

TECHNICAL RESEARCH REPORT

On the Effective Bandwidth of the Output Process of a Single Server Queue

by L. Banege, A.M. Makowski

**CSHCN T.R. 95-23
(ISR T.R. 95-65)**



The Center for Satellite and Hybrid Communication Networks is a NASA-sponsored Commercial Space Center also supported by the Department of Defense (DOD), industry, the State of Maryland, the University of Maryland and the Institute for Systems Research. This document is a technical report in the CSHCN series originating at the University of Maryland.

Web site <http://www.isr.umd.edu/CSHCN/>

On the Effective Bandwidth of the Output Process of a Single Server Queue

LIONEL BANEGE
lionel@src.umd.edu
(301) 405-2932
FAX: (301) 405-6707

ARMAND M. MAKOWSKI ^{††}
armand@src.umd.edu
(301) 405-6844
FAX: (301) 314-9281

July, 18 1995

Abstract

We show that the initial condition of the buffer content in a $\mathbf{G}/\mathbf{G}/1$ queue satisfies a Sample Path Large Deviations Principle with convex good rate function, provided it has an exponential decay rate. This result is then used to derive conditions under which the transient and stationary output processes satisfy the same Large Deviations Principle. The relationship between the Large Deviations Principle and the effective bandwidth of a queue is discussed.

KEYWORDS: *output process; large deviations; effective bandwidths.*

[†]Electrical Engineering Department and Institute for Systems Research, University of Maryland, College Park, MD 20742. The work of these authors was supported partially through NSF Grant NSFD CDR-88-03012 and through NASA Grant NAGW277S.

^{††}Corresponding author

1 Introduction

A great deal of attention has recently been devoted to the development of the notion of effective bandwidth as a means to help resolve bandwidth allocation issues in high speed networks. In this approach, the network is viewed as a collection of interconnected nodes, each of them modelled as a discrete-time queue. Let $\{a_{t+1}, t = 0, 1, \dots\}$ denote a traffic stream into any one of the queues, with the usual understanding that a_{t+1} cells arrive into the node during time slot $[t, t + 1)$. To define the effective bandwidth of this arrival stream we offer it to a fictitious single server queue with constant release rate r cells/slot. If the fictitious server is equipped with an infinite buffer, the buffer content sequence $\{x_t, t = 0, 1, \dots\}$ evolves according to

$$x_0 = x; \quad x_{t+1} = [x_t + a_{t+1} - r]^+, \quad t = 0, 1, \dots \quad (1.1)$$

If the arrival process $\{a_{t+1}, t = 0, 1, \dots\}$ is stationary and ergodic, and if $\mathbf{E}[a_1] < c$, then steady state is eventually reached in the sense that $x_t \Rightarrow_t x_\infty$ for some \mathbb{R}_+ -valued rv x_∞ . Several authors [4, 12, 10, 11, 16, 17] have shown that under reasonably mild assumptions, the following buffer asymptotics

$$\lim_{b \rightarrow \infty} \frac{1}{b} \ln \mathbf{P}[x_\infty > b] = -\theta^* \quad (1.2)$$

hold for some positive constant θ^* , i.e.,

$$\mathbf{P}[x_\infty > b] \sim \exp(-b\theta^*), \quad b \rightarrow \infty. \quad (1.3)$$

The constant θ^* is an increasing function $\theta^*(r)$ of the release rate r , and is determined by the statistics of the arrival process $\{a_{t+1}, t = 0, 1, \dots\}$. The effective bandwidth $\alpha(\cdot)$ is then simply

$$\alpha(\delta) \equiv \inf\{r \in (0, \mathbf{E}[a_1]) : \delta \leq \theta^*(r)\}, \quad \delta > 0. \quad (1.4)$$

In view of (1.3), we can interpret $\alpha(\delta)$ as the smallest release rate that supports the QoS level characterized by

$$\mathbf{P}[x_\infty > b] \sim \exp(-b\delta), \quad b \rightarrow \infty. \quad (1.5)$$

The use of effective bandwidth for bandwidth allocation has been discussed in [14].

It is plain however that for this notion to be operationally useful, it is crucial to understand how the effective bandwidth of a given stream is altered as the

stream traverses the network, through its interaction for network resources with other streams at various nodes. The simplest possible situation that addresses this issue is obtained when considering the output process from a single node network.

This has recently been addressed by several authors [3, 4, 5, 9, 15, 17]. The difference between those treatments lie in the model chosen, e.g., discrete-time vs. continuous-time single-server queues, constant vs. time-varying capacities, and stationary vs. transient output processes. In what appears to be the first paper on this issue, de Veciana et al. [17] study the large deviations behavior of the stationary output process of a discrete-time $\mathbf{G}/\mathbf{D}/1$ queue. As pointed out in [3], the result of [17] holds under certain technical assumptions on the arrival process which are developed in [7]; however, the authors in [17] do not show that these technical assumptions hold for the departure process, thereby precluding their result to be applied inductively to networks of queues.

In [4], Chang derives closure properties (sum, reduction, composition and reflection mapping) for sample path large deviations. The model considered there is still a discrete-time single-server queue with deterministic capacity, but the *transient* output process (output process resulting from a queue with empty initial buffer content) is considered. Chang assumes that the arrival process is stationary, ergodic, *bounded* (a very restricted assumption) and adapted to a filtration. The existence and differentiability of a limiting log moment generating function are also assumed. Under some technical assumptions, it is shown that the transient output process satisfies a Sample Path Large Deviations Principle (SPLDP) and that the buffer asymptotics (1.3) hold for the queue fed by the transient output process. In [5], the effective bandwidth of the *stationary* output process of a discrete-time queue with bounded arrivals and time-varying capacity is derived. However the argument does not make use of the SPLDP and, as pointed out by the authors, cannot be used inductively. An important point made in [5] is that, except when the queue has constant capacity, the effective bandwidth of the stationary and transient output processes are in general different.

In [3] Berstimas, Paschalidis and Tsitsiklis study the Large Deviations behavior of networks of $\mathbf{G}/\mathbf{G}/1$ queues and establish, under some technical assumptions, that the stationary output process of a continuous-time single server queue with i.i.d. service times satisfies the Large Deviation Principle (LDP). They also show that their assumptions are satisfied by the output process, a fact which allows them to propagate the result. Finally, O’Connell and Duffield [10, 15] also study the large

deviations behavior of the transient output process of a single-server queue.

In this paper, we study sufficient conditions for the stationary and transient output process of a $\mathbf{G}/\mathbf{G}/1$ queue to have the same effective bandwidth. Assuming the statistics of the system to satisfy the LDP together with more technical assumptions, we show that those processes satisfy the same LDP. We also show that the initial queue length satisfies by itself a SP LDP provided it has an exponential decay rate. We rely on results from [1, 2] where it is shown that in all cases where a SPLDP hold for the arrival and capacity process, the transient and stationary output process satisfy the same LDP; this contradicts the statement made in [5].

The paper is organized as follows: The general setup of the $\mathbf{G}/\mathbf{G}/1$ queue and the buffer asymptotics are recalled in Section 2, the effective bandwidth of a stream is presented in Section 3, and the output process is studied in Section 4. Next, conditions of existence of an effective bandwidth for the output process are discussed in Sections 5 and 6. Finally conditions under which the stationary and transient output process satisfy the same LDP are given in Section 7.

A few words on the notation and convention used in this paper: We denote the set of non-negative integers by \mathbb{N} , and the set of all real (resp. non-negative real) numbers by \mathbb{R} (resp. \mathbb{R}_+). All rvs are defined on some probability triple $(\Omega, \mathcal{F}, \mathbf{P})$, and \mathbf{E} denotes the corresponding expectation operator. For any random sequence $\{x_t, t = 0, 1, \dots\}$ on $(\Omega, \mathcal{F}, \mathbf{P})$, we define

$$X(t_1, t_2) \equiv \begin{cases} \sum_{i=t_1}^{t_2} x_i & \text{if } t_1 \leq t_2 \\ 0 & \text{otherwise} \end{cases}. \quad (1.6)$$

2 Asymptotics for Lindley processes

We begin with some general results concerning a class of processes often referred to as Lindley processes: If $\{\xi_{t+1}, t = 0, 1, \dots\}$ denotes a sequence of \mathbb{R} -valued rvs, then the corresponding Lindley process is the sequence of rvs $\{x_t, t = 0, 1, \dots\}$ generated through the recursion

$$x_0 = x; \quad x_{t+1} = [x_t + \xi_{t+1}]^+, \quad t = 0, 1, \dots, \quad (2.1)$$

where the initial condition x is some \mathbb{R}_+ -valued rv. Many instances of this recursion, sometimes referred to as the Lindley recursion, occur in the study of queueing systems.

Consider now a second process $\{\xi_{t+1}^*, t = 0, 1, \dots\}$ which is also defined on the same probability space $(\Omega, \mathcal{F}, \mathbf{P})$ as $\{\xi_{t+1}, t = 0, 1, \dots\}$. We say that the process $\{\xi_{t+1}, t = 0, 1, \dots\}$ couples with the process $\{\xi_{t+1}^*, t = 0, 1, \dots\}$ if there exists an \mathbf{N}^* -valued rv τ such that $\tau < \infty$ a.s. and $\xi_t = \xi_t^*$ for $\tau \leq t$. If the process $\{\xi_{t+1}, t = 0, 1, \dots\}$ couples with each of the stationary and ergodic processes $\{\xi_{t+1}^*, t = 0, 1, \dots\}$ and $\{\xi_{t+1}^{**}, t = 0, 1, \dots\}$, then $\{\xi_{t+1}^*, t = 0, 1, \dots\} =_{st} \{\xi_{t+1}^{**}, t = 0, 1, \dots\}$. In other words, if the process $\{\xi_{t+1}, t = 0, 1, \dots\}$ couples with a stationary and ergodic process $\{\xi_{t+1}^*, t = 0, 1, \dots\}$, then the latter is essentially unique within the class of stationary and ergodic processes, and $\{\xi_{t+1}^*, t = 0, 1, \dots\}$ can be viewed as the stationary and ergodic version of $\{\xi_{t+1}, t = 0, 1, \dots\}$.

The following result, which is originally due to Loynes, addresses the stability of Lindley processes and is by now well known [13, 18, 19].

Proposition 2.1 *Assume the driving sequence $\{\xi_{t+1}, t = 0, 1, \dots\}$ to couple with a stationary and ergodic sequence $\{\xi_{t+1}^*, t = 0, 1, \dots\}$. If $\mathbf{E}[\xi_1^*] < 0$, then the system is stable in the sense that $x_t \Rightarrow_t x_\infty$ for some \mathbb{R}_+ -valued rv x_∞ .*

Under the stationarity assumption made above, the sequence $\{\xi_{t+1}^*, t = 0, 1, \dots\}$ can always be embedded into a bi-infinite stationary sequence $\{\xi_t^*, t = 0, \pm 1, \pm 2, \dots\}$, possibly by enlarging the original probability space $(\Omega, \mathcal{F}, \mathbf{P})$. With this in mind, it is easy to check that $x_\infty =_{st} x_{st}$ where

$$x_{st} = \left[\max_{t=0,1,\dots} \Xi^*(-t, 0) \right]^+ \quad (2.2)$$

and when $x = x_{st}$ in the recursion (2.1), the sequence $\{x_t, t = 0, 1, \dots\}$ is stationary. It is plain that the equilibrium rv x_∞ is determined solely by the stationary version $\{\xi_{t+1}^*, t = 0, 1, \dots\}$ of the original process $\{\xi_{t+1}, t = 0, 1, \dots\}$. Therefore, in studying the properties of x_∞ , there is no loss of generality in assuming that the driving sequence $\{\xi_{t+1}, t = 0, 1, \dots\}$ is indeed stationary and ergodic.

Recently there has been considerable interest in estimating the tail probabilities $\mathbf{P}[x_\infty > b]$ for large b . These asymptotics made use of the representation (2.2), and relied on the existence of large deviations estimates for the driving sequence. The main result along this line is summarized below, and was obtained in various degrees of generality by several authors [4, 9, 11, 12, 16, 17]. We write

$$\Lambda_t(\theta) \equiv \frac{1}{t} \ln \mathbf{E}[\exp(\theta(\xi_1 + \dots + \xi_t))], \quad \theta \in \mathbb{R} \quad (2.3)$$

for each $t = 1, 2, \dots$

Proposition 2.2 Assume the sequence $\{\xi_{t+1}, t = 0, 1, \dots\}$ to be stationary and ergodic, and to satisfy the following conditions:

1. The limit $\Lambda(\theta) \equiv \lim_{t \rightarrow \infty} \Lambda_t(\theta)$ exists (possibly as an extended real number) for all θ in \mathbb{R} ;
2. The set $\Theta \equiv \{\theta > 0 : \Lambda(\theta) < 0\}$ is non-empty, and $\Lambda_t(\theta) < \infty$ for all θ in Θ and $t = 1, 2, \dots$;
3. The process $\{t^{-1}\Xi(1, t), t = 1, 2, \dots\}$ satisfies the LDP with good rate function Λ^* .

Then

$$\lim_{b \rightarrow \infty} \frac{1}{b} \ln \mathbf{P}[x_\infty > b] = -\theta^* \quad (2.4)$$

where

$$\theta^* = \sup\{\theta > 0 : \Lambda(\theta) < 0\} = \inf_{y > 0} \frac{\Lambda^*(y)}{y}. \quad (2.5)$$

3 Effective bandwidth

This last result paves the way for the definition of the effective bandwidth of a source or traffic stream: Consider the traffic stream $\{a_{t+1}, t = 0, 1, \dots\}$, with the usual interpretation that a_{t+1} cells arrive at a network node during time slot $[t, t+1)$, $t = 0, 1, \dots$. We offer this arrival stream to a fictitious work-conserving single server queue with constant release rate of r cells/slot and infinite buffer capacity. Under these operational assumptions, the buffer content sequence $\{x_t, t = 0, 1, \dots\}$ evolves according to the Lindley recursion

$$x_0 = x; \quad x_{t+1} = [x_t + a_{t+1} - r]^+, \quad t = 0, 1, \dots \quad (3.1)$$

This recursion is of the form (2.1) with driving sequence $\{a_{t+1} - r, t = 0, 1, \dots\}$. By Proposition 2.1, if the arrival process $\{a_{t+1}, t = 0, 1, \dots\}$ is stationary and ergodic, and if $\mathbf{E}[a_1] < c$, then steady state is eventually reached in the sense that $x_t \Rightarrow_t x_\infty$ for some \mathbb{R}_+ -valued rv x_∞ . Specializing Proposition 2.2 to this setup, under appropriate conditions we can obtain the buffer asymptotics

$$\lim_{b \rightarrow \infty} \frac{1}{b} \ln \mathbf{P}[x_\infty > b] = -\theta_a^* \quad (3.2)$$

for some positive constant θ_a^* . More precisely, we write

$$\Lambda_t^a(\theta) \equiv \frac{1}{t} \ln \mathbf{E}[\exp(\theta(a_1 + \dots + a_t))], \quad \theta \in \mathbb{R}. \quad (3.3)$$

Proposition 3.1 *Assume the sequence $\{a_{t+1}, t = 0, 1, \dots\}$ to be stationary and ergodic, and to satisfy the following conditions:*

1. *The limit $\Lambda_a(\theta) \equiv \lim_{t \rightarrow \infty} \Lambda_t^a(\theta)$ exists (possibly as an extended real number) for all θ in \mathbb{R} ;*
2. *The set $\Theta_a \equiv \{\theta > 0 : \Lambda_a(\theta) < r\theta\}$ is non-empty, and $\Lambda_t^a(\theta) < \infty$ for all θ in Θ_a and $t = 1, 2, \dots$;*
3. *The process $\{t^{-1}A(1, t), t = 1, 2, \dots\}$ satisfies the LDP with good rate function Λ_a^* .*

Then, the buffer asymptotics (3.2) hold with θ_a^ given by*

$$\theta_a^* = \sup\{\theta > 0 : \Lambda_a(\theta) < r\theta\} = \inf_{y>0} \frac{\Lambda_a^*(r+y)}{y}. \quad (3.4)$$

The constant θ_a^* is an increasing function $\theta^*(r)$ of the release rate r , and is determined by the statistics of the arrival process $\{a_{t+1}, t = 0, 1, \dots\}$. Informally, we can rewrite (3.2) in the form

$$\mathbf{P}[x_\infty > b] \sim \exp(-b\theta_a^*), \quad b \rightarrow \infty. \quad (3.5)$$

With this in mind, the effective bandwidth $\alpha_a(\cdot)$ of the traffic stream $\{a_{t+1}, t = 0, 1, \dots\}$ is then simply

$$\alpha_a(\delta) \equiv \inf\{r \in (0, \mathbf{E}[a_1]) : \delta \leq \theta_a^*(r)\}, \quad \delta > 0 \quad (3.6)$$

so that $\alpha(\delta)$ represents the smallest release rate that supports the QoS level characterized by

$$\mathbf{P}[x_\infty > b] \sim \exp(-b\delta), \quad b \rightarrow \infty. \quad (3.7)$$

To the best of our knowledge, and with the exception of [4], Proposition 2.2 has only been established for stationary and ergodic sequences, or for sequences which couple with stationary and ergodic ones, satisfying the assumption of the proposition.

4 The output process of a discrete-time G/G/1 queue

With these preliminaries now out of the way, we consider a discrete-time G/G/1 queue with arrival and capacity sequences $\{a_{t+1}, t = 0, 1, \dots\}$ and $\{c_{t+1}, t = 0, 1, \dots\}$, where a_{t+1} (resp. c_{t+1}) denotes the number of arrivals (resp. capacity) in

the time interval $[t, t+1)$, $t = 0, 1, \dots$. The queue length sequence $\{q_t, t = 0, 1, \dots\}$ is then generated through the Lindley recursion

$$q_0 = q; \quad q_{t+1} = [q_t + \xi_{t+1}]^+, \quad t = 0, 1, \dots, \quad (4.1)$$

where we have set

$$\xi_{t+1} = a_{t+1} - c_{t+1}, \quad t = 0, 1, \dots \quad (4.2)$$

The motivation for considering time-varying single-server queues comes from the possible application of this model to satellite links or impaired wireless channels.

Throughout the discussion the \mathbb{R}_+^2 -valued process $\{(a_{t+1}, c_{t+1}), t = 0, 1, \dots\}$ is assumed to be (jointly) stationary and ergodic, and can thus be embedded into a bi-infinite stationary and ergodic sequence $\{(a_t, c_t), t = 0, \pm 1, \pm 2, \dots\}$. The queueing system (4.1) is stable if $\mathbf{E}[a_1 - c_1] < 0$, a condition enforced thereafter, in which case $q_t \Rightarrow_t q_\infty$ for some \mathbb{R}_+ -valued rv q_∞ . Note from (2.2) that $q_\infty =_{st} q_{st}$ with

$$q_{st} \equiv \left[\max_{t=0,1,\dots} A(-t, 0) - C(-t, 0) \right]^+ \quad (4.3)$$

and that with $q = q_{st}$ in the recursion (4.1), the process $\{(q_t, a_{t+1}, c_{t+1}), t = 0, 1, \dots\}$ is jointly stationary.

The problem considered in this paper is that of characterizing the effective bandwidth of the output process $\{b_{t+1}, t = 0, 1, \dots\}$, where the number b_{t+1} of departures in the time interval $[t, t+1)$ is given by

$$b_{t+1} = a_{t+1} - (q_{t+1} - q_t), \quad t = 0, 1, \dots \quad (4.4)$$

Naturally we have in mind to apply Proposition 3.1, to the sequence $\{b_{t+1}, t = 0, 1, \dots\}$. To help pinpoint the difficulties associated with this approach, we begin by developing a representation of the output process in terms of the basic input rvs of the model: First, upon iterating (4.1) it is easy to see that for each $t = 0, 1, \dots$, the relations

$$\begin{aligned} q_t &= \max \left\{ 0, q + A(1, t) - C(1, t), \max_{s=2,\dots,t} (A(s, t) - C(s, t)) \right\} \\ &= A(1, t) - C(1, t) + q - \min \left\{ 0, q + \min_{s=1,\dots,t} (A(1, s) - C(1, s)) \right\} \end{aligned}$$

hold, and simple algebra then readily shows that

$$B(1, t) = C(1, t) + \min \left\{ 0, q + \min_{s=1,\dots,t} (A(1, s) - C(1, s)) \right\}. \quad (4.5)$$

In order to write this last expression more compactly, we define the rvs $\{m(1, t), t = 0, 1, \dots\}$ by

$$m(1, t) \equiv \min_{s=1, \dots, t} (A(1, s) - C(1, s)), \quad t = 1, 2, \dots \quad (4.6)$$

and note that

$$B(1, t) = F(C(1, t), q + m(1, t)), \quad t = 1, 2, \dots \quad (4.7)$$

where the continuous mapping $F : \mathbb{R}^2 \rightarrow \mathbb{R}$ is given by

$$F(x, y) = x + \min(0, y), \quad x, y \in \mathbb{R}. \quad (4.8)$$

5 On applying Proposition 3.1 to $\{b_{t+1}, t = 0, 1, \dots\}$

We now discuss some of the difficulties in applying Proposition 3.1 to the output process $\{b_{t+1}, t = 0, 1, \dots\}$, and we do so under a set of assumptions which naturally imply those of Proposition 2.2 in the context of the discrete-time $\mathbf{G}/\mathbf{G}/1$ (4.1) with time-varying capacity: As said before, we assume the \mathbb{R}_+^2 -valued process $\{(a_{t+1}, c_{t+1}), t = 0, 1, \dots\}$ to be (jointly) stationary and ergodic. In addition, we assume the following conditions: Set

$$\Lambda_t^{a,c}(\alpha, \gamma) \equiv \frac{1}{t} \ln \mathbf{E} [\exp(\alpha A(1, t) + \gamma C(1, t))], \quad (\alpha, \gamma) \in \mathbb{R}^2 \quad (5.1)$$

for each $t = 1, 2, \dots$

1. The limit $\Lambda_{a,c}(\alpha, \gamma) \equiv \lim_{t \rightarrow \infty} \Lambda_t^{a,c}(\alpha, \gamma)$ exists (possibly as an extended real number) for all (α, γ) in \mathbb{R}^2 ;

2. The process $\{t^{-1}(A(1, t), C(1, t)), t = 1, 2, \dots\}$ satisfies the LDP with good rate function $\Lambda_{a,c}^*$.

Obviously, for all θ in \mathbb{R} and $t = 1, 2, \dots$, we have $\Lambda_t^{a-c}(\theta) = \Lambda_t^{a,c}(\theta, -\theta)$ and $\Lambda_t^a(\theta) = \Lambda_t^{a,c}(\theta, 0)$. The existence of the limits

$$\Lambda_{a-c}(\theta) \equiv \lim_{t \rightarrow \infty} \Lambda_t^{a-c}(\theta) = \Lambda_{a,c}(\theta, -\theta), \quad \theta \in \mathbb{R} \quad (5.2)$$

and

$$\Lambda_a(\theta) \equiv \lim_{t \rightarrow \infty} \Lambda_t^a(\theta) = \Lambda_{a,c}(\theta, 0), \quad \theta \in \mathbb{R} \quad (5.3)$$

is now immediate. It also plain that the processes $\{t^{-1}(A(1, t) - C(1, t)), t = 1, 2, \dots\}$ and $\{t^{-1}A(1, t), t = 1, 2, \dots\}$ each satisfy the LDP with good rate functions Λ_{a-c}^* and Λ_a^* , respectively, which can easily obtained from the Contraction Principle

[7]. Furthermore, we assume the set $\Theta_{a-c} = \{\theta > 0 : \Lambda_{a-c}(\theta) < 0\}$ to be non-empty, and $\Lambda_t^{a-c}(\theta) < \infty$ for all θ in Θ_{a-c} and $t = 1, 2, \dots$. Similarly, we assume that $\Theta_a = \{\theta > 0 : \Lambda_a(\theta) < r\theta\}$ is non-empty, and $\Lambda_t^a(\theta) < \infty$ for all θ in Θ_a and $t = 1, 2, \dots$. Under these assumptions, both Proposition 2.2 and Proposition 3.1 hold.

In order to apply Proposition 3.1 to the output process $\{b_{t+1}, t = 0, 1, \dots\}$ we need to check several assumptions:

a. The first requirement is that the output process $\{b_{t+1}, t = 0, 1, \dots\}$ be stationary and ergodic. It should be clear that this process is *not* necessarily stationary, not to say ergodic, even though the process $\{(a_{t+1}, c_{t+1}), t = 0, 1, \dots\}$ has been assumed (jointly) stationary and ergodic. Only when the system is in statistical equilibrium or steady-state, i.e., when $q = q_{st}$, is the output process $\{b_{t+1}, t = 0, 1, \dots\}$ stationary as can be seen from (4.4).

b. With $q = q_{st}$, we now set

$$\Lambda_t^b(\theta) \equiv \frac{1}{t} \ln \mathbf{E} [\exp(\theta B(1, t))], \quad \theta \in \mathbb{R} \quad (5.4)$$

for all $t = 1, 2, \dots$. The limit

$$\Lambda_b(\theta) \equiv \lim_{t \rightarrow \infty} \Lambda_t^b(\theta), \quad \theta \in \mathbb{R} \quad (5.5)$$

needs to exist.

c. Next, the set $\Theta_b = \{\theta > 0 : \Lambda_b(\theta) < 0\}$ should be non-empty, and $\Lambda_t^b(\theta) < \infty$ for all θ in Θ_b and all $t = 1, 2, \dots$.

d. Finally, we need to establish that the process $\{t^{-1}B(1, t), t = 1, 2, \dots\}$ satisfies a LDP with good rate function Λ_b^* . That such a property holds is far from obvious as the process $\{B(1, t), t = 1, 2, \dots\}$ is a complicated function of the process $\{(a_t, c_t), t = 0, \pm 1, \pm 2, \dots\}$. However, one possible avenue of progress on this issue is to observe from (4.7) that

$$\frac{B(1, t)}{t} = F\left(\frac{C(1, t)}{t}, \frac{q}{t} + \frac{m(1, t)}{t}\right), \quad t = 1, 2, \dots \quad (5.6)$$

Therefore, if the process $\{t^{-1}(C(1, t), m(1, t), q), t = 1, 2, \dots\}$ satisfies a joint LDP with good rate function $J : \mathbb{R}^3 \rightarrow [0, \infty]$, then by the Contraction Principle, so does $\{t^{-1}B(1, t), t = 1, 2, \dots\}$ with good rate function $J_b : \mathbb{R} \rightarrow [0, \infty]$ given by

$$J_b(z) \equiv \inf_{(c, q, m) \in \mathbb{R}^3} \{J(c, q, m) : z = F(c, q + m)\}, \quad z \in \mathbb{R}. \quad (5.7)$$

Of course, this approach still leaves unanswered whether the good rate function J_b is indeed Λ_b^* , the Legendre–Fenchel transform of Λ_b . We should also point out that **b.** and **c.** do not appear to be simple consequences of the enforced assumptions. Moreover, obtaining the LDP for $\{t^{-1}m(1, t), t = 1, 2, \dots\}$, whence a fortiori jointly for $\{t^{-1}(C(1, t), m(1, t), q), t = 1, 2, \dots\}$, is not an easy task, and to our knowledge, has been done only through sample path Large Deviations arguments [4, 17].

This suggests that an approach at the sample path level might be more promising, with all discrete-time processes of interest being embedded into continuous-time ones.

6 A sample path approach to effective bandwidth

Let $(D[0, 1], d_\infty)$ denote the space of right-continuous functions on $[0, 1]$ with left-hand limits, endowed with the uniform norm, and for any sequence $\{x_t, t = 0, 1, \dots\}$, define the rv $X_n(\cdot)$ on $(D[0, 1], d_\infty)$ by setting

$$\begin{aligned} X_n(t) &\equiv \frac{1}{n}X(1, \lfloor nt \rfloor) \\ &= \begin{cases} \frac{1}{n} \sum_{i=1}^{\lfloor nt \rfloor} x_i & \text{if } \lfloor nt \rfloor \geq 1 \\ 0 & \text{otherwise} \end{cases}, \quad t \in [0, 1], \quad n = 1, 2, \dots \end{aligned} \quad (6.1)$$

Following [7], we refer to $X_n(\cdot) \equiv \{X_n(t), t \in [0, 1]\}$ as the partial sum process associated with $\{x_t, t = 0, 1, \dots\}$. We can then use (4.7) to write

$$\frac{B(1, \lfloor nt \rfloor)}{n} = F\left(\frac{C(1, \lfloor nt \rfloor)}{n}, \frac{q}{n} + \frac{m(1, \lfloor nt \rfloor)}{n}\right), \quad t \in [0, 1], \quad (6.2)$$

where

$$\begin{aligned} \frac{m(1, \lfloor nt \rfloor)}{n} &= \min_{s=1, \dots, \lfloor nt \rfloor} \left(n^{-1}(A(1, s) - C(1, s)) \right) \\ &= \inf_{0 \leq s \leq t} \left(n^{-1}(A(1, s) - C(1, s)) \right), \quad t \in [0, 1]. \end{aligned} \quad (6.3)$$

As usual, the minimum over the empty set is taken to be ∞ , a convention which is consistent with our earlier definitions.

Therefore, defining the families of random processes $\{q_n(\cdot), n = 1, 2, \dots\}$ and $\{m_n(\cdot), n = 1, 2, \dots\}$ by

$$q_n(t) \equiv \frac{q}{n} \quad \text{and} \quad m_n(t) \equiv \inf_{0 \leq s \leq t} (A_n(s) - C_n(s)), \quad t \in [0, 1], \quad (6.4)$$

we can rewrite (6.2) as

$$B_n(\cdot) = \tilde{F}(C_n(\cdot), q_n(\cdot) + m_n(\cdot)), \quad n = 1, 2, \dots, \quad (6.5)$$

where the (continuous) mapping $\tilde{F} : D[0, 1]^2 \rightarrow D[0, 1]$ is defined by

$$\tilde{F}(\psi_1, \psi_2) \equiv \psi_1 + \min\{0, \psi_2\}, \quad \psi_1, \psi_2 \in D[0, 1].$$

In principle, the sample path LDP for the transient output process ($q = 0$) as well as for its stationary version ($q = q_{st}$) can now be derived from the joint LDP for the family of partial sum processes $\{(C_n(\cdot), m_n(\cdot), q_n(\cdot)), n = 1, 2, \dots\}$ through a simple application of the Contraction Principle [7].

In the sequel we will often require the sequences $\{a_{t+1} : t = 1, 2, \dots, 0\}$ and $\{c_{t+1} : t = 1, 2, \dots\}$ to satisfy a SPLDP: Let $AC_0[0, 1]$ denote the space of functions $\varphi : [0, 1] \rightarrow \mathbb{R}$ which are absolutely continuous and such that $\varphi(0) = 0$.

Assumption (E) *The partial sum processes $\{A_n(\cdot), n = 1, 2, \dots\}$ and $\{C_n(\cdot), n = 1, 2, \dots\}$ satisfy the LDP on $(D[0, 1], d_\infty)$ with good rate functions given respectively by*

$$I_A(\varphi) = \begin{cases} \int_0^1 \Lambda_A^*(\dot{\varphi}(t)) dt, & \varphi \in AC_0[0, 1] \\ \infty & \text{otherwise} \end{cases} \quad (6.6)$$

and

$$I_C(\varphi) = \begin{cases} \int_0^1 \Lambda_C^*(\dot{\varphi}(t)) dt, & \varphi \in AC_0[0, 1] \\ \infty & \text{otherwise} \end{cases} \quad (6.7)$$

where Λ_A^* and Λ_C^* are the convex good rate functions associated with the LDP satisfied by $\{t^{-1}A(1, t), t = 1, 2, \dots\}$ and $\{t^{-1}C(1, t), t = 1, 2, \dots\}$, respectively.

Conditions for this assumption to hold are presented in [8]. Although the conditions there do not cover all the cases where a stationary sequence satisfy a LDP, they are quite general.

The following result from [1, 2] shows that a sample path approach would also yield directly the buffer asymptotics and the effective bandwidth, bypassing Proposition 2.2 and 3.1

Proposition 6.1 *Under Assumption (E) we have*

$$\lim_{b \rightarrow \infty} \frac{1}{b} \log \mathbf{P}[q_\infty > b] = -\theta^*, \quad (6.8)$$

where

$$\theta^* \equiv \inf_{\theta > 0} \frac{1}{\theta} \inf \{ \Lambda_A^*(x) + \Lambda_C^*(x - \theta) : x \in \mathbb{R} \}.$$

The next result shows that (6.8) is enough to ensure the existence of a LDP for the sequence $\{q_n(\cdot), n = 1, 2, \dots\}$; its proof is given in Section 8. This assumption is satisfied in most cases of interest, as shown by Proposition 2.2 and 6.1.

Let \mathcal{C}_+ denote the subset of $C[0, 1]$ consisting of all non-negative constant functions, and denote by c_M its generic element, i.e., $c_M(t) = M$ for all t in $[0, 1]$.

Proposition 6.2 *Under (6.8), the sequence $\{q_n(\cdot), n = 1, 2, \dots\}$ satisfies the LDP on $(D[0, 1], d_\infty)$ with good rate function $I_q : D[0, 1] \rightarrow [0, \infty]$ given by*

$$I_q(\varphi) = \begin{cases} M\theta^* & \text{if } \varphi = c_M, M \geq 0 \\ \infty & \text{otherwise.} \end{cases} \quad (6.9)$$

It is actually shown in [1] that under Assumption **(E)** (6.8) is *equivalent* to the existence of the LDP for $\{q_n(\cdot) : n = 1, 2, \dots\}$. We point out that the existence of a LDP for $\{(C_n(\cdot), m_n(\cdot)) : n = 1, 2, \dots\}$ together with that of $\{q_n(\cdot), n = 1, 2, \dots\}$ is not enough, in general, to yield the joint LDP for $\{(C_n(\cdot), m_n(\cdot), q_n(\cdot)), n = 1, 2, \dots\}$. From [7, Exercise 4.2.7 p. 106], it does so for i.i.d. arrivals and capacities. It turns out that the joint LDP for $\{(C_n(\cdot), m_n(\cdot), q_n(\cdot)), n = 1, 2, \dots\}$ also holds under the much broader Assumption **(E)**. A proof of this result as well as a detailed account of this sample path approach can be found in [1] and in the forthcoming paper [2].

7 Stationary vs. Transient

Let $\{B^{st}(1, t), t = 1, 2, \dots\}$ and $\{B^{tr}(1, t), t = 1, 2, \dots\}$ denote the stationary ($q = q_{st}$) and transient ($q = 0$) output sequences. As seen from (4.3), the rv q_{st} , hence $B^{st}(1, t)$, depend on the entire past of the driving sequence $\{\xi_t, t = 0, \pm 1, \pm 2, \dots\}$, whereas the expression for $B^{tr}(1, t)$ depends only on the driving sequence through $\{\xi_1, \dots, \xi_t\}$. It is thus in general much easier to obtain the LDP and to compute the associated rate function and effective bandwidth for the transient output process, than for the stationary one. Relatively simple expressions for the effective bandwidth of the transient output process can be found in [4, 10, 15].

Proposition 7.1 *If $\mathbf{E} \left[e^{\theta q_{st}} \right] < \infty$ for all $\theta > 0$, then $\{B^{st}(1, t), t = 1, 2, \dots\}$ and $\{B^{tr}(1, t), t = 1, 2, \dots\}$ are exponentially equivalent. Thus, if one satisfies the LDP, so does the other with the same rate function.*

The condition on the exponential moment is quite strong and will not, in general, be satisfied.

Proof. It is readily seen from (4.5) that

$$0 \leq B^{st}(1, t) - B^{tr}(1, t) \leq q_{st}, \quad t = 1, 2, \dots \quad (7.1)$$

Thus, from Chebycheff inequality, for any $\delta > 0$, we have

$$\begin{aligned} \mathbf{P} \left[\left| \frac{1}{t} B^{st}(1, t) - \frac{1}{t} B^{tr}(1, t) \right| > \delta \right] &\leq \mathbf{P}[q_{st} > t\delta] \\ &\leq e^{-t\delta\theta} \mathbf{E} \left[e^{\theta q_{st}} \right], \quad \theta > 0. \end{aligned} \quad (7.2)$$

Taking the log, dividing by t and letting t goes to infinity in the last inequality yields

$$\limsup_{t \rightarrow \infty} \frac{1}{t} \log \mathbf{P} \left[\left| \frac{1}{t} B^{st}(1, t) - \frac{1}{t} B^{tr}(1, t) \right| > \delta \right] \leq -\delta\theta, \quad \theta > 0, \quad (7.3)$$

where the last step is obtained from the finiteness of the exponential moment. The result then follows easily upon letting θ go to infinity in (7.3). \blacksquare

As the next proposition shows, the existence of a joint LDP for the process $\{t^{-1}(C(1, t), m(1, t), q), \quad t = 1, 2, \dots\}$ is in some cases enough for the stationary and transient output processes to satisfy the same LDP.

Proposition 7.2 *For $q = q_{st}$, assume $\{t^{-1}(C(1, t), m(1, t), q), \quad t = 1, 2, \dots\}$ to jointly satisfy a LDP on \mathbb{R}^3 with good rate function $J : \mathbb{R}^3 \rightarrow [0, \infty]$. If the relation*

$$J(x, y - z, z) = J(x, y, z), \quad x, y, z \in \mathbb{R}, \quad (7.4)$$

holds, then the processes $\{\frac{B^{st}(1, t)}{t}, \quad t = 1, 2, \dots\}$ and $\{\frac{B^{tr}(1, t)}{t}, \quad t = 1, 2, \dots\}$ satisfy the same LDP on \mathbb{R} .

Proof. From the Contraction Principle [7] and expression (4.7), the transient and stationary output processes satisfy the LDP with respective good rate functions

$$J_b^{tr}(x) = \inf \{J(y_1, y_2, y_3) : x = y_1 + \min\{0, y_2\}, \quad y_i \in \mathbb{R}, \quad i = 1, 2, 3\}$$

and

$$J_b^{st}(x) = \inf \{ J(y_1, y_2, y_3) : x = y_1 + \min\{0, y_3 + y_2\}, y_i \in \mathbb{R}, i = 1, 2, 3 \}$$

and the proof is completed upon noting from (7.4) that

$$\begin{aligned} J_b^{st}(x) &= \inf \{ J(y_1, y_2, y_3) : x = y_1 + \min\{0, y_3 + y_2\}, y_i \in \mathbb{R}, i = 1, 2, 3 \} \\ &= \inf \{ J(y_1, y_2 - y_3, y_3) : x = y_1 + \min\{0, y_2\}, y_i \in \mathbb{R}, i = 1, 2, 3 \} \\ &= \inf \{ J(y_1, y_2, y_3) : x = y_1 + \min\{0, y_2\}, y_i \in \mathbb{R}, i = 1, 2, 3 \} \\ &= J_b^{tr}(x), x \in \mathbb{R} . \end{aligned} \tag{7.5}$$

■

This last result is preserved when a sample path Large Deviation is assumed to hold, as shown below. We consider $(D[0, 1], d_\infty)^3$ endowed with the metric d_0 given by

$$d_0((\varphi_1, \varphi_2, \varphi_3); (\psi_1, \psi_2, \psi_3)) \equiv \max_{i=1,2,3} d_0(\varphi_i, \psi_i), \quad \varphi_i, \psi_i \in C[0, 1] . \tag{7.6}$$

Let $B_n^{st}(\cdot)$ and $B_n^{tr}(\cdot)$ denote the partial sum processes associated with the stationary and transient departure sequences, i.e. the processes obtained from (6.5) with $q = q_\infty$ and $q = 0$, respectively.

Proposition 7.3 *For $q = q_{st}$, assume $\{(C_n(\cdot), m_n(\cdot), q_n(\cdot)), n = 1, 2, \dots\}$ to jointly satisfy a LDP on $(D[0, 1], d_\infty)^3$ with good rate function $I : D[0, 1]^3 \rightarrow [0, \infty]$. Under (6.8), if for all $M \geq 0$, the relation*

$$I(\varphi_1, \varphi_2 - c_M, c_M) = I(\varphi_1, \varphi_2, c_M), \quad \varphi_i \in C[0, 1], i = 1, 2 \tag{7.7}$$

holds, then the families of partial sum processes $\{B_n^{st}(\cdot), n = 1, 2, \dots\}$ and $\{B_n^{tr}(\cdot), n = 1, 2, \dots\}$ satisfy the same LDP on $(D[0, 1], d_\infty)$.

Proof. Under the enforced assumptions, the Contraction Principle yields the LDP for $\{q_n(\cdot), n = 1, 2, \dots\}$ on $(D[0, 1], d_\infty)$ with good rate function

$$I_q(\phi) = \inf \{ I(\psi_1, \psi_2, \psi_3) : \phi \equiv \psi_3, \psi_i \in D[0, 1], i = 1, 2, 3 \} . \tag{7.8}$$

Thus, Proposition 6.2 and the uniqueness of the rate function [7, p. 103] already imply

$$I(\psi_1, \psi_2, \psi_3) = \infty, \psi_i \in D[0, 1], i = 1, 2, \psi_3 \notin \mathcal{C}_+ . \tag{7.9}$$

Next, because $B_n^{st}(\cdot) = \tilde{F}(C_n(\cdot), m_n(\cdot) + q_n(\cdot))$ and $B_n^{tr}(\cdot) = \tilde{F}(C_n(\cdot), m_n(\cdot))$, the Contraction Principle yields the LDP for $\{B_n^{st}(\cdot), n = 1, 2, \dots\}$ and $\{B_n^{tr}(\cdot), n = 1, 2, \dots\}$ on $(D[0, 1], d_\infty)$, with good rate functions $I_{st}, I_{tr} : D[0, 1] \rightarrow [0, \infty]$ given respectively by

$$I_{st}(\varphi) = \inf_{\substack{\psi_i \in D[0, 1] \\ i=1, 2, 3}} \{I(\psi_1, \psi_2, \psi_3) : \varphi = \psi_1 + \min\{0, \psi_2 + \psi_3\}\}$$

and

$$I_{tr}(\varphi) = \inf_{\substack{\psi_i \in D[0, 1] \\ i=1, 2, 3}} \{I(\psi_1, \psi_2, \psi_3) : \varphi = \psi_1 + \min\{0, \psi_2\}\}.$$

To complete the proof, we note from (7.7), (7.9) and from the expressions above that

$$\begin{aligned} I_{st}(\varphi) &= \inf_{\substack{\psi_i \in D[0, 1] \\ i=1, 2, M \geq 0}} \{I(\psi_1, \psi_2, c_M) : \varphi = \psi_1 + \min\{0, \psi_2 + M\}\} \\ &= \inf_{\substack{\psi_i \in D[0, 1] \\ i=1, 2, M \geq 0}} \{I(\psi_1, \psi_2 - c_M, c_M) : \varphi = \psi_1 + \min\{0, \psi_2 + M\}\} \\ &= \inf_{\substack{\psi_i \in D[0, 1] \\ i=1, 2, M \geq 0}} \{I(\psi_1, \psi_2, c_M) : \varphi = \psi_1 + \min\{0, \psi_2\}\} \\ &= \inf_{\substack{\psi_i \in D[0, 1] \\ i=1, 2, 3}} \{I(\psi_1, \psi_2, \psi_3) : \varphi = \psi_1 + \min\{0, \psi_2\}\} \\ &= I_{tr}(\varphi), \quad \varphi \in D[0, 1]. \end{aligned} \tag{7.10}$$

■

As would be expected, by the Contraction Principle and simple algebra, the assumptions of Proposition 7.3 implies that of Proposition 7.2.

We conclude with some comments as to when the conditions of Proposition 7.3 are indeed satisfied:

7.1 GI/GI/1 queue

Let d_1 denote the Skorohod metric which makes $D[0, 1]$ into a Polish space; a detailed account of the properties of this topology can be found in [6].

The **GI/GI/1** queue is characterized by the sequences $\{a_{t+1}, t = 0, 1, \dots\}$ and $\{c_{t+1}, t = 0, 1, \dots\}$ being independent i.i.d. sequences. In that case, it is easily seen [7, Exercise 4.2.7 p. 106] that, the partial sum processes $q_n(\cdot)$ and $m_n(\cdot)$ being independent, the LDP for the joint family of processes $\{(C_n(\cdot), m_n(\cdot), q_n(\cdot)), n =$

$1, 2, \dots\}$ follows from that of $\{(C_n(\cdot), m_n(\cdot)), n = 1, 2, \dots\}$ and Proposition 6.2. Indeed, under suitable conditions on $\{a_{t+1}, t = 0, 1, \dots\}$ and $\{c_{t+1}, t = 0, 1, \dots\}$, the family of partial sum processes $\{A_n(\cdot), n = 1, 2, \dots\}$ and $\{C_n(\cdot), n = 1, 2, \dots\}$ satisfy the LDP on $(D[0, 1], d_\infty)$ and thus on $(D[0, 1], d_1)$. Then, a simple application of [7, exercise 4.2.7 p. 106] and of the Contraction Principle yields the joint LDP for $\{(C_n(\cdot), m_n(\cdot)), n = 1, 2, \dots\}$ on $(D[0, 1], d_1)^2$. The use of the metric space $(D[0, 1], d_1)$ rather than $(D[0, 1], d_\infty)$ is motivated by the separability requirement in Exercise 4.2.7 of [7]. However, this is not at all restrictive, as it is shown in [1] how to go from one space to the other.

Because the rate function associated with the LDP of $\{(C_n(\cdot), m_n(\cdot), q_n(\cdot)), n = 1, 2, \dots\}$ can be expressed as ([1])

$$I(\psi_1, \psi_2, \psi_3) = \begin{cases} \int_0^1 \Lambda^*(\dot{\psi}_1(t), \dot{\psi}_2(t), \dot{\psi}_3(t)) dt & \text{if } \psi_i \in D[0, 1], i = 1, 2, 3 \\ \infty & \text{otherwise} \end{cases},$$

for some good rate function $\Lambda^* : \mathbb{R} \rightarrow [0, \infty]$, assumption (7.7) holds and Proposition 7.3 yields the result.

7.2 Sample Path Large Deviations approach

As mentioned earlier, it is shown in [1, 2] that under Asumption **(E)** the sequence $\{(C_n(\cdot), m_n(\cdot), q_n(\cdot)), n = 1, 2, \dots\}$ with $q = q_{st}$ satisfies the LDP on $(D[0, 1], d_1)^3$ with good rate function satisfying (7.7). Therefore, by Propositions 6.1, 6.2 and 7.3 the stationary and transient output process satisfy the same sample path LDP, whence the same LDP. The generality of the assumptions under which this result holds allows its propagation along **G/G/1** queues in series.

Noteworthy is the fact (shown in [1, 2]) that the expression obtained for the rate function associated with the LDP of $\{(C_n(\cdot), m_n(\cdot), q_n(\cdot)), n = 1, 2, \dots\}$ yields a direct proof of Proposition 6.2.

8 A proof of Proposition 6.2

The proof of Proposition 6.2 passes through a series of easy Lemmas. Let $B_\alpha(\varphi)$ denote the open ball of radius α centered at φ , i.e. $B_\alpha(\varphi) \equiv \{\psi \in C[0, 1] : d_\infty(\varphi, \psi) < \alpha\}$.

Lemma 8.1 For $0 < x_- < x_+$, we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \ln \mathbf{P} [q < nx_+] = 0 \quad (8.1)$$

and

$$\lim_{n \rightarrow \infty} \frac{1}{n} \ln \mathbf{P} [nx_- < q < nx_+] = -\theta^* x_-. \quad (8.2)$$

The limits (8.1)–(8.2) are simple consequences of (6.8) and of the elementary fact

$$\lim_{n \rightarrow \infty} \frac{1}{n} \ln (1 - e^{-nz}) = 0, \quad z > 0. \quad (8.3)$$

The arguments are elementary and are omitted for the sake of brevity; details are available in [1].

Lemma 8.2 The sequence $\{q_n(\cdot), n = 1, 2, \dots\}$ is exponentially tight in $(C[0, 1], d_\infty)$.

Proof. We define

$$K_\alpha \equiv B_\alpha(0) \cap \mathcal{C}_+, \quad \alpha > 0 \quad (8.4)$$

and note that $[q_n(\cdot) \in K_\alpha^c] = [q > n\alpha] \cup [q_n(\cdot) \notin \mathcal{C}_+]$. Hence, under (6.8),

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{n} \ln \mathbf{P} [q_n(\cdot) \in K_\alpha^c] &\leq \lim_{n \rightarrow \infty} \frac{1}{n} \ln [\mathbf{P} [q > n\alpha] + \mathbf{P} [q_n(\cdot) \notin \mathcal{C}_+]] \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \ln \mathbf{P} [q > n\alpha] \\ &= -\theta^* \alpha, \end{aligned} \quad (8.5)$$

and the exponential tightness follows once it is seen that the set K_α is homeomorphic to the compact interval $[0, M]$, thus compact in $(C[0, 1], d_\infty)$. \blacksquare

For φ in $C[0, 1]$ and $\delta > 0$, we set

$$\varphi_+^\delta \equiv \inf_{t \in [0, 1]} \varphi(t) + \delta \quad \text{and} \quad \varphi_-^\delta \equiv \sup_{t \in [0, 1]} \varphi(t) - \delta. \quad (8.6)$$

Lemma 8.3 For φ in $C[0, 1]$ and $\delta > 0$, we have

$$\mathcal{L}(B_\delta(\varphi)) \equiv - \lim_{n \rightarrow \infty} \frac{1}{n} \ln \mathbf{P} [q_n(\cdot) \in B_\delta(\varphi)] = \begin{cases} \theta^* \varphi_-^\delta & \text{if } 0 < \varphi_-^\delta < \varphi_+^\delta \\ 0 & \text{if } \varphi_-^\delta \leq 0 < \varphi_+^\delta \\ \infty & \text{otherwise} \end{cases} \quad (8.7)$$

with φ_-^δ and φ_+^δ defined by (8.6).

Proof. Using the continuity of φ , we note the equality

$$[q_n(\cdot) \in B_\delta(\varphi)] = [n\varphi_-^\delta < q < n\varphi_+^\delta] \quad (8.8)$$

for each $n = 1, 2, \dots$, so that

$$\mathbf{P}[q_n(\cdot) \in B_\delta(\varphi)] = \begin{cases} 0 & \text{if } \varphi_+^\delta \leq 0 \\ 0 & \text{if } 0 < \varphi_+^\delta \leq \varphi_-^\delta \\ \mathbf{P}[q < n\varphi_+^\delta] & \text{if } \varphi_-^\delta \leq 0 < \varphi_+^\delta \\ \mathbf{P}[n\varphi_-^\delta < q < n\varphi_+^\delta] & \text{if } 0 < \varphi_-^\delta < \varphi_+^\delta. \end{cases} \quad (8.9)$$

By letting n go to infinity in (8.9), we readily obtain (8.7) from Lemma 8.1. \blacksquare

Proof of Proposition 6.2. Clearly, $\mathcal{B} \equiv \{B_\delta(\varphi) : \varphi \in C[0, 1], \delta > 0\}$ is a base for $(C[0, 1], d_\infty)$. Therefore, by Theorem 4.1.11 of [7, p. 106] and Lemma 8.3, the sequence $\{q_n(\cdot), n = 1, 2, \dots\}$ satisfies the weak LDP on $(C[0, 1], d_\infty)$ with rate function

$$I_q(\varphi) = \sup\{\mathcal{L}(B_\delta(\psi)) : \varphi \in B_\delta(\psi), \psi \in C[0, 1], \delta > 0\} \quad (8.10)$$

where $\mathcal{L}(B_\delta(\psi))$ is given by (8.6). In computing $I_q(\varphi)$ for each φ in $C[0, 1]$, several cases arise:

1. $\varphi \geq 0$, i.e., $\varphi(t) \geq 0$ for all t in $[0, 1]$: In that case, whenever φ lies in the open ball $B_\delta(\psi)$ for some ψ in $C[0, 1]$ and $\delta > 0$, we find $\psi(t) - \delta < \varphi(t) < \psi(t) + \delta$ for all t in $[0, 1]$, so that by continuity,

$$\psi_+^\delta \equiv \inf_{t \in [0, 1]} \psi(t) + \delta > \inf_{t \in [0, 1]} \varphi(t) \geq 0. \quad (8.11)$$

Two sub-cases then arise:

1.a. $\inf_{t \in [0, 1]} \varphi(t) < \sup_{t \in [0, 1]} \varphi(t)$, i.e., φ is not a constant: If we take $\psi = \varphi$ and δ such that $0 < \delta < \frac{1}{2} \left(\sup_{t \in [0, 1]} \varphi(t) - \inf_{t \in [0, 1]} \varphi(t) \right)$, then φ belongs to $B_\delta(\psi)$, and we have

$$\begin{aligned} \psi_-^\delta &= \sup_{t \in [0, 1]} \psi(t) - \delta \\ &= \sup_{t \in [0, 1]} \varphi(t) - \delta \\ &> 2\delta + \inf_{t \in [0, 1]} \varphi(t) - \delta = \psi_+^\delta. \end{aligned} \quad (8.12)$$

Therefore, $\psi_+^\delta < \psi_-^\delta$, so that $\mathcal{L}(B_\delta(\psi)) = \infty$ by Lemma 8.3, and $I_q(\varphi) = \infty$ follows.

1.b. $\inf_{t \in [0,1]} \varphi(t) = \sup_{t \in [0,1]} \varphi(t)$, i.e., $\varphi \equiv M \geq 0$: In that case, whenever φ lies in the open ball $B_\delta(\psi)$ for some ψ in $C[0,1]$ and $\delta > 0$, we find $M - \delta < \psi(t) < M + \delta$ for all t in $[0,1]$, whence by continuity, we see that

$$\psi_-^\delta = \sup_{t \in [0,1]} \psi(t) - \delta < M < \inf_{t \in [0,1]} \psi(t) + \delta = \psi_+^\delta. \quad (8.13)$$

In short, $\psi_-^\delta < \psi_+^\delta$, so that $\mathcal{L}(B_\delta(\psi)) = \theta^*(\psi_-^\delta)^+$ by Lemma 8.3, and (8.10) reduces to

$$I_q(\varphi) = \sup\{\theta^*(\psi_-^\delta)^+ : \varphi \in B_\delta(\psi), \psi \in C[0,1], \delta > 0\}. \quad (8.14)$$

It is now plain from (8.13) and (8.14) that $I_q(\varphi) \leq \theta^*M$. On the other hand, upon taking $\psi = \varphi = M$ in (8.13), we get $(\psi_-^\delta)^+ = (M - \delta)^+$ for all $\delta > 0$, so that $I_q(\varphi) \geq \theta^*M$, and we conclude $I_q(\varphi) = \theta^*M$.

2. $\varphi(s) < 0$ for some s in $[0,1]$: If we take $\psi = \varphi$, then φ lies in $B_\delta(\varphi)$ and $\inf_{t \in [0,1]} \psi(t) < 0$. Upon selecting $\delta > 0$ such that $\psi_+^\delta = \inf_{t \in [0,1]} \psi(t) + \delta < 0$, we get $\mathcal{L}(B_\delta(\psi)) = \infty$ by Lemma 8.3, whence $I_q(\varphi) = \infty$.

The expression (6.9) for I_q is obtained by combining the various cases. To conclude, we recall that d_∞ is the induced metric on the closed set $(C[0,1], d_\infty)$ of $(D[0,1], d_\infty)$ (or for that matter, d_1 or d_∞). The sequence $\{q_n(\cdot), n = 1, 2, \dots\}$ is exponentially tight by Lemma 8.2, and therefore it satisfies the (strong) LDP with *good* rate function I_q on $(C[0,1], d_\infty)$ [7, Lemma 1.2.18, p. 8]. Therefore, $C[0,1]$ being a closed measurable subset of $(D[0,1], d_\infty)$ with $\mathbf{P}[q_n(\cdot) \in C[0,1]] = 1$ for all $n = 1, 2, \dots$ and $\mathcal{D}_{I_q} \subseteq C[0,1]$, the desired result follows from Lemma 4.1.5 of [7, p. 104]. ■

References

- [1] L. Banege, *Large Deviations and Effective Bandwidth for Queues in Series*, Ph.D. Thesis, Electrical engineering Department, University of Maryland, College Park, (MD), 1995.
- [2] L. Banege and A.M. Makowski, "A sample path approach to Large Deviations and effective bandwidths for $\mathbf{G}/\mathbf{G}/1$ queues in series," *In preparation*.

- [3] D. Bertsimas, I.Ch. Paschalidis and J.N. Tsitsiklis, "On the large deviations behavior of acyclic networks of G/G/1 queues," Technical Report **LIDS-P-2278**, Laboratory for Information and Decision Systems and Operations Research Center, Massachusetts Institute of Technology, Cambridge (MA), 1994.
- [4] C.-S. Chang, "Sample path large deviations and intree Networks," *Queuing Systems – Theory & Applications*, forthcoming 1995.
- [5] C.-S. Chang and T. Zajic, "Effective bandwidths of departure processes from queues with time varying capacities," In IEEE Infocom (1995), pp. 1001–1009.
- [6] P. Billingsley, *Convergence of Probability Measures*, J. Wiley & Sons, New York (NY), 1968.
- [7] A. Dembo and O. Zeitouni, *Large Deviation Techniques and Applications*, Jones and Bartlett, Boston (MA), 1993.
- [8] A. Dembo and T. Zajic, "Large deviations: From empirical mean and measure to partial sum process," *Stochastic Processes and Their Applications* **57** (1995), pp. 191–224.
- [9] N.G. Duffield and N. O’Connell, "Large deviations and overflow probabilities for the general single server queue, with applications," Technical Report **DIAS-STP-93-30**, Dublin Institute for Advanced Studies, Dublin (Ireland), 1993. *Proceedings of the Cambridge Philosophical Society*, forthcoming 1995.
- [10] N.G. Duffield and N. O’Connell, "Large deviations for arrivals, departures, and overflow in some queues of interacting traffic," Technical Report **DIAS-APG-94-08**, Dublin Institute for Advanced Studies, Dublin (Ireland), 1994.
- [11] P.W. Glynn and W. Whitt, "Logarithmic asymptotics for steady-state tail probabilities in a single-server queue," *Journal of Applied Probability* **31A** (1994), pp. 131–156.
- [12] G. Kesidis, J. Walrand and C.-S. Chang, "Effective bandwidths for multi-class Markov fluids and other ATM sources," *IEEE/ACM Transactions on Networking* **1** (1993), pp. 424–428.
- [13] R.M. Loynes, "The stability of a queue with non-independent inter-arrival and service times," *Proceedings of the Cambridge Philosophical Society* **58** (1962), pp. 497–520.

- [14] R. Guérin, H. Ahmadi and M. Naghshineh, "Equivalent capacity and its application to bandwidth allocation in high-speed networks," *IEEE Journal on Selected Areas in Communications* **JSAC-9** (1991), pp. 968–981.
- [15] N. O'Connell, "Large deviations in queueing networks," *Technical Report DIAS-STP-94-13*, Dublin Institute for Advanced Studies, 1994.
- [16] G. de Veciana and J. Walrand, "Effective bandwidths: call admission, traffic policing & filtering for ATM Networks," *Technical Report M93/47*, University of California at Berkeley/Electronics Research Laboratory, 1994.
- [17] G. de Veciana, C. Courcoubetis and J. Walrand, "Decoupling bandwidths for networks: A decomposition approach to resource management," *Technical Report M93/50*, University of California at Berkeley/Electronics Research Laboratory, 1995.
- [18] W. Szczotka, "Stationary representation of queues. I," *Advances in Applied Probability* **18** (1986), pp. 815–846.
- [19] K. Sigman, *Stationary Marked Point Processes: An Intuitive Approach*, Chapman & Hall, New York (NY), 1994.
- [20] W. Whitt, "Tail probability with statistical multiplexing and effective bandwidths in multi-class queues," *Telecommunication Systems*, to appear.