Joint Distribution of Instantaneous and Averaged Queue Length in an M/M/1/K System

E. Kuumola[1], J. Resing[2] and J. Virtamo[1]

[1]Networking Laboratory
Helsinki University of Technology
P.O. Box 3000, FIN-02015-HUT, Finland
Email: {Eemeli.Kuumola, Jorma.Virtamo}@hut.fi

[2]Department of Mathematics and Computer Science
Eindhoven University of Technology
P.O. Box 513, 5600 MB Eindhoven, The Netherlands
Email: j.a.c.resing@tue.nl

Abstract

We consider the joint dynamics of the instantaneous and exponentially averaged queue length in an M/M/1/K queue. A system of ordinary differential equations is derived for the joint stationary distribution of the instantaneous and the exponentially averaged queue length. The solution of the system of equations is obtained in a few special cases. Three different numerical approaches are presented to find the stationary distribution in the general case. Some results obtained with the numerical methods are presented and the efficiency of the numerical approaches is discussed. In addition, we describe how the model can be extended to a more complex situation which contains a rejection mechanism that randomly drops incoming customers with a dropping probability that depends on the current state of the averaged queue length.

Keywords: Exponential averaging, fluid queue, RED, stochastic discretization, embedded Markov chain, method of characteristics

1 Introduction

Congestion control and service differentiation in the Internet are central problem areas for current teletraffic research. The goal is to develop mechanisms which, while being scalable and easy to implement, allow providing certain level of Quality of Service assurances for the users of the network.

Several active queue management methods have been proposed for congestion control purposes. Notably, the random early detection (RED) mechanism was suggested by Floyd and Jacobson [8] as a means to avoid global synchronization of the TCP sources, which may happen in an uncontrolled queue, where upon a buffer overflow all the sources first halve their sending rates and then gradually increase the rates more or less synchronously. In RED, some packets are dropped randomly even before the buffer is full, thus spreading out the phases of the cycling TCP sources. In order to make the behaviour of this mechanism smoother, not reacting too aggressively to short time fluctuations of the instantaneous queue, it was proposed that the packet dropping is controlled by an average queue. The dropping probability was chosen to be a deterministic function of the average queue as explained in [8]. Several variants of the RED mechanism with different refinements have also been developed, such as weighted RED (WRED), cf. e.g. [5], RED with an in/out bit (RIO) [6], adaptive RED (ARED) [7], stabilized RED (SRED) [15] and flow RED (FRED) [14]. The use of RED has also been proposed in the Assured Forwarding (AF), which is one of the differentiated services (DiffServ) traffic handling mechanisms. In fact, some of these mechanisms are today already being deployed in routers of the Internet.

Obviously, due to these developments it is important to understand the joint dynamics of the instantaneous and average queues. This is the problem addressed in the present paper. In particular, our aim is to define a specific system model, elaborate the equations that govern the system and study different approaches for solving these equations. In earlier work [11]-[13] the dynamics has been studied in terms of the behavior of the expected values of the instantaneous and average queue lengths. In this paper we focus on the full joint distribution of these quantities.

To be specific, we consider an ordinary $M/M/1/K$ queue and the related average queue. By an average queue we mean an exponentially weighted moving average of the instantaneous queue. For brevity, this is called either exponentially averaged queue or just averaged queue. The averaging here refers to a time average. This differs from the definition of the average queue in RED, where the average queue is updated upon arrival of each new packet with no regard to the time elapsed between the arrivals. Such an event driven average changes more slowly when there are few arrivals, and more rapidly when the arrivals are frequent. Time average is, however, easier to analyze and is studied here. It can also be claimed that the event based approach has been adopted solely by implementation considerations, while the time average is perhaps more desirable. Notably, in RED an exceptional handling of the average queue is specified for the case of an empty queue, which makes the average look more like the time average with an exponential weight.

In our model, packet arrivals to the queue are assumed to constitute a Poisson process. While this assumption is primarily made for the tractability of the problem, it can be argued that the short term behaviour of a packet stream may not be too far from a Poisson process. A more serious shortcoming appears to be that this assumption neglects the packet dropping the average queue is supposed to control. While this is true, the aim of the paper really is to establish the methodology and to demonstrate how the problem can be solved in the simpler case. The analysis of the case where packet dropping is taken into account can obtained as a rather straightforward generalization of the method presented in this paper, as will be discussed. We will also point out a possible extension of the work to the case where the external arrival process is a Markov modulated Poisson process.

The paper focuses on the study of the dynamics of the joint process of the instantaneous and average queues of an $M/M/1/K$ system. Despite of the 'classical' nature of this problem, it has not been analyzed before, as far as the authors are aware. The state of the system is specified by a pair of a discrete and a continuous variable. The setting is similar to a fluid queue driven by a Markov modulated rate process (MMRP). Indeed, our analysis draws much on the seminal paper [4] on fluid queues; the equations are very similar in both systems.

Our main interest is in the stationary joint distribution of the state variables. This is governed by a system of coupled ordinary differential equations (ODEs). An analytical solution to these equations is found in a few special cases. Unfortunately, a general solution has still evaded us. In fact, despite of the deceiving simplicity of the found exact solutions, we believe that the general solution does not have a simple form. Furthermore, it also turns out that a direct numerical solution of the system of ODEs is unstable. Therefore, we develop different alternative approximate ways for finding the stationary distribution.

Two of the methods consider the temporal behaviour of the state distribution. By Kolmogorov's theorem, starting from any initial distribution, the system will eventually approach an equilibrium, i.e. integrating the equations in time is inherently stable. The first of the methods considers the evolution of the system in continuous time, while in the second approach an embedded system in discrete time is studied. A disadvantage of these methods is that the equilibrium is only approached asymptotically, and with a long averaging time the convergence is slow. The third method focuses directly on the equilibrium distribution but using an approximation. The method applies the stochastic discretization approach, previously presented in [2, 3], where the deterministic evolution of the continuous variable is replaced by small stochastic transitions, thus allowing the use of standard methods of Markovian systems.

The rest of the paper is organized as follows. First the model for the combined system of instantaneous and exponentially averaged queue is introduced in section 2. We also derive a system of ordinary differential equations governing the joint distribution of instantaneous and exponentially averaged queue length. Analytical solutions for the joint stationary distribution function in a few specific cases are obtained in section 3. In section 4, the three different numerical methods for approximately finding the joint stationary distribution function are presented. A comparison of the numerical approaches and some other numerical results are given in section 5. In section 6 we discuss extensions of the current model. Conclusions are given in section 7.

2 Model description

We consider an $M/M/1/K$ queueing system. Customers arrive at the system according to a Poisson process with arrival rate λ and have exponentially distributed service times with parameter μ. With $L(t)$ we denote the instantaneous queue length at time t. In addition, we define the exponentially averaged queue length $S(t)$ at time t by,

$$S(t) = \int_0^\infty L(t-u)\alpha e^{-\alpha u} du. \tag{1}$$

Here, α is a weighting (or averaging) parameter and by definition $L(t) = 0$ for $t \leq 0$.

It is readily seen that the process $S(t)$ obeys the differential equation

$$\frac{d}{dt}S(t) = -\alpha(S(t) - L(t)), \tag{2}$$

i.e. the rate at which $S(t)$ changes is proportional to the difference at time t between the instantaneous and the exponentially averaged queue length. The influence of the instantaneous queue length process $L(t)$ on the exponentially averaged queue length process $S(t)$ is illustrated in Figure 1.

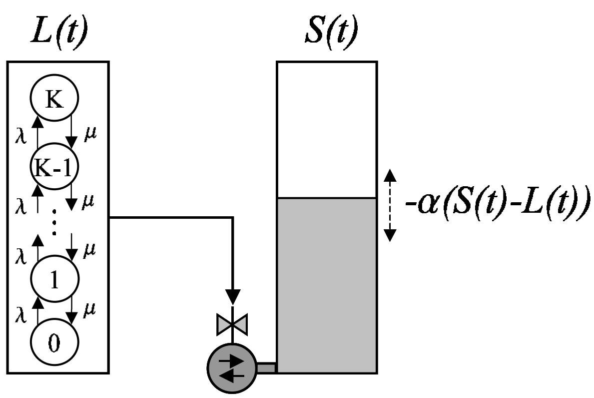


Figure 1: Influence of instantaneous queue on exponentially averaged queue

From equation (2) and the definition of $L(t)$ it is easily seen that the two-dimensional process $(L(t), S(t))$ is a Markov process with state space

$$\mathcal{S} = \{(i, x) : i \in \{0, \ldots, K\}, x \in [0, K]\}.$$

In the sequel we study the joint distribution of the process $(L(t), S(t))$. As already mentioned in the introduction, the setting resembles the situation in classical Markov modulated fluid queues: a continuous-state process $S(t)$ regulated by a discrete-state Markov process $L(t)$. However, unlike as in the classical models, the rate at which the process $S(t)$ changes at time t not only depends on $L(t)$ but also on $S(t)$ itself (see (2)).

Define the partial cumulative distribution functions

$$F_i(t,x) = P\{L(t) = i, S(t) \leq x\}, \qquad i = 0,\dots,K. \tag{3}$$

The time evolution of the partial cumulative distribution functions are governed by the forward Kolmogorov equations, which can be written as the system of partial differential equations,

$$\begin{aligned} &\frac{\partial}{\partial t}F_i(t,x) - \alpha(x-i)\frac{\partial}{\partial x}F_i(t,x) = \\ &\quad \lambda_{i-1}F_{i-1}(t,x) - (\lambda_i+\mu_i)F_i(t,x) + \mu_{i+1}F_{i+1}(t,x), \qquad i = 0,\dots,K, \end{aligned} \tag{4}$$

where

$$\lambda_i = \begin{cases} \lambda, & 0 \leq i \leq K-1, \\ 0, & i = K, \end{cases} \qquad \text{and} \quad \mu_i = \begin{cases} \mu, & 1 \leq i \leq K, \\ 0, & i = 0. \end{cases} \tag{5}$$

Introducing the notation

$$\begin{cases} \mathbf{F}(t,x) = (F_0(t,x), F_1(t,x), ..., F_K(t,x))^{\mathrm{T}}, \\ \\ \mathbf{D}(x) = \alpha\ \mathbf{diag}(0-x, 1-x, 2-x, ..., K-x), \\ \\ \mathbf{Q} = \begin{pmatrix} -\lambda & \lambda & 0 & \cdots & 0 \\ \mu & -(\lambda+\mu) & \lambda & & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & & \mu & -(\lambda+\mu) & \lambda \\ 0 & \cdots & 0 & \mu & -\mu \end{pmatrix}, \end{cases}$$

equation (4) can be alternatively written as

$$\frac{\partial}{\partial t}\mathbf{F}(t,x) + \mathbf{D}(x)\frac{\partial}{\partial x}\mathbf{F}(t,x) = \mathbf{Q}^{\mathrm{T}}\mathbf{F}(t,x). \tag{6}$$

In particular, we are interested in studying the stationary distribution of the process $(L(t), S(t))$. Defining $\mathbf{F}(x) = \lim_{t\to\infty}\mathbf{F}(t,x)$, and using $\lim_{t\to\infty}\frac{\partial}{\partial t}\mathbf{F}(t,x) = 0$ we obtain for $\mathbf{F}(x)$ the system of differential equations

$$\mathbf{D}(x)\frac{d}{dx}\mathbf{F}(x) = \mathbf{Q}^{\mathrm{T}}\mathbf{F}(x). \tag{7}$$

The boundary conditions for the differential equations are given by

$$F_i(0) = 0, \qquad i = 0,\dots,K, \tag{8}$$

$$F_i(K) = \pi_i, \qquad i = 0,\dots,K, \tag{9}$$

where π_i denotes the stationary probability of having i customers in an $M/M/1/K$ queue. Clearly, with $\rho = \lambda/\mu$, the probabilities π_i are equal to

$$\pi_i = \frac{\rho^i}{\sum_{j=0}^K \rho^j} = \frac{1-\rho}{1-\rho^{K+1}}\,\rho^i, \qquad i = 0,1,\dots,K. \tag{10}$$

3 Analytical solution for some special cases

In general, the solution of equation (7) together with boundary conditions (8) and (9) is difficult to find. However, in some special cases we are able to find an analytical solution. In section 3.1, we present the solution for the case $K = 1$, i.e., the situation in which the instantaneous queue length can only be in two different states. After that, in section 3.2, we present the solution for some special choices of the parameters α, λ and μ in the case $K = 2$.

3.1 The $M/M/1/1$ system

In this case, the system of differential equations (7) is given by

$$\begin{cases} \alpha(0-x)\frac{d}{dx}F_0(x) &= -\lambda F_0(x) + \mu F_1(x), \\ \alpha(1-x)\frac{d}{dx}F_1(x) &= \lambda F_0(x) - \mu F_1(x). \end{cases} \tag{11}$$

Now, it is straightforward to check that the solution of (11) together with boundary conditions (8) and (9) is given by

$$\begin{cases} F_0(x) &= \dfrac{\mu}{\lambda+\mu} \cdot \dfrac{B(x, \lambda/\alpha, \mu/\alpha+1)}{B(\lambda/\alpha, \mu/\alpha+1)}, \quad 0 \le x \le 1, \\ F_1(x) &= \dfrac{\lambda}{\lambda+\mu} \cdot \dfrac{B(x, \lambda/\alpha+1, \mu/\alpha)}{B(\lambda/\alpha+1, \mu/\alpha)}, \quad 0 \le x \le 1, \end{cases} \tag{12}$$

where $B(\cdot,\cdot)$ and $B(\cdot,\cdot,\cdot)$ are the beta function and incomplete beta function, respectively, defined by

$$\begin{cases} B(z_1, z_2) &= \displaystyle\int_0^1 y^{z_1-1}(1-y)^{z_2-1}\,dy, \\ B(x, z_1, z_2) &= \displaystyle\int_0^x y^{z_1-1}(1-y)^{z_2-1}\,dy. \end{cases}$$

Remark that this implies that in the stationary situation, given that the instantaneous queue length is 0, the exponentially averaged queue length has a beta distribution with parameters λ/α and $\mu/\alpha + 1$. Similarly, given that the instantaneous queue length is 1, the exponentially averaged queue length has a beta distribution with parameters $\lambda/\alpha + 1$ and μ/α. This result coincides with formula (4.8) in Kella and Stadje [10].

3.2 Some special cases of the $M/M/1/2$ system

In the case $K = 2$, the system of differential equations (7) is given by

$$\begin{cases} \alpha(0-x)\frac{d}{dx}F_0(x) &= -\lambda F_0(x) + \mu F_1(x), \\ \alpha(1-x)\frac{d}{dx}F_1(x) &= \lambda F_0(x) - (\lambda+\mu)F_1(x) + \mu F_2(x), \\ \alpha(2-x)\frac{d}{dx}F_2(x) &= \lambda F_1(x) - \mu F_2(x). \end{cases} \tag{13}$$

Now, let us restrict our attention to the case $\lambda/\alpha = \mu/\alpha = m$, where m is some arbitrary non-negative integer. The way we proceed is that we try to find a solution of equations (13), (8) and (9) of the form

$$F_i(x) = \sum_{k=0}^{\infty} a_{k,i} x^k, \quad 0 \le x \le 2, \quad i = 0, 1, 2. \tag{14}$$

Clearly, from boundary conditions (8) we obtain $a_{0,0} = a_{0,1} = a_{0,2} = 0$. Furthermore, substitution of (14) into (13) yields, for $k \ge 0$, the following recursive relations for the coefficients $a_{k,i}$:

$$-ka_{k,0} = -ma_{k,0} + ma_{k,1}, \tag{15}$$

$$(k+1)a_{k+1,1} - ka_{k,1} = ma_{k,0} - 2ma_{k,1} + ma_{k,2}, \tag{16}$$

$$2(k+1)a_{k+1,2} - ka_{k,2} = ma_{k,1} - ma_{k,2}. \tag{17}$$

From these relations, we can obtain successively for $k = 1, 2, 3, \ldots$ all the coefficients $a_{k,i}$: first $a_{k,2}$ from (17), then $a_{k,1}$ from (16) and finally $a_{k,0}$ from (15). The coefficients obtained in this way satisfy:

- $a_{k,0} = a_{k,1} = a_{k,2} = 0$, for $k < m$;
- $a_{m,2} = a_{m,1} = 0$;
- $a_{m,0}$ can be chosen arbitrarily, say $a_{m,0} = c$;
- $a_{k,0} = a_{k,1} = a_{k,2} = 0$, for $k > 3m$.

Hence, we find polynomials $F_0(x), F_1(x)$ and $F_2(x)$ of degree $3m$ satisfying (13) and boundary conditions (8). The question remains whether or not the functions $F_0(x), F_1(x)$ and $F_2(x)$ also satisfy boundary conditions (9).

It turns out that for m odd, indeed we can choose the value of c such that also boundary conditions (9) are satisfied. Below we show, for the special cases $m = 1$ and $m = 3$ the obtained solutions for $f_i(x) = \frac{d}{dx}F_i(x)$, $i = 0, 1, 2$. The reason that we show $f_i(x)$ and not $F_i(x)$ itself is that $f_i(x)$ partly factorizes in terms x and $(2-x)$.

$m = 1:$

$$f_0(x) = \tfrac{1}{8}(2-x)^2, \quad f_1(x) = \tfrac{1}{4}x(2-x), \quad f_2(x) = \tfrac{1}{8}x^2, \quad 0 \le x \le 2$$

$m = 3:$

$$\left\{ \begin{array}{lcl} f_0(x) &=& \frac{105}{2048}x^2(2-x)^4(5x^2 - 8x + 8), \\ f_1(x) &=& \frac{105}{1024}x^3(2-x)^3(5x^2 - 10x + 8), \\ f_2(x) &=& \frac{105}{2048}x^4(2-x)^2(5x^2 - 12x + 12). \end{array} \right. \qquad 0 \le x \le 2$$

Remark that the functions $f_0(x), f_1(x)$ and $f_2(x)$ satisfy the relations

$$\left\{ \begin{array}{lcll} f_0(x) &=& f_2(2-x), & 0 \le x \le 2, \\ f_1(x) &=& f_1(2-x), & 0 \le x \le 2, \end{array} \right.$$

which is, of course, due to symmetry in the case $\lambda = \mu$.

For m even unfortunately we are not able to choose the value of c such that also boundary conditions (9) are satisfied. However, in this case we make use of the symmetry argument mentioned before to obtain the solution. Similar as before, we construct $F_i(x)$ as in (14) but now only for $0 \le x \le 1$. For $1 \le x \le 2$, we set $F_i(x) = \pi_i - F_{2-i}(2-x), \quad i = 0, 1, 2$. In this way, we can choose the value of c such that also boundary conditions (9) are satisfied. We also automatically obtain continuity of the functions $F_i(x)$ at $x = 1$. Below we show, for the case $m = 2$, again the obtained solutions for $f_i(x) = \frac{d}{dx}F_i(x)$.

$m = 2:$

$$\begin{cases} f_0(x) &= \frac{1}{15}x(3x^4 - 20x^3 + 50x^2 - 60x + 30), \\ f_1(x) &= \frac{2}{15}x^2(-3x^3 + 15x^2 - 25x + 15), \\ f_2(x) &= \frac{1}{15}x^3(3x^2 - 10x + 10), \end{cases} \qquad 0 \le x \le 1$$

$$\begin{cases} f_0(x) &= \frac{1}{15}(2-x)^3(3x^2 - 2x + 2), \\ f_1(x) &= \frac{2}{15}(2-x)^2(3x^3 - 3x^2 + x - 1), \\ f_2(x) &= \frac{1}{15}(2-x)(3x^4 - 4x^3 + 2x^2 + 4x - 2). \end{cases} \qquad 1 \le x \le 2$$

4 Different numerical approaches

As mentioned in the introduction, we have little hope of finding an analytical solution for equations (7) in the general case. Furthermore, direct integration schemes for (7) are numerically unstable. Therefore, in this section we discuss three stable methods to find, approximatively, the stationary partial cumulative distribution functions.

4.1 Method of characteristics

The time evolution of the partial cumulative distribution functions $F_i(t, x)$ is given by the forward Kolmogorov equations (4). Kolmogorov's theorem guarantees that the solutions $F_i(t, x)$ of (4) will approach the stationary distribution $F_i(x)$ independent of the chosen initial distribution function $F_i(0, x)$. This provides an approach to numerically approximate the stationary solutions.

The Kolmogorov equations constitute a set of linear partial differential equations. They can be transformed into a set of ordinary differential equations using Cauchy's method of characteristics (see, e.g. [16]). To this end, we find for each i the characteristic curve $x_i(t)$ such that $\frac{d}{dt}x_i(t) = -\alpha(x_i(t) - i)$, namely

$$x_i(t) \equiv x_i(t, x_0) = i + (x_0 - i)e^{-\alpha t}, \tag{18}$$

where $x_0 = x_i(0)$. Along the characteristic curves the $F_i(t, x_i(t))$ are just functions of time satisfying

$$\frac{d}{dt}F_i(t, x_i(t)) = \frac{\partial}{\partial t}F_i(t, x_i(t)) + \frac{d}{dt}x_i(t)\frac{\partial}{\partial x}F_i(t, x_i(t))$$

$$= \frac{\partial}{\partial t} F_i(t, x_i(t)) - \alpha\,(x_i(t) - i)\frac{\partial}{\partial x} F_i(t, x_i(t)).$$

Thus equations (4) become ordinary differential equations,

$$\begin{aligned} \frac{d}{dt} F_i(t, x_i(t)) &= \lambda_{i-1} F_{i-1}(t, x_i(t)) - (\lambda_i + \mu_i) F_i(t, x_i(t)) + \mu_{i+1} F_{i+1}(t, x_i(t)) \qquad (19) \\ &= \mathbf{Q}_i^{\mathrm{T}} \mathbf{F}(t, x_i(t)), \end{aligned}$$

where $\mathbf{Q}_i^{\mathrm{T}}$ is the ith row of the matrix $\mathbf{Q}^{\mathrm{T}}$. Now, these equations can be solved by discretizing the time, $t_n = n\Delta t$, and considering the evolution over one time step. By the simplest integration scheme we have

$$\begin{aligned} F_i(t_{n+1}, x) &\approx F_i(t_n, x_i(-\Delta t, x)) + \Delta t \cdot \mathbf{Q}_i^{\mathrm{T}} \mathbf{F}(t_n, x_i(-\Delta t, x)) \\ &= F_i(t_n, i + (x-i)e^{\alpha \Delta t}) + \Delta t \cdot \mathbf{Q}_i^{\mathrm{T}} \mathbf{F}(t_n, i + (x-i)e^{\alpha \Delta t}), \qquad (20) \end{aligned}$$

where $x_i(-\Delta t, x)$ is the position of the characteristic curve at time t_n, given that at time t_{n+1} it is at x, which by (18) means that $x_i(-\Delta t, x)) = i + (x-i)e^{\alpha \Delta t}$. A better scheme is provided by e.g. the second order Runge-Kutta method [1], which comprises of first calculating the $F_i(t_{n+1}, x)$ from (20), and then recalculating them but now approximating the derivative by $\frac{1}{2}\mathbf{Q}_i^{\mathrm{T}}(\mathbf{F}(t_n, i + (x-i)e^{\alpha \Delta t}) + \mathbf{F}(t_{n+1}, x))$. A higher order Runge-Kutta method could also be employed. In practice, also the state variable x has to be discretized in the interval $(0, K)$, $x_k = k\Delta x$. Then (20) is used to calculate the $F_i(t_{n+1}, x_k)$ at the discretization points. An interpolation, e.g. using third order polynomial interpolation, can be used to calculate the values of the function for points between the discretization points. The interpolation is needed in the next time step since $(x_k - i)e^{\alpha \Delta t}$ can refer to any point between the discretization points (or to a point outside the range, where we always have $F_i(t_{n+1}, x) = 0$ for $x < 0$, and $F_i(t_{n+1}, x) = \pi_i$ for $x > K$).

4.2 The embedded process approach

Instead of integrating the time evolution of $\mathbf{F}(t, x)$ in real time from equation (4), one can look at the Markov process only at suitably chosen embedded points. One alternative is to consider the state of the system just before the instants, t_n, when a transition occurs from a given state, i, of the process $L(t)$. The stationary distribution of $S(t_n^-)$ is then equal to the stationary conditional distribution of $S(t)$, given $L(t) = i$. The disadvantage of this choice is that the transition matrix of the embedded chain becomes complex for large K.

A second possibility is to consider the full jump chain, i.e. the chain over all transitions of $L(t)$. A slight problem associated with this choice is that these points do not represent arbitrary time instants. To obtain the stationary distribution at arbitrary time instants, an extra integral operation is needed.

Thirdly, and most simply, one can modify the process in such a way that the transition instants become representative time points in continuous time. This is obtained by uniformization, i.e., by introducing in the continuous time process self transitions in state

0 at rate μ and in state K at rate λ. Then the transitions occur at the arrival instants of two independent, uninterrupted Poisson processes running in parallel: the arrival process at rate λ and the service process at rate μ. The intervals between the embedding points are distributed exponentially as $\mathrm{Exp}(\lambda+\mu)$. With this choice, the transition probability matrix of the Markov chain $L(t_n^-)$ becomes

$$\mathbf{P} = \begin{pmatrix} \frac{\mu}{\lambda+\mu} & \frac{\lambda}{\lambda+\mu} & 0 & \dots & 0 \\ \frac{\mu}{\lambda+\mu} & 0 & \frac{\lambda}{\lambda+\mu} & & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & & \frac{\mu}{\lambda+\mu} & 0 & \frac{\lambda}{\lambda+\mu} \\ 0 & \dots & 0 & \frac{\mu}{\lambda+\mu} & \frac{\lambda}{\lambda+\mu} \end{pmatrix}. \tag{21}$$

The chain $\{L(t_n^-)\}$ has the same stationary distribution (10) as the continuous time process $\{L(t)\}$. In the sequel, we take this third possibility as the choice of our embedded points.

Let $F_{i,n}^-(x)$ and $F_{i,n}^+(x)$ be the partial cumulative distribution functions of the process $(L(t), S(t))$ at the time points $t = t_n^-$ and $t = t_n^+$, respectively, and denote $\mathbf{F}_n^-(x) = (F_{0,n}^-(x), F_{1,n}^-(x), ..., F_{K,n}^-(x))^{\mathrm{T}}$ and similarly for $\mathbf{F}_n^+(x)$. Clearly, we have

$$\mathbf{F}_n^+(x) = \mathbf{P}^{\mathrm{T}}\mathbf{F}_n^-(x). \tag{22}$$

Because there are no transitions between t_n^+ and t_{n+1}^- the partial distribution function is constant along the characteristic curve (18) and $F_{i,n+1}^-(x)$ is obtained from $F_{i,n}^+(x)$ by

$$\begin{aligned} F_{i,n+1}^-(x) &= \int_0^\infty F_{i,n}^+(i + (x-i)e^{\alpha t})\,(\lambda+\mu)e^{-(\lambda+\mu)t}\,dt \\ &= \int_0^1 F_{i,n}^+(i + (x-i)z^{-a})\,dz, \end{aligned}$$

where the latter form, obtained by a change of variable $z = e^{-(\lambda+\mu)t}$, shows that the relation depends only on $a = \alpha/(\lambda+\mu)$ and not separately on α and $(\lambda+\mu)$. With the aid of the conditions $F_{i,n}^+(x) = 0$ for $x < 0$ and $F_{i,n}^+(x) = F_{i,n}^+(K)$ for $x > K$, the integral can be split as follows

$$F_{i,n+1}^-(x) = \begin{cases} \int_{\left(\frac{i-x}{i}\right)^{1/a}}^1 F_{i,n}^+(i-(i-x)z^{-a})dz, & x \le i, \\ \left(\frac{x-i}{K-1}\right)^{1/a} F_{i,n}^+(K) + \int_{\left(\frac{x-i}{K-i}\right)^{1/a}}^1 F_{i,n}^+(i+(x-i)z^{-a})dz, & x > i. \end{cases} \tag{23}$$

Hence, if we define the operator

$$U_{i,a}F(x) = \begin{cases} \int_{\left(\frac{i-x}{i}\right)^{1/a}}^1 F(i-(i-x)z^{-a})dz, & x \le i, \\ \left(\frac{x-i}{K-1}\right)^{1/a} F(K) + \int_{\left(\frac{x-i}{K-i}\right)^{1/a}}^1 F(i+(x-i)z^{-a})dz, & x > i, \end{cases} \tag{24}$$

and the vector operator $\mathbf{U}$,

$$\mathbf{U}\mathbf{F}(x) = (U_{0,a}F_0(x),\ U_{1,a}F_1(x), \dots, U_{K,a}F_K(x))^{\mathrm{T}},$$

combination of (22), (23) and (24) shows that the evolution of the distribution of the embedded chain, $\mathbf{F}_n^-(x)$, can be written as

$$\mathbf{F}_{n+1}^-(x) = \mathbf{U}(\mathbf{P}^{\mathrm{T}}\mathbf{F}_n^-(x)).$$

4.3 The stochastic discretization approach

In the stochastic discretization approach we discretize the state space of the exponentially averaged queue length process. So, instead of studying the continuous-state process $(L(t), S(t))$ we study the discrete-state process $(L(t), N(t))$, where $N(t)$ denotes the number of discrete items in the system at time t. Each discrete item represents an amount of $1/N$ in the exponentially averaged queue. We assume that the process $(L(t), N(t))$ is a Markov process with state space $\{(i, j) : i = 0, 1, \dots, K;\ j = 0, 1, \dots, KN\}$. The dynamics of $N(t)$ is described by the following rules:

- In state (i, j) with $j < iN$, items are added with rate $\alpha(iN - j)$.
- In state (i, j) with $j > iN$, items are removed with rate $\alpha(j - iN)$.

The balance equations for the equilibrium probabilities $p(i, j)$ of the Markov process $(L(t), N(t))$ are for $i = 0$ and $i = K$ and $j = 0, \dots, KN$ given by

$$\begin{aligned}(\lambda + j\alpha)p(0, j) &= \mu p(1, j) + 1_{[j<KN]}(j + 1)\alpha p(0, j + 1),\\ (\mu + (KN - j)\alpha)p(K, j) &= \lambda p(K - 1, j) + 1_{[j>0]}(KN - j + 1)\alpha p(K, j - 1),\end{aligned}$$

and for $i = 1, \dots, K - 1$ and $j = 0, \dots, KN$ by

$$\begin{aligned}(\lambda + \mu + |iN - j|\alpha)p(i, j) &= \lambda p(i - 1, j) + \mu p(i + 1, j) +\\ &\quad 1_{[0<j\le iN]}(iN - j + 1)\alpha p(i, j - 1) + 1_{[iN\le j<KN]}(j + 1 - iN)\alpha p(i, j + 1).\end{aligned}$$

Remark that the process $(L(t), N(t))$ is a truncated quasi birth-death process with level dependent transition rates. Hence, efficient methods exist to find the limiting probabilities of the process (see [9, 17]). Once we have found the limiting probabilities $p(i, j)$, we approximate the partial cumulative distribution functions $F_i(x)$ of the continuous-state process by linear interpolation,

$$F_i(x) = \sum_{j=0}^{\lfloor Nx\rfloor} p(i, j) + (x - \lfloor Nx\rfloor) \cdot p(i, \lfloor Nx\rfloor + 1). \tag{25}$$

5 Numerical results

In this section we use the three numerical approaches to approximate the joint stationary distribution of instantaneous and exponentially averaged queue length. First, we focus on a specific case for which the analytical solution is known. For this case, the performance of the numerical approaches is evaluated in terms of the accuracy of the solution and the computation time required to obtain the solution. Next, we give some numerical results in cases for which analytical results are not available.

5.1 Comparison of numerical approaches

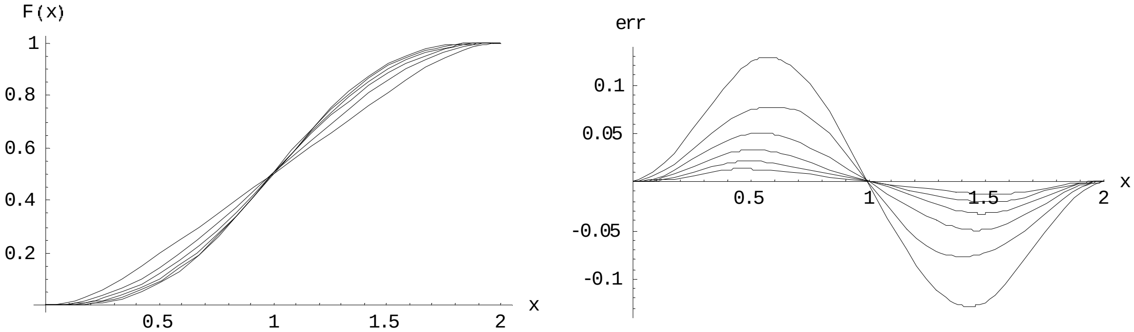
We have computed the cumulative distribution function $F(x) = F_0(x)+F_1(x)+F_2(x)$ in the case $K = 2$, $\lambda = \mu = 1$ and $\alpha = 1/5$ with the three different numerical approaches. In this case we can find the solution in analytical form as described in section 3.2. The accuracy of the numerical solution $\tilde{F}(x)$ was measured with the maximum absolute difference from the corresponding analytical solution $F(x)$,

$$\varepsilon = \max_{0\leq x\leq 2} \mid \tilde{F}(x) - F(x) \mid .$$

In order to compare the efficiency of the different numerical approaches, we measured the computation time required to numerically solve the cumulative distribution function with a desired accuracy level ε. The resolution of discretization and the other parameters of the numerical approaches affect both the accuracy and the computational efficiency of the method. With the stochastic discretization approach one has to determine the number of the intervals N the state space is divided into. The method of characteristics and the embedded process approaches are based on solving the time dependent equations on certain interval $[0, T]$ and using the solution at time instant T as an approximation for the stationary distribution. In addition to setting the length of time frame T, one has to choose the initial distribution function and also to set the resolution of the state space discretization in both of these methods. In the embedded process approach the time evolution of the distribution functions is inherently computed over the time step between successive embedded time point. However, in the case of the method of characteristics one needs to define a time step Δt over which the evolution of the distribution functions are computed.

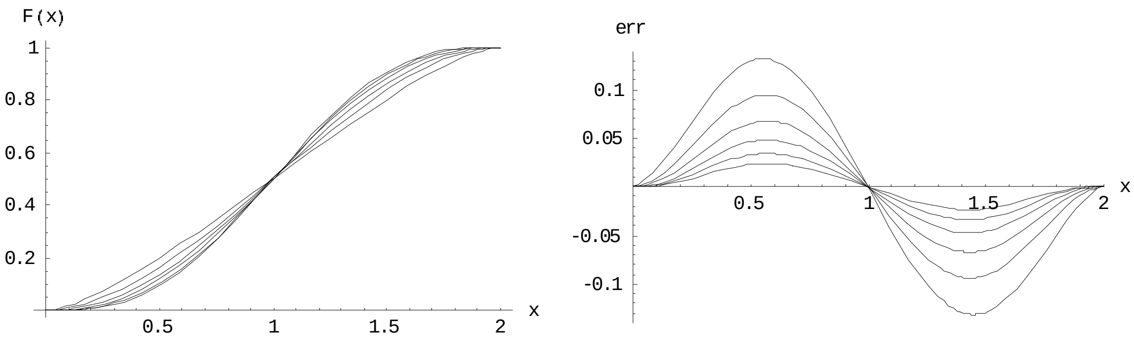
The differences between the numerical approaches complicate the comparison of the methods. The discretization of both state and time space are chosen as coarse as possible and the time frame T as small as possible such that the desired accuracy level can still be achieved. When the parameters of the numerical methods are chosen in this manner the computation times correspond to the minimum time required to solve the stationary distributions. However, we must emphasize that the actual implementation of the methods may still effect the computation time considerably.

The effect of the discretization and the length of the time frame on the numerical solutions is illustrated in Figures 2, 3 and 4. Figure 2 illustrates the evolution of the numerical solution $\tilde{F}(x)$ and the error curve $\tilde{F}(x) - F(x)$ for the method of characteristics with time frames $T = \{1, 2, 3, 4, 5, 6\}$ and Figure 3 respectively for the embedded process approach. The state space is divided into 40 intervals each of length $\Delta x = 0.05$ in both approaches. The time step for the method of characteristics was set to $\Delta t = 0.05$. The initial distribution function was taken to be uniform both in the method of characteristics and in the embedded process approach. The error decreases roughly exponentially as the time frame T increases. Figure 4 shows the evolution of the numerical solution $\tilde{F}(x)$ and the error curve $\tilde{F}(x) - F(x)$ for the stochastic discretization approach. The state space is discretizated with $N = \{2, 4, 6, 8, 10, 12\}$, corresponding to intervals of length $\Delta x \approx \{0.5, 0.25, 0.167, 0.125, 0.1, 0.083\}$.



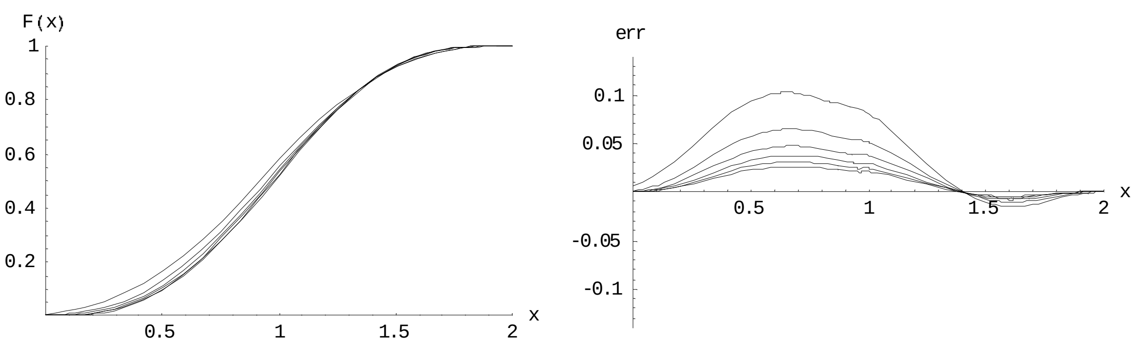
(a) Evolution of $\tilde{F}(x)$ (b) Evolution of the error $\tilde{F}(x) - F(x)$

Figure 2: Method of characteristics: numerical solution with $T = \{1, 2, 3, 4, 5, 6\}$.



(a) Evolution of $\tilde{F}(x)$ (b) Evolution of the error $\tilde{F}(x) - F(x)$

Figure 3: Embedded process approach: numerical solution with $T = \{1, 2, 3, 4, 5, 6\}$.



(a) Evolution of $\tilde{F}(x)$ (b) Evolution of the error $\tilde{F}(x) - F(x)$

Figure 4: Stochastic discretization approach: numerical solution with $N = \{2, 4, 6, 8, 10, 12\}$.

We have chosen the maximum allowed error level $\varepsilon = 10^{-3}$ in the further analysis. We evaluated the method of characteristics, embedded process and stochastic discretization approaches. The numerical approaches were implemented with Mathematica 4.1 and the test runs were carried out on a PC equipped with AMD Athlon 1400 Mhz processor and 512 Mb of memory. In Table 1, we show the computation time required to solve the problem with desired accuracy level ε. Also the required resolution of the discretization and the length of a time frame is shown in the table. Based on the results, the method of characteristics and the stochastic discretization approach were roughly ten times faster than the embedded process approach. One can note that the stochastic discretization approach requires more dense state space discretization than the other two methods.

Method	**Computation Time (s)**	Δx	T	Δt
Method of characteristics	7	0.05	16	0.02
Embedded process approach	103	0.05	15	-
Stochastic discretization	12	0.001	-	-

Table 1: Comparison of different numerical approaches.

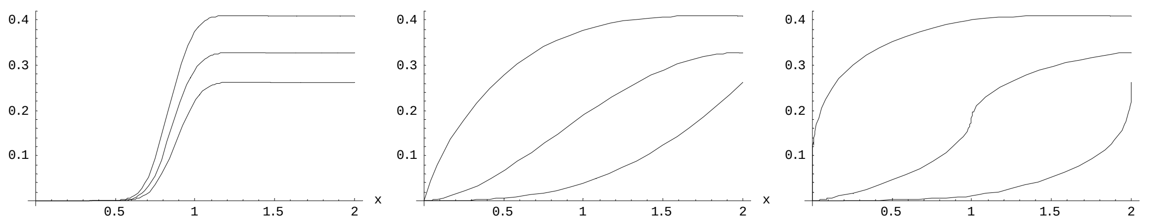
The implementation of the numerical methods can effect the computation times considerably. Because of this, the results give only a hint about the magnitude of computation time required by the approaches. In addition, the parameters of the queueing system itself affect the computation times. With larger buffer size K the computation time increases with all the methods. Furthermore, small values of the parameter α may also increase the computation time.

5.2 Numerical results in special cases

We have chosen the method of characteristics to compute the solution for the stationary distribution in some specific scenarios. In the first scenario we illustrate the effect of α on the solutions for a three state system. In the second scenario we show the solution for a system with larger buffer size.

Consider a system with $K = 2$, $\mu = 1$ and $\lambda = 0.8$. In Figure 5, we show the effect of α on the partial cumulative distribution functions $F_0(x), F_1(x)$ and $F_2(x)$. The example demonstrates a qualitative change in the curves when α passes the value 1. For $\alpha < 1$ the pdf's behave smoothly. Also all the curves are similar as $S(t)$ reacts more slowly to changes of the values of $L(t)$. When $\alpha > 1$ the derivative of each $F_i(x)$ becomes infinite at the point $x = i$ and the probability mass is more concentrated around this point.

A system with a larger buffer, $K = 20$, and $\mu = 1, \lambda = 0.8$ and $\alpha = 0.1$ is illustrated in Figure 6. In this figure the conditional cumulative distribution functions $\bar{F}_i(x) = F_i(x)/\pi_i$, $i = \{0, 1, ..., K\}$ are shown. In this case, even if α is small, the instantaneous queue length has effect on the conditional distribution of averaged queue length, which is due to the broad range of possible values of the queue length. While this example demonstrates that



(a) $\alpha = 0.1$ (b) $\alpha = 1.0$ (c) $\alpha = 3.0$

Figure 5: Effect of α on the partial cumulative distribution functions, $F_i(x), i = 0, 1, 2$ (from top to bottom)

it is feasible to handle a system of this size by the method of characteristics, however, the computation time may become considerably larger for bigger systems. In this case the computation time was over 200 times longer than with the three stage system described in section 5.1.

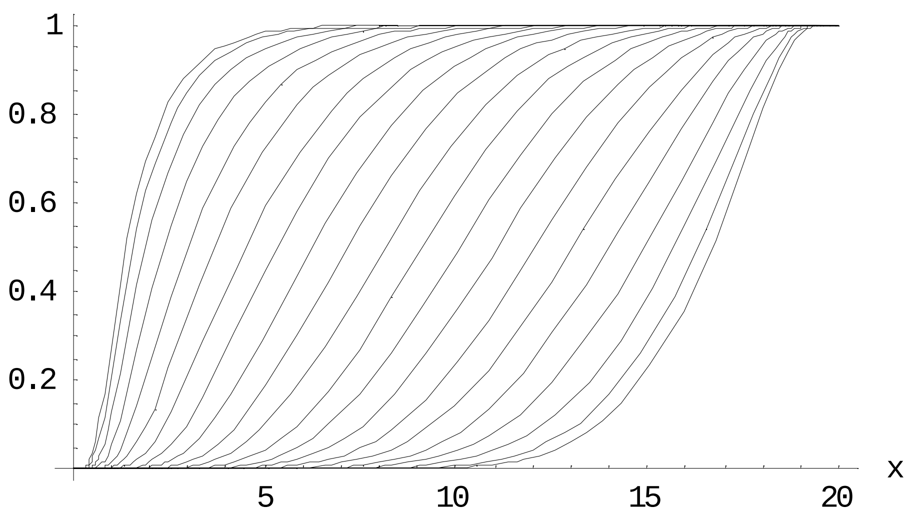


Figure 6: Conditional cumulative distribution functions $\bar{F}_i(x), i = 0, \ldots, 20$ (from left to right) for a larger buffer system, $K = 20$.

6 Model extensions

The model presented in this paper for an $M/M/1/K$ queueing system can be extended to a system in which the buffer has a rejection mechanism that randomly drops arriving customers with a probability that depends on the current state of the exponentially averaged queue length. This resembles a buffer system with a RED mechanism as discussed in the introduction.

Assume that customers arrive to the system according to a Poisson process with constant intensity λ and let $p(x)$ denote the customer dropping probability when $S(t) = x$. We denote the arrival intensity of customers admitted into the queue by $\lambda(x) = \lambda(1 - p(x))$.

The two-dimensional process $(L(t), S(t))$ still constitutes a Markov process. The forward Kolmogorov equations can now be written more conveniently for the partial probability density functions $f_i(t, x)$,

$$\frac{\partial}{\partial t}\mathbf{f}(t,x) + \mathbf{D}(x)\frac{\partial}{\partial x}\mathbf{f}(t,x) = \mathbf{Q}^{\mathrm{T}}(x)\mathbf{f}(t,x), \tag{26}$$

in which $\mathbf{f}(t, x)$ is column vector containing the $f_i(t, x)$, $\mathbf{D}(x)$ is as defined in section 2, and

$$\mathbf{Q}(x) \quad = \quad \begin{pmatrix} -(\lambda(x)-\alpha) & \lambda(x) & 0 & \cdots & 0 \\ \mu & -(\lambda(x)+\mu-\alpha) & \lambda(x) & \cdots & \vdots \\ 0 & \mu & -(\lambda(x)+\mu-\alpha) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \lambda(x) \\ 0 & \cdots & 0 & \mu & -(\mu-\alpha) \end{pmatrix}. \tag{27}$$

The numerical approaches described in section 4 require only slight modifications to be used in solving the stationary distributions in this case. The method of characteristics can be applied directly for solving the equations (26). The only difference from the original method is that matrix $\mathbf{Q}(x)$ now depends on the state variable x. Also in the embedded process approach the transition probability matrix $\mathbf{P}$ becomes dependent on the state variable x. More specifically, the upward transition probabilities $\lambda/(\lambda + \mu)$ are now split into the probability of an upward transition $\lambda(1 - p(x))/(\lambda + \mu)$ that replaces the old $\lambda/(\lambda + \mu)$ on the upper diagonal of (21) and a non-transition probability $\lambda p(x)/(\lambda + \mu)$ that has to be added to the diagonal elements. Finally, in the stochastic discretization approach the upward transition rate λ becomes dependent on the discretized state variable $N(t)$. However, no other modifications are required.

Another direction of extending the model is to use a more complex traffic model, e.g. a Markov modulated Poisson process (MMPP). Assume that the customer arrival rate is regulated by a Markov process $M(t)$ with state space $\{0, 1, ..., M\}$. The three dimensional process $(M(t), L(t), S(t))$ is a Markov process for which we can write the Kolmogorov equations. The structure of the model remains the same and we get $(M+1)(K+1)$ partial differential equations that can be solved in the stationary case, e.g., by using one of the numerical methods described in section 4.

7 Conclusions

We have developed a model for the joint dynamics of the instantaneous and the exponentially averaged queue length in an $M/M/1/K$ queue. The time evolution of the joint distribution functions of the instantaneous and averaged queue length was described with Kolmogorov equations. In the stationary case this leads to a system of ordinary differential equations for the joint distribution functions.

The analytical solution for the distribution functions was found only in a few special cases. The numerical solution of the ODE system turned out to be unstable with traditional integration schemes and thus other numerical methods were considered. Three different numerical approaches were presented to approximate the stationary solutions. Two of the methods were based on solving the time evolution of the distribution functions either over continuous time or over embedded time points. The third method is based on a stochastic discretization method. The stationary distributions were obtained in some special cases using the numerical methods and some comparison between the methods was carried out. Based on the results, we concluded that both time dependent methods as well as the stochastic discretization approach can be used to get accurate approximations for the stationary distributions.

The aim of the paper was to construct a model that describes the joint stationary properties of the instantaneous and exponentially averaged queue length in an $M/M/1/K$ queue. In addition, we described some extensions for the current model. The traffic model can be easily extended to a Markov modulated Poisson process. Furthermore, the model can also be extended to a buffer system where the customers are rejected with a dropping probability that depends on the current value of the averaged queue length. The numerical methods described in the paper can be easily adapted to find the joint stationary distributions also for the extended models.

Acknowledgments

This work was done during J. Resing's visit to the Networking Laboratory of Helsinki University of Technology. He greatly acknowledges the Networking Laboratory for their kind hospitality and financial support. Furthermore, the authors wish to thank S. Aalto for helpful discussions and comments on the work.

References

[1] M. Abramowitz and I.A. Stegun, Handbook of mathematical functions, Dover Publications Inc., New York (1972), p 896.

[2] I.J.B.F. Adan, E.A. van Doorn, J.A.C. Resing and W.R.W. Scheinhardt, Analysis of a single-server queue interacting with a fluid reservoir, Queueing Systems, Vol. 29 (1998), pp 313-336.

[3] I.J.B.F. Adan and J.A.C. Resing, A two-level traffic shaper for an on-off source, Performance Evaluation, Vol. 42 (2000), pp 279-298.

[4] D. Anick, D. Mitra and M.M. Sondhi, Stochastic theory of a data handling system with multiple sources, The Bell System Technical Journal, Vol. 61 (1982), pp 1871-1894.

[5] G. Armitage, Quality of service in IP networks, McMillan Technical Publishing, USA (2000).

[6] D. Clark and W. Fang, Explicit allocation of best effort packet delivery service, IEEE/ACM Transactions on Networking, Vol. 6, No. 4 (1998), pp 362-373.

[7] W. Feng, D. Kandlur, D. Saha and K. Shin, A self-configuring RED gateway, Proceedings of INFOCOM '99, Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies, Vol. 3 (1999), pp 1320 -1328.

[8] S. Floyd and V. Jacobson, Random early detection gateways for congestion avoidance, IEEE/ACM Transactions on Networking, Vol. 1, No. 4 (1993), pp 397-413.

[9] D.P. Gaver, P.A. Jacobs and G. Latouche, Finite birth-and-death models in randomly changing environments, Advances in Applied Probability, Vol. 16 (1984) pp 715-731.

[10] O. Kella, and W. Stadje, Exact results for a fluid model with state-dependent flow rates, Submitted for publication (2001).

[11] P. Kuusela, P. Lassila, J. Virtamo and P. Key, Modeling RED with Idealized TCP Sources, in Proceedings of IFIP ATM & IP 2001, June 27-29, 2001, Budapest, Hungary, pp 155-166.

[12] P. Kuusela, P. Lassila and J. Virtamo, Stability of TCP-RED Congestion Control, in Proceedings of ITC-17, December 2-7, 2001, Salvador da Bahia, Brazil, pp 655-666.

[13] P. Lassila and J. Virtamo, Modeling the dynamics of the RED algorithm, in Proceedings of Quality of future Internet Services, QofIS'2000, 25-26 September, 2000, Berlin, Germany, pp 28-42.

[14] D. Lin and R. Morris, Dynamics of random early detection, Proceedings of ACM SIGCOMM97, Cannes, France (1997), pp 127-137.

[15] T. Ott, T. Lakshman and H. Wong, SRED: stabilized RED, Proceedings of INFOCOM '99, Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies, Vol. 3 (1999), pp 1346 -1355.

[16] I. Sneddon, Elements of partial differential equations, McGraw-Hill (1957), pp 61-66.

[17] J. Ye and S.Q. Li, Folding algorithm: a computational method for finite QBD processes with level dependent transitions, IEEE Transactions on Communications, Vol. 42 (1994), pp 625-639.