Journal of Industrial and Systems Engineering
Vol. 1, No. 3, pp 260-280
Fall 2007

JISE

A Non-Preemptive Two-Class M/M/1 System with Prioritized Real-Time Jobs under Earliest-Deadline-First Policy

Mehdi Kargahi[1*], Ali Movaghar[2]

[1]Department of Electrical and Computer Engineering, University of Tehran, Tehran, Iran
kargahi@ece.ut.ac.ir

[2]Department of Computer Engineering, Sharif University of Technology, Tehran, Iran
movaghar@sharif.edu

ABSTRACT

This paper introduces an analytical method for approximating the performance of a two-class priority M/M/1 system. The system is fully non-preemptive. More specifically, the prioritized class-1 jobs are real-time and served with the non-preemptive earliest-deadline-first (EDF) policy, but despite their priority cannot preempt any non real-time class-2 job. The waiting class-2 jobs can only be served from the time instant that no class-1 job is in the system. The service discipline of the class-2 jobs is FCFS. The required mean service times may depend on the class of the jobs. The real-time jobs have exponentially distributed relative deadlines until the end of service. The system is approximated by a Markovian model in the long run, which can be solved numerically using standard Markovian solution techniques. The performance measures of the system are the loss probability of the class-1 jobs and the mean sojourn (waiting) time of the class-2 jobs. Comparing the numerical and simulation results, we find that the existing errors are relatively small.

Keywords: Approximation methods; Earliest-deadline-first (EDF) policy; Non-preemptive services; Queueing; Real-time jobs; Two-class priority M/M/1 system.

1. INTRODUCTION

Multi-priority demands for computation and communication are required by many applications of the newly developed systems such as wireless sensor networks or high speed packet switching networks (e.g., a DiffServ Router), which are usually referred to as *multi-class traffics*. This is particularly evident in the era of growing real-time, multimedia, and telecommunication systems, with both real-time and non real-time classes of traffic, in which the *quality of service* (QoS) of the applications is to be guaranteed. While certain timing constraints exist for the real-time incoming demands, where violating them beyond certain thresholds is unacceptable, the average traffic delay of the non real-time applications is also an important performance metric to be considered. A real-time job has a *deadline* before which it is available for service and after which it must leave the system. (This is the property of firm real-time (FRT) systems (Bernat et al., 2001) which are considered in this paper, while in soft real-time (SRT) systems; a late job which has missed its deadline continues to get service until completion.) Two models of job behavior are usually

*Corresponding Author.

considered: *deadlines until the beginning of service* (DBS) and *deadlines until the end of service* (DES). In the former model, a job keeps its deadline only until the beginning of service. Accordingly, jobs remain in the system while being served until they complete their service requirements. In the latter model, a job retains its deadline until the end of service. Accordingly, jobs may discontinue their service because they have missed their deadlines. For the class of real-time jobs, the loss probability, which is the fraction of jobs missing their deadlines, is an important performance measure. On the other hand, the interdependency among the traffics of different classes may affect the performance of non real-time as well as the real-time demands and is central to both the design and analysis of such systems. For the class of non real-time jobs, some performance measures such as the average *sojourn time* (the interval of time between the arrival and departure of a job) and *waiting time* (the interval of time between the arrival of a job and the instant of starting service of that job) are of high importance. Beside the respective priority of the classes, the scheduling policy within each class of jobs which assigns priorities to the jobs in the same class and constitutes the scheduling decisions also strongly influences the overall performance of the system. The scheduling policies can be classified into two broad categories: *preemptive* and *non-preemptive*. In preemptive scheduling, processing of the currently running job can be interrupted by a higher priority job, whereas in non-preemptive scheduling, an arriving job can be scheduled only when the running job leaves the system. Though preemptive scheduling can guarantee better system utilization and is usually more desirable, there are many scenarios where the properties of some hardware or software devices make preemption either impossible or prohibitively expensive. For example, in high speed packet switching networks, preemption requires the retransmission of the preempted packet. Scheduling over a shared media such as LAN, WLAN and field buses (EN 50170, 1996) such as CAN bus (CAN-CIA, 1992; Livani and Kaiser, 1998) is inherently non-preemptive, because each node in the network has to ensure that the shared channel is free before it can begin transmission. Besides its extensive use in communication systems, non-preemptive processor scheduling is also used in light-weight multi-tasking kernels and is beneficial in multimedia applications (Dolev and Keizelman, 1999). Non-preemptive scheduling for real-time embedded systems has also some benefits such as the ease of implementation, reduced run-time overhead, and guaranteeing exclusive access to shared resources and data which eliminates both the need for synchronization and its associated overhead.

According to the above discussion, the scheduling policy used within each class of the jobs in a real-time system strongly influences the performance of the system. Among such policies, the earliest-deadline-first (EDF) policy (Liu and Layland, 1973), which schedules the jobs in the ascending order of their deadlines, is known to be an optimal scheduling policy within the class of non-idling service time independent scheduling policies (George et al.,1995 and 1996) and also stochastically minimizes the fraction of lost jobs in the same class of policies (Towsley and Panwar, 1990 and 1992).

In a more general view, most of the applications in current computing and communicating systems have more than one class of traffic and the real-time demands in such systems make an important portion of the multi-class traffics. As an example for such a system, consider a DiffServ supported network, which offers service differentiation for different classes of flows at each network node (May et al., 1999). In such a system, the traffic is categorized into different classes at the ingress edge nodes (which is implemented by priority queues). As an example for the applications in such a system, we can assume voice/video messages which are useless unless they are transmitted before their deadlines, and data messages which should be transmitted with no limitation on their sojourn times (no loss). To analyze the performance of such systems with multiple classes of jobs, the interaction of the jobs in the classes should be taken into account. More critical examples on the matter can also be found in the applications of wireless sensor networks.

In this paper, we present an approximation method for the performance analysis of a non-preemptive two-class M/M/1 system. The class-1 jobs with the higher priority are real-time and have exponentially distributed *relative deadlines* (where the relative deadline is the interval of time between the arrival of a job and its deadline). On the other hand, the lower priority class-2 jobs are non real-time. The model of exponential relative deadlines is more suitable for approximating the properties of applications with unpredictable input patterns, which are most common in intermediate nodes in wireless sensor networks or intermediate routers in high-speed packet switching networks as well as military and avionics-related systems. The class-1 jobs in the system have DES and are served according to EDF, while the class-2 jobs are served according to FCFS. Due to the optimality of the EDF policy, the performance analysis of the two-class system with this policy for scheduling of the class-1 jobs can be very important. In this paper, we assume that no service, either real-time or non real-time, can be preempted and the service discipline is totally non-preemptive. The proposed approximation method to analyze the two-class M/M/1 system uses a key parameter, namely, the rate of missing deadlines (loss rate), which primarily depends on the number of class-1 jobs in the system. This important parameter is estimated using an upper bound and a lower bound for the case of non-preemptive EDF with DBS. The resulting formulation is then generalized to the case of non-preemptive EDF with DES for both single-class and two-class systems using some heuristics. Such results are finally used in a Markov chain model of the two-class M/M/1 system. To the best of our knowledge, no other analytical or approximation method exists for this problem. Comparison of the analytical and simulation results shows that the presented method is relatively accurate.

The rest of this paper is organized as follows. Section 2 presents some related works. Section 3 describes the basic system model and the proposed analytical method for modeling the system and extracting the required performance measures. This is followed in Section 4 by explaining our method of estimating the loss rate of the class-1 jobs for non-preemptive model of the EDF scheduling policy as well as the same parameter of the non-preemptive two-class system. Section 5 provides some numerical examples and the comparison of the analytical and simulation results. Summary, concluding remarks, and future works are finally presented in Section 6.

2. RELATED WORK

The performance analysis of systems with a single class of real-time jobs was well investigated for the FCFS scheduling policy in several studies such as (Palm,1953; Barrer, 1957; Daley, 1965; Baccelli et al., 1984; Zhao and Stankovic, 1989; Boxma and Wall, 1994; Movaghar, 1998, 2006; Brandt and Brandt, 1999a; Brandt and Brandt, 2002) and the references therein. However, in spite of the optimality of both preemptive and non-preemptive EDF policies (Towsley and Panwar, 1990 and 1992) there exist relatively few papers on the probabilistic analysis of EDF. This may be due to the complexity of such analysis. Some of the works done in this area such as (Leulseged and Nissanke, 2004; Nissanke et al., 2002) have concentrated on the probabilistic analysis of EDF for periodic task arrivals. For non-periodic arrivals, Hong et al. (Hong et al., 1989) first introduced upper and lower bounds for the performance of an M/M/*m*/EDF+M queue in a FRT system, where the last M specifies that the distribution of the relative deadlines is exponential. The accuracy of their approximation method is very good for small values of relative input rates as well as for small mean relative deadlines of jobs with DBS. The results presented in (Hong et al., 1989) are only for relative input rates up to 1.2. It is mentioned that the accuracy of the method may decrease for higher relative input rates and also for preemptive EDF with DES. These results were later improved in (Kargahi and Movaghar, 2004) and also extended to M/M/*m*/EDF+G queues in (Kargahi and Movaghar, 2006) (where m=1 for DES has been assumed in all these three studies). Moreover, an approximation method for the analysis of an M/M/1/EDF+M queue in the case of non-preemptive EDF scheduling of jobs with DES has been

presented in (Kargahi and Movaghar, 2005). This latter work has also extended to multi-server queues in (Kargahi and Movaghar, 2007). On the other hand, Lehoczky and his colleagues in (Lehoczky,1996a and1996b; Doytchinov et al., 2001) have developed an approximation method to compute the fraction of late tasks in a SRT system for the *heavy traffic* case (where the traffic intensity converges to 1 and the system has high average utilization). In their model, it is assumed that all jobs are processed to completion. The method is called real-time queueing theory (RTQT), which is an extension of the traditional queueing theory where it takes the timing requirements of tasks into account, and its performance metric is the fraction of the offered load that completes within its deadline. RTQT was first introduced by Lehoczky (1996a) for M/M/1 queues with the EDF scheduling policy. The single queue case was also put on a firm mathematical foundation in the paper by Doytchinov et al. (2001) for GI/G/1 queues. It should be noted that the EDF scheduling policy considered in (Hong et al., 1989; Kargahi and Movaghar, 2004 and 2005), and also in the current paper, differs from the one analyzed in (Lehoczky,1996a and 1996b; Doytchinov et al., 2001). This is due to the fact that unlike the latter works, the former works never schedule jobs which are already past their deadlines (due to the FRT nature of the system). Furthermore, the latter works have only focused on the heavy traffic intensities.

The above studies have been for systems with a single class of jobs. A number of references have investigated some systems with priority queues (Miller,1960; Jaiswal, 1968; Brandt and Brandt, 1999b; Choi et al, 2001; Kruk et al, 2003; Brandt and Brandt, 2004; Kargahi and Movaghar, 2007). There also exist few papers in the literature on the priority queues with some classes of real-time jobs. Brandt and Brandt (1999b) first considered a two-queue priority system with multi-servers, where the real-time jobs in the first queue (the class-1 jobs) have priority over the non real-time jobs in the second queue (the class-2 jobs) and also have generally distributed relative deadlines until the beginning of service. Some approximations for the performance of the class-2 jobs in the system are presented in there. Such results are later improved in an exact form for a two-class single server queue in (Brandt and Brandt, 2004). Similar results for deterministic relative deadlines and both cases of DBS and DES are presented in (Choi et al, 2001). In all of these studies, the scheduling policy of the class-1 jobs is considered to be FCFS in a FRT system. For the EDF scheduling of multi-class traffics, Kruk, et al. in (Kruk et al, 2003) first used RTQT (Lehoczky,1996a and1996b; Doytchinov et al., 2001) for the analysis of a SRT system of K input streams (each with the EDF or FCFS policy) with a shared processor across the streams. RTQT has also been extended in (Kruk et al, 2004) to the case of open queueing networks with multiple independent traffic flows, each with the EDF policy. Both of these latter works also assume that due to the SRT nature of the system, all jobs are processed to completion (even if they are late). Likewise, they model the system only for the heavy traffic intensities. Whereas, the work presented in this paper considers FRT systems and never schedules the real-time (class-1) jobs which are already past their deadlines, and also covers almost all of the input rates with which the system still remains stable.

3. SYSTEM MODEL AND SOLUTION

This section initially describes the general system model, and then solves it with respect to some performance measures, namely the loss probability of real-time (class-1) jobs and the average sojourn (waiting) time of non real-time (class-2) jobs.

3.1. System Model

We consider a two-class M/M/1 system, i.e., a single server with an infinite-capacity queue. Two Poisson streams (classes) of jobs with positive intensity λ_i, $i \in \{1,2\}$, arrive to the system, which require exponential service times with mean $1/\mu_i$, $i \in \{1,2\}$, respectively. The class-1 jobs are served

with a non-preemptive priority discipline over the class-2 jobs. More precisely, if a class-1 job arrives before the service completion of a class-2 job, (in spite of their respective priority) the service will not be interrupted and will continue to completion. Afterwards, the remaining class-2 jobs (if any) can only be served from the time instant that no real-time job is in the system. Furthermore, a relative deadline is associated with each class-1 job. We assume that the relative deadlines are random variables of an exponential distribution with rate ν (i.e., $\theta=1/\nu$ is the mean value of relative deadlines). Since deadlines are until the end of service, a job is thrown away if it cannot complete execution before its deadline. This can occur while the job waits in the queue or while it is in service. If the job is waiting in the queue at the time when the deadline is reached, the job is thrown out. If the job is in service at the time that the job reaches its deadline, it is aborted and then thrown out. In either case, the job that is thrown away is considered *lost*. The class-1 and class-2 jobs are served according to the earliest-deadline-first (EDF) and first-come-first-served (FCFS) scheduling policies, respectively. As specified in the definition of the EDF policy, the job closest to its deadline is to be served. Since the service desciplines are non-preemptive, no job can preempt the serving job. It is proved in (Towsley and Panwar, 1990 and 1992) that the EDF scheduling policy stochastically maximizes the fraction of jobs meeting their deadlines for DES within the class of non-idling service time independent non-preemptive scheduling policies.

According to the relation between the two classes of jobs, the behavior and performance of both class-1 and class-2 jobs are totally affected by each other. In order to model the system, the state of the two-class system is represented by $\mathbf{n}{:}i{=}(n_1,n_2){:}i$, where n_1 is the current number of class-1 jobs and n_2 is the number of existing class-2 jobs in there. Moreover, i shows that a class-i job, i=1, 2, is running when the system is in the state.

The approach presented in this paper is based on using a state-dependent loss rate function γ_n for class-1 jobs to be defined below. (Currently, the formulations related to γ_n are presented for a system only of class-1 jobs which will be adapted later in a proper manner for the two-class system.) Let $\boldsymbol{N}$ be the set of natural numbers and $\boldsymbol{R}^+$ the set of positive real numbers. For $t, \varepsilon \in \boldsymbol{R}^+$ and $n \in \boldsymbol{N}$, let

$\Psi_n(t,\varepsilon)$ ≡ the probability that a class-1 job misses its deadline during $[t,t+\varepsilon)$, given there are $n>0$ real-time (class-1) jobs in the system at time t and one of them is serving.

Define

$$\gamma_n(t) = \lim_{\varepsilon \to 0} \frac{\Psi_n(t,\varepsilon)}{\varepsilon} \tag{1}$$

Assuming statistical equilibrium, let

$$\gamma_n = \lim_{t \to \infty} \gamma_n(t) \tag{2}$$

γ_n is the (steady-state) rate of missing deadlines when there are n class-1 jobs in the system (including the real-time job being served). Accordingly, we define the loss rate of the system at state $\mathbf{n}{:}1{=}(n_1,n_2){:}1$ as $\gamma_{\mathbf{n}:1} = \gamma_{n_1}$. Furthermore, since the system is fully non-preemptive, the loss rate

function of the system in state **n:**2=(n_1,n_2):2, namely, $\gamma_{\mathbf{n}:2}$ will be different of γ_{n_1}, which will be defined later in Section 4.

Barrer (1957) was the first to introduce the idea of γ_n for deterministic relative deadlines of real-time jobs in a single-class system. The idea was extended in (Movaghar, 1998 and 2006; Brandt and Brandt, 2002) to a larger class of models when relative deadlines have a general distribution and jobs arrive according to a *state-dependent* Poisson process. These latter results assume the FCFS policy, and show that γ_n is independent of the input rate and only depends on the number of jobs in the system. In (Movaghar, 1998; Brandt and Brandt, 2002), the description of how to calculate γ_n for DBS is given. The calculation of γ_n for the case of DES is presented in (Movaghar, 2006). Moreover, a method for estimating γ_n of an M/M/*m*/EDF+G system with non-preemptive services for DBS and preemptive services for DES (with *m*=1) is presented in (Kargahi and Movaghar, 2006), and also a method for estimating that of an M/M/1/EDF+M queue with non-preemptive services for DES is proposed in (Kargahi and Movaghar, 2005). This latter method is also extended to multi-server non-preemptive queues in (Kargahi and Movaghar, 2007).

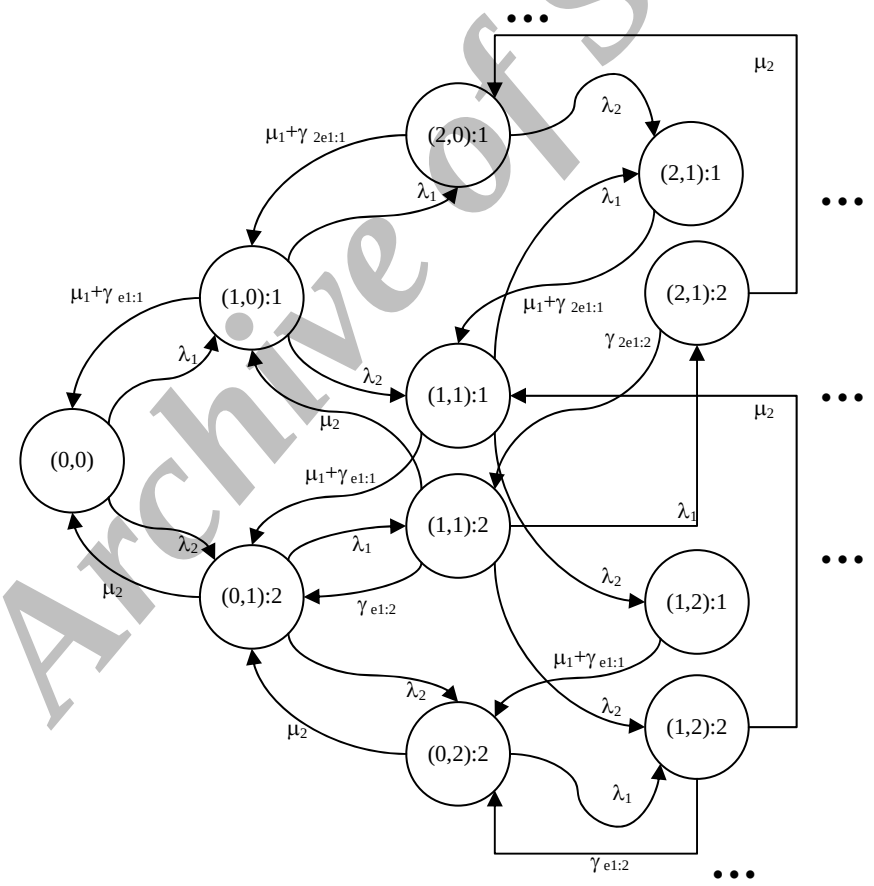


Figure 1. Partial state-transition-rate diagram for Markov chain **M**.

Considering the above notations, the resulting Markov chain model of the two-class system, **M**, may partially be shown as in Figure 1. Let e_1 = (1,0), e_2 = (0,1), and **0** = (0,0). Moreover, let **n**=(n_1,n_2)≥**n'**=(n'_1,n'_2) if and only if $n_1 \geq n'_1$ and $n_2 \geq n'_2$. Assuming that the system is in state

n:1=(n_1,n_2):1, the state of the system can be changed to **n**+e_1**:**1=(n_1+1,n_2):1 or **n**+e_2**:**1=(n_1,n_2+1):1 with rates λ_1 or λ_2, respectively. When the system is in state **n:**1, the state can be changed to **n**-e_1**:1,** if **n**>e_1, to **n**-e_1**:**2**,** if **n**≥e_1+e_2 but not **n**>e_1, or to **0** if **n**=e_1, because of either completing the service requirements of a class-1 job (with rate μ_1) or missing a real-time job's deadline (with rate $\gamma_{\mathbf{n}:1}$). On the other hand, when the system is in state **n:**2=(n_1, n_2):2, the state of the system can be changed to **n**+e_1**:**2=(n_1+1,n_2):2 or **n**+e_2**:**2=(n_1,n_2+1):2 with rates λ_1 or λ_2, respectively. Moreover, when the system is in state **n:**2, the state can be changed to **n**-e_2**:1,** if **n**≥e_1, to **n**-e_2**:2,** if **n**>e_2 but not **n**≥e_1, or to **0** if **n**=e_2, due to completing the service requirements of the class-2 job (with rate μ_2) or to **n**-e_1**:**2, if **n**≥e_1, due to missing a real-time job's deadline (with rate $\gamma_{\mathbf{n}:2}$).

3.2. Model Solution

In the following, the required equations for solving the system model **M** are presented and the equilibrium state probabilities will be obtained. Using such information, the target performance measures, namely the loss probability of the class-1 jobs and the average sojourn (waiting) time of class-2 jobs will be calculated. Let

$p(\mathbf{n}:i)$ ≡ the (steady-state) probability that the system is

in state n=(n1,n2) and serving a class-i job. (3)

The balance equations for the system, in equilibrium, can be written as (we assume $p(\mathbf{0}) = p(\mathbf{0}:1) = p(\mathbf{0}:2)$ in the notation):

$$(\lambda_1+\lambda_2)\,p(\mathbf{0}) = (\mu_1+\gamma_{e_1:1})\,p(e_1:1) + \mu_2 p(e_2:2)$$

$$\begin{aligned}(\lambda_1+\lambda_2+\gamma_{\mathbf{n}})\,p(\mathbf{n}:1) = {} & (\mu_1+\gamma_{\mathbf{n}+e_1:1})\,p(\mathbf{n}+e_1:1) + \mu_2 p(\mathbf{n}+e_2:2) + \lambda_1 p(\mathbf{n}-e_1:1), \quad \text{if } \mathbf{n}\geq \mathrm{e}_1 \\ & + \lambda_2 p(\mathbf{n}-e_2:1), \quad \text{if } \mathbf{n}\geq \mathrm{e}_2\end{aligned} \tag{4}$$

$$\begin{aligned}(\lambda_1+\lambda_2+\gamma_{\mathbf{n}:2})\,p(\mathbf{n}:2) = {} & \gamma_{\mathbf{n}+e_1:2}\,p(\mathbf{n}+e_1:2) + \lambda_2 p(\mathbf{n}-e_2:2), && \text{if } \mathbf{n}\geq \mathrm{e}_2 \\ & + \lambda_1 p(\mathbf{n}-e_1:2), && \text{if } \mathbf{n}\geq \mathrm{e}_1 \\ & + (\mu_1+\gamma_{e_1:1})\,p(\mathbf{n}+e_1:1) + \mu_2 p(\mathbf{n}+e_2:2), && \text{if } \textit{not } \mathbf{n}\geq \mathrm{e}_1\end{aligned}$$

The normalizing condition is also as follows:

$$\sum_{i=1}^{2}\sum_{n_1=0}^{\infty}\sum_{n_2=0}^{\infty} p(\mathbf{n}=(n_1,n_2):i) = 1 \tag{5}$$

Solving the equilibrium in (4) and using (5), we find the state probabilities of the system, namely, $p(\mathbf{n}=(n_1, n_2) : i)$, $n_1, n_2 \in \{0, 1, 2, \ldots\}$, i=1, 2.

Assuming a stable system (the stability conditions are presented later in this section), the desired performance measures can be calculated as follows. The loss probability of class-1 jobs in the system may be obtained as

$$\alpha_d = \frac{\sum_{n_1=1}^{\infty}\sum_{n_2=0}^{\infty} p(\mathbf{n}=(n_1,n_2):1)\gamma_{\mathbf{n}:1} + \sum_{n_1=1}^{\infty}\sum_{n_2=1}^{\infty} p(\mathbf{n}=(n_1,n_2):2)\gamma_{\mathbf{n}:2}}{\lambda_1} \tag{6}$$

which is the average rate of missing deadlines divided by the average rate of class-1 job arrivals. Whereas for the class-1 jobs, identifying the loss probability is quite valuable, for the class-2 jobs, the average sojourn (waiting) time is of high importance. Assume that $\overline{N_i}$, $i \in \{1,2\}$, is the average number of class-*i* jobs in the system. Then, we have

$$\overline{N_i} = \sum_{k=1}^{2}\sum_{n_1=0}^{\infty}\sum_{n_2=0}^{\infty} n_i\, p(\mathbf{n}=(n_1,n_2):k) \tag{7}$$

Using Little's formula, we obtain

$$V_i = \bar{N}_i / \lambda_i \tag{8}$$

where V_i is the average sojourn time of class-*i* jobs. The average waiting time of the class-2 jobs can also be derived as

$$W_2 = V_2 - \frac{1}{\mu_2} \tag{9}$$

Due to the fact that the stability of the two-class system should be preserved, we try to find an estimation of the maximum permitted input rate of class-2 jobs, above which the system becomes unstable. This will be done by putting the system into a saturated condition with respect to class-2 jobs and finding the desired maximum input rate using some heuristics. Figure 2 shows a simplified presentation of the state transition-rate diagram of Markov chain **M** when it is in a saturated condition. The idea for such simplification is as follows. Since the system is assumed to be saturated (the processor is fully utilized), whenever there is no class-1 job in the system and one arrives, it should wait for the service completion of the serving class-2 job. This behavior can be observed in transitions among the states of type **n**:2 in Figure 2. When the service is completed, the system behaves normally with respect to class-1 jobs (see the transitions among states **n**:1 in Figure 2). Due to the assumption of saturation, as soon as the system becomes empty of class-1 jobs, there would certainly be further class-2 job waiting, which will start its service immediately. Note that the number of waiting class-2 jobs in the system is not of high importance in this part of our study. Rather, it should be considered that whenever the system becomes empty of class-1 jobs, there is always at least one waiting class-2 job in the system. Let

$$p_{u_1} = \sum_{n_1=1}^{\infty} p'(\mathbf{n}=(n_1,0):1) \tag{10}$$

be the fraction of time that the processor is utilized by the class-1 jobs, where $p'(\mathbf{n}:i)$ is the (steady-state) probability of being in state $\mathbf{n}:i$ in the simplified version of Markov chain **M** (Figure 2). $p_{u_2} = \lambda_2 / \mu_2$ is also the fraction of time that the processor can be utilized by the class-2 jobs. The system is stable if and only if $p_{u_1} + p_{u_2} < 1$, or equivalently

$$\lambda_2 < (1 - p_{u_1})\mu_2 \tag{11}$$

where p_{u_1} can simply be calculated using (10) after solving the simplified Markov chain using the standard Markovian solution techniques.

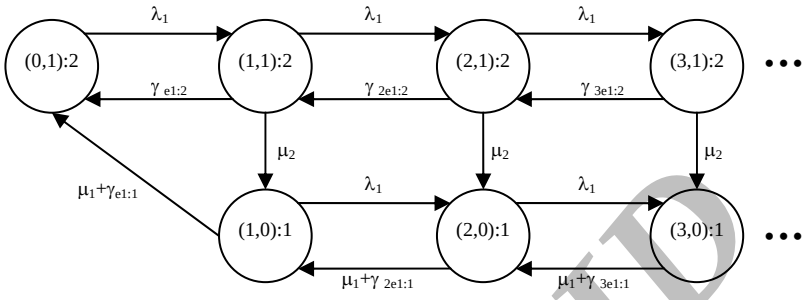


Figure 2. Partial simplified state-transition-rate diagram for Markov chain **M** in a saturated condition.

To analyze the system with the EDF policy for the class-1 jobs, we need to have formulations of $\gamma_{\mathbf{n}:1}$ and $\gamma_{\mathbf{n}:2}$ (for EDF) as defined in (2) and the following paragraph. Next, we review a method for estimating $\gamma_{\mathbf{n}:1}$ and $\gamma_{\mathbf{n}:2}$ (of EDF) for an infinite-capacity system.

4. DETERMINATION OF LOSS RATES

In this section, we present methods for estimating $\gamma_{\mathbf{n}:1}^{\text{EDF}} = \gamma_{n_1}^{\text{EDF}}$ in the cases of DBS and DES. The former case will be used in the estimation of the latter case as well as $\gamma_{\mathbf{n}:2}^{\text{EDF}}$. First, we will have an overview on a method to estimate $\gamma_{n_1}^{\text{EDF}}$ for the case of DBS, namely, $\gamma_{n_1}^{\text{EDF-DBS}}$. To do so, some bounds for $\gamma_{n_1}^{\text{EDF}}$ will be defined. Combining the bounds will result in an estimation of the required parameter for DBS. Then, we will use some ideas to present a method for estimating the same parameter for DES, namely, $\gamma_{n_1}^{\text{EDF-DES}}$. The resulting formulation for the case of DBS besides a different view to the system will be used to estimate the required parameter for DES. Both of these estimations are also used together to estimate $\gamma_{\mathbf{n}:2}^{\text{EDF}}$.

As indicated in (Movaghar, 2006), for a specified mean relative deadline (θ) in a FCFS system, deterministic relative deadlines generate the minimum loss probability among all distributions of relative deadlines. Accordingly, we assume that such a property is also valid for the EDF scheduling algorithm. Since for deterministic relative deadlines, EDF is the same as FCFS, we can assume the loss probability of FCFS scheduling algorithm for deterministic relative deadlines as the lower bound of the loss probability of EDF scheduling algorithm for exponentially distributed relative deadlines. On the other hand, since EDF is an optimal scheduling algorithm for both deadline models (see Towsley and Panwar, 1990 and 1992), it can minimize the loss probability among all other scheduling algorithms, especially FCFS. Therefore, we will have

$$\alpha_d^{\text{FCFS-det}} \le \alpha_d^{\text{EDF-exp}} \le \alpha_d^{\text{FCFS-exp}} \tag{12}$$

where $\alpha_d^{\text{FCFS-det}}$ and $\alpha_d^{\text{FCFS-exp}}$ represent the loss probabilities of the system with deterministic and exponential relative deadlines for the FCFS scheduling algorithm, respectively. We also assume that such ordering is valid for loss rates in the FCFS and EDF scheduling algorithms. Such validity is strongly confirmed by simulation results presented in part in (Kargahi and Movaghar, 2006). Therefore, we will have

$$\gamma_{n_1}^{\text{FCFS-det}} \le \gamma_{n_1}^{\text{EDF-exp}} \le \gamma_{n_1}^{\text{FCFS-exp}} \tag{13}$$

where the functions describing the above two bounds of $\gamma_{n_1}^{\text{EDF-exp}}$ are given in (Movaghar, 1998) for a multi-sever system with DBS and in (Movaghar, 2006) for a single-server system with DES.

The above two bounds are linearly combined using a multiplier to obtain an appropriate estimation of $\gamma_{n_1}^{\text{EDF-DBS}}$. More explanation of this approach for an infinite-capacity queue and the DBS model is given in the following section. Consequently, such estimation will be used in a different manner to estimate the loss rates of a system with non-preemptive EDF scheduling algorithm and the DES model, namely, $\gamma_{n_1}^{\text{EDF-DES}}$. Afterwards, the solution will be extended to the two-class system defined in the previous section by finding an estimation for $\gamma_{\mathbf{n}:2}^{\text{EDF}}$.

4.1. Non-preemptive EDF with DBS

In this section, we propose a multiplier to linearly combine the two bounds indicated above in the case of DBS to estimate $\gamma_{n_1}^{\text{EDF-DBS}}$.

As defined in (Kargahi and Movaghar, 2006), contrary to the FCFS scheduling algorithm, the simulation results strongly indicate that the state-dependent loss rates depend on λ_1 for the EDF scheduling algorithm. Accordingly, advantages of some properties of EDF and some simulation results can be used to make a multiplier which linearly combines the bounds defined previously. The multiplier must be adjusted to a function of λ_1 to get a more accurate estimation of $\gamma_{n_1}^{\text{EDF-DBS}}$. The multiplier, $\xi_{\text{DBS}}(.)$, combines the bounds as follows:

$$\gamma_{n_1}^{\text{EDF-DBS}} = \frac{\left(\xi_{\text{DBS}}(.)\gamma_{n_1}^{\text{FCFS-exp-DBS}} + \gamma_{n_1}^{\text{FCFS-det-DBS}}\right)}{\xi_{\text{DBS}}(.)+1} \tag{14}$$

where $\xi_{\text{DBS}}(.)$, which defines the effective ratio of each of the bounds on $\gamma_{n_1}^{\text{EDF-DBS}}$, is to be specified.

As discussed previously, it has been shown that for the FCFS scheduling algorithm, the loss rate is independent of λ_1 (see Movaghar, 1998 and 2006; Brandt and Brandt, 2002). Therefore, such parameters can be calculated as

$$\gamma_{n_1}^{\text{FCFS-exp-DBS}} = \begin{cases} 0, & n_1 \le 1 \\ \dfrac{n_1 - 1}{\theta}, & n_1 > 1 \end{cases} \tag{15}$$

and

$$\gamma_{n_1}^{\text{FCFS-det-DBS}} = \begin{cases} 0, & n_1 \le 1 \\ \mu_1 \left(\dfrac{F_{E_{n_1-2}}(\theta)}{F_{E_{n_1-1}}(\theta)} - 1 \right) & n_1 > 1 \end{cases} \tag{16}$$

where

$$F_{E_n}(\theta) = 1 - e^{-\mu_1 \theta} \sum_{i=0}^{n-1} \frac{(\mu_1 \theta)^i}{i!} \tag{17}$$

for exponential and deterministic relative deadlines until the beginning of service, respectively, which are obtained from (Movaghar, 1998). As defined in (Kargahi and Movaghar, 2006), $\xi_{\text{DBS}}(.)$ is a function of three parameters, namely, n_1, $\rho_1 = \lambda_1/\mu_1$, and $\mu_1\theta$ for exponential relative deadlines, where n_1 is the number of waiting class-1 jobs in the queue, ρ_1 is the normalized arrival rate (normalized λ_1 with respect to μ_1), and $\mu_1\theta$ is the normalized mean relative deadline with respect to the mean service time $1/\mu_1$. The function describing the behavior of $\xi_{\text{DBS}}(.)$ with respect to the above three parameters, i.e., $\xi_{\text{DBS}}(n_1, \rho_1, \mu_1\theta)$, is as follows (obtained from Kargahi and Movaghar, 2006):

$$\xi_{\text{DBS}}(n_1, \rho_1, \mu_1\theta) = \frac{6.7}{n_1 \rho_1^{1.25} \sqrt{\mu_1\theta}} \tag{18}$$

Substituting $\xi_{\text{DBS}}(.)$ above in (14), we can find $\gamma_{n_1}^{\text{EDF-DBS}}$.

The way that $\xi_{\text{DBS}}(.)$ depends on the normalized arrival rate (ρ_1) can be explained by some properties of EDF. Due to the dynamics of the EDF scheduling algorithm with respect to different values of ρ_1, for very small values of ρ_1 where $\rho_1 \to 0$, $\gamma_{n_1}^{\text{EDF-DBS}}$ converges to $\gamma_{n_1}^{\text{FCFS-exp-DBS}}$; therefore, $\xi_{\text{DBS}}(.)$ tends to be very large as $\xi_{\text{DBS}}(.) \to +\infty$. The reason is that for very light traffic intensities (where the average population is very low), EDF behaves very similar to FCFS and the improvements of EDF over FCFS are quite limited. On the other hand, for large values of ρ_1, the behavior of EDF becomes more similar to that of FCFS with deterministic relative deadlines, where $\gamma_{n_1}^{\text{EDF-DBS}}$ converges to the lower bound; therefore, $\xi_{\text{DBS}}(.)$ tends to be very small as $\xi_{\text{DBS}}(.) \to 0$. Next, we use the recent formulations of $\gamma_{n_1}^{\text{EDF-DBS}}$, as obtained from (18) and (14), to estimate $\gamma_{n_1}^{\text{EDF-DES}}$ for non-preemptive EDF scheduling algorithm.

4.2. Non-preemptive EDF with DES

In spite of the DBS model, even the serving jobs may miss their deadlines in the DES model. However, although the deadlines are until the end of service, due to the non-preemptive nature of the scheduling algorithm, even if the deadline of an arriving job is earlier than that of the job in the server, the serving job will not be preempted and continues to get service. It has been proven in

(Towsley and Panwar, 1990 and 1992) that the non-preemptive EDF scheduling algorithm also stochastically minimizes the fraction of lost jobs in the class of non-idling service time independent non-preemptive scheduling algorithms. In spite of its optimality, to the best of our knowledge, other than the approximation method proposed in (Kargahi and Movaghar, 2005) for a single-server queue, which is also extended in (Kargahi and Movaghar, 2007) for a multi-server queue by the same authors, no other analytical or approximation method for the probabilistic analysis of this algorithm exists. In the following, we present a method for estimating $\gamma_{n_1}^{\text{EDF}}$ for non-preemptive EDF with DES model, namely $\gamma_{n_1}^{\text{EDF-DES}}$, which results in approximating the performance of non-preemptive EDF scheduling algorithm. To do so, we propose another view to the system as in the following paragraphs.

The main idea of the proposing estimation method is to break the system into two subsystems. Afterwards, two loss rates will be calculated for the subsystems, which adding them together will result in an estimation of the desired parameter, namely, $\gamma_{n_1}^{\text{EDF-DES}}$.

Due to the fact that the serving job is non-preemptive, after starting service, the behavior of the system with respect to this job is similar to that of a system with a stand-alone server (no waiting rooms) and the FCFS scheduling algorithm. On the other hand, if the number of available class-1 jobs in the system (n_1) is greater than two, the remaining n_1-1 job(s) in the system follow the EDF scheduling algorithm. Therefore, the system can be broken into two subsystems (see Figure 3): the first one (Subsystem-1) containing the non-preemptive server with rate μ_1, which can be considered as a FCFS queue with capacity 1 (no waiting room), and the second one (Subsystem-2) which can virtually be assumed as a non-preemptive EDF queue with DBS and a servers with a virtual service rate μ_1', to be determined.

First, we study Subsystem-1. Since Subsystem-1 can be assumed as a FCFS queue (with capacity 1), the loss rate of this subsystem will be simply $\gamma_1' = \gamma_1^{\text{FCFS-exp-DES}}$, where we have $\gamma_1^{\text{FCFS-exp-DES}} = \gamma_2^{\text{FCFS-exp-DBS}}$ as can be found in (Movaghar, 2006).

Second, we consider Subsystem-2. As defined previously, we can assume of this subsystem as a virtual queue with the DBS model. According to such view to Subsystem-2 and due to the fact that the loss rate in the server is taken into account in Subsystem-1, it can virtually be assumed that no jobs of Subsystem-2 will miss their deadlines after starting service. Now, we use the method presented in Section 4.1 to calculate the loss rate of Subsystem-2. First, the lower and upper bounds should be specified (since this subsystem is assumed as a virtual system with DBS, we use the respective bounds for the required calculations). As indicated previously, deterministic relative deadlines construct the lower bound and exponentially distributed relative deadlines construct the upper bound. Since the serving job leaves the server due to service completion or deadline miss, the virtual service rate of the server for the lower bound can be assumed as $\mu_{\text{L}} = \mu_1 + \gamma_1^{\text{FCFS-det-DES}}$, where we have $\gamma_1^{\text{FCFS-det-DES}} = \gamma_2^{\text{FCFS-det-DBS}}$ as can be found in (Movaghar, 2006). Similarly, the virtual service rate of the servers for the upper bound can be assumed as $\mu_{\text{U}} = \mu_1 + \gamma_1^{\text{FCFS-exp-DES}}$. Due to the fact that $\gamma_{n_1}^{\text{FCFS-exp-DBS}}$ (and therefore $\gamma_1^{\text{FCFS-exp-DES}}$) is independent of the service rate, μ_{U} does not affect the upper bound. Substituting μ_{L} for μ_1 in (16) and (17), we obtain the lower bound. Moreover, (15) simply gives the upper bound. On the other hand, since the distribution of relative deadlines is exponential for the EDF scheduling algorithm, the virtual service rate of the server of

Subsystem-2 can also be assumed as $\mu_1' = \mu_1 + \gamma_1^{\text{FCFS-exp-DES}}$. Accordingly, substituting μ_1' for μ_1 and $\rho_1' = \lambda_1 / \mu_1'$ for ρ_1 in (18), and then using (14) with the bounds specified above, we obtain the loss rate for Subsystem-2, namely, γ_{n_1}''. Consequently, we have

$$\gamma_{n_1}^{\text{EDF-DES}} = \begin{cases} \gamma_1', & \text{if } n_1 \le 1 \\ \gamma_1' + \gamma_{n_1}'', & \text{if } n_1 > 1 \end{cases} \tag{19}$$

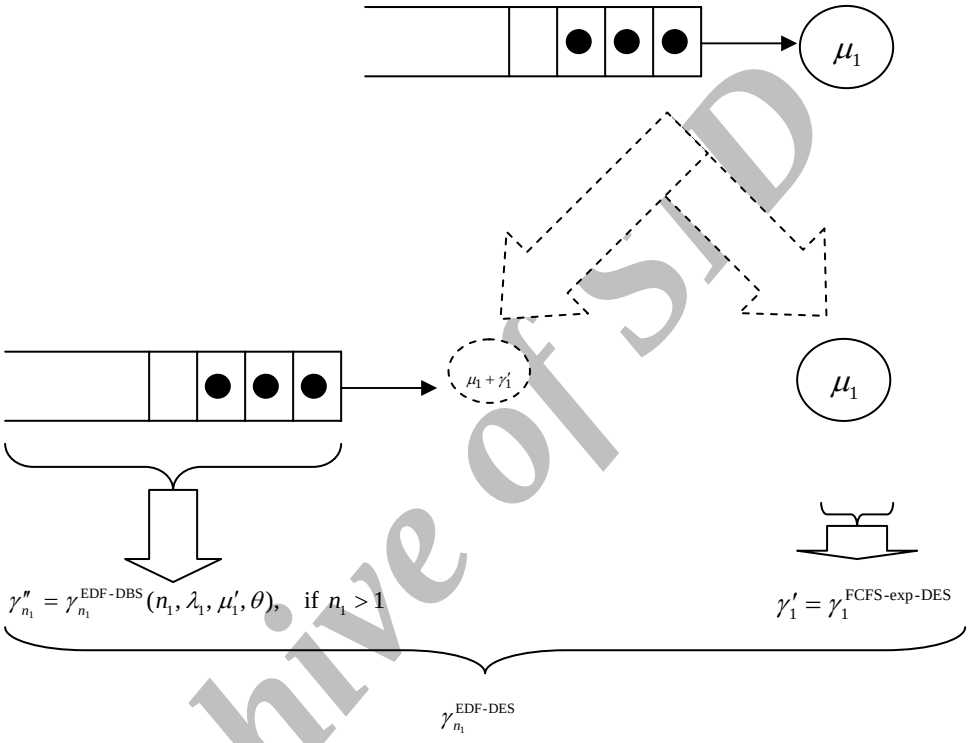


Figure 3. The modified view to the system with non-preemptive EDF and the model of DES.

Whenever a class-2 job is in service, the loss rate will be different and is shown as $\gamma_{\mathbf{n}:2}$. In such conditions, we assume that the system works with the model of DBS (because the serving job has no deadline and the waiting jobs in such conditions can miss their deadlines before starting service) where the instantaneous service rate in the respective states is μ_2. Accordingly, substituting $\mu_1'' = \mu_2$ for μ_1 and $\rho_1'' = \lambda_1 / \mu_1''$ for ρ_1 in (18), and then using (14) with the bounds specified above (note that the respective bounds should also be calculated by substituting μ_1'' for μ_1), we obtain the loss rate mentioned above, namely, $\gamma_{\mathbf{n}:2}$ as (assuming $n_1 \ge 1$):

$$\gamma_{\mathbf{n}:2} = \gamma_{n_1+1}^{\text{EDF-DBS}}(\mu_1'') \tag{20}$$

using $\xi_{\text{DBS}}(n_1 + 1, \rho_1'', \mu_1''\theta)$ as in (18) and the respective bounds with the mentioned parameters.

Substituting $\gamma_{n_1}^{\text{EDF-DES}}$ above for $\gamma_{\mathbf{n}:1}$ and using $\gamma_{\mathbf{n}:2}$ above in **M** and then solving the resulting Markov chain using the method presented in Section 3.1, we find the desired performance measures for the system with the non-preemptive service discipline.

5. NUMERICAL EXAMPLES

In this section, we study examples to verify the presented ideas and to illustrate the accuracy of the proposed approximation method. We consider the system for two configurations: one with purely class-1 jobs, denoted as SYS1, and another with both class-1 and class-2 jobs, referred to as SYS2. The examples for SYS1 have been studied for three values of mean relative deadline θ, namely 2, 4, and 8, denoted as type I, II, and III, respectively. Moreover, the examples for SYS2 have been studied for θ=4 (a type II system). For all the examples, θ is normalized with respect to $1/\mu_1$. Furthermore, a broad range of normalized class-1 input rates ($\rho_1=\lambda_1/\mu_1$) for SYS1 is considered, while the normalized class-2 input rates for SYS2 ($\rho_2=\lambda_2/\mu_2$) are given some values from almost no traffic up to $(1-p_{u_1})$ obtained from (11). In other words, $\rho_2<(1-p_{u_1})$ should be held to maintain the stability of the system (or equivalently the normalized class-2 input rate for saturation, namely, ρ_2^{sat} is approximated by $(1-p_{u_1})$).

In order to find the accuracy of the analytical results, we have also simulated the above systems through an event-driven simulator, written in C++. Two job generators are considered: one that generates the real-time jobs with the specification indicated above, and another that generates the non real-time jobs. The simulator supports the non-preemptive scheduling of both real-time and non-real-time jobs. Other details of the system are as indicated above. The length of the waiting queue in the simulator changes dynamically up to the available memory of the system (to approximate the unlimited capacity of the desired system with a good estimation). All the experiments (for each data point) have been done for at least 5 million customers in each run, within a 0.01 of relative confidence interval, and with a 99.5% confidence level.

At first, we investigate the loss probability of a system with purely class-1 jobs, namely SYS1 (i.e., ρ_2=0). Note that due to having no concern about the instability of a SYS1, we can do the experiments for a broad range of input rates in here. These results are obtained from the analytical modeling and simulation for a wide range of normalized class-1 input rates (ρ_1) from almost no traffic to very heavy traffic intensity, i.e., for the interval (0, 3]. In the analytical modeling, the capacity of the system is taken to be large enough to be approximated as infinite. The loss probabilities obtained from the analytical modeling as well as simulation and their respective errors are presented in Table 1 for the non-preemptive EDF policy. At the bottom of the table, the *maximum relative error*, *average relative error*, and *root mean square error* (*RMSE*[*]) are also presented for the respective group of data. Figure 4 illustrates the same information graphically showing that the analytical and simulation results almost overlap in all cases.

As can be observed in Table 1, the worst relative error of the analytical and simulation results for a non-preemptive model of the EDF policy is about 1.42 %, which happens when θ=8 and ρ is about 0.9. As can be observed, the analytical results are closer to the simulation results for smaller values of mean relative deadline, namely θ, i.e., the maximum relative error is lower for smaller values of θ. Since the relative errors may cancel each other out, RMSEs have also been shown in the tables.

[*] To calculate RMSE, the square root of the mean value of the squares of relative errors is calculated.

Table 1. Loss probabilities obtained from the analytical method and simulation and their respective errors for a SYS1 with non-preemptive EDF

ρ_1	α_d								
	θ = 2 (Type I)			θ = 4 (Type II)			θ = 8 (Type III)		
	Simulation	**Analytic**	**Err.%**	**Simulation**	**Analytic**	**Err.%**	**Simulation**	**Analytic**	**Err.%**
0.1	0.3445	0.3445	-0.0058	0.2107	0.2108	0.0456	0.1192	0.1192	-0.0386
0.3	0.3663	0.3666	0.0846	0.2321	0.2326	0.1818	0.1355	0.1353	-0.1439
0.5	0.3887	0.3885	-0.0594	0.2548	0.2544	-0.1425	0.1519	0.1514	-0.3627
0.7	0.4110	0.4102	-0.2090	0.2784	0.2771	-0.4408	0.1707	0.1684	-1.3199
0.9	0.4338	0.4320	-0.4062	0.3038	0.3020	-0.5840	0.1929	0.1901	-1.4145
1.1	0.4563	0.4542	-0.4451	0.3328	0.3307	-0.6321	0.2253	0.2240	-0.5894
1.3	0.4794	0.4771	-0.4912	0.3667	0.3647	-0.5288	0.2771	0.2776	0.1844
1.5	0.5035	0.5006	-0.5647	0.4051	0.4040	-0.2785	0.3453	0.3461	0.2271
1.7	0.5278	0.5248	-0.5529	0.4472	0.4467	-0.1181	0.4146	0.4146	0.0019
1.9	0.5519	0.5494	-0.4640	0.4904	0.4900	-0.0830	0.4747	0.4742	-0.1032
2.1	0.5762	0.5739	-0.3897	0.5307	0.5310	0.0535	0.5237	0.5239	0.0338
2.6	0.6337	0.6323	-0.2192	0.6162	0.6162	0.0006	0.6150	0.6154	0.0579
3.0	0.6738	0.6735	-0.0497	0.6671	0.6668	-0.0498	0.6662	0.6667	0.0700
	Max relative error= -0.5647 %			Max relative error= -0.6321 %			Max relative error= -1.4145 %		
	Average relative error= -0.2881 %			Average relative error= -0.1701 %			Average relative error= -0.2044 %		
	RMSE=0.3515 %			RMSE=0.2956 %			RMSE=0.5072 %		

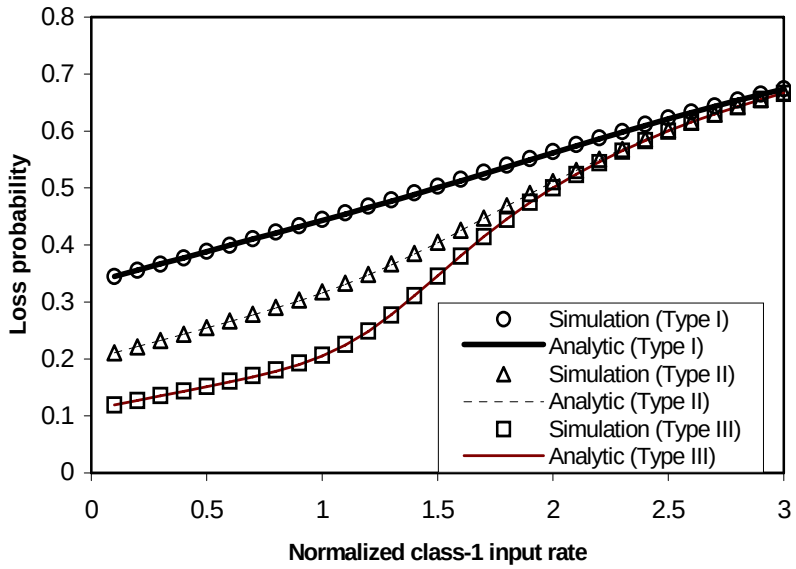
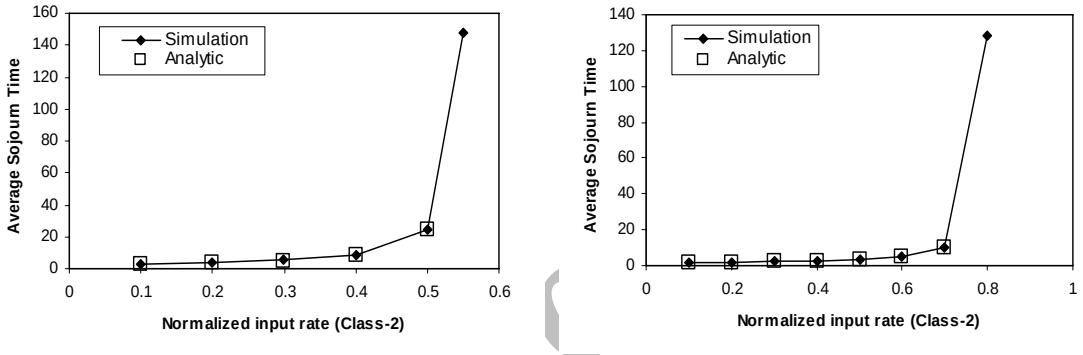


Figure 4. Loss probability (α_d) for a SYS1 with non-preemptive EDF scheduling policy.

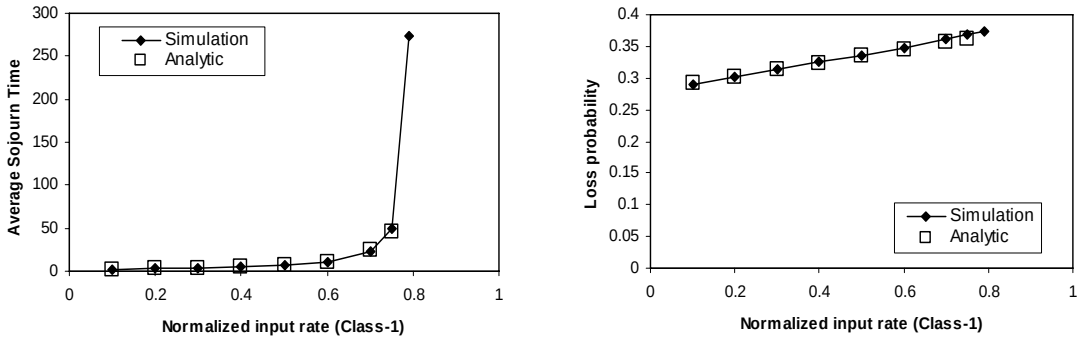
Second, we investigate a type II system (θ=4) with both class-1 and class-2 jobs ($\rho_1,\rho_2\neq0$), namely, a SYS2. The experiments have been done for the traffic intensities which do not violate the stability conditions of the system. As the first example, we consider a system with two fixed values of ρ_1, namely, 0.7 and 0.3, for $\mu_1=\mu_2=1$. Using the solution technique described in Section 3 and the loss rates described by (19) and (20), we can see that ρ_2 should be below $(1-p_{u_1})=0.557$ (ρ_2^{sat}=0.557) for ρ_1=0.7 and below $(1-p_{u_1})=0.8087$ (ρ_2^{sat}=0.8087) for ρ_1=0.3 to maintain the stability of the system. The analytical and simulation results of the average sojourn time of class-2 jobs for $\rho_2<(1-p_{u_1})$ and the two values of ρ_1, namely, 0.7 and 0.3, have been shown graphically in

Figure 5. As can be observed, the analytical and simulation results almost overlap in all cases. (The analytical results for the class-2 input rates close to ρ_2^{sat} have not been calculated. The reason is that for such values of input rates, solving a Markov chain that approximates an infinite-capacity queue becomes very hard to compute numerically due to the large capacity of the respective queues.) As another example, we consider a system with a fixed value of ρ_2=0.5 and μ_1=μ_2=1. For such a system, the loss probability of class-1 jobs and the average sojourn time of class-2 jobs for different values of ρ_1 (with which the system still remains stable) have been shown in Figure 6, where the respective analytical and simulation results also almost overlap in all cases. Similar results for a SYS2 with μ_1=2μ_2=1 have been presented in Figure 7 for two values of ρ_1, namely, 0.6 and 0.3, and also in Figure 8 for a fixed value of ρ_2=0.3. As can be observed, the accuracies of the results are similar to the respective ones of the μ_1=μ_2=1.



(a) (b)

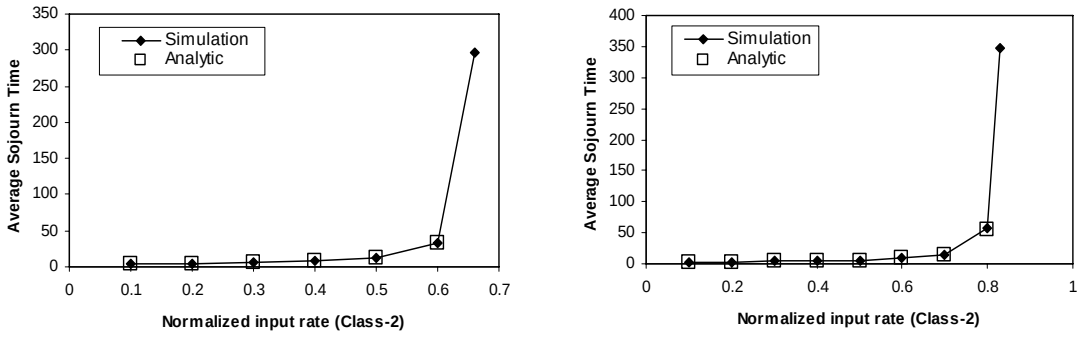
Figure 5. Average sojourn time of class-2 jobs (V_2) for a type II SYS2 with non-preemptive EDF, μ_1=μ_2=1, and (a) ρ_1=0.7, (b) ρ_1=0.3.



(a) (b)

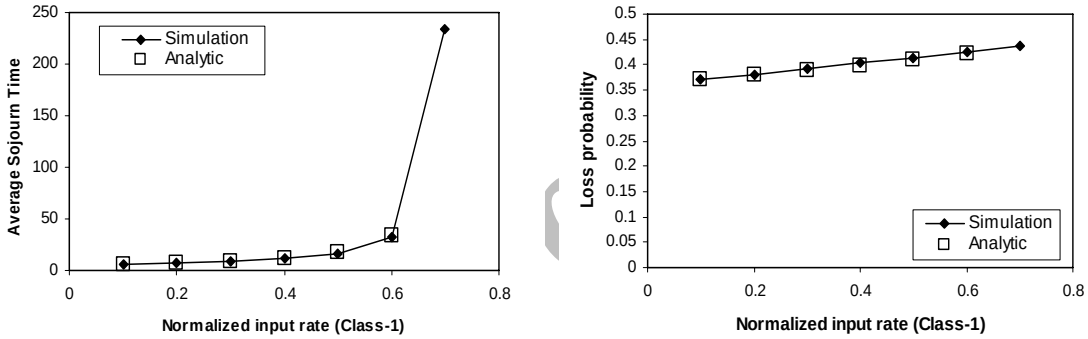
Figure 6. The behavior of a type II SYS2 with non-preemptive EDF, μ_1=μ_2=1, and ρ_2=0.5 for different values of ρ_1, (a) average sojourn time of class-2 jobs (V_2), (b) loss probability of class-1 jobs (α_d).

To illustrate the accuracy of our approximation method with respect to the class-2 job performance measures, the simulation and analytical results of Figure 5 and Figure 7 for the corresponding normalized input rates are shown in Table 3 and Table 4, respectively. As also indicated above, due to the complexity of approximating an infinite-capacity queue, the relative errors for the average



(a) (b)

Figure 7. Average sojourn time of class-2 jobs (V_2) for a type II SYS2 with non-preemptive EDF, $\mu_1=2\mu_2=1$, and (a) $\rho_1=0.6$, (b) $\rho_1=0.3$.



(a) (b)

Figure 8. The behavior of a type II SYS2 with non-preemptive EDF, $\mu_1=2\mu_2=1$, and $\rho_2=0.3$ for different values of ρ_1, (a) average sojourn time of class-2 jobs (V_2), (b) loss probability of class-1 jobs (α_d).

Table 2. Average sojourn time obtained from the analytical method and simulation and their respective errors for a type II SYS2 with $\rho_1=0.7$ or $\rho_1=0.3$, and $\mu_1=\mu_2=1$

ρ_2	Average Sojourn Time (V_2)					
	$\rho_1 = 0.7 \Rightarrow \rho_2^{sat} = 0.557$			$\rho_1 = 0.3 \Rightarrow \rho_2^{sat} = 0.8087$		
	Simulation	Analytic	Error (%)	Simulation	Analytic	Error (%)
0.1	2.7398	2.7006	-1.43076	1.4786	1.4715	-0.48018
0.2	3.6053	3.5607	-1.23707	1.7414	1.7347	-0.38475
0.3	5.1822	5.089	-1.79846	2.1095	2.1015	-0.37924
0.4	8.6429	8.5596	-0.9638	2.6572	2.6477	-0.35752
0.5	24.0235	24.11	0.360064	3.571	3.5478	-0.64968
0.6				5.3593	5.3106	-0.9087
0.7				10.4321	10.3125	-1.14646
	Average relative error= -1.014 %			Average relative error= -0.6152 %		
	RMSE= 1.1453 %			RMSE= 0.7317 %		

sojourn times may increase when the class-2 input rate approaches to ρ_2^{sat}. However, the average relative error and RMSE for the presented data points have been shown at the bottom of the tables. As can also be seen in the tables, while the accuracies of the results are in the acceptable range for most applications, they may be more accurate for the case of class-1 and class-2 jobs with more similar service rates, e.g., $\mu_1=\mu_2=1$.

Table 3. Average sojourn time obtained from the analytical method and simulation and their respective errors for a type II SYS2 with $\rho_1=0.6$ or $\rho_1=0.3$, and $\mu_1=2\mu_2=1$

ρ_2	Average Sojourn Time (V_2)					
	$\rho_1 = 0.6 \Rightarrow \rho_2^{sat} = 0.663$			$\rho_1 = 0.3 \Rightarrow \rho_2^{sat} = 0.835$		
	Simulation	Analytic	Error (%)	Simulation	Analytic	Error (%)
0.1	3.4159	3.4162	0.008782	2.6046	2.6065	0.072948
0.2	4.2508	4.2424	-0.19761	3.0454	3.0446	-0.02627
0.3	5.5211	5.5242	0.056148	3.6428	3.6465	0.10157
0.4	7.7527	7.7818	0.375353	4.5284	4.525	-0.07508
0.5	12.6942	12.8141	0.944526	5.9207	5.928	0.123296
0.6	32.238	33.5555	4.086792	8.47	8.5246	0.644628
0.7				14.7368	14.9654	1.551219
0.8				56.7058	56.2121	-0.87063
	Average relative error= 0.8790 %			Average relative error= 0.1902 %		
	RMSE= 1.7213 %			RMSE= 0.7764 %		

6. CONCLUDING REMARKS AND FUTURE WORK

In this paper, we have presented a method for approximating the performance of a two-class M/M/1 system. The prioritized class-1 jobs are considered to be real-time and served according to the earliest-deadline-first (EDF) scheduling policy, and the non real-time class-2 jobs are served according to the FCFS policy. The service discipline of the system is non-preemptive. The system has been solved for real-time jobs with deadlines until the end of service and non-preemptive model of the EDF policy. The performance measure of class-1 jobs is the loss probability and that of the class-2 jobs is the average sojourn (waiting) time. Moreover, the stability conditions of the system are considered. The importance of the problem arises from the fact that EDF is an optimal policy which minimizes the fraction of lost real-time (class-1) jobs. The analysis is done by estimating an important parameter called the loss rate of real-time jobs. To the best of our knowledge, in spite of its importance, there has been no exact analytical solution for the analysis of EDF, even for a system with purely real-time jobs. We have proposed an approximation method for a two-class system which we believe is quite accurate and very simple. The proposed method can also simply be extended to real-time jobs with deadlines until the beginning of service using the respective loss rates presented in (Kargahi and Movaghar, 2006).

Some of the future works to continue this study include extending the presented approach to other patterns of input traffic, multi-server systems, and more general distribution of relative deadlines.

REFERENCES

[1] Baccelli F., Boyer P., Hebuterne G. (1984), Single-server queues with impatient customers; *Advanced Applied Probability* 16; 887-905.

[2] Barrer D.Y. (1957), Queueing with impatient customers and ordered service; *Operational Research*5; 650-656.

[3] Bernat G., Burns A., Llamosi, A. (2001), Weakly hard real-time systems; *IEEE Transactions on Computers* 50(4); 308-321.

[4] Boxma O.J., Wall P.R. (1994), Multiserver queues with impatient customers; Proceedings of the *14th International Teletraffic Congress*, 14, Antibes, France, pp 743-756.

[5] Brandt A. Brandt M. (1999a), On the M(n)/M(n)/s Queue with impatient calls; *Performance Evaluation* 35; 1-18.

[6] Brandt A. Brandt M. (1999b), On a two-queue priority system with impatience and its application to a call center; *Methodology and Computing in Applied Probability* 1; 191-210.

[7] Brandt A. Brandt M. (2002), Asymptotic results and a Markovian approximation for the M(n)/M(n)/s+GI system; *Queueing Systems* 41; 73-94.

[8] Brandt A. Brandt M. (2004), On the two-class M/M/1 system under preemptive resume and impatience of the prioritized customers; *Queueing Systems* 47(1-2), 147-168.

[9] CAN-CIA, CAN specification 2.0 Part B, http://www.cancia.org/downloads/ ciaspecifications, 1992.

[10] Choi B.D., Kim B., Chung J. (2001), M/M/1 Queue with impatient customers of higher priority; *Queueing Systems* 38; 49-66.

[11] Daley D.J. (1965), General customer impatience in queue GI/G/1; *Journal of Applied Probability* 2; 186-205.

[12] Dolev S. Keizelman A. (1999), Non-preemptive real-time scheduling of multimedia tasks; *Real-Time Systems* 17(1); 23–39.

[13] Doytchinov B., Lehoczky J., Shreve S. (2001), Real-time queues in heavy traffic with earliest-deadline-first queue discipline; *Annals of Applied Probability* 11; 332-379.

[14] EN 50170, General purpose field communication system, In *European Standard*, CENELEC, 1996.

[15] George L., Muhlethaler P., Rivierre N. (1995), Optimality and non-preemptive real-time scheduling revisited; Rapport de Recherche RR-2516, INRIA, Le Chesnay Cedex, France.

[16] George L., Rivierre N., Spuri M. (1996), Preemptive and non-preemptive real-time uni-processor scheduling; Rapport de Recherche RR-2966, INRIA, Le Chesnay Cedex, France.

[17] Hong J., Tan X., Towsley D. (1989), A performance analysis of minimum laxity and earliest deadline scheduling in a real-time system; *IEEE Transactions on Computers* 38(12); 1736-1744.

[18] Jaiswal N. (1968), Priority queues; Academic Press, New York.

[19] Kargahi M., Movaghar A. (2004), A method for performance analysis of earliest-deadline-first scheduling policy; Proceedings of the *2004 IEEE International Conference on Dependable Systems and Networks*, Florence, Italy, pp 826-834.

[20] Kargahi M., Movaghar A. (2005), Non-preemptive earliest-deadline-first scheduling policy: A performance study; Proceedings of IEEE International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems, Georgia, Atlanta, USA, pp 201-210.

[21] Kargahi M., Movaghar A. (2006), A method for performance analysis of earliest-deadline-first scheduling policy; *Journal of Supercomputing* 37(2); 197-222.

[22] Kargahi M., Movaghar A. (2007), A multiprocessor system with non-preemptive earliest deadline first scheduling policy: A performability study; *Journal of Industrial and Systems Engineering* 1(1); 37-55.

[23] Kruk L., Lehoczky J., Shreve S., Yeung S.N. (2003), Multiple-input heavy-traffic real-time queues; *Annals of Applied Probability* 13(1); 54-99.

[24] Kruk L., Lehoczky J.P., Shreve S., Yeung S.N. (2004), Earliest-deadline-first service in heavy-traffic acyclic networks; *Annals of Applied Probability* 14(3); 1306-1352.

[25] Lehoczky J.P. (1996a), Real-time queueing theory; Proceedings of the *17th IEEE Real-Time Systems Symposium*, Washington, D.C., USA, pp 186-195.

[26] Lehoczky J.P. (1996b), Using real-time queueing theory to control lateness in real-time systems; *Performance Evaluation Review* 25(1); 158-168.

[27] Leulseged A., Nissanke N. (2004), Probabilistic analysis of multi-processor scheduling of tasks with uncertain parameters;, Proceedings of the 9th International Conference on Real-Time and Embedded Computing Systems and Applications, pp 103-122. (LNCS 2968)

[28] Liu C.L., Layland J.W. (1973), Scheduling algorithms for multiprogramming in a hard real-time environment; *Journal of the ACM* 20(1); 46-61.

[29] Livani M.A., Kaiser J. (1998), EDF consensus on CAN bus access for dynamic real-time applications; Proceedings of *IEEE Workshop on Parallel and Distributed Computing Systems in conjunction with 12th International Parallel Processing Symposium* / 9th Symposium on Parallel and Distributed Processing, pp 1088-1097.

[30] May M., Bolot J., Jean-Marie A., Diot C. (1999), Simple performance models of differentiated services schemes for the Internet; Proceedings of the *18th Annual Joint Conference of the IEEE Computer and Communications Societies, INFOCOM'99*, New York, NY, pp 1385-1394 (vol. 3).

[31] Miller R.G. (1960), Priority queues; *Annals of Mathematical Statistics* 31; 86-103.

[32] Movaghar A. (1998), On queueing with customer impatience until the beginning of service; *Queueing Systems* 29; 337-350.

[33] Movaghar A. (2006), On queueing with customer impatience until the end of service; *Stochastic Models* 22; 149-173.

[34] Nissanke N., Leulseged A., Chillara S. (2002), Probabilistic performance analysis in multiprocessor scheduling; *Journal of Computing and Control Engineering* 13(4); 171-179.

[35] Palm C. (1953), Methods for judging the annoyance caused by congestion; *Tele 2*; 1-20.

[36] Towsley D., Panwar S.S. (1990), On the optimality of minimum laxity and earliest deadline scheduling for real-time multiprocessors; Proceedings of *IEEE EUROMICRO-90 Workshop on Real-Time*, pp 17-24.

[37] Towsley D., Panwar S.S. (1992), Optimality of the stochastic earliest deadline policy for the G/M/c queue serving customers with deadlines; Proceedings of the *Second ORSA Telecommunications Conference*.

[38] Zhao W., Stankovic J.A. (1989), Performance analysis of FCFS and improved FCFS scheduling algorithms for dynamic real-time computer systems; Proceedings of *IEEE Real-Time Systems Symposium*, California, USA, pp 156-165.