Analytical Results on the Stochastic Behaviour of an Averaged Queue Length

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

The joint dynamics of the instantaneous and exponentially averaged queue length in an M/M/1/K queue is studied. A system of ordinary differential equations is derived for the joint stationary distribution of the instantaneous and the exponentially averaged queue length. The equations are similar to those governing an MMRP driven fluid queue. An analytical solution to the equations is obtained in a few special cases. Mean and variance of the marginal distribution of the averaged queue length, both unconditional and conditional on the value of the instantaneous queue length, are derived for the general case.

Keywords: Exponential averaging, fluid queue, RED

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 [5] 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 [5]. Several variants of the RED mechanism with different refinements have also been developed, such as weighted RED (WRED), cf. e.g. [2], RED with an in/out bit (RIO) [3], adaptive RED (ARED) [4], stabilized RED (SRED) [13] and flow RED (FRED) [12]. 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 give some analytical results relating to their solution. In earlier work [9]-[11] 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.

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. Thus we are led to 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.

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 [1] on fluid queues; the equations are very similar in both systems.

The stationary joint distribution of the state variables is governed by a system of coupled ordinary differential equations (ODEs). An analytical solution to these equations is found in a few special cases. These solutions may give some hint about the structure of the general solution, though finding such a solution has evaded us. In the general case, however, we can find the mean and variance of the conditional (conditioned on the value of the instantaneous queue) as well as the unconditional distribution of the averaged queue length. It also turns out that a direct numerical solution of the system of ODEs is unstable and needs some special techniques. These are not developed in this paper but the interested reader is referred to a companion paper [8].

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 analytical solutions for the mean and variance of the exponentially averaged queue length are derived. Conclusions are given in section 5.

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.

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

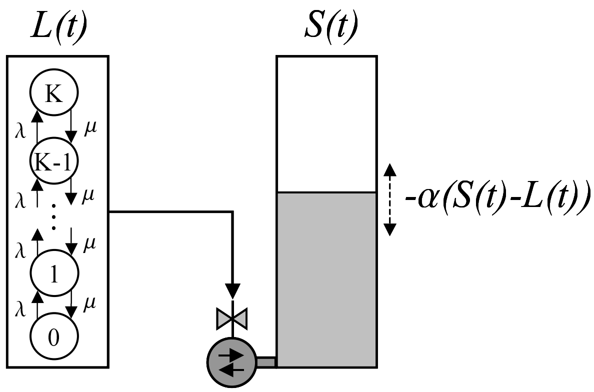


Figure 1: Influence of instantaneous queue on exponentially averaged queue

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

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

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

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

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

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

equation (4) can alternatively be 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

$$\begin{aligned} F_i(0) &= 0, \qquad i = 0,\ldots,K, && (8)\\ F_i(K) &= \pi_i, \qquad i = 0,\ldots,K, && (9)\end{aligned}$$

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

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

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

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

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

$$\left\{\begin{aligned} F_0(x) &= \frac{\mu}{\lambda+\mu}\cdot\frac{B(x,\lambda/\alpha,\mu/\alpha+1)}{B(\lambda/\alpha,\mu/\alpha+1)}, \quad 0\le x\le 1,\\ F_1(x) &= \frac{\lambda}{\lambda+\mu}\cdot\frac{B(x,\lambda/\alpha+1,\mu/\alpha)}{B(\lambda/\alpha+1,\mu/\alpha)}, \quad 0\le x\le 1,\end{aligned}\right. \tag{12}$$

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

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

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 [6].

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 $(2 - x)$.

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

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

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

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

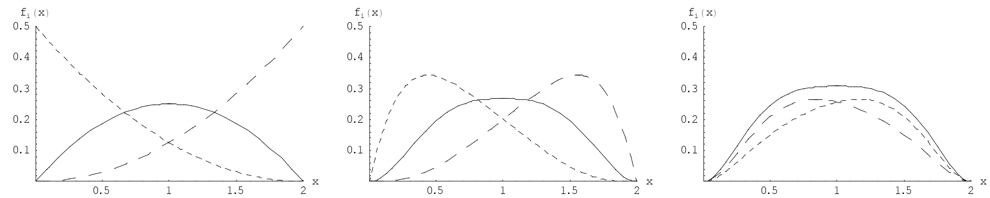
which is, of course, due to symmetry in the case $\lambda = \mu$.

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 $f_i(x) = \frac{d}{dx}F_i(x)$.

$$m = 2: \quad \left\{ \begin{array}{lcl} f_0(x) & = & \frac{1}{15}x(3x^4-20x^3+50x^2-60x+30), \\ f_1(x) & = & \frac{2}{15}x^2(-3x^3+15x^2-25x+15), \\ f_2(x) & = & \frac{1}{15}x^3(3x^2-10x+10), \end{array} \right. \qquad 0 \le x \le 1$$

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

The partial probability density functions $f_i(x), i = 0, 1, 2$ are illustrated in Figure 2.



(a) $m = 1$ (b) $m = 2$ (c) $m = 3$

Figure 2: Partial probability density functions, $f_i(x), i = 0, 1, 2$ (small dashed line, solid line, large dashed lined) with $m = 1, 2, 3$

4 Mean and variance of the averaged queue

The previous section represents our current knowledge about the exact analytical solution of the joint stationary distribution of instantaneous and averaged queue length. In this section we will derive results that are generally valid for the first two moments of the marginal distribution of the averaged queue, both conditional and unconditional on the value of the instantaneous queue. We consider the stationary version of the process $(L(t), S(t))$, i.e. for all $t \in (-\infty, \infty)$, $\mathbf{F}(t, x) = \mathbf{F}(x)$. In the sequel 0 represents an arbitrary point in time and so we are interested in the mean and variance of $S(0)$.

The following notation will be used

$$
\begin{aligned}
\boldsymbol{\pi} &= (\pi_0, \pi_1, \ldots, \pi_K), \quad \mathbf{e}_i = (0, \ldots, 0, \underbrace{1}_{i^{\text{th}} \text{ position}}, 0, \ldots, 0), \quad i = 0, 1, \ldots, K, \\
\mathbf{e} &= (1, 1, 1, \ldots, 1), \quad \mathbf{n} = (0, 1, 2, \ldots, K), \\
\mathbf{I} &= \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \\ 0 & 0 & & 1 \end{pmatrix} = \operatorname{diag}(\mathbf{e}), \quad \mathbf{N} = \begin{pmatrix} 0 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \\ 0 & 0 & & K \end{pmatrix} = \operatorname{diag}(\mathbf{n}).
\end{aligned}
$$

Then it holds

$$
\mathbf{n} = \mathbf{e} \cdot \mathbf{N}, \quad \mathbf{n} \cdot \mathbf{N} = \mathbf{N} \cdot \mathbf{n}^{\mathrm{T}} = (0^2, 1^2, \ldots, K^2), \quad \mathrm{E}[\mathbf{e}_{L(t)}] = \boldsymbol{\pi}, \text{ for all } t.
$$

4.1 Unconditional mean and variance

Consider $S(0)$, where 0 represents a random point of time. Note that the mean of $S(0)$ is

$$
s = \mathrm{E}[S(0)] = \mathrm{E}[L(t)] = \boldsymbol{\pi} \cdot \mathbf{n}^{\mathrm{T}}.
$$

The variance of $S(0)$ is

$$
\mathrm{V}[S(0)] = \mathrm{E}[S(0)^2] - s^2.
$$

Exploiting time reversibility [7] of $L(t)$ we get, $E[L(-u)L(-v)] = E[L(u)L(v)]$ and using (1), the second moment can be calculated as follows

$$\begin{aligned}
\mathrm{E}[S(0)^2] &= \alpha^2 \int_0^\infty du \int_0^\infty dv\, \mathrm{E}[L(u)L(v)]\, e^{-\alpha u} e^{-\alpha v} \\
&= 2\alpha^2 \int_0^\infty du \int_u^\infty dv\, \mathrm{E}[L(u)L(v)]\, e^{-\alpha u} e^{-\alpha v} && \mid \text{ by symmetry} \\
&= 2\alpha^2 \int_0^\infty du \int_0^\infty dv\, \mathrm{E}[L(u)L(u+v)]\, e^{-\alpha u} e^{-\alpha(u+v)} && \mid\ v \leftarrow (v+u) \\
&= 2\alpha^2 \int_0^\infty du\, e^{-2\alpha u} \int_0^\infty dv\, \mathrm{E}[L(0)L(v)]\, e^{-\alpha v} && \mid \text{ by stationarity} \\
&= \alpha \int_0^\infty \mathrm{E}[L(0)L(v)]\, e^{-\alpha v}\, dv. && (18)
\end{aligned}$$

The expectation in the integrand can be calculated by conditioning

$$\mathrm{E}[L(0)L(v)] = \mathrm{E}[L(0) \cdot \mathrm{E}\,[L(v) \,|\, L(0)]].$$

Denote the state probability vector at time v by $\boldsymbol{\pi}(v)$. Conditioned on the value $L(0)$ of the initial queue length, i.e. on the initial state $\boldsymbol{\pi}(0) = \mathbf{e}_{L(0)}$, $\boldsymbol{\pi}(v)$ evolves as

$$\boldsymbol{\pi}(v) = \mathbf{e}_{L(0)} e^{\mathbf{Q}v}.$$

Thus,

$$\mathrm{E}\,[L(v) \,|\, L(0)] = \boldsymbol{\pi}(v)\mathbf{n}^{\mathrm{T}} = \mathbf{e}_{L(0)} e^{\mathbf{Q}v} \mathbf{n}^{\mathrm{T}},$$

and

$$\begin{aligned}
\mathrm{E}[L(0) \cdot \mathrm{E}\,[L(v) \,|\, L(0)]] &= \mathrm{E}[L(0)\mathbf{e}_{L(0)}] e^{\mathbf{Q}v} \mathbf{n}^{\mathrm{T}} = \mathrm{E}[\mathbf{e}_{L(0)}\mathbf{N}] e^{\mathbf{Q}v} \mathbf{n}^{\mathrm{T}} \\
&= \boldsymbol{\pi}\mathbf{N} e^{\mathbf{Q}v} \mathbf{n}^{\mathrm{T}}.
\end{aligned}$$

Substituting this into (18) finally yields

$$\mathrm{V}[S(0)] = \alpha \boldsymbol{\pi}\mathbf{N}(\alpha\mathbf{I} - \mathbf{Q})^{-1}\mathbf{n}^{\mathrm{T}} - (\boldsymbol{\pi}\mathbf{n}^{\mathrm{T}})^2.$$

Note that in the limit $\alpha \to \infty$ (implying $S(0) \to L(0)$) this correctly reduces to

$$\mathrm{V}[S(0)] \to \boldsymbol{\pi}\mathbf{N}\mathbf{n}^{\mathrm{T}} - (\boldsymbol{\pi}\mathbf{n}^{\mathrm{T}})^2 = \mathrm{E}[L(0)^2] - \mathrm{E}[L(0)]^2 = \mathrm{V}[L(0)].$$

Example: $K = 1$

For a two-state system $K = 1$, with

$$\mathbf{Q} = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix} \quad \text{and} \quad \boldsymbol{\pi} = \frac{1}{1+\rho} \begin{pmatrix} 1 & \rho \end{pmatrix}$$

we have

$$\mathrm{V}[S(0)] = \frac{\bar{\alpha}}{1+\bar{\alpha}} \frac{\rho}{(1+\rho)^2},$$

where $\bar{\alpha} = \alpha/(\lambda + \mu)$.

4.2 Conditional mean and variance

Now consider $S(0)$ conditioned on the queue length at time 0, $L(0) = i$. Utilizing the reversibility property of $L(t)$ the conditional mean of $S(0)$ is

$$\begin{aligned} s_i &= \mathrm{E}\left[S(0)\,|\,L(0)=i\right] = \alpha \int_0^\infty \mathrm{E}\left[L(u)\,|\,L(0)=i\right] e^{-\alpha u}\,du \\ &= \alpha \int_0^\infty \mathbf{e}_i\, e^{\mathbf{Q}u}\mathbf{n}^{\mathrm{T}}\, e^{-\alpha u}\,du \qquad (19)\\ &= \alpha\,\mathbf{e}_i\,(\alpha\mathbf{I}-\mathbf{Q})^{-1}\mathbf{n}^{\mathrm{T}}. \end{aligned}$$

Defining the vector $\mathbf{s} = (s_0, \ldots, s_K)$ we have

$$\mathbf{s}^{\mathrm{T}} = \alpha\,(\alpha\mathbf{I}-\mathbf{Q})^{-1}\mathbf{n}^{\mathrm{T}}.$$

As a check, note that by unconditioning the correct unconditional mean is regained,

$$\boldsymbol{\pi}\mathbf{s}^{\mathrm{T}} = \boldsymbol{\pi}(\mathbf{I}-\mathbf{Q}/\alpha)^{-1}\mathbf{n}^{\mathrm{T}} = \boldsymbol{\pi}\mathbf{n}^{\mathrm{T}},$$

since $\boldsymbol{\pi}\mathbf{Q} = \mathbf{0}$.

Similarly, again using the reversibility property of $L(t)$ we can calculate the conditional second moments

$$s_i^{(2)} = \mathrm{E}\left[S(0)^2\,|\,L(0)=i\right] = 2\alpha^2\int_0^\infty du\, e^{-2\alpha u}\int_0^\infty dv\,\mathrm{E}\left[L(u)L(u+v)\,|\,L(0)=i\right] e^{-\alpha v}.$$

The conditional expectation is developed as before

$$\begin{aligned} \mathrm{E}\left[L(u)L(u+v)\,|\,L(0)=i\right] &= \mathrm{E}\left[L(u)\mathrm{E}\left[L(u+v)\,|\,L(u)\right]|\,L(0)=i\right] \\ &= \mathrm{E}\left[L(u)\mathbf{e}_{L(u)}\,|\,L(0)=i\right] e^{\mathbf{Q}v}\mathbf{n}^{\mathrm{T}} \\ &= \mathrm{E}\left[\mathbf{e}_{L(u)}\,|\,L(0)=i\right] \mathbf{N}e^{\mathbf{Q}v}\mathbf{n}^{\mathrm{T}} \\ &= \mathbf{e}_i e^{\mathbf{Q}u}\mathbf{N}e^{\mathbf{Q}v}\mathbf{n}^{\mathrm{T}}. \end{aligned}$$

Substitution into (19) yields

$$s_i^{(2)} = 2\alpha^2\mathbf{e}_i(2\alpha\mathbf{I}-\mathbf{Q})^{-1}\mathbf{N}(\alpha\mathbf{I}-\mathbf{Q})^{-1}\mathbf{n}^{\mathrm{T}},$$

and the corresponding vector $\mathbf{s}^{(2)}$,

$$\mathbf{s}^{(2)\,\mathrm{T}} = 2\alpha^2(2\alpha\mathbf{I}-\mathbf{Q})^{-1}\mathbf{N}(\alpha\mathbf{I}-\mathbf{Q})^{-1}\mathbf{n}^{\mathrm{T}}.$$

Now the conditional variances are obtained by,

$$v_i = \mathrm{V}[S(0)\,|\,L(0)=i] = s_i^{(2)} - s_i^2.$$

Example: $K = 1$

For the two-state system $K = 1$, the vector of conditional means, $\mathbf{s} = (s_0, s_1)$, reads

$$\begin{aligned}\mathbf{s} &= (\frac{\lambda}{\alpha+\lambda+\mu}, \frac{\alpha+\lambda}{\alpha+\lambda+\mu}) \\ &= \frac{1}{1+\bar{\alpha}}(\frac{\rho}{1+\rho}, \bar{\alpha}+\frac{\rho}{1+\rho}),\end{aligned}$$

and the conditional variance vector, $\mathbf{v} = (v_0, v_1)$, is

$$\begin{aligned}\mathbf{v} &= (\frac{\alpha\lambda(\alpha+\mu)}{(\alpha+\lambda+\mu)^2(2\alpha+\lambda+\mu)}, \frac{\alpha\mu(\alpha+\lambda)}{(\alpha+\lambda+\mu)^2(2\alpha+\lambda+\mu)}) \\ &= \frac{\bar{\alpha}}{(1+\bar{\alpha})^2(1+2\bar{\alpha})(1+\rho)}(\rho\bar{\alpha}+\frac{\rho}{1+\rho}, \bar{\alpha}+\frac{\rho}{1+\rho}).\end{aligned}$$

Note that in the limit $\alpha \to 0$ we have $\mathbf{s} \to \mathrm{E}[S(0)](1,1)$ and $\mathbf{v} \to \mathrm{V}[S(0)](1,1)$, i.e. when the averaging time is long conditioning on the current state has no effect. Note also that when $\alpha \to \infty$, $\mathbf{s} \to (0,1)$ and the variance tends to zero, $\mathbf{v} \to \frac{1}{2\alpha}(\frac{\rho}{1+\rho}, \frac{1}{1+\rho})$, because conditioning on the current state makes the short time average deterministic.

The mean and variance can be easily calculated using the analytical solutions for the distribution functions derived in section 3. In these cases it is easy to verify that the solutions are in agreement with the mean and variance results above.

5 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.

An analytical solution for the distribution functions was found only in a few special cases. The general formula was however derived for the conditional and unconditional mean and variance of the exponentially averaged queue length.

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] 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.

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

[3] 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.

[4] 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.

[5] 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.

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

[7] F. Kelly, Reversibility and stochastic networks, Wiley, (1979).

[8] E. Kuumola, J. Resing, J. Virtamo, Joint Distribution of Instantaneous and Averaged Queue Length in an M/M/1/K System, to be presented in the 15th ITC Specialist Seminar, July 21-24, 2002, Würzburg, Germany.

[9] 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.

[10] 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.

[11] 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.

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

[13] 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.