First- and Second-Derivative Estimators for Cyclic Closed-Queueing Networks

Gang Bao, Christos G. Cassandras, *Fellow, IEEE*, and Michael A. Zazanis

***Abstract*—We consider a cyclic closed-queueing network with arbitrary service time distributions and derive first- and second-derivative estimators of some *finite horizon* performance metrics with respect to a parameter of any one of the service distributions. Our approach is based on observing a single sample path of this system and evaluating first- and second-order effects on departure times as a result of the parameter perturbation. We then define an estimator as a conditional expectation over appropriate observable quantities, using smoothed perturbation analysis techniques. This process recovers the first-derivative estimator along the way and gives new insights into event order change phenomena which are of higher order. Despite the complexity of the analysis, the final algorithms we obtain are relatively simple. Further, we show that our estimators are unbiased and include some numerical examples. We also show the use of our estimators in obtaining approximations of the *entire system response surface* as a function of system parameters.**

I. Introduction

In dealing with stochastic discrete-event dynamic systems (DEDS), we are often faced with situations where the functional relationship between design or control parameters and performance metrics of interest are unknown. Still, by observing a single sample path of such a system (in a simulation or in an actual operating environment), it is often possible to estimate efficiently gradients of performance metrics with respect to various parameters. This can be accomplished through techniques such as perturbation analysis (PA) [21], [24], [15] and the likelihood ratio (LR) methodology [16], [22], [23]. These techniques provide an alternative to costly (sometimes infeasible in real-time) simulation where sensitivity estimation requires multiple sample path generations. In addition, they can often be integrated into gradient-based optimization algorithms (e.g., [8] and [9]) for problems of considerable complexity.

Techniques such as perturbation analysis can generally be used to estimate not only the first, but also higher-order derivatives of performance metrics with respect to some parameters. For example, for a GI/G/1 queueing system Zazanis and Suri [26] introduced the idea of using conditional expectation to calculate the expected effects of second-order contributions due to parameter perturbations along a sample path and obtained second derivative estimates for the customer-stationary flow time. Using the framework of smoothed perturbation analysis (SPA) [17], Fu and Hu [10] have derived a second-derivative estimator for the GI/G/m queue. Some recent work by Fu and Hu [11] has also extended SPA into a more general framework. However, there is little other work pursued along the lines of second-order derivative estimation, the main reason being that for complex systems, second-derivative estimators become difficult to obtain and are harder to implement in practice (compared to first-derivative estimators).

Recently, however, two developments have provided renewed motivation for seeking higher-order derivative estimates for performance metrics of DEDS. First, as first-derivative estimators are used in gradient-based optimization, we are often faced with the practical problem of instabilities in the form of large oscillations [8], [3]. To alleviate this problem, it is known that algorithms using second-derivative information may be used [4]. The second, perhaps more important, development is the emergence of Padé approximation techniques as viable means to accurately estimate the entire response surface of a complex system with respect to some parameter. Given some function $J(\theta)$, a Padé approximant is a rational function of the form $P_L(\theta)/Q_M(\theta)$, where $P_L(\theta)$ and $Q_M(\theta)$ are appropriately selected polynomials of degree L and M, respectively (see [1]). First- and higher-order derivative information at a single point θ provides one effective way to obtain the coefficients of these polynomials. As was recently shown in [18], Padé approximants of performance metrics of GI/G/1 systems show remarkable accuracy using first- and second-derivative information alone. This opens up a range of exciting possibilities for estimating global response surfaces of more complex systems based on information extracted from a single sample path observed under a parameter setting θ. Lastly, it is worth mentioning that a by-product of sample-path-based first- and second-derivative estimators is the fact that they sometimes lead directly to the establishment of structural properties of a system such as monotonicity or convexity/concavity of performance metrics with respect to parameters (if, for example, it turns out that the expectation of an unbiased such estimator is always positive/negative).

In this paper, we consider a closed-queueing network consisting of m servers connected in series, providing service to a fixed population of n customers. This is a model often used for serial transfer lines in manufacturing constrained to operate with a limited number of fixtures, n, or virtual circuits in communication networks with flow control limiting the

Manuscript received February 17, 1995; revised October 6, 1995 and March 5, 1996. Recommended by Associate Editor, E. K. P. Chong. This work was supported in part by the National Science Foundation under Grants ECS-8801912 and EID-9212122, by AFOSR under Contract F49620-95-1-0131, and by a grant from United Technologies/Otis Elevator Co.

G. Bao and C. G. Cassandras are with the Department of Electrical and Computer Engineering, University of Massachusetts, Amherst, MA 01003 USA.

M. A. Zazanis is with the Department of Industrial Engineering and Operations Research, University of Massachusetts, Amherst, MA 01003 USA.

Publisher Item Identifier S 0018-9286(96)05907-7.

number of packets to n and assuming no interfering traffic at any link. Other than some mild technical constraints, all service time distributions are arbitrary. PA techniques were first applied to this type of system in [20] to approximate first derivatives of the throughput. In [19] and [6] estimators for the throughput of closed Jackson queueing networks using infinitesimal perturbation analysis (IPA) were derived and extended by Cao [7] to general service time distributions. However, IPA cannot generally be applied to second-derivative estimation. The main reason is that IPA is based on limited information obtained from the observed sample path; to estimate second derivatives one needs significantly more information to be extracted from this same sample path to account for second-order effects in event order changes.

The main contribution of this paper is the derivation of unbiased second-derivative estimators for two types of performance metrics: *throughput* and *mean delay* between any two points in the closed serial-queueing network described above, over a finite time horizon. Our basic approach is to evaluate all second-order effects on departure times as a result of a parameter perturbation in a sample path of this network. Our analysis uses SPA techniques [17] in a way that provides new insights into the type of information on which one needs to condition. In particular, it is common in SPA to impose a conditioning on all past history of a process when an event occurs as well as the identity of the next event (e.g., see [14]); our approach is different in that our conditioning includes residual event lifetimes at the time of the next event, but not the time of the next event. As we will see, this allows us to obtain conditional expectations of performance metrics which overcome difficulties caused by event order changes on any given sample path. While the analysis is at times tedious and the form of the second-derivative estimator appears complex, its implementation, as we will show, turns out to be quite simple. Finally, even though we do not address the issue of consistency of the estimators we derive, we refer the reader to [2], where consistency is shown to hold for a two-node cyclic Markovian network for which second derivatives of the throughput at steady state can be analytically obtained.

The paper is organized as follows. In Section II, we set up the estimation problem and present some notation. In Section III, we derive a second-derivative estimator for the throughput; in the process, our approach recovers the first derivative which can also be obtained through standard IPA techniques. We also present an algorithm for implementing our first- and second-derivative estimators in Section IV. Some numerical examples are then presented in Section V. In Section VI, we show how our analysis is extended to first- and second-derivative estimators of mean delays. In Section VII, we present an application to the estimation of the throughput over all parameter values, based on the analysis recently provided in [18]. Finally, Section VIII contains a summary and discussion of future research in this area.

II. Notation and Estimation Problem Setup

We consider a serial closed-queueing network consisting of m servers (nodes) and a finite population of n customers. Each customer joins the queue in front of the next node as soon as it completes service at the current node, all queues have infinite capacity, and all nodes serve customers in FCFS fashion. Let $S_{i,k}$ denote the service time of the ith customer served at node k. We assume that the service times $\{S_{i,k}; i = 1, 2, \cdots\}$ are an i.i.d. sequence of random variables with distribution $F_k(\cdot), k = 1, \cdots, m$. The sequences $\{S_{i,k}; i = 1, 2, \cdots\}$, $k = 1, \cdots, m$, are also assumed independent.

Our objective is to estimate the first and second derivatives of the expected departure time of the Nth customer served at a node, say node 1, with respect to a parameter θ of the service time distribution of one of the nodes based on observations extracted from single sample path (the "nominal sample path"). Without loss of generality (as it will become apparent from our analysis), we assume that $\theta \in \Theta$ is a parameter of $F_1(\cdot)$ and Θ is an interval in $\boldsymbol{R}$.

Suppose that our probability space $(\Omega, \mathcal{F}, P)$ supports m sequences of i.i.d. random variables $\{U_{i,k}; i = 1, 2, \cdots, k = 1, \cdots, m\}$, uniformly distributed on [0, 1]. Let $F_1^{-1}(u, \theta) = \inf\{x\colon F_1(x, \theta) > u\}, F_k^{-1}(u) = \inf\{x\colon F_k(x) > u\}, k = 2, \cdots, m$. Thus, letting $S_{i,1}(\theta) = F_1^{-1}(U_{i,1}, \theta), S_{i,k} = F_k^{-1}(U_{i,k}), k = 2, \cdots, m$ defines a family of sample paths parameterized by θ. (For more details we refer the reader to [15] and [25].)

We are now ready to state the assumptions under which we carry out our analysis and derive the first- and second-derivative estimators. For the sake of (relative) simplicity we have not chosen the most general assumptions under which our results hold, but rather those that would simplify the proofs to the extent possible. In the remarks that follow we indicate ways to extend our results beyond the class of systems that satisfy the assumptions stated here.

Assumption A.1: $S_{i,1}(\theta)$ is an increasing function of θ, i.e., $\Delta S_{i,1} \overset{\text{def}}{=} S_{i,1}(\theta + \Delta\theta) - S_{i,1}(\theta) \geq 0$ w.p. 1, for $\Delta\theta \geq 0$.

This assumption simplifies the sample path analysis since it guarantees that a positive change in the parameter θ will result in positive perturbations. It can be relaxed via the approach described in [26, Sec. 8].

Assumption A.2: The derivative

$$\frac{\partial S_{i,1}}{\partial\theta}(\theta) = \lim_{\Delta\theta \to 0} \frac{S_{i,1}(\theta + \Delta\theta) - S_{i,1}(\theta)}{\Delta\theta}$$

exists and is a continuous function of $\theta \in \Theta$ w.p. 1. Furthermore, there exist positive constants c_1, c_2, such that

$$\frac{\partial S_{i,1}}{\partial\theta} \leq c_1 + c_2 S_{i,1} \text{ w.p.1}, \quad \text{for all} \quad \theta \in \Theta.$$

The above assumption (together with the mean value theorem) implies that

$$\frac{\Delta S_{i,1}}{\Delta\theta} \leq c_1 + c_2 S_{i,1}. \tag{1}$$

This assumption is introduced purely for convenience in our analysis. Note that it is a condition which is easy to verify for any given distribution and is satisfied by most commonly encountered parametric distribution families. In particular, it is always satisfied when θ is a scale or a location parameter (see [25]).

Assumption A.3: The second derivative

$$\frac{\partial^2 S_{i,1}}{\partial \theta^2}(\theta) = \lim_{\Delta\theta \to 0} \frac{1}{\Delta\theta}\left[\frac{\partial S_{i,1}}{\partial \theta}(\theta + \Delta\theta) - \frac{\partial S_{i,1}}{d\theta}(\theta)\right]$$

exists and is continuous for all $\theta \in \Theta$ w.p. 1. Furthermore

$$E\left|\sup_{\theta\in\Theta} \frac{\partial^2 S_{i,1}}{\partial \theta^2}(\theta)\right|^3 < \infty.$$

This is a technical condition required to prove unbiasedness of our estimators and is commonly encountered in the PA literature (e.g., see [15]). Its interpretation is that the second derivative $(\partial^2 S_{i,1}/\partial\theta^2)(\theta)$ should not be "excessively large." Note that for scale and location parameters, as in the case of Assumption A.2 above, this condition is automatically satisfied since $(\partial^2 S_{i,1}/\partial\theta^2)(\theta) = 0$ w.p. 1.

Assumption A.4: The distributions $F_k, k = 1, \cdots, m$, are absolutely continuous with density $f_k(t)$ and corresponding hazard rate $f_k(t)/1 - F_k(t)$ bounded above by γ for all $t \geq 0$. In particular, $f_1(t,\theta)/1 - F_1(t,\theta) \leq \gamma$ for all $\theta \in \Theta$.

This assumption may also be relaxed, though with considerable effort. The reader is referred to [12], [13], and [27].

The following notation will be needed in the sample path analysis of the following section:

$C_{i,k}$: the ith customer served at node k in the nominal path;

$e_{i,k}$: departure event of $C_{i,k}$ in the nominal path;

$A_{i,k}$: arrival time of $C_{i,k}$ in the nominal path;

$D_{i,k}$: departure time of $C_{i,k}$ in the nominal path;

where $i = 1, 2, \cdots, k = 1, \cdots, n$.

In this paper we focus on finite horizon performance metrics. With the above notation, $D_{N,1}$ is the time to observe N departures from node 1. The main performance metric we consider is

$$\overline{D}(\theta) = \frac{1}{N}E[D_{N,1}]$$

which can be thought of as the mean interdeparture time. (The term is justified if we assume that at $t = 0$ a departure occurs at node 1.) When the value of the parameter θ changes to $\theta + \Delta\theta$, the service times at node 1 are increased by $\Delta S_{i,1}(\theta), i = 1, \cdots, N$. This in turn causes $D_{N,1}$ to increase by $\Delta D_{N,1} = D_{N,1}(\theta + \Delta\theta) - D_{N,1}(\theta)$. In a nutshell, our analysis provides an expression for

$$\Delta\overline{D}(\theta) = \overline{D}(\theta + \Delta\theta) - \overline{D}(\theta) = \frac{1}{N}E[\Delta D_{N,1}].$$

As we will see in the next section, the first step toward this end is to show that $\Delta D_{N,1}$ can be expressed in the form $\Sigma_{(j,l)\in B_{N,1}} \Delta^1_{j,l}$, where (j,l) denotes the jth departure event at node l, $B_{N,1}$ is a subset of the events that have occurred up to time $D_{N,1}$ (defined later), and $\Delta^1_{j,l}$ a quantity related to the jth departure event at node l. Therefore, we have

$$\Delta\overline{D}(\theta) = \frac{1}{N}E\left[\sum_{(j,l)\in B_{N,1}} \Delta^1_{j,l}\right].$$

In general, $\Sigma_{(j,l)\in B_{N,1}} \Delta^1_{j,l}$ is not a "smooth" enough function of θ for the purpose of deriving second-derivative estimators. Thus, in the spirit of SPA [14], [17], we need to differentiate a corresponding conditional expectation. As we will see in the next section, the information on which it is necessary to condition consists of the age of various service processes at critical events as well as the length of certain idle periods and waiting times.

III. Derivation of Estimators

When θ is increased to $\theta + \Delta\theta$, we get a perturbed sample path. We use the superscript p to denote various quantities in the perturbed path (for example, $D^p_{i,k}$ denotes a departure time in the perturbed path). Further, we will assume that our m-node, n-customer system is such that $m > 1$ and $n > 1$, and we will always use $k+1$ to refer to the node where a customer departing from node k goes next, i.e., $k \pm 1 = (k \pm 1) \bmod m$.

The analysis that follows involves a significant amount of notation and becomes complicated at times. For ease of exposition, it is organized in a number of subsections.

A. Lindley Recursions for Perturbed Event Times

The following recursive equations describe the evolution of the nominal and perturbed paths:

$$\begin{aligned} D_{i,k} &= S_{i,k} + \max\{D_{i-1,k}, A_{i,k}\} \\ D^p_{i,k} &= S^p_{i,k} + \max\{D^p_{i-1,k}, A^p_{i,k}\} \end{aligned}$$

where $D_{0,k} = D^p_{0,k} = 0$ for all $k = 1, \cdots, m$. Defining $\Delta D_{i,k} = D^p_{i,k} - D_{i,k}$, we have

$$\Delta D_{i,k} = \Delta S_{i,k} + \max\{D^p_{i-1,k}, A^p_{i,k}\} - \max\{D_{i-1,k}, A_{i,k}\}. \tag{2}$$

Note that $A_{i,k} = D_{j,k-1}$ for some departure event $e_{j,k-1}$ at node $k-1$. We set $j = \hat{\imath}$ to denote the index of $C_{i,k}$ when this customer is at $k-1$, i.e., $C_{\hat{\imath},k-1}$ becomes $C_{i,k}$ immediately after this event. For consistency of notation, if there are initially n_k customers at node k, we simply set $\hat{\imath} = 0$ (i.e., $A_{i,k} = D_{0,k} = 0$) for all $i = 1, \cdots, n_k$.

Note that as a result of Assumption A.4, the probability that two events occur simultaneously is zero, and denote by $[x]^+ = \max\{0, x\}$ the positive part of the real number x. We now consider two cases in (2).

Case 1: $D_{i-1,k} > A_{i,k}$ (or $D_{i-1,k} > D_{\hat{\imath},k-1}$).

$$\begin{aligned} \Delta D_{i,k} &= \Delta S_{i,k} + \max\{D^p_{i-1,k}, A^p_{i,k}\} - D_{i-1,k} \\ &= \Delta S_{i,k} + (D^p_{i-1,k} - D_{i-1,k}) \\ &\quad + \max\{0, A^p_{i,k} - D^p_{i-1,k}\} \\ &= \Delta S_{i,k} + \Delta D_{i-1,k} + [A^p_{i,k} - D^p_{i-1,k}]^+ \\ &= \Delta S_{i,k} + \Delta D_{i-1,k} + [D^p_{\hat{\imath},k-1} - D^p_{i-1,k}]^+ \\ &= \Delta S_{i,k} + \Delta D_{i-1,k} \\ &\quad + [\Delta D_{\hat{\imath},k-1} - \Delta D_{i-1,k} - (D_{i-1,k} - D_{\hat{\imath},k-1})]^+ \\ &= \Delta S_{i,k} + \Delta D_{i-1,k} \\ &\quad + [\Delta D_{\hat{\imath},k-1} - \Delta D_{i-1,k} - W_{i,k}]^+ \end{aligned}$$

where $W_{i,k} = D_{i-1,k} - D_{\hat{\imath},k-1}$ is the *waiting time* of $C_{i,k}$.

Case 2: $D_{i-1,k} < A_{i,k}$ (or $D_{i-1,k} < D_{\hat{\imath},k-1}$).

$$\begin{aligned}\Delta D_{i,k} &= \Delta S_{i,k} + \max\{D^p_{i-1,k}, A^p_{i,k}\} - A_{i,k}\\ &= \Delta S_{i,k} + (A^p_{i,k} - A_{i,k}) + \max\{0, D^p_{i-1,k} - A^p_{i,k}\}\\ &= \Delta S_{i,k} + (D^p_{\hat{\imath},k-1} - D_{\hat{\imath},k-1}) + [D^p_{i-1,k} - D^p_{\hat{\imath},k-1}]^+\\ &= \Delta S_{i,k} + \Delta D_{\hat{\imath},k-1} + [D^p_{i-1,k} - D^p_{\hat{\imath},k-1}]^+\\ &= \Delta S_{i,k} + \Delta D_{\hat{\imath},k-1}\\ &\quad + [\Delta D_{i-1,k} - \Delta D_{\hat{\imath},k-1} - (D_{\hat{\imath},k-1} - D_{i-1,k})]^+\\ &= \Delta S_{i,k} + \Delta D_{\hat{\imath},k-1}\\ &\quad + [\Delta D_{i-1,k} - \Delta D_{\hat{\imath},k-1} - I_{i,k}]^+\end{aligned}$$

where $I_{i,k} = D_{\hat{\imath},k-1} - D_{i-1,k}$ is the *length of the idle period* terminated by the arrival of $C_{i,k}$ at node k.

Combining the above two cases we have

$$\begin{aligned}\Delta D_{i,k} &= \Delta S_{i,k} + (\Delta D_{i-1,k} + [\Delta D_{\hat{\imath},k-1}\\ &\quad - \Delta D_{i-1,k} - W_{i,k}]^+)\mathbf{1}(D_{i-1,k} > D_{\hat{\imath},k-1})\\ &\quad + (\Delta D_{\hat{\imath},k-1} + [\Delta D_{i-1,k} - \Delta D_{\hat{\imath},k-1} - I_{i,k}]^+)\\ &\quad \cdot \mathbf{1}(D_{i-1,k} < D_{\hat{\imath},k-1})\end{aligned}$$

where $\mathbf{1}(\cdot)$ denotes the indicator function.

Observing that the parameter θ affects only service times at node 1, we have $\Delta S_{i,k} = S^p_{i,k} - S_{i,k} = 0$ for $k \neq 1$, and $\Delta S_{i,1} = S^p_{i,1} - S_{i,1}$ generally $\neq 0$. Therefore

$$\begin{aligned}\Delta D_{i,k} &= \Delta S_{i,k}\mathbf{1}(k = 1) + (\tilde{W}_{i,k}\mathbf{1}(D_{i-1,k} > D_{\hat{\imath},k-1})\\ &\quad + \tilde{I}_{i,k}\mathbf{1}(D_{i-1,k} < D_{\hat{\imath},k-1}))\\ &\quad + (\Delta D_{i-1,k}\mathbf{1}(D_{i-1,k} > D_{\hat{\imath},k-1})\\ &\quad + \Delta D_{\hat{\imath},k-1}\mathbf{1}(D_{i-1,k} < D_{\hat{\imath},k-1}))\end{aligned}$$

where

$$\begin{aligned}\tilde{W}_{i,k} &= [\Delta D_{\hat{\imath},k-1} - \Delta D_{i-1,k} - W_{i,k}]^+\\ \tilde{I}_{i,k} &= [\Delta D_{i-1,k} - \Delta D_{\hat{\imath},k-1} - I_{i,k}]^+. \end{aligned} \tag{3}$$

We then define

$$\begin{aligned}\Delta^1_{i,k} &= \Delta S_{i,k}\mathbf{1}(k = 1) + \tilde{W}_{i,k}\mathbf{1}(D_{i-1,k} > D_{\hat{\imath},k-1})\\ &\quad + \tilde{I}_{i,k}\mathbf{1}(D_{i-1,k} < D_{\hat{\imath},k-1})\end{aligned} \tag{4}$$

and

$$\begin{aligned}\Delta^2_{i,k} &= \Delta D_{i-1,k}\mathbf{1}(D_{i-1,k} > D_{\hat{\imath},k-1})\\ &\quad + \Delta D_{\hat{\imath},k-1}\mathbf{1}(D_{i-1,k} < D_{\hat{\imath},k-1})\end{aligned} \tag{5}$$

so that

$$\Delta D_{i,k} = \Delta^1_{i,k} + \Delta^2_{i,k}. \tag{6}$$

B. The Induced Event Set $B_{i,k}$

We now express $\Delta^2_{i,k}$ in (5) recursively in terms of $\Delta^1_{j,l}$'s corresponding to events that occur prior to $e_{i,k}$. To facilitate this process we introduce the following terminology: we say that event $e_{i,k}$ is *induced* by another event $e_{j,l}$ if $e_{i,k}$ becomes a feasible event at time $D_{j,l}$. In particular, $e_{i,k}$ is either a) induced by $e_{i-1,k}$ if $D_{i-1,k} > D_{\hat{\imath},k-1}$ or b) induced by $e_{\hat{\imath},k-1}$ if $D_{i-1,k} < D_{\hat{\imath},k-1}$ (i.e., $e_{i,k}$ is the first departure in a busy period of node k initiated by $e_{\hat{\imath},k-1}$). Thus, looking at (5), we observe that $\Delta^2_{i,k} = \Delta D_{i-1,k}$, or $\Delta^2_{i,k} = \Delta D_{\hat{\imath},k-1}$, depending on whether $e_{i-1,k}$ or $e_{\hat{\imath},k-1}$ induces $e_{i,k}$. If $e_{i-1,k}$ induces $e_{i,k}$, (6) becomes

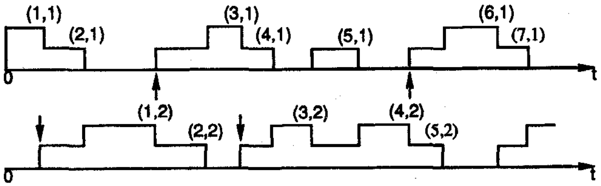


Fig. 1. Nominal sample path for a two-node two-customer system.

$$\Delta D_{i,k} = \Delta^1_{i,k} + \Delta D_{i-1,k} = \Delta^1_{i,k} + \Delta^1_{i-1,k} + \Delta^2_{i-1,k}.$$

The term $\Delta^2_{i-1,k}$ can again be expressed in terms of some $\Delta D_{j,l}$ such that $e_{j,l}$ induces $e_{i-1,k}$. By repeating this process backward in time until the start of the sample path, we can ultimately express $\Delta D_{i,k}$ as a sum of terms of the form $\Delta^1_{j,l}$. A similar process applies if $e_{\hat{\imath},k-1}$ induces $e_{i,k}$.

With this discussion in mind, we construct a set $B_{i,k}$ associated with event $e_{i,k}$ as follows.

1) $(i,k) \in B_{i,k}$.
2) If e_{i_1,k_1} induces $e_{i,k}$, then $(i_1,k_1) \in B_{i,k}$.
3) For all $j = 2, 3, \cdots$, if e_{i_j,k_j} induces $e_{i_{(j-1)},k_{(j-1)}}$ and $(i_{(j-1)},k_{(j-1)}) \in B_{i,k}$, then $(i_j,k_j) \in B_{i,k}$.
4) The procedure ends at the beginning of the sample path with $(i_s,k_s) \in B_{i,k}$ such that $D_{i_s,k_s} = \min\{D_{j,l}\colon (j,l) \in B_{i,k}\}$.

Therefore, $B_{i,k}$ is of the form

$$B_{i,k} = \{(i_s,k_s),(i_{(s-1)},k_{(s-1)}),\cdots,(i_2,k_2),(i_1,k_1),(i,k)\}$$

and thus, returning to the expression in (6), we can write

$$\Delta D_{i,k} = \sum_{(j,l)\in B_{i,k}} \Delta^1_{j,l}. \tag{7}$$

Example: In a two-node two-customer system, a typical nominal sample path is shown in Fig. 1 where (i,k) denotes the departure event $e_{i,k}$. Using the recursive construction of the set $B_{i,k}$ described above, we have

$$B_{7,1} = \{(1,1),(1,2),(3,1),(3,2),(4,2),(6,1),(7,1)\}.$$

Before proceeding, let us also expand $\Delta D_{i,k}$ into an easier-to-handle recursive form. In particular, we decompose the set $B_{i,k}$ into the following three subsets:

$$\begin{aligned}P_{i,k} &= \{(j,l) \in B_{i,k}\colon l = 1\}\\ Q_{i,k} &= \{(j,l) \in B_{i,k}\colon D_{j-1,l} > D_{\hat{\jmath},l-1}\}\\ R_{i,k} &= \{(j,l) \in B_{i,k}\colon D_{j-1,l} \leq D_{\hat{\jmath},l-1}\}.\end{aligned}$$

By combining (4) and (7), we then have

$$\Delta D_{i,k} = \sum_{(j,l)\in P_{i,k}} \Delta S_{j,l} + \sum_{(j,l)\in Q_{i,k}} \tilde{W}_{j,l} + \sum_{(j,l)\in R_{i,k}} \tilde{I}_{j,l}. \tag{8}$$

Clearly, $P_{i,k}$ simply contains all elements of $B_{i,k}$ corresponding to events at node 1, where service time perturbations are generated. An element (j,l) of $R_{i,k}$, on the other hand, corresponds to any event in $B_{i,k}$ which happens to be the *first departure in a busy period* at node l. Thus, $e_{j,l}$ with $(j,l) \in R_{i,k}$ is always induced by $e_{\hat{\jmath},l-1}$. We note that the set $R_{i,k}$ captures the well-known "perturbation propagation" phenomena due to idling at nodes as identified in early work on PA (e.g., [20]). All other events in $B_{i,k}$ not contained in $R_{i,k}$ belong to the set $Q_{i,k}$.

As an example, one can easily check that in Fig. 1

$$\begin{aligned}P_{7,1} &= \{(1,1),(3,1),(6,1),(7,1)\}\\ Q_{7,1} &= \{(4,2),(7,1)\}\\ R_{7,1} &= \{(1,1),(1,2),(3,1),(3,2),(6,1)\}.\end{aligned}$$

The two last terms on the right-hand side (RHS) of (8) correspond to the effect of "changes in the order of events." More specifically, the sum of $\tilde{I}_{j,l}$'s [defined in (3)] in the third term captures the effect of idle periods disappearing as a result of perturbations, while the sum of $\tilde{W}_{j,l}$'s in the second term captures the effect of idle periods being created as a result of perturbations. Using (8) we obtain the following expressions for these terms:

$$\tilde{I}_{j,l} = [\Delta_{j,l} + \tilde{T}_{j,l} - I_{j,l}]^+ \tag{9}$$

$$\tilde{W}_{j,l} = [-\Delta_{j,l} - \tilde{T}_{j,l} - W_{j,l}]^+ \tag{10}$$

where we have set

$$\Delta_{j,l} = \sum_{(q,r)\in P_{j-1,l}} \Delta S_{q,r} - \sum_{(q,r)\in P_{\hat{\jmath},l-1}} \Delta S_{q,r} \tag{11}$$

and

$$\tilde{T}_{j,l} = \left[\sum_{(q,r)\in Q_{j-1,l}} \tilde{W}_{q,r} - \sum_{(q,r)\in Q_{\hat{\jmath},l-1}} \tilde{W}_{q,r}\right] + \left[\sum_{(q,r)\in R_{j-1,l}} \tilde{I}_{q,r} - \sum_{(q,r)\in R_{\hat{\jmath},l-1}} \tilde{I}_{q,r}\right]. \tag{12}$$

Although these expressions appear to be prohibitively complicated, we will soon show that all $\tilde{T}_{j,l}$ terms do not contribute to the estimators we will derive.

C. The Critical Event Sets $R^*_{i,k}, Q^*_{i,k}$

We define two new sets, $Q^*_{i,k}$ and $R^*_{i,k}$, which exclude those elements (j,l) of $Q_{i,k}, R_{i,k}$ for which (as shown in subsequent sections) contributions of $E[\tilde{W}_{j,l}]$ and $E[\tilde{I}_{j,l}]$, respectively, are of order higher than $O(\Delta\theta^2)$ in the estimators we will derive

$$\begin{aligned}Q^*_{i,k} = \{&(j,l) \in B_{i,k}\colon D_{j-1,l} > D_{\hat{\jmath}l-1}\\ &\text{and no events occur in } (D_{\hat{\jmath},l-1}, D_{j-1,l})\}\\ R^*_{i,k} = \{&(j,l) \in B_{i,k}\colon D_{j-1,l} \le D_{\hat{\jmath}l-1}\\ &\text{and no events occur in } (D_{j-1,l}, D_{\hat{\jmath},l-1})\}.\end{aligned}$$

Recalling the definition of $R_{i,k}$ for example, it should be clear that an element (j,l) of $R^*_{i,k}$ corresponds to any event in $B_{i,k}$ which happens to be the first departure in a busy period at node l such that no event takes place in the preceding idle period $(D_{j-1,l}, D_{\hat{\jmath},l-1})$. As we will see, events such that $(j,l) \in B_{i,k}$ which are not "critical," i.e., not belonging to $R^*_{i,k}$ or $Q^*_{i,k}$, can be ignored as far as perturbation propagation is concerned.

Finally, define the set

$$\begin{aligned}\Gamma = \{&(j,l)\colon j = 1,2,\cdots,l = 1,\cdots,m\\ &\text{and no events occur in } (\min\{D_{j-1,l}, D_{\hat{\jmath},l-1}\},\\ &\max\{D_{j-1,l}, D_{\hat{\jmath},l-1}\})\}\end{aligned} \tag{13}$$

and observe that $R^*_{i,k} = R_{i,k} \cap \Gamma$, and $Q^*_{i,k} = Q_{i,k} \cap \Gamma$.

D. The Conditions $z_{j,l}$

Denote by $\mathcal{M}(t)$ the set of nodes that are busy at time t. The process $\{\mathcal{M}(t); t \ge 0\}$ taking values on the class of subsets of the set of nodes, $\{1,2,\cdots,m\}$, is assumed to have *right–continuous paths.* In particular, $\mathcal{M}(D_{j-1,l})$ is the set of nodes that are busy immediately after the $(j-1)$th departure from node l. Next, define

$$S^a_k(t) = \begin{cases} \textit{age} \text{ of service time of customer} & \\ \quad\text{present at node } k \text{ at time } t, & \text{if } k \in \mathcal{M}(t)\\ 0, & \text{if } k \notin \mathcal{M}(t)\end{cases}$$

$$S^r_k(t) = \begin{cases} \textit{residual} \text{ service time of customer} & \\ \quad\text{present at node } k \text{ at time } t, & \text{if } k \in \mathcal{M}(t)\\ 0, & \text{if } k \notin \mathcal{M}(t)\end{cases}$$

$$S_k(t) = \begin{cases} \textit{total service time} \text{ of customer} & \\ \quad\text{present at node } k \text{ at time } t, & \text{if } k \in \mathcal{M}(t)\\ 0, & \text{if } k \notin \mathcal{M}(t)\end{cases}$$

where, of course, $S_k(t) = S^a_k(t) + S^r_k(t)$. Hence, we define the families of processes $\{S^a_k(t); t \ge 0, k = 1,\cdots,m\}, \{S^r_k(t); t \ge 0, k = 1,\cdots,m\}, \{S_k(t); t \ge 0, k = 1,\cdots,m\}$, and again, we consider the *right-continuous* versions of these processes.

We now impose a sequence of conditions, $z_{j,l}$, associated with events $e_{j,l}$ as follows: Let $e'_{j,l}$ be the event immediately following $e_{j,l}$. Let $\Lambda_{j,l}$ denote all events and event times before $D_{j,l}$, as well as *the identity of the next event*, $e'_{j,l}$. In particular we point out that $\Lambda_{j,l}$ contains all *ages* of feasible events at $D_{j,l}$.

In addition, let $D'_{j,l}$ denote the time of the next event, $e'_{j,l}$. Then, we define the condition $z_{j,l}$ as follows:

$$z_{j,l} = \{\Lambda_{j,l}, \mathcal{M}(D'_{j,l}), S^r_q(D'_{j,l}); q \in \mathcal{M}(D'_{j,l})\}. \tag{14}$$

Note that, in addition to the history of the process up to the event time and the identity of the next event, $z_{j,l}$ includes the list of active nodes at the time the next event occurs, as well as the residual service times at the time of the next event. It does not, however, contain the *time of the next event*, $D'_{j,l}$. Readers familiar with the smoothed perturbation analysis (SPA) methodology will appreciate the importance of carefully selecting the conditions $z_{j,l}$: "just enough" information from the observed sample path is included in (14) to allow us to "smooth out" discontinuities that prevent a PA estimator from being unbiased. The justification for the specific choices made here will become clear in the analysis that follows.

Remark: In a simulation setting, conditioning on information such as the identity of the next event or residual service times presents no implementation problem, since this information is routinely available. In a real-time setting, on the other hand, future information is not available. This, however, presents no problem for the implementation of the estimators we will derive; it simply requires additional memory associated with certain events. For example, if one requires residual time data for certain active nodes when event $e_{j,l}$ occurs, one should wait until all corresponding service completions take place and then proceed with whatever computation is involved associated with $e_{j,l}$.

Returning to (8), our objective now becomes to replace the terms involving $\tilde{I}_{j,l}$ and $\tilde{W}_{j,l}$ by appropriate conditional expectations and explicitly evaluate them. Our main result, Theorem 1, will provide an expression for $E[\Delta D_{i,k}]$ in terms of these conditional expectations.

E. Conditional Expectations for Noncritical Events

We begin by considering all noncritical events in the induced event set $B_{i,k}$, i.e., those that do not belong to $R^*_{i,k}$ or $Q^*_{i,k}$ as defined in Section III-C. The following lemma shows that the contribution of such events is of order $o(\Delta\theta^2)$. This is done by providing bounds for the conditional expectations $E[\tilde{I}_{j,l}|\Lambda_{j-1,l}]$ and $E[\tilde{W}_{j,l}|\Lambda_{\hat{j},l-1}]$. Observe that in these conditional expectations *only the first part of the condition* defined in (14) is used. Thus, if $e_{j,l}$ is the first event in a busy period at node l, we condition only on $\Lambda_{j-1,l}$, where $e_{j-1,l}$ is the event that initiates the preceding idle period.

Lemma 1:

1) If at least one event occurs during an idle period of length $I_{j,l}$ (i.e., if $e_{\hat{j},l-1}$ is not the event immediately following $e_{j-1,l}$), then

$$\frac{1}{\Delta\theta^2}E[\tilde{I}_{j,l}|\Lambda_{j-1,l}] \leq K^I_{j,l}(\Delta\theta) \tag{15}$$

where $K^I_{j,l}(\Delta\theta)$ is such that $K^I_{j,l}(\Delta\theta) \in \Lambda_{j-1,l}, EK^I_{j,l}(\Delta\theta) < \infty$, and $\lim_{\Delta\theta\to 0} K^I_{j,l}(\Delta\theta) = 0$.

2) If at least one event occurs during the waiting time of length $W_{j,l}$ (i.e., if $e_{j-1,l}$ is not the event immediately following $e_{\hat{j},l-1}$), then

$$\frac{1}{\Delta\theta^2}E[\tilde{W}_{j,l}|\Lambda_{\hat{j},l-1}] \leq K^W_{j,l}(\Delta\theta) \tag{16}$$

where $K^W_{j,l}(\Delta\theta) \in \Lambda_{\hat{j},l-1}, EK^W_{j,l}(\Delta\theta) < \infty$ and $\lim_{\Delta\theta\to 0} K^W_{j,l}(\Delta\theta) = 0$.

Proof: See the Appendix.

The above lemma implies of course that the contribution of noncritical events is of order $o(\Delta\theta^2)$ and hence that these events will not play a role in the final expression for the second derivative of $E[D_{i,k}]$ and the resulting estimator.

F. Conditional Expectations for Critical Events

Next, we consider critical events in the induced event set $B_{i,k}$, i.e., events belonging to $R^*_{i,k}$ or $Q^*_{i,k}$ as defined in Section III-C. Before obtaining the analog of Lemma 1 for critical events, i.e., evaluating the corresponding conditional expectations, we first need to obtain some conditional densities. In particular, suppose event $e_{j-1,l}$ occurs at time $D_{j-1,l}$ and is followed by an idle period of duration $I_{j,l}$. In the following lemma, an expression for the conditional density of $I_{j,l}$ given $z_{j-1,l}, g^I_{j,l}(x|z_{j-1,l})$, is obtained for the case where no events occur during the idle period, or, equivalently, $e_{\hat{j},l-1}$ occurs immediately after $e_{j-1,l}$ (with an obvious dual result for $W_{j,l}$).

Lemma 2:

1) If event $e_{j-1,l}$ initiates an idle period of length $I_{j,l}$ at node l and the next event to occur is $e_{\hat{j},l-1}$, i.e., the event terminating the idle period, then the conditional density of this idle period given $z_{j-1,l}, g^I_{j,l}(x|z_{j-1,l})$, is shown in (17) at the bottom of the page.
2) If customer $C_{j,l}$ upon arrival to node l at time $D_{\hat{j},l-1}$ finds the server busy and the next event to occur is $e_{j-1,l}$, then the conditional density of this customer's waiting time given $z_{\hat{j},l-1}, g^W_{j,l}(x|z_{\hat{j},l-1})$, is shown in (18) at the bottom of the next page.

Proof: See the Appendix. ■

We now consider case 1) of Lemma 2 to obtain an analog of case 1) in Lemma 1. In this case, we consider the conditional expectation $E[\tilde{I}_{j,l}|z_{j-1,l}]$. Unlike Lemma 1, however, the condition here is $z_{j-1,l}$, not just $\Lambda_{j-1,l}$. In the next lemma we show that the contribution of critical events is no longer of order $o(\Delta\theta^2)$, but instead it depends on the conditional density function $g^I_{j,l}(\cdot|z_{j-1,l})$ and a quantity $Y^I_{j,l}$ defined next. Let $S^a_{\hat{j},l-1}$ denote the *age* of the service time of customer $C_{\hat{j},l-1}$ at the time $e_{j-1,l}$ occurs. In other words, using the definition of a service time age in Section III-D, we simply set $S^a_{\hat{j},l-1} = S^a_{l-1}(D_{j-1,l})$. As illustrated in Fig. 2, this corresponds to any event $e_{j,l}$ which is the first departure in a busy period of node l initiated by $e_{\hat{j},l-1}$ following an idle period of length $I_{j,l}$. The crucial observation here is that $S^a_{\hat{j},l-1}$

$$g^I_{j,l}(x|z_{j-1,l}) = \frac{f_{l-1}(S^a_{l-1}(D_{j-1,l})+x))\left[\prod_{\substack{q\in\mathcal{M}(D_{j-1,l})\\ q\neq l-1}} f_q(S_q(D_{j-1,l})-I_{j,l}+x)\right]}{\int_0^\infty f_{l-1}(S^a_{l-1}(D_{j-1,l})+u)\left[\prod_{\substack{q\in\mathcal{M}(D_{j-1,l})\\ q\neq l-1}} f_q(S_q(D_{j-1,l})-I_{j,l}+u)\right]du} \tag{17}$$

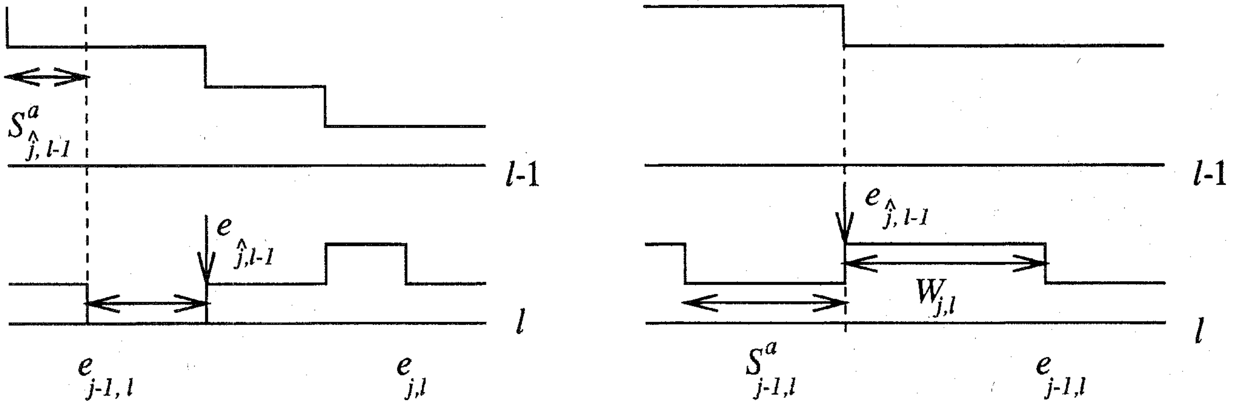


Fig. 2. Illustrating the definitions of $S^a_{\hat{j},l-1}$ and $S^a_{j-1,l}$.

belongs to $z_{j-1,l}$, the condition imposed in $E[\tilde{I}_{j,l}|z_{j-1,l}]$. Let us then define

$$Y^I_{j,l} = \sum_{(q,r)\in P_{j-1,l}} \frac{\partial}{\partial\theta} S_{q,r} - \sum_{\substack{(q,r)\in P_{\hat{j},l-1}\\ (q,r)\neq(\hat{j},l-1)}} \frac{\partial}{\partial\theta} S_{q,r} - \frac{\partial}{\partial\theta} S^a_{\hat{j},l-1}. \tag{19}$$

Similarly, we use $S^a_{j-1,l}$ to denote the *age* of the service time of customer $C_{j-1,l}$ at the time $e_{\hat{j},l-1}$ occurs (see Fig. 2). Accordingly, we define

$$Y^W_{j,l} = \sum_{(q,r)\in P_{\hat{j},l-1}} \frac{\partial}{\partial\theta} S_{q,r} - \sum_{\substack{(q,r)\in P_{j-1,l}\\ (q,r)\neq(j-1,l)}} \frac{\partial}{\partial\theta} S_{q,r} - \frac{\partial}{\partial\theta} S^a_{j-1,l}. \tag{20}$$

Lemma 3:

1) If no events occur in the system during an idle period of length $I_{j,l}$ (i.e., if the event that occurs immediately after $e_{j-1,l}$ is $e_{\hat{j},l-1}$), then

$$\lim_{\Delta\theta\to 0} \frac{1}{\Delta\theta^2} E[\tilde{I}_{j,l}|z_{j-1,l}] = \frac{1}{2} g^I_{j,l}(0|z_{j-1,l})([Y^I_{j,l}]^+)^2. \tag{21}$$

2) If no events occur in the system during a waiting period of length $W_{j,l}$ (i.e., if the event that occurs immediately after $e_{\hat{j},l-1}$ is $e_{j-1,l}$), then

$$\lim_{\Delta\theta\to 0} \frac{1}{\Delta\theta^2} E[\tilde{W}_{j,l}|z_{\hat{j},l-1}] = \frac{1}{2} g^W_{j,l}(0|z_{\hat{j},l-1})([Y^W_{j,l}]^+)^2. \tag{22}$$

Proof: See the Appendix.

G. Second-Derivative Estimators Using Conditional Expectations

Returning to (8) and taking expectations on both sides, we get

$$E[\Delta D_{i,k}] = E \sum_{(j,l)\in P_{i,k}} \Delta S_{j,l} + E \sum_{(j,l)\in Q_{i,k}} \tilde{W}_{j,l} + E \sum_{(j,l)\in R_{i,k}} \tilde{I}_{j,l}. \tag{23}$$

Combining the results from the previous sections, we will now replace the RHS above by terms involving conditional expectations. This leads to our main result, Theorem 1 below.

Theorem 1:

$$E[\Delta D_{i,k}] = E \sum_{(j,l)\in B_{i,k}} \Delta S_{j,l} \mathbf{1}(l=1) + E\left[\sum_{(j,l)\in R^*_{i,k}} \frac{1}{2} g^I_{j,l}(0|z_{j-1,l})([Y^I_{j,l}]^+)^2 + \sum_{(j,l)\in Q^*_{i,k}} \frac{1}{2} g^W_{j,l}(0|z_{\hat{j},l-1})([Y^W_{j,l}]^+)^2\right]\Delta\theta^2 + o(\Delta\theta^2). \tag{24}$$

$$g^W_{j,l}(x|z_{\hat{j},l-1}) = \frac{f_l(S^a_l(D_{\hat{j},l-1}) + x))\left[\prod_{\substack{q\in\mathcal{M}(D_{\hat{j},l-1})\\ q\neq l}} f_q(S_q(D_{\hat{j},l-1}) - W_{j,l} + x)\right]}{\int_0^\infty f_l(S^a_l(D_{\hat{j},l-1}) + u))\left[\prod_{\substack{q\in\mathcal{M}(D_{\hat{j},l-1})\\ q\neq l}} f_q(S_q(D_{\hat{j},l-1}) - W_{j,l} + u)\right] du} \tag{18}$$

Proof: Recalling (4) and (7), we have

$$\begin{aligned}\Delta D_{i,k} = \sum_{(j,l)\in B_{i,k}} \Delta^1_{j,l} &= \sum_{(j,l)\in B_{i,k}} \Delta S_{j,l}\mathbf{1}(l=1)\\ &+ \sum_{(j,l)\in B_{i,k}} \tilde{W}_{j,l}\mathbf{1}(D_{j-1,l} > D_{\hat{j},l-1})\\ &+ \sum_{(j,l)\in B_{i,k}} \tilde{I}_{j,l}\mathbf{1}(D_{j-1,l} < D_{\hat{j},l-1}).\end{aligned}$$

Comparing RHS above with the RHS of (24), note that the first summations are identical. Let us next consider the remaining two sums above. The last one can be written as

$$\begin{aligned}&\sum_{(j,l)\in B_{i,k}} \tilde{I}_{j,l}\mathbf{1}(D_{j-1,l} < D_{\hat{j},l-1})\\ &\quad = \sum_{l=1}^{m}\sum_{j=1}^{\infty} \tilde{I}_{j,l}\mathbf{1}(D_{j-1,l} < D_{\hat{j},l-1})\mathbf{1}[(j,l) \in B_{i,k}].\end{aligned}$$

Taking expectations and using Fubini's theorem, we obtain

$$\begin{aligned}&E \sum_{(j,l)\in B_{i,k}} \tilde{I}_{j,l}\mathbf{1}(D_{j-1,l} < D_{\hat{j},l-1})\\ &\quad = \sum_{l=1}^{m}\sum_{j=1}^{\infty} E[\tilde{I}_{j,l}\mathbf{1}(D_{j-1,l} < D_{\hat{j},l-1})\mathbf{1}((j,l) \in B_{i,k})].\end{aligned} \tag{25}$$

Recalling the definitions of the sets $R_{i,k}$ and $R^*_{i,k}$ in Sections B and C, we have

$$\mathbf{1}((j,l) \in B_{i,k})\mathbf{1}(D_{j-1,l} < D_{\hat{j},l-1}) = \mathbf{1}((j,l) \in R_{i,k}) \tag{26}$$

$$\mathbf{1}((j,l) \in R_{i,k})\mathbf{1}((j,l) \in \Gamma) = \mathbf{1}((j,l) \in R^*_{i,k}). \tag{27}$$

We next note that the information in $\Lambda_{j-1,l}$ is enough to determine whether $D_{j-1,l} < D_{\hat{j},l-1}$, and whether (j,l) belongs to Γ or not, that is

$$\mathbf{1}(D_{j-1,l} < D_{\hat{j},l-1}) \in \Lambda_{j-1,l} \subset z_{j-1,l} \tag{28}$$

and

$$\mathbf{1}((j,l) \in \Gamma) \in \Lambda_{j-1,l} \subset z_{j-1,l}. \tag{29}$$

The above two remarks are crucial in what follows. Equation (28) is simply due to the fact that $\Lambda_{j-1,l}$ includes the whole history of the process up to time $D_{j-1,l}$, so that by that time we obviously know whether $D_{\hat{j},l-1}$ has occurred (in which case the inequality is satisfied) or not. To check (29) we need to recall that $\Lambda_{j-1,l}$ contains the identity of the next event: If $D_{j-1,l} < D_{\hat{j},l-1}$, observe that no events occur in $(D_{j-1,l}, D_{\hat{j},l-1})$ iff $e_{\hat{j},l-1}$ is the next event. If on the other hand $D_{j-1,l} > D_{\hat{j},l-1}$, then we can tell whether any events occurred in $(D_{\hat{j},l-1}, D_{j-1,l})$ or not since this time interval clearly belongs to the past history of the process.

With these observations in mind, let us now return to the RHS of (25) and examine a typical term in the double summation. Noting that $\mathbf{1}((j,l) \in B_{i,k}) = \mathbf{1}((j,l) \in B_{i,k} \cap \Gamma) + \mathbf{1}((j,l) \in B_{i,k} \cap \Gamma^c)$, we have

$$\begin{aligned}&E[\tilde{I}_{j,l}\mathbf{1}((j,l) \in B_{i,k})\mathbf{1}(D_{j-1,l} < D_{\hat{j},l-1})]\\ &\quad = E[\mathbf{1}(D_{j-1,l} < D_{\hat{j},l-1})\tilde{I}_{j,l}\mathbf{1}((j,l) \in B_{i,k} \cap \Gamma)]\\ &\quad + E[\mathbf{1}(D_{j-1,l} < D_{\hat{j},l-1})\tilde{I}_{j,l}\mathbf{1}((j,l) \in B_{i,k} \cap \Gamma^c)].\end{aligned}$$

Next, take conditional expectations on the RHS of the above display, conditioning on $z_{j-1,l}$ for the first term above and on $\Lambda_{j-1,l} \subset z_{j-1,l}$ for the second term

$$\begin{aligned}&E[\tilde{I}_{j,l}\mathbf{1}((j,l) \in B_{i,k})\mathbf{1}(D_{j-1,l} < D_{\hat{j},l-1})]\\ &\quad = E[E[\mathbf{1}(D_{j-1,l} < D_{\hat{j},l-1})\\ &\qquad \cdot \tilde{I}_{j,l}\mathbf{1}((j,l) \in B_{i,k} \cap \Gamma)|z_{j-1,l}]]\\ &\quad + E[E[\mathbf{1}(D_{j-1,l} < D_{\hat{j},l-1})\\ &\qquad \cdot \tilde{I}_{j,l}\mathbf{1}((j,l) \in B_{i,k} \cap \Gamma^c)|\Lambda_{j-1,l}]]\\ &\quad = E[\mathbf{1}(D_{j-1,l} < D_{\hat{j},l-1})\\ &\qquad \cdot E[\tilde{I}_{j,l}\mathbf{1}((j,l) \in B_{i,k} \cap \Gamma)|z_{j-1,l}]]\\ &\quad + E[\mathbf{1}(D_{j-1,l} < D_{\hat{j},l-1})\\ &\qquad \cdot E[\tilde{I}_{j,l}\mathbf{1}((j,l) \in B_{i,k} \cap \Gamma^c)|\Lambda_{j-1,l}]]\end{aligned} \tag{30}$$

where the last step above follows from (28).

Examine now the two terms on the RHS of (30) separately, starting with the second term. First, note that $\tilde{I}_{j,l} \geq 0$ w.p. 1. Hence the conditional expectation in the second term is dominated by $E[\tilde{I}_{j,l}|\Lambda_{j-1,l}]$. By virtue of Lemma 1 this, in turn, is smaller than $K^I_{j,l}(\Delta\theta)\Delta\theta^2$. It follows from a straightforward application of the dominated convergence theorem (e.g., see [5]) that the second term is of order $o(\Delta\theta^2)$.

Turning our attention to the first term, we next argue that the random variables $\mathbf{1}((j,l) \in B_{i,k})$ and $\tilde{I}_{j,l}$ are conditionally independent on the event $\{(j,l) \in \Gamma\}$ given $z_{j-1,l}$. More precisely we will show that

$$\begin{aligned}&E[\tilde{I}_{j,l}\mathbf{1}((j,l) \in B_{i,k} \cap \Gamma)|z_{j-1,l}]\\ &\quad = \mathbf{1}((j,l) \in \Gamma)E[\tilde{I}_{j,l}|z_{j-1,l}]\\ &\qquad \cdot E[\mathbf{1}((j,l) \in B_{i,k})|z_{j-1,l}].\end{aligned} \tag{31}$$

To this end it will be helpful to think of the network as a continuous-time Markov process. Let $X_i(t)$ denote the number of customers in node i at time t and $X(t) := (X_1(t), \cdots, X_m(t))$. Similarly, $S^r_i(t)$ is the residual service time of the customer present in node i at time t (with $S^r_i(t) = 0$ when node i is idle) and $S^r(t) := (S^r_1(t), \cdots, S^r_m(t))$. Then $\xi(t) = (X(t), S^r(t))$, the vector of queue lengths and residual service times at time t, is a continuous time Markov process. We will denote by $\mathcal{F}^\xi_t := \sigma - \{\xi(u); 0 \leq u \leq t\}$ the history of the Markov process up to time t. Observe the following.

a) When $(j,l) \in \Gamma$, then the next event after $D_{j-1,l}$ is $D_{\hat{j},l-1}$ and thus $z_{j-1,l}$ is in this case the history of $X(t)$ up to $D_{j-1,l}$ together with $X(D_{\hat{j},l-1})$ and $S^r(D_{\hat{j},l-1})$. Thus, when $(j,l) \in \Gamma$, the full state of the Markov process at time $D_{\hat{j},l-1}$ belongs to $z_{j-1,l}$, i.e.,

$$\xi(D_{\hat{j},l-1}) \in z_{j-1,l} \subset \mathcal{F}^\xi_{D_{\hat{j},l-1}}. \tag{32}$$

b) Recalling the definition of $\tilde{I}_{j,l}$ in (9), we see that it clearly depends only on events that have occurred up to time $D_{j-1,l}$ and on the length of the idle period ending at time $D_{\hat{j},l-1}$, i.e., it depends only on events up to time $D_{\hat{j},l-1}$, so that

$$\tilde{I}_{j,l} \in \mathcal{F}^\xi_{D_{\hat{j},l-1}}. \tag{33}$$

c) The event $\{(j,l) \in B_{i,k}\}$ is determined from the segment of the sample path of $X(t)$ between $D_{j,l}$ and $D_{i,k}$. (Recall the construction of $B_{i,k}$: We start with the event $e_{i,k}$ and, going backward, find the event that induces it, then the one that induces that, and so forth.) Going backward in this fashion we can determine whether (j,l) belongs to $B_{i,k}$ or not from the information in $\sigma - \{X(u); D_{j,l} \le u \le D_{i,k}\}$. Therefore

$$\begin{aligned}\mathbf{1}\{(j,l) \in B_{i,k}\} &\in \sigma - \{X(u); u > D_{\hat{j},l-1}\}\\ &\subset \sigma - \{\xi(u); u > D_{\hat{j},l-1}\}\end{aligned} \tag{34}$$

since $D_{\hat{j},l-1} < D_{j,l}$.

d) $D_{\hat{j},l-1}$ is clearly a stopping time with respect to $\mathcal{F}_t^{\xi}$.

Based on a), b), c), d), and the strong Markov property, we can then see that, given $z_{j-1,l}, \tilde{I}_{j,l}$ and $\mathbf{1}((j,l) \in B_{i,k})$ are conditionally independent on Γ [i.e., (31) holds]. First, given $z_{j-1,l}$, we know $\xi(D_{\hat{j},l-1})$. Second, $\{(j,l) \in B_{i,k}\}$ is an event that depends on the future of the process, *after* $D_{\hat{j},l-1}$. Finally, $\tilde{I}_{j,l}$ depends on the evolution of the sample path of $X(t)$ *up to* $D_{\hat{j},l-1}$. Hence (31) follows from the fact that the future is conditionally independent from the past, given the present state $\xi(D_{\hat{j},l-1})$.

Next, to keep notation manageable, set

$$X_{j,l}^I = \tfrac{1}{2} g_{j,l}^I(0|z_{j-1,l})([Y_{j,l}^I]^+)^2$$

and recall Lemma 3 to write

$$E[\tilde{I}_{j,l}|z_{j-1,l}] = X_{j,l}^I \Delta\theta^2 + o(\Delta\theta^2) \text{ on } \{(j,l) \in \Gamma\}$$

where $X_{j,l}^I$ belongs to $z_{j-1,l}$. This in turn yields

$$\begin{aligned}&E[\tilde{I}_{j,l}\mathbf{1}((j,l) \in B_{i,k} \cap \Gamma)|z_{j-1,l}]\\ &\quad = \mathbf{1}((j,l) \in \Gamma)E[\mathbf{1}((j,l) \in B_{i,k})|z_{j-1,l}]X_{j,l}^I\Delta\theta^2\\ &\qquad + o(\Delta\theta^2).\end{aligned}$$

Therefore, returning to (25) and combining all of the above results

$$\begin{aligned}&E \sum_{(j,l)\in B_{i,k}} \tilde{I}_{j,l}\mathbf{1}(D_{j-1,l} < D_{\hat{j},l-1})\\ &\quad = \sum_{l=1}^{m}\sum_{j=1}^{\infty} E[E[\mathbf{1}((j,l) \in B_{i,k})|z_{j-1,l}]\mathbf{1}((j,l) \in \Gamma)\\ &\qquad \cdot X_{j,l}^I \mathbf{1}(D_{j-1,l} < D_{\hat{j},l-1})]\Delta\theta^2 + o(\Delta\theta^2)\\ &\quad = \sum_{l=1}^{m}\sum_{j=1}^{\infty} E[E[X_{j,l}^I\mathbf{1}((j,l) \in B_{i,k} \cap \Gamma)\\ &\qquad \cdot \mathbf{1}(D_{j-1,l} < D_{\hat{j},l-1})|z_{j-1,l}]]\Delta\theta^2 + o(\Delta\theta^2)\\ &\quad = \Delta\theta^2 E \sum_{l=1}^{m}\sum_{j=1}^{\infty} X_{j,l}^I \mathbf{1}((j,l) \in B_{i,k} \cap \Gamma)\\ &\qquad \cdot \mathbf{1}(D_{j-1,l} < D_{\hat{j},l-1}) + o(\Delta\theta^2)\\ &\quad = \Delta\theta^2 E \sum_{(j,l)\in R_{i,k}^*} X_{j,l}^I + o(\Delta\theta^2).\end{aligned}$$

Similarly, we can show that

$$\begin{aligned}&E \sum_{(j,l)\in B_{i,k}} \tilde{W}_{j,l}\mathbf{1}(D_{j-1,l} > D_{\hat{j},l-1})\\ &\quad = E \sum_{(j,l)\in Q_{i,k}^*} \tfrac{1}{2} g_{j,l}^W(0|z_{\hat{j},l-1})([Y_{j,l}^W]^+)^2\Delta\theta^2 + o(\Delta\theta^2)\end{aligned}$$

which completes the proof. ■

We are now in a position to obtain the desired first- and second-derivative estimators. Using Assumption A.3 and the dominated convergence theorem, (24) can be written as

$$\begin{aligned}E[\Delta D_{i,k}] = {} & E\left[\sum_{(j,l)\in P_{i,k}} \frac{\partial}{\partial\theta} S_{j,l}\right]\Delta\theta + \frac{1}{2}E\left[\sum_{(j,l)\in P_{i,k}} \frac{\partial^2}{\partial\theta^2} S_{j,l}\right.\\ &+ \sum_{(j,l)\in Q_{i,k}^*} g_{j,l}^W(0|z_{\hat{j},l-1})([Y_{j,l}^W]^+)^2\\ &\left.+ \sum_{(j,l)\in R_{i,k}^*} g_{j,l}^I(0|z_{j-1,l})([Y_{j,l}^I]^+)^2\right]\Delta\theta^2\\ &+ o(\Delta\theta^2).\end{aligned} \tag{35}$$

From the expression above, letting $(i,k) = (N,1)$ and recalling (7), we obtain the following estimator for the first-order derivative of $\overline{D}$

$$\begin{aligned}\left[\frac{\partial}{\partial\theta}\overline{D}\right]_{\text{est.}} &= \frac{1}{N} \lim_{\Delta\theta\to 0} \frac{1}{\Delta\theta} \sum_{(j,l)\in B_{N,1}} \Delta_{j,l}^1\\ &= \frac{1}{N} \sum_{(j,1)\in P_{N,1}} \frac{\partial}{\partial\theta} S_{j,1}.\end{aligned} \tag{36}$$

Thus, we have recovered the standard IPA first-derivative estimator for this type of network (e.g., [7]).

Similarly, our second-order derivative estimator of $\overline{D}$ is given by

$$\begin{aligned}\left[\frac{\partial^2}{\partial\theta^2}\overline{D}\right]_{\text{est.}} &= \frac{2}{N} \lim_{\Delta\theta\to 0} \frac{1}{\Delta\theta^2}\\ &\quad \cdot \left[\sum_{(j,l)\in B_{N,1}} \Delta_{j,l}^1 - \sum_{(j,1)\in P_{N,1}} \frac{\partial}{\partial\theta} S_{j,1}\Delta\theta\right]\\ &= \frac{1}{N}\left\{ \sum_{(j,1)\in P_{N,1}} \frac{\partial^2}{\partial\theta^2} S_{j,1}\right.\\ &\quad + \sum_{(j,l)\in Q_{N,1}^*} g_{j,l}^W(0|z_{\hat{j},l-1})\left(\left[\sum_{(q,1)\in P_{\hat{j},l-1}} \frac{\partial}{\partial\theta} S_{q,1}\right.\right.\\ &\quad \left.\left. - \sum_{\substack{(q,1)\in P_{j-1,l}\\ (q,1)\neq(j-1,l)}} \frac{\partial}{\partial\theta} S_{q,1} - \frac{\partial}{\partial\theta} S_{j-1,l}^a\right]^+\right)^2\\ &\quad + \sum_{(j,l)\in R_{N,1}^*} g_{j,l}^I(0|z_{j-1,l})\left(\left[\sum_{(q,1)\in P_{j-1,l}} \frac{\partial}{\partial\theta} S_{q,1}\right.\right.\end{aligned}$$

$$- \sum_{\substack{(q,1)\in P_{\hat{j},l-1}\\(q,1)\neq(\hat{j},l-1)}} \frac{\partial}{\partial\theta}S_{q,1} - \frac{\partial}{\partial\theta}S^a_{\hat{j},l-1}\Bigg]^+\Bigg)^2\Bigg\}. \tag{37}$$

It readily follows from Theorem 1 that the two derivative estimators above are indeed unbiased.

It is worth pointing out that in the second and third terms of (37) above, not all (j,l) belonging to the critical even sets $Q^*_{N,1}$ and $R^*_{N,1}$ contribute to the estimator. This is because the corresponding differences in the $[\cdot]^+$ in these terms may be negative. As an example, consider the two-node two-customer system of Fig. 1, where it is easy to check that $Q^*_{7,1} = Q_{7,1} = \{(4,2),(7,1)\}$. However, only $(4,2)$ has a nonzero contribution in the second term of (37), whereas for $(j,l) = (7,1)$ we get

$$\begin{aligned}\Bigg[\sum_{(q,1)\in P_{4,2}} \frac{\partial}{\partial\theta}S_{q,1} - \sum_{(q,1)\in P_{6,1},(q,1)\neq(6,1)} \frac{\partial}{\partial\theta}S_{q,1}\\ - \frac{\partial}{\partial\theta}S^a_{j-1,l}\Bigg]^+ = \Bigg[\left(\frac{\partial S_{3,1}}{\partial\theta} + \frac{\partial S_{1,1}}{\partial\theta}\right)\\ - \left(\frac{\partial S_{3,1}}{\partial\theta} + \frac{\partial S_{1,1}}{\partial\theta}\right) - \frac{\partial S^a_{6,1}}{\partial\theta}\Bigg]^+ = 0.\end{aligned}$$

An explicit algorithm for implementing the second-order derivative estimator above is provided in the next section.

IV. The Estimation Algorithm

Although the expression for the second-derivative estimator in (37) is rather complicated, we will present in this section an algorithm for implementing both first- and second-derivative estimators which is quite simple.

We begin with the first-derivative estimator in (36). Defining

$$L_1(i,k) = \sum_{(j,1)\in P_{i,k}} \frac{\partial}{\partial\theta}S_{j,1} \tag{38}$$

we have

$$\left[\frac{\partial}{\partial\theta}\overline{D}\right]_{\text{est.}} = \frac{1}{N}L_1(N,1).$$

Now, let e_{i_1,k_1} be the event that induces $e_{i,k}$. From the definition of $B_{i,k}$ in Section III-B, we know that

$$B_{i,k} = B_{i_1,k_1} \cup \{(i,k)\}; \quad P_{i,k} = P_{i_1,k_1} \cup \{(i,k)\mathbf{1}(k=1)\}$$

where we agree to let $\{(i,k)\mathbf{1}(k=1)\}$ be the empty set if $\mathbf{1}(k=1) = 0$. Therefore, we have the following iterative scheme for obtaining $L_1(i,k)$ from $L_1(i_1,k_1)$:

$$\begin{aligned}L_1(i,k) &= \sum_{(j,1)\in P_{i_1,k_1}\cup\{(i,k)\mathbf{1}(k=1)\}} \frac{\partial}{\partial\theta}S_{j,1}\\ &= L_1(i_1,k_1) + \mathbf{1}(k=1)\frac{\partial}{\partial\theta}S_{i,k}.\end{aligned} \tag{39}$$

We can obtain a similar iterative scheme for the second derivative. Define

$$L_2(i,k) = N\left[\frac{\partial^2}{\partial\theta^2}\overline{D}(\theta)\right]_{\text{est}} \tag{40}$$

and

$$\begin{aligned}L_1^a(i,k) &= \sum_{\substack{(j,1)\in P_{i,k}\\(j,1)\neq(i,k)}} \frac{\partial}{\partial\theta}S_{j,1} + \frac{\partial}{\partial\theta}S^a_{i,k}\\ &= L_1(i_1,k_1) + \mathbf{1}(k=1)\frac{\partial}{\partial\theta}S^a_{i,k}.\end{aligned} \tag{41}$$

Then, from the definition of $L_1(i,k)$ in (38), we have

$$\begin{aligned}L_2(i,k) = &\sum_{(j,1)\in P_{i,k}} \frac{\partial^2}{\partial\theta^2}S_{j,1}\\ &+ \sum_{(j,l)\in Q^*_{i,k}} g^W_{\hat{j},l-1}(0|z_{\hat{j},l-1})([L_1(\hat{j},l-1)\\ &- L_1^a(j-1,l)]^+)^2 + \sum_{(j,l)\in R^*_{i,k}} g^I_{j-1,l}(0|z_{j-1,l})\\ &\cdot([L_1(j-1,l) - L_1^a(\hat{j},l-1)]^+)^2\\ = &L_2(i_1,k_1) + \mathbf{1}(k=1)\frac{\partial^2}{\partial\theta^2}S_{i,1}\\ &+ \mathbf{1}[(i,k)\in Q^*_{i,k}]g^W_{\hat{i},k-1}(0|z_{\hat{i},k-1})\\ &\cdot([L_1(\hat{i},k-1) - L_1^a(i-1,k)]^+)^2\\ &+ \mathbf{1}[(i,k)\in R^*_{i,k}]g^I_{i-1,k}(0|z_{i-1,k})\\ &\cdot([L_1(i-1,k) - L_1^a(\hat{i},k-1)]^+)^2.\end{aligned} \tag{42}$$

The last expression above corresponds to the following three cases.

Case 1: $D_{i-1,k} > D_{\hat{i},k-1}$ and no event occurs during the waiting time of customer $C_{i,k}$, i.e., during the time interval $(D_{\hat{i},k-1}, D_{i-1,k})$.

It follows from the definition of $Q^*_{i,k}$ and $R^*_{i,k}$ that (42) yields:

$$\begin{aligned}L_2(i,k) = &L_2(i_1,k_1) + \mathbf{1}(k=1)\frac{\partial^2}{\partial\theta^2}S_{i,1}\\ &+ g^W_{\hat{i},k-1}(0|z_{\hat{i},k-1})\\ &\cdot([L_1(\hat{i},k-1) - L_1^a(i-1,k)]^+)^2.\end{aligned}$$

Case 2: $D_{i-1,k} \leq D_{\hat{i},k-1}$ and no event occurs during the idle period at node k, i.e., during the time interval $(D_{i-1,k}, D_{\hat{i},k-1})$. Then

$$\begin{aligned}L_2(i,k) = &L_2(i_1,k_1) + \mathbf{1}(k=1)\frac{\partial^2}{\partial\theta^2}S_{i,1}\\ &+ g^I_{i-1,k}(0|z_{i-1,k})\\ &\cdot([L_1(i-1,k) - L_1^a(\hat{i},k-1)]^+)^2.\end{aligned}$$

Case 3: Otherwise

$$L_2(i,k) = L_2(i_1,k_1) + \mathbf{1}(k=1)\frac{\partial^2}{\partial\theta^2}S_{i,1}.$$

Using (39) and the three cases in (42), we have the following.

First and Second Derivative Estimation Algorithm

1) Initialize: $L_1(1,k) := 0, L_1^a(1,k) := 0, L_2(1,k) := 0$, for $k = 1, \cdots, n$.

2) At each (departure) event $e_{i,k}$ induced by e_{i_1,k_1}:

a) $L_1(i,k) := L_1(i_1,k_1), L_2(i,k) := L_2(i_1,k_1)$.

b) If $k = 1$

$$L_1(i,k) := L_1(i,k) + \frac{\partial}{\partial\theta} S_{i,1}$$

$$L_2(i,k) := L_2(i,k) + \frac{\partial^2}{\partial\theta^2} S_{i,1}$$

$$L_1^a(i,k) := L_1(i,k) + \frac{\partial}{\partial\theta} S_{i,1}^a.$$

c) If $D_{i-1,k} > D_{\hat{i},k-1}$ and no event has occurred in $(D_{\hat{i},k-1}, D_{i-1,k})$ (i.e., if $C_{i,k}$ had to wait and no event occurred in the network during his waiting time)

$$L_2(i,k) := L_2(i,k) + g_{i,k-1}^W(0|z_{i,k-1}) \cdot ([L_1(\hat{i},k-1) - L_1^a(i-1,k)]^+)^2.$$

d) If $D_{i-1,k} \le D_{\hat{i},k-1}$ and no event has occurred in $(D_{i-1,k}, D_{\hat{i},k-1})$ (i.e., if $C_{i,k}$ did not have to wait and no event occurred in the network during the idle period in node k preceding his arrival)

$$L_2(i,k) := L_2(i,k) + g_{i-1,k}^I(0|z_{i-1,k}) \cdot ([L_1(i-1,k) - L_1^a(\hat{i},k-1)]^+)^2.$$

3) If N events have occurred at node 1, stop and set

$$L_1(N,1) := \frac{1}{N} L_1(N,1), L_2(N,1) := \frac{1}{N} L_2(N,1).$$

The calculations of $g_{i,k-1}^W(0|z_{i,k-1})$ and $g_{i-1,k}^I(0|z_{i-1,k})$ whenever required in the algorithm for a particular (j,l), are carried out by using (17) and (18). Observe, however, that the expressions for these conditional densities are predetermined off line. They may be quite complicated, but, once available, these on-line calculations are simply a matter of evaluating given expressions for specific numerical values of observed quantities such as customer waiting times. As noted in Section III-D, these calculations require the residual service times of all active nodes at $D_{j,l}$. Also, note that $g_{i-1,k}^I(0|z_{i-1,k})$ may be calculated by using (53) instead of (17), where the latter requires the length of observed idle periods, whereas the former requires residual service time observations. The calculation of the first and second derivatives of the service times $S_{i,1}(\theta)$ is done using standard techniques (e.g., see [26, Sect. II]).

Finally, when the algorithm stops we have

$$L_1(N,1) = \left[\frac{\partial \overline{D}(\theta)}{\partial\theta}\right]_{\text{est.}}, \quad L_2(N,1) = \left[\frac{\partial^2 \overline{D}(\theta)}{\partial\theta^2}\right]_{\text{est}}.$$

V. Simulation Results

Here we present numerical results obtained by applying the algorithm of Section IV to two different serial closed-queueing networks. The first simulation experiment was conducted for a network with exponential service times to compare the output of our algorithm with analytical results. The second experiment was conducted for a network with hyperexponential service times (H_2). In this case, we compare our estimators with "brute force" simulation results, i.e., finite-difference estimators for a given value of $\Delta\theta$. In what follows, the quantities with the subscript "*est*" represent estimates which are compared to the corresponding analytical results or brute force simulation results (with subscript "*b*"). Finally, $\overline{D}'(\theta)$ and $\overline{D}''(\theta)$ represent the first and second derivatives with respect to parameter θ. The 95% confidence intervals included were obtained from a batch of 25 simulation runs in each case. Additional numerical results may be found in [2].

Case 1: Exponential system with three nodes and five customers.

N	D	$D_{est.}$	$D'(\theta)$	$D'(\theta)_{est.}$	$D''(\theta)$	$D''(\theta)_{est.}$
case 1						
10^3	1.40000	1.40357 ±0.01224	0.46667	0.47230 ±0.02004	0.62222	0.53745 ±0.13390
10^4	1.40000	1.39952 ±0.00417	0.46667	0.46411 ±0.00525	0.62222	0.61406 ±0.03369
10^5	1.40000	1.39951 ±0.00107	0.46667	0.46636 ±0.00133	0.62222	0.61900 ±0.01110
10^6	1.40000	1.40015 ±0.00026	0.46667	0.46650 ±0.00049	0.62222	0.62056 ±0.00417
case 2						
10^3	12.0455	12.0467 ±0.13675	0.05164	0.05052 ±0.00384	0.01401	0.01376 ±0.00559
10^4	12.0455	12.0404 ±0.04351	0.05164	0.05138 ±0.00118	0.01401	0.01402 ±0.00197
10^5	12.0455	12.0414 ±0.01188	0.05164	0.05159 ±0.00033	0.01401	0.01409 ±0.00061
10^6	12.0455	12.0468 ±0.00284	0.05164	0.05166 ±0.00010	0.01401	0.01391 ±0.00016
case 3						
10^3	1.00005	0.99607 ±0.01494	0.99981	0.99579 ±0.01491	0.00086	0.00083 ±0.00096
10^4	1.00005	0.99716 ±0.00296	0.99981	0.99683 ±0.00295	0.00086	0.00086 ±0.00042
10^5	1.00005	0.99943 ±0.00097	0.99981	0.99919 ±0.00096	0.00086	0.00087 ±0.00012
10^6	1.00005	1.00009 ±0.00021	0.99981	0.99986 ±0.00021	0.00086	0.00088 ±0.00005

with corresponding parameters:

case	$1/\mu_1$	$1/\mu_2$	$1/\mu_3$
1	1.0	1.0	1.0
2	1.0	0.1	0.1
3	1.0	10.0	10.0

Case 2: H_2 system with three nodes and three customers. In this system, the service time at each node is H_2 with service time distributions $f_1(x) = \alpha\mu_1 e^{-\mu_1 x} + (1-\alpha)\mu_2 e^{-\mu_2 x}, f_2(x) = \beta\lambda_1 e^{-\lambda_1 x} + (1-\beta)\lambda_2 e^{-\lambda_2 x}$ and $f_3(x) = \gamma\nu_1 e^{-\nu_1 x} + (1-\gamma)\nu_2 e^{-\nu_2 x}$ respectively. Here,

TABLE I

μ_1	μ_2	$Y(\theta)$	$Y(\theta)_{\text{est.}}$	$\partial Y(\theta)/\partial\theta$	$(\partial Y(\theta)/\partial\theta)_{\text{est.}}$	$\partial^2 Y(\theta)/\partial\theta^2$	$(\partial^2 Y(\theta)/\partial\theta^2)_{\text{est.}}$
1.0	1.0	1.50000	1.49998	1.75000	1.74934	0.25000	0.24976
1.0	0.1	10.0909	10.0923	1.17355	1.17309	0.15026	0.15004
1.0	10.0	1.00909	1.00882	1.99174	1.98467	0.01502	0.01489

we chose $\theta = 1/\mu_1$ and $\Delta\theta = 0.02$.

N	D_b	$D_b'(\theta)$	$D'(\theta)_{est.}$	$D_b''(\theta)$	$D''(\theta)_{est.}$
			case 1		
10^3	2.78209 ±0.05279	0.41893 ±0.01226	0.42339 ±0.01098	0.26856 ±0.05987	0.27737 ±0.02150
10^4	2.74555 ±0.01094	0.42213 ±0.00401	0.42459 ±0.00372	0.27511 ±0.01987	0.27454 ±0.00838
10^5	2.74012 ±0.00264	0.42117 ±0.00128	0.42312 ±0.00122	0.27384 ±0.00722	0.27025 ±0.00238
			case 2		
10^3	1.27548 ±0.04328	0.46483 ±0.01123	0.46785 ±0.01099	3.02916 ±0.52520	3.00300 ±0.30233
10^4	1.29267 ±0.01805	0.47128 ±0.00424	0.47471 ±0.00399	2.86930 ±0.19665	2.82898 ±0.11949
10^5	1.29084 ±0.00603	0.46949 ±0.00098	0.47123 ±0.00097	2.85956 ±0.08343	2.82155 ±0.03936
			case 3		
10^3	12.0389 ±0.14017	0.07550 ±0.00540	0.07636 ±0.00220	0.02683 ±0.01887	0.03313 ±0.00208
10^4	11.9517 ±0.03103	0.07838 ±0.00162	0.07787 ±0.00102	0.04059 ±0.00854	0.03437 ±0.00066
10^5	11.9672 ±0.01022	0.07741 ±0.00057	0.07775 ±0.00023	0.03420 ±0.00267	0.03424 ±0.00025

with corresponding parameters:

case	α	$1/\mu_1$	$1/\mu_2$	β	$1/\lambda_1$	$1/\lambda_2$	γ	$1/\nu_1$	$1/\nu_2$
1	0.96	1.0	5.0	0.90	1.0	5.0	0.80	1.0	5.0
2	0.98	0.1	10.0	0.95	0.1	10.0	0.95	0.1	10.0
3	0.3	1.0	10.0	0.5	1.0	10.0	0.45	1.0	10.0

VI. Derivative Estimators of Mean Delays

One interesting performance metric for serial closed networks is the mean delay over several nodes. For simplicity, we will only show how to use our results from Section III to obtain first- and second-derivative estimators of mean delays over a single node. Mean waiting times can of course be handled similarly. The extension to mean delays over several nodes is straightforward.

For an m-node n-customer network, denoting the delay experienced by the ith customer at node k by $T_{i,k}$, we have

$$\begin{aligned} T_{i,k} &= D_{i,k} - D_{\hat{\imath},k-1} \\ T_{i,k}^p &= D_{i,k}^p - D_{\hat{\imath},k-1}^p. \end{aligned}$$

Therefore

$$\Delta T_{i,k} = T_{i,k}^p - T_{i,k} = \Delta D_{i,k} - \Delta D_{\hat{\imath},k-1}.$$

Now, for N customers served at node k, define

$$\overline{T}_N = \frac{1}{N}\sum_{i=1}^{N} T_{i,k}.$$

Therefore

$$E[\Delta\overline{T}_N] = \frac{1}{N}\left\{\sum_{i=1}^{N} E[\Delta D_{i,k}] - \sum_{i=1}^{N} E[\Delta D_{\hat{\imath},k-1}]\right\}.$$

By expanding each term on the RHS above, we obtain the following first-derivative estimator:

$$[\overline{T}'_N]_{\text{est.}} = \frac{1}{N}\sum_{i=1}^{N}\left[\sum_{(j,1)\in P_{i,k}} \frac{\partial}{\partial\theta} S_{j,1} - \sum_{(j,1)\in P_{\hat{\imath},k-1}} \frac{\partial}{\partial\theta} S_{j,1}\right].$$

Similarly, the second derivative estimator is given by

$$[T''_N]_{\text{est.}} = \frac{1}{N}\sum_{i=1}^{N}([T''_{i,k}]_{\text{est.}} - [T''_{\hat{\imath},k}]_{\text{est.}})$$

where

$$\begin{aligned} [T''_{i,k}]_{\text{est.}} = & \sum_{(j,1)\in P_{i,k}} \frac{\partial^2}{\partial\theta^2} S_{j,1} + \sum_{(j,l)\in Q^*_{i,k}} g^W_{j,l}(0|z_{\hat{\jmath},l-1}) \\ & \cdot\left(\left[\sum_{(q,1)\in P_{\hat{\jmath},l-1}} \frac{\partial}{\partial\theta} S_{q,1} - \sum_{\substack{(q,1)\in P_{j-1,l} \\ (q,1)\neq(j-1,l)}} \frac{\partial}{\partial\theta} S_{q,1}\right.\right. \\ & \left.\left. - \frac{\partial}{\partial\theta} S^a_{j-1,l}\right]^+\right)^2 + \sum_{(j,l)\in R^*_{i,k}} g^I_{j,l}(0|z_{j-1,l}) \\ & \cdot\left(\left[\sum_{(q,1)\in P_{j-1,l}} \frac{\partial}{\partial\theta} S_{q,1} - \sum_{\substack{(q,1)\in P_{\hat{\jmath},l-1} \\ (q,1)\neq(\hat{\jmath},l-1)}} \frac{\partial}{\partial\theta} S_{q,1}\right.\right. \\ & \left.\left. - \frac{\partial}{\partial\theta} S^a_{\hat{\jmath},l-1}\right]^+\right)^2. \end{aligned}$$

Table I provides experimental results for one of the systems considered in the previous section.

Case 1: Exponential system with two nodes and two customers ($N = 10^6$).

Note that the second derivative of the mean cycle time is the same as the second derivative of the mean interdeparture time (see Section VI, Case 1) which of course is to be expected from Little's law.

VII. Padé Approximation of the Mean Interdeparture Time Response Curve Through Its Derivatives

As already mentioned in the introduction, one of the reasons for seeking efficient means to estimate second derivatives is to provide estimates of entire response surfaces of DEDS with respect to various parameters. As recently shown in [18], Padé approximants of various performance metrics can be remarkably accurate based on first- and second-derivative information alone (compared to polynomial approximations whose accuracy is inherently limited as discussed in [18]). The benefit here is obvious, since this approach requires only observation of a single sample path (e.g., one simulation run) with some additional computational effort for first- and second-derivative estimation to determine an entire response curve; this is in contrast to an approach based on multiple sample path observations.

In this section, we will generate the response curve of the mean interdeparture time through a Padé approximation based on the results of Section III. Since the throughput is the inverse of the mean interdeparture time, this is equivalent to obtaining the response curve of the throughput with respect to the mean of one of the service time distributions of our system. Before doing so, we give a brief description of the Padé approximation as it applies to our problem (for a detailed description, see [1] and [18]).

Suppose that the function $J(\theta)$ can be written as a power series

$$J(\theta) = \sum_{i=0}^{\infty} c_i \theta^i.$$

Let L, M be two integers. Then a $[L/M]$ Padé approximant to $J(\theta)$ can be expressed as a rational function

$$[L/M] = \frac{P_L(\theta)}{Q_M(\theta)} \tag{43}$$

where $P_L(\theta) = a_0 + a_1\theta + \cdots + a_L\theta^L$ and $Q_M(\theta) = b_0 + b_1\theta + \cdots + b_M\theta^M$, and $a_0, a_1, \cdots, a_L$ and $b_1, b_2, \cdots, b_M$ are coefficients to be determined (b_0 is always set to one).

The important requirement for $[L/M]$ is that the first $L + M + 1$ coefficients of its Taylor expansion agree with those of $J(\theta)$. When the first $L + M + 1$ coefficients of $J(\theta), c_0, c_1, \cdots, c_{L+M+1}$, are given through the derivatives of $J(\theta)$ with respect to θ, the coefficients $a_0, a_1, \cdots, a_L$ and $b_1, b_2, \cdots, b_M$ can be determined through the following two sets of equations (see [1] for details):

$$\begin{pmatrix} c_{L-M+1} & c_{L-M+2} & \cdots & c_L \\ c_{L-M+2} & c_{L-M+3} & \cdots & c_{L+1} \\ \vdots & \vdots & \ddots & \vdots \\ c_L & c_{L+1} & \cdots & c_{L+M-1} \end{pmatrix} \begin{pmatrix} b_M \\ b_{M-1} \\ \vdots \\ b_1 \end{pmatrix} = -\begin{pmatrix} c_{L+1} \\ c_{L+2} \\ \vdots \\ c_{L+M} \end{pmatrix} \tag{44}$$

and

$$\begin{aligned} a_0 &= c_0, \\ a_1 &= c_1 + b_1 c_0, \\ \vdots & \;\vdots\; \vdots \\ a_L &= c_L + \sum_{i=1}^{\min(L,M)} b_i c_{L-i}. \end{aligned} \tag{45}$$

Thus, for a fixed number of given derivatives of $J(\theta)$, say T derivatives, we can always choose some L and M such that $T = L + M$ and determine $a_0, a_1, \cdots, a_L$ and $b_1, b_2, \cdots, b_M$ through (44) and (45). In fact, the power of the Padé approximation is that it provides the flexibility of choosing L and M as long as $T = L + M$ holds which means that it is possible to approximate a very general class of curves for a fixed T. Obviously, the selection of appropriate L and M is very important for a good approximation given some T. However, this task is greatly facilitated in the case of many DEDS where one can exploit some basic properties of $J(\theta)$.

In general, the more derivatives we can provide, the more accurate the approximation becomes. In our system, we have seen that estimating higher than second derivatives becomes very involved. However, we have found that first- and second-derivative estimates alone provide excellent approximations for a large range of parameter values around a nominal point.

We will now derive a Padé approximant for $\overline{D}(\theta)$ based on $\overline{D}(\theta_1)_{\text{est.}}, \overline{D}'(\theta_1)_{\text{est.}}$, and $\overline{D}''(\theta_1)_{\text{est.}}$, the estimates of the mean interdeparture time and its first two derivatives obtained from a single sample path observed under $\theta = \theta_1$ (where θ is the mean service time at node one). Since we know that the mean interdeparture time approaches the mean service time as $\theta \to \infty$, i.e.,

$$\lim_{\theta\to\infty} [\overline{D}(\theta) - \theta] = 0$$

instead of approximating $\overline{D}(\theta)$ we will approximate $[\overline{D}(\theta) - \theta]$. In view of this observation, we derive a [0/2] ($L = 0$ and $M = 2$) Padé approximant. Let

$$\begin{aligned} c_0 &= \overline{D}(\theta_1)_{\text{est.}} - \theta_1 \\ c_1 &= \overline{D}'(\theta_1)_{\text{est.}} - 1 \\ c_2 &= \tfrac{1}{2}\overline{D}''(\theta_1)_{\text{est.}}. \end{aligned}$$

By applying (44) and (45), we can solve for a_0, b_1, and b_2 to get

$$[\overline{D}(\theta) - \theta]_{\text{est.}} = \frac{c_0^3}{c_0^2 - c_0c_1(\theta - \theta_1) + (c_1^2 - c_0c_2)(\theta - \theta_1)^2}.$$

Therefore

$$\begin{aligned} [\overline{D}(\theta)]_{\text{approx.}} = \theta &+ [2(\overline{D}(\theta_1)_{\text{est.}} - \theta_1)^3]\{2(\overline{D}(\theta_1)_{\text{est.}} - \theta_1)^2 \\ &- 2(\overline{D}(\theta_1)_{\text{est.}} - \theta_1)(\overline{D}'(\theta_1)_{\text{est.}} - 1)(\theta - \theta_1) \\ &+ [2(\overline{D}'(\theta_1)_{\text{est.}} - 1)^2 - (\overline{D}(\theta_1)_{\text{est.}} - \theta_1) \\ &\cdot (\overline{D}''(\theta_1)_{\text{est.}})](\theta - \theta_1)^2\}^{-1} \end{aligned} \tag{46}$$

which is the Padé approximant.

In Fig. 3, we show the response curve of the mean interdeparture time at node 1 with respect to the mean service time in a two-node two-customer system. The service times of both nodes are exponentially distributed with mean 1. Using the results of Section V, we have included both a Padé approximation and a polynomial approximation. It can be

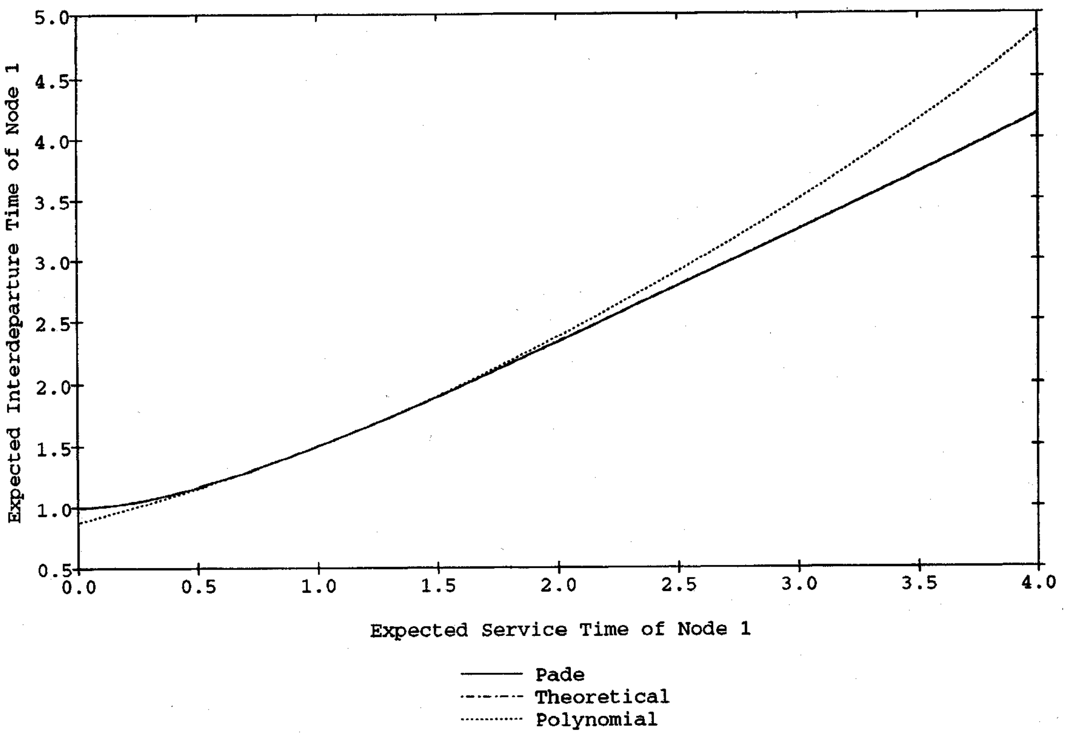


Fig. 3. Padé approximation for a two-node two-customer system (1).

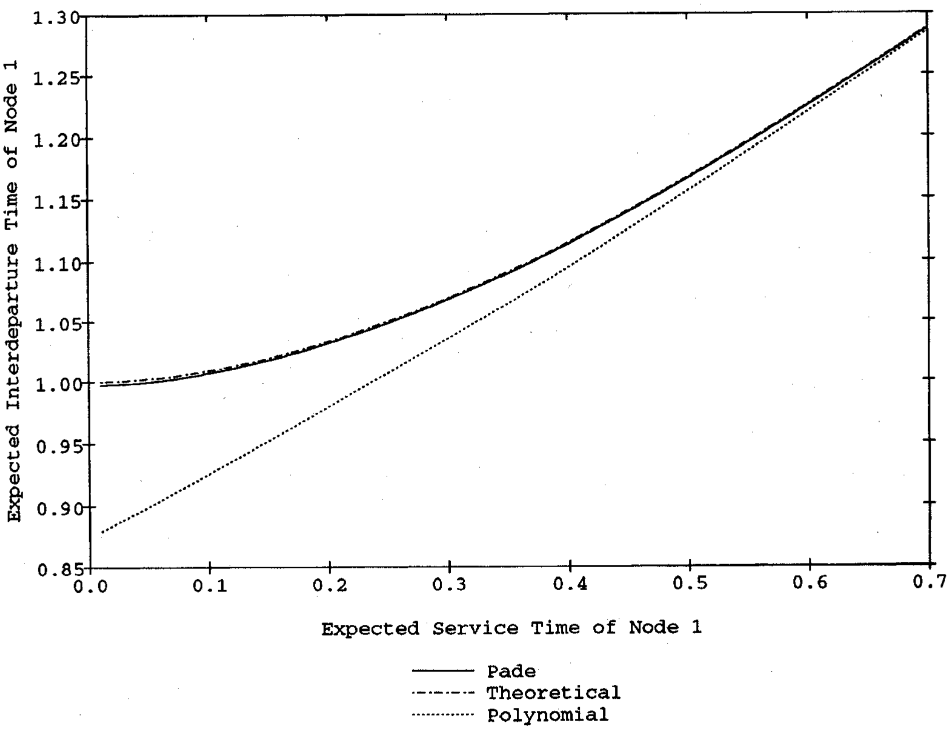


Fig. 4. Padé approximation for a two-node two-customer system (2).

seen from this figure that the Padé approximation is virtually indistinguishable from the curve representing the analytically evaluated mean interdeparture time. The accuracy of the polynomial approximation, on the other hand, is limited to a small range of values around the nominal point ($\theta = 1$). In Fig. 4, we have substantially enlarged part of Fig. 3 to further show that it is in fact difficult to find an error in the Padé approximation compared to the analytically obtained curve.

VIII. Conclusions and Future Work

We have considered a serial closed-queueing network with arbitrary service time distributions and derived an unbiased second-derivative estimator of the throughput over N customers served at some node with respect to a parameter of the service distribution at that node. Our approach is based on observing a single sample path of this system and evaluating all second-order effects on interdeparture times as a result of the parameter perturbation. We then define an estimator as a conditional expectation over appropriate observable quantities, as in smoothed perturbation analysis (SPA). Along the way, we have also recovered the first-derivative estimator of the throughput which can also be derived using other techniques (e.g., [7]). Our results are easily extended to the second derivative of the mean delay of customers between any two points in the network, as shown in Section VI.

Along the way, our analysis has provided some new insights regarding the type of sample path information we need to condition on to estimate higher-order performance derivatives. As seen in Section IV, even though the derivation of the second-order derivative estimator is fairly elaborate, the actual algorithm for its on-line implementation is relatively simple.

As mentioned in the introduction, a major motivation for this work is the possibility of using the first and second derivatives of performance metrics of complex DEDS to construct a global response surface. Recent developments exploiting Padé approximation techniques [18] have made this possibility very real. Our results in Section VII indicate that the entire throughput response surface of a serial closed-queueing network can be constructed with remarkable accuracy using only the first- and second-derivative estimates we have developed. Moreover, we believe that the basic approach presented here may be extended to more complex network topologies which is the subject of ongoing research.

Finally, establishing the consistency of the estimators we have derived remains the subject of future research. Nonetheless, it has been shown by direct computation (see [2]) that consistency holds for two-node cyclic Markovian networks in which case second derivatives of the throughput at steady state can be analytically obtained.

Appendix
Proofs of the Lemmas

Before proceeding with the proof of Lemma 1, we need the following result regarding the conditional density of the idle period $I_{j,l}$ (respectively, waiting time $W_{j,l}$) for noncritical (j,l), i.e., $(j,l) \notin Q^*_{i,k}$ (respectively, (j,l), i.e., $(j,l) \notin R^*_{i,k}$), as defined in Section III-C.

Lemma 4:

1) Suppose that an idle period immediately follows a departure at node l at time $D_{j-1,l}$. If $e_{j,l-1}$ is not the next event to occur, then

$$g^I_{j,l}(0|\Lambda_{j-1,l}) = \lim_{x\to 0} \frac{1}{x} P(I_{j,l} \le x|\Lambda_{j-1,l}) = 0. \tag{47}$$

2) Suppose that the customer $C_{j,l}$ upon arrival at node l at time $D_{j,l-1}$ finds the server busy. If $e_{j-1,l}$ is not the next event to occur, then

$$g^W_{j,l}(0|\Lambda_{j,l-1}) = \lim_{x\to 0} \frac{1}{x} P(W_{j,l} \le x|\Lambda_{j,l-1}) = 0. \tag{48}$$

Proof: We only give the proof of 1). The arguments for 2) are identical. First, suppose that node $l-1$ is busy. In that case, $I_{j,l} = S^r_{l-1}(D_{j-1,l})$, the residual service time of the upstream server at the beginning of the idle period. Then

$$\begin{aligned} &P(I_{j,l} \le x|\Lambda_{j-1,l}) \\ &\quad = P(S^r_{l-1}(D_{j-1,l}) \le x|\mathcal{M}(D_{j-1,l}), \\ &\qquad \{S^a_q(D_{j-1,l}); q \in \mathcal{M}(D_{j-1,l})\}, \\ &\qquad S^r_{l-1}(D_{j-1,l}) > \min_{\{q\in\mathcal{M}(D_{j-1,l})\}} S^r_q(D_{j-1,l})). \end{aligned} \tag{49}$$

Keeping in mind that conditioned on $\mathcal{M}(D_{j-1,l})$, $\{S^a_q(D_{j-1,l}); q \in \mathcal{M}(D_{j-1,l}\}$, the random variables $\{S^r_q(D_{j-1,l});\ q \in \mathcal{M}(D_{j-1,l})\}$ are independent with distributions

$$\left\{ \frac{F_q(S^a_q(D_{j-1,l}) + x)}{\overline{F}_q(S^a_q(D_{j-1,l}))}; q \in \mathcal{M}(D_{j-1,l}) \right\}$$

where, as usual, $\overline{F}_q \overset{\text{def}}{=} 1 - F_q$ denotes the survivor function of the distribution F_q, we obtain via a straightforward computation (similar to one found in [14]) expression (50), shown at the bottom of the page, for the conditional distribution of $I_{j,l}$.

Hence the conditional distribution of $I_{j,l}$ is obviously absolutely continuous with density $g^I_{j,l}(\cdot|\Lambda_{j-1,l})$ obtained by differentiating (50) w.r.t. x. Then, an application of the mean value theorem yields

$$\frac{1}{x} P(I_{j,l} \le x|\Lambda_{j-1,l}) = g^I_{j,l}(\xi|\Lambda_{j-1,l})$$

for some $\xi \in [0,x]$. Then, since $g^I_{j,l}(0|\Lambda_{j-1,l}) = 0$ (the term inside the square brackets in the numerator of (50) vanishes

$$P(I_{j,l} \le x|\Lambda_{j-1,l}) = \frac{\displaystyle\int_0^x f_{l-1}(S^a_{l-1}(D_{j-1,l}) + u) \left[1 - \prod_{\substack{q\in\mathcal{M}(D_{j-1,l}) \\ q\neq l-1}} \frac{\overline{F}_q(S^a_q(D_{j-1,l}) + u)}{\overline{F}_q(S^a_q(D_{j-1,l}))}\right] du}{\displaystyle\int_0^\infty f_{l-1}(S^a_{l-1}(D_{j-1,l}) + u) \left[1 - \prod_{\substack{q\in\mathcal{M}(D_{j-1,l}) \\ q\neq l-1}} \frac{\overline{F}_q(S^a_q(D_{j-1,l}) + u)}{\overline{F}_q(S^a_q(D_{j-1,l}))}\right] du} \tag{50}$$

for $u = 0$), taking the limit as $x \to 0$ in the above display establishes part 1) of the lemma when the upstream node is busy.

To complete the proof we need to also examine the possibility that the upstream node $l-1$ is idle. Suppose then that at time $D_{j-1,l}$ the first a upstream nodes, $l-1, \cdots, l-a$, are idle, where $1 \leq a \leq m-1$. In that case, $P(I_{j,l} \leq x|\Lambda_{j-1,l})$ is given by the convolution of the distribution of the residual service time at node $l-a-1$ with the distributions of the service times at the intervening nodes, $l-a, \cdots, l-1$, all of which have hazard rates bounded above by γ (by Assumption A.4). Hence

$$P(I_{j,l} \leq x|\Lambda_{j-1,l}) \leq \frac{(\gamma x)^{a+1}}{(a+1)!}e^{-\gamma x}$$

from which the lemma statement follows immediately since $a \geq 1$. ∎

We can now proceed with the proof of Lemma 1 in Section III-E

Proof of Lemma 1: We only give the proof of 1), since 2) is obtained by identical arguments. From the definitions of $\Delta_{j,l}$ and $\tilde{T}_{j,l}$ in (11) and (12), respectively, in conjunction with Assumption A.2 and the fact that $S_{q,r} \leq D_{\hat{j},l-1}$ for any $(q,r) \in B_{j,l}$, we obtain the following inequalities:

$$|\Delta_{j,l}| \leq 2(j+n)c_1\Delta\theta + 2c_2D_{\hat{j},l-1}\Delta\theta \tag{51}$$

$$|\tilde{T}_{j,l}| \leq 4(j+n)c_1\Delta\theta + 4c_2D_{\hat{j},l-1}\Delta\theta. \tag{52}$$

Then, since $I_{j,l} = D_{\hat{j},l-1} - D_{j-1,l}$, we get

$$\begin{aligned}\tilde{I}_{j,l} &= [\Delta_{j,l} + \tilde{T}_{j,l} - I_{j,l}]^+\\ &\leq [6(j+n)c_1\Delta\theta + 6c_2D_{j-1,l}\Delta\theta - (1-6c_2\Delta\theta)I_{j,l}]^+.\end{aligned}$$

Moreover, since $[x-y]^+ \leq x\mathbf{1}(y \leq x)$, for $x, y \geq 0$

$$\begin{aligned}\tilde{I}_{j,l} \leq{}& \Delta\theta(6(j+n)c_1 + 6c_2D_{j-1,l})\\ &\cdot \mathbf{1}\left(I_{j,l} \leq \frac{6(j+n)c_1 + 6c_2D_{j-1,l}}{(1-6c_2\Delta\theta)}\Delta\theta\right).\end{aligned}$$

Since $D_{j-1,l}$ belongs to $\Lambda_{j-1,l}$, taking conditional expectations in the above expression gives

$$\begin{aligned}&E[\tilde{I}_{j,l}|\Lambda_{j-1,l}]\\ &\quad\leq \Delta\theta(6(j+n)c_1 + 6c_2D_{j-1,l})\\ &\quad\quad\cdot P\left(I_{j,l} \leq \frac{6(j+n)c_1 + 6c_2D_{j-1,l}}{(1-6c_2\Delta\theta)}\Delta\theta|\Lambda_{j-1,l}\right).\end{aligned}$$

Letting

$$x(\Delta\theta) := \frac{6(j+n)c_1 + 6c_2D_{j-1,l}}{(1-6c_2\Delta\theta)}\Delta\theta$$

and dividing both sides of the above inequality by $\Delta\theta^2$, we obtain

$$\begin{aligned}&\frac{1}{\Delta\theta^2}E[\tilde{I}_{j,l}|\Lambda_{j-1,l}]\\ &\quad\leq \frac{(6(j+n)c_1 + 6c_2D_{j-1,l})^2}{(1-6c_2\Delta\theta)}\ \frac{1}{x(\Delta\theta)}\\ &\quad\quad\cdot P(I_{j,l} \leq x(\Delta\theta)|\Lambda_{j-1,l})\\ &\quad= \frac{(6(j+n)c_1 + 6c_2D_{j-1,l})^2}{(1-6c_2\Delta\theta)}g^I(\xi|\Lambda_{j-1,l})\end{aligned}$$

for some $\xi \in [0, x(\Delta\theta)]$. Note that $g^I(y|\Lambda_{j-1,l}) \leq c'$ for some constant (this immediately follows from (50), where differentiating to obtain $g^I(\cdot|\Lambda_{j-1,l})$ gives a numerator bounded by γ by Assumption A.4 and a denominator which is some constant). Since, in addition, $E[D^2_{j-1,l}] < \infty$, setting

$$K^I_{j,l} = \frac{(6(j+n)c_1 + 6c_2D_{j-1,l})^2}{(1-6c_2\Delta\theta)}g^I(\xi|\Lambda_{j-1,l})$$

completes the proof. ∎

Proof of Lemma 2: We examine only the idle period case in detail, as the waiting time case is similar. We begin with the observation that necessarily $l-1 \in \mathcal{M}(D_{j-1,l})$, i.e., the node immediately preceding node l must be busy at time $D_{j-1,l}$. Hence, the idle period of length $I_{j,l}$ is the residual service time of the customer in the upstream node, $l-1$, at time $D_{j-1,l}$, defined as $S^r_{l-1}(D_{j-1,l})$. Equivalently, we write

$$I_{j,l} = S_{l-1}(D_{j-1,l}) - S^a_{l-1}(D_{j-1,l}).$$

Because of the independence assumptions regarding the service processes at the nodes of the network, the relevant part of $z_{j-1,l}$ is the set of active nodes, $\mathcal{M}(D_{j-1,l})$, the ages of the service processes at the active nodes given by $\{S^a_q(D_{j-1,l}); q \in \mathcal{M}(D_{j-1,l})\}$, the identity of the next event, and the residual service times *at the end of the idle period* $I_{j,l}, \{S^r_q(D_{\hat{j},l-1}); q \in \mathcal{M}(D_{j-1,l})\}$.

Under the assumption of the lemma, $e_{\hat{j},l-1}$ is the next event, i.e., no events occur during the idle period $I_{j,l}$. This translates into the condition

$$\begin{aligned}&S_{l-1}(D_{j-1,l}) - S^a_{l-1}(D_{j-1,l})\\ &\quad= S_q(D_{j-1,l}) - S^a_q(D_{j-1,l}) - S^r_q(D_{\hat{j},l-1})\\ &\quad\quad\text{for all } q \in \mathcal{M}(D_{j-1,l})\backslash\{l-1\}\end{aligned}$$

where $S_q(D_{j-1,l}) - S^a_q(D_{j-1,l})$ is the residual service time at node q at the *beginning* of the idle period, and $S^r_q(D_{\hat{j},l-1})$ is the residual service time at node q at the *end* of the idle period.

Denoting by $g^I(\cdot|z_{j-1,l})$ the conditional density of $I_{j,l}$, we then have

$$\begin{aligned}&g^I(x|z_{j-1,l})\,dx\\ &\quad\propto P(S_{l-1}(D_{j-1,l}) - S^a_{l-1}(D_{j-1,l}) \in dx\\ &\quad\quad S_q(D_{j-1,l}) - S^a_q(D_{j-1,l}) - S^r_q(D_{\hat{j},l-1}) \in dx\\ &\quad\quad q \in \mathcal{M}(D_{j-1,l})\backslash\{l-1\}|S^a_{l-1}(D_{j-1,l})\\ &\quad\quad S^a_q(D_{j-1,l}), S^r_q(D_{\hat{j},l-1}), q \in \mathcal{M}(D_{j-1,l})\backslash\{l-1\}).\end{aligned}$$

It is then straightforward to obtain (53), shown at the bottom of the next page.

Taking into account that

$$\begin{aligned}S_q(D_{j-1,l}) &= S^a_q(D_{j-1,l}) + I_{j,l} + S^r_q(D_{\hat{j},l-1})\\ &\quad q \in \mathcal{M}(D_{j-1,l})\end{aligned}$$

(53) can be rewritten as (54), shown at the bottom of the next page. ∎

Before proceeding with the proof of Lemma 3, we need the following result stated as Lemma 5.

Lemma 5: Let supp$F_1(\cdot;\theta)$ denote the support of $F_1(\cdot;\theta)$ and $\Phi_{\Delta\theta}$:supp $F_1(\cdot;\theta) \to \boldsymbol{R}^+$ the function $F_1^{-1}(F_1(\cdot;\theta);\theta+\Delta\theta)$. Keeping θ fixed, consider the family $\{\Phi_{\Delta\theta};\Delta\theta \geq 0\}$. Then, for any $S \in$ supp $F_1(\cdot;\theta)$

$$\lim_{\Delta\theta\to 0} \frac{1}{\Delta\theta}[\Phi_{\Delta\theta}(S+y\Delta\theta) - S] = \frac{\partial S}{\partial\theta} + y.$$

Proof: Observe first that Φ_0 is the identity map on the support of $F_1(\cdot;\theta)$ and

$$\lim_{\Delta\theta\to 0} \Phi_{\Delta\theta}(x) = x \quad \text{for } x \in \text{supp } F_1. \tag{55}$$

Since $\Phi_{\Delta\theta}(S+y\Delta\theta) - S = \Phi_{\Delta\theta}(S+y\Delta\theta) - \Phi_{\Delta\theta}(S) + \Phi_{\Delta\theta}(S) - S$, and

$$\frac{\Phi_{\Delta\theta}(S) - S}{\Delta\theta} \to \frac{\partial S}{\partial\theta} \quad \text{for all } S \in \text{supp } F_1(\cdot;\theta)$$

it is enough to show that

$$\frac{\Phi_{\Delta\theta}(S+y\Delta\theta) - \Phi_{\Delta\theta}(S)}{\Delta\theta} \to y$$
$$\text{for all } S \in \text{supp } F_1(\cdot;\theta). \tag{56}$$

Suppose that for some $x_0, f(x_0;\theta) > 0$. Note that, in view of the continuity of $f(x;\theta)$, there exists ϵ such that when $|x-x_0| < \epsilon, \Delta\theta < \epsilon, f(x;\theta+\Delta\theta) > 0$. Also, (55) implies that there exists $\delta > 0$ such that for $\Delta\theta < \delta, |\Phi_{\Delta\theta}(x_0) - x_0| < \epsilon$. From this follows that for $\Delta\theta < \delta, f(\Phi_{\Delta\theta}(x_0);\theta+\Delta\theta) > 0$, and hence that for $x_0 \in$ supp $F_1(\cdot;\theta)$

$$\frac{\partial}{\partial x}\Phi_{\Delta\theta}(x) = \frac{f(x;\theta)}{f(\Phi_{\Delta\theta}(x);\theta+\Delta\theta)} \quad \text{for} \quad \Delta\theta < \delta.$$

Therefore, from the mean value theorem

$$\frac{1}{\Delta\theta}[\Phi_{\Delta\theta}(S+y\Delta\theta) - \Phi_{\Delta\theta}(S)]$$
$$= y\frac{f(\zeta;\theta)}{f(\Phi_{\Delta\theta}(\zeta);\theta+\Delta\theta)} \quad \text{where} \quad \zeta \in [S, S+y\Delta\theta].$$

Letting $\Delta\theta \to 0$, and invoking the continuity of f and (55), establishes (56). ■

We can now provide the proof of Lemma 3.

Proof of Lemma 3: Let us begin by defining

$$\Delta R \overset{\text{def}}{=} \sum_{(q,r)\in P_{j-1,l}} \Delta S_{q,r} - \sum_{\substack{(q,r)\in P_{\hat{\jmath},l-1}\\(q,r)\neq(\hat{\jmath},l-1)}} \Delta S_{q,r} + \tilde{T}_{j,l}. \tag{57}$$

We will then obtain a limit for $\Delta R/\Delta\theta$ as $\Delta\theta \to 0$.

Since $D_{j-1,l}$ immediately precedes $D_{\hat{\jmath},l-1}$ we can see from (12) and (3) that $\tilde{T}_{j,l}$ belongs to $z_{j-1,l}$. Moreover, observe in (12) that $\tilde{T}_{j,l}$ consists of a sum of $\tilde{I}_{q,r}$ and $\tilde{W}_{q,r}$ terms. Also, $\tilde{W}_{q,r} = [\Delta D_{\hat{q},r-1} - \Delta D_{q-1,r} - W_{q,r}]^+$ and $\tilde{I}_{q,r} = [\Delta D_{q-1,r} - \Delta D_{\hat{q},r-1} - I_{q,r}]^+$ from (3). In addition, when $(q,r) \in Q_{j-1,l} \cup Q_{\hat{\jmath},l-1}$, we have $W_{q,r} > 0$. Hence

$$\lim_{\Delta\theta\to 0} \frac{\tilde{W}_{q,r}}{\Delta\theta} = \lim_{\Delta\theta\to 0} \left[\frac{\Delta D_{\hat{q},r-1}}{\Delta\theta} - \frac{\Delta D_{q-1,r}}{\Delta\theta} - \frac{W_{q,r}}{\Delta\theta}\right]^+.$$

However, w.p. 1, $\lim_{\Delta\theta\to 0} \Delta D_{\hat{q},r-1}/\Delta\theta = \partial/D_{\hat{q},r-1}/\partial\theta$ and $\lim_{\Delta\theta\to 0} \Delta D_{q-1,r}/\Delta\theta = \partial D_{q-1,r}/\partial\theta$. In view of the fact that $W_{q,r} > 0$ and the continuity of the positive part function $[\cdot]^+$, w.p. 1

$$\lim_{\Delta\theta\to 0} \frac{1}{\Delta\theta}\tilde{W}_{q,r} = 0.$$

Similarly, for $(q,r) \in R_{j-1,l} \cup R_{\hat{\jmath},l-1}$, we have $I_{q,r} > 0$ and get

$$\lim_{\Delta\theta\to 0} \frac{1}{\Delta\theta}\tilde{I}_{q,r} = 0.$$

From (12), this establishes that

$$\lim_{\Delta\theta\to 0} \frac{1}{\Delta\theta}\tilde{T}_{j,l} = 0 \text{ w.p. } 1. \tag{58}$$

Therefore, returning to (57)

$$\lim_{\Delta\theta\to 0} \frac{\Delta R}{\Delta\theta} = \sum_{(q,r)\in P_{j-1,l}} \frac{\partial S_{q,r}}{\partial\theta} - \sum_{\substack{(q,r)\in P_{\hat{\jmath},l-1}\\(q,r)\neq(\hat{\jmath},l-1)}} \frac{\partial S_{q,r}}{\partial\theta} \text{ w.p. } 1. \tag{59}$$

$$g^I(x|z_{j-1,l}) = \frac{f_{l-1}(S^a_{l-1}(D_{j-1,l})+x)\left[\displaystyle\prod_{\substack{q\in\mathcal{M}(D_{j-1,l})\\q\neq l-1}} f_q(S^a_q(D_{j-1,l}) + S^r_q(D_{\hat{\jmath},l-1}) + x)\right]}{\displaystyle\int_0^\infty f_{l-1}(S^a_{l-1}(D_{j-1,l})+u)\left[\prod_{\substack{q\in\mathcal{M}(D_{j-1,l})\\q\neq l-1}} f_q(S_q(D_{j-1,l}) + S^r_q(D_{\hat{\jmath},l-1}) + u)\right] du} \tag{53}$$

$$g^I(x|z_{j-1,l}) = \frac{f_{l-1}(S^a_{l-1}(D_{j-1,l})+x)\left[\displaystyle\prod_{\substack{q\in\mathcal{M}(D_{j-1,l})\\q\neq l-1}} f_q(S_q(D_{j-1,l}) - I_{j,l} + x)\right]}{\displaystyle\int_0^\infty f_{l-1}(S^a_{l-1}(D_{j-1,l})+u)\left[\prod_{\substack{q\in\mathcal{M}(D_{j-1,l})\\q\neq l-1}} f_q(S_q(D_{j-1,l}) - I_{j,l} + u)\right] du} \tag{54}$$

Next, combining the definition of $\Delta_{j,l}$ in (11) with that of ΔR in (57), we get

$$\Delta_{j,l} + \tilde{T}_{j,l} = \Delta R + \Delta S_{\hat{\jmath},l-1}.$$

With the above definitions, we have

$$\begin{aligned} E[\tilde{I}_{j,l}|z_{j-1,l}] &= E[[\Delta_{j,l} + \tilde{T}_{j,l} - I_{j,l}]^{+}|z_{j-1,l}] \\ &= \int_0^\infty [\Delta R + \Delta S_{\hat{\jmath},l-1} - x]^{+} g_{j,l}^{I}(x|z_{j-1,l})\, dx. \end{aligned} \tag{60}$$

We now distinguish two cases.

Case 1—$l \neq 2$: In this case, since we have assumed that the parameter θ only affects the service distribution of node 1, we have $\Delta S_{j,l-1} = 0$ for all $j = 1, 2, \cdots$, and hence $\Delta S_{\hat{\jmath},l-1}/\Delta\theta = 0$. Then, with the change of variables $x = y\Delta\theta$, (60) becomes

$$E[\tilde{I}_{j,l}|z_{j-1,l}] = \Delta\theta^2 \int_0^\infty \left[\frac{\Delta R}{\Delta\theta} - y\right]^{+} g_{j,l}^{I}(y\Delta\theta|z_{j-1,l})\, dy. \tag{61}$$

From (57) and Assumption A.2, we can proceed exactly as in (51) and (52) of Lemma 1 and obtain the bound

$$\frac{\Delta R}{\Delta\theta} \leq 6((j+n)c_1 + D_{j-1,l})\Delta\theta \tag{62}$$

which allows us to apply the dominated convergence theorem

$$\begin{aligned} &\lim_{\Delta\theta\to 0} \frac{1}{\Delta\theta^2} E[\tilde{I}_{j,l}|z_{j-1,l}] \\ &\quad= \int_0^\infty \left[\lim_{\Delta\theta\to 0} \frac{\Delta R}{\Delta\theta} - y\right]^{+} g_{j,l}^{I}(0|z_{j-1,l})\, dy. \end{aligned} \tag{63}$$

Finally, observe that since in this case $\Delta S_{\hat{\jmath},l-1}/\Delta\theta = 0$, (19) reduces to

$$Y_{j,l}^{I} = \sum_{(q,r)\in P_{j-1,l}} \frac{\partial S_{q,r}}{\partial\theta} - \sum_{\substack{(q,r)\in P_{\hat{\jmath},l-1} \\ (q,r)\neq(\hat{\jmath},l-1)}} \frac{\partial S_{q,r}}{\partial\theta}. \tag{64}$$

This observation, combined with (64) and the fact that for any real a, $\int_0^\infty [a-y]^{+}\, dy = \frac{1}{2}(a^{+})^2$, yields from (63)

$$\lim_{\Delta\theta\to 0} \frac{1}{\Delta\theta^2} E[\tilde{I}_{j,l}|z_{j-1,l}] = \frac{1}{2} g_{j,l}^{I}(0|z_{j-1,l})([Y_{j,l}^{I}]^{+})^2. \tag{65}$$

Case 2—$l = 2$: Using the same notation as before, let us set in addition

$$\Delta S_{\hat{\jmath},l-1} = \begin{cases} F_1^{-1}(F_1(S_{\hat{\jmath},1};\theta);\theta+\Delta\theta), & \text{if } l = 2 \\ 0, & \text{otherwise.} \end{cases} \tag{66}$$

Then, using the definition in Lemma 5, and observing that $S_{\hat{\jmath},l-1} = S_{\hat{\jmath},l-1}^{a} + I_{j,l}$ (since $D_{j-1,l}$ immediately precedes $D_{\hat{\jmath},l-1}$), we have

$$\Delta S_{\hat{\jmath},l-1} = \Phi_{\Delta\theta}(S_{\hat{\jmath},l-1}^{a} + I_{j,l}) - S_{\hat{\jmath},l-1}^{a} - I_{j,l}.$$

Thus, (60) becomes

$$\begin{aligned} E[\tilde{I}_{j,l}|z_{j-1,l}] = E \int_0^\infty &[\Delta R + \Phi_{\Delta\theta}(S_{\hat{\jmath},l-1}^{a} + x) \\ &- S_{\hat{\jmath},l-1}^{a} - 2x]^{+} g_{j,l}^{I}(x|z_{j-1,l})\, dx. \end{aligned} \tag{67}$$

Arguing as in Case 1, with the change of variable $x = y\Delta\theta$, (67) becomes

$$\begin{aligned} &E[\tilde{I}_{j,l}|z_{j-1,l}] \\ &\quad= \Delta\theta^2 \int_0^\infty \left[\frac{\Delta R}{\Delta\theta} + \frac{\Phi_{\Delta\theta}(S_{\hat{\jmath},l-1}^{a} + y\Delta\theta) - S_{\hat{\jmath},l-1}^{a}}{\Delta\theta} - 2y\right]^{+} \\ &\qquad\cdot g_{j,l}^{I}(y\Delta\theta|z_{j-1,l})\, dy. \end{aligned} \tag{68}$$

However, Assumption A.2 and the triangle inequality lead to the bound

$$\begin{aligned} &|\Phi_{\Delta\theta}(S_{\hat{\jmath},l-1}^{a} + y\Delta\theta) - S_{\hat{\jmath},l-1}^{a}| \\ &\quad\leq |\Phi_{\Delta\theta}(S_{\hat{\jmath},l-1}^{a} + y\Delta\theta) - S_{\hat{\jmath},l-1}^{a} - y\Delta\theta| + y\Delta\theta \\ &\quad\leq c_1\Delta\theta + c_2 S_{\hat{\jmath},l-1}^{a} + (c_2+1)y\Delta\theta. \end{aligned} \tag{69}$$

Furthermore, from Lemma 5

$$\lim_{\Delta\theta\to 0} \frac{\Phi_{\Delta\theta}(S_{\hat{\jmath},l-1}^{a} + y\Delta\theta) - S_{\hat{\jmath},l-1}^{a}}{\Delta\theta} = \frac{\partial S_{\hat{\jmath},l-1}^{a}}{\partial\theta} + y. \tag{70}$$

The above equation together with (59) and (19) shows that

$$\begin{aligned} &\lim_{\Delta\theta\to 0} \left[\frac{\Delta R}{\Delta\theta} + \frac{\Phi_{\Delta\theta}(S_{\hat{\jmath},l-1}^{a} + y\Delta\theta) - S_{\hat{\jmath},l-1}^{a}}{\Delta\theta}\right] \\ &\quad= Y_{j,l}^{1} + y \text{ w.p.1.} \end{aligned} \tag{71}$$

We can now divide (68) by $\Delta\theta^2$ and take the limit as $\Delta\theta \to 0$. The dominated convergence theorem can be applied here in view of (69) and (62). Then, making use of (71) and the same argument as in (65), we obtain

$$\begin{aligned} &\lim_{\Delta\theta\to 0} \frac{1}{\Delta\theta^2} E[\tilde{I}_{j,l}|z_{j-1,l}] \\ &\quad= \int_0^\infty [Y_{j,l}^{I} - y]^{+} g_{j,l}^{I}(0|z_{j-1,l})\, dy \\ &\quad= \frac{1}{2} g_{j,l}^{I}(0|z_{j-1,l})([Y_{j,l}^{I}]^{+})^2. \end{aligned}$$

■

Acknowledgment

The authors would like to thank W.-B. Gong for several suggestions and helpful comments, especially for the work presented in Section VII.

References

[1] G. A. Baker, *Essentials of Padé Approximants.* New York: Academic, 1975.

[2] G. Bao, C. G. Cassandras, and M. A. Zazanis, "Second derivative estimators for serial closed queueing networks," Dept. Electrical Computer Engineering, Univ. Massachusetts, Tech. Rep., Feb. 1996.

[3] D. P. Bertsekas, "Dynamic behavior of shortest path routing algorithms for communication networks," *IEEE Trans. Automat. Contr.*, vol. AC-27, pp. 60–74, 1982.

[4] D. P. Bertsekas, E. M. Gafni, and R. G. Gallager, "Second derivative algorithms for minimum delay distributed routing in networks," *IEEE Trans. Commun.*, vol. COM-32, pp. 911–919, 1984.

[5] P. Billingsley, *Probability and Measure.* New York: Wiley, 1979.

[6] X. R. Cao, "Realization probability in closed Jackson queueing networks and its application,"*Adv. Appl. Prob.*, vol. 19, pp. 708–738, 1987.

[7] ———, "The basic concepts of perturbation analysis of closed queueing networks with general service time distributions," in *Proc. 29th IEEE Conf. Decision Contr.*, 1990, pp. 2833–2838.

[8] C. G. Cassandras, M. Abidi, and D. Towsley, 'Distributed routing with on-line marginal delay estimation," *IEEE Trans. Commun.*, vol. 38, pp. 348–359, 1990.

[9] E. K. P. Chong and P. J. Ramadge, "Convergence of recursive optimization algorithms using IPA derivative estimates," *J. Discrete-Event Dynamic Syst.*, vol. 1, no. 4, pp. 339–372, 1992.

[10] M. C. Fu and J. Q. Hu, "Second derivative sample path estimators for the GI/G/m queue," *Magmt. Sci.*, submitted.

[11] ———, "Extensions and generalizations of smoothed perturbation analysis in a generalized semi-Markov process framework," *IEEE Trans. Automat. Contr.*, vol. 37, pp. 1483–1450, 1992.

[12] ———, "Addendum to 'Extensions and generalizations of smoothed perturbation analysis in a generalized semi-Markov process framework,'" *IEEE Trans. Automat. Contr.*, submitted.

[13] ———, "On unbounded hazard rates for smoothed perturbation analysis," *J. Appl. Prob.*, to appear.

[14] P. Glasserman and W.-B. Gong, "Smoothed perturbation analysis for a class of discrete-event systems," *IEEE Trans. Automat. Contr.*, vol. 35, pp. 1218–1230, 1991.

[15] P. Glasserman, *Gradient Estimation Via Perturbation Analysis.* Boston, MA: Kluwer, 1991.

[16] P. Glynn, "Likelihood ratio gradient estimation: An overview," in *Proc. 1987 Winter Simulation Conf.*, pp. 336–375.

[17] W. B. Gong and Y. C. Ho, "Smoothed perturbation analysis of discrete-event dynamic systems," *IEEE Trans. Automat. Contr.*, vol. AC-32, pp. 858–866, 1987.

[18] W. Gong, S. Nananukul, and A. Yan, "Padé approximation for stochastic discrete-event systems," *IEEE Trans. Automat. Contr.*, vol. 40, no. 8, pp. 1349–1358, 1995.

[19] Y. C. Ho and X. R. Cao, "Perturbation analysis and optimization of queueing networks," *J. Optim. Theory Applicat.*, vol. 40, pp. 559–582, 1983.

[20] Y. C. Ho, X. R. Cao, and C. G. Cassandras, "Infinitesimal and finite perturbation analysis for queueing networks," *Automatica*, vol. 19, pp. 439–445, 1983.

[21] Y. C. Ho and X. R. Cao, *Perturbation Analysis of Discrete-Event Dynamic Systems.* Boston, MA: Kluwer, 1991.

[22] M. Reiman and A. Weiss, "Sensitivity analysis for simulations via likelihood ratios," *Operations Res.*, vol. 37, pp. 830–844, 1989.

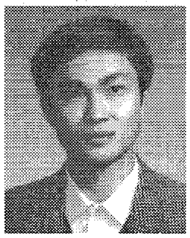
[23] R. Y. Rubinstein, *Monte Carlo Optimization, Simulation, and Sensitivity Analysis of Queueing Networks.* New York: Wiley, 1986.

[24] R. Suri, "Perturbation analysis: The state of the art and research issues explained via the GI/G/1 queue," *Proc. IEEE*, vol. 77, no. 1, pp. 114–137, 1989.

[25] R. Suri and M. A. Zazanis, "Perturbation analysis gives strongly consistent sensitivity estimates for the M/G/1 queue," *Mgmt. Sci.*, vol. 34, no. 1, pp. 39–64, 1988.

[26] M. A. Zazanis and R. Suri, "Perturbation analysis of the GI/GI/1 queue," *Queueing Syst.*, vol. 18, pp. 199–248, 1994.

[27] M. A. Zazanis, "Derivative estimation via compensators: Event averages in queueing systems," *Lett. Operations Res.*, vol. 17, pp. 77–84, 1995.



Gang Bao received the B.E. degree from Huazhong University of Science and Technology, Wuhan, China, in 1984, the M.S. degree from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 1987, and the Ph.D. degree from the University of Massachusetts, Amherst, in 1994.

He joined Qualcomm Inc., San Diego, CA, in 1994, where he is currently working on modeling and performance evaluation of CDMA wireless communication systems. His interests include discrete-event systems, wireless communication, and computer networks.

Christos G. Cassandras (S'82–M'82–SM'91–F'95) received the B.S. degree from Yale University, New Haven, CT, in 1977, the M.S.E.E. degree from Stanford University, CA, in 1978, and the S.M. and Ph.D. degrees in 1979 and 1982, respectively, from Harvard University, Cambridge, MA.

From 1982–84, he was with ITP Boston, Inc., where he worked on control systems for computer-integrated manufacturing. He is a Professor of Electrical and Computer Engineering at the University of Massachusetts at Amherst, where he has been since 1984. His research interests include discrete-event systems, stochastic optimization, computer simulation, and performance evaluation and control of computer networks and manufacturing systems. He is the author of over 100 technical publications and a textbook.

Dr. Cassandras is on the Board of Governors of the Control Systems Society and Editor of Technical Notes and Correspondence for this TRANSACTIONS. He serves on several other editorial boards and has been a guest editor for various journals. He was awarded a Lilly Fellowship in 1991, and he is a member of Phi Beta Kappa and Tau Beta Pi.

Michael A. Zazanis was born in Athens, Greece, in 1959. He received the Diploma in electrical engineering from the National Technical University of Athens in 1982 and the M.S. and Ph.D. degrees in applied mathematics from Harvard University, Cambridge, MA, in 1983 and 1986, respectively.

He taught at Northwestern University, Chicago, IL, from 1986 to 1993. His interests include applied probability, queueing systems, and sensitivity analysis of discrete-event simulations.

Dr. Zazanis received the best publication award from The Institute of Management Sciences in 1990.