Probability in the Engineering and Informational Sciences, **12**, 1998, 479–507. Printed in the U.S.A.

FUNCTIONAL LARGE DEVIATION PRINCIPLES FOR WAITING AND DEPARTURE PROCESSES

ANATOLII A. PUHALSKII

Department of Mathematics
University of Colorado at Denver
Denver, Colorado 80217-3364
puhalski@math.cudenver.edu
and
Institute for Problems in Information Transmission
19 Bolshoi Karetnii
Moscow 101447, Russia

WARD WHITT

AT&T Labs
Florham Park, New Jersey 07932-0971
wow@research.att.com

We establish functional large deviation principles (FLDPs) for waiting and departure processes in single-server queues with unlimited waiting space and the first-in first-out service discipline. We apply the extended contraction principle to show that these processes obey FLDPs in the function space D with one of the nonuniform Skorohod topologies whenever the arrival and service processes obey FLDPs and the rate function is finite for appropriate discontinuous functions. We apply our previous FLDPs for inverse processes to obtain an FLDP for the waiting times in a queue with a superposition arrival process. We obtain FLDPs for queues within acyclic networks by showing that FLDPs are inherited by processes arising from the network operations of departure, superposition, and random splitting. For this purpose, we also obtain FLDPs for split point processes. For the special cases of deterministic arrival processes and deterministic service processes, we obtain convenient explicit expressions for the rate function of the departure process, but not more generally. In general, the rate function for the departure process evidently must be calculated numerically. We also obtain an FLDP for the departure process of completed work, which has important application to the concept of effective bandwidths for admission control and capacity planning in packet communication networks.

0269-9648/98 $12.50

1. INTRODUCTION

The purpose of this paper is to establish functional (or sample path) large deviation principles (FLDPs) for stochastic processes arising in queues and acyclic networks of queues. A distinguishing feature from previous work in this direction, notably by de Veciana, Courcoubetis, and Walrand [37] and Chang [7], is our focus on FLDPs in the function space D with the (nonuniform) Skorohod [33] topologies, where the rate functions may be finite on some discontinuous functions.

Establishing such general FLDPs is challenging and interesting mathematically, but there also is substantial practical engineering motivation, which we first describe. In recent years there has been great interest in large deviations principles (LDPs) for queueing models, primarily motivated by the problems of admission control and capacity planning in emerging high-speed packet communication networks. These LDPs were especially important because they provide a theoretical framework supporting a concept known as an *effective bandwidth* (see Chang and Thomas [9], Kelly [19], de Veciana et al. [38], and Whitt [40]).

In a packet network, sources do not receive dedicated bandwidth (e.g., circuits) for the entire duration of a connection, but instead emit packets at a variable rate. However, admission control and capacity planning in a packet network can be greatly simplified if each connection can be treated as if it required a constant "effective" bandwidth throughout the active period of the connection. A given set of connections can then be deemed feasible if the sum of the effective bandwidths is less than the total available capacity. By using effective bandwidths in this manner, the problems of admission control and capacity planning can be addressed as in circuit-switched networks. For capacity planning, we can apply stochastic loss networks, as in Ross [31].

It is evident that an effective bandwidth should be some value between the average rate and the peak (maximum) rate of the connection, but any actual value must be an approximation. In this setting, a commonly expressed goal is to admit as many connections as possible, subject to the constraint that the long-run average probability of packet loss is suitably small. Since this loss probability target is usually set very small, e.g., at 10^{-9}, it is natural to consider large deviations theory. The problem of identifying appropriate effective bandwidths has been approached by considering a fluid queueing model with unlimited buffer, constant output rate, the first-come first-served (FCFS) service discipline and an input composed of the superposition of several independent nondecreasing stochastic processes each with stationary increments. With this model, the loss probability constraint is represented by the constraint

$$P(L > b) \leq p, \tag{1.1}$$

where L is the steady-state workload (buffer content). The large deviations analysis is based on the limit as $b \to \infty$ and $p \to 0$ in Eq. (1.1). The large deviations analysis indicates that, under regularity conditions, the effective bandwidth of source i should be

$$e_i = \alpha_i(\theta^*)/\theta^* \quad \text{for} \quad \theta^* = -(\log p)/b \tag{1.2}$$

where p and b come from Eq. (1.1) and $\alpha_i(\theta)$ is the asymptotic logarithmic moment generating function (almgf), i.e.,

$$\alpha_i(\theta) = \lim_{t\to\infty} \frac{1}{t} \log E[\exp(\theta A_i(t))], \tag{1.3}$$

with $A_i(t)$ representing the input (arrivals) from source i in the interval $[0,t]$. (We assume that $A_i(t)$ has stationary increments.)

The present paper was motivated in part by two remaining problems. The first is the desire to extend the effective bandwidth concept from a single queue to a network of queues (because a communication network does not act as a single queue). The second is the desire to extend the effective bandwidth concept from the FCFS service discipline to other service disciplines such as priorities and generalized processor sharing, which are very important for providing appropriate grades of service to very different sources, e.g., voice, data, and video.

It turns out that both problems can be approached by establishing LDPs for departure processes. If we can establish an LDP for a departure process, then we can extend the effective bandwidth concept to acyclic networks of queues. Significant progress on that program was carried out by de Veciana et al. [37] and Chang [7]. They found, again under regularity conditions, that the departure process (of completed work) $D(t)$ has the almgf

$$\delta(\theta) = \begin{cases} \alpha(\theta), & \theta < \hat{\theta} \\ \alpha(\hat{\theta}) + c(\theta - \hat{\theta}), & \theta > \hat{\theta}, \end{cases} \tag{1.4}$$

where c is the constant output rate from the queue, $\alpha(\theta) \equiv \Sigma\alpha_i(\theta)$ is the almgf for the aggregate arrival process, defined as in Eq. (1.3), and $\hat{\theta}$ is a "decoupling" bandwidth defined by $\alpha'(\hat{\theta}) = c$.

A key to establishing Eq. (1.4) was exploiting a functional (or sample path) LDP (FLDP). However, the FLDP used, involving the uniform topology, places strong restrictions on the input processes for which Eq. (1.4) can be established. In particular, it was necessary to work in discrete time and the almgf in Eq. (1.3) is required to be finite everywhere. This finiteness requirement is satisfied if the increments of $A_i(t)$ are bounded, which is perhaps an acceptable condition from an engineering perspective, but we want to know what happens more generally. For example, that FLDP does not imply Eq. (1.4) even for the M/D/1 fluid queue (with a single Poisson arrival process).

In this paper (Section 5) we show that Eq. (1.4) is valid much more generally. In particular, it suffices to assume that FLDPs hold for the input processes $A_i(t)$ in the function space D with a nonuniform Skorohod topology. Our LDP for departure processes (based on an assumed FLDP for the arrival processes) is applied in Berger and Whitt [3,4] in order to establish the exact large-buffer-asymptotic admissible set when there are several priority classes. Unfortunately, this admissible set with priorities does not have a single linear boundary, so it does not directly support the concept of effective bandwidths. However, a natural approximation for the exact admissible set has a linear constraint for each priority class, which supports a new notion of effective bandwidths. With priorities, this analysis indicates that there should be multiple effective bandwidths, one for the given priority class and one for each lower priority class.

The model of interest for effective bandwidths, for which we establish Eq. (1.4), is a fluid queue. It is natural to ask what happens in the standard G/GI/1 queue, which has i.i.d. service times with a general distribution and a general stationary arrival process. As should be expected, we show that the discussion above applies essentially unchanged to the G/D/1 model with deterministic service times. However, we find that the departure-process LDP is much more complicated with nondeterministic service times. (A related observation has been made by Chang and Zajic [10].) We show that, in general, it is necessary to solve an optimization problem in order to calculate the LDP rate function for the departure process. We demonstrate that the departure-time rate function does not simplify by deriving the rate function in the special case of deterministic interarrival times (Thm. 4, Cor. 2). We propose using an upper bound for the departure-time rate function as an approximation (Rmk. 4.4). A promising direction for future research is the application of mathematical programming to calculate these rate functions systematically.

Our goals here extend beyond the communications network application to try to establish FLDPs in D with appropriate nonuniform topologies for queueing processes in queueing networks. Here we focus successively on waiting times, departure times, the departure process of completed work, and split point processes. This paper parallels that of Chang [7], which establishes FLDPs with the uniform topology for discrete-time processes in acyclic queueing networks, which in turn parallels much earlier heavy-traffic FCLTs for queues in Iglehart and Whitt [18]. This paper complements that of Puhalskii [28], where FLDPs were obtained for the queue length process and the virtual waiting time process in the GI/GI/1 queue. This paper is also a sequel to that of Puhalskii and Whitt [30], showing how FLDPs for inverse processes established in [30] can be applied to queueing models. The FLDPs for inverse processes enable us to obtain an FLDP for the waiting times in the $\sum_{i=1}^{k} G_i/G/1$ queue, which has an arrival process that is a superposition of arrival processes (Thms. 3.2 and 3.3, below). The inverse FLDPs also play a role in establishing FLDPs for randomly split point processes, which arise when departures from one queue are routed to one of several other queues or leave the network.

Just as functional central limit theorems (FCLTs) are useful to establish ordinary central limit theorems for various functionals of stochastic processes [6,39], so are functional (or sample-path) large deviation principles (FLDPs) useful to establish ordinary large deviations principles (LDPs) for various functionals of stochastic processes [25]. The contraction principle and its extensions play the role for FLDPs that the continuous mapping theorem and its extensions play for FCLTs. The nonuniform Skorohod topologies are important in part to avoid measurability problems for continuous-time stochastic processes with discontinuous sample paths using the uniform topology (see Billingsley [6, Sect. 18]).

We close this introduction by mentioning a few other related papers, in particular, Anantharam [1], Bertsimas, Paschalidis, and Tsitsiklis [5], Chang, Heidelberger, Juneja, and Shahabuddin [8], Chen [11], Dobrushin and Pechersky [14], O'Connell [23], and Tsoucas [34].

2. TECHNICAL PRELIMINARIES

We shall work in the function space $D \equiv D([0,\infty), R)$ of right-continuous real-valued functions with left limits, endowed with the Skorohod [33] J_1 or M_1 topologies, or a modification of the M_1 topology denoted by M_1'; we refer to Billingsley [6], Lindvall [20], Puhalskii and Whitt [30], and Whitt [39] for details. (We take this opportunity to correct here a slip on p. 365 of [30]: The M_1' topology is stronger, not weaker, than the weak topology; that was the purpose of introducing it.) These spaces are metrizable as separable metric spaces and have Borel σ-fields coinciding with the usual Kolmogorov σ-field generated by the coordinate projections. We shall also use the subset $E^{\uparrow}$ of nondecreasing nonnegative functions x with $x(t) \to \infty$ as $t \to \infty$. We shall exploit continuity properties of standard functions on D such as addition. Continuity results for the most familiar J_1 topology are established in Whitt [39], but analogs also hold in the other topologies; e.g., such continuity results were established by Pomarede [24].

We say that a function $I(x)$ defined on a metric space S and taking values in $[0,\infty]$ is a rate function if the sets $\{x \in S : I(x) \le a\}$ are compact for all $a \ge 0$, and a sequence $\{P_n, n \ge 1\}$ of probability measures on the Borel σ-field of S (or a sequence of random elements $\{X_n, n \ge 1\}$ with values in S and distributions P_n) obeys the LDP with the rate function I if

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \log P_n(F) \le -\inf_{x\in F} I(x) \tag{2.1}$$

for all closed $F \subset S$, and

$$\underline{\lim_{n\to\infty}} \frac{1}{n} \log P_n(G) \ge -\inf_{x\in G} I(x) \tag{2.2}$$

for all open $G \subset S$. Here we call the LDP an FLDP if it is for a sequence of normalized processes in the function space D, i.e., given a stochastic process $(X(t), t \ge 0)$, the normalized processes are $(n^{-1}X(nt), t \ge 0), n \ge 1$. We refer to Dembo and Zeitouni [13], Puhalskii [25–29], Shwartz and Weiss [32], and Varadhan [35,36] for additional background. We remark that it is possible to express the LDP with incompatible topology and σ-field [13, p. 5], but we always use the Borel σ-field. So far we see little advantage in having a non-Borel σ-field. In particular, for applications in D we want the Kolmogorov σ-field. On D we could use a topology such as the uniform topology (corresponding to uniform convergence on bounded intervals) which makes the Kolmogorov σ-field non-Borel, but this does not seem helpful. On the other hand, whenever we have an FLDP with rate function that is equal to infinity at discontinuous elements of D, we can extend it to an FLDP for uniform topology and Kolmogorov σ-field (see, e.g., [27, Thm. C]). However, it is often important to consider rate functions that are finite for some discontinuous functions.

We establish new FLDPs from previously established ones by applying the contraction principle or an extension [25, 28, Sect. 2]. The contraction principle states that if $\{X_n, n \geq 1\}$ obeys an LDP with rate function I and if f is continuous, then $\{f(X_n), n \geq 1\}$ obeys an LDP with rate function

$$I'(y) \equiv \inf_{x:f(x)=y} I(x). \tag{2.3}$$

The extended contraction principle states that if $\{X_n, n \geq 1\}$ obeys an LDP with rate function I, if $\{f_n, n \geq 1\}$ is a sequence of measurable functions, if the function f is continuous when restricted to the sets $\{x : I(x) \leq a\}$, $a \geq 0$, and if $f_n(x_n) \to f(x)$ as $n \to \infty$ for all x_n for which $x_n \to x$ as $n \to \infty$ for all x for which $I(x) < \infty$, then $\{f_n(X_n), n \geq 1\}$ obeys an LDP with rate function (2.3). An important special case is $f_n = f$, as in the contraction principle, where f is continuous at each x with $I(x) < \infty$. In either case, if in addition f is a bijection, then we can write $I'(y) = I(f^{-1}(y))$. The applications here illustrate the importance of the extended contraction principle.

3. WAITING TIMES

We now establish FLDPs for waiting times in a single-server queue with unlimited waiting room. We use lowercase letters to define the basic random variables and associated capital letters for the associated normalized processes in the function space D. Let v_n be the service time of the nth customer and let u_n be the interarrival time between customers n and $(n+1)$. Let a 0th customer arrive at time 0 to find an empty system. (More general initial conditions can be treated too; e.g., [28, Rmk. 5].) Then the waiting time of the nth customer satisfies

$$w_n = [w_{n-1} + x_{n-1}]^+ = s_n - \min\{s_k : 0 \leq k \leq n\},$$

where $x_n = v_n - u_n$, $s_n = x_0 + \cdots + x_{n-1}$ and $s_0 = 0$. Hence, introducing the processes

$$U_n(t) = n^{-1} \sum_{i=1}^{\lfloor nt \rfloor} u_{i-1}, \quad V_n(t) = n^{-1} \sum_{i=1}^{\lfloor nt \rfloor} v_{i-1}, \quad W_n(t) = n^{-1} w_{\lfloor nt \rfloor}, \qquad t \geq 0, \tag{3.1}$$

we have

$$W_n = R(V_n - U_n), \qquad n \geq 1,$$

where R is the one-dimensional reflection map on D, defined by

$$R(x)(t) = x(t) - \inf\{x(s) : 0 \leq s \leq t\} \wedge 0, \qquad t \geq 0. \tag{3.2}$$

We start by stating a technical lemma about the reflection map, which is a version of Lemma 4.6 of [28]. For a function x of finite variation on bounded intervals, we denote by $x = x_1^l + x_2^l$ its Lebesgue decomposition with respect to Lebesgue measure, where x_1^l is the absolutely continuous component with $x_1^l(0) = 0$ and x_2^l is the singular component.

LEMMA 3.1: *Let a function x have finite variation on bounded intervals with x_2^l increasing. Then a function z is the reflection of x, i.e., $z = R(x)$ where R is the reflection map in Eq.* (3.2), *if and only if z is nonnegative, has finite variation on bounded intervals, $z_2^l = x_2^l$, $\dot{z}_1^l(t) = \dot{x}_1^l(t)$ a.e. on the set $\{z(t) > 0\}$ and $\dot{x}_1^l(t) \leq 0$ a.e. on the set $\{z(t) = 0\}$. Also $\dot{z}_1^l(t) = 0$ a.e. on the set $\{z(t) = 0\}$.*

Below we mostly apply the lemma to the case when x is absolutely continuous. Note that if we introduce

$$\dot{y}(t) = \dot{z}(t) - \dot{x}(t), \qquad t \geq 0, \qquad y(0) = 0, \tag{3.3}$$

then we get the standard characterization of the reflection map as $z = x + y$, where y is nondecreasing and increases only when $z(t) = 0$ (e.g., [17, p. 19]).

Let 1_A be the indicator function of the set A, which is 1 on A and 0 elsewhere. In the next theorem and below we set by definition $\infty \cdot 0 = 0$.

THEOREM 3.1:

(a) *If $\{(U_n, V_n), n \geq 1\}$ for U_n and V_n in Eq.* (3.1) *obeys an FLDP in $D \times D$ for (the product topology associated with) one of the topologies J_1, M_1, or M_1' with rate function $I_{U,V}$ where $I_{U,V}(u,v) = \infty$ if both u and v are discontinuous, then $\{W_n, n \geq 1\}$ in Eq.* (3.1) *obeys an FLDP in D for the same topology with rate function*

$$I_W(w) = \inf_{\substack{u,v \in D \times D: \\ w = R(v-u)}} \{I_{U,V}(u,v)\}. \tag{3.4}$$

(b) *In addition, suppose that $I_{U,V}(u,v) = I_U(u) + I_V(v)$, where I_U and I_V are integrals of* ([0,∞]-*valued, lower semicontinuous with compact level sets*) *local rate functions λ_U and λ_V, e.g.,*

$$I_U(u) = \int_0^\infty \lambda_U(\dot{u}(t))dt \tag{3.5}$$

for absolutely continuous nondecreasing u with $u(0) = 0$ and $I_U(u) = \infty$ otherwise, then for nonnegative absolutely continuous w with $w(0) = 0$

$$I_W(w) = \int_0^\infty 1_{\{w(t)>0\}} \inf_{z \geq 0 \vee \dot{w}(t)} \{\lambda_U(z - \dot{w}(t)) + \lambda_V(z)\}dt$$
$$+ \inf_{0 \leq z \leq y} \{\lambda_U(y) + \lambda_V(z)\} \int_0^\infty 1_{\{w(t)=0\}}dt, \tag{3.6}$$

while $I_W(w) = \infty$ otherwise.

(c) *If, in addition,*

$$\lambda_U(z) = \sup_{\alpha\in\mathbb{R}}\{\alpha z - \psi_U(\alpha)\} \quad \text{and} \quad \lambda_V(z) = \sup_{\alpha\in\mathbb{R}}\{\alpha z - \psi_V(\alpha)\}, \tag{3.7}$$

where $\psi_U(\alpha)$ and $\psi_V(\alpha)$ are convex, nondecreasing finite real-valued functions with $\psi_U(0) = \psi_V(0) = 0$, then for nonnegative absolutely continuous w with $w(0) = 0$

$$I_W(w) = \int_0^\infty 1_{\{w(t)>0\}} \sup_{\beta\in\mathbb{R}}\{\beta\dot{w}(t) - \psi_U(-\beta) - \psi_V(\beta)\}dt$$

$$+ 1_{\{\rho>1\}} \sup_{\beta\le 0}\{-\psi_U(-\beta) - \psi_V(\beta)\} \int_0^\infty 1_{\{w(t)=0\}}dt, \tag{3.8}$$

where $\rho = \dot{\psi}_V(0)/\dot{\psi}_U(0)$ with $\dot{\psi}_V(0)$ and $\dot{\psi}_U(0)$ denoting left and right derivatives, respectively, and $I_W(w) = \infty$ otherwise.

Proof: As $I_{U,V}(u,v)$ is infinite when both u and v are discontinuous, as subtraction $v - u$ is continuous when one of v and u is continuous by [39, Thm. 4.1] (the continuity is only established for the J_1 topology in [39], but Theorem 4.1 there holds for the other two topologies as well), and as the reflection map is continuous ([39, Thm. 6.4]), we can apply the extended contraction principle in [25, Thm. 2.2] (see also [28, Sect. 2, 30]) to obtain the FLDP for $\{W_n, n \ge 1\}$ in D for the same topology with rate function I_W in Eq. (3.4).

Turning to part (b), we apply Eq. (3.4), Lemma 3.1, and [28, Lem. 3.3] to obtain (with $\dot{y}$ from Eq. (3.3))

$$I_W(w) = \inf_{\substack{u,v\in D\times D\\ \dot{w}=\dot{v}-\dot{u}+\dot{y}}}\left\{\int_0^\infty [\lambda_U(\dot{u}(t)) + \lambda_V(\dot{v}(t))]\,dt\right\}$$

$$= \int_0^\infty \inf_{\substack{\dot{u},\dot{v}\\ \dot{w}=\dot{v}-\dot{u}+\dot{y}}} \{\lambda_U(\dot{u}(t)) + \lambda_V(\dot{v}(t))\}\,dt$$

$$= \int_0^\infty 1_{\{w(t)>0\}} \inf_{\substack{\dot{u},\dot{v}\\ \dot{w}=\dot{v}-\dot{u}}} \{\lambda_U(\dot{u}(t)) + \lambda_V(\dot{v}(t))\}\,dt$$

$$+ \int_0^\infty 1_{\{w(t)=0\}} \inf_{\substack{\dot{u},\dot{v}\\ 0\le\dot{v}\le\dot{u}}} \{\lambda_U(\dot{u}(t)) + \lambda_V(\dot{v}(t))\}\,dt,$$

from which Eq. (3.6) follows.

Finally, for (c), apply the argument of [28, Sect. 4] (which includes a minimax theorem on the second line, see, e.g., [2, Thm. 7, Ch. 6, Sect. 2]) to obtain, in analogy with (3.6),

$$\begin{aligned}
I_W(w) &= \int_0^\infty \inf_{\substack{\dot{u},\dot{v} \\ \dot{w}=\dot{v}-\dot{u}+\dot{y}}} \left\{ \sup_{\alpha,\beta\in\mathbf{R}} \{\alpha\dot{u}(t) - \psi_U(\alpha) + \beta\dot{v}(t) - \psi_V(\beta)\} \right\} dt \\
&= \int_0^\infty \sup_{\alpha,\beta\in\mathbf{R}} \left\{ \inf_{\substack{\dot{u},\dot{v}: \\ \dot{w}=\dot{v}-\dot{u}+\dot{y}}} \{\alpha\dot{u}(t) - \psi_U(\alpha) + \beta\dot{v}(t) - \psi_V(\beta)\} \right\} dt \\
&= \int_0^\infty 1_{\{w(t)>0\}} \sup_{\alpha,\beta\in\mathbf{R}} \left\{ \beta\dot{w}(t) - \psi_U(\alpha) - \psi_V(\beta) + \inf_{\dot{u}(t)\geq 0} \{(\alpha+\beta)\dot{u}(t)\} \right\} dt \\
&\quad + \int_0^\infty 1_{\{w(t)=0\}} \sup_{\alpha,\beta} \left\{ -\psi_U(\alpha) - \psi_V(\beta) + \inf_{\dot{u}(t)\geq\dot{v}(t)\geq 0} \{\alpha\dot{u}(t) + \beta\dot{v}(t)\} \right\} dt \\
&= \int_0^\infty 1_{\{w(t)>0\}} \sup_{\alpha+\beta\geq 0} \{\beta\dot{w}(t) - \psi_U(\alpha) - \psi_V(\beta)\} dt \\
&\quad + \sup_{\substack{\alpha\geq 0 \\ \alpha+\beta\geq 0}} \{-\psi_U(\alpha) - \psi_V(\beta)\} \int_0^\infty 1_{\{w(t)=0\}} dt,
\end{aligned}$$

which equals Eq. (3.8), because $\psi_U(\alpha)$ and $\psi_V(\beta)$ are nondecreasing. For the second term in Eq. (3.8), note that $\phi(\beta) \equiv -\psi_U(-\beta) - \psi_V(\beta)$ has left derivative $\dot{\phi}(0) = \dot{\psi}_U(0) - \dot{\psi}_V(0)$. If $\dot{\phi}(0) \geq 0$, then the supremum is attained at $\beta = 0$, and equals 0 since $\phi(0) = 0$. On the other hand $\dot{\phi}(0) < 0$ if and only if $\dot{\phi}_V(0) > \dot{\psi}_U(0)$. ■

Remark 3.1: The natural sufficient condition for $\{(U_n, V_n), n \geq 1\}$ to obey an FLDP for the M_1'-topology with rate function $I_{U,V}(u,v) = I_U(u) + I_V(v)$ is for $\{U_n, n \geq 1\}$ and $\{V_n, n \geq 1\}$ to be independent and separately obey FLDPs with rate functions $I_U(u)$ and $I_V(v)$. The familiar special case is an i.i.d. sequence: If $\{u_n\}$ is i.i.d. and $E\exp(\alpha u_1) < \infty$ for some $\alpha > 0$, then $\{U_n\}$ obeys an FLDP for the M_1'-topology with rate function

$$I_U(u) = \int_0^\infty \sup_{\alpha<\alpha^*} \{\alpha\dot{u}_1^\ell(t) - \log E\exp(\alpha u_1)\} dt + \alpha^* u_2^\ell(\infty), \tag{3.9}$$

where, as above, $u = u_1^\ell + u_2^\ell$ is the Lebesgue decomposition of u with u_1^ℓ being the absolutely continuous component with $u_1^\ell(0) = 0$, $\dot{u}_1^\ell(t)$ its derivative and u_2^ℓ is the singular component, $\alpha^* = \sup\{\alpha : E\exp(\alpha u_1) < \infty\}$; see [30, Eq. (6.5), 28, Lem. 3.2, 21, 22]. If $\alpha^* = \infty$, then $I_U(u) = \infty$ whenever x is not absolutely continuous and the FLDP holds for the J_1-topology; otherwise this is not the case. If $\alpha^* < \infty$, then we can have $I_U(u) < \infty$ for discontinuous u. We thus obtain an FLDP for the GI/GI/1 queue if both $E\exp(\alpha u_1) < \infty$ and $E\exp(\alpha v_1) < \infty$ for some α and one holds for all α. In order for the extra condition in Theorem 3.1(b) to hold, we thus need to have both finite for all α. Then, the condition of Theorem 3.1(c) holds as well. FLDPs for partial sums of dependent variables have also been established; e.g., see Chang [7], Dembo and Zajic [12], and Puhalskii [26, Cor. 6.6].

Remark 3.2: The conclusion of Theorem 3.1(c) is consistent with the more general formula given for the GI/GI/1 queue without proof at the end of [28, Sect. 1]. The more general formula allows both $I_U(u)$ and $I_V(v)$ to be finite for discontinuous arguments. Then $I_W(w)$ is finite for discontinuous arguments, too. It remains to prove the more general result.

Remark 3.3: We will show in Section 4 below that the rate function of the departure process has the form required for the arrival process in part (a), but not in parts (b) and (c), of Theorem 3.1.

Remark 3.4: We can also obtain an FLDP in D for the virtual waiting time process in a general single-server queue by essentially the same argument, extending the GI/GI/1 result in Puhalskii [28] (the key parts of the proof are already given in [28]).

Because the rate function I_W in Eq. (3.6) is infinite at discontinuous arguments, we can apply the extended contraction principle with the projection map to obtain an LDP for the sequences $\{w_{\lfloor nt \rfloor}/n, n \geq 1\}$ in $\mathbb{R}$. For sufficiently large t, we can directly read off (but not rigorously prove) an LDP for the steady-state waiting time distribution. (One cannot even guarantee the existence of a steady-state waiting time distribution under the hypotheses.)

COROLLARY 3.1: *If, in addition to the assumptions of Theorem* 3.1(*b*), λ_V *and* λ_U *are convex functions with* $\lambda_V(v) = \lambda_U(u) = 0$ *for some points* u *and* v *with* $0 \leq v \leq u$ *(as anticipated for* $\rho \leq 1$*), then for each* $t > 0$ $\{w_{\lfloor nt \rfloor}/n, n \geq 1\}$ *obeys an LDP in* $\mathbb{R}$ *with rate function*

$$I_{w(t)}(z) = \inf_{\substack{x \in D \\ z = x(t)}} \{I_W(x)\} = z\Big(\inf_{y \geq z/t} \{\lambda_X(y)/y\}\Big), \tag{3.10}$$

where

$$\lambda_X(z) = \inf_{y \geq 0} \{\lambda_U(y) + \lambda_V(y+z)\}. \tag{3.11}$$

If $\lambda_X(0) > 0$, *then, for all* t *sufficiently large,*

$$I_{w(t)}(z) = z\Big(\inf_{y \geq 0} \{\lambda_X(y)/y\}\Big). \tag{3.12}$$

PROOF: Under the assumptions, the second term in Eq. (3.6) vanishes. Next, a direct argument shows that the function λ_X in Eq. (3.11) is convex. Suppose that we stipulate that y is the measure of the set in $[0,t]$ on which $w(t) > 0$. Then, we can take the infimum in the first integral in Eq. (3.6) to obtain

$$\inf_{\substack{x \in D \\ z = x(t)}} \{I_W(x)\} = y\lambda_X(z/y). \tag{3.13}$$

We thus obtain Eq. (3.10) by taking the infimum of Eq. (3.13) over all y, $0 \leq y \leq t$, and replacing z/y by y. Formula (3.12) follows since $\lambda_X(y)/y \to \infty$ as $y \to 0$ if $\lambda_X(0) > 0$. ■

Remark 3.5: Note that Eq. (3.11) simplifies to $\lambda_X(z) = \lambda_U(v - z)$ when $z \leq v$ and $\lambda_X(z) = \infty$ when $z > v$ if the service times are deterministic with value v, and to $\lambda_X(z) = \lambda_V(u + z)$ if the interarrival times are deterministic with value u.

COROLLARY 3.2: *Under the assumptions of Theorem* 3.1(*c*), *Eq.* (3.11) *becomes*

$$\lambda_X(z) = \sup_{\beta \in \mathbf{R}} \{\beta z - \psi_U(-\beta) - \psi_V(\beta)\}. \tag{3.14}$$

If, in addition, $\psi_U(-\beta) + \psi_V(\beta) < 0$ *for some* β, *then*

$$I_{w(t)}(z) = zx^* \tag{3.15}$$

for t sufficiently large, where $x^* = \sup\{\beta : \psi_U(-\beta) + \psi_V(\beta) \leq 0\}$.

PROOF: The proof of Eq. (3.14) follows by applying a minimax theorem and was essentially carried out in the proof of Theorem 3.1(c). Equality (3.15) follows by substituting (3.14) into (3.12) and applying a minimax theorem if one notes that $\lambda_X(0) = 0$ if and only if $\psi_U(-\beta) + \psi_V(\beta) \geq 0$ for all β.

Remark 3.6: Note that the large t result in Corollary 3.2 is consistent with the result for the steady-state waiting time in Glynn and Whitt [16].

Remark 3.7: We can also apply the contraction principle to obtain an FLDP for the maximum waiting time process $M_n(t) = n^{-1} m_{\lfloor nt \rfloor}$, $t \geq 0$, where $m_n = \max\{w_k : 0 \leq k \leq n\}$. It is easy to see that the rate function has the same forms as Eqs. (3.6) and (3.8) for nondecreasing functions w. The rate function for m_n/n is the same as in Eq. (3.10) for $t = 1$. ■

We now apply Theorem 3.1 and our previous paper [30] to obtain an FLDP for waiting times in a queue with a superposition arrival process. We start with FLDPs for the component arrival times; we apply the inverse map to get FLDPs for the associated arrival counting processes; we add to get an FLDP for the aggregate superposition counting process; and we apply the inverse map once again to get an FLDP for the arrival times of the superposition process. This program parallels previous heavy-traffic FCLTs for queueing networks [18,39].

Assume that there are k component arrival processes and let u_i^j be the ith interarrival time in component process j. Let U_n^j be the normalized arrival time process for component process j, defined as in Eq. (3.1), and U_n be the normalized arrival time process for the superposition process.

For $x \in E^\uparrow$, we define the inverse function x^{-1} by $x^{-1}(t) = \inf(s > 0 : x(s) > t)$.

THEOREM 3.2: *Consider the* $\sum_{i=1}^k G_i/G/1$ *model in which the k component arrival processes and the service-time sequence are mutually independent. Assume that the processes* U_n^j *and* V_n *satisfy FLDPs in* $E^\uparrow$ *and D, respectively, with one of the* J_1, M_1, *or* M_1' *topologies and rate functions* I_{U^j} *and* I_V, $1 \leq j \leq k$.

(a) *If* $I_{U^j}(u) = \infty$ *for all u that are not strictly increasing, for all but one j,* $1 \leq j \leq k$, *then* $\{U_n\}$ *obeys an FLDP in* (D, M_1') *with rate function* I_U.

(b) *If $I_{U^j}(u) = \infty$ for all u that are not strictly increasing for all j, then $I_U(u) = \infty$ if u is not strictly increasing.*

(c) *If, in addition to the condition of part (a), either $I_{U^j}(u) = \infty$ for all discontinuous $u, 1 \le j \le k$, or $I_V(v) = \infty$ for all discontinuous v, then the conditions of Theorem* 3.1*(a) hold with $I_{U,V}(u,v) = I_U(u) + I_V(v)$.*

(d) *If, in addition to the condition of part (a),*

$$I_{U^j}(x) = \int_0^\infty \lambda_{U^j}(\dot{x}(t))\,dt, \qquad 1 \le j \le k, \tag{3.16}$$

and

$$I_V(x) = \int_0^\infty \lambda_V(\dot{x}(t))\,dt \tag{3.17}$$

with $I_{U^j}(x) = \infty$ and $I_V(x) = \infty$ if x is not absolutely continuous with $x(0) = 0$, where

$$\lambda_{U^j}(z) = \sup_{\alpha\in\mathbb{R}}\{\alpha z - \psi_{U^j}(\alpha)\} \quad \textit{and} \quad \lambda_V(z) = \sup_{\alpha\in\mathbb{R}}\{\alpha z - \psi_V(\alpha)\} \tag{3.18}$$

with ψ_{U^j} and ψ_V convex, nondecreasing finite real-valued functions with $\psi_{U^j}(0) = 0$ and $\psi_V(0) = 0$, then the conditions of Theorem 3.1*(c) hold with*

$$\psi_U(\alpha) = -\psi_N^{-1}(-\alpha), \qquad \psi_N(\alpha) = \psi_{N^1}(\alpha) + \cdots + \psi_{N^k}(\alpha) \tag{3.19}$$

and

$$\psi_{N^j}(\alpha) = -\psi_{U^j}^{-1}(-\alpha). \tag{3.20}$$

Proof: (a) We apply the inverse map and [30, Thm. 3.3] to get FLDPs in $E^\uparrow$ for the associated counting processes with rate functions I_{N^j}, which satisfy $I_{N^j}(x) = I_{U^j}(x^{-1})$. As $I_{N^j}(x) = \infty$ for discontinuous x for all but one j and addition is continuous in the M_1'-topology at summands with no common discontinuity (the proof is similar to the one in Pomarede [24]), we obtain an FLDP for the superposition process by applying the extended contraction principle with addition, yielding

$$I_N(x) = \inf_{\substack{(x_1,\ldots,x_k):\\ x=x_1+\cdots+x_k}} \{I_{N^1}(x_1) + \cdots + I_{N^k}(x_k)\}.$$

Finally, we obtain an FLDP for $\{U_n\}$ by applying the inverse map again. As we need not have $I_N(x) = \infty$ for strictly increasing x, we apply the contraction principle with [30, Thm. 3.3] to get the FLDP for $\{U_n\}$ in $E^\uparrow$ with the M_1' topology.

(b) Under the extra condition, $I_N(x) = \infty$ if x is not continuous, so that $I_U(x) = \infty$ if x is not strictly increasing.

(c) If, in addition, $I_{U^j}(u) = \infty$ for all discontinuous u, $1 \le j \le k$, then $I_{N^j}(x) = I_{U^j}(x^{-1}) = \infty$ for all x that are not strictly increasing. Hence, $I_N(x) = \infty$ for all x not strictly increasing, so that $I_U(x) = I_N(x^{-1}) = \infty$ for all discontinuous x. Hence, the condition here ensures that the conditions of Theorem 3.1(a) are satisfied.

(d) The conditions here imply the conditions in (c). The form of the rate function is determined by [15, Thm. 1] and [30, Thm. 3.4]. The argument is as in the proof in [30, Thm. 3.4]. ■

We now consider the case in which the interarrival times and service times are i.i.d.

THEOREM 3.3: *Consider the* $\sum_{i=1}^{k} GI_i/GI/1$ *model. Let* $E\exp(\alpha u_1^j) < \infty$, $1 \le j \le k$, *and* $E\exp(\alpha v_1) < \infty$ *for some* $\alpha > 0$.

(a) *An FLDP holds for* $\{U_n^j\}$ *in* $(E^{\uparrow}, M_1')$ *with* $I_{U^j}(x) = \infty$ *for all* x *that are not strictly increasing, as needed for Theorem* 3.2(*a*), *if and only if* $P(u_1^j = 0) = 0$.

(b) *In addition,* $I_{U^j}(x) = \infty$ *for all discontinuous* x *if and only if* $E\exp(\alpha u_1^j) < \infty$ *for all* α. *Then, Eqs.* (3.16) *and* (3.18) *hold with*

$$\psi_{U^j}(\alpha) = \log E\exp(\alpha u_1^j),$$

so that the conditions on U_n^j *in Theorem* 3.2(*c*) *hold.*

(c) *An FLDP holds for* $\{V_n\}$ *in* $(E^{\uparrow}, M_1')$ *with* $I_V(x) = \infty$ *for all discontinuous* x *if and only if* $E\exp(\alpha v_1) < \infty$ *for all* α.

PROOF: The general FLDPs allowing $I(x) < \infty$ for discontinuous x are stated in [30, Eqs. (6.4) and (6.5)]. See Eq. (3.9) for one. ■

4. DEPARTURE TIMES

In order to obtain LDPs for waiting times of queues within an acyclic network, we need to obtain an LDP for departure processes. We consider these processes as random elements of $E^{\uparrow}$. Recall that the departure time of the nth customer is

$$d_n = \sum_{i=1}^{n} u_{i-1} + w_n + v_n, \qquad n \ge 0. \tag{4.1}$$

Let D_n be the associated normalized process, defined by

$$D_n(t) = n^{-1} d_{\lfloor nt \rfloor}, \qquad t \ge 0. \tag{4.2}$$

THEOREM 4.1: *Assume that the conditions of Theorem* 3.1 (*a*) *hold with* $I_{U,V}(u,v) = \infty$ *if* v *either is not continuous or does not start at* 0. *Assume that the* U_n *have paths unbounded above with probability* 1.

(a) *Then* $\{(W_n, D_n), n \ge 1\}$ *obeys an FLDP in* $D \times E^{\uparrow}$ *for the same topology with rate function*

$$I_{W,D}(w,d) = \inf_{\substack{u,v:\\ w=R(v-u),\\ d=R(v-u)+u}} \{I_{U,V}(u,v)\}. \tag{4.3}$$

Moreover, $\{D_n, n \geq 1\}$ *obeys an FLDP in* $E^{\uparrow}$ *for the same topology with rate function*

$$I_D(d) = \inf_{\substack{u,v:\\ d=u+R(v-u)}} \{I_{U,V}(u,v)\}. \tag{4.4}$$

(b) *If, in addition, the conditions of Theorem* 3.1(*b*) *hold, then*

$$\begin{aligned} I_{W,D}(w,d) = &\int_0^{\infty} 1_{\{w(t)>0\}}[\lambda_U(\dot{d}(t) - \dot{w}(t)) + \lambda_V(\dot{d}(t))]\,dt \\ &+ \int_0^{\infty} 1_{\{w(t)=0\}}\Big[\lambda_U(\dot{d}(t)) + \inf_{0\leq y\leq \dot{d}(t)} \lambda_V(y)\Big]\,dt \end{aligned} \tag{4.5}$$

if w *and* d *are absolutely continuous,* w *is nonnegative with* $w(0) = d(0) = 0$ *and* $I_{W,D}(w,d) = \infty$ *otherwise.*

(c) *If, in addition, the conditions of Theorem* 3.1(*c*) *hold, then*

$$\begin{aligned} I_{W,D}(w,d) = &\int_0^{\infty} 1_{\{w(t)>0\}} \sup_{\alpha,\beta\in\mathbb{R}} \{(\alpha+\beta)\dot{d}(t) - \alpha\dot{w}(t) \\ &\qquad - \psi_U(\alpha) - \psi_V(\beta)\}\,dt \\ &+ \int_0^{\infty} 1_{\{w(t)=0\}} \sup_{\alpha\in\mathbb{R},\beta\leq 0} \{(\alpha+\beta)\dot{d}(t) - \psi_U(\alpha) - \psi_V(\beta)\}\,dt. \end{aligned} \tag{4.6}$$

Proof: The argument for part (a) follows the proof of Theorem 3.1(a) using $D_n = U_n + W_n + Z_n$, where

$$Z_n(t) = n^{-1}v_{\lfloor nt\rfloor}, \qquad t \geq 0.$$

We first note that D_n is a random element of $E^{\uparrow}$, as U_n has unbounded paths, and W_n and Z_n are nonnegative. Next, $Z_n(t)$ is dominated by the largest jump in V_n. Because, by the contraction principle, $\{V_n, n \geq 1\}$ obeys an FLDP with $I_V(v) = \infty$ if v is not continuous or $v(0) \neq 0$, $Z_n \xrightarrow{P^{1/n}} \theta$, where $\theta(t) = 0$, $t \geq 0$, and $P^{1/n}$ denotes super exponential convergence in probability [30]; see the proof of [30, Thm. 5.1]. (The argument there is for the J_1 topology, but essentially the same argument applies to the other topologies.) Hence, $\{(U_n, V_n, Z_n), n \geq 1\}$ obeys an FLDP with rate function

$$I_{U,V,Z}(u,v,z) = I_{U,V}(u,v) + \delta(z),$$

where $\delta(z) = 0$ if $z = \theta$ and $\delta(z) = \infty$ otherwise by [30, Lem. 4.1(b)]. As $D_n = V_n + (U_n - V_n)^{\uparrow} \vee 0 + Z_n$, where $x^{\uparrow}(t) \equiv \sup_{s\leq t} x(s)$, $I_{U,V}(u,v) = \infty$ when v is discontinuous, and both supremum and reflection map are continuous functions on D ([39, Thms. 6.2, 6.3, and 6.4]), applying the extended contraction principle concludes the proof of part (a). For part (b), we apply the reasoning in the proof of Theorem 3.1(b) to get

$$I_{W,D}(w,d) = \int_0^\infty \inf_{\substack{\dot u, \dot v: \\ \dot w = \dot v - \dot u + \dot y \\ \dot d = \dot w + \dot u}} \{\lambda_U(\dot u(t)) + \lambda_V(\dot v(t))\}\, dt,$$

which implies Eq. (4.5). Finally, part (c) follows directly from (b). ■

Remark 4.1: By Remark 3.1, the condition in Theorem 4.1(a) holds in the GI/GI/1 queue if $E\exp(\alpha u_1) < \infty$ for some $\alpha > 0$ and $E\exp(\alpha v_1) < \infty$ for *all* α. Theorems 3.2 and 3.3 provide sufficient conditions for the conditions in Theorem 4.1 to be satisfied in the $\sum_{i=1}^k G_i/G/1$ queue.

In general, an explicit expression for the rate function $I_D(d)$ in Eq. (4.4) seems difficult to obtain analytically from Eqs. (4.4)–(4.6). However, those expressions provide a basis for calculating the rate function $I_D(d)$ numerically. We also can deduce the form of the rate function in special cases. In particular, we now consider deterministic service times, as in de Veciana, Courcoubetis, and Walrand [37] and Chang [7]. The rate function of the departure times then is identical to the rate function of the arrival times, in the region where it is finite. The following result enables us to obtain FLDPs for waiting times at each queue of an acyclic network of queues with all service times deterministic.

COROLLARY 4.1: *If, in addition to the assumptions of Theorem* 4.1(*b*), *the service times are deterministic, so that* $\lambda_V(x) = 0$ *for* $x = v$ *and* $\lambda_V(x) = \infty$ *otherwise,* λ_U *is convex and* $\lambda_U(u) = 0$ *for some* $u \geq v$ ($\rho \leq 1$), *then*

$$I_D(d) = \int_0^\infty \lambda_D(\dot d(t))\, dt, \tag{4.7}$$

where

$$\lambda_D(z) = \begin{cases} \lambda_U(z), & z \geq v \\ \infty, & z < v. \end{cases} \tag{4.8}$$

PROOF: The infimum in Eq. (4.5) is attained by assigning $w(t) = 0$ for all t. To see this, note that we must have $\dot d(t) = v$ a.e. on the set $\{w(t) > 0\}$ in order for the rate function to be finite. On the set $\{t : w(t) > 0\}$ we must have the average of $\dot w(t)$ always nonnegative. Hence, by convexity of λ_U and the inequality $v \leq u$, the first integral is bounded below by

$$\lambda_U(v) \int_0^\infty 1_{\{w(t)>0\}} 1_{\{\dot d(t)=0\}}\, dt$$

so that taking $\{w(t) = 0\}$ gives the infimum of the rate function. ■

We now consider the case of deterministic interarrival times. This case illustrates that the infimum of Eq. (4.5) over w need not be attained at $w(t) = 0$, $t \geq 0$. This case also demonstrates that the rate function I_D cannot always be written as the integral of a local rate function λ_D that depends only on the derivative $\dot{d}(t)$. In this case we can use a local rate function that is a function of the two variables $\dot{d}(t)$ and $d(t) - ut$.

COROLLARY 4.2: *In addition to the assumptions of Theorem* 4.1(*b*), *assume that the interarrival times are deterministic, so that* $\lambda_U(x) = 0$ *for* $x = u$ *and* $\lambda_U(x) = \infty$ *otherwise, and that* $\lambda_V(v) = 0$ *for some* $v \leq u$. *Then* $I_D(d) = 0$ *when* $d(t) = ut$, $t \geq 0$, *and, provided that* $d(t) - ut \geq 0$ *for all* t,

$$I_D(d) = \int_0^{\infty} \lambda_D(d(t) - ut, \dot{d}(t))\, dt, \tag{4.9}$$

where

$$\lambda_D(y,z) = \lambda_V(z) 1_{\{y>0\}} + \inf_{0 \leq x \leq z} \lambda_V(x) 1_{\{y=0\}}. \tag{4.10}$$

Otherwise $I_D(d) = \infty$.

PROOF: As $\dot{w}(t) = 0$ a.e. on $\{w(t) = 0\}$, we must have $\dot{d}(t) - \dot{w}(t) = u$ a.e., so $w(t) = d(t) - ut$, hence, $d(t) \geq ut$. The result follows. ■

Remark 4.2: In Corollary 4.2, in order to obtain a finite rate function, the derivative $\dot{d}(t)$ is required to alternate between intervals on which it is equal to u and $d(t) = ut$ (which contribute 0 to $I_D(d)$) and intervals over which its averages starting from the left end point of the interval exceed u. The number of intervals on which $\lambda_D(\dot{d}(t))$ can be positive can be finite or infinite. To illustrate, if $\dot{d}(t) = 2u1_{[0,1)}(t) + u1_{[2,\infty)}(t)$, then $I_D(d) = \lambda_V(2u) + \lambda_V(0)$. On the other hand, if $\dot{d}(t) = 2u1_{[1,2)}(t) + u1_{[2,\infty)}(t)$, then $I_D(d) = \infty$.

Remark 4.3: Given FLDPs for the departure times in Theorem 4.1 and Corollaries 4.1 and 4.2, we obtain FLDPs for the corresponding continuous-time departure processes by applying [30].

We tend to have more arrivals when the interarrival times are small. The following result gives a general result for small interarrival times.

COROLLARY 4.3: *In the general setting of Theorem* 4.1(*b*), *if* λ_U *is convex,* $\lambda_U(u) = 0$ *for some* u *and* $\lambda_V(v)$ *for some* $v \leq u$, $\dot{d}(t) = c < u$ *for* $0 \leq t < t_0$ *and* $d(t) = u$ *for* $t \geq t_0$, *then*

$$I_D(d) = t_0 \lambda_D(c) \tag{4.11}$$

for

$$\lambda_D(z) = \lambda_U(z) + \inf_{0 \leq y \leq z} \lambda_V(y). \tag{4.12}$$

PROOF: Under the assumed condition on $\dot{d}(t)$, the infimum over w in Eq. (4.5) is attained at $w(t) = 0, t \geq 0$. ■

Remark 4.4: Under the conditions of Theorem 4.1(b), an upper bound for the rate function I_D is

$$I_D(d) \leq \int_0^\infty \lambda_D(\dot{d}(t))\,dt, \tag{4.13}$$

where λ_D is given in Eq. (4.12), which is obtained by having $w(t) = 0$ for all t. Corollary 4.2 shows that the infimum need not be attained at this expression. Nevertheless, Eq. (4.12) seems like a good basis for an approximate rate function. Thus, for the departure process from n queues in series, we suggest Eq. (4.13) as an *approximation* with

$$\lambda_{D_n}(z) \approx \lambda_U(z) + \sum_{i=1}^{n} \inf_{0 \leq y \leq z} \lambda_{V_i}(y), \tag{4.14}$$

where λ_{V_i} is the service-time local rate function at queue i.

The approximation (4.14) helps show how the service times can make the LDP behavior of the departure process different from the LDP behavior of the arrival process. Assuming that $\lambda_{V_i}(v_i) = 0$ for $0 < v_i < u$ for all i, we see that the likelihood of long service times play no role in long interdeparture times from the perspective of the LDP ($z > u$ in Eq. (4.14)) whereas the likelihood of short service times can influence short departure times ($z < v_i < u$ in Eq. (4.14)). As λ_{D_n} in Eq. (4.14) is increasing in n, we anticipate that large waiting times are less likely at later queues (given the same service-time distribution).

Remark 4.5: We can also establish FLDPs for the waiting times at all queues for n queues in series. For example, the appropriate continuous mapping for the waiting times at the second queue of two queues in series is

$$w_2 = R(v_2 - u_1 - R(v_1 - u_1))$$

which, using Lemma 3, can be expressed as

$$\dot{w}_2 = \dot{v}_2 - \dot{v}_1 - \dot{y}_1 + \dot{y}_2$$

where all the functions on the right side are nonnegative and $1_{\{w_i(t)>0\}}\,\dot{y}_i = 0$ for $i = 1, 2$. However, the challenge is to determine the rate functions.

5. THE DEPARTURE PROCESS OF COMPLETED WORK

Motivated by communication network models [7,10,37], in this section we consider an autonomous service model (which we will relate to the previous model). Let $a(t)$ denote the input in work and $s(t)$ the potential processing of work in the interval

$[0,t]$ for $t \geq 0$. We assume that $(a(t), t \geq 0)$ and $(s(t), t \geq 0)$ are nonnegative, nondecreasing stochastic processes. Assuming that the system starts empty, we define the workload at time t and the completed work in $[0,t]$ by

$$\ell(t) = R(a-s)(t), \qquad t \geq 0, \tag{5.1}$$

and

$$c(t) = a(t) - \ell(t), \qquad t \geq 0, \tag{5.2}$$

where R is the reflection map in Eq. (3.2). In the standard, single-server queueing model,

$$a(t) = \sum_{i=1}^{N(t)} v_{i-1}, \qquad t \geq 0, \tag{5.3}$$

where v_n is the service time of the nth customer,

$$N(t) = \max\left\{k : \sum_{i=1}^{k} u_{i-1} \leq t\right\}, \qquad t \geq 0, \tag{5.4}$$

u_n is the interarrival time between customers n and $n+1$ and $s(t) = t, t \geq 0$. The communication network models may have input of random jumps at random times as in Eq. (5.3) or input continuously at a random rate, or both. The communication network models typically have $s(t) = rt, t \geq 0$, for some constant r, but Chang and Zajic [10] have considered generalizations. When $s(t)$ is random, we can think of the server as working at a random rate.

Paralleling Eq. (3.1), we now introduce the normalized processes

$$\begin{aligned} A_n(t) &= n^{-1}a(nt), \qquad S_n(t) = n^{-1}s(nt), \\ L_n(t) &= n^{-1}\ell(nt), \qquad C_n(t) = n^{-1}c(nt). \end{aligned} \tag{5.5}$$

From Eqs. (5.1) and (5.2), we get

$$L_n = R(A_n - S_n) \quad \text{and} \quad C_n = A_n - R(A_n - S_n).$$

An FLDP for L_n was established in Puhalskii [28]. It was established for the GI/GI/1 model, but it is easily extended to the case in which an FLDP holds for (A_n, S_n) with appropriate conditions on the rate function. Hence, here we are primarily interested in the normalized completed work process C_n.

THEOREM 5:

(a) *If* $\{(A_n, S_n), n \geq 1\}$ *for* A_n *and* S_n *in Eq. (5.5) obeys an FLDP in* $D \times D$ *for (the product topology associated with) one of the topologies* J_1, M_1, *or* M_1' *with rate function* $I_{A,S}$ *where* $I_{A,S}(a,s) = \infty$ *if* s *is discontinuous, then* $\{L_n, n \geq 1\}$ *and* $\{C_n, n \geq 1\}$ *obey FLDPs in* D *for the same topology with rate functions*

$$I_L(\ell) = \inf_{\substack{(a,s)\in D\times D:\\ \ell=R(a-s)}} \{I_{A,S}(a,s)\} \tag{5.6}$$

and

$$I_C(c) = \inf_{\substack{(a,s)\in D\times D:\\ c=a-R(a-s)}} \{I_{A,S}(a,s)\}. \tag{5.7}$$

(b) *Assume, in addition, that $s(t) = rt, t \geq 0$, and $\{A_n, n \geq 1\}$ obeys the FLDP with rate function*

$$I_A(a) = \int_0^\infty \lambda_A(\dot{a}_1^l(t))\,dt + \alpha^* a_2^l(\infty), \tag{5.8}$$

where $\lambda_A(x) = \sup_{\alpha<\alpha^*}(\alpha x - \psi_A(\alpha))$ with $\psi_A(\alpha)$ convex taking on values in $(-\infty,\infty]$ finite in a neighborhood of 0, $\psi_A(0) = 0$ and $\alpha^* \in (0,\infty]$.

Let $m_A = \sup\{x : \lambda_A(x) = 0\}$ and $k(c) = \text{ess sup}\,\{t > 0 : \dot{c}(t) < r\}$. If $m_A < r$, then

$$I_C(c) = \int_0^\infty \lambda_A(\dot{c}(t))\,dt,$$

when c is absolutely continuous, $c(0) = 0$, $\dot{c}(t) \leq r$ a.e. and $k(c) = \infty$, and $I_C(c) = \infty$ otherwise. If $m_A \geq r$, then

$$I_C(c) = \int_0^{k(c)} \lambda_A(\dot{c}(t))\,dt,$$

when c is absolutely continuous, $c(0) = 0$, $\dot{c}(t) \leq r$ a.e. and $I_C(c) = \infty$ otherwise.

PROOF: Part (a) follows by continuity of the reflection map and the contraction principle. For part (b), first note that $I_{A,S}(a,s) = I_A(a)$ when $s = re$, where $e = (t, t \geq 0)$, and $I_{A,S}(a,s) = \infty$, otherwise so that $s = re$ in the infimum in Eq. (5.7). Next, by Lemma 3, if $c = a - R(a - re)$ for some a with $I_A(a) < \infty$, then c is absolutely continuous, $c(0) = 0$, $\dot{c}(t) \leq r$ a.e. and $a(t) \geq c(t)$. Moreover, Lemma 3 implies that $\dot{a}_1^l(t) = \dot{c}(t)$ a.e. on the event $\{a(t) = c(t)\}$ and $a(t) = c(t)$ a.e. on the event $\{\dot{c}(t) < r\}$. Hence, up to a set of Lebesgue measure zero,

$$\{t : \dot{c}(t) < r\} \subset \{t : \dot{a}_1^l(t) = \dot{c}(t)\} \cap \{t : a(t) = c(t)\}. \tag{5.9}$$

Let $k(c) = \infty$. We will now show that the infimum in Eq. (5.7) is attained at absolutely continuous a such that $\dot{a}(t) = \dot{c}(t)$ for almost all t. First, the definition of $k(c)$ and Eq. (5.9) imply that there exists a sequence $\{t_n, n \geq 1\}$ of numbers with $t_n \to \infty$ such that $a(t_n) = c(t_n)$. By the hypothesis on λ_A and Eq. (5.9), we have for $\alpha < \alpha^*$

$$\begin{aligned}
&\int_0^{t_n} \lambda_A(\dot a_1^l(t))1(\dot c(t) = r)\,dt + \alpha^* \int_0^{t_n} 1(\dot c(t) = r)\,da_2^l(t) \\
&\quad \ge \int_0^{t_n} (\alpha \dot a_1^l(t) - \psi_A(\alpha))1(\dot c(t) = r)\,dt + \alpha \int_0^{t_n} 1(\dot c(t) = r)\,da_2^l(t) \\
&\quad = \alpha \int_0^{t_n} 1(\dot c(t) = r)\,da(t) - \psi_A(\alpha) \int_0^{t_n} 1(\dot c(t) = r)\,dt \\
&\quad = \alpha a(t_n) - \alpha \int_0^{t_n} 1(\dot c(t) < r)\,da(t) - \psi_A(\alpha) \int_0^{t_n} 1(\dot c(t) = r)\,dt \\
&\quad = \alpha c(t_n) - \alpha \int_0^{t_n} 1(\dot c(t) < r)\dot c(t)\,dt - \alpha \int_0^{t_n} 1(\dot c(t) < r)\,da_2^l(t) - \psi_A(\alpha) \\
&\qquad \times \int_0^{t_n} 1(\dot c(t) = r)\,dt \\
&\quad = (\alpha r - \psi_A(\alpha)) \int_0^{t_n} 1(\dot c(t) = r)\,dt - \alpha \int_0^{t_n} 1(\dot c(t) < r)\,da_2^l(t)
\end{aligned}$$

with $\psi_A(\alpha) < \infty$. Therefore, as $\alpha < \alpha^*$,

$$\begin{aligned}
&\int_0^{t_n} \lambda_A(\dot a_1^l(t))\,dt + \alpha^* a_2^l(t_n) \\
&\quad \ge (\alpha r - \psi(\alpha)) \int_0^{t_n} 1(\dot c(t) = r)\,dt - \alpha \int_0^{t_n} 1(\dot c(t) < r)\,da_2^l(t) \\
&\qquad + \int_0^{t_n} \lambda_A(\dot a_1^l(t))1(\dot c(t) < r)\,dt + \alpha^* \int_0^{t_n} 1(\dot c(t) < r)\,da_2^l(t) \\
&\quad \ge (\alpha r - \psi_A(\alpha)) \int_0^{t_n} 1(\dot c(t) = r)\,dt + \int_0^{t_n} \lambda_A(\dot c(t))1(\dot c(t) < r)\,dt.
\end{aligned}$$

Taking on the right-most side supremum over $\alpha < \alpha^*$, we arrive at the inequality

$$\int_0^{t_n} \lambda_A(\dot a_1^l(t))\,dt + \alpha^* a_2^l(t_n) \ge \int_0^{t_n} \lambda_A(\dot c(t))\,dt,$$

which proves the claim.

Finally, let $k(c) < \infty$. Then, for $T > k(c)$, by Jensen's inequality,

$$\int_{k(c)}^{\infty} \lambda_A(\dot a(t))\,dt + \alpha^* a_2^l(\infty) \ge \lambda_A\left(\frac{a(T) - a(k(c))}{T - k(c)}\right)(T - k(c)).$$

As $a(T) \geq c(T) = r(T - k(c)) + c(k(c))$ for $m_A < r$, we have that, for all large T, $(a(T) - a(k(c)))/(T - k(c)) > m_A$ so that $\lambda_A((rT - a(k(c))/(\dot{T} - k(c))$ is bounded away from zero. This proves that $\int_{k(c)}^{\infty} \lambda_A(\dot{a}(t))\, dt + \alpha^* a_2^l(\infty) = \infty$.

If $m_A \geq r$, then, by the preceding argument and the fact that $\lambda_A(m_A) = 0$ (since $\lambda_A(\alpha)$ is lower semicontinuous), it is optimal to take $a(t) = c(t)$ for $t \leq k(c)$ and $\dot{a}(t) = m_A$ for $t \geq k(c)$ which implies $I_C(c) = \int_0^{k(c)} \lambda_A(\dot{c}(t))\, dt$.

Remark 5.1: It follows from [28, Lem. 4.3] that, for the GI/GI/1 model, if $P(u_n > 0) = 1$ and $E\exp(\theta u_n) < \infty$ and $E\exp(\theta v_n) < \infty$ for some $\theta > 0$, then $\{A_n, n \geq 1\}$ obeys an FLDP in (D, M_1') with rate function

$$I_A(a) = \int_0^{\infty} \sup_{\substack{\theta_2 < \beta^*, \theta_1 < \alpha^* \\ \beta(\theta_2) + \alpha(\theta_1) \leq 0}} \{\theta_2 \dot{a}_1^{\ell}(t) + \theta_1\}\, dt + \beta^* a_2^{\ell}(\infty),$$

where $\alpha^* = \sup\{\theta : E\exp(\theta u_n) < \infty\}$, $\beta^* = \sup\{\theta : E\exp(\theta v_n) < \infty\}$, $\alpha(\theta) = \log E\exp(\theta u_n)$ and $\beta(\theta) = \log E\exp(\theta v_n)$. To apply [28, Lem. 4.3], we apply the contraction principle with the coordinate projection map (noting that the infimum over a of the rate function $I_a^S(f) + I^A(a)$ there is not attained at the function a making $I^A(a) = 0$). In the special case of M/G/1, the input process $(A(t), t \geq 0)$ is compound Poisson, $\alpha(\theta) = \log(\lambda/(\lambda - \theta))$ and $\alpha^* = \lambda$ for some λ. Then

$$I_A(a) = \int_0^{\infty} \sup_{\theta < \beta^*} \{\theta \dot{a}_1^{\ell}(t) - \psi_a(\theta)\}\, dt + \beta^* a_2^{\ell}(\infty),$$

where $\psi_a(\theta) = \lambda(E\exp(\theta v_n) - 1)$. The FLDP for a compound Poisson process was obtained earlier by Lynch and Sethuraman [21]. ■

We can apply Theorem 5 to obtain an LDP in $\mathbb{R}$ for the departure process of completed work and an associated limit for the cumulant generating functions. This corollary gives the same answer as obtained by de Veciana et al. [38] and Chang [7] but under more general conditions (note that if in the conditions of Theorem 5 $\alpha^* = \infty$, then $\lambda_A(x)$ is an arbitrary nonnegative, convex and lower semicontinuous function with $\min_{x \in \mathbb{R}} \lambda_A(x) = 0$). In particular, now the cumulant generating function of the input process need not be finite everywhere.

Corollary 5: *Under the assumptions of Theorem* 5(*b*), $t^{-1}c(t)$ *satisfies an LDP in* $\mathbb{R}$ *as* $t \to \infty$ *with rate function*

$$I_c(z) = \begin{cases} \lambda_A(z), & z \leq r, \\ \infty, & z > r. \end{cases}$$

Moreover,

$$\lim_{t \to \infty} t^{-1} \log E e^{\theta c(t)} = \psi_c(\theta) = \begin{cases} \psi_a(\theta), & \theta \leq \hat{\theta}, \\ \psi_a(\hat{\theta}) + (\theta - \hat{\theta}) r, & \theta > \hat{\theta}, \end{cases} \tag{5.10}$$

where $\psi_a(\theta) = \sup_{z\in\mathbb{R}}(\theta z - \lambda_A(z))$ for λ_A from Eq. (5.8) *and $\hat{\theta} = \lambda'_A(r)$ with λ'_A denoting the left derivative (equivalently, $\psi'_a(\hat{\theta}) = r$).*

PROOF: Apply the contraction principle with the projection map to get the LDP in $\mathbb{R}$. Then apply Varadhan's integral lemma (noting that $c(t) \le rt$) to get Eq. (5.10). ■

6. SPLIT POINT PROCESSES

We now discuss random splitting, which can be regarded as the inverse of superposition. Random splitting arises in a queueing network when departures from one queue are routed to one of several other queues or depart from the network. Obtaining FLDPs for split processes enables us to obtain FLDPs for arrival processes within an acyclic network. Random splitting is an alternative to the deterministic routing of multiple streams through a queue considered by O'Connell [23].

Given a point process or counting process $(N(t), t \ge 0)$, let each successive point be randomly assigned one of k labels. Let Y_j be a k-dimensional random vector with a 1 in the ith place and 0's elsewhere if the jth point is assigned label i. Then the resulting k-dimensional counting process obtained from the splitting is

$$[N^1(t),\dots,N^k(t)] = \sum_{j=1}^{N(t)} Y_j, \qquad t \ge 0. \tag{6.1}$$

We have in mind i.i.d. splitting in which $(N(t), t \ge 0)$ is independent of $\{Y_j, j \ge 1\}$ and $\{Y_j, j \ge 1\}$ is i.i.d. Note that an independent splitting of a superposition process typically does not reproduce the original component processes. However, this does occur in the special case of independent Poisson processes with rates λ_i, $1 \le i \le k$, when $P(Y_j = i) = \lambda_i / \sum_{l=1}^{k} \lambda_l$.

Let N_n, Z_n, and N_n^i be the normalized processes defined by

$$N_n(t) = n^{-1}N(nt), \qquad t \ge 0, \tag{6.2}$$

$$Z_n(t) = n^{-1}\sum_{j=1}^{\lfloor nt \rfloor} Y_j, \qquad t \ge 0, \tag{6.3}$$

and

$$N_n^i(t) = n^{-1}N^i(nt), \qquad t \ge 0. \tag{6.4}$$

Also let ξ_j^i be the jth interval between points in the ith split stream and let X_n^i be the normalized process

$$X_n^i(t) = n^{-1}\sum_{j=1}^{\lfloor nt \rfloor} \xi_j^i, \qquad t \ge 0. \tag{6.5}$$

The key to establishing FLDPs for the processes $(N_n^1,\dots,N_n^k)$ and $(X_n^1,\dots,X_n^k)$ of interest is the recognition that they are related to previous processes by the composition and inverse maps, i.e.

$$(N_n^1,\dots,N_n^k) = Z_n \circ N_n,$$

where $\circ$ is the composition map as in [39, Sect. 3], and $X_n^i = (N_n^i + n^{-1})^{-1}$ as in [30, Eq. 7.4].

THEOREM 6.1: *Assume that $(N(t), t \geq 0)$ and $\{Y_j, j \geq 1\}$ are independent. Also assume that N_n in Eq. (6.2) and Z_n in Eq. (6.3) obey FLDPs in $E^{\uparrow}$ and $(E^{\uparrow})^k$ (with product topology), respectively, for one of the J_1, M_1, or M_1' topologies with rate functions I_N and I_Z, where either $I_N(x) = \infty$ for discontinuous or not strictly increasing x, or $I_Z(z) = \infty$ for $z = (z_1, \ldots, z_k)$ with at least one discontinuous component if the topology is J_1, $I_N(x) = \infty$ for discontinuous x and either $I_Z(z) = \infty$ for $z = (z_1, \ldots, z_k)$ with at least one discontinuous component or $I_N(x) = \infty$ for not strictly increasing x if the topology is M_1, $I_N(x) = \infty$ for discontinuous x and for x with $x(0) \neq 0$ and either $I_Z(z) = \infty$ for $z = (z_1, \ldots, z_k)$ with at least one discontinuous component or $I_N(x) = \infty$ for not strictly increasing x if the topology is M_1'. Then $(N_n^1, \ldots, N_n^k)$ in Eq. (6.4) obeys an FLDP in $(E^{\uparrow})^k$ for the product topology associated with the same topology with rate function*

$$I_{N^1,\ldots,N^k}(x_1, \ldots, x_k) = \inf_{\substack{z_1,\ldots,z_k,x: \\ x_j = z_j \circ x}} \{I_Z(z_1, \ldots, z_k) + I_N(x)\}. \tag{6.6}$$

PROOF: By the assumed independence, (Z_n, N_n) obeys an FLDP in $(E^{\uparrow})^k \times E^{\uparrow}$ with rate function

$$I_{Z,N}(z_1, \ldots, z_k, x) = I_Z(z_1, \ldots, z_k) + I_N(x). \tag{6.7}$$

Next, as in [30, Lem. 4.3], apply the extended contraction principle with the composition map to obtain Eq. (6.6) from (6.7). ■

COROLLARY 6.1: *If, in addition to the assumptions of Theorem 6.1, $\{Y_n, n \geq 1\}$ is i.i.d. with an assignment of label j with probability p_j and*

$$I_N(x) = \int_0^{\infty} \sup_{\alpha < \alpha^*} \{\alpha - \dot{x}(t)\psi(\alpha)\}\, dt, \tag{6.8}$$

when x is absolutely continuous with $x(0) = 0$ and $I_N(x) = \infty$ otherwise, where $\psi(0) = 0$ and $\psi(\alpha) < \infty$ in a neighborhood of 0, as is typical of renewal processes and superpositions of renewal processes (Puhalskii [28, Thm. 3.1], Puhalskii and Whitt [30, Thms. 6.1, 7.1]), then the FLDP holds for the J_1-topology and the rate function for the split process assumes the form

$$I_{N^1,\ldots,N^k}(x_1, \ldots, x_k) = \int_0^{\infty} \left[\sum_{j=1}^{k} \dot{x}_j(t) \log \frac{\dot{x}_j(t)}{p_j \sum_{i=1}^{k} \dot{x}_i(t)} + \sup_{\alpha < \alpha^*} \left\{ \alpha - \sum_{j=1}^{k} \dot{x}_j(t)\psi(\alpha) \right\} \right] dt,$$

when x_j is absolutely continuous with $x_j(0) = 0, 1 \leq j \leq k$, where $0 \log 0 = 0$, and $I_{N^1,\ldots,N^k}(x_1, \ldots, x_k) = \infty$ otherwise.

PROOF: For the case in which $\{Y_n, n \geq 1\}$ is i.i.d. with an assignment of label j with probability p_j,

$$Ee^{(s \cdot Y_1)} = \sum_{j=1}^{k} p_j e^{s_j}$$

so that $\{Z_n, n \geq 1\}$ in Eq. (6.3) obeys the FLDP in $E^{\uparrow}(J_1)$ with rate function

$$I_Z(z_1, \ldots, z_k) = \int_0^\infty \sup_{\alpha_1, \ldots, \alpha_k} \left\{ \sum_{j=1}^{k} \alpha_j \dot{z}_j(t) - \log \sum_{j=1}^{k} p_j e^{\alpha_j} \right\} dt,$$

when z_j is absolutely continuous with $z_j(0) = 0$, $1 \leq j \leq k$ and $I_Z(z_1, \ldots, z_k) = \infty$ otherwise (e.g., [27, Thm. 2.3] applies). Straightforward calculations yield

$$I_Z(z_1, \ldots, z_k) = \sum_{j=1}^{k} \int_0^\infty \dot{z}_j(t) \log \frac{\dot{z}_j(t)}{p_j}\, dt,$$

when z_j is absolutely continuous with $z_j(0) = 0$, $1 \leq j \leq k$ and $\sum_{j=1}^{k} \dot{z}_j(t) = 1$ a.e., and $I_Z(z_1, \ldots, z_k) = \infty$ otherwise.

Therefore, the infimum in Eq. (6.6) can be taken over x such that $x(t) = \sum_{j=1}^{k} x_j(t)$ so that the claim follows by the fact that if $x_j = z_j \circ x$, then

$$\int_0^\infty \sum_{j=1}^{k} \dot{z}_j(t) \log \frac{\dot{z}_j(t)}{p_j}\, dt = \int_0^\infty \sum_{j=1}^{k} \dot{z}_j(x(t))\dot{x}(t) \log \frac{\dot{z}_j(x(t))}{p_j}\, dt$$

$$= \int_0^\infty \sum_{j=1}^{k} \dot{x}_j(t) \log \frac{\dot{x}_j(t)}{p_j \sum_{j=1}^{k} \dot{x}_j(t)}\, dt. \qquad \blacksquare$$

In applications, we will be interested in the arrival process to another queue, which is one component of the vector $[N^1(t), \ldots, N^k(t)]$.

COROLLARY 6.2: *Under the assumptions of Corollary* 6.1, N_n^j *obeys an FLDP in* $E^{\uparrow}$ *for the* J_1*-topology with rate function*

$$I_{N^j}(x) = \int_0^\infty \sup_{\alpha < \alpha_j} (\alpha - \dot{x}(t)\psi_j(\alpha))\, dt,$$

where

$$\alpha_j = \sup\{\alpha : \psi(\alpha) < -\log(1 - p_j)\},$$

$$\psi_j(\alpha) = \psi(\alpha) + \log \frac{p_j}{1 - (1 - p_j)\exp\psi(\alpha)},$$

when x is absolutely continuous with $x(0) = 0$, $I_{N^j}(x) = \infty$ otherwise.

PROOF: The proof follows because, by Corollary 6.1, the contraction principle and the minimax theorem

$$I_{N^j}(x) = \inf_{\substack{x_1,\dots,x_k,\\ x_j=x}} I_{N^1,\dots,N^k}(x_1,\dots,x_k)$$

$$= \int_0^\infty \sup_{\alpha<\alpha^*} \inf_{\substack{\dot{x}_1,\dots,\dot{x}_k,\\ \dot{x}_j=\dot{x}}} \left[\sum_{l=1}^k \dot{x}_l(t) \log \frac{\dot{x}_l(t)}{p_l \sum_{i=1}^k \dot{x}_i(t)} + \alpha - \sum_{l=1}^k \dot{x}_l(t)\psi(\alpha) \right] dt. \quad ■$$

Note that the form of I_{N^j} is the same as if N were a renewal process.

Now vector analogs of [30, Thms. 3.1–3.3] can be applied to yield FLDPs for the sequence $\{(X_n^1,\dots,X_n^k), n \geq 1\}$ defined in Eq. (6.5), assuming that $I_{N^1,\dots,N^k}$ in Eq. (6.6) satisfies the conditions there. To illustrate, we state the result for the M_1' topology.

THEOREM 6.2: *Assume that the conditions of Theorem* 6.1 *hold for the M_1' topology. Then $(X_n^1,\dots,X_n^k)$ obeys the LDP in $(E^{\uparrow}(M_1'))^k$ with rate function*

$$I_{X^1,\dots,X^k}(x_1,\dots,x_k) = I_{N^1,\dots,N^k}(x_1^{-1},\dots,x_k^{-1}). \tag{6.9}$$

We conclude by considering FLDPs for centered processes. As in [30] we will work in the framework of triangular arrays for the initial point process, i.e., instead of a single point process $(N(t), t \geq 0)$ we consider a sequence of point processes $(N_n'(t), t \geq 0), n = 1,2,\dots$. The split processes $[N_n'^1(t),\dots,N_n'^k(t)]$ are still defined by Eq. (6.1) with $N_n'(t)$ substituted for $N(t)$, and the normalized processes are defined as in [30]: given $a_n > 0$,

$$N_n(t) = a_n^{-1} N_n'(a_n t), \qquad t \geq 0,$$

$$Z_n(t) = a_n^{-1} \sum_{j=1}^{\lfloor a_n t \rfloor} Y_j, \qquad t \geq 0,$$

and

$$N_n^i(t) = a_n^{-1} N'^i(a_n t), \qquad t \geq 0.$$

For appropriate normalizing constants c_n below, see [30, Thm. 6.2] and Corollary 6.3.

THEOREM 6.3: *Assume that $\{Y_j, j \geq 1\}$ and $(N_n'(t), t \geq 0)$ are independent for $n = 1,2,\dots$. Also assume that there are k-tuples $\lambda_n \to \lambda$ and constants $\mu_n \to \mu > 0$ and $c_n \to \infty$ such that $\{c_n(Z_n - e\lambda_n), n \geq 1\}$ and $\{c_n(N_n - e\mu_n), n \geq 1\}$ obey FLDPs in D^k (with product topology) and D for one of the J_1, M_1, and M_1' topologies with rate functions I_Z and I_N, respectively, where either $I_Z(z_1,\dots,z_k) = \infty$ if z_j is discontinuous for some j, $1 \leq j \leq k$, or $I_N(x) = \infty$ if x is discontinuous. Then*

$\{c_n[(N_n^1,\dots,N_n^k) - \mu_n e\lambda_n],\ n \geq 1\}$ obeys an FLDP in D^k for the same topology with rate function

$$I_{N^1,\dots,N^k}(x_1,\dots,x_k) = \inf_{\substack{(z_1,\dots,z_k,x)\in D^{k+1}:\\ x_i=z_i\circ\mu e+\lambda_i x}} \{I_Z(z_1,\dots,z_k) + I_N(x)\}. \tag{6.10}$$

Proof: Note that

$$c_n((N_n^1,\dots,N_n^k) - \mu_n e\lambda_n) = c_n(Z_n - e\lambda_n)\circ N_n + \lambda_n c_n(N_n - \mu_n e),$$

so that we can apply the maps $h_n(x,y,z) = x\circ y + \lambda_n z$ and $h(x,y,z) = x\circ y + \lambda z$ as in [39, Thm. 5.1(i)]. If, for one of the topologies J_1, M_1, or M_1', $x_n \to x$, $y_n \to y$ and $z_n \to z$, then, as $\lambda_n \to \lambda$, we have that $h_n(x_n, y_n, z_n) \to h(x,y,z)$ when y is continuous, strictly increasing and equals 0 at 0 and no discontinuities of $x\circ y$ coincide with discontinuities of z. (As above, though, the continuity is only established for the J_1 topology in [39], Theorem 4.1 carries over to the other two topologies as well, cf. [30, Lem. 4.3].) By the assumed independence, the pair $[c_n(Z_n - e\lambda_n), c_n(N_n - \mu_n e)]$ obeys an FLDP in $D^k \times D$ with rate function $I_Z + I_N$. By [30, Lem. 4.2(b)] and the convergence $\mu_n \to \mu, N_n \xrightarrow{p^{1/n}} \mu e$. By [30, Lem. 4.1(a,b)], $[c_n(Z_n - e\lambda_n), N_n, c_n(N_n - \mu_n e)]$ obeys an FLDP in $D^k \times D \times D$ with rate function $I_Z(z) + \delta(y - \mu e) + I_N(x)$, where $\delta(y - \mu e) = 0$ if $y = \mu e$ and ∞ otherwise. Finally, by the extended contraction principle, we obtain

$$I_{N^1,\dots,N^k}(x_1,\dots,x_k) = \inf_{\substack{(z_1,\dots,z_k,y,x)\in D^{k+2}:\\ x_i=z_i\circ y+\lambda_i x}} \{I_Z(z_1,\dots,z_k) + \delta(y - \mu e) + I_N(x)\}$$

which reduces to Eq. (6.10). ■

Corollary 6.3 *If, in addition to the assumptions of Theorem 6.3, $\{Y_n, n \geq 1\}$ is i.i.d. with an assignment of label j with probability p_j, $c_n = \sqrt{a_n/n}$ where $a_n/n \to \infty$ as $n \to \infty$, and*

$$I_N(x) = \frac{1}{2\sigma^2}\int_0^\infty \dot{x}(t)^2\, dt, \tag{6.11}$$

when x is absolutely continuous with $x(0) = 0$ and $I_N(x) = \infty$ otherwise, as is typical of renewal processes and superpositions of renewal processes (Puhalskii and Whitt [30, Thms. 6.2 and 7.2]), then the LDP holds for the J_1-topology and the rate function for the split process assumes the form

$$I_{N^1,\dots,N^k}(x_1,\dots,x_k) = \frac{1}{2\sigma^2}\int_0^\infty \left(\sum_{j=1}^k \dot{x}_j(t)\right)^2 dt$$
$$+ \frac{1}{2\mu}\int_0^\infty \left(\sum_{j=1}^k \frac{\dot{x}_j^2(t)}{p_j} - \left(\sum_{j=1}^k \dot{x}_j(t)\right)^2\right) dt,$$

when x_j is absolutely continuous with $x_j(0) = 0$, $1 \leq j \leq k$, and $I_{N^1,\dots,N^k}(x_1,\dots,x_k) = \infty$ otherwise.

PROOF: We take $\lambda_n = (p_1, \ldots, p_k)$ so that by [26, Cor. 6.7], $\{c_n(Z_n - e\lambda_n), n \geq 1\}$ obeys an FLDP for the J_1-topology with rate function

$$I_Z(z_1, \ldots, z_k) = \frac{1}{2} \int_0^\infty \sum_{j=1}^k \frac{\dot{z}_j(t)^2}{p_j}\, dt$$

if z_j is absolutely continuous with $z_j(0) = 0$, $1 \leq j \leq k$, and $\sum_{j=1}^k \dot{z}_j(t) = 0$ a.e., and $I_Z(z_1, \ldots, z_k) = \infty$ otherwise. Therefore, in the infimum in Eq. (6.10) $x(t) = \sum_{j=1}^k x_j(t)$ and $z_j(t) = x_j(\mu^{-1}t) - p_j \sum_{i=1}^k x_i(\mu^{-1}t)$. Substituting this into $I_Z(z_1, \ldots, z_k) + I_N(x)$ yields the result. ■

Straightforward minimization over the other components in $I_{N^1,\ldots,N^k}(x_1, \ldots, x_k)$ provides an LDP for one component of the vector $[N^1(t), \ldots, N^k(t)]$.

COROLLARY 6.4: *Under the assumptions of Corollary* 6.3, $\{c_n(N_n^j - \mu_n p_j e), n \geq 1\}$ *obeys an FLDP in D for the J_1-topology with rate function*

$$I_{N^j}(x) = \frac{1}{2(\mu p_j(1 - p_j) + p_j^2 \sigma^2)} \int_0^\infty \dot{x}(t)^2\, dt,$$

when x is absolutely continuous with $x(0) = 0$, $I_{N^j}(x) = \infty$ otherwise.

Note that $\mu p_j(1 - p_j) + p_j^2\sigma^2$ is "the variance per unit time" in the CLT for the jth component of the split process if the initial process obeys the CLT with mean μt and variance $\sigma^2 t$.

Again, vector analogs of [30, Thms. 5.1, 5.3, and 5.4] can be applied to give FLDPs for the centered processes of partial sums of time intervals between the events in the components of the split process.

Acknowledgment

The first author was partly supported by Bell Laboratories, first as part of AT&T and then as part of Lucent Technologies.

References

1. Anantharam, V. (1989). How large delays build up in a GI/G/1 queue. *Queueing Systems* 5: 345–367.
2. Aubin, J.-P. & Ekeland, I. (1984). *Applied nonlinear analysis.* New York: Wiley.
3. Berger, A.W. & Whitt, W. (1997). *A general framework for effective bandwidths with priority and loss criteria.* Florham Park, NJ: AT&T Labs.
4. Berger, A.W. & Whitt, W. (1998). *Effective bandwidths with priorities. IEEE/ACM Transactions on Networking*, to appear.
5. Bertsimas, D., Paschalidis, I., & Tsitsiklis, J. (1995). *On the large deviations behavior of acyclic networks of* G/G/1 *queue.* Cambridge, MA: MIT.
6. Billingsley, P. (1968). *Convergence of probability measures.* New York: Wiley.
7. Chang, C.S. (1995). Sample path large deviations and intree networks. *Queueing Systems* 20: 7–36.

8. Chang, C.S., Heidelberger, P., Juneja, S., & Shahabuddin, P. (1994). Effective bandwidth and fast simulation of ATM intree networks. *Performance Evaluation* 20: 45–66.
9. Chang, C.S. & Thomas, J.A. (1995). Effective bandwidths in high-speed digital networks. *IEEE Journal of Selected Areas in Communications* 13: 1091–1100.
10. Chang, C.S. & Zajic, T. (1995). Effective bandwidths of departure processes from queues with time-varying capacities. *Proceedings of IEEE INFOCOM '95*: 1001–1009.
11. Chen, H. (1996). Rate of convergence of the fluid approximation for generalized Jackson networks. *Journal of Applied Probability* 33: 804–814.
12. Dembo, A. & Zajic, T. (1995). Large deviations: From empirical mean and measure to partial sums processes. *Stochastic Processes and Their Applications* 57: 91–124.
13. Dembo, A. & Zeitouni, O. (1993). *Large deviations, techniques and applications.* Boston: Jones and Bartlett.
14. Dobrushin, R.L. & Pechersky, E.A. (1995). Large deviations for tandem queueing systems. *Journal of Applied Mathematics and Stochastic Analysis* 7: 301–330.
15. Glynn, P.W. & Whitt, W. (1994). Large deviations behavior of counting processes and their inverses. *Queueing Systems* 17: 107–128.
16. Glynn, P.W. & Whitt, W. (1994). Logarithmic asymptotics for steady-state tail probabilities in a single-server queue. *Journal of Applied Probability* 31A: 131–156.
17. Harrison, J.M. (1985). *Brownian motion and stochastic flow systems.* New York: Wiley.
18. Iglehart, D.L. & Whitt, W. (1970). Multiple channel queues in heavy traffic, I and II. *Advances in Applied Probability* 2: 150–177, 355–369.
19. Kelly, F.P. (1996). Notes on effective bandwidths. In F.P. Kelly, S. Zachary, & I. Ziedins (eds.), *Stochastic networks.* Oxford: Clarendon Press, pp. 141–168.
20. Lindvall, T. (1973). Weak convergence of probability measures and random functions in the function space $D[0,\infty)$. *Journal of Applied Probability* 10: 109–121.
21. Lynch, J. & Sethuraman, J. (1987). Large deviations for processes with independent increments. *Annals of Probability* 15: 610–627.
22. Mogulskii, A. (1993). Large deviations for processes with independent increments. *Annals of Probability* 21: 202–215.
23. O'Connell, N. (1997). Large deviations for departures from a shared buffer. *Journal of Applied Probability* 34: 753–766.
24. Pomarede, J.L. (1976). *A unified approach via graphs to Skorohod's topologies on the function space D.* Ph.D. dissertation, Yale University, New Haven, CT.
25. Puhalskii, A. (1991). On functional principle of large deviations. In V. Sazonov & T. Shervashidze (eds.), *New trends in probability and statistics*, Vol. 1, VSP/Mokslas, pp. 198–218.
26. Puhalskii, A. (1994). Large deviations of semimartingales via convergence of the predictable characteristics. *Stochastics* 49: 27–85.
27. Puhalskii, A. (1994). The method of stochastic exponentials for large deviations. *Stochastic Processes and Their Applications* 54: 45–70.
28. Puhalskii, A. (1995). Large deviation analysis of the single server queue. *Queueing Systems* 21: 5–66.
29. Puhalskii, A. (1997). Large deviations of semimartingales: A maxingale problem approach. I. Limits as solutions to a maxingale problem. *Stochastics* 61: 141–243.
30. Puhalskii, A. & Whitt, W. (1997). Functional large deviation principles for first-passage-time processes. *Annals of Applied Probability* 7: 362–381.
31. Ross, K.W. (1995). Multiservice loss models for broadband telecommunications. London: Springer.
32. Shwartz, A. & Weiss, A. (1995). Large deviations for performance analysis. London: Chapman and Hall.
33. Skorohod, A.V. (1956). Limit theorems for stochastic processes. *Theory of Probability and Its Applications* 1: 261–290.
34. Tsoucas, P. (1992). Rare events in series of queues. *Journal of Applied Probability* 29: 168–175.
35. Varadhan, S.R.S. (1966). Asymptotic probabilities and differential equations. *Communications in Pure Applied Mathematics* 19: 261–286.

36. Varadhan, S.R.S. (1984). Large deviations and applications. Philadelphia: SIAM.
37. de Veciana, G., Courcoubetis, C., & Walrand, J. (1994). Decoupling bandwidths for networks: A decomposition approach to resource management for networks. *Proceedings of IEEE INFOCOM '94* 2: 466–474.
38. de Veciana, G., Kesidis, G., & Walrand, J. (1995). Resource management in wide-area ATM networks using effective bandwidths. *IEEE Journal of Selected Areas in Communications* 13: 1081–1090.
39. Whitt, W. (1980). Some useful functions for functional limit theorems. *Mathematics of Operations Research* 1: 67–85.
40. Whitt, W. (1993). Tail probabilities with statistical multiplexing and effective bandwidths in multiclass queues. *Telecommunications Systems* 2: 71–107.