Proceedings of the
47th IEEE Conference on Decision and Control
Cancun, Mexico, Dec. 9-11, 2008

IPA for Fluid Queue with Delayed Flow Control

Y. Wardi and G. Riley

Abstract— **This paper develops a general framework for Infinitesimal Perturbation Analysis (IPA) for fluid queues, and applies it to a queue with flow control. Unlike previous works, we consider signal delays in the feedback channel. The performance measure of interest is the loss volume, and the variable parameters are the buffer size and the feedback gain. The IPA derivatives are characterized, and simulation results are presented.**

Index Terms— **Infinitesimal perturbation analysis, stochastic hybrid systems, stochastic flow models.**

I. INTRODUCTION

Fluid queues have been considered as a natural setting for *Infinitesimal Perturbation Analysis (IPA)* in the past few years (see [4] for a recent survey). The reason is that, for a large class of systems and functions, they yield statistically unbiased derivatives (gradients), whose formulas are model-free and simple to compute. Moreover, they have been shown experimentally to have a certain measure of robustness with respect to modeling variations, and hence the IPA formulas that are derived from fluid-queue models can provide reliable sensitivity estimates when applied to sample paths that are observed from discrete systems. All of this suggests that the analysis of IPA in the fluid-queue setting has the potential for sample-based optimization of high-speed networks, be it in the setting of simulation or real-time parameter control [3].

The earliest results concerned IPA for loss-related performance measures as functions of the buffer size and other parameters [3], and the resulting IPA derivatives were shown to be computable via simple formulas. Extensions to queueing networks were carried out in [7], and further extensions to flow-control systems were obtained in [10], [11], [8], [1].

An abstract, hybrid-system framework was developed in [9], and it seems to cover many of the previous results concerning IPA in the setting of a single-stage SFM. However, it assumes that the feedback signal is delivered instantly to the source, and hence it excludes delayed control. The purpose of this paper is to extend that framework to the case of delayed feedback signals, and apply it to a queue with flow control. In particular, we consider the IPA derivative of the loss volume as a function of the buffer size and feedback gain. It must be pointed out that the presence of delays in the feedback loop results not only in more tedious analysis, but also in more complicated IPA derivatives, whose computation requires an iterative algorithm instead of a simple formula.

The authors are with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, Georgia, 30332, USA {ywardi,riley}@ece.gatech.edu

The rest of the paper is organized as follows: Section II presents the abstract framework for IPA, while Section III concerns a fluid queue with additive flow control. Section IV provides simulation examples, and Section V concludes the paper.

II. ABSTRACT FRAMEWORK FOR IPA

Consider a hybrid system with bi-level dynamics, having continuous-time dynamics at the lower level and discrete-event dynamics at the higher level. The system is assumed to operate over a given finite-length time-interval $[0,T]$, where a finite number (N) of events occur. We are not concerned here with the details of the discrete-event dynamics, but only with the times at which the events occur, which we denote by $\tau_1,\ldots,\tau_N$, in increasing order. We further define $\tau_0 := 0$ and $\tau_{N+1} := T$, and denote the interval $[\tau_{i-1},\tau_i]$ by I_i, $i=1,\ldots,N+1$.

Let θ be a scalar parameter assumed to be confined to a closed, finite-length interval Θ, and suppose that the event-times $\tau_i(\theta)$ are functions of θ; we will use the simpler notation $\tau_i := \tau_i(\theta)$, and similarly for the derivatives, $\frac{d\tau_i}{d\theta} := \frac{d\tau_i(\theta)}{d\theta}$, when no confusion arises.

The continuous-time dynamics evolve in the interval I_i in the following way. Let $x\in R^n$ denote the continuous state variable. There exists a function $f_i : R^n\times\Theta\times[0,T]\rightarrow R^n$ such that

$$\dot{x} \;=\; f_i(x,\theta,t), \tag{1}$$

where"dot" denotes derivative with respect to t. The state variable x depends on θ and t, namely $x=x(\theta,t)$, and $\dot{x} := \frac{d}{dt}x(\theta,t)$. Suppose that $\theta\in\Theta$ is fixed throughout the evolution of the system in the interval $t\in[0,T]$, and assume that the functions f_i $(i=1,\ldots,N+1)$ are continuously differentiable. We also assume that the state variable $x(\theta,t)$ is continuous in t at the event-times τ_i, and therefore it is uniquely defined by (1) once the initial state $x(\theta,0)$ is specifiied. Finally, we mention that the event-times $\tau_i(\theta)$ and the functions $f_i(x,\theta,t)$ are all random and defined over a common probability space $(\Omega,\mathcal{F},P)$. In the forthcoming description we will consider a realization corresponding to a sample path $\omega\in\Omega$, and hence regard the various functions as implicitly dependent upon that sample path. We further assume that all of the derivative terms, mentioned in the sequel, exist, and later will state assumptions guaranteeing their existence w.p.1.

Let $L_i : R^n\times\Theta\times I_i\rightarrow R$ be a continuously-differentiable function, and let $J_i(\theta)$ be defined by

$$J_i(\theta) \;=\; \int_{\tau_{i-1}}^{\tau_i} L_i(x(\theta,t),\theta,t)dt. \tag{2}$$

978-1-4244-3124-3/08/$25.00 ©2008 IEEE

The cost function we are interested in is $J(\theta)$, defined by

$$J(\theta) \ = \ \sum_{i=1}^{N+1} J_i(\theta). \quad (3)$$

Our main interest is in the sample derivative $\frac{dJ}{d\theta} := \frac{dJ(\theta)}{d\theta}$ (the explicit dependence on θ will be suppressed when no confusion arises), and this is the IPA derivative. According to (3),

$$\frac{dJ}{d\theta} \ = \sum_{i=1}^{N+1} \frac{dJ_i}{d\theta}, \quad (4)$$

so the main task before us is to compute the derivative terms $\frac{dJ_i}{d\theta}$, $i = 1, \ldots, N+1$. Fix $\theta \in \Theta$, and define $x_i := x(\theta, \tau_i)$. Taking derivatives with respect to θ in (2), we obtain,

$$\frac{dJ_i}{d\theta} \ = \ \int_{\tau_{i-1}}^{\tau_i} \Big(\frac{\partial L_i}{\partial x}(x,\theta,t)\frac{\partial x}{\partial \theta}(\theta,t) + \frac{\partial L_i}{\partial \theta}(x,\theta,t)\Big)dt + L_i(x_i,\theta,\tau_i)\frac{d\tau_i}{d\theta} - L_i(x_{i-1},\theta,\tau_{i-1})\frac{d\tau_{i-1}}{d\theta}. \quad (5)$$

The Right-Hand Side (RHS) of (5) depends on the derivative terms $\frac{\partial x}{\partial \theta}$ and $\frac{d\tau_i}{d\theta}$. The former one is basically the linearized (continuous-time) state, and the latter term will be shown to be obtained from linearizing the discrete-event dynamic equations. Taking derivatives with respect to θ in (1) we obtain the following linear equation in the interval I_i,

$$\frac{d}{dt}\frac{\partial x}{\partial \theta}(\theta,t) \ = \ \frac{\partial f_i}{\partial x}(x,\theta,t)\frac{\partial x}{\partial \theta}(\theta,t) + \frac{\partial f_i}{\partial \theta}(x,\theta,t), \quad (6)$$

with the boundary condition

$$\frac{\partial x}{\partial \theta}(\theta,\tau_{i-1}^{+}) \ = \ \frac{\partial x}{\partial \theta}(\theta,\tau_{i-1}^{-}) + \big(f_{i-1}(x_{i-1},\theta,\tau_{i-1}) - f_i(x_{i-1},\theta,\tau_{i-1})\big)\frac{d\tau_{i-1}}{d\theta} \quad (7)$$

(see [9]). The terms $\frac{d\tau_i}{d\theta}$, $i = 1, 2, \ldots$, which appear in Equations (5) and (7), have special forms that are next described.

We classify the events of the system into three categories: exogenous events, endogenous events, and induced events. These categories are defined as follows.

Definition 2.1: (i). An event is exogenous if its event-time, τ_i, satisfies the equation $\frac{d\tau_i}{d\theta} = 0$ (see [9]). (ii). An event is endogenous if there exists a continuously-differentiable function $g_i : R^n \times \Theta \rightarrow R^n$ such that the event-time τ_i is defined by the following equation,

$$\tau_i \ := \ \min\{t > \tau_{i-1} \ : \ g_i(x(\theta,t),\theta) = 0\} \quad (8)$$

(see [9]). (iii). An event occurring at time τ_j induces an event at time $\tau_i \geq \tau_j$ if the former event triggers the latter one. Moreover, there is a quantity $Q(\tau_j)$ that is transferred from time τ_j to time τ_i; this quantity is computable at time τ_j, and then made available for the required computations at time τ_i.

We denote the time-lag between τ_j and τ_i by $s(\tau_j)$ and also by $S(\tau_i)$, and hence, $s(\tau_j) = S(\tau_i) = \tau_i - \tau_j$. Both notations are used since we will refer to the time-lag at both times τ_j and τ_i. We call the event at time τ_j the inducing event, and the associated event at time τ_i, the induced event.

Definition 2.1 sheds light on the derivative terms $\frac{d\tau_i}{d\theta}$. For an exogenous event at time τ_i, $\frac{d\tau_i}{d\theta} = 0$ by definition, and for an endogenous event, (8) implies that $g_i(x_i,\theta) = 0$, and taking derivatives with respect to θ and using (1), we obtain that

$$\frac{\partial g_i}{\partial x}(x_i,\theta)\Big(\frac{\partial x}{\partial \theta}(\theta,\tau_i) + f_i(x_i,\theta,\tau_i)\frac{d\tau_i}{d\theta}\Big) + \frac{\partial g_i}{\partial \theta}(x_i,\theta) \ = \ 0. \quad (9)$$

As for induced events, the term $\frac{d\tau_i}{d\theta}$ generally is computable by the quantity $Q(\tau_j)$ that was evaluated at time τ_j; the details of these computations depend on the specific system under investigation.

In order to be useful, the IPA derivative has to be statistically unbiased, namely satisfy the relation $E(\frac{dJ}{d\theta}(\theta)) = \frac{d}{d\theta}E\big(J(\theta)\big)$ (see [5]). Unbiasedness is usually associated with continuity of the sample performance functions $J(\theta)$, and it is often easily ascertained in the fluid-queue setting [4]. It will become evident that the functions that are defined in the next section are indeed continuous, and moreover, that the simple proofs of unbiasedness that are presented in [9] can be applied to them as well. However, due to space limitations, we will not further discuss the issue of unbiasedness in this paper.

III. IPA for Fluid Queues with delayed Flow Control

To exemplify the IPA framework defined in the last section, we consider a fluid queue with flow control, where the inflow rate is adjusted by the loss rate. The feedback law is linear and the feedback signal is subjected to some delay, so that the inflow rate at time t is reduced by an amount proportional to the loss rate at time $t - S(t)$ for some (random) $S(t) > 0$.

We will investigate the IPA derivative of the loss-volume function with respect to two parameters: the buffer size, and the feedback gain. Results for the former case were developed in [2] via ad-hoc analysis that is specific to the particular problem. Here we rederive them by using the framework developed in Section II, in order to exhibit its generality. It will become evident that the principles of the analysis can be applied to the case of feedback-gain parameter, where we state the results without proofs.

A. Fluid-Queue Formulation

Consider the fluid queue shown in Figure 1, where $\sigma(t)$ and $\beta(t)$ are the external inflow rate (offered load) and server's rate, respectively, and $b > 0$ is the buffer size. Such a queue has been called a *Stochastic Flow Model (SFM)* [4]. Suppose that the processes $\{\sigma(t)\}$ and $\{\beta(t)\}$ are random and defined over a given time-interval (horizon) $[0,T]$ and over a common probability space $(\Omega,\mathcal{F},P)$, and assume that w.p.1 the functions $\sigma(t)$ and $\beta(t)$ are piecewise continuous. Let $\alpha(t)$, $x(t)$, $\gamma(t)$, and $\delta(t)$ denote, respectively, the inflow rate to the buffer, amount of fluid in the queue (buffer contents), the spillover rate due to buffer overflow, and the outflow rate from the server, all at time $t \in [0,T]$. The

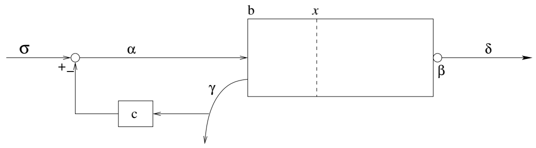

Fig. 1. Stochastic Fluid Model

process $\{\delta(t)\}$ will not interest us in this paper, and hence will not be further discussed. As for the other processes, $\{x(t)\}$ and $\{\gamma(t)\}$ are related to $\{\alpha(t)\}$ and $\{\beta(t)\}$ by the following equations:

$$\alpha(t) \;=\; \max\{\sigma(t) - c\gamma(t - S(t)), 0\} \tag{10}$$

where $c > 0$ is a given constant;

$$\dot{x} \;=\; \begin{cases} 0, & \text{if } x(t) = 0 \text{ and } \alpha(t) \leq \beta(t), \\ 0, & \text{if } x(t) = b \text{ and } \alpha(t) \geq \beta(t), \\ \alpha(t) - \beta(t), & \text{otherwise,} \end{cases} \tag{11}$$

with a given initial condition $x_0 := x(0)$; and

$$\gamma(t) \;=\; \begin{cases} \alpha(t) - \beta(t), & \text{if } x(t) = b, \\ 0, & \text{otherwise.} \end{cases} \tag{12}$$

We further make the simplifying assumption that $\alpha(t) > 0$ always in (10); this is a reasonable assumption that will simplify the analysis without detracting from its salient features, and it implies that (10) has the following form,

$$\alpha(t) \;=\; \sigma(t) - c\gamma(t - S(t)). \tag{13}$$

In the literature on IPA in the SFM setting it is common to consider (primarily) two performance measures, the loss volume defined by

$$J_L \;=\; \int_0^T \gamma(t)dt, \tag{14}$$

and the cumulative workload, defined by

$$J_Q \;=\; \int_0^T x(t)dt; \tag{15}$$

see [3] for their justification in applications. Due to space limitation we consider only the former performance measure in this paper, and we denote it by $J := J_L$. Note that $T^{-1}J$ is the average loss rate, and $(\int_0^T \sigma(t)dt)^{-1}J$ is the loss probability.

Next, suppose that some of the traffic processes depend on a control parameter θ, and hence are denoted by $\alpha(\theta, t)$, $\gamma(\theta, t)$, etc. We will consider θ to be the buffer size and the feedback gain c, and in both cases it is assumed to be constrained by a compact interval Θ with a positive left point. In either case $\sigma(t)$ and $\beta(t)$ are independent of θ while the other processes are functions of θ. Equations (11) – (14) now are valid with the dependence of their relevant terms on θ, and in particular, (14) provides the definition of the sample performance function $J(\theta)$. The sample-path derivative term $\frac{dJ}{d\theta}$ is the IPA derivative (gradient) that we seek.

Fix $\theta \in \Theta$, and consider the state trajectory developed according to (11). An *empty period* is a maximal period during which $x(\theta, t) = 0$, and a *full period* is a maximal period during which $x(\theta, t) = b$. Empty periods and full periods are labeled as *boundary periods*. A *nonboundary period* is a supremal period during which $0 < x(\theta, t) < b$, and a *nonempty period* is a supremal period during which $x(\theta, t) > 0$. Now *events* constitute the following occurrences: (i). The start of boundary periods constitute endogenous events. (ii). The end of boundary periods are events and they can be exogenous or induced. (iii). The beginning of full periods, as well as induced events during full periods, constitute inducing events. These inducing events are marked by discontinuities (jumps) in $\gamma(\theta, t)$, and these cause (induce) jumps in $\alpha(\theta, t)$ at the time of the corresponding induced event. More specifically, for an inducing event at time ρ, the quantity $Q(\theta, \rho) := \big(\gamma(\theta, \rho^-) - \gamma(\theta, \rho^+)\big)\frac{d\rho}{d\theta}$ will be shown to be computable at time ρ, and by virtue of (13), it will be transferred to time $\tau := \rho + s(\rho)$ via the relation

$$\begin{aligned} &\big(\alpha(\theta, \tau^-) - \alpha(\theta, \tau^+)\big)\frac{d\tau}{d\theta} \\ &= \; -c\big(\gamma(\theta, \rho^-) - \gamma(\theta, \rho^+)\big)\frac{d\rho}{d\theta}; \end{aligned} \tag{16}$$

this defines the essence of induced events. We remark that the delay term $s(\rho)$ is a random variable that may be a function of θ, but practically its dependence on θ can be so weak that we can assume that $\frac{ds}{d\theta} = 0$.

The following assumption, or variants thereof, are made routinely in the literature on IPA in the SFM setting; see, e.g., [3], [9] for discussion and justifications.

Assumption 3.1: For every $\theta \in \Theta$, w.p.1:

(i). The functions $\sigma(t)$ and $\beta(t)$ are piecewise continuously differentiable, the numbers of points where they are discontinuous have finite first moments, and the number of points where their first derivatives change signs have finite first moments.

(ii). For every open interval $I \subset (0, T)$, it does not happen that $\sigma(t) = \beta(t)$ for every $t \in I$.

(iii). The functions $\sigma(t)$ and $\beta(t)$ are continuous at time-points when boundary periods begin.

(iv). It does not happen that $\sigma(t) = \beta(t)$ at any point when a boundary period begins.

(v). No boundary period consists of a single point.

(vi). No induced event co-occurs with either another induced event, an endogenous event, or a jump in $\sigma(t)$ or $\beta(t)$.

Part (i) of the assumption generally guarantees statistical unbiasedness of the IPA derivative; see [3] for a detailed explanation. Parts (ii) – (vi) guarantee that all the derivatives mentioned throughout the discussion indeed exist. If one or more of these parts of the assumption is not satisfied, then the one-sided derivatives still exist, and the results derived below pertain to them instead of the derivatives.

The next two subsections will discuss the two parameters separately, namely $\theta = b$ and $\theta = c$, under Assumption 3.1. We also make the implicit assumption that initially the queue is empty, i.e., $x(\theta, 0) = 0$.

B. Buffer size as the parameter

Consider the case where $\theta = b$, and recall that $J(\theta)$ is the loss volume as a function of the buffer size. Fix $\theta \in \Theta$, and suppose that Assumption 3.1 is satisfied and hence all of the derivative terms mentioned below exist. We next apply Equations (5) – (9) to derive recursive relations among the various quantities that can be used in the computation of $\frac{dJ}{d\theta}$. First, a number of preliminary results are established.

Lemma 3.1: For every $t \in [0,T]$ such that $\alpha(\theta,t)$ is continuous at t, we have that $\frac{\partial\alpha}{\partial\theta}(\theta,t) = 0$.

Proof: If $t-S(t)$ was not contained in a full period then $\alpha(\theta,t) = \sigma(t) - \beta(t)$, and hence $\frac{\partial\alpha}{\partial\theta}(\theta,t) = 0$. If $t - S(t)$ was contained in a full period then $\alpha(\theta,t) = \sigma(t) - \big(\alpha(\theta, t - S(t)) - \beta(t - S(t))\big)$; now a recursive argument implies the desired result. ∎

Lemma 3.2: The function $\alpha(\theta,\cdot) - \beta(\cdot)$ is discontinuous only at points t when either (i) $\sigma(\cdot)$ or $\beta(\cdot)$ is discontinuous, or (ii) an induced event occurs.

Proof: Immediate by Assumption 3.1(iii) and (vi). ∎

Lemma 3.3: Let τ be the end-time of a boundary period. (i). If τ is the time of an exogenous event, then $\frac{d\tau}{d\theta} = 0$. (ii). If no exogenous or induced event occurs at time τ, then the function $\alpha(\theta,t) - \beta(t)$ is continuous at $t = \tau$, and $\alpha(\theta,t) - \beta(t) = 0$.

Proof: (i) is true by definition. (ii). The function $\alpha(\theta,t) - \beta(t)$ is continuous at $t = \tau$ (see Lemma 3.2), and since it changes signs at $t = \tau$, we have that $\alpha(\theta,\tau) - \beta(\tau) = 0$. ∎

If the end of a boundary period is neither an exogenous nor an induced event, we label it an endogenous event even though we will not make use of Equation (9). This is a semantic point that will simplify the notation while remaining consistent with in the coming discussion.

Let τ_{i-1} and τ_i be the times of an event and the next event, respectively, let $I_i := [\tau_{i-1}, \tau_i]$, and let $I_i^o := (\tau_{i-1}, \tau_i)$.

Lemma 3.4: (i). If I_i^o is a subset of an empty period then for every $t \in I_i^o$, $\frac{\partial x}{\partial\theta}(\theta,t) = 0$. (ii). If I_i^o is a subset of a full period then for every $t \in I_i^o$, $\frac{\partial x}{\partial\theta}(\theta,t) = 1$.

Proof: Immediate from the fact that $x(\theta,t) = 0$ whenever the buffer is empty, and $x(\theta,t) = \theta$ whenever the buffer is full. ∎

We next derive the linearized system via Equations (5) – (7). While keeping the notation $I_i = [\tau_{i-1}, \tau_i]$, we will denote by ϕ the starting time of boundary periods, and by ψ and ζ the end times of boundary periods.

Lemma 3.5: (i). For every interval I_i, $\frac{\partial x}{\partial\theta}(\theta,t)$ has a constant value for all $t \in I_i$. (ii). If τ_{i-1} is contained in a nonboundary period, then

$$\frac{\partial x}{\partial\theta}(\theta,\tau_{i-1}^+) = \frac{\partial x}{\partial\theta}(\theta,\tau_{i-1}^-) + \big(\alpha(\theta,\tau_{i-1}^-) - \alpha(\theta,\tau_{i-1}^+)\big)\frac{d\tau_{i-1}}{d\theta}. \quad (17)$$

(iii). Let $\tau_{i-1} := \zeta$ be the end of a boundary period. Then,

$$\frac{\partial x}{\partial\theta}(\theta,\zeta^+) = \frac{\partial x}{\partial\theta}(\theta,\zeta^-) - \big(\alpha(\theta,\zeta^+) - \beta(\zeta^+)\big)\frac{d\zeta}{d\theta}, \quad (18)$$

and unless ζ is the time of an induced event, $\big(\alpha(\theta,\zeta^+) - \beta(\zeta^+)\big)\frac{d\zeta}{d\theta} = 0$ and hence $\frac{\partial x}{\partial\theta}(\theta,\zeta^+) = \frac{\partial x}{\partial\theta}(\theta,\zeta^-)$. (iv). Let $\tau_i := \phi$ be the starting time of a full period. Then the function $\alpha(\theta,t) - \beta(t)$ is continuous at $t = \phi$, and

$$\big(\alpha(\theta,\phi) - \beta(\phi)\big)\frac{d\phi}{d\theta} = 1 - \frac{\partial x}{\partial\theta}(\theta,\phi^-). \quad (19)$$

Proof: (i). Either $f_i(x,\theta,t) = 0$ or $f_i(x,\theta,t) = \alpha(\theta,t) - \beta(t)$, and hence, and by Lemma 3.1, $\frac{\partial f_i}{\partial x}(x,\theta,t) = \frac{\partial f_i}{\partial\theta}(x,\theta,t) = 0$ for all $t \in I_i^o$. Equation (6) now implies the desired result.

(ii). $f_{i-1}(x,\theta,t) = \alpha(\theta,t) - \beta(t)$ and $f_i(x,\theta,t) = \alpha(\theta,t) - \beta(t)$. By Assumption 3.1(vi), $\beta(t)$ is continuous at $t = \tau_i$. (7) now implies (17).

(iii). $f_{i-1}(x,\theta,t) = 0$ and $f_i(x,\theta,t) = \alpha(\theta,t) - \beta(t)$, and hence (7) implies (18). The second statement follows from (18) and Lemma 3.3.

(iv). By Assumption 3.1(iii), and (vi), the function $\alpha(\theta,t) - \beta(t)$ is continuous at $t = \phi$. Note that, in (8), $g_i(x,\theta) = x - \theta$, and hence, $\frac{\partial g_i}{\partial x} = -\frac{\partial g_i}{\partial\theta} = 1$. Moreover, $f_i(x,\theta,t) = \alpha(\theta,t) - \beta(t)$. Plug all of this in (9), to obtain (19). ∎

Let $F := [\phi,\psi]$ denote a generic full period, and define $\Lambda(\theta) := \int_F \gamma(\theta,t)dt$. Obviously $\frac{dJ}{d\theta}$ is the sum of the terms $\frac{d\Lambda}{d\theta}$ corresponding to the various full periods, hence we will focus on the IPA term $\frac{d\Lambda}{d\theta}$ in the sequel. Let $P := (\zeta,\phi)$ denote the nonboundary period preceding F, and define π by $\pi = 0$ if the boundary period ending at time $t = \zeta$ was full, and $\pi = 1$ if the boundary period ending at time ζ was empty. Furthermore, let $\zeta = \tau_{k_1}$, $\phi = \tau_{k_2}$, and $\psi = \tau_{k_3}$, for some $k_1 > 0$, $k_2 > k_1$, and $k_3 > k_2$; let τ_i, $i = k_1 + 1, \ldots, k_2 - 1$, be the times of induced events in P, and let τ_i, $i = k_2 + 1, \ldots, k_3 - 1$ be the times of induced events in F^o (the interior of F).

Proposition 3.1: (i).

$$\frac{d\Lambda}{d\theta} = \big(\alpha(\theta,\psi^-) - \beta(\psi^-)\big)\frac{d\psi}{d\theta} + \sum_{i=k_2+1}^{k_3-1}\big(\alpha(\theta,\tau_i^-) - \alpha(\theta,\tau_i^+)\big)\frac{d\tau_i}{d\theta} - \big(\alpha(\theta,\phi) - \beta(\phi)\big)\frac{d\phi}{d\theta}, \quad (20)$$

where

$$\big(\alpha(\theta,\phi) - \beta(\phi)\big)\frac{d\phi}{d\theta} = \pi - \sum_{i=k_1+1}^{k_2-1}\big(\alpha(\theta,\tau_i^-) - \alpha(\theta,\tau_i^+)\big)\frac{d\tau_i}{d\theta} + \big(\alpha(\theta,\zeta^+) - \beta(\zeta^+)\big)\frac{d\zeta}{d\theta}. \quad (21)$$

(ii). Unless ψ is the time of an induced event, $\big(\alpha(\theta,\psi^-) - \beta(\psi^-)\big)\frac{d\psi}{d\theta} = 0$; and unless ζ is the time of an induced event, $\big(\alpha(\theta,\zeta^+) - \beta(\zeta^+)\big)\frac{d\zeta}{d\theta} = 0$.

Proof: (i). Note that $F = \cup_{i=k_2+1}^{k_3} I_i$, and for all $i = k_2+1, \ldots, k_3$, $L_i(x,\theta,t) = \alpha(\theta,t) - \beta(t)$ (see (2) and (12)). Therefore, and by Lemma 3.1, the integral term in (5) is zero, and (5) yields (20).

Next, by Lemma 3.5(i), $\frac{\partial x}{\partial \theta}(\theta,\tau_i^-) = \frac{\partial x}{\partial \theta}(\theta,\tau_{i-1}^+)$ for all i. Applying (19), then repeatedly (17) with $i = k_2,\ldots,k_1+2$, followed by (18), and noting (Lemma 3.4) that $1 - \frac{\partial x}{\partial \theta}(\theta,\zeta^-) = \pi$, Equation (21) follows.

(ii). Follows immediately from Lemma 3.3. ∎

Proposition 3.1 has a recursive structure that leads to a computation of the IPA derivative, as can be seen from the following result.

Proposition 3.2: For every ρ being the time of an inducing event, the quantity $\big(\gamma(\theta,\rho^-) - \gamma(\theta,\rho^+)\big)\frac{d\rho}{d\theta}$ is computable at time $t = \rho$ without having to compute the term $\frac{d\rho}{d\theta}$.

Proof: The proof is by induction. The first inducing event is the starting time of the first full period, and denoting this point by ϕ, (21) implies that $\big(\gamma(\theta,\phi^-) - \gamma(\theta,\phi^+)\big)\frac{d\phi}{d\theta} = 1$.

Next, let ρ be the time of an inducing event, and suppose that, for any past inducing event occurring at any time $\rho_1 < \rho$, the quantity $\big(\gamma(\theta,\rho_1^-) - \gamma(\theta,\rho_1^+)\big)\frac{d\rho_1}{d\theta}$ was computable at time ρ_1. Consequently, and by (13), for every induced event occurring at any time $\tau \le \rho$ (including the time $\tau = \rho$), the quantity $\big(\alpha(\theta,\tau^-) - \alpha(\theta,\tau^+)\big)\frac{d\tau}{d\theta}$ was computable at time τ.

There are three possibilities regarding the inducing event at time ρ: (1). It is contained in the interior of a full period. (2). It is the start of a full period. (3). It is an induced event that ends a full period. In case (1), by (12), $\big(\gamma(\theta,\rho^-) - \gamma(\theta,\rho^+)\big)\frac{d\rho}{d\theta} = \big(\alpha(\theta,\rho^-) - \alpha(\theta,\rho^+)\big)\frac{d\rho}{d\theta}$, and by the inductions hypothesis, the latter quantity is computable at time $t = \rho$. In case (2), let us define $\phi = \rho$ to comply with the notation of (21). Now $\gamma(\theta,\rho^-) - \gamma(\theta,\rho^+) = -\big(\alpha(\theta,\phi) - \beta(\phi)\big)$, and Equation (21) is in force. In this equation, π is certainly computable from the past; the sum-terms are computable at the (past) times τ_i by the induction's hypothesis, and the last term is equal to

$$\begin{aligned}&\big(\alpha(\theta,\zeta^+) - \beta(\zeta^+)\big)\frac{d\zeta}{d\theta}\\ &= \frac{\alpha(\theta,\zeta^+) - \beta(\zeta^+)}{\alpha(\theta,\zeta^-) - \beta(\zeta^-)} \times \big(\alpha(\theta,\zeta^-) - \beta(\zeta^-)\big)\frac{d\zeta}{d\theta}. \qquad (22)\end{aligned}$$

There may be no escape from computing the fraction term, which involves explicit rates (these can be approximated via moving averages - see [9]), but the last term in (22) is computable at time ζ by the inductions hypothesis. Finally, in case (3), we use the notation $\psi = \rho$. Since ψ is the end of a full period, we have that $\gamma(\theta,\psi^-) - \gamma(\theta,\psi^+) = \alpha(\theta,\psi^-) - \beta(\psi^-)$, and hence,

$$\begin{aligned}&\big(\gamma(\theta,\psi^-) - \gamma(\theta,\psi^+)\big)\frac{d\psi}{d\theta}\\ &= \frac{\alpha(\theta,\psi^-) - \beta(\psi^-)}{\alpha(\theta,\psi^-) - \alpha(\theta,\psi^+)} \times \big(\alpha(\theta,\psi^-) - \alpha(\theta,\psi^+)\big)\frac{d\psi}{d\theta}. \quad (23)\end{aligned}$$

Again, there appears to be no way to get around the computation of the fraction term in (23), which requires explicit flow rates. However, the last term in (23) is computable at time $\psi = \rho$ by the induction's hypothesis, and this completes the proof. ∎

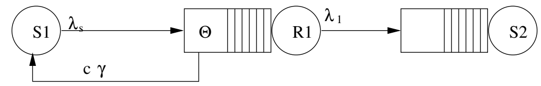

Fig. 2. Simulation Topology

C. Feedback gain as the parameter

Consider the case where $\theta = c$, the feedback gain, and the buffer size is a given constant $b > 0$. As a result, Equation (13) assumes the form $\alpha(\theta,t) = \sigma(t) - \theta\gamma(\theta, t - S(t))$, and hence,

$$\frac{\partial \alpha}{\partial \theta}(\theta,t) = -\gamma(\theta,t-S(t)) - \theta\frac{\partial \alpha}{\partial \theta}(\theta,t-S(t)) \qquad (24)$$

if $t - S(t)$ was contained in the interior of a full period, and $\frac{\partial \alpha}{\partial \theta}(\theta,t) = 0$ if $t = S(t)$ was not contained in a full period. The main result concerning the IPA derivative, Proposition 3.3, is similar to Proposition 3.1 in the last subsection, and hence it will be stated without a proof.

Proposition 3.3: (i).

$$\begin{aligned}\frac{d\Lambda}{d\theta} = &\int_\phi^\psi \frac{\partial \alpha}{\partial \theta}(\theta,t)dt + \big(\alpha(\theta,\psi^-) - \beta(\psi^-)\big)\frac{d\psi}{d\theta}\\ &+ \sum_{i=k_2+1}^{k_3-1}\big(\alpha(\theta,\tau_i^-) - \alpha(\theta,\tau_i^+)\big)\frac{d\tau_i}{d\theta}\\ &-\big(\alpha(\theta,\phi) - \beta(\phi)\big)\frac{d\phi}{d\theta}, \qquad (25)\end{aligned}$$

where

$$\begin{aligned}\big(\alpha(\theta,\phi) - \beta(\phi)\big)\frac{d\phi}{d\theta} = &\int_\zeta^\phi \frac{\partial \alpha}{\partial \theta}(\theta,t)dt\\ &- \sum_{i=k_1+1}^{k_2-1}\big(\alpha(\theta,\tau_i^-) - \alpha(\theta,\tau_i^+)\big)\frac{d\tau_i}{d\theta}\\ &+\big(\alpha(\theta,\zeta^+) - \beta(\zeta^+)\big)\frac{d\zeta}{d\theta}. \qquad (26)\end{aligned}$$

(ii). Unless ψ is the time of an induced event, $\big(\alpha(\theta,\psi^-) - \beta(\psi^-)\big)\frac{d\psi}{d\theta} = 0$; and unless ζ is the time of an induced event, $\big(\alpha(\theta,\zeta^+) - \beta(\zeta^+)\big)\frac{d\zeta}{d\theta} = 0$.

We remark that Proposition 3.2 provides the recursive structure for computing the various terms in the RHS of (25) AND (26).

IV. Simulation Results

The IPA derivatives were analyzed in the fluid-queue setting, but their eventual implementation may be on traffic processes observed from packet-based networks. For these reason we conducted simulation experiments on a packet multiplexor, where we computed the IPA derivative of the loss volume as function of the buffer size. The experiments were performed using the *Georgia Tech Network Simulator* (*GTNetS*) [6], a general-purpose packet-network simulator, modified to include a loss feedback loop with delays.

The simulation setup is shown in Figure 2. Bits are generated at the source $S1$ according to an on-off process, where both "on" and "off" periods are exponentially distributed

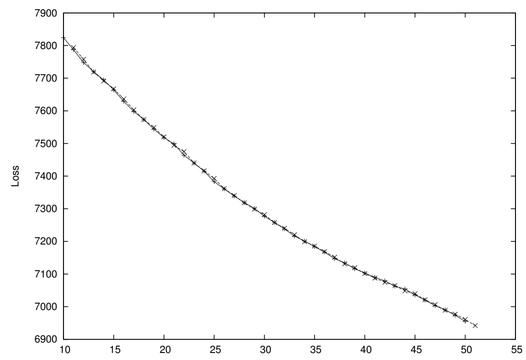

Fig. 3. Loss Volume and Predicted Loss Volume

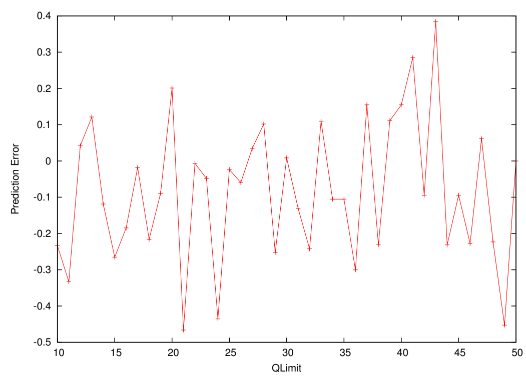

Fig. 4. Loss Prediction Error

with mean 100 milliseconds each; the source generates the bits at the rate of 8.281 Megabits per second during the "on" periods. The bits are assembled at the source into UDP packets, which are encapsulated into IP datagrams and Ethernet packets, so that, counting the overhead associated with the encapsulation, each packet consists of 556 bytes. These packets are then transmitted to the router $R1$ at the rate of $\lambda_s := 10$ Megabits per second. Unless they are dropped by the router, the packets are transmitted from there to the sink $S2$ at the rate of $\lambda_1 := 5$ Megabits per second, so that the traffic intensity at the queue is 0.9. The variable parameter θ is the buffer size in units of packets, and the feedback gain is $c = 0.3$. Each time a packet is dropped, a feedback signal is sent back to the source, and the feedback law is implemented by discarding the next-generated packet at the source with a probability of 0.3. The latency of the feedback signal was (somewhat arbitrarily) set to $S(\tau) = 10$ milliseconds. The horizon interval over which we computed the sample performance function $J(\theta)$ and its IPA derivative $\frac{dJ}{d\theta}$ is $[0, T]$, with $T = 10$ seconds.

We performed 41 individual simulations, with buffer sizes at the router $R1$ ranging from 10 to 50 inclusive, all with the same seed. For each run, the actual loss volume (in packets) was calculated and reported, as well as the IPA derivative $\frac{dJ}{d\theta}$. We then used it to compute a predicted loss volume at the next value of θ via a first-order linearization with $\Delta\theta = 1$, namely the predicted value of $J(\theta + 1)$ is $J(\theta) + \frac{dJ}{d\theta}(\theta)$, where the latter IPA derivative was computed via Equations (20) and (212). As a matter of fact, in these computations we neglected the problematic terms, namely the first term in the RHS of (20) and the last term in the RHS of (21). These terms generally arise infrequently, and neglecting them reduced the complexity of the simulation program while yielding good results.

The simulation results are shown in Figure 3, where we plotted the value of the computed loss volumes (the continuous curve) as well as their predicted values (indicated by stars); it is hard to discern any difference. Figure 4 shows a plot of the relative prediction error, which is usually under 0.3, always under 0.5, and whose average is -0.0856 . A provably-convergent gradient-descent algorithm is guaranteed to converge under such error bounds.

V. Conclusions

This paper explores IPA for the loss volume in fluid queues with flow control where, in contrast to earlier results, the feedback signals incur delays. It first proposes a general framework for modeling and analysis of IPA in a fluid-queue setting, and then applies it to specific examples. It appears that the framework can be generalized to networks of fluid queues, and this provides an avenue for future research.

References

[1] R. Adams, Y. Wardi, and G. Riley, "Sensitivity Analysis of Loss Volume in Stochastic Fluid Systems with Flow Control", in *Proc. 46th CDC*, New Orleans, Louisiana, December 12-14, 2007.

[2] R. Adams and Y. Wardi, "IPA For Flow-Control Fluid Queues with Delays", *Proc. 9th Intl. Workshop on Discrete Event Dynamic Systems, WODES'08*, Gothenburg, Sweden, May 28-30, 2008.

[3] C.G. Cassandras, Y. Wardi, B. Melamed, G. Sun, and C.G. Panayiotou, "Perturbation Analysis for On-Line Control and Optimization of Stochastic Fluid Models", *IEEE Transactions on Automatic Control*, Vol. AC-47, No. 8, pp. 1234-1248, 2002.

[4] C.G. Cassandras, "Stochastic Flow Systems: Modeling and Sensitivity Analysis", in *Stochastic Hybrid Systems: Recent Developments and Research Trends*, Eds. C.G. C.assandras and J. Lygeros, CRC Press, New York, New York, pp. 139-167, 2006.

[5] Y.C. Ho and X.R. Cao, *Perturbation Analysis of Discrete Event Dynamic Systems*, Kluwer Academic Publishers, Boston, Massachusetts, 1991.

[6] G.F. Riley, "The Georgia Tech Network Simulator", in *Proc. ACM SIGCOMM workshop on Models, methods and tools for reproducible network research*, pp. 5–12, 2003.

[7] G. Sun, C.G. Cassandras, Y. Wardi, C.G. Panayiotou, and G. Riley, "Perturbation Analysis and Optimization of Stochastic Flow Networks", *IEEE Transactions on Automatic Control*, Vol. 49, No. 12, pp. 2143-2159, 2004.

[8] Y. Wardi, G. Riley, and R. Adams, "Perturbation Analysis for Stochastic Flow Systems with Feedback", in *Stochastic Hybrid Systems: Recent Developments and Research Trends*, Eds. C.G. Cassandras and J. Lygeros, CRC Press, New York, New York, pp. 167-188, 2006.

[9] Y. Wardi and R. Adams, "A Unified Approach to Infinitesimal Perturbation Analysis in Stochastic Flow Models: The Single-Stage Case", Manuscript, School of Electrical and Computer Engineering, Georgia Institute of Technology, November 2007; under review by the *IEEE Transactions on Automatic Control*.

[10] H. Yu and C.G. Cassandras, "Perturbation Analysis of Feedback-Controlled Stochastic Flow Systems", *IEEE Transactions on Automatic Control*, Vol. 49, No. 8, pp. 1317-1332, 2004.

[11] H. Yu and C.G. Cassandras, "A New Paradigm for an On-Line Management of Communication Networks with Multiplicative Feedback Control", in *Performance Evaluation and Planning Methods for the Next Generation Internet*, Eds. A. Girard, B. Sanso', and F. Vasquez-Abad, pp. 297-332, Springer-Verlag, 2005.