

LARGE DEVIATION PROPERTIES OF CONSTANT RATE DATA STREAMS SHARING A BUFFER WITH LONG-RANGE DEPENDENT TRAFFIC IN CRITICAL LOADING

KURT MAJEWSKI,* *Siemens AG*

Abstract

We consider a constant rate traffic which shares a buffer with a random cross traffic. A first come first served or priority service discipline is applied at the buffer. After service at the first buffer the constant rate traffic moves to a play-out buffer. Both buffers provide service at constant rate and infinite waiting room. We investigate logarithmic large and moderate deviation asymptotics for the tail probabilities of the steady-state queue length distribution at the play-out buffer for long-range dependent cross traffic in critical loading. We characterize the asymptotic behavior of the cross traffic which leads to a large queue length at the play-out buffer and compare it to the one for renewal cross traffic.

Keywords: Moderate deviations; first come first served; priority service discipline; heavy traffic; queueing network

2000 Mathematics Subject Classification: Primary 60F10

Secondary 60K25; 90B15; 68M20

1. Introduction

Up to now, only a few explicit analytical results have been obtained for the behavior of queueing systems with more than one buffer *and* long-range dependent random traffic. Here processes are called *long-range dependent* when the auto-correlation function of their increments is not integrable [4, p. 4], [6].

Heavy traffic convergence in distribution [16], [21] constitutes a general framework in which such an analysis is reduced to an investigation of the properties of reflected multi-dimensional fractional Brownian motion when the queues approach critical load. We call a process *reflected* if it is the image of a process under a Skorokhod map which restricts the movements of the process to the positive orthant by shifting it at the border [9], [14]. (A definition of the one-dimensional Skorokhod map is contained in Section 2.) However, again, only a few explicit results about the behavior of multi-dimensional reflected fractional Brownian motion have been obtained so far. Most of them can be classified as logarithmic asymptotics for certain tail probabilities of their distribution [20], [26].

To some extent these two observations (which also partially apply to a single queue and short-range dependent traffic) have motivated a large number of publications which address the analysis of queueing systems by investigating large deviations of reflected Gaussian processes [1], [20], [26], [28], [31], [32]. In such models the Gaussian input approximates statistical properties of the traffic processes, and the Skorokhod map captures the response of the queueing system to this input. However, the use of Gaussian input requires additional justification, and

Received 5 December 2005; revision received 27 September 2006.

* Postal address: Siemens AG, CT SE 6, 81730 München, Germany.

the class of queueing networks which can be directly modeled through a Skorokhod map is rather restricted.

In contrast to this, moderate deviations in critical loading constitute a direct approach to expressing logarithmic asymptotics of tail probabilities of queueing systems through those of reflected Gaussian processes. We refer to *moderate deviations* when the scaling is situated between an ordinary large deviation principle and a central limit theorem, such that the rate function of a Gaussian process appears; see [7], [30], [36]. This connection has been established for single class queueing networks (i.e. generalized Jackson) in [36] and for multi-class feedforward queueing networks in [24]. In [25] and [36] a moderate deviation principle for the input processes is combined with a heavy traffic convergence. The result is the characterization of logarithmic tail asymptotics for the behavior of queueing networks in critical loading. The rate function in the examples of [24] and [36] can be identified with the rate function for the tail probabilities of a (multi-dimensional) reflected Brownian motion, for which numerous analytical results are available [3], [10], [13], [19]. This approach takes advantage of both the simplified heavy traffic model and the explicit rate function of a Gaussian process. It can be used to rigorously justify the calculation of large deviations of reflected Gaussian processes as a means of analyzing moderate deviations of queueing networks in critical loading. However, it has not yet been applied to queueing networks with long-range dependent input. In this work we exemplify how this can be done and find, surprisingly, that with long-range dependent input we can obtain both *large* and *moderate* deviation principles in critical loading.

Our example network consists of two queueing nodes with constant service rates and infinite waiting rooms. A constant rate fluid stream must visit both nodes. It shares the service and buffer capacity of the first node with a random cross traffic.

For cross traffic, we discuss a long-range dependent superposition of fluid sources with exponential inter start times and Pareto distributed durations, and, for the purpose of comparison, short-range dependent renewal traffic. Such cross traffic processes can satisfy sample path large and moderate deviation principles when they are centered and suitably scaled [22], [23], [34], [35]. The rate functions of these moderate deviation principles are well known from large deviations of fractional Brownian motion in the long-range dependent case and of Brownian motion in the renewal cross traffic case.

We use the framework of [21] and [24] to characterize the steady-state buffer contents at the second node. This queue length distribution is a measure of the cell delay variation induced on the constant rate stream by sharing the first buffer with random cross traffic. The results of [24] yield a sample path large deviation principle for the queue length processes when the sequence of centered arrival processes satisfies a large deviation principle, and the queues approach critical load with a certain speed.

In the main result of this work we characterize moderate and large deviation asymptotics for the steady-state queue length distribution at the second node by solving two-dimensional minimization problems. This result makes it possible to expose the transition of moderate deviations for the behavior of the network from short-range to long-range dependent input, and to reveal the distinction between large and moderate deviations for the behavior of the queueing network with long-range dependent input. Figures containing logarithmic decay rates visualize the findings for first come first served (FCFS) and priority service disciplines at the first node. Furthermore, we calculate certain minimizing paths which display the 'most likely' asymptotic behavior of the cross traffic and the queue length processes conditioned to a large queue length at the second buffer.

Large deviation asymptotics for this queueing system with short-range dependent input processes and FCFS service discipline can also be found in [37]. A scaling is applied in [37], in which time and state are scaled with the same increasing factor. The traffic intensity at the nodes remains constant and away from a critical load. In contrast to this, we apply a scaling where time is scaled faster than state and a heavy traffic condition compensates for this difference. Similar remarks apply to the work of [5], which focused on large deviations for queueing systems with a priority service discipline.

We refer the reader to [12] for further references and background on large and moderate deviations for queueing systems with short- and long-range dependent input. Furthermore, the ‘infinite intersections approach’ in [27] could be of use in identifying moderate deviation rates and minimizing paths for variants of the example network considered here.

An overview of this work is as follows. We recall necessary basic definitions and facts in Section 2. We introduce a sequence of networks with two queues approaching heavy traffic in Section 3. We characterize large deviation rates and minimizing paths leading to a large queue length in the second buffer in Section 4. Three sequences of cross traffic processes satisfying the assumptions are presented in Section 5. We calculate and display large deviation rates and minimizing paths for these sequences in Section 6. Appendix A contains the proof of our main result.

2. Preliminaries

We first recall the following elements of large deviations theory for which [8] is a standard reference. A $[0, \infty]$ -valued lower semicontinuous function on a topological space is called a *rate function*. A rate function is *good* if it has compact level sets. A sequence $(X_k)_{k \in \mathbb{N}}$ of random elements with values in a measurable space (E, \mathcal{E}) satisfies a *large deviation principle* with normalizing sequence $(b_k)_{k \in \mathbb{N}}$ and rate function I in the topology \mathcal{T} on E if $\lim_{k \rightarrow \infty} b_k = \infty$ and, for every measurable set $A \in \mathcal{E}$,

$$\limsup_{k \rightarrow \infty} \frac{1}{b_k} \log P(X_k \in A) \leq - \inf_{x \in A^c} I(x), \tag{1}$$

$$\text{and } \liminf_{k \rightarrow \infty} \frac{1}{b_k} \log P(X_k \in A) \geq - \inf_{x \in A^o} I(x), \tag{2}$$

where A^c (A^o) is the closure (interior) of A in the topology \mathcal{T} , and P is the underlying probability measure. This work deals with large deviation principles for the sample paths of random processes. The underlying function spaces are therefore called *path spaces* and are introduced next.

We let \mathcal{D} be the set of *paths* $d: \mathbb{R} \rightarrow \mathbb{R}$ which are right continuous, possess left-hand limits and have limits

$$\underline{d} := \lim_{t \rightarrow -\infty} \frac{d(t)}{t} \quad \text{and} \quad \bar{d} := \lim_{t \rightarrow \infty} \frac{d(t)}{t},$$

in \mathbb{R} . (On product path spaces these limits are taken componentwise.) With \mathcal{I} we denote the subset of \mathcal{D} consisting of nondecreasing (continuous) paths. With \mathcal{C} we denote the subset of \mathcal{D} consisting of continuous paths. We provide these (and further) path spaces with the σ -algebras generated by the family of one-dimensional projections and topology induced by the norm $\|\cdot\|$, given by

$$\|d\| := \sup_{t \in \mathbb{R}} \frac{|d(t)|}{1 + |t|}.$$

Product path spaces are equipped with corresponding product σ -algebras and topologies. With pointwise and componentwise addition and scalar multiplication, product path spaces formed from \mathcal{D} become topological vector spaces. We use bold-face symbols for elements of, or mappings into, path spaces.

We let \mathcal{D}_0 be the subset of \mathcal{D} containing all paths $\mathbf{d} \in \mathcal{D}$ satisfying $\underline{\mathbf{d}} = \overline{\mathbf{d}} = 0$. Clearly, if a path $\mathbf{d} \in \mathcal{D}$ satisfies $\underline{\mathbf{d}} = \overline{\mathbf{d}} =: \delta$, then the *centered* path $\mathbf{d} - \delta \mathbf{id}$ is an element of \mathcal{D}_0 , where $\mathbf{id}: \mathbb{R} \rightarrow \mathbb{R}, t \mapsto t$ denotes the identity map of \mathbb{R} .

For $\tau \in \mathbb{R}$, we define the (time-)shift $\Theta_\tau: \mathcal{D} \rightarrow \mathcal{D}$ by $(\Theta_\tau \mathbf{d})(t) := \mathbf{d}(t + \tau)$ for every $t \in \mathbb{R}$. We say that a process \mathbf{x} on \mathcal{D} is *stationary*, when the distribution of $\Theta_\tau \mathbf{x}$ is the same for every $\tau \in \mathbb{R}$. We say that a process \mathbf{x} on \mathcal{D} has *stationary increments* when the distribution of $\Theta_\tau \mathbf{x} - \mathbf{x}(\tau)$ is the same for every $\tau \in \mathbb{R}$. We define, for $\tau > 0$, the scaling mapping $\Gamma_\tau: \mathcal{D} \rightarrow \mathcal{D}$ by $(\Gamma_\tau \mathbf{d})(t) := \mathbf{d}(\tau t)/\tau$ for $t \in \mathbb{R}$.

On the set $\mathcal{D}_{\text{sup}} := \{\mathbf{d} \in \mathcal{D} : \underline{\mathbf{d}} > 0\}$ we consider the ‘running supremum’ $\mathbf{sup}: \mathcal{D}_{\text{sup}} \rightarrow \mathcal{D}_{\text{sup}} \cap \mathcal{I}$ defined for $\mathbf{d} \in \mathcal{D}_{\text{sup}}$ and $t \in \mathbb{R}$ by

$$(\mathbf{sup} \mathbf{d})(t) := \sup_{\tau \in]-\infty, t]} \mathbf{d}(\tau).$$

This mapping is continuous (see [11]), and satisfies $\mathbf{sup} \mathbf{d} = \underline{\mathbf{d}}$ and $\overline{\mathbf{sup} \mathbf{d}} = \max\{0, \overline{\mathbf{d}}\}$. (For a finite subset \mathcal{K} of \mathbb{R} , we let $\max \mathcal{K}$ denote the maximum of its elements.)

For a path $\mathbf{d} \in \mathcal{D}$ with $\underline{\mathbf{d}} < 0$, the path $\mathbf{w} := \mathbf{d} + \mathbf{sup}(-\mathbf{d})$ is the unique path in \mathcal{D}_0 for which there exists a nondecreasing path $\mathbf{y} \in \mathcal{I}$ such that $\mathbf{w} = \mathbf{d} + \mathbf{y} \geq 0$ and $\int_{\mathbb{R}} \mathbf{w}(t) \, d\mathbf{y}(t) = 0$; see Lemma 5 in [20]. (Inequalities with paths are satisfied if they hold for every argument.) The backlogged workload $\mathbf{w} \in \mathcal{D}_0$ and cumulative idle time $\mathbf{y} \in \mathcal{I}$ at a nonidling queueing node naturally satisfy these two properties, namely, the workload remains nonnegative and the cumulative idle time increases only during time intervals in which the workload is zero. Therefore, a (one-dimensional) *Skorokhod* or *reflection* map of the form $\mathbf{d} \mapsto \mathbf{d} + \mathbf{sup}(-\mathbf{d})$ can often be found in the definition of the behavior of a queueing system or its heavy traffic approximation; compare (3), (4), (6), (10), and (11) below. The image of a process under the Skorokhod map is frequently called (somewhat imprecisely) a *reflected* process: it is constrained to remain nonnegative, but is pushed upwards only at zero.

3. Two queues approaching heavy traffic

We consider the queueing system depicted in Figure 1. It possesses two queueing nodes $N := \{\text{switch}, \text{bucket}\}$ with infinite buffer capacities and constant service rates. It is populated with customers of two types $C := \{\text{long}, \text{cross}\}$. Customers of type *long* (= longitudinal) must queue up for service at node *switch* and subsequently at node *bucket*. Customers of type *cross* must visit only node *switch* before leaving the system. We are interested in the impact of sharing queue *switch* with the random cross-traffic consisting of type *cross* customers on the traffic of type *long* customers. In order to assess this impact we calculate large (and moderate) deviation asymptotics for large queue lengths at queue *bucket*. We distinguish two different service disciplines $S := \{\text{FCFS}, \text{PRIO}\}$ at the first node. In case *FCFS* a *FCFS* service discipline is applied at the node *switch*. In case *PRIO* customers of type *cross* have high priority and customers of type *long* have low priority at node *switch*.

We let \mathbf{a}_{long} and $\mathbf{a}_{\text{cross}}$ be two independent random processes on \mathcal{I} with stationary increments, which model the cumulative number of arrivals of type *long* and *cross* customers

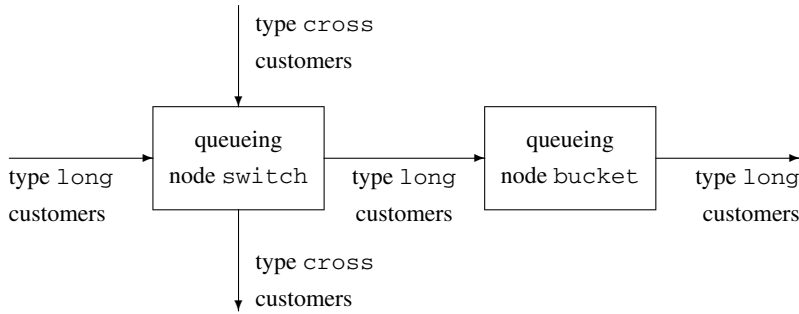


FIGURE 1: Queuing network with two nodes.

respectively, at node `switch` as a function of time. We assume that there are arrival rates $\alpha_{\text{long}} > 0$ and $\alpha_{\text{cross}} > 0$, such that

$$\alpha_{\text{long}} = \underline{a}_{\text{long}} = \overline{a}_{\text{long}} \quad \text{and} \quad \alpha_{\text{cross}} = \underline{a}_{\text{cross}} = \overline{a}_{\text{cross}}$$

have probability 1. We let $\sigma^{\text{switch}} > 0$ and $\sigma^{\text{bucket}} > 0$ be the constant service rate at the first and second queue respectively, and impose the stability conditions

$$\alpha_{\text{long}} + \alpha_{\text{cross}} < \sigma^{\text{switch}} \quad \text{and} \quad \alpha_{\text{long}} < \sigma^{\text{bucket}}.$$

For the FCFS service discipline, we define the workload process (the amount of unfinished work measured in units of time) at the first queue by

$$w^{\text{switch,FCFS}} := \frac{a_{\text{long}} + a_{\text{cross}}}{\sigma^{\text{switch}}} - \text{id} + \sup \left(\text{id} - \frac{a_{\text{long}} + a_{\text{cross}}}{\sigma^{\text{switch}}} \right). \tag{3}$$

This standard construction of the workload process of a nonidling queueing node on the entire time interval \mathbb{R} has its root in Loynes’s lemma ([2, Section IX.2], [18]) and is also used in (4) and (6).

If the first queue applies a FCFS service discipline, and the values $\tau \leq t$ satisfy

$$a_{\text{long}}(t) + a_{\text{cross}}(t) - a_{\text{long}}(\tau) - a_{\text{cross}}(\tau) \geq \sigma^{\text{switch}} w^{\text{switch,FCFS}}(t),$$

then all the customers which arrive up to time τ complete their service at the first queue until time t . Otherwise some of these customers are still waiting for service or service completion. This motivates the following definition of the cumulative departure process $d_c^{\text{switch,FCFS}}$ of type $c \in C$ customers at the first node under the FCFS service discipline: For $t \in \mathbb{R}$, we set

$$d_c^{\text{switch,FCFS}}(t) := \sup_{\tau \leq t, a_{\text{long}}(t) + a_{\text{cross}}(t) - a_{\text{long}}(\tau) - a_{\text{cross}}(\tau) \geq \sigma^{\text{switch}} w^{\text{switch,FCFS}}(t)} a_c(\tau).$$

It is not difficult to check that the processes $d_{\text{cross}}^{\text{switch,FCFS}}$ and $d_{\text{long}}^{\text{switch,FCFS}}$ are nondecreasing and right continuous. As the same set of τ -values appears in the supremum for both customer types $c \in C$, the departures are in the order of the arrivals. As intended these properties characterize a FCFS service discipline. See [21] and [24] for a more general version of this construction, which is similarly applied in (5) and (7).

For the priority service discipline, the workload generated by the high priority customers (i.e. the time needed to complete service for all the high priority customers that are currently at the node) of type `cross` at node `switch` is given by

$$w_{\text{cross}}^{\text{switch,PRIO}} := \frac{a_{\text{cross}}}{\sigma_{\text{switch}}} - \text{id} + \sup \left(\text{id} - \frac{a_{\text{cross}}}{\sigma_{\text{switch}}} \right). \tag{4}$$

The workload generated by the low priority customers of type `long` is

$$w_{\text{long}}^{\text{switch,PRIO}} := w_{\text{long}}^{\text{switch,FCFS}} - w_{\text{cross}}^{\text{switch,PRIO}}.$$

Hence, the cumulative departure process of type $c \in C$ customers at queue `switch` can be defined by

$$d_c^{\text{switch,PRIO}}(t) := \sup_{\tau \leq t, a_c(t) - a_c(\tau) \geq \sigma_{\text{switch}} w_c^{\text{switch,PRIO}}(\tau)} a_c(\tau). \tag{5}$$

In particular, the behavior of type `cross` customers at the first node does not depend on that of type `long` customers, and type `long` customers are only served when no type `cross` customers are present at node `switch`. Such behavior is usually called a preemptive resume priority service discipline.

The queue length process formed by type $c \in C$ customers under the service discipline $s \in S$ at node `switch` is given by

$$q_c^{\text{switch},s} := a_c^{\text{switch},s} - d_c^{\text{switch},s}.$$

The workload process at the second node for the service discipline $s \in S$ is defined as

$$w_{\text{long}}^{\text{bucket},s} := \frac{d_{\text{long}}^{\text{switch},s}}{\sigma_{\text{bucket}}} - \text{id} + \sup \left(\text{id} - \frac{d_{\text{long}}^{\text{switch},s}}{\sigma_{\text{bucket}}} \right). \tag{6}$$

The departure process of type `long` customers at queue `bucket` under the service discipline $s \in S$ at queue `switch` can be defined as

$$d_{\text{long}}^{\text{bucket},s}(t) := \sup_{\tau \leq t, d_{\text{long}}^{\text{switch},s}(t) - d_{\text{long}}^{\text{switch},s}(\tau) \geq \sigma_{\text{bucket}} w_{\text{long}}^{\text{bucket},s}(t)} d_{\text{long}}^{\text{switch},s}(\tau). \tag{7}$$

This eventually leads to the definition of the queue length process at queue `bucket` under the service discipline $s \in S$ at queue `switch`, given by

$$q_{\text{long}}^{\text{bucket},s} := d_{\text{long}}^{\text{switch},s} - d_{\text{long}}^{\text{bucket},s}.$$

We note that these definitions support both discrete customers and continuous fluids. Jumps in the arrival or departure process represent an arrival or departure respectively, of a discrete customer with a size corresponding to the height of the jump at the time of the jump. Whereas a continuous increase in the arrival or departure process can be interpreted as an arrival or departure respectively, of a certain amount of fluid. In this way discrete customer and fluid models are unified.

The use of the infinite negative time interval permits a direct modeling of steady-state behavior. In fact, as a consequence of Lemma 9 in [24] we find that the processes $q_{\text{long}}^{\text{switch},s}$,

$q_{\text{cross}}^{\text{switch},s}$, and $q_{\text{long}}^{\text{bucket},s}$ are jointly stationary for every service discipline $s \in S$. In particular, the random variable $q_{\text{long}}^{\text{bucket},s}(0)$ models the stationary queue length at queue `bucket` for the service discipline $s \in S$ at queue `switch`.

Next, we consider a sequence of the networks discussed above, in which a critical load is approached at both queues. Hence, the rates $\alpha_{\text{long}}, \alpha_{\text{cross}}, \sigma^{\text{switch}}, \sigma^{\text{bucket}}$, and all defined random processes (for instance \mathbf{a}_{long} and $\mathbf{a}_{\text{cross}}$) receive an additional index k distinguishing their position in the sequence. The assumptions and definitions detailed in the first part of this section are assumed to hold for each index $k \in \mathbb{N}$ individually.

The load at queue `switch` and `bucket` in the k th network is, respectively, given by

$$\rho_k^{\text{switch}} := \frac{\alpha_{\text{long},k} + \alpha_{\text{cross},k}}{\sigma_k^{\text{switch}}} < 1 \quad \text{and} \quad \rho_k^{\text{bucket}} := \frac{\alpha_{\text{long},k}}{\sigma_k^{\text{bucket}}} < 1.$$

We assume that there exist values $\gamma > 0$, $\kappa^{\text{switch}} > 0$, and $\kappa^{\text{bucket}} > 0$, such that

$$\begin{aligned} \lim_{k \rightarrow \infty} \alpha_{\text{cross},k} &= \infty, & (8) \\ \lim_{k \rightarrow \infty} (\alpha_{\text{long},k} - \gamma \alpha_{\text{cross},k}) &= 0, \\ \lim_{k \rightarrow \infty} (\sigma_k^{\text{switch}} - \alpha_{\text{long},k} - \alpha_{\text{cross},k}) &= \kappa^{\text{switch}}, \\ \lim_{k \rightarrow \infty} (\sigma_k^{\text{bucket}} - \alpha_{\text{long},k}) &= \kappa^{\text{bucket}}. \end{aligned}$$

These conditions imply that

$$\begin{aligned} \lim_{k \rightarrow \infty} \frac{\alpha_{\text{long},k}}{\alpha_{\text{cross},k}} &= \gamma, \\ \lim_{k \rightarrow \infty} \rho_k^{\text{switch}} = \lim_{k \rightarrow \infty} \rho_k^{\text{bucket}} &= 1, \\ \lim_{k \rightarrow \infty} (\alpha_{\text{long},k} + \alpha_{\text{cross},k})(1 - \rho_k^{\text{switch}}) &= \kappa^{\text{switch}}, \\ \text{and } \lim_{k \rightarrow \infty} \alpha_{\text{long},k}(1 - \rho_k^{\text{bucket}}) &= \kappa^{\text{bucket}}. \end{aligned}$$

Under these assumptions a critical load at node `switch` is approached with speed $\kappa^{\text{switch}} / ((1 + \gamma)\alpha_{\text{cross},k})$ and a critical load at node `bucket` is approached with speed $\kappa^{\text{bucket}} / (\gamma\alpha_{\text{cross},k})$. This justifies speaking of *heavy traffic* or *critical loading*.

4. Main result

In this section we assume that the centered sequences of arrival processes satisfy sample path large deviation principles. This leads to the main result of this work, namely, the characterization of logarithmic asymptotics for the steady-state probability of a large queue length at node `bucket` in heavy traffic.

We let $(b_k)_{k \in \mathbb{N}}$ be a sequence in $\mathbb{R}_+ \setminus \{0\}$ converging to infinity. We assume that the sequence $(\mathbf{a}_{\text{cross},k} - \alpha_{\text{cross},k} \mathbf{id})_{k \in \mathbb{N}}$ satisfies a large deviation principle with normalizing sequence $(b_k)_{k \in \mathbb{N}}$ and good rate function I_{cross} on \mathcal{D}_0 . We assume that $I_{\text{cross}}(\mathbf{a}) = \infty$, whenever $\mathbf{a} \in \mathcal{D}_0$ is not continuous or $\mathbf{a}(0) \neq 0$. We note that the stationary increments of the process $\mathbf{a}_{\text{cross},k}$, for every $k \in \mathbb{N}$, implies that the rate function I_{cross} is invariant under time shifts, i.e. for every $\tau \in \mathbb{R}$ and $\mathbf{a} \in \mathcal{D}_0$,

$$I_{\text{cross}}(\Theta_\tau \mathbf{a} - \mathbf{a}(\tau) + \mathbf{a}(0)) = I_{\text{cross}}(\mathbf{a}). \tag{9}$$

In Section 5 we give examples of cross traffic processes satisfying these assumptions.

As indicated in the introduction, the sequence of cumulative arrival processes $(\mathbf{a}_{\text{long},k})_{k \in \mathbb{N}}$ should model a constant rate traffic. In order to make this notion precise we assume that the sequence $(\mathbf{a}_{\text{long},k} - \alpha_{\text{long},k} \mathbf{id})_{k \in \mathbb{N}}$ satisfies a sample path large deviation principle in \mathcal{D}_0 with normalizing sequence $(b_k)_{k \in \mathbb{N}}$ and good rate function I_{long} given by

$$I_{\text{long}}(\mathbf{a}) := \begin{cases} 0 & \text{if } \mathbf{a} = 0, \\ \infty & \text{otherwise.} \end{cases}$$

This assumption is, for example, satisfied when $\mathbf{a}_{\text{long},k} := \alpha_{\text{long},k} \mathbf{id}$. (It is also satisfied when $\mathbf{a}_{\text{long},k}(t) := \lfloor kt\alpha_{\text{long},k} \rfloor / k$ for $k \in \mathbb{N}$ and $t \in \mathbb{R}$, if we want to model a constant rate traffic with discrete customers. Here $\lfloor x \rfloor$ is the largest integer $z \in \mathbb{Z}$ with $z \leq x$ for $x \in \mathbb{R}$.)

We define the values

$$\begin{aligned} g_{\text{long}}^{\text{FCFS}} &:= \gamma, \\ g_{\text{cross}}^{\text{FCFS}} &:= 1, \\ g_{\text{long}}^{\text{PRIO}} &:= 1 + \gamma, \\ g_{\text{cross}}^{\text{PRIO}} &:= 0. \end{aligned}$$

Furthermore, we set, for $\mathbf{a} \in \mathcal{D}_0$ and $s \in S$,

$$\begin{aligned} X^{\text{switch}}(\mathbf{a}) &:= \frac{\mathbf{a} - \kappa^{\text{switch}} \mathbf{id}}{1 + \gamma}, \\ W^{\text{switch}}(\mathbf{a}) &:= X^{\text{switch}}(\mathbf{a}) + \sup(-X^{\text{switch}}(\mathbf{a})), \\ Q_{\text{cross}}^{\text{switch},s}(\mathbf{a}) &:= g_{\text{cross}}^s W^{\text{switch}}(\mathbf{a}), \\ Q_{\text{long}}^{\text{switch},s}(\mathbf{a}) &:= g_{\text{long}}^s W^{\text{switch}}(\mathbf{a}), \\ X^{\text{bucket},s}(\mathbf{a}) &:= -\frac{\kappa^{\text{bucket}}}{\gamma} \mathbf{id} - \frac{g_{\text{long}}^s}{\gamma} W^{\text{switch}}(\mathbf{a}), \\ W^{\text{bucket},s}(\mathbf{a}) &:= X^{\text{bucket},s}(\mathbf{a}) + \sup(-X^{\text{bucket},s}(\mathbf{a})), \\ Q_{\text{long}}^{\text{bucket},s}(\mathbf{a}) &:= \gamma W^{\text{bucket},s}(\mathbf{a}). \end{aligned} \tag{10}$$

$$\tag{11}$$

Application of statement 6 in Theorem 2 of [24] yields that for each service discipline $s \in S$, the sequence

$$(\mathbf{a}_{\text{cross},k} - \alpha_{\text{cross},k} \mathbf{id}, \mathbf{q}_{\text{cross},k}^{\text{switch},s}, \mathbf{q}_{\text{long},k}^{\text{switch},s}, \mathbf{q}_{\text{long},k}^{\text{bucket},s})_{k \in \mathbb{N}}, \tag{12}$$

satisfies a sample path large deviation principle with normalizing sequence $(b_k)_{k \in \mathbb{N}}$ and good rate function I^s on \mathcal{D}_0^4 given by

$$I^s(\mathbf{a}, \mathbf{q}_{\text{cross}}^{\text{switch}}, \mathbf{q}_{\text{long}}^{\text{switch}}, \mathbf{q}_{\text{long}}^{\text{bucket}}) := \inf_{\substack{\mathbf{q}_{\text{cross}}^{\text{switch}} = \mathbf{Q}_{\text{cross}}^{\text{switch},s}(\mathbf{a}), \\ \mathbf{q}_{\text{long}}^{\text{switch}} = \mathbf{Q}_{\text{long}}^{\text{switch},s}(\mathbf{a}), \\ \mathbf{q}_{\text{long}}^{\text{bucket}} = \mathbf{Q}_{\text{long}}^{\text{bucket},s}(\mathbf{a})}} I_{\text{cross}}(\mathbf{a}).$$

This sample path large deviation principle directly implies the following convergence of conditional distributions to minimizing paths of the rate function; compare Corollary 1 in [25].

If for a service discipline $s \in S$ the measurable set $\mathcal{Q} \subset \mathcal{D}_0^4$ satisfies

$$\begin{aligned} & \inf_{(a, q_{\text{cross}}^{\text{switch}}, q_{\text{long}}^{\text{switch}}, q_{\text{long}}^{\text{bucket}}) \in \mathcal{Q}^c} I^s(a, q_{\text{cross}}^{\text{switch}}, q_{\text{long}}^{\text{switch}}, q_{\text{long}}^{\text{bucket}}) \\ &= \inf_{(a, q_{\text{cross}}^{\text{switch}}, q_{\text{long}}^{\text{switch}}, q_{\text{long}}^{\text{bucket}}) \in \mathcal{Q}^o} I^s(a, q_{\text{cross}}^{\text{switch}}, q_{\text{long}}^{\text{switch}}, q_{\text{long}}^{\text{bucket}}) < \infty, \end{aligned}$$

and there is a unique element $(a^*, q_{\text{cross}}^{*,\text{switch}}, q_{\text{long}}^{*,\text{switch}}, q_{\text{long}}^{*,\text{bucket}}) \in \mathcal{Q}^c$ which attains the infimum in the last display, the distribution of the process $(a_{\text{cross},k} - \alpha_{\text{cross},k} \mathbf{id}, q_{\text{cross},k}^{\text{switch},s}, q_{\text{long},k}^{\text{switch},s}, q_{\text{long},k}^{\text{bucket},s})$ conditioned to the event \mathcal{Q} converges to the Dirac measure of the path $(a^*, q_{\text{cross}}^{*,\text{switch}}, q_{\text{long}}^{*,\text{switch}}, q_{\text{long}}^{*,\text{bucket}})$ in distribution as $k \rightarrow \infty$. Furthermore, $q_{\text{cross}}^{*,\text{switch}} = \mathcal{Q}_{\text{cross},s}^{\text{switch},s}(a^*)$, $q_{\text{long}}^{*,\text{switch}} = \mathcal{Q}_{\text{long}}^{\text{switch},s}(a^*)$, and $q_{\text{long}}^{*,\text{bucket}} = \mathcal{Q}_{\text{long}}^{\text{bucket},s}(a^*)$.

For $t < u \in \mathbb{R}$, we let $I_{\text{cross}}^{t,u} : \mathbb{R}^2 \rightarrow \mathbb{R}_+ \cup \{\infty\}$ be the good rate function defined by

$$I_{\text{cross}}^{t,u}(x, y) := \inf_{a \in \mathcal{D}_0, a(t)=x, a(u)=y} I_{\text{cross}}(a).$$

The following theorem shows that large deviation asymptotics for the steady-state queue length distribution at node `bucket` can be obtained by solving two-dimensional minimization problems defined in terms of the rate function $I_{\text{cross}}^{t,u}$. This constitutes the main result of this work. Its proof is relegated to Appendix A.

Theorem 1. For $\delta > 0$ and $s \in S$, we set

$$\lambda^s := \frac{1 + \gamma}{g_{\text{long}}^s} \delta \quad \text{and} \quad \mu^s := \frac{1 + \gamma}{g_{\text{long}}^s} \kappa^{\text{bucket}} - \kappa^{\text{switch}}.$$

Then

$$\begin{aligned} \limsup_{k \rightarrow \infty} \frac{1}{b_k} \log \mathbb{P}(q_{\text{long},k}^{\text{bucket},s}(0) \geq \delta) &\leq - \inf_{t < u < 0, x \leq \kappa^{\text{switch}} t, y \geq \lambda^s - \mu^s u} I_{\text{cross}}^{t,u}(x, y) \\ &= - \inf_{t < u < 0} I_{\text{cross}}^{t,u}(\kappa^{\text{switch}} t, \lambda^s - \mu^s u), \end{aligned} \tag{13}$$

$$\liminf_{k \rightarrow \infty} \frac{1}{b_k} \log \mathbb{P}(q_{\text{long},k}^{\text{bucket},s}(0) \geq \delta) \geq - \inf_{t < u < 0, x < \kappa^{\text{switch}} t, y > \lambda^s - \mu^s u} I_{\text{cross}}^{t,u}(x, y). \tag{14}$$

Furthermore,

$$\lim_{k \rightarrow \infty} \frac{1}{b_k} \log \mathbb{P}(q_{\text{long},k}^{\text{bucket},\text{FCFS}}(0) \geq \delta) \leq \lim_{k \rightarrow \infty} \frac{1}{b_k} \log \mathbb{P}(q_{\text{long},k}^{\text{bucket},\text{PRIO}}(0) \geq \delta). \tag{15}$$

If there exist unique values $t^* < u^* < 0$ satisfying

$$I_{\text{cross}}^{t^*,u^*}(\kappa^{\text{switch}} t^*, \lambda^s - \mu^s u^*) = \inf_{t < u < 0, x < \kappa^{\text{switch}} t, y > \lambda^s - \mu^s u} I_{\text{cross}}^{t,u}(x, y) < \infty, \tag{16}$$

and there exists a unique path $a^* \in \mathcal{D}_0$ satisfying $a^*(t^*) = \kappa^{\text{switch}} t^*$, $a^*(u^*) = \lambda^s - \mu^s u^*$, and

$$I_{\text{cross}}(a^*) = \inf_{a \in \mathcal{D}_0, a(t^*) = \kappa^{\text{switch}} t^*, a(u^*) = \lambda^s - \mu^s u^*} I_{\text{cross}}(a),$$

the distribution of the process $(\mathbf{a}_{\text{cross},k} - \alpha_{\text{cross},k} \mathbf{id}, \mathbf{q}_{\text{cross},k}^{\text{switch},s}, \mathbf{q}_{\text{long},k}^{\text{switch},s}, \mathbf{q}_{\text{long},k}^{\text{bucket},s})$ conditioned to the event

$$\{(\mathbf{a}, \mathbf{q}_{\text{cross}}^{\text{switch}}, \mathbf{q}_{\text{long}}^{\text{switch}}, \mathbf{q}_{\text{long}}^{\text{bucket}}) \in \mathcal{D}_0^4 : \mathbf{q}_{\text{long}}^{\text{bucket}}(0) \geq \delta\},$$

converges to the Dirac measure of the path

$$(\mathbf{a}^*, \mathbf{Q}_{\text{cross}}^{\text{switch},s}(\mathbf{a}^*), \mathbf{Q}_{\text{long}}^{\text{switch},s}(\mathbf{a}^*), \mathbf{Q}_{\text{long}}^{\text{bucket},s}(\mathbf{a}^*))$$

in distribution as $k \rightarrow \infty$.

If the rate function $I_{\text{cross}}^{t,u}$ is finite and continuous for every $t < u < 0$, the upper and lower bounds in (13) and (14) agree, and the existence (but not the uniqueness) of t^* and u^* and a path \mathbf{a}^* satisfying the conditions of the last statement in Theorem 1, follows from (13) and the goodness of the rate function I_{cross} . This is the case for the example rate functions presented in Section 5.

5. Examples of sequences of input processes

Theorem 1 calls for a sequence of cross traffic arrival processes satisfying a sample path large deviation principle. Furthermore, (8) requires that the arrival rates converge to infinity. In this section we give three examples of such sequences: $E := \{\text{lrdlp}, \text{lrmdp}, \text{renmdp}\}$, which can be classified as follows. In example `lrdlp` we specify a long-range dependent sequence of cross traffic processes satisfying a large deviation principle. Under a slightly different scaling this example is modified to example `lrmdp`, where a sequence of long-range dependent processes satisfies a moderate deviation principle. The third example `renmdp` with a sequence of renewal processes satisfying a moderate deviation principle provides a contrast to the long-range dependent cross traffic of the first two examples. We specify the appropriate rate functions.

5.1. Large deviations for long-range dependent arrivals

We say that a nonnegative random variable τ is *Pareto* distributed [15, p. 222] with index $\eta > 1$ and mean $\xi > 0$ when, for every $t \geq 0$,

$$P(\tau \geq t) := \left(\frac{(\eta - 1)\xi}{t + (\eta - 1)\xi} \right)^\eta.$$

For $\eta < 2$ the Pareto distribution is *heavy-tailed*; see [39]. We fix a value $\xi > 0$ and a value η with

$$1 < \eta < 2.$$

For $k \in \mathbb{N}$, we let $(\tau_{k,i})_{i \in \mathbb{Z}}$ be a sequence of independent and identically Pareto distributed random variables with index η and mean $\xi/k > 0$. In addition, we let $(\theta_{k,i})_{i \in \mathbb{Z}}$ represent the random time points of a Poisson process with intensity $r_k > 0$ on \mathbb{R} . We assume that for every $k \in \mathbb{N}$ the sequences $(\theta_{k,i})_{i \in \mathbb{Z}}$ and $(\tau_{k,i})_{i \in \mathbb{Z}}$ are independent. Each tuple $(\theta_{k,i}, \tau_{k,i})$ is interpreted as a source which produces fluid with rate one between time $\theta_{k,i} - \tau_{k,i}$ and $\theta_{k,i}$.

For $t \in \mathbb{R}$, we define the continuous and bounded kernel $g_t : \mathbb{R}_+ \times \mathbb{R} \rightarrow \mathbb{R}$ by

$$g_t(\tau, \theta) := \begin{cases} \tau \wedge t \wedge (t + \tau - \theta)_+ \wedge \theta_+ & \text{if } t \geq 0, \\ -(\tau \wedge (-t) \wedge (\tau - \theta)_+ \wedge (\theta - t)_+) & \text{if } t < 0, \end{cases}$$

where, as usual, $s \wedge t$ ($s \vee t$) is the minimum (maximum) of two numbers $s, t \in \mathbb{R}$. For every $s < t$ in \mathbb{R} , the value $g_t(\tau, \theta) - g_s(\tau, \theta)$ can be interpreted as the cumulative output during the time interval $[s, t]$ of a source which produces fluid from time $\theta - \tau$ to time θ . See [22] for properties and graphical representations of this kernel.

For every $k \in \mathbb{N}$, we let $\mathbf{x}_{\text{lrd},k}$ be the random cumulative traffic generated by superimposing the fluid produced by the random sources $(\theta_{k,i}, \tau_{k,i})_{i \in \mathbb{Z}}$ with index $k \in \mathbb{N}$, that is,

$$\mathbf{x}_{\text{lrd},k}(t) := \sum_{i \in \mathbb{Z}} g_t(\tau_{k,i}, \theta_{k,i}).$$

We have $E(\mathbf{x}_{\text{lrd},k}(t)) = \xi k^{-1} r_k t$ for every $t \in \mathbb{R}$ and from [22] we have

$$\frac{\xi r_k}{k} = \underline{\mathbf{x}_{\text{lrd},k}} = \overline{\mathbf{x}_{\text{lrd},k}} \quad \text{with probability 1.}$$

Furthermore, the process $\mathbf{x}_{\text{lrd},k}$ has stationary increments and is long-range dependent [29].

For $n \in \mathbb{N}$ and $t_1 < t_2 < \dots < t_n$ in \mathbb{R} , we define the function $\Lambda_{\text{lrd},\eta}^{t_1, \dots, t_n} : \mathbb{R}^n \rightarrow \mathbb{R}_+$ by

$$\begin{aligned} &\Lambda_{\text{lrd},\eta}^{t_1, \dots, t_n}(y_1, \dots, y_n) \\ &:= \int_{\mathbb{R}} \int_{\mathbb{R}_+} \left(\exp\left(\sum_{i=1}^n y_i g_{t_i}(\tau, \theta)\right) - 1 - \sum_{i=1}^n y_i g_{t_i}(\tau, \theta) \right) \frac{\eta(\eta - 1)^\eta d\tau d\theta}{\tau^{\eta+1}}. \end{aligned}$$

In [23, Lemma 4] we proved that the function $\Lambda_{\text{lrd},\eta}^{t_1, \dots, t_n}$ is well defined and convex. We also showed that its Fenchel–Legendre transform $I_{\text{lrd},\eta}^{t_1, \dots, t_n} : \mathbb{R}_k \rightarrow \mathbb{R}_+ \cup \{\infty\}$ of $\Lambda_{\text{lrd},\eta}^{t_1, \dots, t_n}$ defined by

$$I_{\text{lrd},\eta}^{t_1, \dots, t_n}(x_1, \dots, x_n) := \sup_{y_1, \dots, y_n \in \mathbb{R}} \left\{ \sum_{i=1}^n x_i y_i - \Lambda_{\text{lrd},\eta}^{t_1, \dots, t_n}(y_1, \dots, y_n) \right\}$$

is a good and convex rate function. Hence, we can define a rate function $I_{\text{lrd},\eta} : \mathcal{D}_0 \rightarrow [0, \infty]$ by

$$I_{\text{lrd},\eta}(\mathbf{x}) := \sup_{n \in \mathbb{N}, t_1 < \dots < t_n \in \mathbb{R}} I_{\text{lrd},\eta}^{t_1, \dots, t_n}(\mathbf{x}(t_1), \dots, \mathbf{x}(t_n)).$$

In [23] we verified that $I_{\text{lrd},\eta}$ has compact level sets and is infinite for paths which are not continuous or are nonzero at 0. In [25] we showed that under the prerequisites

$$b_k = \frac{r_k}{k^\eta} \quad \text{for } k \in \mathbb{N}, \tag{17}$$

$$\lim_{k \rightarrow \infty} \frac{b_k}{\log k} = \infty, \tag{18}$$

the sequence

$$\left(\frac{k^\eta}{r_k} \mathbf{x}_{\text{lrd},k} - \xi k^{\eta-1} \mathbf{id} \right)_{k \in \mathbb{N}},$$

satisfies a large deviation principle with normalizing sequence $(b_k)_{k \in \mathbb{N}}$ and rate function $\xi^\eta I_{\text{lrd},\eta}(\cdot / \xi^\eta)$ in the topology induced by the norm $\|\cdot\|$ on \mathcal{D}_0 . Hence, if we assume (17) and (18) hold, set

$$\begin{aligned} \alpha_{\text{cross},k} &= \frac{k^\eta}{r_k} \mathbf{x}_{\text{lrd},k}, \\ \alpha_{\text{cross},k} &= \xi k^{\eta-1}, \end{aligned}$$

for every $k \in \mathbb{N}$, and have

$$I_{\text{cross}} = \xi^\eta I_{\text{rd},\eta} \left(\frac{\cdot}{\xi^\eta} \right),$$

we satisfy the assumptions of Section 3 and Section 4. We call this set-up example `1rd1dp`.

5.2. Moderate deviations for long-range dependent arrivals

We define the covariance function $f_H: \mathbb{R}^2 \rightarrow \mathbb{R}_+$ for $0 < H < 1$ by

$$f_H(s, t) := \frac{1}{2}(|t|^{2H} + |s|^{2H} - |t - s|^{2H}).$$

From the function f_H we derive the symmetric matrix $M_{\text{fBm},H}^{t_1, \dots, t_n} \in \mathbb{R}^{n \times n}$ with $n \in \mathbb{N}$ and $t_1 < t_2 < \dots < t_n$ in \mathbb{R} by defining its i th row and j th column entry as

$$(M_{\text{fBm},H}^{t_1, \dots, t_n})_{i,j} := f_H(t_i, t_j).$$

If $t_1 \neq 0, \dots, t_n \neq 0$, the matrix $M_{\text{fBm},H}^{t_1, \dots, t_n}$ is positive definite [25]. In this case we let $(M_{\text{fBm},H}^{t_1, \dots, t_n})_{i,j}^{-1}$ denote the i th row and j th column entry of the inverse of the matrix $M_{\text{fBm},H}^{t_1, \dots, t_n}$ and define the good rate function $I_{\text{fBm},H}^{t_1, \dots, t_n}: \mathbb{R}^n \rightarrow [0, \infty]$ by

$$I_{\text{fBm},H}^{t_1, \dots, t_n}(x_1, \dots, x_n) = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (M_{\text{fBm},H}^{t_1, \dots, t_n})_{i,j}^{-1} x_i x_j.$$

We define the function $I_{\text{fBm},H}: \mathcal{D}_0 \rightarrow [0, \infty]$ by

$$I_{\text{fBm},H}(\mathbf{x}) := \sup_{n \in \mathbb{N}, t_1 < \dots < t_n \in \mathbb{R} \setminus \{0\}} I_{\text{fBm},H}^{t_1, \dots, t_n}(\mathbf{x}(t_1), \dots, \mathbf{x}(t_n)).$$

We note that this definition implies that $I_{\text{fBm},H}(\mathbf{x}) = \infty$ if either $\mathbf{x}(0) \neq 0$ or \mathbf{x} is not continuous. The function $I_{\text{fBm},H}$ is a good rate function [17].

We know from Theorem 1 of [22] that, if $(b_k)_{k \in \mathbb{N}}$ is a sequence satisfying

$$\lim_{k \rightarrow \infty} \frac{b_k k^\eta}{r_k} = 0, \tag{19}$$

$$\lim_{k \rightarrow \infty} \frac{b_k}{\log r_k} = \infty, \tag{20}$$

and the process $\mathbf{x}_{\text{rd},k}$ is defined as in Section 5.1 for every $k \in \mathbb{N}$, the sequence

$$\left(\sqrt{\frac{k^\eta}{b_k r_k}} \mathbf{x}_{\text{rd},k} - \sqrt{\frac{k^{\eta-1} r_k}{b_k}} \xi \mathbf{id} \right)_{k \in \mathbb{N}},$$

satisfies a sample path large deviation principle with normalizing sequence $(b_k)_{k \in \mathbb{N}}$ and good rate function $\xi^{-\eta} \Delta_\eta I_{\text{fBm},(3-\eta)/2}$ with

$$\Delta_\eta := \frac{(2 - \eta)(3 - \eta)}{2(\eta - 1)^{\eta-1}}.$$

Hence, if we assume (19) and (20) hold, set

$$\begin{aligned} \mathbf{a}_{\text{cross},k} &= \sqrt{\frac{k^\eta}{b_k r_k}} \mathbf{x}_{\text{Ird},k}, \\ \alpha_{\text{cross},k} &= \sqrt{k^{\eta-1} r_k} \xi, \end{aligned}$$

for $k \in \mathbb{N}$, and have

$$I_{\text{cross}} = \xi^{-\eta} \Delta_\eta I_{\text{fBm},(3-\eta)/2},$$

we satisfy the assumptions of Section 3 and Section 4. We call this set-up example `1rd.mdp`. We make the following remarks concerning this example:

1. If \mathbf{B}_H is a standard fractional Brownian motion [38, Section 7.7.2, pp. 318–339] (i.e. a zero expectation Gaussian process with stationary increments and autocovariance function f_H), then for every sequence $(b_k)_{k \in \mathbb{N}}$ converging to infinity the sequence $(\mathbf{B}_H / \sqrt{b_k})_{k \in \mathbb{N}}$ satisfies a sample path large deviation principle with normalizing sequence $(b_k)_{k \in \mathbb{N}}$ and good rate function $I_{\text{fBm},H}$ [33]. We note that $\mathbf{B}_{1/2}$ is a standard Brownian motion with time interval \mathbb{R} .
2. In [29] it has been proved that the condition

$$\lim_{k \rightarrow \infty} \frac{r_k}{k^\eta} = \infty,$$

implies that the sequence

$$\left(\sqrt{\frac{k^\eta}{r_k}} \mathbf{x}_{\text{Ird},k} - \sqrt{k^{\eta-1} r_k} \xi \mathbf{id} \right)_{k \in \mathbb{N}},$$

converges to the fractional Brownian motion $\sqrt{\xi^\eta / \Delta_\eta} \mathbf{B}_{(3-\eta)/2}$ in distribution with respect to the topology of uniform convergence on compacts.

3. As the scaling in the large deviation principle of this section is situated between the scaling in the functional central limit theorem of the previous remark and the scaling in the large deviation principle of the previous section, and the rate function of a Gaussian process shows up, it is called a moderate deviation principle.

5.3. Moderate deviations for renewal arrivals

We let $(Z_i)_{i \in \mathbb{Z}}$ be a sequence of independent and identically distributed nonnegative random variables with mean $\zeta := E(Z_0) > 0$ and standard deviation $\Sigma := \sqrt{E(Z_0^2) - \zeta^2} > 0$. We let $\Lambda_{\text{ren}} : \mathbb{R} \rightarrow]-\infty, \infty]$ be the logarithmic moment generating function of Z_0 given by

$$\Lambda_{\text{ren}}(y) := \log E(\exp(yZ_0)),$$

and assume that Λ_{ren} is finite in a neighborhood of zero. Hence, Z_0 is light-tailed [39].

We let N be a random variable which is uniformly distributed on $[0, 1]$ and independent of the sequence $(Z_i)_{i \in \mathbb{Z}}$. We define for $k \in \mathbb{N}$ the process $\mathbf{x}_{\text{ren},k}$ on \mathcal{I} by

$$\mathbf{x}_{\text{ren},k}(t) := \begin{cases} 0 & \text{if } (N - 1)Z_0 \leq kt < NZ_0, \\ m & \text{if } NZ_0 + \sum_{h=1}^{m-1} Z_h \leq kt < NZ_0 + \sum_{h=1}^m Z_h, \\ -m & \text{if } (N - 1)Z_0 - \sum_{h=-m}^{-1} Z_h \leq kt < (N - 1)Z_0 - \sum_{h=1-m}^{-1} Z_h. \end{cases}$$

(Here an empty sum is zero.) Clearly, for every $k \in \mathbb{N}$ the process $\mathbf{x}_{\text{ren},k}$ has stationary increments and satisfies

$$\frac{k}{\zeta} = \underline{\mathbf{x}_{\text{ren},k}} = \overline{\mathbf{x}_{\text{ren},k}}$$

with probability 1.

As a consequence of Corollary 7 in [24] we find that if $(b_k)_{k \in \mathbb{N}}$ is a sequence satisfying

$$\lim_{k \rightarrow \infty} \frac{b_k}{k} = 0, \tag{21}$$

$$\text{and } \lim_{k \rightarrow \infty} \frac{b_k}{\log k} = \infty, \tag{22}$$

the sequence of centered and scaled renewal processes

$$\left(\frac{1}{\sqrt{b_k k}} \mathbf{x}_{\text{ren},k} - \frac{\sqrt{k}}{\zeta \sqrt{b_k}} \mathbf{id} \right)_{k \in \mathbb{N}},$$

satisfies a sample path large deviation principle on \mathcal{D}_0 with normalizing sequence $(b_k)_{k \in \mathbb{N}}$ and good rate function $\Sigma^{-2} \zeta^3 I_{\text{fBm},1/2}$. We note that more general results of this type are available for the topology of uniform convergence on compacts [34].

Hence, if we assume (21) and (22) hold, set

$$\begin{aligned} \mathbf{a}_{\text{cross},k} &= \frac{1}{\sqrt{b_k k}} \mathbf{x}_{\text{ren},k}, \\ \alpha_{\text{cross},k} &= \frac{\sqrt{k}}{\zeta \sqrt{b_k}}, \end{aligned}$$

for $k \in \mathbb{N}$, and have

$$I_{\text{cross}} = \frac{\zeta^3}{\Sigma^2} I_{\text{fBm},1/2},$$

we satisfy the assumptions of Section 3 and Section 4. We call this set-up example `ren.mdp`.

6. Large deviation rates and minimizing paths

In this section we calculate for each example $e \in E$ and service discipline $s \in S$ the large deviation bounds and minimizing paths of Theorem 1. In Table 1 we recall the particularities of the three examples.

TABLE 1: Characteristics of the exemplified choices of cross-traffic.

Example name e	lrd ldp	lrd mdp	ren mdp
Cross traffic $\mathbf{a}_{\text{cross},k}$	$\mathbf{x}_{\text{lrd},k} k^\eta / r_k$	$\mathbf{x}_{\text{lrd},k} \sqrt{k^\eta} / \sqrt{b_k r_k}$	$\mathbf{x}_{\text{ren},k} / \sqrt{b_k k}$
Arrival rate $\alpha_{k,1}$	$k^{\eta-1} \xi$	$b_k^{-1/2} k^{\eta/2-1} r_k^{1/2} \xi$	$b_k^{-1/2} k^{1/2} \zeta^{-1}$
Dependence	long-range	long-range	renewal
Deviations type	large	moderate	moderate
Restrictions on $(b_k)_{k \in \mathbb{N}}$	(17) and (18)	(19) and (20)	(21) and (22)
Rate function I_{cross}	$\xi^\eta I_{\text{lrd}}(\cdot / \xi^\eta)$	$\xi^{-\eta} \Delta_\eta I_{\text{fBm},(3-\eta)/2}$	$\zeta^3 \Sigma^{-2} I_{\text{fBm},1/2}$

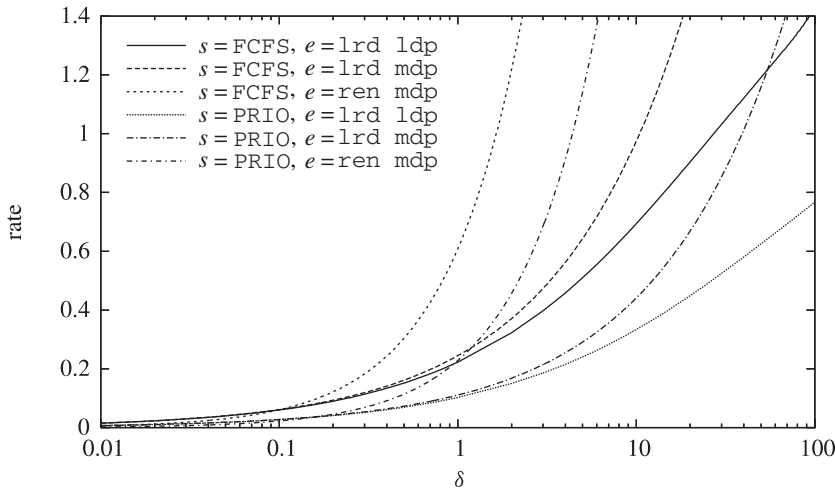


FIGURE 2: Large and moderate deviation rates as a function of δ for $\eta = 0.7$.

For every example $e \in E$ and time points $t < u$, the function $I_{\text{cross}}^{t,u}$ is convex, finite, and continuous. Hence, the upper and lower bounds in (13) and (14) match for every $\delta > 0$ and $s \in S$. The value of the rate function $I_{\text{cross}}^{t,u}(x, y)$ for $e = \text{lrd ldp}$ can be calculated numerically using the power series representation developed in Lemma 6 of [25]. For the examples $e \in \{\text{lrd mdp}, \text{ren mdp}\}$, we can use the formula

$$I_{\text{fBm},(3-\eta)/2}^{t,u}(x, y) = \frac{x^2 |u|^{3-\eta} + y^2 |t|^{3-\eta} - 2xy f_{(3-\eta)/2}(t, u)}{2|t|^{3-\eta} |u|^{3-\eta} - 2f_{(3-\eta)/2}^2(t, u)},$$

in order to numerically calculate the value of $I_{\text{cross}}^{t,u}(x, y)$. Thus, we can apply standard optimization algorithms in order to solve the two-dimensional minimization problem on the right-hand side of (13) for every cross traffic example $e \in E$ and service discipline $s \in S$.

In the calculations we use the specific parameters $\zeta = \xi = \gamma = 1$, $\kappa^{\text{switch}} = 0.2$, and $\kappa^{\text{bucket}} = 0.1$. Furthermore, we set $\Sigma^2 := 1/\Delta_\eta$ in order for the asymptotic variance of the renewal cross traffic at time one to agree with the long-range dependent cross traffic. In Figure 2 to Figure 6, one graph has been calculated for each service discipline $s \in S$ and cross traffic example $e \in E$.

In Figure 2 we plot the values of the infimum on the right-hand side of (13) as a function of δ for $\eta = 0.7$, whereas in Figure 3 we plot these values as a function of η for $\delta = 10$. These

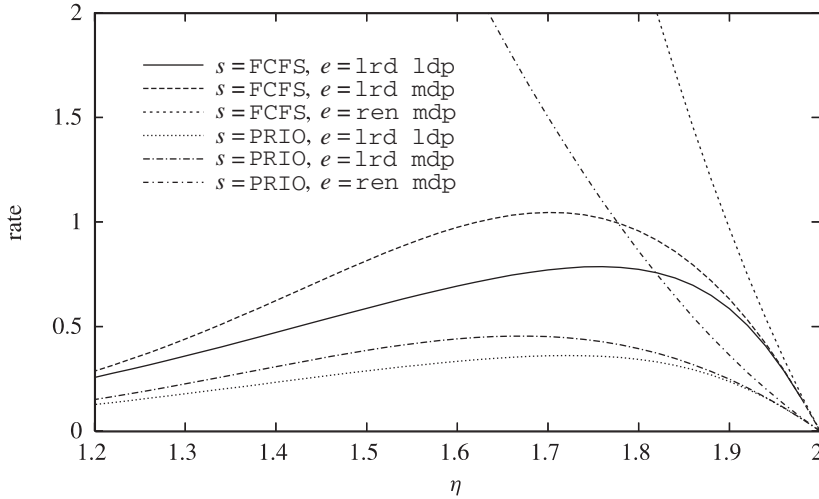


FIGURE 3: Large and moderate deviation rates as a function of η for $\delta = 10$.

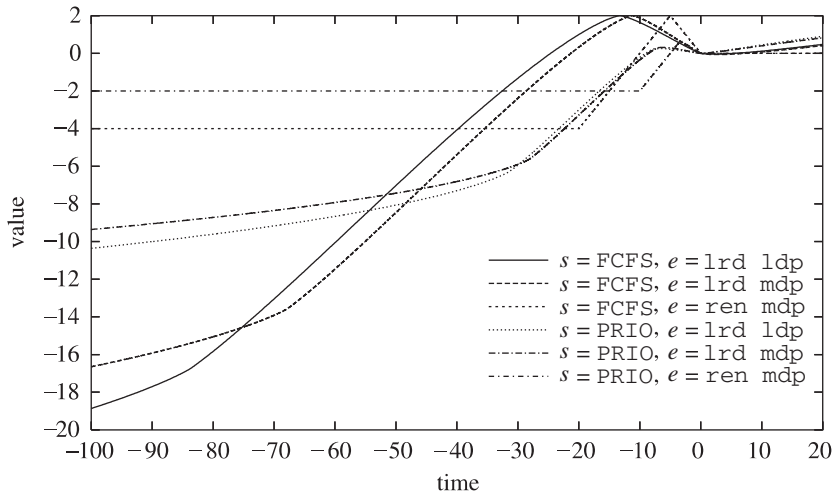


FIGURE 4: Minimizing paths \mathbf{a}^* for $\eta = 0.7$.

figures illustrate that the decay rates increase much faster for the renewal than for the long-range dependent cross traffic when δ grows. The rate functions for the tail probabilities of the queue length distribution at node 2 behave like those of a Pareto distribution for $e = \text{lrd ldp}$ (for large δ), a Weibull distribution for $e = \text{lrd mdp}$, and an exponential distribution for $e = \text{ren mdp}$. When η increases to 2 (which corresponds to a Hurst parameter of 0.5) the decay rates of all three types of asymptotic approach. Furthermore, large deviation rates are smaller than comparable moderate deviation rates for the long-range dependent cross traffic. They approach when δ is small or η is close to 1 or 2. As predicted by Theorem 1, the rates are larger for the FCFS service discipline than for the priority service discipline.

The numerical solution of the two-dimensional minimization problems indicates that for both service disciplines $s \in S$ and all three cross traffic examples $e \in E$ there exist unique

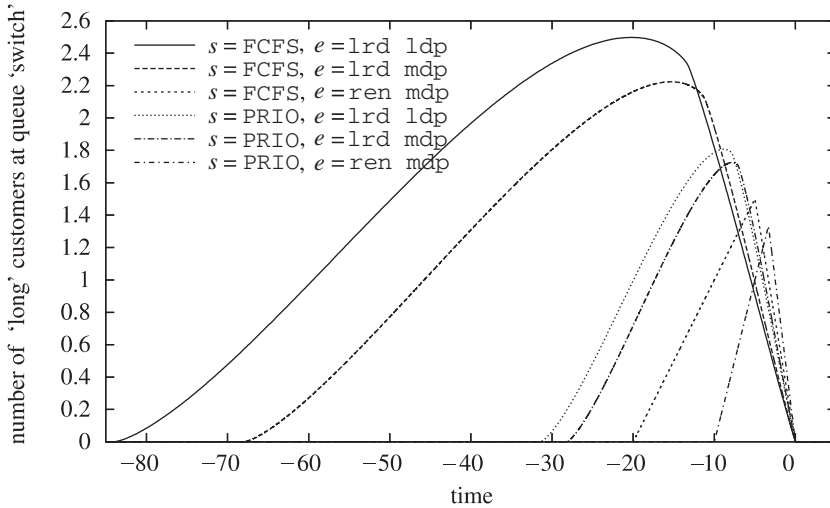


FIGURE 5: Queue length behavior $Q_{\text{long}}^{\text{switch},s}(a^*)$ of type long customers at queue switch.

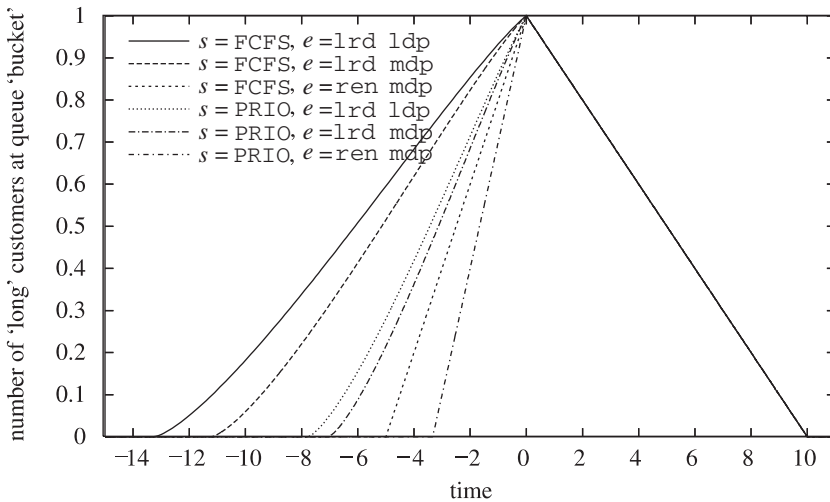


FIGURE 6: Queue length behavior $Q_{\text{long}}^{\text{bucket},s}(a^*)$ of type long customers at queue bucket.

values $t^* < u^* < 0$ such that (16) is satisfied. It determines the values of these minimizing time points as a by-product. Given these time points there exists a unique path a^* minimizing the rate function I_{cross} subject to the conditions $a^*(t^*) = \kappa^{\text{switch}}$ and $a^*(u^*) = \lambda^s - \mu^s u^*$. This is proved in [25] and [22]. According to Theorem 1, this unique path a^* characterizes the asymptotically most likely behavior of the centered and scaled cross traffic which leads to a large queue length at the second node. This asymptotically most likely conditional behavior has a period with an extraordinarily large number of arrivals in the cross traffic followed by a period with unusually few arrivals. In Figure 6 we plot the minimizing path a^* for each service discipline $s \in S$ and cross traffic example $e \in E$ for $\eta = 0.7$. The minimizing paths for the renewal cross traffic show the typical piecewise linear behavior. Analogously to [25], the

large deviation minimizing paths for the long-range dependent cross traffic ($e = \text{lr d ldp}$) are relatively close to the corresponding moderate deviation paths ($e = \text{lr d mdp}$). For each fixed cross traffic example $e \in E$, smaller movements are needed to provoke a large queue length at the second queue with the priority service discipline ($s = \text{PRIO}$) than with the FCFS service discipline ($s = \text{FCFS}$).

In Figure 5 and Figure 6 we plot the queue length behavior $Q_{\text{long}}^{n,s}(a^*)$ of type long customers corresponding to the minimizing path a^* at node $n = \text{switch}$ and node $n = \text{bucket}$, respectively. (The corresponding queue length behavior $Q_{\text{cross}}^{\text{switch},s}(a^*)$ of type cross customers at node switch is equal to $Q_{\text{long}}^{\text{switch},s}(a^*)/\gamma$ for $s = \text{FCFS}$, and 0 for $s = \text{PRIO}$.) In these two figures we can observe how the deviation a^* from the average behavior of the type cross customers leads to a backlog of type long customers first at queue switch and then at queue bucket. It is interesting to note that the behavior of these queue lengths after time zero does not depend on the service discipline s and the cross traffic example e .

Appendix A. Proof of Theorem 1

In order to prove (14) we can assume without loss of generality that its right-hand side is finite. For $\varepsilon > 0$ we can therefore find values $t' < u' < 0$, $x' < \kappa^{\text{switch}}t'$, and $y' > \lambda^s - \mu^s u'$, such that

$$I_{\text{cross}}^{t',u'}(x', y') \leq \varepsilon + \inf_{t < u < 0, x < \kappa^{\text{switch}}t, y > \lambda^s - \mu^s u} I_{\text{cross}}^{t,u}(x, y) < \infty.$$

As the function I_{cross} has compact level sets we can find a path $a' \in \mathcal{D}_0$ such that $I_{\text{cross}}(a')$ equals the value of the left-hand side of this display and the equations $a'(t') = x'$ and $a'(u') = y'$ are satisfied. In particular, the function a' is continuous and satisfies $a'(0) = 0$.

We let v be the smallest time with $u' \leq v \leq 0$ and $a'(v) = \kappa^{\text{switch}}v$. As

$$a'(u') = y' > \lambda^s - \mu^s u' > \kappa^{\text{switch}}u',$$

and a' is continuous, we know that $v > u'$. We set $t^* := t' - v$, $u^* := u' - v$, $x^* := x' - \kappa^{\text{switch}}v$, $y^* := y' - \kappa^{\text{switch}}v$, and $a^* := \Theta_{-v}a' - a'(-v) \in \mathcal{C} \cap \mathcal{D}_0$. These quantities have the following properties:

$$\begin{aligned} & t^* < u^* < 0, \\ & a^*(t^*) = x^* < \kappa^{\text{switch}}t^*, \\ & a^*(u^*) = y^* > \lambda^s - \mu^s u^*, \\ & a^*(u) > \kappa^{\text{switch}}u \quad \text{for } u \in [u^*, 0[, \\ & a^*(0) = 0, \\ & I_{\text{cross}}(a^*) = I_{\text{cross}}(a') \leq \varepsilon + \inf_{t < u < 0, x < \kappa^{\text{switch}}t, y > \lambda^s - \mu^s u} I_{\text{cross}}^{t,u}(x, y). \end{aligned}$$

The last line follows from (9).

Given these properties of the path a^* , the path $x^{*,\text{switch}} := X^{\text{switch}}(a^*)$ satisfies

$$\begin{aligned} x^{*,\text{switch}}(t^*) &< 0, \\ x^{*,\text{switch}}(u^*) &= \frac{a^*(u^*)}{1 + \gamma} - \frac{\kappa^{\text{switch}}u^*}{1 + \gamma} > \frac{\delta}{g_{\text{long}}^s} - \frac{\kappa^{\text{bucket}}u^*}{g_{\text{long}}^s}, \\ x^{*,\text{switch}}(u) &> 0 \quad \text{for } u \in [u^*, 0[, \\ x^{*,\text{switch}}(0) &= 0. \end{aligned}$$

Furthermore, since $\dot{x}^{*,\text{switch}} = -\kappa^{\text{switch}}/(1 + \gamma) < 0$, there exists a time $v^* \leq 0$ that minimizes $x^{*,\text{switch}}$ on $] -\infty, 0]$. This time v^* satisfies $v^* < u^*$ and $x^{*,\text{switch}}(v^*) < 0$. This implies that, for the path $w^{*,\text{switch}} := W^{\text{switch}}(a^*)$,

$$\begin{aligned} w^{*,\text{switch}}(0) &= -x^{*,\text{switch}}(v^*), \\ \text{and } w^{*,\text{switch}}(u^*) &= x^{*,\text{switch}}(u^*) - x^{*,\text{switch}}(v^*). \end{aligned}$$

In particular, the path $x^{*,\text{bucket}} := X^{\text{bucket},s}(a^*)$ satisfies

$$\begin{aligned} x^{*,\text{bucket}}(0) - x^{*,\text{bucket}}(u^*) &= \frac{\kappa^{\text{bucket}}u^*}{\gamma} - \frac{g_{\text{long}}^s}{\gamma}(w^{*,\text{switch}}(0) - w^{*,\text{switch}}(u^*)) \\ &= \frac{\kappa^{\text{bucket}}u^*}{\gamma} + \frac{g_{\text{long}}^s}{\gamma}x^{*,\text{switch}}(u^*) > \frac{\delta}{\gamma}. \end{aligned}$$

This in turn implies that

$$Q_{\text{long}}^{\text{bucket},s}(a^*)(0) = \gamma W^{\text{bucket},s}(a^*)(0) \geq \gamma(x^{*,\text{bucket}}(0) - x^{*,\text{bucket}}(u^*)) > \delta.$$

As the mapping $Q_{\text{long}}^{\text{bucket},s}$ is continuous, we conclude that there is an environment of a^* such that all elements a in this environment satisfy $Q_{\text{long}}^{\text{bucket},s}(a)(0) > \delta$. Application of the lower bound of the large deviation principle, (2), for the sequence defined in (12) yields

$$\begin{aligned} \liminf_{k \rightarrow \infty} \frac{1}{b_k} \log P(q_{\text{long},k}^{\text{bucket},s}(0) \geq \delta) &\geq -I_{\text{cross}}(a^*) \\ &\geq -\varepsilon - \inf_{t < u < 0, x < \kappa^{\text{switch}}t, y > \lambda^s - \mu^s u} I_{\text{cross}}^{t,u}(x, y). \end{aligned}$$

Letting ε decrease to zero we obtain (14).

Next, we prove the inequality in (13). As the mapping $Q_{\text{long}}^{\text{bucket},s}$ is continuous and the set $\{q \in \mathcal{D}_0 : q(0) \geq \delta\}$ is closed, application of the upper bound of the large deviation principle, (1), for the sequence defined in (12) yields

$$\limsup_{k \rightarrow \infty} \frac{1}{b_k} \log P(q_{\text{long},k}^{\text{bucket},s}(0) \geq \delta) \leq - \inf_{a \in \mathcal{D}_0, Q_{\text{long}}^{\text{bucket},s}(a)(0) \geq \delta} I_{\text{cross}}(a). \tag{23}$$

Without loss of generality we can assume that the right-hand side is finite.

Because the rate function I_{cross} has compact level sets, there exists a path $a' \in \mathcal{D}_0$ satisfying $Q_{\text{long}}^{\text{bucket},s}(a')(0) \geq \delta$ and being such that the value of $-I_{\text{cross}}(a')$ is equal to the right-hand side of (23). As this value is finite the path a' is continuous and satisfies $a'(0) = 0$.

There exists a largest time $\tau < 0$ at which the continuous path $Q_{\text{long}}^{\text{bucket},s}(a')$ (or equivalently $W^{\text{bucket},s}(a')$) is zero. Otherwise we would find that the path $\sup X^{\text{bucket},s}(a')$ is constant on the negative time interval (since it can only increase on time intervals, where $W^{\text{bucket},s}(a')$ is zero), which would lead to the contradiction

$$\underline{\sup X^{\text{bucket},s}(a')} = 0 \neq \frac{-\kappa^{\text{bucket}}}{\gamma} = \underline{X^{\text{bucket},s}(a')} = \underline{\sup X^{\text{bucket},s}(a')}.$$

We let u' be the smallest time with $\tau \leq u' \leq 0$ and $Q_{\text{long}}^{\text{bucket},s}(a') = \delta$. We define $a^* := \Theta_{-u'} a' - a'(-u')$. Since $Q_{\text{long}}^{\text{bucket},s}(a^*) = \Theta_{-u'} Q_{\text{long}}^{\text{bucket},s}(a')$ we conclude that $Q_{\text{long}}^{\text{bucket},s}(a^*)(t) < \delta$ for every $t \in [\tau - u', 0]$ and that $Q_{\text{long}}^{\text{bucket},s}(a^*)(0) = \delta$. Furthermore, (9) yields $I_{\text{cross}}(a^*) = I_{\text{cross}}(a')$.

This set-up implies that the path $w^{*,\text{bucket}} := W^{\text{bucket},s}(a^*)$ satisfies $w^{*,\text{bucket}}(0) = \delta/\gamma$. Hence, there exists a $u < 0$ such that the path $x^{*,\text{bucket}} := X^{\text{bucket},s}(a^*)$ satisfies

$$x^{*,\text{bucket}}(0) - x^{*,\text{bucket}}(u) = \frac{\delta}{\gamma}.$$

By definition this equation is equivalent to

$$\frac{g_{\text{long}}^s}{\gamma}(w^{*,\text{switch}}(u) - w^{*,\text{switch}}(0)) = \frac{\delta}{\gamma} - \frac{\kappa^{\text{bucket}}u}{\gamma}, \tag{24}$$

where we have set $w^{*,\text{switch}} := W^{\text{switch}}(a^*)$.

If we assume that there is a $v \in [u, 0[$ with $w^{*,\text{switch}}(v) = 0$, we obtain

$$\begin{aligned} w^{*,\text{bucket}}(v) &= x^{*,\text{bucket}}(v) + (\sup(-x^{*,\text{bucket}}))(v) \geq x^{*,\text{bucket}}(v) - x^{*,\text{bucket}}(u) \\ &> \frac{\kappa^{\text{bucket}}(u - v)}{\gamma} + \frac{g_{\text{long}}^s}{\gamma} w^{*,\text{switch}}(u) \geq \frac{\delta}{\gamma}. \end{aligned}$$

We used (24) in the last step. This yields the contradiction $Q_{\text{long}}^{\text{bucket},s}(a^*)(v) \geq \delta$. We conclude that $w^{*,\text{switch}}$ is nonzero on the time interval $[u, 0[$. Therefore, we can find a $t < u$ such that we obtain, with the definition $x^{*,\text{switch}} := X^{\text{switch}}(a^*)$,

$$\sup(-x^{*,\text{switch}})(0) = \sup(-x^{*,\text{switch}})(u) = -x^{*,\text{switch}}(t) = \frac{a^*(t)}{1 + \gamma} - \frac{\kappa^{\text{switch}}t}{1 + \gamma}.$$

In particular,

$$\begin{aligned} w^{*,\text{switch}}(0) &= x^{*,\text{switch}}(0) + \sup(-x^{*,\text{switch}})(0) = \frac{a^*(t)}{1 + \gamma} - \frac{\kappa^{\text{switch}}t}{1 + \gamma}, \\ w^{*,\text{switch}}(u) &= x^{*,\text{switch}}(u) + \sup(-x^{*,\text{switch}})(u) \\ &= \frac{a^*(u)}{1 + \gamma} - \frac{a^*(t)}{1 + \gamma} - \frac{\kappa^{\text{switch}}(u - t)}{1 + \gamma} \\ &= -\frac{\kappa^{\text{switch}}u}{1 + \gamma}. \end{aligned}$$

Using (24), this implies that

$$\begin{aligned} \mathbf{a}^*(u) &= (1 + \gamma)(\mathbf{w}^{*,\text{switch}}(u) - \mathbf{w}^{*,\text{switch}}(0)) + \kappa^{\text{switch}}u \\ &= \frac{(1 + \gamma)\delta}{g_{\text{long}}^s} - \frac{(1 + \gamma)\kappa^{\text{bucket}}u}{g_{\text{long}}^s} + \kappa^{\text{switch}}u \\ &= \lambda^s - \mu^s u. \end{aligned}$$

Furthermore, we obtain

$$\mathbf{a}^*(t) = \kappa^{\text{switch}}t - (1 + \gamma)\mathbf{w}^{\text{switch}}(0) \leq \kappa^{\text{switch}}t.$$

In particular,

$$\begin{aligned} \limsup_{k \rightarrow \infty} \frac{1}{b_k} \log P(q_{\text{long},k}^{\text{bucket},s}(0) \geq \delta) &\leq -I_{\text{cross}}(\mathbf{a}^*) \\ &\leq - \inf_{t < u < 0, x \leq \kappa^{\text{switch}}t, y \geq \lambda^s - \mu^s u} I^{t,u}(x, y), \end{aligned}$$

which completes the proof of the inequality in (13).

In order to prove the right-hand side equality in (13), we have only to show ‘ \geq ’. Hence, we can assume without loss of generality that the second last term is finite. For every $\varepsilon > 0$ we can therefore find values $t' < u' < 0$, $x' \leq \kappa^{\text{switch}}t'$, $y' \geq \lambda^s - \mu^s u'$ and a path $\mathbf{a} \in \mathcal{C}_0$ such that $\mathbf{a}(t') = x'$, $\mathbf{a}(u') = y'$, and

$$I_{\text{cross}}(\mathbf{a}) = I_{\text{cross}}^{t',u'}(x', y') \leq \varepsilon + \inf_{t < u < 0, x \leq \kappa^{\text{switch}}t, y \geq \lambda^s + \mu^s u} I_{\text{cross}}^{t,u}(x, y) < \infty.$$

In particular, \mathbf{a} is continuous. As $\lambda^s > 0$ and $\mu^s > -\kappa^{\text{switch}}$, there exist values $t'' < u'' < 0$ such that $\mathbf{a}(t'') = \kappa^{\text{switch}}t''$ and $\mathbf{a}(u'') = \lambda^s - \mu^s u''$. This implies that

$$I_{\text{cross}}^{t'',u''}(\kappa^{\text{switch}}t'', \lambda^s - \mu^s u'') \leq I_{\text{cross}}(\mathbf{a}) \leq \varepsilon + \inf_{t < u < 0, x \leq \kappa^{\text{switch}}t, y \geq \lambda^s + \mu^s u} I_{\text{cross}}^{t,u}(x, y).$$

Letting ε tend to zero, we therefore recover the equality in (13). The second statement, (15), of Theorem 1 follows from the first statement and the inequalities $\lambda^{\text{FCFS}} > \lambda^{\text{PRIO}}$ and $\mu^{\text{FCFS}} > \mu^{\text{PRIO}}$. The last statement of Theorem 1 is a consequence of Corollary 1 of [25].

References

- [1] ADDIE, R., MANNERSALO, P. AND NORROS, I. (2002). Most probable paths and performance formulae for buffers with Gaussian input traffic. *Europ. Trans. Telecommun.* **13**, 183–196.
- [2] ASMUSSEN, S. (2003). *Applied Probability and Queues* (Appl. Math. (New York) **51**), 2nd edn. Springer, New York.
- [3] AVRAM, F., DAI, J. AND HASENBEIN, J. J. (2001). Explicit solutions for variational problems in the quadrant. *Queueing Systems* **37**, 261–291.
- [4] BERAN, J. (1994). *Statistics for Long-Memory Processes*. Chapman and Hall, New York.
- [5] BERGER, A. W. AND WHITT, W. (1998). Effective bandwidth with priorities. *IEEE/ACM Trans. Networking* **6**, 447–460.
- [6] COX, D. (1984). Long-range dependence: a review. In *Statistics: An Appraisal* (Proc. 50th Anniv. Conf. Iowa State Statist. Lab.), ed. H. David, The Iowa State University Press, Ames, pp. 55–74.
- [7] DEMBO, A. AND ZAJIC, T. (1997). Uniform large and moderate deviations for functional empirical processes. *Stoch. Process. Appl.* **71**, 195–211.
- [8] DEMBO, A. AND ZEITOUNI, O. (1993). *Large Deviations Techniques and Applications*. Jones and Bartlett, London.

- [9] DUPUIS, P. AND ISHII, H. (1991). On Lipschitz continuity of the solution mapping to the Skorokhod problem, with applications. *Stoch. Stoch. Reports* **35**, 31–62.
- [10] DUPUIS, P. AND RAMANAN, K. (2002). A time-reversed representation for the tail probabilities of stationary reflected Brownian motion. *Stoch. Process. Appl.* **98**, 253–288.
- [11] GANESH, A. AND O'CONNELL, N. (2002). A large deviation principle with queueing applications. *Stoch. Stoch. Reports* **73**, 25–35.
- [12] GANESH, A., O'CONNELL, N. AND WISCHIK, D. (2004). *Big Queues* (Lecture Notes Math. **1838**). Springer, Berlin.
- [13] HARRISON, J. M. AND WILLIAMS, R. (1992). Brownian models of feedforward queueing networks: quasi-reversibility and product form solutions. *Ann. Appl. Prob.* **2**, 263–293.
- [14] HARRISON, J. M. AND REIMAN, M. I. (1981). Reflected Brownian motion on an orthant. *Ann. Prob.* **9**, 302–308.
- [15] HOGG, R. V. AND KLUGMAN, S. A. (1984). *Loss Distributions*. John Wiley, New York.
- [16] KONSTANTOPOULOS, T. AND LIN, S.-J. (1996). Fractional Brownian motion approximations of queueing networks. In *Stochastic Networks: Stability and Rare Events* (Lecture Notes Statist. **117**), eds P. Glasserman, K. Sigman and D. D. Yao, Springer, New York, pp. 257–274.
- [17] KOZACHENKO, Y., VASYLYK, O. AND SOTTINEN, T. (2002). Path space large deviations of a large buffer with Gaussian input traffic. *Queueing Systems* **42**, 113–129.
- [18] LOYNES, R. (1962). The stability of a queue with nonindependent inter-arrivals and service times. *Math. Proc. Camb. Philos. Soc.* **58**, 497–520.
- [19] MAJEWSKI, K. (1998). Large deviations of the steady-state distribution of reflected processes with applications to queueing systems. *Queueing Systems* **29**, 351–381.
- [20] MAJEWSKI, K. (2003). Large deviations for multi-dimensional reflected fractional Brownian motion. *Stoch. Stoch. Reports* **75**, 233–257. (Correction: **76** (2004), 479.)
- [21] MAJEWSKI, K. (2005). Fractional Brownian heavy traffic approximations of multiclass feedforward queueing networks. *Queueing Systems* **50**, 199–230.
- [22] MAJEWSKI, K. (2005). Sample path moderate deviations for a family of long-range dependent traffic and associated queue length processes. *Stochastics* **77**, 81–107.
- [23] MAJEWSKI, K. (2006). Sample path large deviations for a family of long-range dependent traffic and associated queue length processes. *Queueing Systems* **52**, 105–118.
- [24] MAJEWSKI, K. (2006). Sample path large deviations for multiclass feedforward queueing networks in critical loading. *Ann. Appl. Prob.* **16**, 1893–1924.
- [25] MAJEWSKI, K. (2007). Minimizing large deviation paths for a family of long-range dependent processes and their fractional Brownian approximations. *Stoch. Models* **23**, 49–77.
- [26] MANDJES, M. AND VAN UITERT, M. (2005). Sample-path large deviations for tandem and priority queues with Gaussian inputs. *Ann. Appl. Prob.* **15**, 1193–1226.
- [27] MANDJES, M., MANNERSALO, P., NORROS, I. AND VAN UITERT, M. (2006). Large deviations of infinite intersections of events in Gaussian processes. *Stoch. Process. Appl.* **116**, 1269–1293.
- [28] MANNERSALO, P. AND NORROS, I. (2002). A most probable path approach to queueing systems with general Gaussian input. *Comput. Networks* **40**, 399–411.
- [29] MIKOSCH, T., RESNICK, S., ROOTZÉN, H. AND STEGEMAN, A. (2002). Is network traffic approximated by stable Lévy motion or fractional Brownian motion? *Ann. Appl. Prob.* **12**, 23–68.
- [30] MUGULSKII, A. (1976). Large deviations for trajectories of multi-dimensional random walks. *Theory Prob. Appl.* **21**, 300–315.
- [31] NORROS, I. (1995). On the use of fractional Brownian motions in the theory of connectionless networks. *IEEE J. Selected Areas Commun.* **13**, 953–962.
- [32] NORROS, I. (1999). Busy periods of fractional Brownian storage: a large deviations approach. *Adv. Performance Anal.* **2**, 1–19.
- [33] NORROS, I. (2002). Large deviations of queues with long-range dependent input. In *Theory and Applications of Long-Range Dependence*, Birkhäuser, Boston, MA, pp. 409–415.
- [34] PUHALSKII, A. A. (1994). Large deviations of semimartingales via convergence of the predictable characteristics. *Stoch. Stoch. Reports* **49**, 27–85.
- [35] PUHALSKII, A. A. (1995). Large deviation analysis of the single server queue. *Queueing Systems* **21**, 5–66.
- [36] PUHALSKII, A. A. (1999). Moderate deviations for queues in critical loading. *Queueing Systems* **31**, 359–392.
- [37] RAMANAN, K. AND DUPUIS, P. (1998). Large deviation properties of data streams that share a buffer. *Ann. Appl. Prob.* **8**, 1070–1129.
- [38] SAMORODNITSKY, G. AND TAQQU, M. S. (1994). *Stable Non-Gaussian Random Processes*. Chapman and Hall, New York.
- [39] SIGMAN, K. (1999). Appendix: a primer on heavy-tailed distributions. *Queueing Systems* **33**, 261–275.