

Analysis of an M/G/1 queue with customer impatience and adaptive arrival process

Onno Boxma, Balakrishna Prabhu

▶ **To cite this version:**

Onno Boxma, Balakrishna Prabhu. Analysis of an M/G/1 queue with customer impatience and adaptive arrival process. NetGCOOP 2011 : International conference on NETwork Games, COntrol and OPtimization, Telecom SudParis et Université Paris Descartes, Oct 2011, Paris, France. hal-00644866

HAL Id: hal-00644866

https://inria.hal.science/hal-00644866

Submitted on 25 Nov 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Analysis of an $M/G/1$ queue with customer impatience and adaptive arrival process

O.J. Boxma[1] and B.J. Prabhu[2,3]
[1] EURANDOM and Department of Mathematics and Computer Science
Eindhoven University of Technology; 5600 MB Eindhoven, The Netherlands.
boxma@win.tue.nl
[2] CNRS ; LAAS ; 7 avenue du colonel Roche; F-31077 Toulouse Cedex 4, France
[3] Université de Toulouse ; UPS, INSA, INP, ISAE ; UT1, UTM, LAAS; F-31077 Toulouse Cedex 4, France
bjprabhu@laas.fr

***Abstract*— We study an $M/G/1$ queue with impatience and an adaptive arrival process. The rate of the arrival process changes according to whether an incoming customer is accepted or rejected. We analyse two different models for impatience : *(i)* based on workload, and *(ii)* based on queue-length. For the workload-based model, we obtain the Laplace-Stieltjes Transform of the joint stationary workload and arrival rate process, and that of the waiting time. For the queue-length based model we obtain the analogous z-transform. These queueing models might also be useful for capturing the interaction between congestion control algorithms and queue management schemes in the Internet.**

I. Introduction

Data traffic in the Internet is regulated by distributed algorithms which require each flow to adapt its sending rate according to the level of congestion in the network. This information is conveyed to a flow by the links in the network using a binary feedback instructing the flow to either increase or decrease its sending rate. The link generates these feedback signals as a function of the occupancy of its input buffer: the higher the occupancy, the higher is the level of congestion, leading to a larger number of decrease signals. A frequently employed binary feedback signal is packet admission/rejection : admission signals an increase and rejection signals a decrease. An easy to implement example of a packet admission control policy is to reject an incoming packet if the buffer is full and to accept it otherwise. However, this policy leads to synchronization among concurrent flows and hence to inefficient utilization of the link bandwidth, *cf.* [1]. To overcome such effects, various alternative policies, including probabilistic admission control, have been proposed, *cf.* [1] and [2].

The aim of this paper is to model and analyze the above described interaction between a data source and a link that has a probabilistic packet admission control policy. The packet arrival process is modelled as a time inhomogeneous Poisson process whose intensity varies depending on whether an incoming packet is accepted or rejected. The packet sizes are assumed to be independent and identically distributed, and the link is modelled as an infinite buffer served at a fixed rate. An incoming packet is accepted to the queue with a probability that is a function of the current workload in the queue. We obtain the Laplace-Stieltjes Transform (LST) of the joint stationary workload and arrival intensity process, which then leads to the LST of the waiting time of accepted arrivals. In addition, we also study the model in which incoming packets are accepted depending upon the number of packets in the system. For this model we give the z-transform of the joint stationary distribution of the queue length and the arrival intensity process, which then leads to the z-transform of the number in the system seen by accepted arrivals.

A. Related work

Performance analysis of congestion control and admission control policies has been a subject of several studies, though most of them have neglected the interaction between the two (*cf.* [3] for a detailed discussion). Due to its complexity, the analytical study of this interaction has been more or less restricted to studying the dynamics of the expected values using deterministic differential equations (with or without feedback delay), *cf.* [4] and [5]. The model considered in the present paper attempts to capture this interaction within a stochastic framework for a certain class of probabilistic admission control policies.

A related work in which this interaction has been studied is that of [6] in which the authors model a TCP (for Transmission Control Protocol, the predominant congestion control protocol in the Internet) source as a fluid source whose rate varies depending upon whether the finite buffer is full or not. As opposed to their fluid model, we model the TCP source as one which emits packets at distinct epochs thereby causing jumps in the buffer content process which no longer has a continuous sample path. Another important difference is in the feedback model itself. In [6] there is a positive feedback if and only if the buffer is not full. It is thus a model for a Drop-Tail policy, whereas we study a probabilistic feedback policy - the feedback is positive with a probability that decreases as the buffer level increases.

The present model is also strongly related to some of the existing models on impatience investigated in queueing theory, and in the following we describe this connection.

B. Connection with queueing theory

The connection between queueing theory and performance analysis of congestion control algorithms has been known since long, *cf.* [7]. In the context of queueing theory, several rejection rules have been studied for the $M/G/1$ queue, see, e.g., [8], but the arrival rate typically does not change with the decision to accept or reject a customer.

The model under investigation in this paper can be seen as a generalization of the $MAP/G/1$ queue with impatient customers which was studied by Combé in [9]. In that model, the arrival process changes states at each packet arrival instant. However, the dynamics of the arrival process do not depend on whether a packet is accepted or rejected, which makes our model a generalization of the one studied in [9]. Our method of analysis is similar to the one in [9] in that we obtain a system of Volterra integral equations of the second kind for the joint stationary distribution of the workload and arrival rate process, from which we obtain a system of recursive equations for the LST of the joint stationary process.

A part of the present work containing the LST of the joint stationary workload and arrival rate process appeared in [10].

C. Organization of the paper

The rest of the paper is organized as follows. In Section II, we describe the system model and state the assumptions. In Section III, we present the analysis of the model with workload-based impatience leading to the computation of the LST of the joint stationary workload and arrival rate process. Based upon this LST, we give the LST of the waiting time of the accepted arrivals. In Section IV, we analyse the queue-length-based model and obtain the z-transform of joint stationary queue-length and arrival rate process. Based upon this z-transform, we give the z-transform of the number of customers seen in the system by accepted arrivals.

Most of the results in this paper are presented without proof. The proofs are provided in the research report [3], which also contains a numerical example.

II. Model description

Consider a variable data rate source which generates packets at Poisson intensity $\lambda(t) \in \mathcal{L}$, where $\mathcal{L}$ is a finite set of cardinality N. The packet sizes are assumed to be i.i.d. with distribution function $B_i(\cdot)$, mean μ_i^{-1}, and LST $\mathsf{B_i}(\cdot)$, when the data source is in state i. These packets arrive at a queue, say a router in the Internet, which admits the packets based on the following admission control policy. An incoming packet which sees a workload level of x is admitted to the queue with probability $\exp(-\nu x)$, and rejected otherwise. We do not model the possibility of a rejected packet re-entering the queue at a later instant. The function $\exp(-\nu x)$ can also be thought of as an impatience function associated with customers arriving to a server. If an incoming customer sees a higher waiting time, then it is less likely to join the queue.

For a router in the Internet the buffer occupancy in bits (the workload) is generally known to the router, and hence a workload-based impatience model is better suited for the analysis of such systems. However, in some cases the impatience probability may depend on the number of customers in the system rather than the total workload which may be unknown. For such a queue-length-based model, we shall assume that if an incoming packet sees n packets in the system, then it is admitted with probability p^n, $0 < p < 1$.

We shall assume that the variable data rate source is informed immediately whether a packet was admitted or rejected. In practice, there is a delay after which the source receives this information. The source reacts to the admission control policy by adapting its data rate in the following way. With state i of the source we associate a Poisson intensity λ_i. The state of the source jumps from i to j with probability p_{ij} if a packet is rejected, and with probability p^*_{ij} if a packet is accepted. Thus, the intensity of the arrival process potentially changes with each arrival to the queue. In a protocol like TCP, the state of the source will jump to a state $j \leq i$ if a packet is rejected and to a state $j \geq i$ if a packet is accepted. However, we shall not assume any particular structure for the matrices $\mathbf{P} = [p_{ij}]$ and $\mathbf{P}^* = [p^*_{ij}]$.

III. Impatience based on workload

Let $V_i(t,x)$ denote the joint probability that at time t the workload is less than or equal to x and the input process is in state i. The server is assumed to work at unit rate. There are three events that can happen in a small interval $[t, t+\delta t)$: (i) there are no arrivals, in which case the workload is drained by an amount δt; (ii) an arrival occurs and is rejected, in which case the input process changes state; and (iii) an arrival occurs and is accepted, in which case the input process changes state and there is a jump in the workload process. The three terms on the RHS in the following equation correspond to the above three possible events.

$$\begin{aligned} V_i(t+\delta t, x) = {} & (1-\lambda_i \delta t) V_i(t, x+\delta t) \\ & + \sum_j p_{ji} \lambda_j \delta t \int_{0^-}^{x} (1-\exp(-\nu y)) dV_j(t,y) \\ & + \sum_j p^*_{ji} \lambda_j \delta t \int_{0^-}^{x} B_j(x-y) \exp(-\nu y) dV_j(t,y), \end{aligned}$$

for $x > 0,\ 1 \leq i \leq N$.

From the above dynamics, we can derive the following integral equation for $i = 1, \ldots, N$:

$$\begin{aligned} \frac{\partial V_i(t,x)}{\partial t} = {} & \frac{\partial V_i(t,x)}{\partial x} - \lambda_i\ V_i(t,x) \\ & + \sum_j p_{ji} \lambda_j \int_{0^-}^{x} (1-\exp(-\nu y)) dV_j(t,y) \\ & + \sum_j p^*_{ji} \lambda_j \int_{0^-}^{x} B_j(x-y) \exp(-\nu y) dV_j(t,y). \end{aligned} \tag{1}$$

Let us now discuss the issue whether the joint steady-state distribution of workload and arrival rate process exists.

Proposition 1 (Stability)**.** *If*

1) $\rho_{max} := \sup_i \lambda_i \mu_i^{-1}$ *is finite, and*
2) $\lim_{x\to\infty} f(x) = 0$*,*

then the joint workload and arrival rate process is stable.

In the sequel we assume that the two conditions of the above proposition hold.

Let $\Phi_i(s) = \int_{0^-}^{\infty} \exp(-sx) dV_i(x)$ denote the LST of the joint distribution function, and let $\overline{\Phi}(s) := [\Phi_i(s)]$ denote the row vector of the LST of the joint stationary distribution. Also, let $V_i(0)$ be the stationary joint probability that the workload is zero and the arrival intensity is λ_i, and $\overline{V}(0) := [V_i(0)]$ be the row vector of these joint probabilities. The following result relates $\overline{\Phi}(s)$ to $\overline{V}(0)$.

Lemma 1. *The LST of the joint distribution function* $\overline{\Phi}(s)$ *is given by the following infinite sum:*

$$\overline{\Phi}(s) = \overline{V}(0) \left[\sum_{i=0}^{\infty} \mathbf{D}(s+i\nu)\mathbf{A}^{-1}(s+i\nu) \left[\prod_{j=0}^{i-1} \mathbf{C}(s+j\nu)\mathbf{A}^{-1}(s+j\nu) \right] \right], \quad (2)$$

where the empty product is assumed to be unity, and

$$\mathbf{A}(s) = s\mathbf{I} - \mathbf{\Lambda}(\mathbf{I} - \mathbf{P}), \quad \mathbf{D}(s) = s\mathbf{I}, \quad \mathbf{C}(s) = \mathbf{\Lambda}(\mathbf{P} - \mathbf{B}(s)\mathbf{P}^*),$$

$\mathbf{\Lambda}$ *is a diagonal matrix with* λ_i *as its* i*th diagonal entry,* $\mathbf{B}(s)$ *is a diagonal matrix with* $\mathsf{B}_i(s)$ *as its* i*th diagonal entry.*

Remark 1. *Instead of rejecting packets based on workload upon arrival, the router could reject packets based on the sum of the workload upon arrival and the service time of the arriving packet. For this rejection policy, the LST of the joint distribution function turns out to have the same form as* (2) *where* $\mathbf{A}(s) = s\mathbf{I} - \mathbf{\Lambda}(\mathbf{I} - \mathbf{P})$*,*

$$\mathbf{D}(s) = s\mathbf{I}, \quad \mathbf{C}(s) = \Lambda(\mathbf{B}(\nu)\mathbf{P} - \mathbf{B}(s+\nu)\mathbf{P}^*).$$

The only difference with respect to rejection based on workload upon arrival is in the matrix $\mathbf{C}$*.*

We next proceed to determine the constants $V_i(0), i = 1, 2, ..., N$, which will then completely characterize $\overline{\Phi}(s)$.

Let γ_i, $i = 1, 2, ..., N$, denote the ith eigenvalue of $\mathbf{\Lambda}(\mathbf{I} - \mathbf{P})$, such that $\gamma_i \leq \gamma_j$ for $i < j$, and $\underline{\boldsymbol{\alpha}}_i$ denote the corresponding right eigenvector. Since $\mathbf{P}$ is a stochastic matrix, the first eigenvector, $\underline{\boldsymbol{\alpha}}_1$, equals $[1\ 1 \ldots 1]^T$ with eigenvalue $\gamma_1 = 0$. For the location of the other $N-1$ eigenvalues of $\mathbf{\Lambda}(\mathbf{I} - \mathbf{P})$, we have the following result.

Lemma 2. *The eigenvalues of the matrix* $\mathbf{\Lambda}(\mathbf{I} - \mathbf{P})$ *have positive real parts.*

Since $\mathbf{A}(s) = s\mathbf{I} - \mathbf{\Lambda}(\mathbf{I} - \mathbf{P})$, $\mathbf{A}(s)$ is singular at the eigenvalues of $\mathbf{\Lambda}(\mathbf{I} - \mathbf{P})$, i.e., $\det(\mathbf{A}(s)) = 0$ at $s = \gamma_i, i = 1, 2, ..., N$. However, $\overline{\Phi}(s)$ is analytic in the half-plane $\mathrm{Re}(s) \geq 0$, and hence the constants $V_i(0), i = 1, 2, ..., N$, are such that the RHS of (2) is finite at $s = \gamma_i, i = 1, 2, ..., N$.

In order to compute $\overline{V}(0)$ we shall make use of the above fact and the following representation

$$\overline{\Phi}(s)\mathbf{A}(s) = \overline{V}(0)\mathbf{M}(s), \quad (3)$$

where

$$\mathbf{M}(s) = \left[\sum_{i=0}^{\infty} \mathbf{D}(s+i\nu) \left[\prod_{j=0}^{i-1} \mathbf{A}^{-1}(s+(j+1)\nu)\mathbf{C}(s+j\nu) \right] \right].$$

Before stating the main result, we first make an assumption under which the result holds.

Assumption 1. *For* $j \geq 1$*,* $\mathbf{A}(s+j\nu)$ *is invertible at* $s = \gamma_i$*, which is equivalent to the condition that* $\gamma_i \neq \gamma_k + j\nu$ *for* $i \neq k$ *and for every* j*, i.e., no two eigenvalues differ by an integer multiple of* ν*.*

The above assumption ensures that $\mathbf{A}(s+j\nu)$ is invertible in the right-half plane for $j \geq 1$. We shall later observe using numerical computations that when two eigenvalues differ by an integer multiple of ν, we can obtain the constants $V_i(0)$ by perturbing the entries of the matrix $\mathbf{\Lambda}(\mathbf{I} - \mathbf{P})$.

Theorem 1. *The joint probability vector* $\overline{V}(0)$ *is the unique solution of the following set of* N *linear equations:*

$$\overline{V}(0) = \begin{bmatrix} (\mathbf{I} + \mathbf{M}(\nu)\mathbf{A}^{-1}(\nu)\mathbf{\Lambda}\boldsymbol{\mu}^{-1})\underline{\boldsymbol{\alpha}}_1 \\ \mathbf{M}(\gamma_2)\underline{\boldsymbol{\alpha}}_2 \\ \vdots \\ \mathbf{M}(\gamma_N)\underline{\boldsymbol{\alpha}}_N \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \quad (4)$$

Proof: Since $\mathbf{A}(s) = s\mathbf{I} - \mathbf{\Lambda}(\mathbf{I} - \mathbf{P})$, every right eigenvector of $\mathbf{\Lambda}(\mathbf{I} - \mathbf{P})$, $\underline{\boldsymbol{\alpha}}_i$, is also a right eigenvector of $\mathbf{A}(s)$ with eigenvalue $(s - \gamma_i)$. For $i = 2, 3, ..., N$, we right multiply (3) by $\underline{\boldsymbol{\alpha}}_i$ and set $s = \gamma_i$ to get the following $N-1$ equations

$$0 = \overline{V}(0)\mathbf{M}(\gamma_i)\underline{\boldsymbol{\alpha}}_i, \ i = 2, 3, ..., N. \quad (5)$$

For the final equation, we first note that $\mathbf{M}(s)$ is singular at $s = 0$. To see this, rewrite $\mathbf{M}(s)$ as

$$\mathbf{M}(s) = \mathbf{D}(s) + \mathbf{M}(s+\nu)\mathbf{A}^{-1}(s+\nu)\mathbf{C}(s), \quad (6)$$

and right multiply by $\underline{\boldsymbol{\alpha}}_1$. Using (1), we see that $\mathbf{M}(s)\underline{\boldsymbol{\alpha}}_1 = s\underline{\boldsymbol{\alpha}}_1 + \mathbf{M}(s+\nu)\mathbf{A}^{-1}(s+\nu)\mathbf{\Lambda}(\mathbf{I} - \mathbf{B}(s))\underline{\boldsymbol{\alpha}}_1$ is equal to zero at $s = 0$, and that

$$\lim_{s\to 0} \frac{\mathbf{M}(s)\underline{\boldsymbol{\alpha}}_1}{s} = (\mathbf{I} + \mathbf{M}(\nu)\mathbf{A}^{-1}(\nu)\mathbf{\Lambda}\boldsymbol{\mu}^{-1})\underline{\boldsymbol{\alpha}}_1, \quad (7)$$

where $\boldsymbol{\mu}$ is a diagonal matrix with μ_i as its ith diagonal entry. We right multiply (3) by $\underline{\boldsymbol{\alpha}}_1$ and use the normalization equation $\overline{\mathsf{V}}(0)\underline{\boldsymbol{\alpha}}_1 = 1$ to obtain

$$1 = \overline{V}(0)(\mathbf{I} + \mathbf{M}(\nu)\mathbf{A}^{-1}(\nu)\mathbf{\Lambda}\boldsymbol{\mu}^{-1})\overline{\boldsymbol{\alpha}}_1. \quad (8)$$

Combining (8) and (5), we obtain the system of equations (4). □

Let W be the waiting time of the accepted customers in steady state. The following result relates the LST of W to Φ_i, which is the LST of the the joint steady-state probability that

the workload is less than or equal to x and the input process is in state i. Combé [9] derived a similar result for the special case when the dynamics of the input process do not depend on whether a packet is accepted or rejected.

Proposition 2.

$$E[\exp(-sW)] = \frac{\sum_i \lambda_i \Phi_i(s+\nu)}{\sum_i \lambda_i \Phi_i(\nu)}.$$

IV. Impatience based on number of customers

In this section we shall consider a discrete state-space model to study the joint behaviour of a variable data rate source and the queue-length at the buffer. As in the previous section, we assume that the variable rate source generates packets according to a Poisson process of rate λ_i when it is in state i, $i = 1, \ldots, N$. The packets arrive at a single server queue which admits the packets based on the following admission control policy. If an incoming packet sees n packets in the system, then it is admitted to the queue with probability p^n, with $0 < p < 1$. The background state now changes to j w.p. $p^*_{i,j}$. It is rejected with probability $1 - p^n$, and then the background state changes to j w.p. $p_{i,j}$. Unlike in the previous section however, we shall restrict ourselves to the case when packet sizes are i.i.d. and exponentially distributed with rate μ_i when the background state is i (i.e., the service speed may depend on the background state).

As in Proposition 1, we may conclude that the system is stable when $p < 1$. Let $\{q_{n,i}\}$, $n = 0, 1, \ldots$, and $i = 1, \ldots, N$, be the steady-state probability that the system contains n customers while the background state is i. In the following, we shall obtain the z-transform of $\{q_{n,i}\}$ defined as

$$Q_i(z) := \sum_{n=0}^{\infty} z^n q_{n,i},\ i = 1, \ldots, N, \tag{9}$$

$$\text{and } \overline{Q}(z) := [Q_1(z)\ Q_2(z)\ \ldots Q_N(z)]. \tag{10}$$

The steps to obtain $\overline{Q}(z)$ closely follow the steps for obtaining $\overline{\Phi}(s)$ in Section III. Let $\boldsymbol{\mu}$ (resp. $\boldsymbol{\Lambda}$) be a $N \times N$ diagonal matrix with μ_i (resp. λ_i) as its ith diagonal element, and let $\overline{q}_0 := \overline{Q}(0)$. Then,

Lemma 3. *The joint transform vector*

$$\overline{Q}(z) = \overline{q}_0 \left[\sum_{i=0}^{\infty} \mathbf{D}(p^i z)\mathbf{A}^{-1}(p^i z) \left[\prod_{j=0}^{i-1} \mathbf{C}(p^j z)\mathbf{A}^{-1}(p^j z) \right] \right], \tag{11}$$

where the empty product is assumed to be unity, and $\mathbf{A}(z) := (z-1)\boldsymbol{\mu} + z\boldsymbol{\Lambda}(\mathbf{I} - \mathbf{P})$,

$$\mathbf{C}(z) := -z\boldsymbol{\Lambda}\mathbf{P} + z^2\boldsymbol{\Lambda}\mathbf{P}^*, \quad \mathbf{D}(z) := (z-1)\boldsymbol{\mu}. \tag{12}$$

In order to determine $\overline{q}_0$ we shall couple the fact that $\overline{Q}(z)$ is analytic in the unit disk $\{z : |z| \leq 1\}$, i.e. it has no poles in the unit disk, with the fact that $\det(\mathbf{A}^{-1}(z))$ has N poles in the unit disk, and deduce that $\overline{q}_0$ is such that the RHS of (11) should remain analytic at these N singularities.

Lemma 4. *The N zeros of the polynomial $det(\mathbf{A}(z))$ lie in the disk $\{z : |z| \leq 1\}$.*

Let $\underline{\alpha}_i$, $i = 1, 2, ..., N$ denote the right eigenvectors of $\mathbf{I} + \boldsymbol{\mu}^{-1}\boldsymbol{\Lambda}(\mathbf{I} - \mathbf{P})$ corresponding to the eigenvalue γ_i, and let

$$\mathbf{M}(z) = \sum_{i=0}^{\infty} \left[\prod_{j=0}^{i-1} \mathbf{C}(p^j z)\mathbf{A}^{-1}(p^{j+1} z) \right] \mathbf{D}(p^i z). \tag{13}$$

In order to determine $\overline{q}_0$, we make the following assumption on the eigenvalues γ_i.

Assumption 2. *For any pair i_1 and i_2 such that $i_1 \neq i_2$,*

$$\gamma_{i_1} \neq p^j \gamma_{i_2},\ \forall j \in \mathbb{Z}.$$

We now have the following result.

Theorem 2. *The probabilities $\overline{q}_0$ are the unique solution to*

$$\overline{q}_0 = \begin{bmatrix} \mathbf{M}(1)\mathbf{A}^{-1}(1)\underline{\boldsymbol{\alpha}}_1 \\ \mathbf{M}(\gamma_2)\underline{\boldsymbol{\alpha}}_2 \\ \vdots \\ \mathbf{M}(\gamma_N)\underline{\boldsymbol{\alpha}}_N \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{14}$$

Let L be the number of customers in the system as seen by an accepted arrival in steady state. We can derive the z-transform of L by using arguments similar to those used for deriving the LST of the waiting time in the previous section.

Proposition 3.

$$E[z^L] = \frac{\sum_i \lambda_i Q_i(zp)}{\sum_i \lambda_i Q_i(p)}.$$

References

[1] S. Floyd and V. Jacobson, "Random early detection gateways for congestion avoidance," *IEEE/ACM Transactions on Networking*, vol. 1, no. 4, pp. 397–413, August 1993.

[2] R. Pan, B. Prabhakar, and K. Psounis, "CHOKe: A stateless active queue management scheme for approximating fair bandwidth allocation," in *Proceedings of the IEEE INFOCOM*, 2000.

[3] O. Boxma and B. Prabhu, "Analysis of an $M/G/1$ queue with customer impatience and an adaptive arrival process," Eurandom Research Report 2009-028, 2009, available at http://www.eurandom.tue.nl/reports.

[4] V. Mishra, W. Gong, and D. Towsley, "Fluid based analysis of a network of AQM routers supporting TCP Flows with an application to RED," in *Proceedings of the ACM SIGCOMM*, 2000.

[5] P. Kuusela, P. Lassila, J. Virtamo, and P. Key, "Modeling RED with idealised TCP sources," in *Proceedings of IFIP ATM and IP*, 2001.

[6] N. van Foreest, M. Mandjes, and W. Scheinhardt, "Analysis of a feedback fluid model for heterogeneous TCP sources," *Stochastic Models*, vol. 19, no. 3, pp. 299–324, 2003.

[7] Y. Wang and B. Sengupta, "Performance analysis of a feedback congestion control policy under non-negligible propagation delay," in *Proceedings of ACM SIGCOMM*, 1991, pp. 149–157.

[8] D. Perry and S. Asmussen, "Rejection rules in the M/G/1 queue," *Queueing Systems - Theory and Applications*, vol. 29, pp. 105–130, 1995.

[9] M. Combé, "Impatient customers in the MAP/G/1 queue," Centrum voor Wiskunde en Informatica (CWI), Amsterdam, Research Report BS-R9413, 1994.

[10] O. Boxma, O. Kella, D. Perry, and B. Prabhu, "Analysis of an M/G/1 queue with impatience and an adaptive arrival process," in *Proceedings of IWAP*, 2008.