the call size variability the better the QoS, Performance Evaluation 52, 4 (2003) 193 – 220.

- [16] R. Litjens, R. Boucherie, J.L. van den Berg, Throughputs in processor sharing models for integrated stream and elastic traffic, Memorandum 1708, Department of Applied Mathematics, University of Twente, 2004.
- [17] R.F. Serfozo, *Introduction to Stochastic Networks*, Springer Verlag, 1999.
- [18] W.R. Stevens, *TCP/IP Illustrated, Volume 1: The Protocols*, Addison-Wesley, 1994.