15th Triennial World Congress, Barcelona, Spain

PERTURBATION ANALYSIS OF MULTICLASS STOCHASTIC FLUID MODELS

Christos G. Cassandras [*,1] **Gang Sun** [*,1]
Christos G. Panayiotou [*,1] **Yorai Wardi** [**,2]

[*] *Dept. of Manufacturing Engineering, Boston University, Boston, MA 02215*
[**] *School of Electrical Engineering, Georgia Institute of Technology, Atlanta, GA*

Abstract: We use Stochastic Fluid Models (SFM) for control and optimization (rather than performance analysis) of communication networks, focusing on problems of admission control. We consider a SFM with an uncontrolled traffic class and a controlled traffic class subject to threshold-based admission control. We derive gradient estimators for packet loss and workload related performance metrics with respect to threshold parameters. These estimators are shown to be unbiased and directly observable from a sample path without any knowledge of underlying stochastic characteristics. This renders them computable in on-line environments and easily implementable for network management and control. We further demonstrate their use in admission control problems where our SFM-based estimators are evaluated based on data from an actual system.

Keywords: Stochastic Fluid Models, Perturbation Analysis, Communication Networks, Discrete Event Systems.

1. INTRODUCTION

A natural modeling framework for communication networks is provided through queueing systems, which capture the discrete event nature of packet-based operations. However, the huge traffic volume that networks are supporting today makes such models highly impractical. Moreover, incorporating sophisticated stochastic processes and modeling buffer overflow phenomena typically defy tractable analytical derivations. This has motivated an alternative modeling paradigm based on Stochastic Fluid Models (SFM). SFMs have recently been shown to be especially useful for simulating various kinds of high-speed networks (Kesidis *et al.*, 1996), (Kumaran and Mitra, 1998), (Miyoshi, 1998), (Liu *et al.*, 1999), (Yan and Gong, 1999), (Wardi and Melamed, 2000).

For the purpose of *performance analysis* with Quality of Service (QoS) requirements, the accuracy of SFMs depends on traffic conditions, the structure of the underlying system, and the nature of the performance metrics of interest. For the purpose of *control and optimization*, on the other hand, as long as a SFM captures the salient features of the underlying "real" system it is possible to obtain accurate solutions to problems even if we cannot estimate the corresponding performance with accuracy. In short, an SFM may be too "crude" for some performance analysis purposes,

[1] Supported in part by the National Science Foundation under grants EEC-0088073 and ACI-9873339, by AFOSR under contract F49620-98-1-0387, by the Air Force Research Laboratory under contract F49620-01-1-0056, by EPRI/ARO under contract WO8333-03 and by the Nokia Research Center (Boston).
[2] Supported in part by the National Science Foundation under grant DMI-0085659 and by DARPA under contract F30602-00-2-0556.

but able to accurately capture sensitivity information. This point of view is taken in (Cassandras *et al.*, 2001), where a SFM is adopted for a network node in which threshold-based admission control is exercised. For the problem of determining a threshold (measured in packets or bytes) that minimizes a weighted sum of loss volume and buffer content, it is shown that a solution based on a SFM recovers or gives close approximations to the solution of the associated queueing model. Since solving such problems usually relies on gradient information, estimating the gradient of a given cost function with respect to key parameters, such as the aforementioned threshold, becomes an essential task. Perturbation Analysis (PA) methods (Ho and Cao, 1991), (Cassandras and Lafortune, 1999) are therefore suitable, if appropriately adapted to a SFM viewed as a discrete-event system. This approach has been used in (Liu and Gong, 1999), where incoming traffic rates were the parameters of interest, and (Cassandras *et al.*, 2001), where threshold parameters are optimized to solve admission control problems. In (Cassandras *et al.*, 2001), in particular, it was shown that Infinitesimal Perturbation Analysis (IPA) yields remarkably simple nonparametric sensitivity estimators for packet loss and workload metrics with respect to threshold or buffer size parameters in a single-node SFM with a single incoming traffic stream. In addition, the estimators obtained are unbiased under very weak structural assumptions on the defining traffic processes.

In this paper, we consider a single node SFM with two traffic streams; one traffic stream is uncontrolled and the other is subject to threshold-based admission control (see Fig. 1). Thus, we model a typical network node where the controlled stream represents a source of new traffic into the network at that node and the uncontrolled stream represents "interfering traffic", i.e., traffic from other nodes on its way to various destinations. Interestingly, this model also captures the operation of the Differentiated Services (DS) protocol that has been proposed for supporting QoS requirements (Blake *et al.*, 1998). We derive IPA gradient estimators for performance metrics related to loss and workload levels with respect to the threshold parameter in our model. These estimators can be evaluated *based on data observed on a sample path of the actual (discrete-event) system.* Thus, we may use the SFM only to obtain a gradient estimator form; the associated value at any operating point is obtained on line from real system data and no simulation is necessary. The estimators derived are also shown to be unbiased. Finally, we use these estimators to illustrate how to solve admission control problems.

2. A MULTICLASS STOCHASTIC FLUID MODEL (SFM)

The SFM studied in this paper is based on the model described in (Cassandras *et al.*, 2001) where a single node and single traffic stream was considered. In our case, as shown in Fig. 1, there are two "classes" of traffic: controlled (class 1) and uncontrolled (class 2). Uncontrolled traffic has an arrival rate $\alpha_2(t)$. A threshold θ is associated with class 1 traffic, which has an arrival rate $\alpha_1(t)$. An admission control policy is exercised so that when the total buffer content reaches a threshold θ, class 1 traffic is rejected, while class 2 traffic is not affected. The two traffic streams share a common FIFO buffer assumed of infinite size. The service rate is denoted by $\beta(t)$. In addition, let $\gamma(\theta;t)$ denote the loss rate when the buffer content exceeds the designated threshold level θ, and let $x(\theta;t)$ denote the buffer content at time t. The notational dependence on θ indicates that we will analyze performance metrics as functions of the given θ.

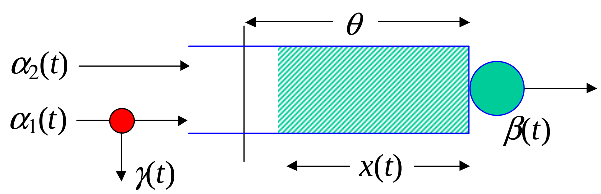

Fig. 1. Stochastic fluid model (SFM) with two traffic classes

We are interested in studying sample paths of the SFM over a time interval $[0,T]$ for a given fixed $0 < T < \infty$. We assume that the processes $\{\alpha_1(t)\}$, $\{\alpha_2(t)\}$, and $\{\beta(t)\}$ are independent of θ and they are piecewise continuously differentiable w.p.1. Note that a typical sample path can be decomposed into two kinds of alternating intervals: *empty periods* and *busy periods*. Empty Periods (EP) are maximal intervals during which the buffer is empty, while Busy Periods (BP) are supremal intervals during which the buffer is nonempty. Observe that during an EP the system is not necessarily idle since the server may be active, processing traffic supplied to it at a rate that does not exceed $\beta(t)$, i.e., $\alpha_1(t) + \alpha_2(t) - \beta(t) \leq 0$.

Viewed as a discrete-event system, an *event* in a sample path of the above SFM may be either *exogenous* or *endogenous*. An exogenous event is a jump in $\alpha_1(t)$, $\alpha_2(t)$, or $\beta(t)$ (if any exist) and any point where the difference function $\{\alpha_1(t) + \alpha_2(t) - \beta(t)\}$ or $\{\alpha_2(t) - \beta(t)\}$ changes sign. An endogenous event is defined to occur whenever the buffer (i) ceases to be empty, (ii) becomes empty, (iii) reaches the value $x(\theta;t) = \theta$ and then maintains it for some finite length of time, (iv)

leaves the value $x(\theta;t)=\theta$ after it has maintained it for some finite length of time, and (*v*) crosses the value $x(\theta;t)=\theta$ from either below or above.

We will assume that the real-valued parameter θ is confined to a closed and bounded (compact) interval Θ. Let $\mathcal{L}(\theta):\Theta\rightarrow\mathbb{R}$ be a random function defined over the underlying probability space $(\Omega,\mathcal{F},P)$. Strictly speaking, we write $\mathcal{L}(\theta,\omega)$ to indicate that this sample function depends on the sample point $\omega\in\Omega$, but will suppress ω unless it is necessary to stress this fact. In what follows, we will consider two performance metrics, the *Loss Volume* $L_T(\theta)$ and the *Cumulative Workload* (or just *Work*) $Q_T(\theta)$, both defined on the interval $[0,T]$ as follows:

$$L_T(\theta) \;=\; \int_0^T \gamma(\theta;t)dt, \qquad (1)$$

$$Q_T(\theta) \;=\; \int_0^T x(\theta;t)dt, \qquad (2)$$

where, for simplicity, we assume that $x(\theta;0)=0$. Observe that $\frac{1}{T}E\left[L_T(\theta)\right]$ is the *Expected Loss Rate* over the interval $[0,T]$, a common performance metric of interest (from which related metrics such as *Loss Probability* can also be derived). Similarly, $\frac{1}{T}E\left[Q_T(\theta)\right]$ is the *Expected Buffer Content* over $[0,T]$. We may then formulate optimization problems such as the determination of θ^* that minimizes a cost function of the form

$$J(\theta)=\frac{1}{T}E\left[Q_T(\theta)\right]+\frac{R}{T}E\left[L_T(\theta)\right] \qquad (3)$$

where R represents a rejection cost due to class 1 loss. In order to accomplish this task, we rely on estimates of $dE\left[L_T(\theta)\right]/d\theta$ and $dE\left[Q_T(\theta)\right]/d\theta$, which we will pursue through Infinitesimal Perturbation Analysis (IPA) techniques (Ho and Cao, 1991), (Cassandras and Lafortune, 1999)). Henceforth we shall use the "prime" notation to denote derivatives with respect to θ. Thus, the sample derivatives $dL_T(\theta)/d\theta$ and $dQ_T(\theta)/d\theta$ are denoted by $L_T^{'}(\theta)$ and $Q_T^{'}(\theta)$, respectively.

3. IPA FOR LOSS VOLUME WITH RESPECT TO THRESHOLD

We proceed by studying a sample path of the SFM over $[0,T]$. For a fixed $\theta\in\Theta$, the interval $[0,T]$ is divided into alternating EPs and BPs. Suppose that a sample path consists of K busy periods denoted by $\mathcal{B}_k$, $k=1,\ldots,K$, in increasing order. Thus, given a BP $\mathcal{B}_k$, its starting point is one where the buffer ceases to be empty, i.e., there is a change in sign of the difference function $\{\alpha_1(t)+\alpha_2(t)-\beta(t)\}$ from non-positive (hence, the buffer was empty) to positive. Since this function is independent of θ, the starting point of $\mathcal{B}_k$ is locally independent of θ. The ending point of $\mathcal{B}_k$ generally depends on θ. Denoting these points by ξ_k and $\eta_k(\theta)$ respectively, we express $\mathcal{B}_k$ as

$$\mathcal{B}_k \;=\; [\xi_k,\eta_k(\theta)), \qquad k=1,\ldots,K$$

for some random integer K. Then, by (1), we may write $L_T(\theta)$ and its derivative (assuming it exists) as

$$L_T(\theta) \;=\; \sum_{k=1}^{K}\int_{\xi_k}^{\eta_k(\theta)}\gamma(\theta;t)dt \qquad (4)$$

$$L_T^{'}(\theta) \;=\; \sum_{k=1}^{K}\frac{d}{d\theta}\int_{\xi_k}^{\eta_k(\theta)}\gamma(\theta;t)dt \qquad (5)$$

where K is locally independent of θ. Let us now focus on a typical $\mathcal{B}_k$ and drop the index k in order to simplify notation. Thus, the BP in question is denoted by $\mathcal{B}=[\xi,\eta(\theta))$. Define the function $\lambda(\theta)$ as

$$\lambda(\theta) \;=\; \int_{\xi}^{\eta(\theta)}\gamma(\theta;t)dt. \qquad (6)$$

and we shall concentrate on evaluating $\lambda^{'}(\theta)$. Let v_i, $i=0,...,S$, be all endogenous event times in the BP (as previously defined). Note that $v_0=\xi$ and $v_S=\eta(\theta)$. Figure 2 shows a typical BP in a sample path of our SFM. According to the different levels of buffer content, we can divide the BP into intervals $[v_{i-1}(\theta),v_i(\theta))$, $i=1,...,S$ so that each belongs to one of the following three sets:

1. **Partial Loss Period Set** $U(\theta)$. During such periods, the buffer content is $x(t;\theta)=\theta$ and class 1 traffic experiences partial loss. In particular,

$$\frac{dx(t)}{dt^+}=0 \qquad (7)$$

$$\alpha_1(t)+\alpha_2(t)-\beta(t)>0 \qquad (8)$$

$$\alpha_2(t)-\beta(t)<0 \qquad (9)$$

$$\gamma(\theta;t)=\alpha_1(t)+\alpha_2(t)-\beta(t) \qquad (10)$$

Formally, we define $U(\theta)$ as follows:

$$U(\theta):=\{[v_{i-1}(\theta),v_i):x(t)=\theta,\quad t\in[v_{i-1}(\theta),v_i)\} \qquad (11)$$

where v_i above is locally independent of θ, since the time when the buffer content leaves θ depends only on a change in sign of the difference function $\{\alpha_1(t)+\alpha_2(t)-\beta(t)\}$ or $\{\alpha_2(t)-\beta(t)\}$, as seen in (8)-(9). In Fig. 2, $[v_3,v_4)$ and $[v_5,v_6)$ are examples of partial loss periods within a BP.

2. **Full Loss Period Set** $V(\theta)$. In a Full Loss period the buffer content is $x(t;\theta)>\theta$ (excluding the starting point $v_{i-1}(\theta)$) and *all* class 1 traffic is lost:

$$V(\theta):=\{[v_{i-1}(\theta),v_i(\theta)):x(v_{i-1}(\theta))=\theta \text{ and } x(t)>\theta,\quad t\in(v_{i-1}(\theta),v_i(\theta))\} \qquad (12)$$

and we have

$$\frac{dx(t)}{dt^+}=\alpha_2(t)-\beta(t) \qquad (13)$$

$$\gamma(\theta;t)=\alpha_1(t) \tag{14}$$

Examples of Full Loss periods are $[v_1,v_2)$ and $[v_6,v_7)$ in Fig. 2.

3. **No Loss Period Set** $W(\theta)$. During such periods the buffer content is $x(t;\theta)<\theta$ (excluding the starting point $v_{i-1}(\theta)$) and no loss occurs:

$$\begin{aligned} W(\theta) :=\{&[v_{i-1}(\theta),v_i(\theta)) : x(\xi)=0 \text{ or} \\ & x(v_{i-1}(\theta))=\theta,\quad i>1 \text{ and} \\ & x(t)<\theta,\quad t\in(v_{i-1}(\theta),v_i(\theta))\} \end{aligned} \tag{15}$$

and we have

$$\frac{dx(t)}{dt^+}=\alpha_1(t)+\alpha_2(t)-\beta(t) \tag{16}$$

$$\gamma(\theta;t)=0 \tag{17}$$

Examples of such periods are $[\xi,v_1)$, $[v_2,v_3)[v_4,v_5)$ and $[v_7,\eta)$ in Fig. 2.

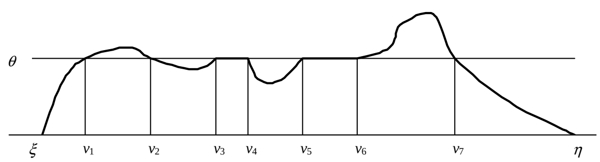

Fig. 2. A typical busy period (BP)

Recalling (6), we can now write $\lambda'(\theta)$ as

$$\begin{aligned} \lambda'(\theta)=\sum_{i=1}^{S-1}\Bigg(&\frac{d}{d\theta}\int_{v_{i-1}(\theta)}^{v_i(\theta)}\gamma(\theta;t)dt\cdot \\ &\mathbf{1}\left[[v_{i-1}(\theta),v_i(\theta))\in U(\theta)\cup V(\theta)\right]\Bigg) \end{aligned} \tag{18}$$

where S is also locally independent of θ. Since we are concerned with the sample derivative $L_T^{'}(\theta)$ we have to identify conditions under which it exists (and, therefore, $\lambda^{'}(\theta)$ also exists). Observe that any endogenous event time (as defined above) is generally a function of θ (with the exceptions of local independence noted above). The derivative $v_i'(\theta)$ exists as long as $v_i(\theta)$ is not a jump point of the difference function $\{\alpha_1(t)+\alpha_2(t)-\beta(t)\}$ or $\{\alpha_2(t)-\beta(t)\}$. Excluding the possibility of the simultaneous occurrence of two (exogenous or endogenous) events, the only situation preventing the existence of the sample derivative $L_T^{'}(\theta)$ involves some t such that $\alpha(t)-\beta(t)=0$ or $\alpha_1(t)+\alpha_2(t)-\beta(t)=0$. In such cases, the one-sided derivative of $L_T(\theta)$ exists and can be obtained through a finite difference analysis as in (Cassandras *et al.*, 2001). However, to keep the analysis simple, we focus only on the differentiable case by proceeding under the following technical conditions:

Assumption 1.
a. W.p.1, $\alpha_2(t)-\beta(t)\neq 0$ and $\alpha_1(t)+\alpha_2(t)-\beta(t)\neq 0$, for all $t\in[0,T]$.
b. For every $\theta\in\Theta$, w.p.1, no two *events* may occur at the same time.

In order to proceed with the detailed derivation of $\lambda^{'}(\theta)$, we begin by simplifying notation through the introduction of the following two operators for $i=1,...,S$:

$$A_i\equiv\alpha_1(v_i(\theta))+\alpha_2(v_i(\theta))-\beta(v_i(\theta)) \tag{19}$$

$$B_i\equiv\alpha_2(v_i(\theta))-\beta(v_i(\theta)) \tag{20}$$

By convention, we shall set $A_0\equiv 1$. The following lemma shows that all event time derivatives of interest, $v_i'(\theta)$, are expressed in terms of these operators. Moreover, we establish the fact that after a Partial Loss period occurs (if any is present), all ensuing event time derivatives are $v_i'(\theta)=0$ (all proofs are omitted due to space limitations).

Lemma 1. Suppose that $[v_m(\theta),v_{m+1}), 1\leq m<S$ is the first Partial Loss period in a BP. Then:

(1) For $v_i\leq v_m$:

$$v_1^{'}(\theta)=\frac{A_0}{A_1} \tag{21}$$

$$v_{2n}^{'}(\theta)=\prod_{i=2,\ldots,2n}\frac{B_{i-1}}{B_i}\cdot\frac{A_{i-2}}{A_{i-1}} \tag{22}$$

where $1\leq n\leq\frac{m}{2}$ if m is even, and $2\leq n\leq\frac{m-1}{2}$ if m is odd, and $m>1$.

$$v_{2n+1}^{'}(\theta)=\frac{A_{2n}}{A_{2n+1}}\cdot\left\{\prod_{i=2,\ldots,2n}\frac{B_{i-1}}{B_i}\cdot\frac{A_{i-2}}{A_{i-1}}\right\} \tag{23}$$

where $1\leq n\leq\frac{m-2}{2}$ if m is even, and $1\leq n\leq\frac{m-1}{2}$ if m is odd, and $m>2$.

(2) For all $v_i\geq v_{m+1}$: $v_i^{'}(\theta)=0$

The next lemma provides an expression for the derivative $\lambda^{'}(\theta)$ in (18).

Lemma 2. For any BP $[\xi,\eta(\theta))$, if at least one Partial Loss period is present, then

$$\lambda^{'}(\theta)=-1. \tag{24}$$

If no Partial Loss period is present, then

$$\lambda^{'}(\theta)=-1+\prod_{i=2,4,\ldots,S-1}\frac{A_i}{B_i}\cdot\frac{B_{i-1}}{A_{i-1}} \tag{25}$$

and

$$-1<\lambda^{'}(\theta)\leq 0. \tag{26}$$

Motivated by our analysis thus far, let $U_k(\theta)$, $V_k(\theta)$, and $W_k(\theta)$ be the Partial Loss, Full Loss, and No Loss period sets respectively in the kth BP, $k=1,\ldots,K$. Similarly, let $v_{k,i}(\theta)$ denote the ith event time in the kth BP, $i=0,\ldots,S_k$. Then, define

$$\Phi(\theta)=\{k\in\{1,\ldots,K\}\ :\ U_k(\theta)\neq\emptyset\} \tag{27}$$

to be the set of BPs containing at least one Partial Loss period, and set

$$\lambda_k^{'}(\theta)=-1+\prod_{i=2,4,\ldots,S_k-1}\frac{A_{k,i}}{B_{k,i})}\cdot\frac{B_{k,i-1}}{A_{k,i-1}} \tag{28}$$

Theorem 1. The sample derivative $L_T^{'}(\theta)$ is given by

$$L_T^{'}(\theta) = -\sum_{k=1}^{K} \mathbf{1}[k \in \Phi(\theta)] + \sum_{k=1}^{K} \mathbf{1}[k \notin \Phi(\theta)]\lambda_k^{'}(\theta) \tag{29}$$

where K is the (random) number of busy periods contained in $[0, T]$, including a possibly incomplete last busy period.

The expression in (29) provides the IPA estimator for the loss metric defined in (1). Note that $L_T^{'}(\theta)$ above does not depend on any distributional information regarding the traffic arrival and service processes except for the rates at event times $v_{k,i}(\theta)$ which may be readily estimated. If BPs include at least one Partial Loss period, then the only implementation requirement is that such a period be detected and the contribution of this entire BP is simply -1.

4. IPA FOR WORK WITH RESPECT TO THRESHOLD

In this section we derive the IPA estimator for the Cumulative Workload (or simply Work) defined in (2) by carrying out an analysis similar to that of the previous section under **Assumption 1**. First, note that we can write

$$Q_T(\theta) = \sum_{k=1}^{K} \int_{\xi_k}^{\eta_k(\theta)} x(\theta;t)dt \tag{30}$$

$$Q_T^{'}(\theta) = \sum_{k=1}^{K} \frac{d}{d\theta} \int_{\xi_k}^{\eta_k(\theta)} x(\theta;t)dt. \tag{31}$$

where, as before, we consider BPs $\mathcal{B}_k = [\xi_k, \eta_k(\theta))$, $k = 1, \dots, K$. Then, focusing on a particular $\mathcal{B}_k$ and dropping the index k, we define

$$q(\theta) = \int_{\xi}^{\eta(\theta)} x(\theta;t)dt. \tag{32}$$

Taking the derivative with respect to θ yields

$$\begin{aligned} q'(\theta) &= \int_{\xi}^{\eta(\theta)} x'(\theta;t)dt + x(\theta;\eta(\theta))\eta'(\theta) \qquad (33) \\ &= \int_{\xi}^{\eta(\theta)} x'(\theta;t)dt \end{aligned}$$

since the BP ends at $\eta(\theta)$, hence $x(\theta;\eta(\theta)) = 0$. We can evaluate $x'(\theta;t)$ by considering all possible cases regarding the location of t in the BP $\mathcal{B} = [\xi, \eta(\theta))$. The following lemma provides an explicit expression for $q'(\theta)$.

Lemma 3. Suppose that $[v_m(\theta), v_{m+1}), 1 \leq m < S$, is the first Partial Loss period in a BP. Then,

$$q^{'}(\theta) = \sum_{i=1}^{m-1} (v_{i+1} - v_i)\phi_i + (v_S - v_m) \tag{34}$$

where

$$\phi_i = \begin{cases} 1 - B_i v_i^{'}, \ i \text{ odd} \\ 1 - A_i v_i^{'}, \ i \text{ even} \end{cases} \tag{35}$$

$$B_1 v_1^{'}(\theta) = \frac{B_1}{A_1} \tag{36}$$

$$A_{2n} v_{2n}^{'}(\theta) = \prod_{i=2,..,2n} \frac{A_i}{B_i} \cdot \frac{B_{i-1}}{A_{i-1}}, \ n > 0 \tag{37}$$

$$B_{2n+1} v_{2n+1}^{'}(\theta) = \frac{B_{2n+1}}{A_{2n+1}} \left\{ \prod_{i=2,..,2n} \frac{A_i}{B_i} \cdot \frac{B_{i-1}}{A_{i-1}} \right\}, \ n > 0 \tag{38}$$

It should be clear that if the BP does not contain a Partial Loss period, then $q^{'}(\theta)$ is given by the sum in (34) evaluated over all $i = 1, \dots, S-1$. Next, letting $v_{k,i}(\theta)$ denote the ith event time in the kth BP, $i = 0, \dots, S_k$, and defining $q_k^{'}(\theta)$ as the obvious extension to $q^{'}(\theta)$, similar to (28), we get the following.

Theorem 2. The sample derivative $Q_T^{'}(\theta)$ is given by

$$Q_T^{'}(\theta) = \sum_{k=1}^{K} q_k^{'}(\theta) \tag{39}$$

where K is the (random) number of busy periods contained in $[0, T]$, including a possibly incomplete last busy period.

The expression in (29) provides the IPA estimator for the work metric defined in (2). Its implementation requires the same information as that for loss metric with the addition of timers to measure the duration of periods $[v_{k,i}, v_{k,i+1})$ within each BP observed in $[0, T]$, as well as $(v_{k,S_k} - v_{k,m_k})$ if one or more Partial Loss periods are included, with the first one starting at v_{k,m_k}.

5. IPA UNBIASEDNESS

In general, the unbiasedness of an IPA derivative $\mathcal{L}'(\theta)$ has been shown to be ensured by the following two conditions (see (Rubinstein and Shapiro, 1993), Lemma A2, p.70):

Condition 1. For every $\theta \in \Theta$, the sample derivative $\mathcal{L}'(\theta)$ exists w.p.1.

Condition 2. W.p.1, the random function $\mathcal{L}(\theta)$ is Lipschitz continuous throughout Θ, and the (generally random) Lipschitz constant has a finite first moment.

Consequently, establishing the unbiasedness of $L_T^{'}(\theta)$ and $Q_T^{'}(\theta)$ as estimators of $dE[L_T(\theta)]/d\theta$ and $dE[Q_T(\theta)]/d\theta$, respectively, reduces to verifying the Lipschitz continuity of $L_T(\theta)$ and $Q_T(\theta)$

with appropriate Lipschitz constants. Due to space limitations, we omit the proof of this fact and only quote our final unbiasedness result.

Theorem 3. Let $N(T)$ be the number of exogenous events in $[0, T]$ and assume $E[N(T)] < \infty$. Then, the derived IPA estimates $L_T^{'}(\theta)$ and $Q_T^{'}(\theta)$ of $dE[L_T(\theta)]/d\theta$ and $dE[Q_T(\theta)]/d\theta$, respectively, are unbiased.

6. OPTIMAL ADMISSION CONTROL USING SFM-BASED IPA ESTIMATORS

As already mentioned, the solution to an optimization problem defined for an actual network node (i.e., a node that operates as a queueing system) may be accurately approximated by the solution to the same problem based on a SFM of the node. However, this may not be always the case. On the other hand, the simple form of the IPA estimators of the Expected Loss Rate and Expected Buffer Content we have obtained allows us to use data from the *actual* (real-world) system in order to estimate sensitivities that, in turn, may be used to solve an optimization problem of interest. In other words, the *form* of the IPA estimators is obtained by analyzing the system as a SFM, but the associated *values* are based on real data.

Let us return to the admission control problem of (3) where we are trading off the expected loss rate of class 1 with a rejection penalty R for the expected queue length. For illustrative purposes, we have applied this approach, using the IPA estimators derived, to a system where $\alpha_1(t)$ is piecewise constant with values uniformly distributed over $[0, 12]$, with each constant rate period exponentially distributed with parameter 0.5; $\alpha_2(t)$ is piecewise constant with values uniformly distributed over $[0, 27]$, with each constant rate period exponentially distributed with parameter 0.3. The service rate is $\beta = 20$ and the rejection cost is $R = 25$. We assume the traffic and service rates are observable and use an estimation period $T = 200K$ time units. The cost curve (labelled DES) is shown in Fig. 3 (average over 30 sample paths). The optimization algorithm applied is similar to the one presented in (Cassandras *et al.*, 2001) and iterates on a discrete threshold parameter using the SFM-based estimators but with all required endogenous events being observed from an actual discrete-event sample path.

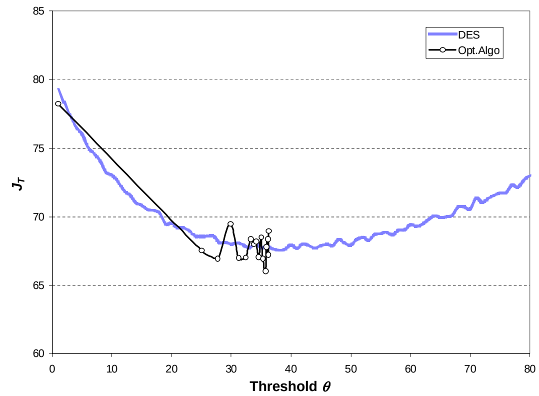

Fig. 3. Optimal threshold determination using SFM-based gradient estimators

7. REFERENCES

Blake, S., D. Black, M. Carlson, E. Davies, Z. Wang and W. Weiss (1998). An architecture for differentiated services. *RFC 2475*. http://www.ietf.org/rfc/rfc2475.txt.

Cassandras, C. G. and S. Lafortune (1999). *Introduction to Discrete Event Systems*. Kluwer Academic Publishers.

Cassandras, C.G., G. Sun and C.G. Panayiotou (2001). Stochastic fluid models for control and optimization of systems with quality of service requirements. In: *Proc. of the IEEE Conf. on Decision and Control*. To appear.

Ho, Y.C. and X.R. Cao (1991). *Perturbation Analysis of Discrete Event Dynamic Systems*. Kluwer Academic Publishers.

Kesidis, G., A. Singh, D. Cheung and W.W. Kwok (1996). Feasibility of fluid-driven simulation for ATM network. In: *Proceedings of IEEE GLOBECOM*. Vol. 3. pp. 2013–2017.

Kumaran, K. and D. Mitra (1998). Performance and fluid simulations of a novel shared buffer management system. In: *Proceedings of IEEE INFOCOM*.

Liu, B., Y. Guo, J. Kurose, D. Towsley and W.B. Gong (1999). Fluid simulation of large scale networks: Issues and tradeoffs. In: *Proc. of the Intl Conf on Parallel and Distributed Processing Techniques and Applications*.

Liu, Y. and W. Gong (1999). Perturbation analysis for stochastic fluid queueing systems. In: *Proc. 38th IEEE Conf. Dec. and Ctrl.* pp. 4440–4445.

Miyoshi, N. (1998). Sensitivity estimation of the cell-delay in the leaky bucket traffic filter with stationary gradual input. In: *Proc. of the Intl Workshop on Discrete Event Systems*.

Rubinstein, R.Y. and A. Shapiro (1993). *Discrete Event Systems: Sensitivity Analysis and Stochastic Optimization by the Score Function Method*. John Wiley and Sons.

Wardi, Y. and B. Melamed (2000). Loss volume in continuous flow models: Fast simulation and sensitivity analysis via IPA. In: *Proc. of IEEE Mediter. Conf. on Control and Automation*.

Yan, A. and W.B. Gong (1999). Fluid simulation for high-speed networks with flow-based routing. *IEEE Transactions on Information Theory* **45**, 1588–1599.