Analysis of an M/M/1+G System Operated under the FCFS Policy with Exact Admission Control

Technical Report No. ASD/2012/7
Dated: 18 May 2012

Sudipta Das, Indian Institute of Science, Bangalore
Debasis Sengupta, Indian Statistical Institute, Kolkata
and
Lawrence Jenkins, Indian Institute of Science, Bangalore

Indian Statistical Institute
Applied Statistics Unit
Kolkata 700 108

Analysis of an $M/M/1+G$ System Operated under the FCFS Policy with Exact Admission Control

Sudipta Das, *Indian Institute of Science, Bangalore*,
Debasis Sengupta, *Indian Statistical Institute, Kolkata*,
Lawrence Jenkins *Indian Institute of Science, Bangalore*

Abstract

In this paper, we present an exact theoretical analysis of an $M/M/1$ system, with arbitrary distribution of relative deadline till the end of service, operated under the first come first served scheduling policy with exact admission control. We provide an explicit solution to the functional equation that must be satisfied by the workload distribution, when the system reaches steady state. We use this solution to derive explicit expressions for the loss ratio and the sojourn time distribution. Finally, we compare this loss ratio with that of a similar system operating without admission control, in the cases of some common distributions of the relative deadline.

Key words and phrases: Firm real time systems; real time queue; performance evaluation; loss ratio.

1 Introduction

In firm real time systems, which mainly consist of aperiodic jobs such as web server, network router or real time database, it is typically not known when the job will arrive or what its service time and deadline will be. If too many jobs arrive simultaneously, the system becomes overloaded and the jobs begin to miss their deadlines. A job may be executed partially before its deadline expires. This occurrence is undesirable from considerations of productive utilization of processor time as well as efficient running of applications. An application can be managed better if it is known at the arrival epoch of a job whether it can be completed. There have been attempts to screen jobs at the arrival stage through a scheduling test, so that the execution of all the admitted jobs (including the latest one) can be guaranteed. Ideally one would like to use an exact admission controller (EAC), which would admit a newly arrived job to the queue only if it is possible to complete this job as well as all the previously admitted jobs. The rejection of a job based upon the workload at arrival epoch would require that the service times of the jobs in the queue and the residual service time of the job in service are known.

A job is supposed to have a deadline either till the beginning of its service or up to the end of its service. There have been extensive studies of systems where jobs keep deadline till the beginning of the service and the underlying scheduling policy is First Come First Served (FCFS) with admission control ([8], [6], [18], [14], [7]) or without it ([17], [3], [4], [13], [2], [15]). Other researchers ([9], [16]) provided exact analysis of an $M/M/1+G$ queue operated under the FCFS scheduling policy without any admission control and job deadline till the end of service.

Cohen [8] investigated $M/M/1$ systems under the FCFS scheduling policy, where customers have constant deadline till the end of service, and an admission controller is used. Gavish et al. [12] partially generalized this work to the $M/G/1$ system and obtained explicit formulae for server utilization, loss ratio, mean sojourn time, mean server busy period and distribution function of virtual waiting time in the $M/M/1$ case. A similar system was investigated by Dijk [11], who obtained an implicit relation satisfied by the workload distribution. Perry and

Asmussen [18] studied a more general system in which jobs have stochastic deadline till the end of the service. They obtained a recursive solution of the workload distribution. This solution does not simplify further for general deadline distribution (except, e.g., in the case of the exponential distribution), and does not appear to be suitable for computing the loss ratio. Also, Bekker [5] derived a formal solution for the steady-state workload density in M/G/1 systems in a generalized way for fixed deadline. In this paper we show that explicit analysis is possible in the case of M/M/1 systems with general stochastic deadline, operating under the FCFS scheduling policy with EAC.

This paper is organized as follows. Section 2 presents the underlying system model for the subsequent analysis. In section 3, we provide an explicit expression for the steady state workload distribution. In Section 4, we present the virtual and the actual sojourn time distributions in the steady state of this system, as well as the loss ratio. Section 5 provides the waiting time distribution of a job under a different implementation of EAC. Finally, in Section 6, we show that the adoption of EAC indeed reduces the loss ratio of a $G/G/1+G$ system operating under the FCFS scheduling policy.

2 System Model and Notations

We model a firm real time system with a single processor and an aperiodic workload as a single server queue with an infinite buffer. The infinite buffer ensures that there is no upper limit on the maximum number of jobs that can remain in the system. We assume that every job is ready as soon as it is released and it never suspends itself. The i^{th} job, J_i, arrives at the instant A_i. It has a service time Y_i and a relative deadline D_i. The i^{th} job requests that it is executed for Y_i time units during the time interval $[A_i, A_i + D_i]$, where $A_i + D_i$ is the absolute deadline (i.e., deadline up to the end of service). We assume that job arrivals follow a Poisson process with rate λ and service times follow common distribution $K(\cdot)$ with mean $1/\mu$. Also, the relative deadlines of the jobs follow common distribution $H(\cdot)$ having mean $1/\delta$. The workload process is $\{V(t)\}_{t\geq 0}$. The workload amount seen by the job J_i upon its arrival at time A_i is $V(A_i-)$. The job J_n is admitted to the system if and only if

$$V(A_n-) + Y_n \leq D_n.$$

Consequently,

$$V(A_n) = V(A_n-) + Y_n I_{\{V(A_n-)+Y_n \leq D_n\}},$$

where $I_{\{\cdot\}}$ is the indicator variable. We define the loss ratio, α, as the limiting probability of rejection at the stage of admission control, i.e.,

$$\alpha = \lim_{n\to\infty} P(V(A_n-) + Y_n > D_n).$$

3 The workload process

We derive the workload distribution as a solution to a functional equation that must hold, when the system reaches steady state. In the next two propositions, we provide the equation for the $M/G/1+G$ queue and its solution for the $M/M/1+G$ queue.

Proposition 1. The steady-state distribution of the workload process $V(t)$ of an $M/G/1+G$ queue with deadline till the end of service, operating under the FCFS scheduling policy with exact admission control, consists of a density $f_V(\cdot)$ on $(0,\infty)$ satisfying

$$f_V(u) = \lambda \int_0^u \int_u^\infty [K(z-v) - K(u-v)]\, dH(z) dF_V(v) \tag{1}$$

and possibly a point mass at 0 given by $\pi = 1 - \int_0^\infty f_V(u)du$.

Proof : The workload seen by the job J_n on its arrival is $V(A_n-)$. For notational convenience, we use V_n for $V(A_n-)$. If X_{n+1} is the time between the arrivals of the job J_n and J_{n+1}, then

$$V_{n+1} = (V_n + Y_n I_{\{V_n+Y_n \le D_n\}} - X_{n+1}) \vee 0.$$

For $u \ge 0$,

$$\begin{aligned} \bar{F}_{V_{n+1}}(u) &= P\Big((V_n + Y_n I_{\{V_n+Y_n \le D_n\}} - X_{n+1}) \vee 0 > u\Big) \\ &= \int_0^\infty \lambda e^{-\lambda t} P(Y_n I_{\{Y_n \le D_n - V_n\}} > u + t - V_n)dt. \end{aligned} \tag{2}$$

Now,

$$\begin{aligned} &P(Y_n I_{\{Y_n \le D_n - V_n\}} > u + t - V_n) \\ &= \int_0^\infty P(Y_n I_{\{Y_n \le D_n - v\}} > u + t - v) dF_{V_n}(v) \\ &= \bar{F}_{V_n}(u+t) + \int_{u+t}^\infty \int_0^{u+t} P(Y_n I_{\{Y_n \le z - v\}} > u + t - v) dF_{V_n}(v) dH(z) \\ &= \bar{F}_{V_n}(u+t) + \int_{u+t}^\infty \int_0^{u+t} P(u + t - v < Y_n \le z - v) dF_{V_n}(v) dH(z). \end{aligned} \tag{3}$$

Assuming steady-state distribution as well as inserting the relation (3) in (2) yields

$$\begin{aligned} &\bar{F}_V(u) \\ &= \int_0^\infty \lambda e^{-\lambda t} \left[\bar{F}_V(u+t) + \int_{u+t}^\infty \int_0^{u+t} \{K(z-v) - K(u+t-v)\} dF_V(v) dH(z)\right] dt \\ &= \int_u^\infty \lambda e^{-\lambda(t-u)} \left[\bar{F}_V(t) + \int_0^t \int_t^\infty \{K(z-v) - K(t-v)\} dH(z) dF_V(v)\right] dt. \end{aligned} \tag{4}$$

Since the last expression is differentiable, the workload distribution function $F_V(\cdot)$, has a density $f_V(\cdot)$ over $(0, \infty)$. The recursive relation (1), given in the proposition is obtained by differentiating (4). □

A different proof of the above proposition, by a level crossing argument can be found in [18]. A solution to this implicit equation is difficult to obtain. Solution through a recursive relation was given in [18]. A recursive relation generally poses difficulty for further computations, such as that of the loss ratio. However, as the next proposition shows, a closed form expression can be obtained in the special case of the $M/M/1+G$ queue.

Proposition 2. The steady-state distribution of the workload process $V(t)$ of an $M/M/1+G$ queue with deadline till the end of service, operating under the FCFS scheduling policy with exact admission control, consists of a point mass π at 0 and a density $f_V(\cdot)$ on $(0, \infty)$ satisfying

$$\frac{1}{\pi} = \mu e^{-B(0)} \int_0^\infty e^{B(u) - \mu u} du, \tag{5}$$

$$f_V(u) = \pi B'(u) e^{B(u) - \mu u - B(0)}, \tag{6}$$

where $B(u) = \lambda(e^{\mu u} \int_u^\infty e^{-\mu z} \bar{H}(z)dz + \int_0^u \bar{H}(z)dz)$ and $B'(u)$ is its first derivative.

Proof. For $K(x) = 1 - e^{-\mu x}$, we get from (1)

$$\begin{aligned} f_V(u) &= \lambda \int_0^u e^{\mu v} \int_u^\infty \left(e^{-\mu u} - e^{-\mu z}\right) dH(z) dF_V(v) \\ &= \lambda\mu \left(\int_0^u e^{\mu v} dF_V(v) \right) \left(\int_u^\infty e^{-\mu z} \bar{H}(z)dz \right). \end{aligned} \tag{7}$$

Without loss of generality we assume that there exists a point mass π at point 0. Then

$$\frac{f_V(u)}{\int_u^\infty e^{-\mu z} \bar{H}(z)dz} = \lambda\mu \left[\pi + \int_0^u e^{\mu v} f_V(v) dv \right]. \tag{8}$$

Note that $f_V(u)$ has a derivative, since RHS of (7) is differentiable. Hence, we can differentiate both sides of (8). After simplifying, we get

$$\begin{aligned} \frac{1}{f_V(u)} \frac{df_V(u)}{du} &= \lambda\mu e^{\mu u} \int_u^\infty e^{-\mu z} \bar{H}(z)dz - \frac{e^{-\mu u} \bar{H}(u)}{\int_u^\infty e^{-\mu z} \bar{H}(z)dz}, \\ \text{i.e.,} \quad d\left[\ln \frac{f_V(u)}{\int_u^\infty e^{-\mu z} \bar{H}(z)dz} \right] &= \lambda\mu \left(e^{\mu u} \int_u^\infty e^{-\mu z} \bar{H}(z)dz \right) du, \\ \text{i.e.,} \quad f_V(u) &= A_H e^{\lambda(e^{\mu u} \int_u^\infty e^{-\mu z} \bar{H}(z)dz + \int_0^u \bar{H}(z)dz)} \int_u^\infty e^{-\mu z} \bar{H}(z)dz \\ &= \frac{A_H}{\lambda\mu} B'(u) e^{B(u) - \mu u}. \end{aligned} \tag{9}$$

By substituting the expression (9) for $f_V(u)$ in (8), we obtain, after simplification,

$$\frac{A_H}{\lambda\mu} = \pi e^{-\lambda \int_0^\infty e^{-\mu z} \bar{H}(z)dz}. \tag{10}$$

Equation (6) is obtained by inserting (10) in (9). Now,

$$\pi + \int_0^\infty f_V(u)du = 1. \tag{11}$$

Hence, equation (5) is obtained by solving (11). □

We now consider models with three types of relative deadline distributions, namely degenerate, exponential and uniform. For each type of distribution, we present the workload distribution functions.

Example 1: Degenerate relative deadline. The distribution function of the relative deadline is given by

$$H(z) = \begin{cases} 0 & \text{if } z < \frac{1}{\delta}, \\ 1 & \text{if } z \geq \frac{1}{\delta}. \end{cases}$$

Hence,

$$\int_u^\infty e^{-\mu z} \bar{H}(z)dz = \begin{cases} \frac{1}{\mu}(e^{-\mu u} - e^{-\frac{\mu}{\delta}}) & \text{if } u \leq \frac{1}{\delta}, \\ 0 & \text{if } u > \frac{1}{\delta}. \end{cases}$$

and

$$\int_0^u \bar{H}(z)dz = \begin{cases} u & \text{if } u \le \frac{1}{\delta}, \\ \frac{1}{\delta} & \text{if } u > \frac{1}{\delta}. \end{cases}$$

Therefore,

$$B(u) = \begin{cases} \lambda\left(\frac{1}{\mu}(1 - e^{\mu(u-\frac{1}{\delta})}) + u\right) & \text{if } u \le \frac{1}{\delta}, \\ \frac{\lambda}{\delta} & \text{if } u > \frac{1}{\delta}. \end{cases}$$

Consequently, the workload distribution is determined by the expressions

$$\frac{1}{\pi} = e^{\frac{\lambda-\mu}{\delta} - \rho(1-e^{-\frac{\mu}{\delta}})}\left[1 + e^{\rho}\rho^{1-\rho}\int_{\rho e^{-\frac{\mu}{\delta}}}^{\rho} x^{\rho-2}e^{-x}dx\right], \tag{12}$$

$$f_V(u) = \begin{cases} \pi\lambda[1 - e^{-\mu(\frac{1}{\delta}-u)}]e^{(\lambda-\mu)u-\rho(e^{\mu x}-1)e^{-\frac{\mu}{\delta}}} & \text{if } u \le \frac{1}{\delta}, \\ 0 & \text{if } u > \frac{1}{\delta}. \end{cases} \tag{13}$$

Equations (12) and (13) are equivalent to the equations (B-1) and (6) of [12]. □

Example 2: Exponential relative deadline. The distribution function of the relative deadline is given by

$$H(z) = 1 - e^{-\delta z}.$$

Hence,

$$\int_u^\infty e^{-\mu z}\bar{H}(z)dz = \frac{e^{-(\mu+\delta)u}}{\mu+\delta}$$

and

$$\int_0^u \bar{H}(z)dz = \frac{1}{\delta}(1 - e^{-\delta u}).$$

Therefore,

$$B(u) = \frac{\lambda}{\delta}(1 - \frac{\mu}{\mu+\delta}e^{-\delta u}).$$

Consequently, the workload distribution is determined by the expressions

$$\frac{1}{\pi} = 1 + \left(\frac{\delta(\mu+\delta)}{\lambda\mu}\right)^{\frac{\mu}{\delta}} e^{\frac{\lambda\mu}{\delta(\mu+\delta)}}\int_0^{\frac{\lambda\mu}{\delta(\mu+\delta)}} x^{\frac{\mu}{\delta}}e^{-x}dx, \tag{14}$$

$$f_V(u) = \frac{\pi\lambda\mu}{\mu+\delta}e^{\frac{\lambda\mu}{\delta(\mu+\delta)} - \frac{\lambda\mu e^{-\delta u}}{\delta(\mu+\delta)} - (\mu+\delta)u}. \tag{15}$$

Equations (14) and (15) can also be found from the Corollary 2.8 of [18]. □

Example 3: Uniform relative deadline. The distribution function of the relative deadline is given by

$$H(z) = \begin{cases} \frac{\delta z}{2} & \text{if } z \le \frac{2}{\delta}, \\ 1 & \text{if } z > \frac{2}{\delta}. \end{cases}$$

Hence,

$$\int_u^\infty e^{-\mu z}\bar{H}(z)dz = \begin{cases} \frac{1}{\mu}\left[(1-\frac{\delta}{2\mu}-\frac{\delta u}{2})e^{-\mu u}+\frac{\delta}{2\mu}e^{\frac{-2\mu}{\delta}}\right] & \text{if } u \le \frac{2}{\delta}, \\ 0 & \text{if } u > \frac{2}{\delta}. \end{cases}$$

and

$$\int_0^u \bar{H}(z)dz = \begin{cases} u-\frac{\delta u^2}{4} & \text{if } u \le \frac{2}{\delta}, \\ \frac{1}{\delta} & \text{if } u > \frac{2}{\delta}. \end{cases}$$

Therefore,

$$B(u) = \begin{cases} \rho\left[(1-\frac{\delta}{2\mu}-\frac{\delta u}{2})+\frac{\delta}{2\mu}e^{\mu(u-\frac{2}{\delta})}\right]+\lambda u\left[1-\frac{\delta u}{4}\right] & \text{if } u \le \frac{2}{\delta}, \\ \frac{\lambda}{\delta} & \text{if } u > \frac{2}{\delta}. \end{cases}$$

Consequently, the workload distribution is determined by the expressions

$$f_V(u) = \begin{cases} \pi C(u) & \text{if } u \le \frac{2}{\delta}, \\ 0 & \text{if } u > \frac{2}{\delta}. \end{cases}$$

and

$$\frac{1}{\pi} = 1 + \int_0^{\frac{2}{\delta}} C(u)du$$

where,

$$C(u) = \frac{\rho}{2}\left[2\mu - \delta(1+\mu u e^{-\mu u} - e^{-\frac{2\mu}{\delta}})\right] e^{\frac{\rho\delta}{2\mu}[-u\mu+(e^{\mu u}-1)e^{-\frac{2\mu}{\delta}}]+\lambda u[1-\frac{\delta u}{4}]}. \qquad \square$$

4 Sojourn Time and Loss Ratio

In order to compute the loss ratio, we need to find the probability of a job being denied admission by the EAC. Hence, we need to calculate the probability that the sum of the workload seen by a job on arrival and its service time is greater than its relative deadline. This sum is called the virtual sojourn time (i.e., the amount of time a job having infinite deadline will spend in the system). We provide the distribution of this quantity in the next proposition.

Proposition 3. The steady-state distribution of the virtual sojourn time T of an incoming job in an $M/M/1+G$ queue with deadline till the end of service, operating under the FCFS scheduling policy with exact admission control, has the density $f_T(\cdot)$ on $[0,\infty)$ given by

$$f_T(s) = \mu\pi e^{B(s)-\mu s-B(0)}, \tag{16}$$

where π and $B(\cdot)$ is defined as in Proposition 2.

Proof. Note that the workload seen by an incoming job is independent of its service time. Therefore, the virtual sojourn time distribution can be obtained by convolving the workload distribution with the service time distribution (assumed to be exponential). Hence, by using the result of Proposition 2, we have

$$\begin{aligned} f_T(s) &= \mu\int_0^s e^{-\mu(s-u)}dF_V(u) \\ &= \mu\pi e^{-\mu s}\left[1+e^{-B(0)}\int_0^s B'(u)e^{B(u)}du\right] \\ &= \mu\pi e^{B(s)-\mu s-B(0)}. \end{aligned} \qquad \square$$

We are now ready for the next proposition, which specifies the loss ratio.

Proposition 4. The loss ratio of an $M/M/1+G$ queue with deadline till the end of service, operating under the FCFS scheduling policy with exact admission control, is given by

$$\alpha = 1 - \left(\int_0^\infty \bar{H}(s) e^{B(s)-\mu s} ds \right) \Big/ \left(\int_0^\infty e^{B(s)-\mu s} ds \right), \tag{17}$$

where $H(\cdot)$ is the deadline distribution and $B(\cdot)$ is specified in Proposition 2.

Proof. Let T and D be the virtual sojourn time and the relative deadline, respectively, of an incoming job. Hence, the loss ratio is

$$\begin{aligned} \alpha &= P(T > D) \\ &= 1 - \mu\pi e^{-B(0)} \int_0^\infty e^{B(s)-\mu s} \bar{H}(s) ds. \end{aligned} \tag{18}$$

The expression given in equation (17) is obtained by substituting (5) in (18), and simplifying the resulting expression. □

Remark 1. It is seen from equation (17), by virtue of the dominated convergence theorem, that

$$\begin{aligned} \lim_{\lambda\to 0} \alpha &= 1 - \lim_{\lambda\to 0} \mu\pi \int_0^\infty e^{B(s)-\mu s - B(0)} \bar{H}(s) ds \\ &= 1 - \left(\lim_{\lambda\to 0} \mu \int_0^\infty e^{B(s)-\mu s-B(0)} \bar{H}(s) ds \right) \Big/ \left(\lim_{\lambda\to 0} \mu \int_0^\infty e^{B(s)-\mu s-B(0)} ds \right) \\ &= 1 - \left(\int_0^\infty \left\{ \lim_{\lambda\to 0} e^{B(s)-\mu s-B(0)} \right\} \bar{H}(s) ds \right) \Big/ \left(\int_0^\infty \left\{ \lim_{\lambda\to 0} e^{B(s)-\mu s-B(0)} \right\} ds \right) \\ &= 1 - \left(\int_0^\infty e^{-\mu s} \bar{H}(s) ds \right) \Big/ \left(\int_0^\infty e^{-\mu s} ds \right) \\ &= 1 - \int_0^\infty \mu e^{-\mu s} \bar{H}(s) ds \\ &= P(D \le X), \end{aligned}$$

where D and X are the (mutually independent) service time and relative deadline of a job. Thus, when the arrival rate is very small, the loss ratio is the probability that the service time of a job exceeds its relative deadline. □

We have obtained, through Propositions 3 and 4, the distribution for the virtual sojourn time and the expression of the loss ratio, respectively. These allows us to specify the actual sojourn time distribution, through the next proposition.

Proposition 5. The steady-state distribution of the actual sojourn time T_a of an incoming job in an $M/M/1+G$ queue with deadline till the end of service, operating under the FCFS scheduling policy with exact admission control, consists of a density $f_{T_a}(\cdot)$ on $(0,\infty)$ satisfying

$$f_{T_a}(s) = \mu\pi \bar{H}(s) e^{B(s)-\mu s-B(0)}, \tag{19}$$

where π and $B(\cdot)$ is defined as in Proposition 2, and a point mass of α (given in Proposition 4) at 0.

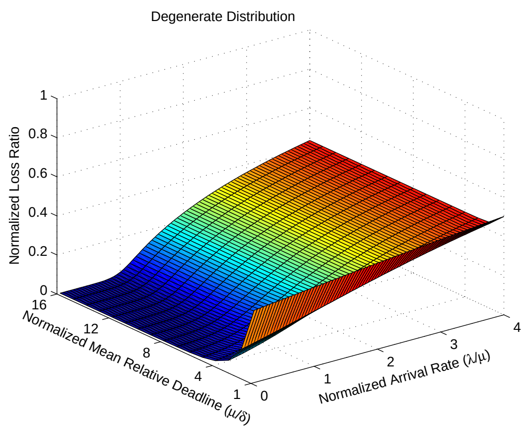

Figure 1: Loss ratios for the FCFC-EAC scheduling policy, for various values of normalized arrival rate (λ/μ) and normalized mean relative deadline (μ/δ), and fixed deadline.

Proof. Let T and D be the virtual sojourn time and the relative deadline, respectively, of an incoming job, when the queue is in steady state. Then, the actual sojourn time of the job, $T_a = TI_{\{T<D\}}$, has the distribution function

$$\begin{aligned} P(T_a \le s) &= 1 - \int_0^\infty f_T(u)P(uI_{\{u<D\}} > s)du \\ &= 1 - \int_s^\infty f_T(u)\bar{H}(u)du. \end{aligned} \tag{20}$$

We obtain equation (19) by differentiating both sides of (20). Since

$$\int_0^\infty f_{T_a}(s)ds < \int_0^\infty f_T(s)ds = 1,$$

we conclude that T_a has a point mass at 0 given by (17). □

We now revert to the examples of relative deadline distributions considered in Section 3, and present the virtual sojourn time distribution, the actual sojourn time distribution and the loss ratio.

Example 1 (contd.): Degenerate relative deadline. The virtual sojourn time distribution has the density

$$f_T(s) = \begin{cases} \pi\mu e^{(\lambda-\mu)s-\rho(e^{\mu s}-1)e^{-\frac{\mu}{\delta}}} & \text{if } s \le \frac{1}{\delta}, \\ \pi\mu e^{\frac{\lambda}{\delta}-\mu s-\rho(1-e^{-\frac{\mu}{\delta}})} & \text{if } s > \frac{1}{\delta}. \end{cases}$$

The actual sojourn time distribution consists of a density on $(0, \infty)$ satisfying

$$f_{T_a}(s) = \pi\mu e^{(\lambda-\mu)s-\rho(e^{\mu s}-1)e^{-\frac{\mu}{\delta}}} I_{\{s\le\frac{1}{\delta}\}}$$

and a point mass at 0, which is the same as the loss ratio, given by

$$\alpha = \pi e^{\frac{\lambda-\mu}{\delta}-\rho(1-e^{-\frac{\mu}{\delta}})}. \tag{21}$$

Equation (21) is equivalent to the equation (12) of [12]. □

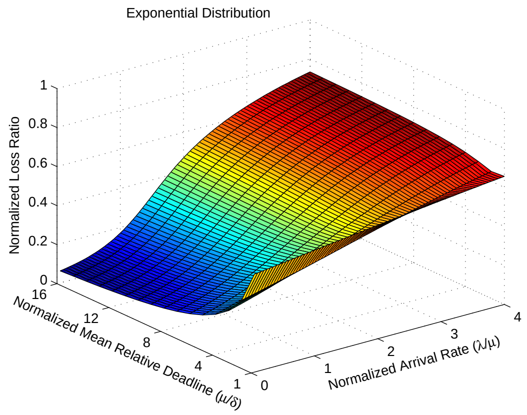

Figure 2: Loss ratios for the FCFC-EAC scheduling policy, for various values of normalized arrival rate (λ/μ) and normalized mean relative deadline (μ/δ), and the Exponential deadline distribution.

Example 2 (contd.): Exponential relative deadline. The virtual sojourn time distribution has the density

$$f_T(s) = \pi\mu e^{\frac{\lambda\mu}{\delta(\mu+\delta)} - \frac{\lambda\mu e^{-\delta s}}{\delta(\mu+\delta)} - \mu s},$$

over $[0,\infty)$. The actual sojourn time distribution consists of a density on $(0,\infty)$ satisfying

$$f_{T_a}(s) = \pi\mu e^{\frac{\lambda\mu}{\delta(\mu+\delta)} - \frac{\lambda\mu e^{-\delta s}}{\delta(\mu+\delta)} - (\mu+\delta)s}$$

and a point mass at 0, namely the loss ratio, given by

$$\alpha = 1 - \frac{\pi}{\delta}\left(\frac{\delta(\mu+\delta)}{\lambda\mu}\right)^{1+\frac{\mu}{\delta}} e^{\frac{\lambda\mu}{\delta(\mu+\delta)}} \int_0^{\frac{\lambda\mu}{\delta(\mu+\delta)}} x^{\frac{\mu}{\delta}} e^{-x} dx.$$

□

Example 3 (contd.): Uniform relative deadline. The virtual sojourn time T has the has the probability density function on $[0,\infty)$, defined by

$$f_T(s) = \begin{cases} \pi\mu e^{C_1(s)} & \text{if } s \leq \frac{2}{\delta}, \\ \pi\mu e^{C_2(s)} & \text{if } s > \frac{2}{\delta}; \end{cases}$$

and the actual sojourn time distribution consists of a density on $(0,\infty)$ of the form

$$f_{T_a}(s) = \pi\mu\left(1 - \frac{\delta z}{2}\right) e^{C_1(s)} I_{\{s \leq \frac{2}{\delta}\}}$$

and a point mass at 0, namely the loss ratio, given by

$$\alpha = 1 - \pi\mu \int_0^{\frac{2}{\delta}} \left(1 - \frac{\delta s}{2}\right) e^{C_1(s)} ds,$$

where

$$C_1(s) = \frac{\rho\delta}{2\mu}\left[(e^{\mu s} - 1)e^{-\frac{2\mu}{\delta}} - s\mu\right] + (\lambda - \mu)s - \frac{\delta s^2}{4}$$

and

$$C_2(s) = \frac{\lambda}{\delta} - \mu s - \rho\left[1 - \frac{\delta}{2\mu}(1 - e^{-\frac{2\mu}{\delta}})\right].$$

□

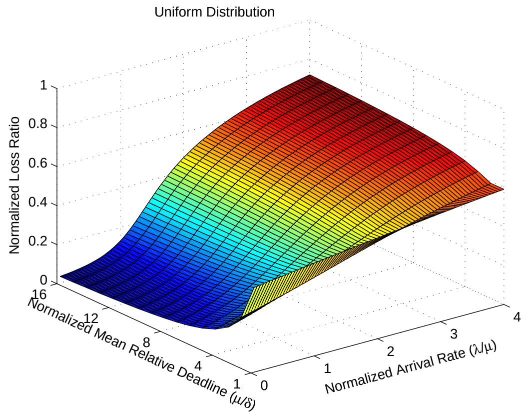

Figure 3: Loss ratios for the FCFC-EAC scheduling policy, for various values of normalized arrival rate (λ/μ) and normalized mean relative deadline (μ/δ), and the Uniform deadline distribution.

Figures 1-3 show the loss ratios of the system for degenerate, exponential and uniform deadline distributions respectively versus normalized arrival rate and normalized mean relative deadline. It is observed that loss ratio increases with the normalized arrival rate. For small normalized rates of arrival, the loss ratio is a decreasing function of the normalized mean deadline. This monotonicity prevails for all normalized arrival rates when the relative deadline is constant. However, when the relative deadline has the exponential or uniform distribution and the normalized arrival rate is large, the loss ratio can increase with the normalized mean relative deadline, particularly for large values of the latter. A possible explanation of this apparently counter-intuitive phenomenon is that stochastic deadlines with large mean lead to admission of some jobs with very large service time, which in turn lead to non-admission of several relatively smaller jobs that could otherwise have been admitted.

5 Deferred Service Decision

One generally needs an admission controller for implementation of the FCFS-EAC scheduling policy. The advantage of this controller is that a decision about the servicing of a job is taken immediately upon its arrival. In the case of FCFS systems, where generating an early decision is not crucial, one may dispense with the admission controller by deferring the decision till the epoch of a job reaching the server. The decision to accept or reject a job may be generated at that epoch by the server itself, depending on the service time of the job and its absolute deadline. Note that this arrangement may be viewed as merely a different implementation of EAC [6], as the set of serviced jobs remains the same. Even though the decision to service a job is generated at a later point of time, the advantage of this implementation is that an admission controller is not needed, as the decision can be taken by the server itself. In any case, sometimes it is only the server that knows the service time of a job [18]. Some researchers have also considered deferment of the service decision in the case of the earliest deadline first (EDF) scheduling policy ([1], [10]).

As indicated above, for any realization of the queue operated under the FCFS-EAC scheduling policy, the actual service status of the jobs do not depend on the time of the service decision. Therefore, even if the service decision is delayed, the loss ratio would continue to be the same. Further, under the deferred decision implementation, the waiting time of a job is identical to the workload seen by it at arrival. However, the departure epochs of the jobs under the two implementations are different, producing different sojourn time distributions. In the next proposition, we provide the sojourn time distribution of the jobs in an $M/M/1+G$ queue operated under the FCFS-EAC policy with deferred service decision.

Proposition 6. The steady-state distribution of the sojourn time T_e of a job in an $M/M/1+G$ queue with deadline till the end of service, operating under the FCFS-EAC policy with deferred service decision, has a density $f_{T_e}(\cdot)$ on $[0,\infty)$ satisfying

$$f_{T_e}(s) = f_T(s)\bar{H}(s) + \mu f_V(s)\int_0^\infty e^{-\mu y}H(s+y)dy, \tag{22}$$

where $f_V(\cdot)$ and $f_T(\cdot)$ are defined as in Proposition 2 and 3, respectively.

Proof. Let V, Y and D be the waiting time, service time and the relative deadline, respectively, of a job, when the queue is in steady state. Then, the sojourn time of a job, $T_e = V + YI_{\{V+Y<D\}}$, has the distribution function

$$\begin{aligned} P(T_e \le s) &= 1 - P(s < V+Y \le D) - P(V > s, V+Y > D) \\ &= 1 - \int_s^\infty f_T(u)\bar{H}(u)du - \int_0^\infty \mu e^{-\mu y}\left(\int_s^\infty f_V(u)H(u+y)du\right)dy. \end{aligned} \tag{23}$$

We obtain equation (22) by differentiating both sides of (23) and then simplifying. □

6 Comparison between loss ratios of FCFS and FCFS-EAC

While EAC is used as a means of ensuring guaranteed service of a job after its admission to the queue, one might expect that such an arrangement would also reduce the loss ratio of the system by eliminating the legacy of unproductive workload. Figures 4, 5 and 6 show surface plots of the loss ratio of an $M/M/1$ system operated under FCFS-EAC, normalized by that of a similar system operated under FCFS alone, when the deadline distribution is degenerate, exponential or uniform. We observe from the figures that EAC indeed reduces the loss ratio. In the next proposition we show that such an order exist under a more general set-up. Here, we use the notations $\alpha^H_{FCFS\text{-}EAC}$ and α^H_{FCFS} to denote the loss ratios of the system, operated under the FCFS scheduling policy with and without EAC, respectively and $H(\cdot)$ is the distribution function the relative deadline.

Proposition 7. In a $G/G/1+G$ queue, the loss ratio under the FCFS scheduling policy can only be reduced when Exact Admission Control is used, i.e.,

$$\alpha^H_{FCFS\text{-}EAC} \le \alpha^H_{FCFS}.$$

Proof. Consider the FCFS-EAC scheduling policy with deferred service decision, as described in Section 5. Consider the first n arrivals. Permit the first job, which is denied service under EAC, to be served. The fact that this job would have been denied service under EAC indicates that this job would miss its deadline. On the other hand, providing service to this job, as is done under the FCFS scheduling policy, can only increase the waiting times of *all* the subsequent jobs in the ordered list, and this increase may trigger further cases of missed deadline. Thus, the act of providing service to the said job can only increase the number of

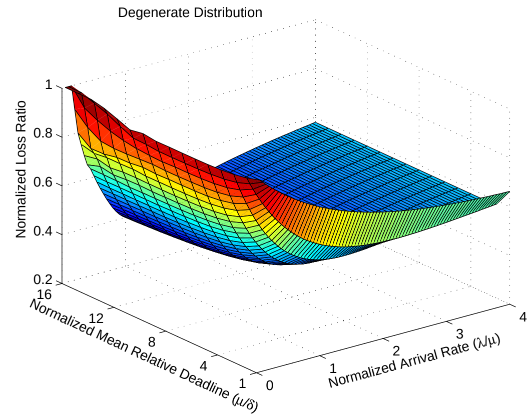

Figure 4: Loss ratios for Degenerate deadline distributions under the FCFS-EAC scheduling policy normalized by loss ratio under the FCFS scheduling policy, for various values of normalized arrival rate (λ/μ) and normalized mean relative deadline (μ/δ).

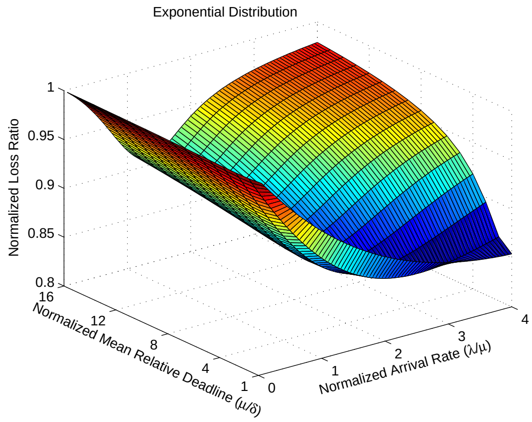

Figure 5: Loss ratios for Exponential deadline distributions under the FCFS-EAC scheduling policy normalized by loss ratio under the FCFS scheduling policy, for various values of normalized arrival rate (λ/μ) and normalized mean relative deadline (μ/δ).

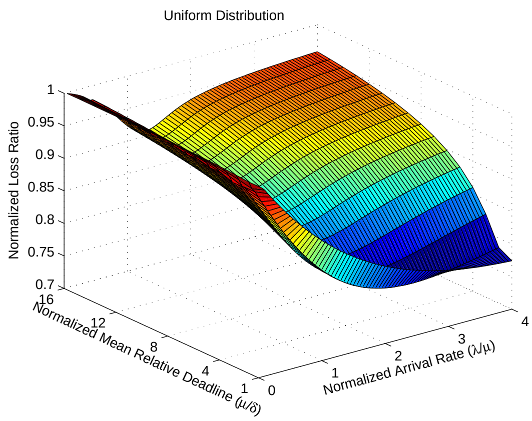

Figure 6: Loss ratios for Uniform deadline distributions under the FCFS-EAC scheduling policy normalized by loss ratio under the FCFS scheduling policy, for various values of normalized arrival rate (λ/μ) and normalized mean relative deadline (μ/δ).

jobs missing their deadlines. This argument holds for every single event of providing service to one of the jobs denied service under FCFS-EAC. Thus, for any given configuration of a finite number of jobs, the proportion of jobs missing their deadlines under FCFS-EAC is less than or equal to that under FCFS. Hence, the expected proportion of jobs missing deadline (out of the first n arrivals) is less for FCFS-EAC than for FCFS. The stated result is obtained by taking the limit of the expected proportion as n goes to infinity. □

Remark 2. Also from equation (4.11) of [16], we can obtain the loss ratio $M/M/1+G$ system under FCFS scheduling policy as follows

$$\alpha^H_{FCFS} = 1 - \mu p_0 \int_0^\infty \bar{H}(s) e^{\lambda \int_0^s \bar{H}(x)dx - \mu s} ds, \tag{24}$$

where $p_0 = \frac{1}{\mu \int_0^\infty e^{\lambda \int_0^s \bar{H}(x)dx - \mu s} ds}$.

Therefore, by applying the dominated convergence theorem, we can show that $\lim_{\lambda\to 0} p_0 = 1$ and

$$\lim_{\lambda\to 0} \alpha^H_{FCFS} = 1 - \int_0^\infty \mu e^{-\mu s} \bar{H}(s) ds.$$

Thus, when the arrival rate is very small, the loss ratio is the probability that the service time of a job exceeds its relative deadline. It has already been observed that the same conclusion holds in respect of an $M/M/1+G$ queue operated under the FCFS-EAC scheduling policy (see Remark 1). Therefore, Proposition 7 holds with equality when $\lambda \approx 0$. This fact explains why the ratio of $\alpha^H_{FCFS-EAC}/\alpha^H_{FCFS}$ plotted in Figures 4-6 approach the value 1 near the edge corresponding to $\lambda/\mu = 0$. □

References

[1] Tarek Abdelzaher and C. Lu. Schedulability analysis and utilization bounds for highly scalable real-time services. In *IEEE Real Time Technology and Applications Symposium*, pages 11–25, June 2001.

[2] F. Baccelli, P. Boyer, and Hebuterne. Single-server queue with impatient customers. *Advances in Applied Probability*, 16(4):887–905, December 1984.

[3] D. Y. Barrer. Queueing with impatient customers and indifferent clerks. *Operation Research*, 5(5):644–649, October 1957.

[4] D. Y. Barrer. Queueing with impatient customers and ordered service. *Operation Research*, 5(5):650–656, October 1957.

[5] Rene Bekker. Finite-buffer queues with workload-dependent service and arrival rates. *Queueing Systems*, 50:231–253, 2005.

[6] O. J. Boxma and P. R. de Waal. Multiserver queues with impatient customers. Technical Report BS-R9319, Centrum voor Wiskunde en Informatica, 1993.

[7] Onno Boxma, David Perry, and Wolfgang Stadje. The M/G/1+G queue revisited. *Queueing System*, 67:207–220, 2011.

[8] J. W. Cohen. Single server queues with restricted accessibility. *Journal of Engineering Mathematics*, 3(4):265–284, October 1969.

[9] D. J. Daley. General customer impatience in queue GI/G/1. *Journal of Applied Probability*, 2(1):186–205, June 1965.

[10] Sudipta Das, Lawrence Jenkins, and Debasis Sengupta. Loss ratios of EDF and FCFS scheduling policies under a performance enhancing modification. In *Proceedings of International Conference on Real-Time and Embedded Systems, RTES*, pages 128–131, November 2010.

[11] Nico M. Van Dijk. Queueing systems with restricted workload: An explicit solution. *Journal of Applied Probability*, 27:393–400, 1990.

[12] Bezalel Gavish and Paul Schweitzer. The markovian queue with bounded waiting time. *Management Science*, 23(12):1349–1357, August 1977.

[13] Rolf B. Haugen and Even Skogan. Queueing systems with stochastic time out. *IEEE Trans on Communications*, 28(12):1984–1989, December 1980.

[14] Liqiang Liu and Vidyadhar G. Kulkarni. Explicit solutions for the steady state distributions in M/PH/1 queues with workload dependent balking. *Queueing System*, 52:251–260, 2006.

[15] Ali Movaghar. On queueing with customer impatience until the beginning of service. *Queueing Systems*, 29:337–350, 1998.

[16] Ali Movaghar. On queueing with customer impatience until the end of service. *Stochastic Models*, 22:149–173, May 2006.

[17] C. Palm. Methods for judging the annoyance caused by congestion. *Tele*, (2):1–20, 1953.

[18] David Perry and Soren Asmussen. Rejection rules in the M/G/1 queue. *Queueing Systems*, 19:105–130, 1995.