Centrum voor Wiskunde en Informatica

REPORT*RAPPORT*

PNA

Probability, Networks and Algorithms

Probability, Networks and Algorithms

Sample-path large deviations for generalized processor sharing queues with Gaussian inputs

M.R.H. Mandjes, M.J.G. van Uitert

REPORT PNA-R0308 JUNE 30, 2003

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

Sample-Path Large Deviations for Generalized Processor Sharing Queues with Gaussian Inputs

Michel Mandjes[*,†] and Miranda van Uitert[*]
email: michel|miranda@cwi.nl

[*] *CWI*
P.O. Box 94079, 1090 GB Amsterdam, The Netherlands
[†] *University of Twente*
P.O. Box 217, 7500 AE Enschede, The Netherlands

ABSTRACT
In this paper we consider the Generalized Processor Sharing (GPS) mechanism serving two traffic classes. These classes consist of a large number of independent identically distributed Gaussian flows with stationary increments. We are interested in the logarithmic asymptotics or exponential decay rates of the overflow probabilities. We first derive both an upper and a lower bound on the overflow probability. Scaling both the buffer sizes of the queues and the service rate with the number of sources, we apply Schilder's sample-path large deviations theorem to calculate the logarithmic asymptotics of the upper and lower bound. We discuss in detail the conditions under which the upper and lower bound match. Finally we show that our results can be used to choose the values of the GPS weights. The results are illustrated by numerical examples.

2000 Mathematics Subject Classification: 60G15, 60K25 (primary), 60G70, 68M20, 90B18.
*Keywords and Phrases:*sample-path large deviations, Gaussian traffic, Schilder's theorem, Generalized Processor Sharing, communication networks, differentiated services, weight setting.
Note: Work carried out under the CWI project P1201 (PNA 2). Financially supported by the Dutch Ministry of Economic Affairs via the project EQUANET (P1230).

1 Introduction

A major trend in communication networking is constituted by the integration of a growing range of traffic types over a common network infrastructure. These traffic types are highly heterogeneous, with respect to both (i) their diverse Quality-of-Service (QoS) requirements in terms of packet delay, loss, and throughput metrics, and (ii) their specific (stochastic) properties. We now comment on both types of heterogeneity.

(i) *Heterogeneous QoS, packet scheduling.* FIFO queues lack the capability of offering multiple QoS levels. Hence, if a FIFO queue is used to support traffic classes with heterogeneous QoS requirements, all classes should be offered the most stringent of these requirements. This approach clearly leads to inefficient use of network resources: some of the classes get a better QoS than requested.

The need for efficient QoS-differentiating mechanisms motivates the development of discriminatory scheduling disciplines that actively distinguish between streams of the various traffic types. Packet versions of the ideal fluid discipline *Generalized Processor Sharing* (GPS), see e.g. [22, 23], are considered to be suitable candidates. In GPS, each class is guaranteed a certain minimum service rate; if one of the classes does not fully use this guaranteed rate, the residual capacity is redistributed among the other classes (in proportion to their guaranteed rates). Note that this makes GPS a work-conserving discipline. GPS is considered as a promising compromise between isolation and sharing: each traffic class is protected against 'misbehavior' of other classes, whereas at the same time significant multiplexing gains between classes can be achieved.

In this paper we focus on two classes sharing the total service capacity C according to GPS. We assign guaranteed rate $\phi_i C$ to class i, which can be claimed by class i at any time – the ϕ_i are referred to as *weights*, $i = 1, 2$. Both classes are assigned a queue, that fills when the input rate temporarily exceeds the capacity available. When both classes are backlogged, i.e., have non-empty queues, both are served at their guaranteed rate. If one of the classes does not fully use its guaranteed rate, then the unused capacity is made available to the other class. It is clear that, in order to fully benefit from GPS, the weights should be chosen appropriately. This is not a straightforward task, that usually relies on expressions (or approximations) for the buffer content distributions of the queues. Weight setting procedures available from the literature are often restricted to special classes of input traffic, see, e.g., [11, 13] for the case of leaky-bucket regulated traffic.

(ii) *Heterogeneous traffic, Gaussian models.* To model the heterogeneity of the input traffic, Gaussian source models have proven to be particularly useful. Traditional traffic models, like for instance Markov-modulated Poisson processes or exponential on-off sources, allow

only a mildly correlated traffic arrival process. As time correlations decay relatively fast in these models, they are referred to as *short-range dependent*. Traffic measurements in the 1990s, however, convincingly showed that for various types of traffic such correlations typically decay relatively slowly, motivating the use of *long-range dependent* models [14]. Gaussian models cover both short-range (for instance Ornstein-Uhlenbeck-type inputs) and long-range dependent traffic (for instance fractional Brownian motion, abbreviated to fBm), and are therefore considered to be extremely useful.

A complicating issue in the choice of the appropriate traffic model is the fact that network traffic is usually influenced by feedback loops (think of TCP), which control how the user's traffic supply is transmitted into the network. Kilpi and Norros [12] argue that (non-feedback) Gaussian traffic models are still justified as long as the aggregation is sufficiently large (both in time and number of flows), due to Central Limit type of arguments.

The Gaussian model is also justified by several theoretical results. Among these we mention Taqqu *et al.* [25], who consider the superposition of many heavy-tailed on-off sources, and prove convergence of the resulting aggregate traffic process to fBm (after rescaling time appropriately). It was recently shown in [8] that this convergence carries over to the queueing process, justifying the choice of fBm as a good approximation of traffic inputs in queueing models.

In this paper we focus on GPS queues with Gaussian inputs. Our framework concentrates on traffic heterogeneity *across* the GPS classes, rather than *within* classes. We assume both classes to consist of superpositions of i.i.d. sources.

Large deviations. Above we motivated the interest in GPS queues with Gaussian inputs. Our study focuses on a *large-deviations* analysis of this model.

Over the past two decades, significant research efforts have been made on the large-deviations analysis of queueing models. These efforts have culminated in a wealth of contributions to the understanding of the occurrence of rare events in queues. In particular, the celebrated *many-sources* scaling, introduced in the seminal paper of Weiss [27], has provided a rich framework for obtaining large-deviations results. In a many-sources setting, one considers a queue fed by the superposition of n i.i.d. traffic sources, with queueing resources (service rates, buffer thresholds) scaled with n as well. This framework is motivated by the fact that the number of sources multiplexed in a network resource (particularly in the core) is typically large. Under mild conditions on the source behavior, explicit expressions are available for the *exponential* decay of the probability that the buffer content in a single FIFO queue exceeds a certain level. Early references in this large-deviations framework are the logarithmic asymptotics found in, e.g., Botvich and Duffield [5] and Courcoubetis and Weber [6].

In contrast, only few large-deviations results are known for queues operating under a non-

FIFO scheduling discipline. In [18], Mannersalo and Norros initiated the study of the priority mechanism, whereas in [19] they examine the GPS discipline. In both papers useful intuition and heuristics were developed. Their results on the priority mechanism were further enhanced in [17]; notably a lower bound on the decay rate of overflow in the low-priority queue was found, as well as conditions under which this lower bound coincides with the exact value. The main goal of the present paper is to obtain similar rigorous many-sources large-deviation results for the two-class GPS system. It is noted that GPS in the large-buffer regime (rather than the many-sources regime) is better understood; significant contributions are, e.g., [4, 26, 28].

Contribution. The results of the paper can be summarized as follows. In the first place, we derive upper and lower bounds for the overflow probabilities in the two-queue GPS system. These are generic in that they do not only apply to Gaussian inputs, but in fact to any input traffic model. Then we evaluate these bounds in the many-sources framework, i.e., we derive their exponential decay rates (in the number of sources n), after rescaling the link speed $C \equiv nc$ as well as the buffer threshold $B \equiv nb$. We do this by using large-deviations machinery, in particular the multi-dimensional version of the classical Cramér result for sample means, and the pathwise large-deviations principle of Schilder. We then prove tightness of the derived bounds under certain conditions, and present an intuitive motivation why tightness can be expected more generally. Finally we address the problem of finding appropriate weights. In particular, we focus on the operational issue of finding weights such that the QoS-requirement is met for all combinations of sources within some predefined region.

The paper is organized as follows. Section 2 deals with preliminaries on GPS, Gaussian sources, and large deviations. Section 3 presents the generic upper and lower bounds on the overflow probability of (without loss of generality) queue 1. We first focus on the regime in which the mean rate of the type-2 sources, $n\mu_2$, is below their guaranteed rate $n\phi_2 c$; lower and upper bounds on the decay rate are derived in Sections 4 and 5. Section 6 deals with the (easier) case $n\mu_2 \geq n\phi_2 c$. A discussion on the results is given in Section 7; it turns out that three generic regimes can be distinguished. Section 8 addresses weight setting procedures. Section 9 concludes.

2 Model and preliminaries

In Section 2.1 we introduce the two-class GPS model with the necessary notation. Then we discuss in Section 2.2 Gaussian sources. The large-deviations theorems of Cramér and Schilder will be presented in Section 2.3.

2.1 Generalized Processor Sharing

We consider a system where traffic is served according to the GPS mechanism, consisting of two queues sharing a link of capacity nc. We assume the system to be fed by traffic from two classes, where class i uses queue i, for $i = 1, 2$. Without loss of generality it is assumed that both classes consist of n flows (see Remark 2.2). We assign a weight $\phi_i \geq 0$ to class i and, again without loss of generality, assume that these add up to 1, i.e., $\phi_1 + \phi_2 = 1$. The GPS mechanism then works as follows. Class i receives service at rate $n\phi_i c$ when both classes are backlogged. Because class i gets at least service at rate $n\phi_i c$ when it has backlog, we will refer to it as the *guaranteed rate* of class i. If one of the classes has no backlog and is transmitting at a rate less than or equal to its guaranteed rate, then this class is served at its transmission rate, while the other class receives the remaining service capacity. If both classes are sending at rates less than their guaranteed rates, then they are both served at their sending rate, and some service capacity is left unused. We assume that the buffer sizes of both queues are infinitely large.

Without loss of generality, we focus on the workload of the first queue. The goal of this paper is to derive the logarithmic asymptotics for the probability that the stationary workload exceeds a threshold nb. Denoting by $Q_{i,n} \equiv Q_{i,n}(0)$ the stationary workload in the i-th GPS queue at time 0, the probability of our interest reads

$$\mathbb{P}(Q_{1,n} \geq nb). \tag{1}$$

We denote by $A_{j,i}(s,t)$ the amount of traffic generated by the j-th flow of class i in the interval $(s,t]$, $j = 1, \ldots, n$, $i = 1, 2$. Defining $B_{i,n}(s,t)$ as the total service that was available for class i in the interval $(s,t]$, we have the following identity:

$$Q_{i,n}(t) = Q_{i,n}(s) + \sum_{j=1}^{n} A_{j,i}(s,t) - B_{i,n}(s,t), \ \ \forall s < t, \text{ with } s, t \in \mathbb{R}. \tag{2}$$

The stationary queue can be represented by:

$$Q_{i,n}(0) = \sup_{t>0} \left\{ \sum_{j=1}^{n} A_{j,i}(-t,0) - B_{i,n}(-t,0) \right\}, \tag{3}$$

where the negative optimizing t corresponds to the beginning of the busy period that includes time 0, as argued in [24]. In Section 3 we rewrite our problem in terms of the *empirical mean processes* $n^{-1}\sum_{j=1}^{n} A_{j,i}(\cdot,\cdot)$, $i = 1, 2$. We define the realization of $n^{-1}\sum_{j=1}^{n} A_{j,i}(0,r)$ by $f_i(r)$, i.e., we speak of $f_i(\cdot)$ as the *path* of the empirical mean process of class i. By $A_i[f_i](s,t)$ we then denote the value of $n^{-1}\sum_{j=1}^{n} A_{j,i}(s,t)$ for the (given) path $f_i(\cdot)$, i.e., $A_i[f_i](s,t) := f_i(t) - f_i(s)$. For notational convenience we use $f(\cdot)$ to denote the two-dimensional path $(f_1(\cdot), f_2(\cdot))$.

2.2 Gaussian processes

We assume the n flows of class i to be i.i.d. Gaussian processes with stationary increments. Let $A_{j,i}(s,t)$ be distributed as $A_i(s,t)$, where $A_i(s,t)$ can be considered as the 'generic' random variable corresponding to the amount of traffic of a single class-i flow arriving in the interval $(s,t]$, $i = 1,2$. We denote the corresponding mean traffic rate and variance function by μ_i and $v_i(\cdot)$ respectively: for all $s < t$, $\mathbb{E}A_i(s,t) = \mu_i(t-s)$ and $\mathbb{V}\mathrm{ar}A_i(s,t) = v_i(t-s)$. We also define the aggregate mean rate $\mu := \mu_1 + \mu_2$ and the aggregate variance function $v(\cdot) := v_1(\cdot) + v_2(\cdot)$. To guarantee stability, we assume that $\mu < c$. In order to apply Schilder's sample-path large-deviations principle (LDP) (Theorem 2.5), we also need to introduce the *centered* process $\bar{A}_i(t) := A_i(0,t) - \mu_i t$. The covariance function $\Gamma_i(s,t)$ is for all $s < t$ defined by

$$\Gamma_i(s,t) := \mathbb{C}\mathrm{ov}[A_i(0,s), A_i(0,t)] = \mathbb{C}\mathrm{ov}[\bar{A}_i(s), \bar{A}_i(t)] = \frac{1}{2}(v_i(s) + v_i(t) - v_i(t-s)).$$

Finally we make the following assumptions on the variance function.

Assumption 2.1 *We assume that, for $i = 1,2$,* (A1) *$v_i(\cdot)$ is continuous, differentiable on $(0,\infty)$;* (A2) *$\sqrt{v_i(\cdot)}$ is strictly increasing and strictly concave;* (A3) *for some $\alpha < 2$ it holds that $v_i(t)t^{-\alpha} \to 0$ as $t \to \infty$.*

Assumptions (A1) and (A3) are required to apply 'Schilder', see [1]. Assumption (A2) is needed in the proof of Lemma 5.9.

Remark 2.2 Above we assumed that both classes consist of n sources, but the analysis can be easily extended to the case of an unequal number of sources. The scenario with class i having $n\beta_i$ flows, mean μ_i and variance $v_i(\cdot)$ is equivalent to a scenario where class i has n flows, mean $\beta_i\mu_i$ and variance $\beta_i v_i(\cdot)$, due to the infinitely divisibility of the Gaussian distribution.

2.3 Sample-path large deviations

The analysis in the present paper relies on a sample-path LDP for (centered) Gaussian processes. This subsection is devoted to a brief description of the main theorem in this field, (the generalized version of) *Schilder's theorem* [3]. However, we start by recalling the multivariate version of the well-known *Cramér's theorem*, see [9, Thm. 2.2.30].

Theorem 2.3 [Multivariate Cramér] *Let $X_j \in \mathbb{R}^d$ be i.i.d. d-dimensional random vectors, $j = 1,\ldots,n$, distributed as a random vector X. Then $n^{-1}\sum_{j=1}^{n} X_j$ satisfies the following LDP:*

(a) For any closed set $F \subset \mathbb{R}^d$,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\frac{1}{n}\sum_{j=1}^{n} X_j \in F\right) \leq -\inf_{x\in F} \Lambda(x);$$

(b) For any open set $G \subset \mathbb{R}^d$,

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\frac{1}{n}\sum_{j=1}^{n} X_j \in G\right) \geq -\inf_{x\in G} \Lambda(x),$$

where the large-deviations rate function $\Lambda(\cdot)$ is given by

$$\Lambda(x) := \sup_{\theta\in\mathbb{R}^d} \left(\langle \theta, x\rangle - \log \mathbb{E}e^{\langle \theta, X\rangle}\right), \tag{4}$$

with the notation $\langle\ ,\ \rangle$ denoting the usual inner product: $\langle a, b\rangle := a^{\mathrm{T}}b = \sum_{i=1}^{d} a_i b_i$.

Remark 2.4 Consider the specific case that X has a multivariate Normal distribution with mean vector μ and $(d \times d)$ non-singular covariance matrix Σ. Using $\log \mathbb{E}e^{\langle \theta, X\rangle} = \langle \theta, \mu\rangle + \frac{1}{2}\theta^{\mathrm{T}}\Sigma\theta$, it is not hard to derive that, with $(x-\mu)^{\mathrm{T}} \equiv (x_1 - \mu_1, \ldots, x_d - \mu_d)$,

$$\theta^\star = \Sigma^{-1}(x-\mu) \ \text{ and } \ \Lambda(x) = \frac{1}{2}(x-\mu)^{\mathrm{T}}\Sigma^{-1}(x-\mu), \tag{5}$$

where $\theta^\star$ optimizes (4); it is well-known that $\Lambda(\cdot)$ is convex.

We now sketch the framework of Schilder's sample-path LDP, as established in [3], see also [10]. We restrict ourselves to the aspects that are relevant in the present study; for more details we refer to [1, 18, 21]. Consider, n i.i.d. centered Gaussian processes $\bar{A}_{j,i}(\cdot)$, for $i = 1, 2$, with stationary increments and covariance $\mathbb{C}\mathrm{ov}[\bar{A}_{j,i}(s), \bar{A}_{j,i}(t)] = \Gamma_i(s,t)$. Define, for $i = 1, 2$, the path space Ω_i as

$$\Omega_i := \left\{\omega_i : \mathbb{R} \to \mathbb{R},\ \text{continuous},\ \omega_i(0) = 0,\ \lim_{t\to\infty} \frac{\omega_i(t)}{1+t} = \lim_{t\to-\infty} \frac{\omega_i(t)}{1+t} = 0\right\},$$

which is a separable Banach space by imposing a specific norm, as explained in [18]. We adhere to the approach in [18] by choosing $\Omega = \Omega_1 \times \Omega_2$ as our path space, where $\left\{\bar{A}_{j,1}(\cdot)\right\}_{j=1}^{n}$ and $\left\{\bar{A}_{j,2}(\cdot)\right\}_{j=1}^{n}$ are independent.

Next we introduce and define the *reproducing kernel Hilbert space* $R_i \subseteq \Omega_i$ – see [2] for a more detailed account – with the property that its elements are roughly as smooth as the covariance functions $\Gamma_i(s,\cdot)$. We start from a 'smaller' space S_i, defined by

$$S_i := \left\{\omega_i : \mathbb{R} \to \mathbb{R},\ \omega_i(\cdot) = \sum_{j=1}^{n} a_{j,i}\Gamma_i(s_j, \cdot),\ \ a_{j,i}, s_j \in \mathbb{R}, j = 1, \ldots, n; n \in \mathbb{N}\right\}.$$

The inner product on this space S_i is, for $\omega_{a,i}, \omega_{b,i} \in S_i$, defined as

$$\langle \omega_{a,i}, \omega_{b,i} \rangle_{R_i} := \left\langle \sum_{j=1}^{n} a_{j,i} \Gamma_i(s_j, \cdot), \sum_{k=1}^{n} b_{k,i} \Gamma_i(s_k, \cdot) \right\rangle_{R_i} = \sum_{j=1}^{n} \sum_{k=1}^{n} a_{j,i} b_{k,i} \Gamma_i(s_j, s_k); \qquad (6)$$

notice that this implies $\langle \Gamma_i(s, \cdot), \Gamma_i(\cdot, t) \rangle_{R_i} = \Gamma_i(s, t)$. We now define the norm $||\omega_i||_{R_i} := \sqrt{\langle \omega_i, \omega_i \rangle_{R_i}}$. The closure of S_i under this norm is defined as the space R_i. Because we have assumed the processes $\bar{A}_{j,1}(\cdot)$ and $\bar{A}_{j,2}(\cdot)$ to be independent, we can define the reproducing kernel Hilbert space of the bivariate process $(\bar{A}_{j,1}(\cdot), \bar{A}_{j,2}(\cdot))$ by $R := R_1 \times R_2$. The inner product in R, with $\omega_{a,i}, \omega_{b,i} \in R_i$, obviously reads

$$\langle (\omega_{a,1}, \omega_{a,2}), (\omega_{b,1}, \omega_{b,2}) \rangle_R = \langle \omega_{a,1}, \omega_{b,1} \rangle_{R_1} + \langle \omega_{a,2}, \omega_{b,2} \rangle_{R_2}.$$

Now we can define the rate function of the sample-path LDP by

$$I(\omega) := \begin{cases} \frac{1}{2} ||\omega||_R^2 & \text{if } \omega \in R; \\ \infty & \text{otherwise.} \end{cases} \qquad (7)$$

Under assumptions (A1) and (A3) the following sample-path LDP holds.

Theorem 2.5 [Generalized Schilder] $n^{-1} \sum_{j=1}^{n} \bar{A}_{j,i}(\cdot)$ *satisfies the following LDP:*
(a) For any closed set $F \subset \Omega$*,*

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P} \left(\frac{1}{n} \sum_{j=1}^{n} \bar{A}_{j,i}(\cdot) \in F \right) \leq - \inf_{\omega \in F} I(\omega);$$

(b) For any open set $G \subset \Omega$*,*

$$\liminf_{n \to \infty} \frac{1}{n} \log \mathbb{P} \left(\frac{1}{n} \sum_{j=1}^{n} \bar{A}_{j,i}(\cdot) \in G \right) \geq - \inf_{\omega \in G} I(\omega).$$

3 Generic upper and lower bound on the probability

In a GPS framework the workloads of the queues are intimately related: it is not possible to write down an explicit expression for $Q_{i,n}(0)$, for $i = 1, 2$, without using the evolution of the workload in the other queue. This makes the analysis of GPS systems hard. In this section we derive explicit upper and lower bounds for $Q_{1,n}(0)$ in terms of the processes $\sum_{j=1}^{n} A_{j,i}(\cdot, \cdot)$, $i = 1, 2$.

In the remainder of this paper, we have to distinguish between two regimes. The most involved regime is $\mu_2 < \phi_2 c$, which we refer to as *underload for class 2*. In this regime, class 2 is stable

regardless of the behavior of the other class. The other regime is $\mu_2 \geq \phi_2 c$, the regime where class 2 is said to be in *overload*. Although the bounds that are derived in this section hold for both regimes, they are only useful in the regime with underload for class 2 – they will be exploited in Sections 4 and 5. The analysis for the regime with class 2 in overload is presented in Section 6.

Note that the results in this section hold regardless of the distribution of the inputs. We mention that bounds similar to the ones that we apply in the next lemmas, have been used in [26, 28].

Trivially, we can rewrite the overflow probability to

$$\mathbb{P}\left(Q_{1,n}(0) \geq nb\right) = \mathbb{P}\left(\bigcup_{x\geq 0} \left\{Q_{1,n}(0) + Q_{2,n}(0) \geq nx + nb, Q_{2,n}(0) \leq nx\right\}\right). \tag{8}$$

Because of the work-conserving nature of GPS, it is easily seen that the following relation holds for the total queue:

$$Q_{1,n}(0) + Q_{2,n}(0) = \sup_{t>0}\left\{\sum_{j=1}^{n}(A_{j,1}(-t,0) + A_{j,2}(-t,0)) - nct\right\}. \tag{9}$$

Substituting this relation for $Q_{1,n}(0) + Q_{2,n}(0)$ in the right-hand side of (8), we find

$$\mathbb{P}\left(\bigcup_{x\geq 0}\left\{\sup_{t>0}\left\{\sum_{j=1}^{n}(A_{j,1}(-t,0) + A_{j,2}(-t,0)) - nct\right\} \geq nx + nb, Q_{2,n}(0) \leq nx\right\}\right). \tag{10}$$

We denote the optimizing t in the above supremum by $t^\star$. Following [24], $-t^\star$ can be interpreted as the beginning of the busy period of the total queue containing time 0. Next we consider $Q_{2,n}(0)$. Let us denote by $-s^\star$ the beginning of the busy period of queue 2 containing time 0. Then clearly, $s^\star \in [0, t^\star]$, since the busy period of the total queue cannot start after the start of the busy period of queue 2. Now using the supremum relation (3), we obtain

$$Q_{2,n}(0) = \sup_{s\in(0,t]}\left\{\sum_{j=1}^{n} A_{j,2}(-s,0) - B_{2,n}(-s,0)\right\}. \tag{11}$$

In order to find bounds for $\mathbb{P}(Q_{1,n}(0) \geq nb)$, it follows from (10) that we need to bound the class-2 workload at time 0, $Q_{2,n}(0)$. Given its representation in (11), this means that we have to find bounds on the service that was available for class 2 during the busy period containing time 0.

We introduce the following additional notation:

$$\mathcal{E}_n := \left\{ \begin{array}{l} \exists x \geq 0, t > 0 : \forall s \in (0,t] : \\ \quad (1/n) \sum_{j=1}^{n} \left(A_{j,1}(-t,0) + A_{j,2}(-t,0) \right) \geq x + b + ct, \\ \quad (1/n) \sum_{j=1}^{n} A_{j,2}(-s,0) \leq x + \phi_2 cs \end{array} \right\};$$

$$\mathcal{F}_n := \left\{ \begin{array}{l} \exists x \geq 0, t > 0 : \forall s \in (0,t] : \exists u \in [0,s) : \\ \quad (1/n) \sum_{j=1}^{n} \left(A_{j,1}(-t,0) + A_{j,2}(-t,0) \right) \geq x + b + ct, \\ \quad (1/n) \sum_{j=1}^{n} \left(A_{j,2}(-s,0) + A_{j,1}(-s,-u) \right) \leq x + \phi_1 cu - cs \end{array} \right\}.$$

In the next lemmas we derive the lower and upper bound for the overflow probability of class 1.

Lemma 3.1 [Lower bound]

$$\mathbb{P}\left(Q_{1,n}(0) \geq nb\right) \geq \mathbb{P}\left(\mathcal{E}_n\right).$$

Proof. Recall that $-s^\star$ denotes the beginning of the busy period of queue 2 that contains time 0. Hence, the workload of class 2 is positive in the interval $(-s^\star, 0]$, indicating that class 2 claims at least its guaranteed rate in this interval: $B_{2,n}(-s^\star, 0) \geq n\phi_2 cs^\star$. Using this lower bound in (11), we derive

$$Q_{2,n}(0) \leq \sup_{s \in (0,t]} \left\{ \sum_{j=1}^{n} A_{j,2}(-s,0) - \phi_2 ncs \right\}. \tag{12}$$

The lower bound for $\mathbb{P}(Q_{1,n}(0) \geq nb)$ is now found by substituting (12) for $Q_{2,n}(0)$ in (10). $\square$

Lemma 3.2 [Upper bound]

$$\mathbb{P}\left(Q_{1,n}(0) \geq nb\right) \leq \mathbb{P}\left(\mathcal{F}_n\right).$$

Proof. From (11) it follows that we need an upper bound for $B_{2,n}(-s^\star, 0)$. We distinguish between two scenarios: (a) queue 1 is strictly positive during $(-s^\star, 0]$ and (b) queue 1 has been empty at some time in $(-s^\star, 0]$.

(a) Since both queues are strictly positive during $(-s^\star, 0]$, both classes claim their guaranteed rate, i.e., $B_{2,n}(-s^\star, 0) = n\phi_2 cs^\star$.

(b) Trivially, $B_{2,n}(-s^\star, 0) \leq ncs^\star - B_{1,n}(-s^\star, 0)$. Bearing in mind that queue 1 has been empty in $(-s^\star, 0]$, we define $u^\star := \inf\{u \in [0, s^\star) : Q_{1,n}(-u) = 0\}$. Hence both queues were strictly positive during $(-u^\star, 0]$, and consequently both classes are assigned their

guaranteed rates. Together with (2) this yields

$$\begin{aligned} B_{1,n}(-s^\star,0) &= B_{1,n}(-s^\star,-u^\star)+B_{1,n}(-u^\star,0) \\ &= Q_{1,n}(-s^\star)+\sum_{j=1}^{n} A_{j,1}(-s^\star,-u^\star)+n\phi_1 cu^\star \\ &\geq \inf_{u\in[0,s^\star)}\left\{\sum_{j=1}^{n} A_{j,1}(-s^\star,-u)+n\phi_1 cu\right\}. \end{aligned}$$

This implies

$$B_{2,n}(-s^\star,0) \leq ncs^\star - \inf_{u\in[0,s^\star)}\left\{\sum_{j=1}^{n} A_{j,1}(-s^\star,-u)+n\phi_1 cu\right\}. \tag{13}$$

As the right hand side of (13) is larger than $n\phi_2 cs^\star$, we derive

$$B_{2,n}(-s^\star,0) \leq ncs^\star - \inf_{u\in[0,s^\star)}\left\{\sum_{j=1}^{n} A_{j,1}(-s^\star,-u)+n\phi_1 cu\right\}.$$

We now use this upper bound in (11) to obtain

$$Q_{2,n}(0) \geq \sup_{s\in(0,t]}\left\{\sum_{j=1}^{n} A_{j,2}(-s,0)-ncs+\inf_{u\in[0,s)}\left\{\sum_{j=1}^{n} A_{j,1}(-s,-u)+n\phi_1 cu\right\}\right\}.$$

Substituting this for $Q_{2,n}(0)$ in (10) then yields the desired upper bound. □

Remark 3.3 Compare the sets $\mathcal{E}_n$ and $\mathcal{F}_n$; evidently, $\mathcal{E}_n \subseteq \mathcal{F}_n$. Any path f of the sample-mean process in $\mathcal{F}_n$ defines epochs $u^\star$ and $s^\star$ (as identified in the proof of Lemma 3.2). It is not hard to see that if these epochs match, f is also in $\mathcal{E}_n$. From the proof of Lemma 3.2, taking $u^\star = s^\star$ means that scenario (a) applies, where queue 1 is strictly positive during the busy period of queue 2 containing time 0. These simple observations turn out to play a crucial role in the discussion presented in Section 7.

4 Lower bound on the decay rate: class 2 in underload

Sections 4 and 5 concern the regime in which class 2 is in underload, i.e., $\mu_2 < \phi_2 c$. In Section 4 we determine the decay rate of the upper bound on $\mathbb{P}(Q_{1,n}(0) \geq nb)$ as presented in Lemma 3.2. Then in Section 5 we calculate the decay rate of the lower bound on $\mathbb{P}(Q_{1,n}(0) \geq nb)$ as presented in Lemma 3.1.

Because of Lemma 3.2,

$$-\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left(Q_{1,n}(0)\geq nb\right)\geq-\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left(\mathcal{F}_n\right).$$

We now investigate the decay rate in the right-hand side of the previous display. Defining the set of paths

$$\mathcal{A}_{b,x}^{s,t,u}:=\left\{f\;\middle|\;\begin{array}{l}A_1[f](-t,0)+A_2[f](-t,0)\geq x+b+ct,\\ A_2[f](-s,0)+A_1[f](-s,-u)\leq x-\phi_1 cu+cs\end{array}\right\},$$

Schilder's sample-path LDP yields

$$-\lim_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left(\mathcal{F}_{\mathbf{n}}\right)=\inf_{x\geq0}J^L(b,x),\;\text{ where }\;J^L(b,x):=\inf_{t>0}\inf_{f\in\bigcap\limits_{s\in(0,t]}\bigcup\limits_{u\in[0,s)}\mathcal{A}_{b,x}^{s,t,u}}I(f).\quad(14)$$

Notice that we used that the decay rate of a union of events is just the infimum of the individual decay rates. Unfortunately, we do not have such a relation for an intersection of events. However, it is possible to find an explicit lower bound, as presented in the next theorem.

Theorem 4.1

$$-\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left(Q_{1,n}(0)\geq nb\right)\geq-\inf_{x\geq0}J^L(b,x)$$

where

$$J^L(b,x)\geq\inf_{t>0}\sup_{s\in(0,t]}\inf_{u\in[0,s)}\inf_{f\in\mathcal{A}_{b,x}^{s,t,u}}I(f).\qquad(15)$$

Proof. The first claim follows directly from the above. Now consider the second claim. Because for all $s\in(0,t]$, for given t,

$$\bigcap_{r\in(0,t]}\bigcup_{u\in[0,r)}\mathcal{A}_{b,x}^{r,t,u}\subseteq\bigcup_{u\in[0,s)}\mathcal{A}_{b,x}^{s,t,u},$$

we have for all $s\in(0,t]$,

$$\inf_{f\in\bigcap\limits_{s\in(0,t]}\bigcup\limits_{u\in[0,s)}\mathcal{A}_{b,x}^{s,t,u}}I(f)\geq\inf_{f\in\bigcup\limits_{u\in[0,s)}\mathcal{A}_{b,x}^{s,t,u}}I(f).$$

Hence, it also holds for the maximizing s,

$$\inf_{f\in\bigcap\limits_{s\in(0,t]}\bigcup\limits_{u\in[0,s)}\mathcal{A}_{b,x}^{s,t,u}}I(f)\geq\sup_{s\in(0,t]}\inf_{f\in\bigcup\limits_{u\in[0,s)}\mathcal{A}_{b,x}^{s,t,u}}I(f).$$

This implies the second claim. □

5 Upper bound on the decay rate: class 2 in underload

This section concentrates on the decay rate of the lower bound on $\mathbb{P}(Q_{1,n}(0) \geq nb)$ as given in Lemma 3.1. The procedure turns out to be more involved than that of Section 4.

Because of Lemma 3.1,

$$-\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(Q_{1,n}(0) \geq nb\right) \leq -\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\mathcal{E}_n\right).$$

We now investigate the decay rate in the right-hand side of the previous display. Define the set of paths

$$\mathcal{A}_{b,x}^{s,t} := \{f \mid A_1[f](-t,0) + A_2[f](-t,0) \geq x + b + ct, A_2[f](-s,0) \leq x + \phi_2 cs\}.$$

Similarly to Theorem 4.1, Schilder's sample-path LDP yields the following upper bound.

Lemma 5.1

$$-\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(Q_{1,n}(0) \geq nb\right) \leq \inf_{x\geq 0} J^U(b,x), \text{ where } J^U(b,x) := \inf_{t>0} \inf_{f \in \bigcap_{s\in(0,t]} \mathcal{A}_{b,x}^{s,t}} I(f).$$

The objective of this section is to prove that, under some assumptions,

$$J^U(b,x) = \inf_{t>0} \sup_{s\in(0,t]} \inf_{f\in\mathcal{A}_{b,x}^{s,t}} I(f). \tag{16}$$

Again, because of the fact that an intersection is involved, no explicit expression for $J^U(b,x)$ is available. We therefore take the following approach: we first derive in Section 5.1 a lower bound for $J^U(b,x)$, and then in Section 5.2 we give conditions under which this lower bound matches the exact value of $J^U(b,x)$.

Remark 5.2 Notice the similarity between the right-hand sides of (15) and (16), in particular if the optimizing s and u in (15) coincide, see also Remark 3.3.

5.1 Lower bound on $J^U(b,x)$

The following lemma gives a lower bound for $J^U(b,x)$. Its proof is analogous to that of the second claim in Theorem 4.1.

Lemma 5.3

$$J^U(b,x) \geq \inf_{t>0} \sup_{s\in(0,t]} \inf_{f\in\mathcal{A}_{b,x}^{s,t}} I(f).$$

The lower bound in Lemma 5.3 can be expressed more explicitly. To this end, we first concentrate on calculating the minimum of $I(f)$ over $f \in \mathcal{A}_{b,x}^{s,t}$, for fixed s and$t$. The result, as stated in Lemma 5.4, requires the introduction of two functions. First recall the large-deviations rate function $\Lambda(\cdot,\cdot)$, of the bivariate Normal random variable $(A_1(-t,0)+A_2(-t,0), A_2(-s,0))$, as given in (5),

$$\Lambda(y_1,y_2) := \frac{1}{2}\begin{pmatrix} y_1-\mu t \\ y_2-\mu_2 s \end{pmatrix}^{\mathrm{T}} \Sigma(s,t)^{-1} \begin{pmatrix} y_1-\mu t \\ y_2-\mu_2 s \end{pmatrix},\ \Sigma(s,t) = \begin{pmatrix} v(t) & \Gamma_2(s,t) \\ \Gamma_2(s,t) & v_2(s) \end{pmatrix}.$$

We also define

$$k_i(x,s,t) := \mu_i s + \frac{(x+b+(c-\mu)t)}{v(t)}\,\Gamma_i(s,t).$$

Lemma 5.4 *For $s \in (0,t]$,*

$$\inf_{f\in\mathcal{A}_{b,x}^{s,t}} I(f) = \Upsilon_{b,x}(s,t) := \begin{cases} \Lambda(x+b+ct, x+\phi_2 cs), & \text{if } k_2(x,s,t) > x+\phi_2 cs; \\ (x+b+(c-\mu)t)^2/2v(t), & \text{if } k_2(x,s,t) \le x+\phi_2 cs. \end{cases}$$

Proof. Using Theorem 2.3,

$$\inf_{f\in\mathcal{A}_{b,x}^{s,t}} I(f) = \inf_{y_1\ge x+b+ct, y_2\le x+\phi_2 cs} \Lambda(y_1,y_2).$$

Because $\Lambda(\cdot,\cdot)$ is convex in y_1 and y_2, we can use the Lagrangian to find the infimum over y_1 and y_2:

$$\mathcal{L}(y_1,y_2,\xi_1,\xi_2) = \Lambda(y_1,y_2) - \xi_1(y_1-x-b-ct) + \xi_2(y_2-x-\phi_2 cs),$$

with $\xi_1,\xi_2 \ge 0$. Two cases may occur, depending on the specific values of x, s and t. (i) If x, s and t are such that $k_2(x,s,t) > x+\phi_2 cs$, then both constraints are binding, i.e., $y_1 = x+b+ct$ and $y_2 = x+\phi_2 cs$. (ii) If x, s and t are such that $k_2(x,s,t) \le x+\phi_2 cs$, then only the first constraint is binding, i.e., $y_1 = x+b+ct$, and $y_2 = k_2(x,s,t)$. □

Remark 5.5 Note that the $\theta^\star$ in Theorem 2.3 are related to the Lagrange multipliers ξ_1 and ξ_2 that are used in the proof of Lemma 5.4. In case (i) $\theta_1^\star(s,t) = \xi_1 > 0$ and $\theta_2^\star(s,t) = -\xi_2 > 0$, whereas in case (ii) $\theta_1^\star(s,t) = \xi_1 > 0$ and $\theta_2^\star(s,t) = -\xi_2 = 0$.

Observe that $\Upsilon_{b,x}(s,t)$ is continuous at $s \downarrow 0$, i.e.,

$$\Upsilon_{b,x}(0,t) = \frac{(x+b+(c-\mu)t)^2}{2v(t)}.$$

Now Lemmas 5.3 and 5.4 yield the final lower bound for $J^U(b,x)$, as stated in the next corollary.

Corollary 5.6

$$J^U(b,x) \geq \inf_{t>0} \sup_{s\in(0,t]} \Upsilon_{b,x}(s,t).$$

Interpretation of $\Upsilon_{b,x}(s,t)$. The decay rate $I(f)$ can be interpreted as the cost of having a path f, and, likewise, $\Upsilon_{b,x}(s,t)$ as the cost of generating a traffic pattern in the set $\mathcal{A}_{b,x}^{s,t}$. The proof of Lemma 5.4 shows that the *first* constraint, i.e., $y_1 \geq x+b+ct$ is always binding, whereas the *second* constraint, i.e., $y_2 \leq x+\phi_2 cs$, is sometimes binding, depending on the value of $k_2(x,s,t)$ compared to $x+\phi_2 cs$. Observe that $k_2(x,s,t)$ is in fact a conditional expectation:

$$k_2(x,s,t) \equiv \mathbb{E}[A_2(-s,0) \mid A_1(-t,0)+A_2(-t,0) = x+b+ct].$$

The two cases of Lemma 5.4 can now be interpreted as follows. (i) The optimal value for y_2 is $x+\phi_2 cs$. In this case, $k_2(x,s,t)$, which is the expected value of the amount of traffic sent by class 2 in $(-s,0]$ given that in total $x+b+ct$ is sent during $(-t,0]$, is larger than $x+\phi_2 cs$: with high probability the second constraint is *not met* just by imposing the first constraint. In terms of cost, this means that in this regime additional cost is incurred by imposing the second constraint. (ii) The optimal value for y_2 is precisely $k_2(x,s,t)$, and is smaller than $x+\phi_2 cs$: $A_1(-t,0)+A_2(-t,0) = x+b+ct$ implies $A_2(-s,0) > x+\phi_2 cs$ with high probability. Intuitively this means that, given that the first constraint is satisfied, the second constraint is already met, with high probability.

Using this reasoning, it follows after some calculations that we can rewrite $\Upsilon_{b,x}(s,t)$ in a helpful way as shown in the next corollary. The first term accounts for the cost of satisfying the first constraint in $\mathcal{A}_{b,x}^{s,t}$, the second term (which is possibly 0) for the second constraint.

Corollary 5.7

$$\Upsilon_{b,x}(s,t) = \frac{(x+b+ct-\mathbb{E}[A_1(-t,0)+A_2(-t,0)])^2}{2\mathbb{V}\mathrm{ar}[A_1(-t,0)+A_2(-t,0)]} + \frac{\max^2\{\mathbb{E}[A_2(-s,0) \mid A_1(-t,0)+A_2(-t,0) = x+b+ct] - x - \phi_2 cs, 0\}}{2\mathbb{V}\mathrm{ar}[A_2(-s,0) \mid A_1(-t,0)+A_2(-t,0) = x+b+ct]}.$$

Two regimes for ϕ_2. Corollary 5.7 implies that

$$\inf_{t>0} \sup_{s\in(0,t]} \Upsilon_{b,x}(s,t) \geq \inf_{t>0} \frac{(x+b+(c-\mu)t)^2}{2v(t)}. \tag{17}$$

Let the optimum in the right-hand side be attained in t^c (which is, in fact, a function of x, but we suppress x here, as x is held fixed in this section). Suppose that for all $s \in (0,t^c]$ it holds

that $k_2(x,s,t^c) < x + \phi_2 cs$, then obviously the inequality in (17) is tight. This corresponds to a critical weight $\phi_2^{c,U}(x)$ above which there is tightness. This critical value is given by

$$\begin{aligned}\phi_2^{c,U}(x) &:= \inf\left\{\phi_2 : \sup_{s\in(0,t^c]}\{k_2(x,s,t^c) - x - \phi_2 cs\} \le 0\right\} \\ &\equiv \sup_{s\in(0,t^c]} \frac{k_2(x,s,t^c) - x}{cs}. \end{aligned} \tag{18}$$

The resulting two regimes can be intuitively described as follows.

Large ϕ_2. If $\phi_2 > \phi_2^{c,U}(x)$, using the interpretation in terms of conditional expectations, the buffer content of queue 2 at time 0 is likely to be below nx. Hence, if in total $n(x + b + ct^c)$ is sent during $(-t^c, 0]$, it is likely that at time 0, the buffer of class 1 has value nb.

Small ϕ_2. If $\phi_2 < \phi_2^{c,U}(x)$ then the guaranteed rate for class 2 is relatively small, meaning that its buffer content may easily grow. Again in total (at least) $n(x + b + ct^c)$ has been sent during the interval $(-t^c, 0]$, but now it is *not* obvious that most of it goes to the buffer of class 1. Class 2 has to be 'forced' to take *at most* its guaranteed rate during this interval.

5.2 Conditions for exactness

As the overflow behavior in case of $\phi_2 \ge \phi_2^{c,U}(x)$ is essentially different from that in case of $\phi_2 < \phi_2^{c,U}(x)$, we will consider in this section the two regimes separately.
The procedure followed will be the same for both regimes. Let us denote the optimizing s and t in Corollary 5.6 by $s^\star$ and $t^\star$, respectively. (Notice that $s^\star$ and $t^\star$ are functions of x, but, for conciseness, we again suppress the argument x.) First we use Schilder's theorem to determine the most probable path in $\mathcal{A}_{b,x}^{s^\star,t^\star}$ for the regime of ϕ_2 under consideration. Denoting this optimal path by $f^\star$ we then check whether

$$f^\star \in \left(\bigcup_{t\ge 0}\bigcap_{s\in(0,t]} \mathcal{A}_{b,x}^{s,t}\right). \tag{19}$$

If so, the optimal path giving rise to the lower bound of Corollary 5.6, is in fact the optimal path for $J^u(b,x)$. Consequently, under the condition (19), the lower bound and $J^u(b,x)$ coincide.

Case A: ϕ_2 larger than critical weight

Because of the definition of $\phi_2^{c,U}(x)$, it holds for all $\phi_2 \ge \phi_2^{c,U}(x)$ that

$$\inf_{t>0}\sup_{s\in(0,t]} \Upsilon_{b,x}(s,t) = \frac{(x + b + (c-\mu)t^c)^2}{2v(t^c)},$$

as identified before. The next theorem states that, for these ϕ_2, the lower bound on $J^U(b,x)$ (see Corollary 5.6) actually *equals* $J^U(b,x)$. We omit its proof because it essentially follows from the proof of Theorem 3.9 in [17].

Theorem 5.8 *If $\phi_2 \geq \phi_2^{c,U}(x)$, then*

$$J^U(b,x) = \inf_{t>0} \sup_{s\in(0,t]} \Upsilon_{b,x}(s,t) = \frac{(x+b+(c-\mu)t^c)^2}{2v(t^c)},$$

and the most probable paths are, for $r \in [-t^c, 0)$,

$$f_1^\star(r) = -\mathbb{E}[A_1(r,0) \mid A_1(-t^c,0) + A_2(-t^c,0) = x+b+ct^c] = -k_1(x,-r,t^c);$$

$$f_2^\star(r) = -\mathbb{E}[A_2(r,0) \mid A_1(-t^c,0) + A_2(-t^c,0) = x+b+ct^c] = -k_2(x,-r,t^c).$$

Case B: ϕ_2 smaller than critical weight

The analysis of this regime is more involved than that of case A. First we will show in the next lemma that in this regime both constraints in Lemma 5.4 are met with equality. Its proof is omitted here, as it is along the lines of Lemma 3.10 in [17]. Note that Assumptions (A2) and (A3) are used in the proof.

Lemma 5.9 *If $\phi_2 < \phi_2^{c,U}(x)$, then $k_2(x,s^\star,t^\star) > x + \phi_2 c s^\star$.*

The next lemma gives the most probable paths in the set $\mathcal{A}_{b,x}^{s,t}$ for the regime where for given x, s and t we have that $k_2(x,s,t) > x + \phi_2 cs$. We give the most probable paths for $r \in [-t, 0)$, but they can be trivially extended to the entire real axis.

Lemma 5.10 *If $k_2(x,s,t) > x + \phi_2 cs$, then, for $r \in [-t,0)$, the most probable paths in $\mathcal{A}_{b,x}^{s,t}$ are*

$$f_1(r) = -\mathbb{E}[A_1(r,0) \mid A_1(-t,0) + A_2(-t,0) = x+b+ct, A_2(-s,0) = x+\phi_2 cs];$$

$$f_2(r) = -\mathbb{E}[A_2(r,0) \mid A_1(-t,0) + A_2(-t,0) = x+b+ct, A_2(-s,0) = x+\phi_2 cs].$$

Proof. This is shown by using the arguments of the proofs of Lemma 3.11 in [17] and Proposition 1 in [21]. □

Easy calculations show that we can rewrite the above-mentioned paths as

$$f_1(r) = \mu_1 r - \theta_1^\star(s,t)\Gamma_1(-r,t);$$

$$f_2(r) = \mu_2 r - \theta_1^\star(s,t)\Gamma_2(-r,t) - \theta_2^\star(s,t)\Gamma_2(-r,s),$$

where the $\theta^\star$ follow from Theorem 2.3 (see also Remark 5.5). Interestingly, only one covariance function is involved in the most probable path of class 1, meaning that its path will be symmetric around $-\frac{1}{2}t$.

Now we present conditions under which the lower bound of Corollary 5.6 matches $J^U(b,x)$, with an approach that is similar to the one followed in [17] for tandem and priority queues. First we introduce new notation. For $r_1 < r_2$,

$$\overline{\mathbb{E}}A_i(r_1,r_2) := \mathbb{E}[A_i(r_1,r_2) \mid A_1(-t^\star,0) + A_2(-t^\star,0) = x + b + ct^\star], \quad i = 1,2,$$

with $\mathbb{V}\overline{\text{ar}}(\cdot)$ and $\mathbb{C}\overline{\text{ov}}(\cdot,\cdot)$ defined similarly. For $r \in (-t^\star, 0)$ we define the functions

$$\overline{m}(r) := \frac{\overline{\mathbb{E}}A_2(r,0) - x + \phi_2 cr}{\sqrt{\mathbb{V}\overline{\text{ar}}A_2(r,0)}}, \qquad m(r) := \frac{\overline{m}(r)}{\overline{m}(-s^\star)},$$

$$\rho(r) := \frac{\mathbb{C}\overline{\text{ov}}(A_2(r,0), A_2(-s^\star,0))}{\sqrt{\mathbb{V}\overline{\text{ar}}A_2(r,0)\,\mathbb{V}\overline{\text{ar}}A_2(-s^\star,0)}}.$$

Again, we should formally write $m_x(\cdot)$ and $\rho_x(\cdot)$ to indicate the dependence on x, but we leave out the subscript x in this section. Both $m(\cdot)$ and $\rho(\cdot)$ attain a maximum 1 at $r = -s^\star$; for $m(\cdot)$ this follows from Corollary 5.7 and Lemma 5.9; for $\rho(\cdot)$ from the fact that it is a correlation coefficient.

Theorem 5.11 *If $\phi_2 < \phi_2^{c,U}(x)$, then*

$$J^U(b,x) = \inf_{t>0} \sup_{s\in(0,t]} \Upsilon_{b,x}(s,t) = \Lambda(x + b + ct^\star, x + \phi_2 cs^\star),$$

under the condition that $m(r) \le \rho(r)$ for all $r \in (-t^\star, 0)$. The corresponding most probable paths are, for $i = 1,2$,

$$f_i^\star(r) = -\overline{\mathbb{E}}[A_i(r,0) \mid A_2(-s^\star,0) = x + \phi_2 cs^\star].$$

Proof. We have to show that (19) holds. Straightforward calculations show that indeed $A_1[f^\star](-t^\star,0) + A_2[f^\star](-t^\star,0) = x + b + ct^\star$, as desired. Now it remains to be shown that, if $m(r) \le \rho(r)$ for all $r \in (-t^\star,0)$, then $A_2[f^\star](r,0) \le x - \phi_2 cr$ for all $r \in (-t^\star,0)$. This follows immediately from the following (standard) decomposition:

$$\begin{aligned} A_2[f^\star](r,0) &= -f_2^\star(r) \\ &= \overline{\mathbb{E}}A_2(r,0) + \frac{\mathbb{C}\overline{\text{ov}}[A_2(r,0), A_2(-s^\star,0)]}{\mathbb{V}\overline{\text{ar}}A_2(-s^\star,0)} \left(x + \phi_2 cs^\star - \overline{\mathbb{E}}A_2(-s^\star,0)\right). \end{aligned}$$

The fact that the decay rate now equals $\Lambda(x + b + ct^\star, x + \phi_2 cs^\star)$ is due to Lemma 5.9. This proves the stated. $\square$

Remark 5.12 Note that the condition $m(r) \leq \rho(r)$ for all $r \in (0, t^\star)$ only involves properties of the class-2 input process. The above Theorem 5.11 therefore holds for *any* class-1 Gaussian process with stationary increments.

Remark 5.13 Following the approach in [17], the optimal input rate paths $g_1(\cdot)$ and $g_2(\cdot)$, which are the first derivatives of $f_1^\star(\cdot)$ and $f_2^\star(\cdot)$, can be calculated. Assuming $v'(0) = 0$, these paths exhibit similar properties as those in [17]: (i) $g_1(-t^\star) + g_2(-t^\star) = c$ and (ii) $g_2(-s^\star) = \phi_2 c$. Hence, at time $-t^\star$ the total input rate is c, making the server operate at full capacity. Then at time $-s^\star$ the total input rate of queue 2 is $\phi_2 c$, meaning that queue 2 starts claiming its guaranteed rate.

6 Analysis of the decay rate: class 2 in overload

In this section the decay rate of $\mathbb{P}(Q_{1,n}(0) \geq nb)$ is calculated for the regime $\phi_2 c \leq \mu_2$.

Theorem 6.1 *If $\phi_2 \leq \mu_2/c$, then*

$$-\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(Q_{1,n}(0) \geq nb) = \inf_{t\geq 0} \frac{(b + (\phi_1 c - \mu_1)t)^2}{2v_1(t)}. \tag{20}$$

Proof. We first show that the desired expression is a lower bound. Denote by $Q_{i,n}^{nc}(0)$ the stationary workload of queue i if it is served (in isolation) at a constant rate nc. Then the lower bound follows from

$$\mathbb{P}(Q_{1,n}(0) \geq nb) \leq \mathbb{P}\left(\exists t > 0 : \frac{1}{n}\sum_{j=1}^{n} A_{j,1}(-t, 0) \geq b + \phi_1 ct\right),$$

due to $Q_{1,n}(0) \leq Q_{1,n}^{n\phi_1 c}(0)$.
The upper bound is a matter of computing the rate function of a feasible path. Let $t^\star$ be the optimizer in the right-hand side of (20). For $r \in [-t^\star, 0)$ define

$$f_1^\star(r) := -\mathbb{E}[A_1(r, 0) \mid A_1(-t^\star, 0) = b + \phi_1 ct^\star] = \mu_1 r - \frac{(b + (\phi_1 c - \mu_1)t^\star)}{v_1(t^\star)} \Gamma_1(-r, t^\star);$$

$$f_2^\star(r) := -\mathbb{E}[A_2(r, 0) \mid A_1(-t^\star, 0) = b + \phi_1 ct^\star] = \mu_2 r.$$

This path clearly leads to overflow in queue 1 of the GPS system (as the type-2 sources claim their weight, such that exactly service rate $n\phi_1 c$ is left for the type-1 sources). The norm of $f_2^\star(\cdot)$ is obviously 0, as these sources are transmitting at mean rate; the rate function corresponding to $f_1^\star(\cdot)$ equals the desired expression. □

7 Discussion of the results

In this section we will discuss the results of the previous sections. We identify three regimes for the value of ϕ_2, corresponding to three generic overflow scenarios. Case (i) directly relates to the overload regime of Section 6; Cases (ii) and (iii) to the underload regime of Sections 4 and 5.

For Case (i) our analysis immediately yields the exact decay rate, see Theorem 6.1. For Cases (ii) and (iii), however, the situation is more complicated. Theorems 4.1, 5.8, and 5.11 provide *bounds* on the decay rate. We strongly believe, however, that under fairly general conditions these bounds coincide. This claim is justified (1) by heuristic arguments in Section 7.1, (2) by extensive numerical experiments, as reported in Section 7.2, and (3) by explicit results for the special case of Brownian motion input in Section 7.3. In this section we use $J(b)$ to denote the decay rate of $\mathbb{P}(Q_{1,n}(0) \geq nb)$, given that it exists.

7.1 Structure of the solution

Ad Case (i): Class 2 in overload. First consider the situation $\phi_2 \leq \mu_2/c =: \phi_2^o$. In this scenario the type-2 sources claim their guaranteed rate $n\phi_2 c$ with overwhelming probability, so that overflow in queue 1 resembles overflow in a FIFO queue with link rate $n\phi_1 c$; this principle plays a crucial role in the proof of Theorem 6.1. We repeat it here for comparison with Cases (ii) and (iii).

For $\phi_2 \in [0, \phi_2^o]$:

$$J(b) = \inf_{t>0} \frac{(b + (\phi_1 c - \mu_1)t)^2}{2v_1(t)}.$$

Ad Case (ii): Class 2 in underload, with ϕ_2 small. As argued in Sections 4 and 5, in this regime it is not sufficient to require that $n(x + b + ct)$ traffic is generated in t units of time, since, with high probability, a considerable amount of traffic will be left in queue 2. Hence, additional effort is required to ensure that queue 2 stays below nx.

Based on heuristic arguments, we present two claims.

A. *Regarding the optimal values of u and s.* Recall the probabilistic upper bound in Lemma 3.2. In the proof of that lemma, $-s^\star$ denotes the beginning of the busy period of the second queue, which contains time 0. Hence, the second queue remains backlogged during the interval $(-s^\star, 0]$ and claims at least its guaranteed rate $n\phi_2 c$, leaving at most rate $n\phi_1 c$ to the first queue. Parallelling the proof of Lemma 3.2, two scenarios are possible: in scenario (a) queue 1 was continuously backlogged during $(-s^\star, 0]$, whereas

in scenario (b), queue 1 has been empty after time $-s^\star$, i.e., queue 1 was empty at some time $-u^\star$ during the busy period of queue 2.

Scenario (b) is not likely to be optimal, for the following reason. As queue 1 was empty at $-u^\star$, it does not benefit from any effort before $-u^\star$; queue 1 has to build up its entire buffer in the interval $(-u^\star, 0]$. Now recall that queue 2 already started to show deviant behavior from time $-s^\star < -u^\star$, claiming its guaranteed rate. However this additional effort of queue 2 before time $-u^\star$ is of no 'benefit' for queue 1. In order for queue 1 to fully exploit that queue 2 takes its guaranteed rate during $(-s^\star, 0]$, it should be continuously backlogged during this interval, as in scenario (a). We therefore expect that in the most likely scenario $u^\star = s^\star$.

B. *Regarding the optimal value of x.* We introduced x in the left-hand side of (8). From this representation it follows immediately that nx can be interpreted as the amount of traffic left in queue 2 (at the epoch when the total queue size reaches $n(x+b)$).

We argued before that queue 2 has to claim its guaranteed rate during $(-s^\star, 0]$. If a positive amount of traffic is left in queue 2 at time 0, the type-2 sources apparently 'generated too much traffic'; the guaranteed rate could have been claimed with less effort. We therefore expect that in the most likely scenario $x^\star = 0$. Notice that an essential condition here is that $\phi_2 > \mu_2/c$, as otherwise a build-up of traffic in queue 2 would not be 'wasted effort'.

Because of Claims A and B, we expect that this regime applies to $\phi_2 \in [\phi_2^o, \phi_2^c]$, with

$$\phi_2^c := \sup_{s \in (0, t^c(0)]} \frac{k_2(0, s, t^c(0))}{cs}.$$

Define

$$\begin{pmatrix} z_1(t) \\ z_2(s) \end{pmatrix} := \begin{pmatrix} b + (c-\mu)t \\ (\phi_2 c - \mu_2)s \end{pmatrix}.$$

We expect that the following relation holds:

For $\phi_2 \in [\phi_2^o, \phi_2^c]$:

$$J(b) = \frac{1}{2} \inf_{t \geq 0} \sup_{s \in [0,t]} \begin{pmatrix} z_1(t) \\ z_2(s) \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} v_1(t) + v_2(t) & \Gamma_2(s,t) \\ \Gamma_2(s,t) & v_2(s) \end{pmatrix}^{-1} \begin{pmatrix} z_1(t) \\ z_2(s) \end{pmatrix},$$

provided that for all $r \in (-t^\star(0), 0)$ it holds that $m_0(r) \leq \rho_0(r)$.

Ad Case (iii): Class 2 in underload, with ϕ_2 large. Here overflow of the total queue implies overflow of queue 1. Consequently we expect the following relation.

For $\phi_2 \in [\phi_2^c, 1]$:

$$J(b) = \inf_{t>0} \frac{(b + (c-\mu)t)^2}{2v(t)}.$$

In Appendix A.1 it is formally shown that this result applies for $\phi_2 \in [\sup_{x\geq 0} \phi_2^{c,U}(x), 1]$. Arguments similar to claim B above (and extensive numerical experiments) however suggest that $\sup_{x\geq 0} \phi_2^{c,U}(x) = \phi_2^c$.

7.2 Numerical results

Section 7.3 verifies the claims of Section 7.1 for the special case of Brownian inputs. Extensive numerical experiments, however, suggest that the claims are valid under considerably more general conditions – we have not found any counterexamples so far. In this subsection we present two numerical examples.

Example 1. In this example type-1 sources are fractional Brownian motion (fBm) with $\mu_1 = 0.2$ and $v_1(t) = t^{2H}$, with Hurst parameter $H = 0.75$, whereas type-2 sources are Ornstein-Uhlenbeck (OU) sources with $\mu_2 = 0.3$ and $v_2(t) = t + e^{-t} - 1$. Take $c = 1$ and $b = 1$. Here $\phi_2^o = 0.3$, while numerical computations yield that $\phi_2^c = 0.4914$. Empirically, it turns out that in Case (ii) where $\phi_2 \in [\phi_2^o, \phi_2^c]$, it holds that $m_0(r) \leq \rho_0(r)$ for all $r \in (-t^\star(0), 0)$. Hence we can compare the upper and lower bounds. As they turn out to match, we conclude that we found the exact value of the decay rate. Regarding Case (iii) where $\phi_2 \in [\phi_2^c, 1]$, we empirically find that indeed $\sup_{x\geq 0} \phi_2^{c,U}(x) = \phi_2^c$, implying the correctness of the relation that we expected.

A specific example is considered in the left panel of Figure 1. There we focus on a situation in which ϕ_2 is in regime (ii): $\phi_2 = 0.4$. Numerical computations yield that $x^\star = 0$, $t^\star = 6.1819$, while $s^\star = u^\star = 5.6853$. The figure shows the traffic rates of both classes as a function of time. The total buffer starts to build up at time $-t^\star$, whereas queue 2 starts a busy period at $-s^\star$. More detailed inspection yields that with these traffic rates, at time 0 the first queue has indeed overflow, whereas the second queue is empty – in other words: the path is feasible.

Example 2. In this example we interchange the two classes of Example 1. Now $\phi_2^o = 0.2$ and $\phi_2^c = 0.7232$. We again find $\sup_{x\geq 0} \phi_2^{c,U}(x) = \phi_2^c$, so that the relation we expected for Case (iii) holds.

For Case (ii) however, we now do *not* find the exact decay rate. Consider the example $\phi_2 = 0.4$. In the computation of the lower bound we find $x^\star = 0$, $t^\star = 5.0723$, and $s^\star = u^\star = 5.0597$. Again we verified the 'exactness condition', but now we found $r \in (-t^\star, 0)$ such that $m(r) > \rho(r)$ – hence, the upper bound does not hold. The right panel of Figure 1 explains what

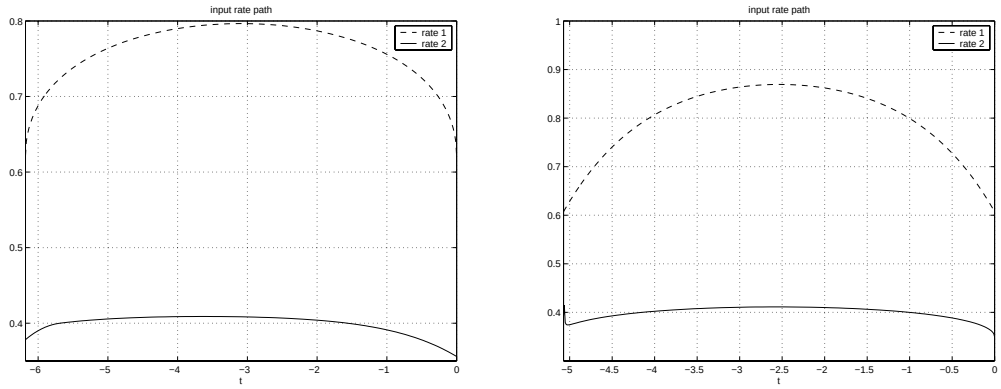

Figure 1: Left panel: type 1 corresponds to fBm and type 2 to OU; right panel: type 1 corresponds to OU and type 2 to fBm.

happens. The corresponding input rate path of the fBm sources has a 'dip' at time $-s^\star$. Consequently this path is *not* feasible: it is true that the sources build up $b + ct^\star$ traffic, as desired, but a positive amount of traffic is left in the second queue at time 0.
Despite the fact that in this case our approach does not yield the exact outcome of the decay rate in Case (ii), it still provides us with useful information. (1) In the first place, we do not have an upper bound, but fortunately the *lower* bound on the decay rate still applies. Such a lower bound corresponds to an upper bound on the probability of interest, which is of practical interest, as typically communication networks have to be designed such that overflow is sufficiently rare. (2) Numerical experiments showed that the amount of fluid left in the second queue at time 0 is usually extremely small. This makes us believe that the lower bound is relatively close to the exact outcome. (3) (Rough) full-link approximations, as introduced in [19], optimize over paths f such that there is a $t > 0$ such that $A_1[f](-t, 0) + A_2[f](-t, 0)$ exceeds $b + ct$, while at the same time $A_2[f](-t, 0) \leq \phi_2 ct$. It is easily seen that this procedure provides a more conservative lower bound (as it *a priori* chooses $s = t$). The observations (1), (2), and (3) justify to use, in Case (ii), the lower bound as an approximation, as is done in Section 8.1.

7.3 Brownian motion input

In this section we consider the special case that both types of sources correspond to Brownian motions: $v_1(t) = \lambda_1 t$, $v_2(t) = \lambda_2 t$. The formulae from the previous subsection can be evaluated explicitly, as shown in [16]. The result is given below. In particular, in the proof of this result – see [16] – it turns out that both Claims A and B hold.

Theorem 7.1 *Suppose $v_i(t) = \lambda_i t$, $i = 1, 2$. Then, with*

$$\phi_2^c = 1 - \frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2} \left(1 - \frac{\mu_1 + \mu_2}{c}\right) - \frac{\mu_1}{c}; \quad \phi_2^o = \frac{\mu_2}{c},$$

it holds that (i) for $\phi_2 \in [0, \phi_2^o]$,

$$J(b) = 2\, \frac{\phi_1 c - \mu_1}{\lambda_1}\, b;$$

(ii) for $\phi_2 \in [\phi_2^o, \phi_2^c]$,

$$t^\star = s^\star = u^\star = b \Big/ \sqrt{(\phi_1 c - \mu_1)^2 + (\phi_2 c - \mu_2)^2 \frac{\lambda_1}{\lambda_2}}\,;$$

$$J(b) = \frac{1}{2} \left(\frac{(b + (\phi_1 c - \mu_1) t^\star)^2}{\lambda_1 t^\star} + \frac{(\phi_2 c - \mu_2))^2}{\lambda_2} t^\star \right);$$

(iii) for $\phi_2 \in [\phi_2^c, 1]$,

$$J(b) = 2\, \frac{c - \mu}{\lambda_1 + \lambda_2}\, b.$$

Notice that in all three Cases (i)-(iii) it holds that $J(b)$ is linear in b; for Case (ii) it takes some simple algebra to see this.

8 Weight setting

This section focuses on the operational issue of selecting appropriate weights in a two-class GPS system. With any set of weights $\phi \equiv (\phi_1, \phi_2)$ an *admissible region* $\mathcal{S}(\phi)$ can be associated, i.e., combinations of sources of both classes such that the required QoS is realized. Obviously the size and shape of $\mathcal{S}(\phi)$ depends critically on the weights ϕ chosen. We refer to, e.g., Zhang *et al.* [29] for a study on these admissible regions $\mathcal{S}(\phi)$ for given weights (in the setting of short-range dependent inputs in the large-buffer regime).

When selecting appropriate weights, various objectives could be chosen. In this section we investigate two approaches. Following Elwalid and Mitra [11], we assume in Section 8.2 that, for practical reasons, it should be avoided to switch between a large number of different weights – in fact, we require that just one set of weights ϕ be used. Therefore, we consider the situation that the user population fluctuates just mildly around some 'operating point' $(\bar{n}_1, \bar{n}_2)$. We develop an algorithm to find a ϕ such that some 'ball' around $\bar{n}_1, \bar{n}_2$ is contained in $\mathcal{S}(\phi)$.

In Section 8.3 we take the opposite approach and allow *infinitely many* weight adaptations, and we compute the resulting admissible region $\mathcal{S} = \cup_\phi \mathcal{S}(\phi)$. Both Sections 8.2 and 8.3 require fast and straightforward approximations of the overflow probabilities in the GPS system. We start with these in Section 8.1.

8.1 Approximation of the overflow probabilities

In this section we develop an approximation for the overflow probabilities in both queues of the GPS system. Recall that for $i = 1, 2$, n_i is the (typically large) number of sources of type i. We denote the stationary buffer content in this GPS model with unequal number of sources by Q_i, the service rate by C, and the buffer threshold of queue i by B_i. Invoking Remark 2.2, the GPS model with $n_1 \neq n_2$ is equivalent to a GPS model with n sources in both classes, mean rates $(n_i/n)\mu_i$ and variance functions $(n_i/n)v_i(\cdot)$. We scale the buffer threshold and service rate with n such that $nb_i \equiv B_i$ and $nc \equiv C$. Now we can apply our earlier results, where we assumed both classes to consist of n sources.

To approximate the overflow probabilities, three regimes are distinguished as in Section 7. Again, we concentrate on the first queue; the second queue can be treated analogously. Define $\Delta_i(n_1, n_2) := -\log \mathbb{P}(Q_i \geq B_i)$. Then it holds that $\Delta_i(n_1, n_2) \equiv -\log \mathbb{P}(Q_{i,n} \geq nb_i)$, with its approximation given by $\bar{\Delta}_i(n_1, n_2) := nJ(b_i)$.

First define $\phi_2^o := n_2\mu_2/C$. Consider

$$\frac{1}{2} \inf_{t>0} \frac{(B_1 + (C - n_1\mu_1 - n_2\mu_2)t)^2}{n_1 v_1(t) + n_2 v_2(t)};$$

denote the minimizer by t^c, and define also

$$\phi_2^c := \sup_{s \in [0,t^c]} \left(\frac{\Gamma_2(t^c, s)}{C\, s\, v(t^c)} \right) (B_1 + (C - n_1\mu_1 - n_2\mu_2)t^c)\,.$$

(i) If $\phi_2 \in [0, \phi_2^o]$, then

$$\bar{\Delta}_1(n_1, n_2) = \frac{1}{2} \inf_{t>0} \frac{(B_1 + (\phi_1 C - n_1\mu_1)t)^2}{n_1 v_1(t)}.$$

(ii) If $\phi_2 \in (\phi_2^o, \phi_2^c)$, then

$$\bar{\Delta}_1(n_1, n_2) =$$

$$\frac{1}{2} \inf_{t>0} \sup_{s \in (0,t]} \begin{pmatrix} z_1(t, n_1, n_2) \\ z_2(s, n_1, n_2) \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} n_1 v_1(t) + n_2 v_2(t) & n_2 \Gamma_2(s,t) \\ n_2 \Gamma_2(s,t) & n_2 v_2(s) \end{pmatrix}^{-1} \begin{pmatrix} z_1(t, n_1, n_2) \\ z_2(s, n_1, n_2) \end{pmatrix},$$

where

$$\begin{pmatrix} z_1(t, n_1, n_2) \\ z_2(s, n_1, n_2) \end{pmatrix} := \begin{pmatrix} B_1 + (C - n_1\mu_1 - n_2\mu_2)t \\ (\phi_2 C - n_2\mu_2)s \end{pmatrix}.$$

(iii) If $\phi_2 \in [\phi_2^c, 1]$, then

$$\bar{\Delta}_1(n_1, n_2) = \frac{1}{2} \inf_{t>0} \frac{(B_1 + (C - n_1\mu_1 - n_2\mu_2)t)^2}{n_1 v_1(t) + n_2 v_2(t)}.$$

8.2 Weight setting algorithm

This subsection focuses on a procedure for finding weights (ϕ_1, ϕ_2) such that both classes receive the desired QoS, despite (mild) fluctuations in the number of sources present. More precisely, for specified (positive) numbers δ_i, we require that $\Delta_i(n_1, n_2) \geq \delta_i$ for all (n_1, n_2) in a 'ball' $\mathcal{B}(\bar{n}_1, \bar{n}_2)$ around $(\bar{n}_1, \bar{n}_2)$:

$$\mathcal{B}(\bar{n}_1, \bar{n}_2) := \left\{(n_1, n_2) \in \mathbb{N}^2 \mid \gamma_1(n_1 - \bar{n}_1)^2 + \gamma_2(n_2 - \bar{n}_2)^2 \leq 1\right\},$$

for positive γ_1, γ_2. It can be easily verified that the procedure described below works, in fact, for any 'target area' $\mathcal{B}$ that is finite and *convex*, rather than just these ellipsoidal sets.
To simplify our algorithm, we use the following expansion of $\bar{\Delta}_i(n_1, n_2)$ around $(n_1, n_2) = (\bar{n}_1, \bar{n}_2)$:

$$\begin{aligned} \bar{\Delta}_i(n_1, n_2) \quad \approx \quad & \bar{\Delta}_i(\bar{n}_1, \bar{n}_2) \; + \; (n_1 - \bar{n}_1) \frac{\partial \bar{\Delta}_i(n_1, n_2)}{\partial n_1}\bigg|_{(n_1,n_2)=(\bar{n}_1,\bar{n}_2)} \\ & + \; (n_2 - \bar{n}_2) \frac{\partial \bar{\Delta}_i(n_1, n_2)}{\partial n_2}\bigg|_{(n_1,n_2)=(\bar{n}_1,\bar{n}_2)}. \end{aligned} \tag{21}$$

This approximation requires the evaluation of two partial derivatives, which can be done relatively explicitly, as described in Appendix A.2.
Relying on (21), we have to verify whether for all $(n_1, n_2) \in \mathcal{B}(\bar{n}_1, \bar{n}_2)$ and $i = 1, 2$,

$$\bar{\Delta}_i(\bar{n}_1, \bar{n}_2) + (n_1 - \bar{n}_1, n_2 - \bar{n}_2)^{\mathrm{T}} e_i \geq \delta_i,$$

where

$$e_i \equiv (e_{i1}, e_{i2}) := \left(\frac{\partial \bar{\Delta}_i(n_1, n_2)}{\partial n_1}\bigg|_{(n_1,n_2)=(\bar{n}_1,\bar{n}_2)}, \frac{\partial \bar{\Delta}_i(n_1, n_2)}{\partial n_2}\bigg|_{(n_1,n_2)=(\bar{n}_1,\bar{n}_2)} \right).$$

Because of the convex shape of $\mathcal{B}(\bar{n}_1, \bar{n}_2)$, we only have to verify this condition for the two points on the boundary $\partial\mathcal{B}(\bar{n}_1, \bar{n}_2)$ having a tangent with slopes equal to $-e_{11}/e_{12}$ and $-e_{21}/e_{22}$ respectively. Denoting these points by $(n^\star_{11}, n^\star_{12})$ and $(n^\star_{21}, n^\star_{22})$, we have

$$(n^\star_{i1}, n^\star_{i2}) := \left(\bar{n}_1 + \sqrt{\left(\gamma_1 + \frac{e_{i2}^2}{e_{i1}^2}\,\frac{\gamma_1^2}{\gamma_2}\right)^{-1}}, \bar{n}_2 + \sqrt{\left(\gamma_2 + \frac{e_{i1}^2}{e_{i2}^2}\,\frac{\gamma_2^2}{\gamma_1}\right)^{-1}} \right),$$

$i = 1, 2$. We say that ϕ is feasible if $K_i := \bar{\Delta}_i(\bar{n}_1, \bar{n}_2) + (n^\star_{i1} - \bar{n}_1, n^\star_{i2} - \bar{n}_2)^{\mathrm{T}} e_i \geq \delta_i$ for both $i = 1$ and 2. Notice that K_i is a function of the weights; as $\phi_1 + \phi_2 = 1$, we can write $K_i(\phi_1)$. $K_1(\phi_1)$ will increase in ϕ_1, whereas $K_2(\phi_1)$ will decrease.
This suggests the following solution to the weight setting problem: (i) First find the smallest ϕ_1 such that $K_1(\phi_1) \geq \delta_1$. If this does not exist, then there is no solution. (ii) If it does

exist, then verify if for this ϕ_1 it holds that $K_2(\phi_1) \geq \delta_2$. If this is true, then the weight setting problem can be solved; if not, then there is no solution (i.e., there is no ϕ such that $\mathcal{B}(\bar{n}_1, \bar{n}_2) \subseteq \mathcal{S}(\phi)$).

Example 3. We first explain how requirements on the admissible numbers of sources naturally lead to a set of the type $\mathcal{B}(\bar{n}_1, \bar{n}_2)$.

- Our analysis assumes fixed numbers of sources of both types, but in practice this number fluctuates in time: sources arrive, and stay in the system for a random amount of time. Now suppose that sources of both types arrive according to Poisson processes (with rates ν_i, for $i = 1, 2$), and that, if admitted, these would require service for some random duration (with finite means $\mathbb{E}D_i$). If there were no admission control, the distributions of the number of jobs of both types are Poisson with means (and variances!) $\bar{n}_i = \nu_i \mathbb{E}D_i$.

- Suppose the system must be designed such that this mean $(\bar{n}_1, \bar{n}_2) \pm$ twice the standard deviation should be in the admissible region, i.e., should be contained in $\mathcal{S}(\phi)$. This suggests choosing

$$\mathcal{B}(\bar{n}_1, \bar{n}_2) = \left\{ (n_1, n_2) \in \mathbb{N}^2 \;\middle|\; \left(\frac{n_1 - \bar{n}_1}{2\sqrt{\bar{n}_1}} \right)^2 + \left(\frac{n_2 - \bar{n}_2}{2\sqrt{\bar{n}_2}} \right)^2 \leq 1 \right\}.$$

In this example we choose $\bar{n}_1 = 900$ and $\bar{n}_2 = 1600$, which leads to:

$$\mathcal{B}(\bar{n}_1, \bar{n}_2) = \mathcal{B}(900, 1600) = \left\{ (n_1, n_2) \in \mathbb{N}^2 \mid 16(n_1 - 900)^2 + 9(n_2 - 1600)^2 \leq 57600 \right\}.$$

We suppose that both types of sources correspond to Brownian motions, with $\mu_1 = 0.2$, $\mu_2 = 0.3$, $v_1(t) = 2t$, and $v_2(t) = t$. We rely on explicit results for Brownian motions, as summarized in Appendix A.2, in particular for the partial derivatives of the $\bar{\Delta}_i(n_1, n_2)$ with respect to the numbers of sources. We choose $C = 1000$, $B_1 = 35$, and $B_2 = 25$.
First suppose the performance targets are $\delta_1 = 9$ and $\delta_2 = 7$ (roughly corresponding to overflow probabilities $1.2 \cdot 10^{-4}$ and $9.1 \cdot 10^{-4}$). Figure 2 shows that no weights ϕ exist to meet this target (to guarantee that the overflow probability in queue 1 is small enough, ϕ_1 should be larger than 0.39, but this implies that $K_2(\phi_1) < 5.7 < \delta_2$). Now suppose that $\delta_1 = 8$ and $\delta_2 = 6$. Then an analogous reasoning gives that ϕ_1 should be chosen in the interval $(0.34, 0.37)$.

8.3 Admissible region

While above we restricted ourselves to just one set of weights, we might allow to switch weights whenever necessary. Clearly, the resulting admissible region can be obtained as the union of the admissible regions for fixed weights.

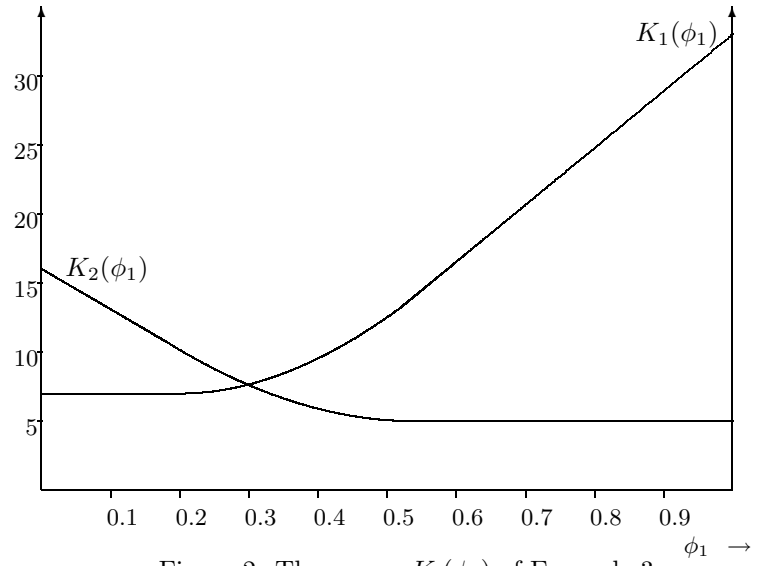

Figure 2: The curves $K_i(\phi_1)$ of Example 3.

Example 4. In Figure 3 we have computed the admissible region for the same types of sources as in the previous subsection, with performance targets $\delta_1 = 5$, $\delta_2 = 7$. We have not succeeded in finding explicit expressions for the boundary of the admissible region.

9 Concluding remarks

We have considered a two-class GPS system with Gaussian inputs in the many-sources regime. We have focused on the asymptotic decay rate of the buffer overflow probability, as function of the number of sources.

We have found the exact value of the decay rate in case one of the classes generates on average more than its guaranteed rate. The opposite case turned out to be considerably harder; there we have developed upper and lower bounds on the decay rate. These appear to be tight under fairly general conditions, as corroborated by extensive numerical experiments, as well as explicit calculations for the special case of Brownian motion sources, and further justified by heuristic arguments. Explicitly finding these conditions, however, remains a challenging problem. The asymptotic results directly lead to approximations for the overflow probability, which we have used to develop weight-setting procedures.

Future research directions include: (1) The results of Section 3 applying to arbitrary sources, it can be expected that our main results hold for more general traffic classes than just Gaussian.

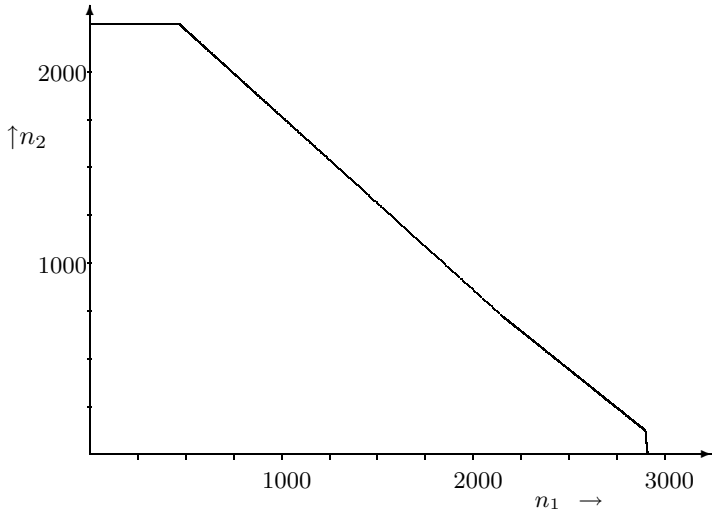

Figure 3: Admissible region of Example 4.

Our analysis depends heavily on the availability of a sample-path LDP ('Schilder'), which suggests the examination of other traffic processes for which such an LDP is known – for instance exponential on-off sources, see [27]. (2) Another possible extension could be GPS systems with more than just two classes. (3) Our results suggest that, under general conditions, the upper and lower bounds, as derived in this paper, coincide. Further analysis is needed to determine the (minimal) conditions under which they match. (4) Section 8 provides a procedure for finding a weight vector ϕ such that some (finite, convex) 'target area' $\mathcal{B}$ is fully contained in the admissible region $\mathcal{S}(\phi)$. Suppose that it is not possible to find a *single* weight such that $\mathcal{B} \subset \mathcal{S}(\phi)$ – for instance because the target area is relatively large – but suppose that $\mathcal{B} \subset \cup_\phi \mathcal{S}(\phi)$. In this case it is necessary to switch between weights to cover $\mathcal{B}$. Now a relevant question is: how to choose a collection of weigths $\phi^{\mathrm{I}}, \phi^{\mathrm{II}}, \ldots$, such that the GPS scheduler has to switch weights as infrequently as possible.

References

[1] R. Addie, P. Mannersalo and I. Norros. Most probable paths and performance formulae for buffers with Gaussian input traffic. *European Transactions on Telecommunications*, 13: 183 – 196, 2002.

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

[3] R. Bahadur and S. Zabell. Large deviations of the sample mean in general vector spaces. *Annals of Probability*, 7: 587 – 621, 1979.

[4] S. Borst, M. Mandjes, and M. van Uitert. Generalized Processor Sharing queues with heterogeneous traffic classes. *IEEE/ACM Transactions on Networking*, to appear, 2003.

[5] 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.

[6] C. Courcoubetis and R. Weber. Buffer overflow asymptotics for a buffer handling many traffic sources. *Journal of Applied Probability,* 33: 886 – 903, 1996.

[7] K. Dębicki and M. Mandjes. Exact overflow asymptotics for queues with many Gaussian inputs. CWI report PNA-R0209. *Journal of Applied Probability*, 40, to appear, 2003. Available at URL `http://www.cwi.nl/ftp/CWIreports/PNA/PNA-R0209.pdf`.

[8] K. Dębicki and M. Mandjes. Traffic with an fBm limit: convergence of the workload process. CWI report PNA-R0220. *Queueing Systems*, to appear, 2003. Available at URL `http://www.cwi.nl/ftp/CWIreports/PNA/PNA-R0220.pdf`.

[9] A. Dembo and O. Zeitouni. Large deviations techniques and applications. Jones and Bartlett Publishers, Boston, USA, 1993.

[10] J.-D. Deuschel and D. Stroock. Large Deviations. Academic Press, London, 1989.

[11] A. Elwalid and D. Mitra. Design of Generalized Processor Sharing schedulers which statistically multiplex heterogeneous QoS classes. *Proceedings IEEE Infocom*, New York, USA, 1220 – 1230, 1999.

[12] J. Kilpi and I. Norros. Testing the Gaussian approximation of aggregate traffic. *Proceedings Internet Measurement Workshop*, Marseille, France, 2002. Available at URL `http://www.vtt.fi/tte/rd/traffic-theory/papers/`

[13] K. Kumaran, G.E. Margrave, D. Mitra and K.R. Stanley. Novel techniques for the design and control of Generalized Processor Sharing schedulers for multiple QoS classes. *Proceedings IEEE Infocom,* Tel Aviv, Israel, 932 – 941, 2000

[14] 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.

[15] M. Mandjes. A note on the benefits of buffering. Submitted, 2002.

[16] M. Mandjes and M. van Uitert. Large deviations for complex buffer architectures: the short-range dependent case. Submitted.

[17] M. Mandjes and M. van Uitert. Sample-path large deviations for tandem and priority queues with Gaussian inputs. CWI report PNA-R0221. Submitted, 2002. Available at URL `http://www.cwi.nl/ftp/CWIreports/PNA/PNA- R0221.pdf`.

[18] P. Mannersalo and I. Norros. Approximate formulae for Gaussian priority queues. *Proceedings ITC 17*, Salvador da Bahia, Brazil, 991 – 1002, 2001.

[19] P. Mannersalo and I. Norros. GPS schedulers and Gaussian traffic. *Proceedings IEEE Infocom*, New York, USA, 1660 – 1667, 2002.

[20] P. Mannersalo and I. Norros. A most probable path approach to queueing systems with general Gaussian input. *Computer Networks*, 40: 399 – 412, 2002.

[21] I. Norros. Most probable path techniques for Gaussian queueing systems. *Proceedings of 2nd International IFIP-TC6 Networking Conference,* Pisa, Italy. Springer, Berlin, 86 – 104.

[22] A. Parekh and R. Gallager. A generalized processor sharing approach to flow control in integrated services networks: The single node case. *IEEE/ACM Transactions on Networking*, 1: 344 – 357, 1993.

[23] A. Parekh and R. Gallager. A generalized processor sharing approach to flow control in integrated services networks: The multiple node case. *IEEE/ACM Transactions on Networking*, 2: 137 – 150, 1994.

[24] E. Reich. On the integrodifferential equation of Takács I. *Annals of Mathematical Statistics*, 29: 563–570, 1958.

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

[26] M. van Uitert and S. Borst. A reduced-load equivalence for Generalised Processor Sharing networks with long-tailed input flows. *Queueing Systems*, 41: 123 – 163, 2002.

[27] A. Weiss. A new technique for analyzing large traffic systems. *Advances in Applied Probability*, 18: 506 – 532, 1986.

[28] Z.-L. Zhang. Large deviations and the processor sharing scheduling for a two-queue system. *Queueing Systems,* 26: 229 – 264, 1997.

[29] Z.-L. Zhang, Z. Liu, J. Kurose and D. Towsley. Call admission control schemes under Generalized Processor Sharing scheduling. *Telecommunication Systems*, 7: 125 – 152, 1997.

A Appendix

A.1 Analysis of underload regime with large ϕ_2

This appendix focuses on the underload regime with large ϕ_2. By deriving the counterpart for $J^L(b,x)$ of Theorem 5.8, we can prove that for a specific range of ϕ_2 the derived upper and lower bounds match. We first introduce some new notation:

$$\bar{k}_2(x,s,t,u) := \mathbb{E}[A_2(-s,0) + A_1(-s,-u) \mid A_1(-t,0) + A_2(-t,0) = x + b + ct],$$

$$\phi_2^{c,L}(x) := \sup_{s\in(0,t^c(x)]} \inf_{u\in[0,s)} \frac{\bar{k}_2(x,s,t^c(x),u) - x + c(u-s)}{cu}.$$

Lemma A.1 *For all $x \geq 0$, it holds that $\phi_2^{c,L}(x) \leq \phi_2^{c,U}(x)$.*

Proof. Notice that $\bar{k}_2(x,s,t,u)$ and $k_2(x,s,t)$ coincide for $u=s$. Then the stated follows directly from the definitions of $\phi_2^{c,L}(x)$ and $\phi_2^{c,U}(x)$. □

The counterpart of Theorem 5.8 follows directly now.

Lemma A.2 *If $\phi_2 \geq \phi_2^{c,L}(x)$, then*

$$J^L(b,x) = \frac{(x+b+(c-\mu)t^c)^2}{2v(t^c)}.$$

This leads to the following result.

Theorem A.3 *If $\phi_2 \in [\sup_{x\geq 0}\phi_2^{c,U}(x), 1]$, then*

$$-\lim_{n\to\infty}\frac{1}{n}\log\mathbb{P}(Q_{1,n}(0)\geq nb) = \inf_{t>0}\frac{(b+(c-\mu)t)^2}{2v(t)}.$$

Proof. Due to Lemma A.1, if $\phi_2 \geq \sup_{x\geq 0}\phi_2^{c,U}(x)$, then also $\phi_2 \geq \phi_2^{c,L}(x)$ for all $x\geq 0$. Now the stated follows directly from the fact that, *for all* $x\geq 0$, Theorem 5.8 and Lemma A.2 apply. Hence the infima over t and x can be interchanged, and the result follows. □

Our numerical experiments suggest that $\sup_{x\geq 0}\phi_2^{c,U}(x) = \phi_2^{c,U}(0)$. The following shows that this property holds under a sufficient condition that can be verified (relatively) easily. Denote by $s^c(x)$ the optimizing $s\in(0,t^c(x)]$ in (18).

Lemma A.4 *If, for all $x\geq 0$,*

$$\frac{v'(t^c(x))}{v(t^c(x))}\,\Gamma_2(s^c(x),t^c(x)) \geq 2\frac{\partial\Gamma_2(s,t)}{\partial t}\bigg|_{s=s^c(x),t=t^c(x)}, \tag{22}$$

then $\sup_{x\geq 0}\phi_2^{c,U}(x) = \phi_2^{c,U}(0)$.

Proof. We prove that $\phi_2^{c,U}(\cdot)$ is decreasing under condition (22). For brevity write

$$\frac{\partial k_2}{\partial s} \equiv \frac{\partial k_2(x,s,t)}{\partial s}\bigg|_{s=s^c(x),t=t^c(x)} \qquad \frac{\partial k_2}{\partial t} \equiv \frac{\partial k_2(x,s,t)}{\partial t}\bigg|_{s=s^c(x),t=t^c(x)}.$$

Notice that $t^c(x)$ and $s^c(x)$ satisfy

$$\frac{x+b+(c-\mu)t^c(x)}{2(c-\mu)} = \frac{v(t^c(x))}{v'(t^c(x))}, \qquad s^c(x)\,\frac{\partial k_2}{\partial s} = k_2(x,s^c(x),t^c(x)) - x. \tag{23}$$

It is easy to check that the derivative of $\phi_2^{c,U}(\cdot)$ is non-positive if

$$s^c(x)\left(\frac{\partial k_2}{\partial s}\frac{\mathrm{d}s^c(x)}{\mathrm{d}x} + \frac{\partial k_2}{\partial t}\frac{\mathrm{d}t^c(x)}{\mathrm{d}x} + \frac{\partial k_2}{\partial x} - 1\right) - \frac{\mathrm{d}s^c(x)}{\mathrm{d}x}\left(k_2(x,s^c(x),t^c(x)) - x\right) \leq 0.$$

Notice that because of the second equation in (23) various terms cancel out. Now due to

$$\frac{\partial k_2}{\partial x} = \frac{\Gamma_2(s,t)}{v(t)} \leq \frac{\Gamma_2(s,t)}{v_2(t)} \leq \frac{\Gamma_2(s,t)}{\sqrt{v_2(s)v_2(t)}} \leq 1,$$

(apply Assumption A2), and $\mathrm{d}t^c(x)/\mathrm{d}x \geq 0$ (see Lemma 3.1 in [15]), it is left to check that $\partial k_2/\partial t \leq 0$. It is a matter of straightforward calculus, using the first equation in (23), to show that this is equivalent to (22). □

A.2 Weight setting algorithm: partial derivatives

In this appendix, we determine expressions for the partial derivatives of $\bar{\Delta}_1(n_1, n_2)$ to the numbers of sources, as required in the weight setting algorithm of Section 8.2. The Cases (i), (ii), (iii) below correspond to the regimes identified in Section 8.1.

(i) Based on Theorem 2.3, in the regime $\phi_2 \in [0, \phi_2^c]$,

$$\bar{\Delta}_1(n_1, n_2) = \inf_{t>0} \sup_{\theta \in \mathbb{R}} \left(\theta(B_1 + (\phi_1 C - n_1\mu_1)t) - \frac{1}{2}\theta^2 n_1 v_1(t) \right).$$

The inner supremum is attained for

$$\theta^\star = \frac{B_1 + (\phi_1 C - n_1\mu_1)t}{n_1 v_1(t)}.$$

Denoting the optimizing t by $t^\star$, we derive

$$\frac{\partial \bar{\Delta}_1(n_1, n_2)}{\partial n_1} = -\theta^\star \mu_1 t^\star - \frac{1}{2}(\theta^\star)^2 v_1(t^\star), \quad \frac{\partial \bar{\Delta}_1(n_1, n_2)}{\partial n_2} = 0.$$

(ii) Similarly, in the regime $\phi_2 \in [\phi_2^c, \phi_2^o]$, $\bar{\Delta}_1(n_1, n_2)$ can be rewritten as

$$\inf_{t>0} \sup_{s\in(0,t]} \sup_{\theta \in \mathbb{R}^2} \left(\theta^{\mathrm{T}} \begin{pmatrix} z_1(t, n_1, n_2) \\ z_2(s, n_1, n_2) \end{pmatrix} - \frac{1}{2}\theta^{\mathrm{T}} \begin{pmatrix} n_1 v_1(t) + n_2 v_2(t) & n_2\Gamma_2(s,t) \\ n_2\Gamma_2(s,t) & n_2 v_2(s) \end{pmatrix} \theta \right).$$

The optimizing θ is given by

$$\theta^\star = \begin{pmatrix} n_1 v_1(t) + n_2 v_2(t) & n_2\Gamma_2(s,t) \\ n_2\Gamma_2(s,t) & n_2 v_2(s) \end{pmatrix}^{-1} \begin{pmatrix} z_1(t, n_1, n_2) \\ z_2(s, n_1, n_2) \end{pmatrix}.$$

Straightforward computations give that, with the optimizing s, t denoted by $s^\star, t^\star$,

$$\frac{\partial \bar{\Delta}_1(n_1, n_2)}{\partial n_1} = -\theta_1^\star \mu_1 t^\star - \frac{1}{2}(\theta_1^\star)^2 v_1(t^\star),$$

$$\frac{\partial \bar{\Delta}_1(n_1, n_2)}{\partial n_2} = -\theta_1^\star \mu_2 t^\star - \theta_2^\star \mu_2 s^\star - \frac{1}{2}\theta^{\star\mathrm{T}} \begin{pmatrix} v_2(t^\star) & \Gamma_2(s^\star, t^\star) \\ \Gamma_2(s^\star, t^\star) & v_2(s^\star) \end{pmatrix} \theta^\star.$$

(iii) In the third regime $\phi_2 \in [\phi_2^o, 1]$,

$$\bar{\Delta}_1(n_1, n_2) = \inf_{t>0} \sup_{\theta \in \mathbb{R}} \left(\theta z_1(t, n_1, n_2) - \frac{1}{2}\theta^2 (n_1 v_1(t) + n_2 v_2(t)) \right).$$

The inner supremum is attained for

$$\theta^\star = \frac{z_1(t, n_1, n_2)}{n_1 v_1(t) + n_2 v_2(t)}.$$

Denoting the optimizing t by $t^\star$, we derive

$$\frac{\partial \bar{\Delta}_1(n_1, n_2)}{\partial n_1} = -\theta^\star \mu_1 t^\star - \frac{1}{2}(\theta^\star)^2 v_1(t^\star), \quad \frac{\partial \bar{\Delta}_1(n_1, n_2)}{\partial n_2} = -\theta^\star \mu_2 t^\star - \frac{1}{2}(\theta^\star)^2 v_2(t^\star).$$

Now we consider the special case that both types of sources correspond to Brownian motions. We assume $v_1(t) = \lambda_1 t$, $v_2(t) = \lambda_2 t$. We again consider the three regimes separately. We have explicit formulae for the 'critical' values of ϕ_2:

$$\phi_2^c = 1 - \frac{n_1\lambda_1 - n_2\lambda_2}{n_1\lambda_1 + n_2\lambda_2}\left(1 - \frac{n_1\mu_1 + n_2\mu_2}{C}\right) - \frac{n_1\mu_1}{C}; \quad \phi_2^o = \frac{n_2\mu_2}{C}.$$

(i) In this case

$$t^\star = \frac{B_1}{\phi_1 C - n_1\mu_1}; \quad \bar{\Delta}_1(n) = 2\,\frac{\phi_1 C - n_1\mu_1}{n_1\lambda_1}\,B_1.$$

This yields:

$$\frac{\partial\bar{\Delta}_1}{\partial n_1} = -2B_1\,\frac{\phi_1 C}{n_1^2\lambda_1}; \quad \frac{\partial\bar{\Delta}_1}{\partial n_2} = 0.$$

(ii) In this case

$$t^\star = B_1 \Bigg/ \sqrt{(\phi_1 C - n_1\mu_1)^2 + (\phi_2 C - n_2\mu_2)^2\frac{n_1\lambda_1}{n_2\lambda_2}}\;;$$

$$\bar{\Delta}_1(n_1, n_2) = \frac{1}{2}\left(\frac{(B_1 + (\phi_1 C - n_1\mu_1)t^\star)^2}{n_1\lambda_1 t^\star} + \frac{(\phi_2 C - n_2\mu_2))^2}{n_2\lambda_2}\,t^\star\right).$$

Also $s^\star = t^\star$. This yields:

$$\frac{\partial\bar{\Delta}_1}{\partial n_1} = -(B_1 + (\phi_1 C - n_1\mu_1)t^\star)\,\frac{\mu_1}{n_1\lambda_1} - \frac{1}{2}\frac{(B_1 + (\phi_1 C - n_1\mu_1)t^\star)^2}{n_1^2\lambda_1 t^\star};$$

$$\frac{\partial\bar{\Delta}_1}{\partial n_2} = -(\phi_2 C - n_2\mu_2)t^\star\,\frac{\mu_2}{n_2\lambda_2} - \frac{1}{2}\frac{(\phi_2 C - n_2\mu_2))^2}{n_2^2\lambda_2}\,t^\star.$$

(iii) In this case

$$t^\star = \frac{B_1}{C - n_1\mu_1 - n_2\mu_2}; \quad \bar{\Delta}_1(n_1, n_2) = 2\,\frac{C - n_1\mu_1 - n_2\mu_2}{n_1\lambda_1 + n_2\lambda_2}\,B_1.$$

This yields:

$$\frac{\partial\bar{\Delta}_1}{\partial n_1} = -\frac{2B_1\mu_1}{n_1\lambda_1 + n_2\lambda_2} - 2B_1\lambda_1\,\frac{C - n_1\mu_1 - n_2\mu_2}{(n_1\lambda_1 + n_2\lambda_2)^2};$$

$$\frac{\partial\bar{\Delta}_1}{\partial n_2} = -\frac{2B_1\mu_2}{n_1\lambda_1 + n_2\lambda_2} - 2B_1\lambda_2\,\frac{C - n_1\mu_1 - n_2\mu_2}{(n_1\lambda_1 + n_2\lambda_2)^2}.$$