A note on the benefits of buffering

Mandjes, M.R.H.

DOI
10.1081/STM-120028390
Publication date
2004

Published in
Stochastic Models

Centrum voor Wiskunde en Informatica

REPORT*RAPPORT*

PNA

Probability, Networks and Algorithms

Probability, Networks and Algorithms

PNA A note on the benefits of buffering

M.R.H. Mandjes

REPORT PNA-E0407 JUNE 2004

CWI is the National Research Institute for Mathematics and Computer Science. It is sponsored by the Netherlands Organization for Scientific Research (NWO).
CWI is a founding member of ERCIM, the European Research Consortium for Informatics and Mathematics.

CWI's research has a theme-oriented structure and is grouped into four clusters. Listed below are the names of the clusters and in parentheses their acronyms.

Probability, Networks and Algorithms (PNA)

Software Engineering (SEN)

Modelling, Analysis and Simulation (MAS)

Information Systems (INS)

P.O. Box 94079, 1090 GB Amsterdam (NL)
Kruislaan 413, 1098 SJ Amsterdam (NL)
Telephone +31 20 592 9333
Telefax +31 20 592 4199

ISSN 1386-3711

A note on the benefits of buffering

ABSTRACT
Gaussian traffic models are capable of representing a broad variety of correlation structures, ranging from short-range dependent (e.g. Ornstein-Uhlenbeck type) to long-range dependent (e.g. fractional Brownian motion, with Hurst parameter H exceeding 1/2). This note focuses on queues fed by a large number (n) of Gaussian sources, emptied at constant service rate nc. In particular, we consider the probability of exceeding buffer level nb, as a function of b. This probability decaying (asymptotically) exponentially in n, the essential information is contained in the exponential decay rate I(b). The main result of this note describes the duality relation between the shape of I(.) and the correlation structure. More specifically, it is shown that the curve I(\cdot) is convex at some buffer size b if and only if there are negative correlations on the time scale at which the overflow takes place.

2000 Mathematics Subject Classification: 60K25
Keywords and Phrases: Queueing, Gaussian sources, correlation structure, large deviations asymptotics

A note on the benefits of buffering

Michel MANDJES

CWI
Advanced Communication Networks
P.O. Box 94079
1090 GB Amsterdam
the Netherlands

Faculty of Mathematical Sciences
University of Twente
P.O. Box 217
7500 AE Enschede
the Netherlands

Abstract Gaussian traffic models are capable of representing a broad variety of correlation structures, ranging from short-range dependent (e.g. Ornstein-Uhlenbeck type) to long-range dependent (e.g. fractional Brownian motion, with Hurst parameter H exceeding $\frac{1}{2}$). This note focuses on queues fed by a large number (n) of Gaussian sources, emptied at constant service rate nc. In particular, we consider the probability of exceeding buffer level nb, as a function of b. This probability decaying (asymptotically) exponentially in n, the essential information is contained in the exponential decay rate $I(b)$. The main result of this note describes the duality relation between the shape of $I(\cdot)$ and the correlation structure. More specifically, it is shown that the curve $I(\cdot)$ is *convex* at some buffer size b if and only if there are *negative* correlations on the time scale at which the overflow takes place.

Keywords Queueing • Gaussian sources • correlation structure • large deviations asymptotics

1 INTRODUCTION

Traffic engineering in communication networks greatly benefits from models that are capable of accurately describing the realized performance. As a broad variety of traffic types are multiplexed in the network, each type having its specific (stochastic) characteristics, this is an extremely challenging task. The nodes of the network are usually modeled as *queues*: traffic arrives, can be stored temporarily, and is served at the link speed. The most fundamental model is the first-in-first-out (FIFO) queue, that is focued on in this note.

Gaussian sources. As indicated above, each type of traffic has its own random properties, often summarized by the *correlation structure.* Traditional traffic models usually allow only a mildly correlated traffic arrival process; think for instance of Markov-modulated Poisson processes or exponential on-off sources. As in these models correlations decay relatively fast, we call them *short-range dependent.* Traffic measurements in the 1990s, however, showed that for various types of traffic *long-range dependent* models are more appropriate. Gaussian models cover both short-range (cf. Ornstein-Uhlenbeck) and long-range dependent models (for instance fractional Brownian motion, see e.g. [6]), which explains their increasing popularity. Kilpi and Norros [5] argue that the use of Gaussian traffic models are justified as long as the aggregation is sufficiently large (both in time and number of flows), due to Central Limit type of arguments.

Many-source large deviations. Assume that n i.i.d. Gaussian sources feed into the queueing system, where the (deterministic) service rates of the queues, as well as the buffer thresholds, are scaled by n, too. We now let n grow large; the resulting framework – introduced in [11] – is often referred to as the *many-sources* scaling.
A vast body of results on this model exists, particularly on the probability $p_n(b,c)$ that the queue (fed by n sources, and emptied at a deterministic rate nc) exceeds level nb. Most notably, under mild conditions on the source behavior, it is possible to prove that $p_n(b,c)$ decays *exponentially* in n; keeping

c fixed, we denote the corresponding decay rate by $I(b)$. An early reference on this large deviations framework is, e.g., Botvich and Duffield [3].

The (marginal) benefits of buffering. The goal of the present note is to study the relationship between the correlation structure of the sources and the shape of the curve $I(\cdot)$. Evidently, $I(\cdot)$ is increasing. It is important to notice that, clearly, the steeper $I(\cdot)$ at some buffer size b, the higher the marginal benefits of an additional unit of buffering (where 'benefits' are in terms of reducing the overflow probability). If $I(\cdot)$ is, for instance, *convex*, this means that adding buffering capacity is getting more and more beneficial. This motivates the examination of the characteristics of the decay rate function $I(\cdot)$.

A key notion in our result is the so-called 'time-scale of overflow' $t(b)$, i.e., the most likely duration of a busy period preceding overflow over buffer level nb. The main contribution of the paper is that we show that the curve $I(\cdot)$ is convex (concave) in b, *if and only if* the sources exhibit negative (positive) correlations on the time-scale $t(b)$. All proofs are elementary, and add insight into the marginal benefits of buffering, i.e., the nature of $I'(\cdot)$.

Organization. This note is organized as follows. Section 2 sketches the model and preliminaries. The main result is derived in Section 3. Section 4 presents a number of insightful analytical examples, whereas numerical examples are found in Section 5.

2 MODEL AND PRELIMINARIES

Consider a Gaussian process $(A(t))_{t\in\mathbb{R}}$ with continuous sample paths and stationary increments, see e.g. [2]. Assume that the process is *centered*, i.e., $\mathbb{E}A(t) = 0$ for all $t \geq 0$. We define the variance function $v(\cdot)$, i.e., $v(t) := \mathbb{V}\mathrm{ar}\, A(t) = \mathbb{E}A^2(t)$. It is elementary to show that $A(t)$ has a quadratic cumulant function: $\log \mathbb{E} \exp(\theta A(t)) = \frac{1}{2}\,\theta^2 v(t)$.

Now consider n i.i.d. copies of $(A(t))_{t\in\mathbb{R}}$, say $(A_1(t))_{t\in\mathbb{R}}$ up to $(A_n(t))_{t\in\mathbb{R}}$. We let these Gaussian processes be inputs of a queueing system: source i generates $A_i(t)-A_i(s)$ in the interval $(s,t]$, with $s < t$. Let the n sources feed into a queue with infinite buffer size and link speed $C \equiv nc$, and consider the probability $p_n(b,c)$ that the buffer content exceeds threshold $B \equiv nb$. Define the exponential decay rate of this overflow probability, for a fixed value of c, as a function of b:

$$
\begin{aligned}
I(b) &:= -\lim_{n\to\infty}\frac{1}{n}\log p_n(b,c) \\
&\equiv -\lim_{n\to\infty}\frac{1}{n}\log \mathbb{P}\left(\exists t>0 : \sum_{i=1}^{n} A_i(-t) - nct \geq nb\right), \qquad (1)
\end{aligned}
$$

assuming that this limit exists. A vast body of literature is available on this large deviations framework, see for instance [3, 11, 12]. For the specific case of Gaussian sources, Addie, Mannersalo, and Norros [1] present (mild) conditions under which

$$
I(b) = \inf_{t\geq 0}\sup_{\theta}\left(\theta(b+ct) - \log \mathbb{E}e^{\theta A(t)}\right); \qquad (2)
$$

most notably, it is required that $v(t)/t^{\alpha} \to 0$ as $t\to\infty$, for some $\alpha < 2$. The optimizing θ, as a function of buffer size b and overflow time t, reads

$$
\theta_t(b) = \frac{b+ct}{v(t)}, \qquad (3)
$$

leading to

$$
I(b) = \inf_{t\geq 0}\frac{(b+ct)^2}{2v(t)}. \qquad (4)
$$

Assumption 2.1 *Two assumptions are imposed on the variance function:* (i) $v(\cdot) \in C_2([0,\infty))$, (ii) $\sqrt{v(\cdot)}$ *is strictly increasing and strictly concave.*

The traffic models of the two examples in Section 5 do not obey (some of) these assumptions. The corresponding numerical results illustrate in detail the impact of this.

Corollary 2.2 — *See Lemma 3.1 in* [4]. *Assumption 2.1 entails that, for any b, minimization* (2) *has a unique minimizer $t(b)$. In fact, $t(b)$ is the unique solution to*

$$
F(b,t) := 2c\,v(t) - (b+ct)v'(t) = 0, \quad \text{or} \quad b = c\left(2\frac{v(t)}{v'(t)} - t\right). \qquad (5)
$$

We use the shorthand notation $\theta(b) \equiv \theta_{t(b)}(b)$. Usually $t(b)$ is interpreted as the most likely duration of a busy period of the queue, preceding overflow over level nb, or, briefly, the *most likely epoch of overflow*, see, e.g., Wischik [12].
This type of asymptotics (1) is, for obvious reasons, usually referred to as *logarithmic* asymptotics. We remark that Dębicki and Mandjes [4] recently improved this to *exact asymptotics*, i.e., they derived a (subexponential) function $f(\cdot)$ such that $f(n)\exp(nI(b))p_n(b,c) \longrightarrow 1$ as $n \longrightarrow \infty$.

3 ANALYSIS

In this section we derive a relation between the shape of the decay rate function $I(\cdot)$, and the correlation structure of the Gaussian sources. Our main result is Theorem 3.3. We first prove two lemmas.

The first lemma says that the most likely epoch of overflow $t(b)$ is an increasing function of the buffer size b.

Lemma 3.1 *$t(\cdot) \in C_1([0,\infty))$, and strictly increasing.*

Proof. Recall the fact that $t(b)$ is the *unique* solution to (5). In conjunction with $v(\cdot) \in C_1([0,\infty))$ and $v'(\cdot) > 0$ (Assumption 2.1), we conclude that $t(\cdot)$ is continuous. From (5), we see that

$$\begin{aligned} t'(b) = -\frac{\partial F/\partial b}{\partial F/\partial t} &= \frac{v'(t(b))}{cv'(t(b)) - (b+ct(b))v''(t(b))} \\ &= \frac{1}{c}\cdot\left(1 - 2\,\frac{v(t(b))v''(t(b))}{v'(t(b))^2}\right)^{-1}, \end{aligned} \tag{6}$$

such that the continuity of $t(\cdot)$, together with $v(\cdot) \in C_2([0,\infty))$, implies that $t'(\cdot)$ is continuous, too.
Assumption 2.1 states that, for all $t \geq 0$,

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\sqrt{v(t)} < 0 \quad \Longleftrightarrow \quad 2\,\frac{v(t)v''(t)}{v'(t)^2} < 1,$$

thus proving the lemma. $\square$

The second lemma states a relation between the derivative of the decay rate and the tilting parameter of the Legendre-Fenchel transform in (2).

Lemma 3.2 *For all $b > 0$, it holds that $I'(b) = \theta(b)$.*

Proof. Recalling that $t(b)$ is the optimizing t, differentiating (4) with respect to b yields

$$I'(b) = \left(\frac{b + c\,t(b)}{v(t(b))}\right) - t'(b)\left(\frac{b + c\,t(b)}{2v^2(t(b))}\right)((b + ct(b))v'(t(b)) - 2c\,v(t(b)))\,.$$

Now note that this equals $\theta(b)$, due to (3) and (5). □

Theorem 3.3 *For all $b > 0$,*

$$I''(b) \geq 0 \quad \Longleftrightarrow \quad v''(t(b)) \leq 0.$$

Proof. Due to Lemma 3.2, $I''(b) = \theta'(b)$. Trivial calculus yields

$$\theta'(b) = \frac{v(t(b))(1 + c\,t'(b)) - 2c\,t'(b)v(t(b))}{v^2(t(b))} = \frac{1 - c\,t'(b)}{v(t(b))},$$

where the last equality is due to (5). As $v(t)$ is non negative for any $t > 0$, conclude that $I''(b) \geq 0$ is equivalent to $c\,t'(b) \leq 1$. So we are left to prove that $c\,t'(b) \leq 1$ is equivalent to $v''(t(b)) \leq 0$.
To show this equivalence, note that relation (6) yields

$$t'(b) = \frac{1}{c}\left(1 - \left(t(b) + \frac{b}{c}\right)\frac{v''(t(b))}{v'(t(b))}\right)^{-1}.$$

Now recall that $t(b) \geq 0$, $t'(b) \geq 0$ (due to Lemma 3.1) and $v'(t(b)) \geq 0$. Conclude that $c\,t'(b) \leq 1$ is equivalent to $v''(t(b)) \leq 0$. □

Remark. If the input process has average traffic rate $\mathbb{E}A(1) = \mu \neq 0$, then we can center this process by looking at $B_i(t) := A_i(t) - \mu t$, with $i = 1, \ldots, n$. Clearly $(A_i(t))_{t\in\mathbb{R}}$ and $(B_i(t))_{t\in\mathbb{R}}$ have the same variance function. Hence we can immediately use all the above results, after replacing c by $c - \mu$.

Obviously, for all b and t, it holds that $I(b) \leq (b + ct)^2/v(t)$. Noticing that both $v(\cdot)$ and $I(\cdot)$ are non negative, this results in the following corollary.

Corollary 3.4 *For all $t > 0$, it holds that*

$$v(t) \leq \inf_{b>0} \frac{(b+ct)^2}{I(b)}. \tag{7}$$

Interestingly, situations may occur in which (7) is an equality, thus constituting a *duality* result. Equality holds for $t > 0$ in the set of most likely epochs, i.e., the t such that $t = t(b)$ for some $b > 0$.

4 ANALYTIC EXAMPLES

In this section we present several different Gaussian input models, that illustrate the theory of the previous section. The first highlights a model in which the type of correlation is determined by the choice of a model parameter. Example 2 relates to negatively correlated input traffic, whereas Examples 3 and 4 focus on positively correlated arrivals. All examples add some specific extra insights.

Example 4.1: fractional Brownian motion

For this process, $v(t) = t^{2H}$, for some $H \in (0,1)$. For $H < \frac{1}{2}$ this function is (uniformly) concave, indicating negative correlations, whereas $H > \frac{1}{2}$ entails that $v(\cdot)$ is convex corresponding to positive correlations — for $H = \frac{1}{2}$, the increments are independent. Assumption 2.1 is fulfilled; notice that $\sqrt{v(t)} = t^H$, which is concave.

If we perform the optimization (4), we get, for $b > 0$,

$$t(b) = \frac{b}{c}\frac{H}{1-H}, \quad I(b) = \left(\left(\frac{b}{1-H}\right)^{1-H}\left(\frac{c}{H}\right)^{H}\right)^2. \tag{8}$$

We see that $I(\cdot)$ is indeed convex (concave) when the Hurst parameter is smaller (larger) than $\frac{1}{2}$, as expected on the basis of Theorem 3.3.

Example 4.2: periodic arrivals

Jobs of size 1 arrive periodically, with a phase that is uniformly distributed on $[0,1]$. Clearly $v(t) = t_e(1-t_e)$, where $t_e := t \bmod 1$. Now consider the

Gaussian counterpart of this model — notice that the input is not centered, so we have to replace c by $c - \mu = c - 1$.
Assumption 2.1 is not satisfied, but it is not hard to verify that we can restrict ourselves to $t \in [0, \frac{1}{2})$, rather than $t > 0$, on which $\sqrt{v(\cdot)}$ is concave. Direct arguments yield that the results of Section 3 go through.
It is straightforward to derive that $t(b) = b(c-1+2b)^{-1}$, and $I(b) = 2b(b+c)$. As expected from Theorem 3.3, the convexity of $v(\cdot)$ translates into $I(\cdot)$ being convex.

Example 4.3: Ornstein-Uhlenbeck input

Following Section 4.3 of [3], we represent a stationary Ornstein-Uhlenbeck process by choosing $v(t) = t - 1 + e^{-t}$, which satisfies Assumption 2.1. It is easy to see that $v(\cdot)$ is convex, so we will have 'decreasing marginal buffering benefits', i.e., $I(\cdot)$ is concave due to Theorem 3.3. This example shows the relation between 'level of positive correlation' and the shape of $I(\cdot)$. The strong convexity for small t indicates strong positive correlation on short time-scale, whereas this positive correlation becomes weaker and weaker as the time-scale increases (reflected by the asymptotically linear shape of $v(\cdot)$ for t large).

First we concentrate on small b. Parametrize $t \equiv \alpha b$. It is a matter of applying Taylor expansions to obtain that (4) reads

$$I(b) = \frac{1}{2}\,c^2 + \sqrt{b}\inf_{\alpha>0}\left(\frac{2c}{\alpha} + \frac{c^2\alpha}{3}\right) + O(b) = \frac{1}{2}\,c^2 + \frac{2}{3}\sqrt{6bc^3} + O(b).$$

for b small. Notice that the infimum over α is attained at $c^{-1}\sqrt{6}$, yielding $t(b) \approx \sqrt{6b/c}$. So $I(\cdot)$ is highly concave (as a square root), expressing the strong positive correlations on a short time-scale, cf. the results in [7, 11].

For large b, the optimum over t in (4) is attained for large t, such that $v(t)$ is approximated by $t - 1$. Doing the calculations, we obtain

$$I(b) - 2c(b+c) \to 0 \quad \text{as } b \to \infty,$$

with $t(b) \approx b/c + 2$. Apparently, for large b, $I(\cdot)$ becomes nearly linear, as expected by the weak correlation on the long time-scale.

Example 4.4: M/G/∞ inputs with Pareto job sizes

A single M/G/∞ source consists of jobs that arrive according to a Poisson process of rate λ. They stay in the system during some holding time, that is distributed as a random variable D (with $\mathbb{E}D < \infty$). During its duration, the job generates fluid at a unit rate. We assume $\mathbb{P}(D > t) = (t+1)^{-\alpha}$, i.e., D has a Pareto distribution. Take $\alpha > 1$; then $\mathbb{E}D = (\alpha - 1)^{-1} < \infty$.
We now consider the Gaussian input process that has the same mean input rate and variance function. The mean input rate is trivially $\mu := \lambda\,\mathbb{E}D$ per unit time, whereas the variance function reads (assume for ease that $\alpha \notin \{2, 3\}$)

$$v(t) = \frac{2\lambda}{(3-\alpha)(2-\alpha)(1-\alpha)}\left(1 - (t+1)^{3-\alpha} + (3-\alpha)t\right),$$

which obeys the requirements stated in Assumption 2.1. As in Example 2, for small b, we put $t \equiv \alpha b$, to get a highly concave shape at $b \downarrow 0$:

$$I(b) = \frac{\alpha - 1}{2}(c-\mu)^2 + \frac{1}{3}\sqrt{6b(\alpha-1)^3(c-\mu)^3} + O(b).$$

The interesting part, however, is $b \to \infty$. If $\alpha > 2$, then $v(t)$ is essentially linear in t, yielding (as in Example 2), $t(b) \approx b(c-\mu)^{-1} - 2(3-\alpha)^{-1}$ and

$$I(b) - 2(\alpha-1)(\alpha-2)(c-\mu)\left(b - \frac{c-\mu}{3-\alpha}\right) \to 0.$$

If $\alpha \in (1, 2)$, then $v(t)$ is roughly of the order $t^{3-\alpha}$. It can be verified easily that $t(b)$ looks like $b\,\mathbb{E}D(c-\mu)^{-1}$, implying that $I(b) = O(b^{\alpha-1})$.
Now an interesting connection can be drawn with results from the literature.

- It has been proven that, for $\alpha \in (1, 2)$, the superposition of many of these M/G/∞ inputs converges (with some appropriate rescaling of time) to fBm with $H = (3-\alpha)/2$, cf. [10]. For these α it can be proven that the input process is long-range dependent (in that the correlation function is not integrable). This persistent correlation is reflected by the concave shape of $I(\cdot)$, even for large b. Also notice the similarity with the formula for $I(b)$ in (8), with $H = (3-\alpha)/2$.

- If on the other hand $\alpha > 2$, then there is convergence to ordinary Brownian motion; the input traffic is short-range dependent, as reflected by the asymptotically linear shape of $I(\cdot)$.

5 NUMERICAL EXAMPLES

In this section we present two numerical examples. The former focuses on sources with positive correlations on a short time-scale, and negative correlations on a longer time-scale, whereas the latter model displays the opposite behavior. In both examples, Assumption 2.1 is only partly met – the impact of this is demonstrated in the numerical results.

Example 5.1: deterministic on-off sources

Consider a source that alternates between being silent and transmitting (at a constant rate of, say, 1). Let the on- and off-periods be deterministic, with lengths σ and τ, respectively. Hence the source is purely periodic; the position within this period, has a uniform distribution on $[0, \sigma+\tau]$. We assume that $\sigma \leq \tau$; a similar reasoning applies to the case $\tau < \sigma$. The variance $v(t)$ of such a source is given by

$$-\frac{t^3}{3(\sigma+\tau)} + \frac{\sigma t^2}{\sigma+\tau} - \frac{\sigma^2 t^2}{(\sigma+\tau)^2} \quad \text{for } t \in [0, \sigma);$$

$$-\frac{\sigma^3}{3(\sigma+\tau)} + \frac{\sigma^2 t}{\sigma+\tau} - \frac{\sigma^2 t^2}{(\sigma+\tau)^2} \quad \text{for } t \in [\sigma, \tau);$$

$$-\frac{\sigma^3}{3(\sigma+\tau)} + \frac{\sigma^2 t}{\sigma+\tau} + \frac{(t-\tau)^3}{3(\sigma+\tau)} - \frac{\sigma^2 t^2}{(\sigma+\tau)^2} \quad \text{for } t \in [\tau, \sigma+\tau).$$

We now focus on the Gaussian counterpart of such a source, i.e., $A(t)$ has a Normal distribution with mean $\sigma\,(\sigma+\tau)^{-1}$ and variance $v(t)$. Notice that the source exhibits positive correlations on a time-scale below σ (as expressed by $v''(t) \geq 0$ for $t \in [0, \sigma)$), whereas on a longer time-scale traffic is negatively correlated (as $v''(t) \leq 0$ for $t \in [\sigma, \tau)$). Similarly to Example 4.2, we can restrict ourselves to $t \in [0, (\sigma+\tau)/2)$. Notice however that $v(\cdot) \in C_1([0, \infty))$ rather than $C_2(([0, \infty))$; hence Assumption 2.1 is *not* met.

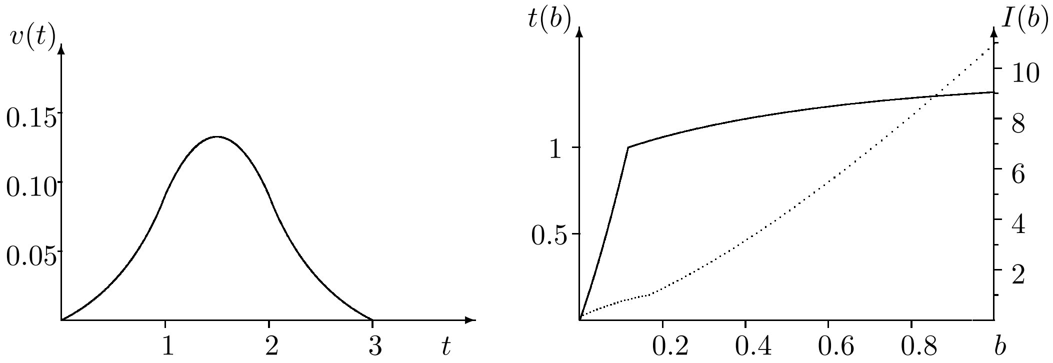

Figure 1: Example 5.1. The left panel displays the variance $v(\cdot)$ as a function of t. The right panel depicts the decay rate $I(\cdot)$ — the dotted line — and the time to overflow $t(\cdot)$ — the solid line — both as functions of the buffer size b.

In the example we let n of these on-off sources feed into a buffered resource. We use the parameters $\sigma = 1$ and $\tau = 2$, and hence the average traffic rate of a single source is $\frac{1}{3}$. We choose the normalized link rate $\frac{1}{2}$. To transform the process into a queue fed by *centered* sources, we take $c = \frac{1}{2} - \frac{1}{3} = \frac{1}{6}$. The function $v(\cdot)$ shifts from convexity to concavity at $t = 1$. We see that $I(\cdot)$ is indeed concave as long as $t(b)$ is below 1, whereas it become convex if $t(b)$ exceeds 1, as implied by Theorem 3.3; the transition takes place at $b = 0.168$. The fact that $v(\cdot) \in C_1([0,\infty))$ – instead of $C_2([0,\infty))$ – is reflected by the fact that $t(\cdot)$ is continuous, but not differentiable.

Example 5.2: a combined packet/burst level model

In this example we consider an on-off source with an explicitly modeled packet and burst level. Within the on-times packets (of fixed size, say 1) are sent periodically (cf. Example 4.2). We normalize time such that the corresponding packet interarrival time is 1. Both the on- and off-times are distributed geometrically, with means p^{-1} and q^{-1}, respectively. We emphasize that — although the on- and off-times are integer-valued — the model is *not* a discrete-time model, as the phase of each source is distributed uniformly on $(0, 1)$.

As before, let $A(t)$ denote the amount of traffic generated by a single source in an interval of length t. In [8] the moment generating function $\mathbb{E}\exp(\theta A(t))$ is given explicitly, from which the moments of $A(t)$ can be derived through differentiation. This enables the computation of $v(t)$. As before, consider the Gaussian counterpart of such a source.

In the example, we choose $p = 0.01$ and $q = 0.02$. This gives $\mu = \frac{2}{3}$. Let the normalized link speed be $\frac{4}{5}$, such that we have to use $c = \frac{2}{15}$. The variance function shows the negative (positive) correlations on a short (longer) time-scale. Notice that Assumption 2.1 is *not* met, see the left panel of Figure 2 — in the right panel we see how this affects the relation between $v(\cdot)$ and $I(\cdot)$.

Figure 2 shows an interesting phenomenon. For small buffers, overflow is mainly caused by 'colliding packets': overflow happens within the packet interarrival time (i.e., 1). For larger buffers it becomes more likely that the number of sources that is in the on-time is higher than on average — then the time to overflow is a multiple of packet interarrival times. A more detailed description of these packet and burst effects is given in [8], cf. also Example 6 in [9].

The curve $I(\cdot)$ is convex for small b, due to the negative correlations on a short time-scale, and has some 'angle' (in the example at a critical buffer level of $b = 0.130$) as soon as the positive correlations kick in. The fact that $v(\cdot) \in C_0([0,\infty))$ – instead of $C_2([0,\infty))$ – is reflected by the fact that $t(\cdot)$ is discontinuous.

Acknowledgment

The author is indebted to Krzysztof Dębicki (CWI) and Laurent Massoulié (Microsoft Research Cambridge, UK) for valuable discussions.

References

[1] R. Addie, P. Mannersalo, and I. Norros. Most probable paths and performance formulae for buffers with Gaussian input traffic. Accepted for publication in *European Transactions on Telecommunications.*

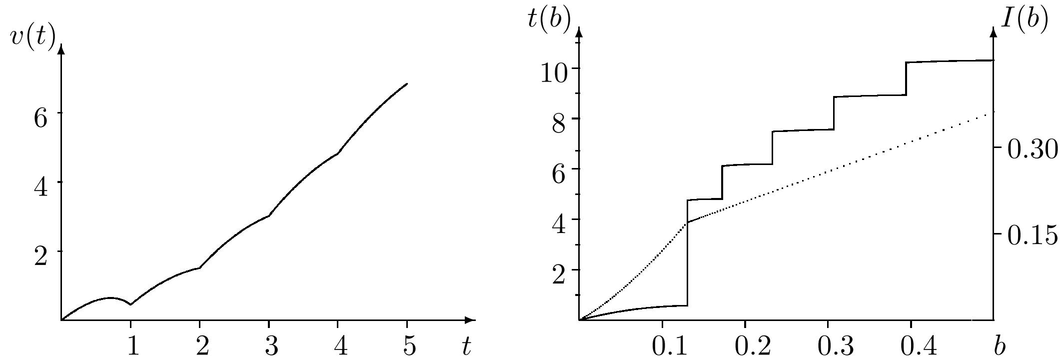

Figure 2: Example 5.2. The left panel displays the variance $v(\cdot)$ as a function of t. The right panel depicts the decay rate $I(\cdot)$ — the dotted line — and the time to overflow $t(\cdot)$ — the solid line — both as functions of the buffer size b.

[2] R. Adler. An introduction to continuity, extrema, and related topics for general Gaussian processes. *IMS Lecture Notes-Monograph Series*, 12, 1990.

[3] D. Botvich and N. Duffield. Large deviations, the shape of the loss curve, and economies of scale in large multiplexers. *Queueing Systems*, 20: 293 – 320, 1995.

[4] K. Dębicki and M. Mandjes. Exact overflow asymptotics for queues with many Gaussian inputs. CWI report PNA-R0209, 2002.

[5] J. Kilpi and I. Norros. Testing the Gaussian character of access network traffic. Proceedings Internet Measurement Workshop, Marseille, France, 2002.

[6] W. Leland, M. Taqqu, W. Willinger, and D. Wilson. On the self-similar nature of Ethernet traffic. *IEEE/ACM Transactions on Networking*, 2: 1 – 15, 1994.

[7] M. Mandjes and J.H. Kim. Large deviations for small buffers: an insensitivity result. *Queueing Systems*, 37: 349 – 362, 2001.

[8] M. Mandjes and J.H. Kim. An analysis of the phase transition phenomenon in packet networks. *Advances in Applied Probability*, 33: 360 – 380, 2001.

[9] P. Mannersalo and I. Norros. A most probable path approach to queueing systems with general Gaussian input. To appear in *Computer Networks*, 2002.

[10] M. Taqqu, W. Willinger, and R. Sherman. Proof of a fundamental result in self-similar treffic modeling. *Computer Communications Review*, 27: 5 – 23, 1997.

[11] A. Weiss. A new technique for analyzing large traffic systems. *Advances in Applied Probability*, 18: 506 – 532, 1995.

[12] D. Wischik. Sample path large deviations for queues with many inputs. *Annals of Applied Probability*, 11: 379 – 404, 2001.