Scientia Iranica, Vol. 15, No. 2, pp 252–265

SCIENTIA
IRANICA

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

M. Kargahi* and A. Movaghar[1]

This paper introduces an analytical method for approximating the performance of a two-class priority $M/M/1$ system. The prioritized class-1 jobs are real-time, served either with the preemptive or non-preemptive Earliest-Deadline-First (EDF) policy and can preempt the non real-time class-2 jobs. The preempted service of the class-2 job is resumed from the time in instances where no class-1 job is in the system. The service discipline of 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 numerical and simulation results, it is found that the existing errors are relatively small.

INTRODUCTION

Multi-priority demand for computation and communication are required in many applications of 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 traffic'. 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 real-time incoming demand, 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 [1], which are considered in this paper, while in Soft Real-Time (SRT) systems, a late job that has missed its deadline, continues to get service until completion.) Two models of job behavior are usually 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 between the traffic of different classes may affect the performance of non real-time, as well as 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 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 first instant of getting 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

*. *Corresponding Author, School of Computer and Electrical Engineering, Faculty of Engineering, University of Tehran, Tehran, I.R. Iran.*

1. *Department of Computer Engineering, Sharif University of Technology, Tehran, I.R. Iran.*

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 has been finished. Though preemptive scheduling can guarantee better system utilization and is usually more desirable, there are 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 [2] (such as CAN bus [3,4]) is inherently non-preemptive. This is due to the fact that 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 [5]. Non-preemptive scheduling for real-time embedded systems has also some benefits, such as ease of implementation, reduced run-time overhead and guaranteed 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 jobs in a real-time system strongly influences the performance of the system. Among such policies, the Earliest-Deadline-First (EDF) policy [6], which schedules 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 [7,8] and, also, stochastically minimizes the fraction of lost jobs in the same class of policies [9,10].

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 traffic. As an example of such a system, consider a DiffServ supported network, which offers service differentiation for different classes of flow at each network node [11]. 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, one 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 job, interaction of the jobs in the classes should be taken into account. More critical examples of the matter can also be found in the applications of wireless sensor networks.

In this paper, an approximation method is presented for the performance analysis of a two-class $M/M/1$ system. 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) and also can preempt the class-2 (non real-time) jobs. This type of relative deadline 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. Class-1 jobs in the system have DES and are served according to EDF, while 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 the scheduling of class-1 jobs can be very important. Since both preemptive and non-preemptive models of the EDF scheduling policy are optimal, either preemptive or non-preemptive EDF scheduling of real-time jobs are considered in our analysis. The proposed approximation method to analyze the two-class $M/M/1$ system uses a key parameter, namely γ_n, which is the rate of missing deadlines when there are n class-1 jobs in the system. This important parameter is estimated using an upper bound and a lower bound for the case of preemptive EDF. The resulting formulation is then generalized into the case of non-preemptive EDF scheduling policy by partitioning the system into two virtual subsystems: One with the FCFS policy and another with the preemptive EDF policy. Such results are then used in a Markov chain model of the two-class $M/M/1$ system. To the best of the authors' 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. First, some related works are presented. Then, basic system model and the proposed analytical method for modeling the system and extracting the required performance measures are described. This is followed by an explanation of a method of estimating the loss rates of class-1 jobs for both preemptive and non-preemptive models of the EDF scheduling policy. After that, some numerical examples and the comparison of the analytical and simulation results are provided. Summary, concluding remarks, and future research grounds are, finally, presented.

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 [12-21] and the references therein. However, in spite of the optimality of both preemptive and non-preemptive EDF policies [9,10], relatively few papers exist on the probabilistic analysis of EDF. This may be due to the complexity of such analysis. Some of the work done in this area, such as [22,23] have concentrated on the probabilistic analysis of EDF for periodic task arrivals. For non-periodic arrivals, Hong et al. [24] first introduced upper and lower bounds for the performance of an $M/M/m/\mathrm{EDF}+M$ 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 [24] are only for relative input rates of 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 [25] and also extended to $M/M/m/\mathrm{EDF}+G$ queues in [26] ($m=1$ for DES has been assumed in all three studies). Moreover, an approximation method for the analysis of an $M/M/1/\mathrm{EDF}+M$ queue, in the case of non-preemptive EDF scheduling of jobs with DES, has been presented in [27]. On the other hand, Lehoczky and his colleagues in [28-30] have developed an approximation method to compute the fraction of late tasks in a SRT system for a heavy traffic case, where the traffic intensity converges to 1 and the system has a high average utilization. In their model, it is assumed that all jobs are processed to completion. The method is called a real-time queuing theory (RTQT), which is an extension of the traditional queuing 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 [29] 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. [28] for $GI/G/1$ queues. It should be noted that the EDF scheduling policy considered in [24,25,27] and, also, in the current paper, differs from the one analyzed in [28-30]. This is due to the fact that, unlike the latter works, the former works never scheduled jobs which were already past their deadlines (due to the FRT nature of the system). Furthermore, the latter works have only focused on 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 [31-37]. Also, few papers exist in the literature on priority queues with some classes of real-time jobs. Brandt and Brandt [31] 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 class-2 jobs in the system are presented therein. Such results are later improved in an exact form for a two-class single server queue in [32]. Similar results for deterministic relative deadlines and for both cases of DBS and DES are presented in [33]. In all of these studies, the scheduling policy of class-1 jobs is considered to be FCFS in a FRT system. For the EDF scheduling of multi-class traffics, Kruk et al. in [34] first used RTQT [28-30] 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 [35] 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 heavy traffic intensities, whereas, the work presented in this paper considers FRT systems and never schedules real-time (class-1) jobs that are already past their deadlines and, also, covers almost all of the input rates with which the system remains stable.

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.

System Model

A two-class $M/M/1$ system is considered, 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. Class-1 jobs are served with a preemptive priority discipline over class-2 jobs. More precisely, if a class-1 job arrives before the service completion of a class-2 job, the service will be interrupted, as long as there are still class-1 jobs in the system. The preempted service of the class-2 job is resumed from the time instant that no class-1 jobs are present in the system. Furthermore, a relative deadline is associated with each class-1 job. It is assumed 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 a job is in service when it reaches its deadline, it is aborted and then thrown out. In either case, the job that is thrown away is considered 'lost'. 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. Two models of class-1 job behavior are considered: Preemptive and non-preemptive. Whereas, in the former model, an arriving job with an earlier deadline than the serving job preempts it, in the latter model, no job can preempt the serving job. It is proved in [9,10] that the EDF scheduling policy stochastically maximizes the fraction of jobs meeting their deadlines for DES, within the classes of non-idling service-time-independent preemptive and non-preemptive scheduling policies.

According to the relation between the two classes of jobs, the behavior and performance of class-1 jobs are not affected by the class-2 ones and, hence, correspond to those of an $M/M/1$/EDF + M system. In contrast, the behavior and performance of class-2 jobs are extremely influenced by the class-1 jobs. In order to model the system, the state of the two-class system is represented by $\mathbf{n} = (n_1, n_2)$, where n_1 is the current number of class-1 jobs and n_2 is the number of existing class-2 jobs in there.

The approach presented in this paper is based on using a state-dependent loss rate function, γ_{n_1}, for class-1 jobs to be defined below. Let, $\mathbb{N}$ be the set of natural numbers and $\mathbb{R}^+$ the set of positive real numbers. For t, $\varepsilon \in \mathbb{R}^+$ and $n_1 \in \mathbb{N}$, let $\Psi_{n_1}(t, \varepsilon) \equiv$ the probability that a class-1 job misses its deadline during $[t, t+\varepsilon)$, given there are n_1 real-time (class-1) jobs in the system at time t.

Define:

$$\gamma_{n_1}(t) = \lim_{\varepsilon \to 0} \frac{\Psi_{n_1}(t, \varepsilon)}{\varepsilon}. \tag{1}$$

Assuming statistical equilibrium, let:

$$\gamma_{n_1} = \lim_{t \to \infty} \gamma_{n_1}(t), \tag{2}$$

where γ_{n_1} is the (steady-state) rate of missing deadlines when there are n_1 class-1 jobs in the system (including the one being served). Accordingly, the resulting Markov chain model of the two-class system, $\mathbf{M}$, may partially be shown as in Figure 1. Assuming that the system is in state $\mathbf{n} = (n_1, n_2)$, the state of the system can be changed to (n_1+1, n_2) or (n_1, n_2+1), with rates λ_1 or λ_2, respectively. When $n_1 > 0$, the state can be changed to $(n_1 - 1, n_2)$, 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 γ_{n_1}). On the other hand, when $n_1 = 0$ and $n_2 > 0$, the state of the system can be changed to $(n_1, n_2 - 1)$ because of completing the service requirements of a class-2 job (with rate μ_2).

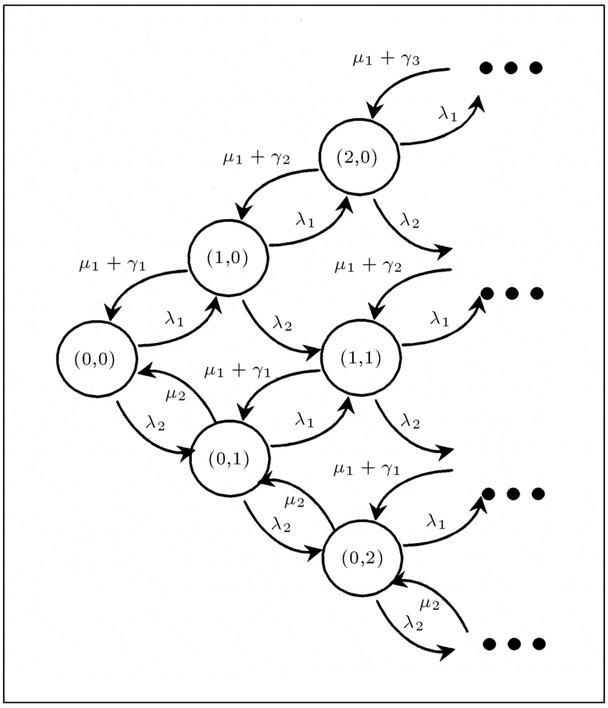


Figure 1. Partial state-transition-rate diagram for Markov chain $\mathbf{M}$.

Barrer [13] 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 [16-19] 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 [16,18], 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 [19]. 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 [26] and, also, a method for estimating that of an $M/M/1$/EDF + M queue, with non-preemptive services for DES, is proposed in [27].

Model Solution

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

$$p(n_1, n_2) \equiv \text{the (steady-state) probability that the system is in state } n = (n_1, n_2). \quad (3)$$

The balance equations for the system, in equilibrium, can be written as:

$$0 = -(\lambda_1 + \lambda_2)p(0,0) + (\mu_1 + \gamma_1)p(1,0) + \mu_2 p(0,1), \quad \text{if } n_1 = n_2 = 0,$$

$$0 = \lambda_2 p(0, n_2 - 1) - (\lambda_1 + \lambda_2 + \mu_2)p(0, n_2) + (\mu_1 + \gamma_1)p(1, n_2) + \mu_2 p(0, n_2 + 1), \quad \text{if } n_1 = 0, n_2 > 0,$$

$$0 = \lambda_1 p(n_1 - 1, 0) - (\lambda_1 + \lambda_2 + \mu_1 + \gamma_{n_1})p(n_1, 0) + (\mu_1 + \gamma_{n_1+1})p(n_1 + 1, 0), \quad \text{if } n_1 > 0, n_2 = 0,$$

$$0 = \lambda_1 p(n_1 - 1, n_2) + \lambda_2 p(n_1, n_2 - 1) - (\lambda_1 + \lambda_2 + \mu_1 + \gamma_{n_1})p(n_1 + n_2) + (\mu_1 + \gamma_{n_1+1})p(n_1 + 1, n_2), \quad \text{if } n_1, n_2 > 0. \quad (4)$$

The normalizing condition is, also, as follows:

$$\sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} p(n_1, n_2) = 1. \quad (5)$$

Solving the equilibrium in Equations 4 and using Equation 5, one finds the state probabilities of the system, namely, $p(n_1, n_2), n_1, n_2 \in \{0, 1, 2, \cdots\}$. Let:

$$p_1(k) = \sum_{n_2=0}^{\infty} p(k, n_2), \quad (6)$$

and:

$$p_2(k) = \sum_{n_1=0}^{\infty} p(n_1, k), \quad (7)$$

be the probabilities of having k class-1 and k class-2 jobs in the system, respectively. It is necessary to find the maximum permitted input rate of class-2 jobs at which the system still remains stable, noting that class-2 jobs have no influence on the behavior and performance of class-1 jobs. Since $p_1(0)$ is the probability of having no class-1 jobs in the system, $1 - p_1(0)$ would be the time fraction for serving the class-1 jobs. Further, the process of class-2 jobs in the server is stable, iff all class-2 jobs are served, i.e., since λ_2/μ_2 is the time fraction where class-2 jobs are served by the server and $1 - p(0,0)$ is the time fraction where the server is busy with serving class-1 or class-2 jobs, in the case of a stable system, it follows that:

$$1 - p_1(0) + \lambda_2/\mu_2 = 1 - p(0,0), \quad (8)$$

or, equivalently,

$$p(0,0) = p_1(0) = \lambda_2/\mu_2. \quad (9)$$

Hence, the process of class-2 jobs is stable, iff $p(0,0)$ in Equation 9 is positive, i.e., iff:

$$\lambda_2/\mu_2 < p_1(0), \quad (10)$$

where $p_1(0)$ can simply be obtained by using Equation 6, assuming $\lambda_2 = 0$ in the Markov chain $\mathbf{M}$ and solving it using the standard Markovian solution techniques.

Assuming a stable system, the desired performance measures can be calculated. The loss probability of class-1 jobs in the system may be obtained as follows:

$$\alpha_d = \frac{\sum_{n_1=1}^{\infty} \sum_{n_2=0}^{\infty} p(n_1, n_2)\gamma_{n_1}}{\sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} p(n_1, n_2)\lambda_1} = \frac{\sum_{n_1=1}^{\infty} \sum_{n_2=0}^{\infty} p(n_1, n_2)\gamma_{n_1}}{\lambda_1}, \quad (11)$$

which is the average rate of missing deadlines divided by the average rate of class-1 job arrivals. Whereas, for 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, one has:

$$\overline{N_i} = \sum_{k=0}^{\infty} k p_i(k). \quad (12)$$

Using Little's formula, one obtains:

$$V_2 = \overline{N_2}/\lambda_2, \quad (13)$$

where V_2 is the average sojourn time of class-2 jobs. The bandwidth of the server for class-2 jobs is the fraction of the total bandwidth (μ_2), in which there are no class-1 jobs in the system, namely, $\mu_2 p_1(0)$. Therefore, the average waiting time of class-2 jobs can be derived as follows:

$$W_2 = V_2 - \frac{1}{\mu_2 p_1(0)} = \frac{1}{\lambda_2}\overline{N_2} - \frac{1}{\mu_2 p_1(0)}, \quad (14)$$

wherein $\mu_2 p_1(0)$ is the average bandwidth (i.e., the actual service rate) of the class-2 jobs and, hence, $1/\mu_2 p_1(0)$ is the average time between the first instant of providing service to them and the completion of their service (considering the preemptions occur due to class-1 job arrivals).

To analyze the system with the EDF policy for class-1 jobs, one needs to have a formulation of γ_{n_1} (for EDF), as defined in Equation 2. Next, a method for estimating γ_{n_1} (of EDF), for an infinite-capacity system, with two cases of preemptive and non-preemptive services of class-1 (real-time) jobs is reviewed.

DETERMINATION OF LOSS RATES

In this section, two methods for estimating $\gamma_{n_1}^{\mathrm{EDF}}$ are presented, in the cases of preemptive and non-preemptive EDF policies and DES. To do so, some bounds for $\gamma_{n_1}^{\mathrm{EDF}}$ in the preemptive case are defined. A combination of these bounds will result in an estimation of the loss rates of EDF for the preemptive case, namely, $\gamma_{n_1}^{\mathrm{P-EDF}}$. Then, a different view of the system is used with the non-preemptive EDF policy, in which the system is split into two subsystems: One with the FCFS policy and another with the preemptive EDF policy. In this regard, the loss rates of the non-preemptive EDF policy, namely, $\gamma_{n_1}^{\mathrm{NP-EDF}}$, will also be estimated.

The ordering $\gamma_{n_1}^{\mathrm{FCFS-det}} \leq \gamma_{n_1}^{\mathrm{P-EDF}} \leq \gamma_{n_1}^{\mathrm{FCFS-exp}}$ is assumed for the loss rate of each state of the Markov chain, $\mathbf{M}$, for the preemptive EDF policy, which is obtained from properties of the FCFS and preemptive EDF policies, for deterministic and exponential relative deadline distributions. Thus, $\gamma_{n_1}^{\mathrm{FCFS-det}}(\gamma_{n_1}^{\mathrm{FCFS}}$ for deterministic relative deadlines) is taken as the lower bound and $\gamma_{n_1}^{\mathrm{FCFS-exp}}(\gamma_{n_1}^{\mathrm{FCFS}}$ for exponential relative deadlines) as the upper bound of $\gamma_{n_1}^{\mathrm{P-EDF}}$. Since the exact values of $\gamma_{n_1}^{\mathrm{FCFS}}$ are known for the general distribution of relative deadlines until the end of service (DES) [19], it is tried to exploit this information to come up with some estimation for $\gamma_{n_1}^{\mathrm{P-EDF}}$. Thus, the above two bounds are linearly combined, using a multiplier, to obtain an appropriate estimation of $\gamma_{n_1}^{\mathrm{P-EDF}}$. (If the exact values of $\gamma_{n_1}^{\mathrm{P-EDF}}$ were to be known, then, solving $\mathbf{M}$ would result in an exact analysis of the system with preemptive EDF). More explanation of the above approach for an infinite-capacity queue is given below.

As indicated in [19], for a fixed value of mean relative deadline (θ) in a FCFS system, deterministic relative deadlines generate the minimum loss probability among all other relative deadline distributions. Similarly, it is assumed that such a property is also valid for the EDF scheduling policy. Since, for the deterministic relative deadlines, EDF is the same as FCFS, the loss probability of the FCFS policy for deterministic relative deadlines can be assumed as a lower bound of the loss probability of the EDF scheduling policy for exponential relative deadlines. On the other hand, owing to the optimality property of the preemptive EDF policy for DES [9,10], the loss probability is minimized by the policy (with respect to all other non-idling service-time-independent scheduling policies). Therefore, one will have:

$$\alpha_d^{\mathrm{FCFS-det}} \leq \alpha_d^{\mathrm{P-EDF}} \leq \alpha_d^{\mathrm{FCFS-exp}}, \tag{15}$$

where $\alpha_d^{\mathrm{FCFS-det}}$, $\alpha_d^{\mathrm{P-EDF}}$ and $\alpha_d^{\mathrm{FCFS-exp}}$ represent the loss probabilities of the system, with deterministic relative deadlines for the FCFS policy, exponential relative deadlines for the preemptive EDF policy and exponential relative deadlines for the FCFS policy, respectively. It is also assumed that such ordering is valid for loss rates in the FCFS and EDF policies. Such validity is strongly confirmed by the simulation results. Therefore, one will have:

$$\gamma_{n_1}^{\mathrm{FCFS-det}} \leq \gamma_{n_1}^{\mathrm{P-EDF}} \leq \gamma_{n_1}^{\mathrm{FCFS-exp}}, \tag{16}$$

where the functions describing the above two bounds of $\gamma_{n_1}^{\mathrm{P-EDF}}$ are given in [19].

Contrary to the FCFS policy, the authors' simulation results strongly indicate that the state-dependent loss rates depend on the input rate of real-time jobs (λ_1) for the EDF policy. Therefore, one can take advantage of some properties of EDF and some simulation results to make a multiplier that linearly combines the bounds defined previously to get an estimation of $\gamma_{n_1}^{\mathrm{P-EDF}}$. The multiplier must be adjusted to a function of λ_1 to get a more accurate estimation of $\gamma_{n_1}^{\mathrm{P-EDF}}$ as discussed in the following. Thereafter, an estimation of $\gamma_{n_1}^{\mathrm{NP-EDF}}$ will also be presented in the subsequent sections.

Preemptive Real-Time Jobs

In the following, a method for estimating $\gamma_{n_1}^{\mathrm{EDF}}$ for preemptive EDF is presented, namely, $\gamma_{n_1}^{\mathrm{P-EDF}}$, in the case of DES [26]. To do so, the above two bounds are linearly combined, using a multiplier to obtain an appropriate estimation of $\gamma_{n_1}^{\mathrm{P-EDF}}$. As discussed previously, it has been shown that, for the FCFS policy, the loss rate is independent of the input traffic intensity and depends on the number of jobs in the system [19]. Therefore, the above two bounds of $\gamma_{n_1}^{\mathrm{P-EDF}}$ can be calculated as follows [19]:

$$\gamma_{n_1}^{\mathrm{FCFS-exp}} = \begin{cases} 0, & n_1 = 0 \\ \frac{n_1}{\theta}, & n_l \geq 1 \end{cases} \tag{17}$$

and:

$$\gamma_{n_1}^{\text{FCFS}-\text{det}} = \begin{cases} 0, & n_1 = 0 \\ \mu_1 \left(\frac{F_{E_{n_1-1}}(\theta)}{F_{E_{n_1}}(\theta)} - 1 \right), & n_l \geq 1 \end{cases} \tag{18}$$

where:

$$F_{E_{n_1}}(\theta) = 1 - e^{-\mu_1 \theta} \sum_{j=0}^{n_1-1} \frac{(\mu_1 \theta)^j}{j!}, \tag{19}$$

for exponential and deterministic relative deadlines until the end of service, respectively.

Unlike the FCFS policy, the authors' simulation results strongly indicate that the state-dependent loss rates depend on the input rate for the EDF policy. Therefore, a multiplier should be made as a function of λ_1, which combines the bounds defined previously to get an appropriate estimation of $\gamma_{n_1}^{\text{P}-\text{EDF}}$. In order to have a simpler combination, a multiplier has been used that linearly combines the bounds as follows:

$$\gamma_{n_1}^{\text{P}-\text{EDF}} = \frac{\left(\xi(.)\gamma_{n_1}^{\text{FCFS}-\text{exp}} + \gamma_{n_1}^{\text{FCFS}-\text{det}}\right)}{\xi(.) + 1}. \tag{20}$$

The multiplier $\xi(.)$, which defines the effective ratio of each of the bounds on $\gamma_{n_1}^{\text{P}-\text{EDF}}$ in Equation 20, can be found as a function of three parameters, namely n_1, $\mu_1\theta$, and $\rho_1 = \lambda_1/\mu_1$ [26]. In the function, n_1 is the number of class-1 jobs in the queue, $\mu_1\theta$ is the normalized mean relative deadline with respect to the mean service time $1/\mu_1$ and ρ_1 is the normalized arrival rate of class-1 jobs (normalized λ_1 with respect to μ_1). However, the function describing the behavior of $\xi(.)$ with respect to the above three parameters i.e., $\xi(n_1, \mu_1\theta, \rho_1)$, is as follows: (more explanation on the method of extracting the actual values of the required information from simulation can be found in [26]):

$$\xi(n_1, \mu_1\theta, \rho_1) = \frac{1}{(n_1+1)\sqrt{\mu_1\theta}} . \frac{6.7}{\rho_1^{1.25}}. \tag{21}$$

Substituting $\xi(.)$ (obtained from Equation 21) in Equation 20 one can find $\gamma_{n_1}^{\text{P}-\text{EDF}}$. Furthermore, for large values of ρ_1, where n_1 becomes very large (i.e., $n_1 \to +\infty$), the numerical calculation of the lower bound of $\gamma_{n_1}^{\text{P}-\text{EDF}}$, i.e., $\gamma_{n_1}^{\text{FCFS}-\text{det}}$ becomes relatively hard. More precisely, since $F_{E_{n_1-1}}(\theta)$ and $F_{E_{n_1}}(\theta)$ (as in Equation 19) converge to very small values, as $n_1 \to +\infty$, the numerical calculation of their ratio (in order to obtain $\gamma_{n_1}^{\text{FCFS}-\text{det}}$) is not so easy. However, using a similar method for DBS as in [16], a limit value for the ratio for DES can be calculated. Accordingly, when $n_1 \to +\infty$, the value of $\gamma_{n_1}^{\text{FCFS}-\text{det}}$ converges to n_1/θ.

Next, such recent formulations of $\gamma_{n_1}^{\text{P}-\text{EDF}}$ are used to estimate $\gamma_{n_1}^{\text{EDF}}$ for the non-preemptive EDF scheduling of class-1 jobs.

Non-Preemptive Real-Time Jobs

Here, a method is presented for estimating the loss rates of the two-class $M/M/1$ system, with the non-preemptive EDF scheduling of class-1 jobs, namely, $\gamma_{n_1}^{\text{NP}-\text{EDF}}$. Contrary to the preemptive EDF policy, even if the deadline of an arriving job is earlier than that of the serving job, the job in the server will not be preempted and continues to get service when the non-preemptive EDF scheduling policy is used. It has been proven in [9,10] that the loss probability of the non-preemptive EDF scheduling policy ($\alpha_d^{\text{NP}-\text{EDF}}$) is also stochastically minimized among all the policies in the class of non-idling service-time-independent non-preemptive scheduling policies. In spite of its optimality, to the best of the authors' knowledge, no other analytical or approximation method for the probabilistic analysis of this policy exists (other than the one proposed in [27] by the same authors). In the following, a method is presented for estimating $\gamma_{n_1}^{\text{EDF}}$, for the non-preemptive EDF, namely, $\gamma_{n_1}^{\text{NP}-\text{EDF}}$, which results in the approximation of the performance of the non-preemptive EDF scheduling policy. To do so, another view of the system is proposed with purely class-1 jobs ($\lambda_2 = 0$), as in the following paragraphs (noting that preemptive class-2 jobs have no influence on class-1 jobs, even if $\lambda_2 > 0$).

Due to the fact that the serving class-1 job is non-preemptive, after the starting service, the behavior of the system, with respect to this job, is similar to that of a system with the FCFS scheduling policy. On the other hand, the remaining $n_1 - 1$ class-1 job(s) in the system (assuming $n_1 > 0$) follow the EDF scheduling policy. Therefore, the system can be broken into two subsystems: the first one 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, which can virtually be assumed as a preemptive EDF queue for the waiting class-1 jobs with a virtual server with rate μ_1' (see Figure 2). Accordingly, the job with the highest priority in the virtual subsystem is entered into the FCFS subsystem, if it has no job to process.

Since the serving class-1 job leaves the FCFS subsystem, due to its service completion with rate μ_1 or deadline miss with rate $\gamma_1^{\text{FCFS}-\text{exp}}$, it is considered that the job leaves the server with rate $\mu_1 + \gamma_1^{\text{FCFS}-\text{exp}}$. It is assumed that this departure rate is equal to the service rate of the virtual server (μ_1') in the preemptive EDF subsystem, at which the waiting class-1 jobs are entered to the actual server. Therefore, one has $\mu_1' = \mu_1 + \gamma_1^{\text{FCFS}-\text{exp}}$. Such a view of the class-1 jobs will result in a loss rate of $\gamma_{n_1-1}^{\prime\text{P}-\text{EDF}}$ in the preemptive EDF subsystem, which can be calculated from Equations 17 to 21 by replacing μ_1 with μ_1' and ρ_1 with $\rho_1' = \lambda_1/\mu_1'$.

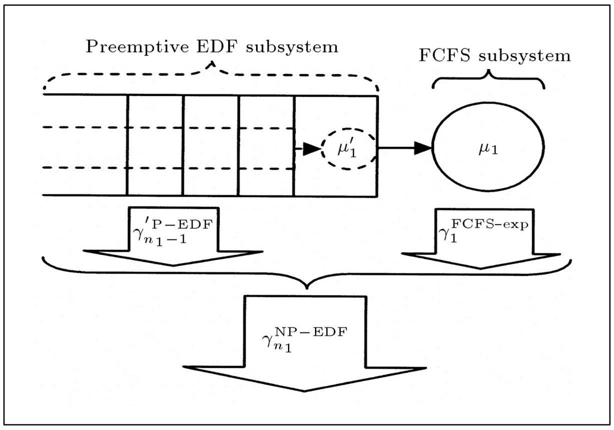


Figure 2. The modified view of a system with the non-preemptive EDF scheduling policy.

Finally, one obtains (assuming $n_1 > 0$):

$$\gamma_{n_1}^{\text{NP-EDF}} = \gamma_{n_1-1}^{\prime\text{P-EDF}} + \gamma_1^{\text{FCFS-exp}}. \tag{22}$$

Substituting $\gamma_{n_1}^{\text{P-EDF}}$ or $\gamma_{n_1}^{\text{NP-EDF}}$ for γ_{n_1} in $\mathbf{M}$ and solving the resulting Markov chain, using the method presented previously, one finds the performance measures of the two-class system with the preemptive or non-preemptive EDF scheduling policies, respectively.

NUMERICAL EXAMPLES

In this section, examples are studied to verify the presented ideas and to illustrate the accuracy of the proposed approximation method. Both preemptive and non-preemptive EDF policies are considered for two system 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 $\theta = 4$ (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 $p_1(0)$. In other words, $\rho_2 < p_1(0)$ 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 equal to $p_1(0)$).

In order to find the accuracy of the analytical results, the above systems have also been simulated 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 both preemptive and non-preemptive EDF scheduling algorithms for 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 10 times for at least 5 million customers in each run and within a 99.5% confidence level.

First, the loss probability of a system with purely class-1 jobs is investigated, namely, SYS1 (i.e., ρ_2=0). Since the performance of class-1 jobs is not influenced by the class-2 ones, these results are also valid for the loss probability of class-1 jobs in a SYS2 (noting that, due to having no concern about the instability of a SYS1, they can be shown for a broad range of input rates in here). These results are obtained from analytical modeling and simulation of 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 analytical modeling, the capacity of the system is taken to be large enough to be approximated as infinite. The experiments have been done for both preemptive and non-preemptive models of the EDF policy. The loss probabilities obtained from the analytical modeling, as well as simulation and their respective errors, are presented in Tables 1 and 2 for the preemptive and non-preemptive EDF policies, respectively. At the bottom of each table, the maximum relative error, average relative error and Root Mean Square Error (RMSE; to calculate RMSE, the square root of the mean value of the squares of relative errors is calculated) are also presented for the respective group of data. Figure 3 illustrates the same information graphically, showing that the analytical and simulation results for both preemptive and non-preemptive models of the EDF policy almost overlap in all cases.

As can be seen in Table 1, the worst relative error of the analytical and simulation results for a preemptive system is about 2.23%, which happens when $\theta = 8$ and ρ is about 0.7. The maximum error of the results in [24] for DES is about 15%. Moreover, the approximation results in [24] are poorer when ρ increases. In the method presented here, however, the approximation results will improve. As can be observed in Table 2, 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 $\theta = 8$ and ρ is about 0.9. As can be observed, the analytical results are closer to the simulation results for smaller values of a 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.

Second, a type II system ($\theta = 4$) is investigated

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

ρ_1	$\alpha_d^{\mathrm{P-EDF}}$								
	$\theta = 2$ (Type I)			$\theta = 4$ (Type II)			$\theta = 8$ (Type III)		
	Simulation	**Analytic**	**Err.%**	**Simulation**	**Analytic**	**Err.%**	**Simulation**	**Analytic**	**Err.%**
0.1	0.3390	0.3410	0.5938	0.2038	0.2059	1.0407	0.1127	0.1144	1.5484
0.3	0.3520	0.3546	0.7469	0.2126	0.2150	1.1411	0.1166	0.1176	0.8456
0.5	0.3670	0.3692	0.6033	0.2243	0.2257	0.6384	0.1224	0.1209	-1.1895
0.7	0.3844	0.3861	0.4503	0.2411	0.2416	0.1908	0.1322	0.1293	-2.2300
0.9	0.4049	0.4050	0.2625	0.2639	0.2653	0.5491	0.1518	0.1505	-0.8386
1.1	0.4275	0.4288	0.2945	0.2961	0.2987	0.8700	0.1930	0.1951	1.0871
1.3	0.4528	0.4541	0.2962	0.3374	0.3409	1.0282	0.2612	0.2639	1.0325
1.5	0.4800	0.4814	0.2952	0.3856	0.3889	0.8488	0.3398	0.3419	0.6109
1.7	0.5077	0.5097	0.4002	0.4362	0.4385	0.5232	0.4129	0.4137	0.1870
1.9	0.5367	0.5382	0.2823	0.4848	0.4860	0.2552	0.4733	0.4741	0.1642
2.1	0.5648	0.5661	0.2300	0.5291	0.5294	0.0478	0.5236	0.5239	0.0548
2.6	0.6289	0.6298	0.1417	0.6156	0.6161	0.0775	0.6158	0.6154	-0.0672
3.0	0.6719	0.6727	0.1134	0.6662	0.6668	0.0887	0.6668	0.6667	-0.0200
	Max relative error= 0.7469%			Max relative error= 1.1411%			Max relative error= -2.2300 %		
	Average relative error= 0.3321%			Average relative error= 0.4554%			Average relative error= 0.0353%		
	RMSE= 0.3719%			RMSE= 0.6023%			RMSE= 0.8709%		

Table 2. Loss probabilities obtained from the analytical method and simulation and their respective errors for a SYS1 with non-preemptive EDF.

ρ_1	$\alpha_d^{\mathrm{NP-EDF}}$								
	$\theta = 2$ (Type I)			$\theta = 4$ (Type II)			$\theta = 8$ (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%		

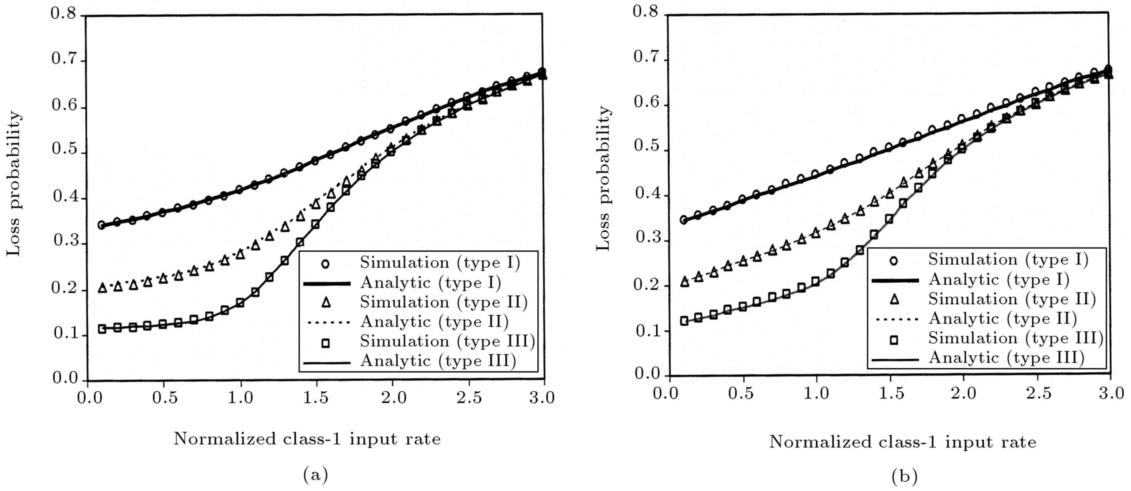


Figure 3. Loss probability (α_d) for a SYS1 with (a) Preemptive EDF policy and (b) Non-preemptive EDF policy.

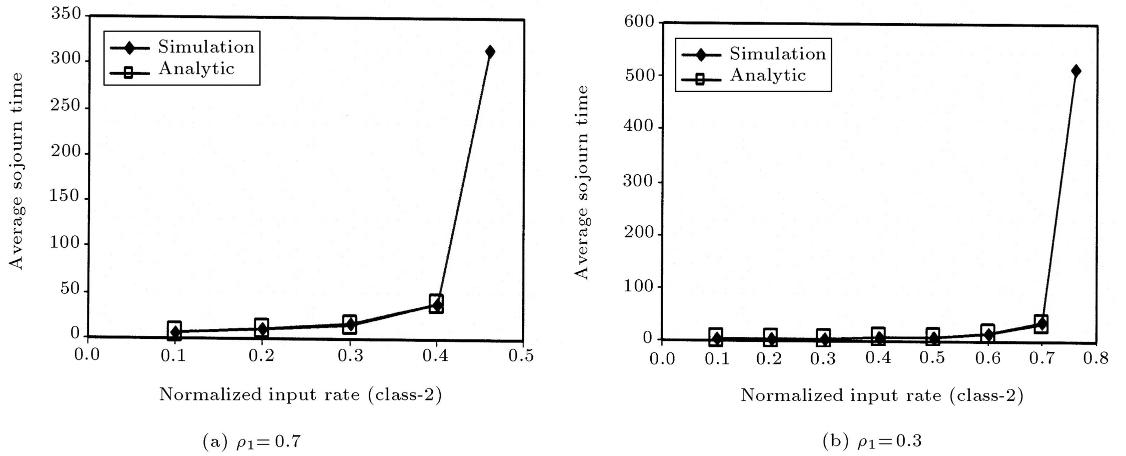
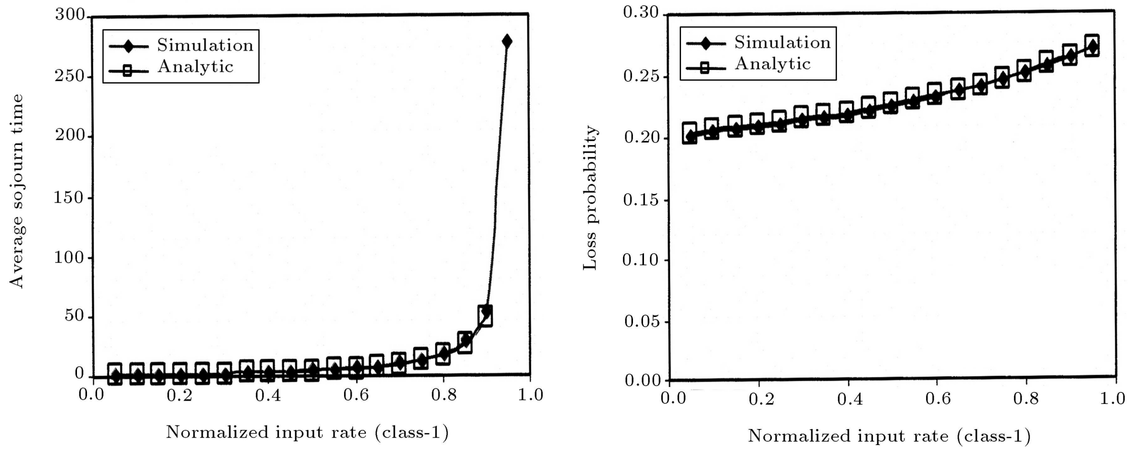


Figure 4. Average sojourn time of class-2 jobs (V_2) for a type II SYS2 with preemptive EDF, $\mu_1 = 2\mu_2 = 1$.

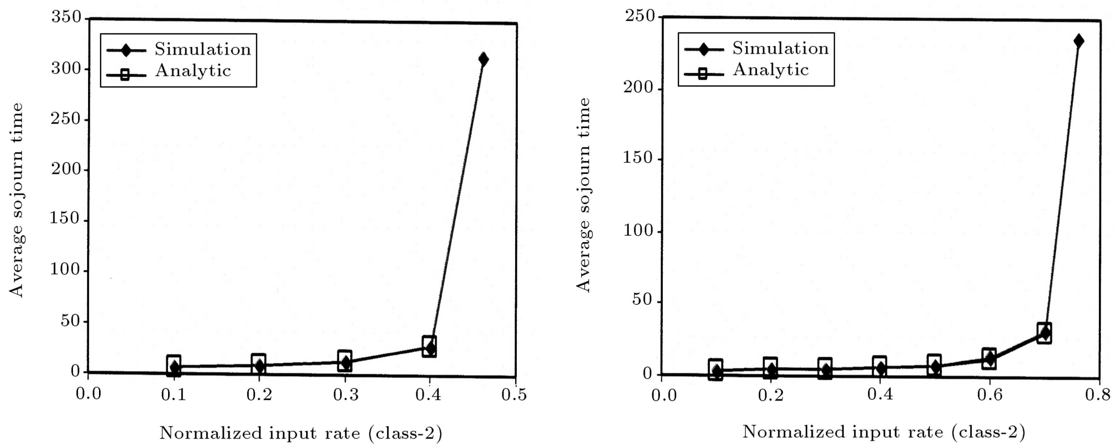
with both class-1 and class-2 jobs ($\rho_1, \rho_2 \neq 0$), namely, a SYS2. The experiments have been done for the traffic intensities which do not violate the stability conditions of the system. As in the first example, a system is considered with the preemptive model of EDF and two fixed values of ρ_1, namely, 0.7 and 0.3 for $\mu_1 = 2\mu_2 = 1$. Using the solution technique described, previously, and the loss rates described by Equations 20 and 21, it can be seen that ρ_2 should be below $p_1(0) = 0.46(\rho_2^{\text{sat}} = 0.46)$ for $\rho_1 = 0.7$ and below $p_1(0) = 0.765(\rho_2^{\text{sat}} = 0.765)$ for $\rho_1 = 0.3$, to maintain the stability of the system. The analytical and simulation results of the average sojourn time of class-2 jobs for $\rho_2 < p_1(0)$ and the two values of ρ_1, namely, 0.7 and 0.3, have been shown graphically in Figure 4. 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 rate, 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, a system is considered with a fixed value of $\rho_2 = 0.3$ and $\mu_1 = \mu_2 = 1$. 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 5, where the respective analytical and simulation results also almost overlap in all cases. Similar results for a SYS2 with the non-preemptive EDF policy have been presented in Figure 6 for two values of ρ_1, namely, 0.7 and 0.3 ($\mu_1 = 2\mu_2 = 1$) and, also, in Figure 7 for a fixed value of $\rho_2 = 0.7(\mu_1 = \mu_2 = 1)$. As can be observed, the accuracies of the results are similar to the respective ones of the preemptive EDF policy.

To illustrate the accuracy of the approximation method, with respect to the class-2 job performance measures, the simulation and analytical results of Figures 4 and 6 for the corresponding normalized input rates are compared with each other in Tables 3 and 4. As can be observed, all the relative errors



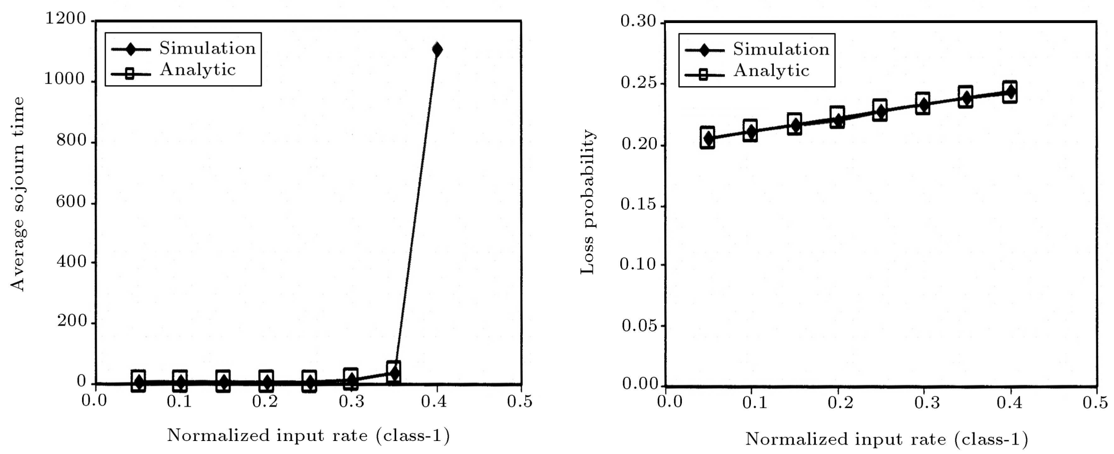
(a) Average sojourn time of class-2 jobs (V_2) (b) Loss probability of class-1 jobs ($\alpha_d^{\text{P-EDF}}$)

Figure 5. The behavior of a type II SYS2 with preemptive EDF, $\mu_1 = \mu_2 = 1$ and $\rho_2 = 0.3$ for different values of ρ_1.



(a) $\rho_1 = 0.7$ (b) $\rho_1 = 0.3$

Figure 6. Average sojourn time of class-2 jobs (V_2) for a type II SYS2 with non-preemptive EDF, $\mu_1 = 2\mu_2 = 1$.



(a) Average sojourn time of class-2 jobs (V_2) (b) Loss probability of class-1 jobs($\alpha_d^{\text{NP-EDF}}$)

Figure 7. The behavior of a type II SYS2 with non-preemptive EDF, $\mu_1 = \mu_2 = 1$ and $\rho_2 = 0.7$ for different values of ρ_1.

Table 3. Average sojourn time obtained from the analytical method and simulation and their respective errors for a type II SYS2 with $\rho_1 = 0.7$ and $\mu_1 = 2\mu_2 = 1$ (see Figures 4a and 6a).

ρ_2	**Average Sojourn Time (V_2)**					
	Preemptive EDF ($\rho_2^{\text{sat}} = 0.46$)			**Non-Preemptive EDF ($\rho_2^{\text{sat}} = 0.494$)**		
	Simulation	**Analytic**	**Error (%)**	**Simulation**	**Analytic**	**Error (%)**
0.05	6.4805	6.3032	-2.7359	5.8162	5.7875	-0.49345
0.1	7.362	7.1569	-2.7859	6.5211	6.5219	0.012268
0.15	8.5452	8.2783	-3.1234	7.5092	7.4698	-0.52469
0.2	10.1433	9.8161	-3.2258	8.7743	8.7401	-0.38977
0.25	12.3987	12.0559	-2.7648	10.4849	10.531	0.43968
0.3	16.1341	15.6196	-3.1889	13.2026	13.245	0.321149
0.35	22.7938	22.1743	-2.7178	17.6924	17.8435	0.854039
0.4	39.3862	38.2078	-2.9919	27.18	27.274	0.345843
0.45	142.5242	137.9149	-3.2341	58.8463	58.3019	-0.92512
	Average relative error = -2.9743%			Average relative error = -0.0400%		
	RMSE = 2.9818%			RMSE = 0.5449%		

Table 4. Average sojourn time obtained from the analytical method and simulation and their respective errors for a type II SYS2 with $\rho_1 = 0.3$ and $\mu_1 = 2\mu_2 = 1$ (see Figures 4b and 6b).

ρ_2	**Average Sojourn Time (V_2)**					
	Preemptive EDF ($\rho_2^{\text{sat}} = 0.765$)			**Non-Preemptive EDF ($\rho_2^{\text{sat}} = 0.770$)**		
	Simulation	**Analytic**	**Error (%)**	**Simulation**	**Analytic**	**Error (%)**
0.1	3.3681	3.3432	-0.7383	3.3052	3.3003	-0.1483
0.2	3.9619	3.9355	-0.6672	3.8682	3.8794	0.2895
0.3	4.8125	4.7827	-0.6196	4.7101	4.7053	-0.1019
0.4	6.1223	6.0947	-0.4500	6.0010	5.9778	-0.3866
0.5	8.4701	8.3989	-0.8411	8.1833	8.1937	0.1271
0.6	13.635	13.5012	-0.9813	13.0832	13.0182	-0.4968
0.7	35.3398	34.4100	-2.6310	31.4836	31.2030	-0.8913
	Average relative error = -0.9898%			Average relative error = -0.2297%		
	RMSE = 1.2053%			RMSE = 0.4347%		

for preemptive EDF in Table 3 are negative (i.e., the values of the analytical results are less than the respective simulation ones). The main reason for such errors is that all the approximated (analytical) values of $\alpha_d^{\text{P-EDF}}$ for a type II system are more than the corresponding simulation values and the corresponding relative errors are positive, as can be observed in Table 1. As also indicated above, due to the complexity of approximating an infinite-capacity queue, the relative errors for the average sojourn times may increase when the class-2 input rate approaches to $\rho_2^{\text{sat}} = 0.46$. However, the average relative error and RMSE for the presented data points have been shown at the bottom of Tables 3 and 4. As can also be seen in the tables, while the accuracies of the results are in the acceptable range for most applications, they are more accurate for the non-preemptive EDF policy.

CONCLUDING REMARKS AND FUTURE WORK

In this paper, a method for approximating the performance of a two-class $M/M/1$ system has been presented. 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. An arriving prioritized class-1 job immediately preempts the serving class-2 job, where the preempted job can resume its service when the system becomes empty of class-1 jobs. The system has been solved for real-time jobs with deadlines until the end of service and two preemptive and non-preemptive models 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 the authors' 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. An approximation method for a two-class system has been proposed, which, it is believed, is quite accurate and, at the same time, 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 [26].

Some future work to continue this study includes extending the presented approach to other patterns of input traffic, multi-server systems and more general distribution of relative deadlines. Likewise, the authors intend to use the introduced method in the design of some embedded real-time systems with multi-class traffics.

REFERENCES

1. Bernat, G., Burns, A. and Llamosi, A. "Weakly hard real-time systems", *IEEE Transactions on Computers*, **50**(4), pp 308-321 (2001).
2. EN 50170 "General purpose field communication system", In *European Standard*, CENELEC (1996).
3. *CAN-CIA*, CAN Specification 2.0, Part B, http://www.cancia.org/downloads/ciaspecifications (1992).
4. Livani, M.A. and Kaiser, J. "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 (1998).
5. Dolev, S. and Keizelman, A. "Non-preemptive real-time scheduling of multimedia tasks", *Real-Time Systems*, **17**(1), pp 23-39 (1999).
6. Liu, C.L. and Layland, J.W. "Scheduling algorithms for multiprogramming in a hard real-time environment", *Journal of the ACM*, **20**(1), pp 46-61 (1973).
7. George, L., Muhlethaler, P. and Rivierre, N. "Optimality and non-preemptive real-time scheduling revisited", *Rapport de Recherche RR-2516*, INRIA, Le Chesnay Cedex, France (1995).
8. George, L., Rivierre, N. and Spuri, M. "Preemptive and non-preemptive real-time uni-processor scheduling", *Rapport de Recherche RR-2966*, INRIA, Le Chesnay Cedex, France (1996).
9. Towsley, D. and Panwar, S.S. "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 (1990).
10. Towsley, D. and Panwar, S.S. "Optimality of the stochastic earliest deadline policy for the G/M/c queue serving customers with deadlines", *Proceedings of the Second ORSA Telecommunications Conference* (1992).
11. May, M., Bolot, J., Jean-Marie, A. and Diot, C. "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, **3**, pp 1385-1394 (1999).
12. Baccelli, F., Boyer, P. and Hebuterne, G. "Single-server queues with impatient customers", *Advanced Applied Probability*, **16**, pp 887-905 (1984).
13. Barrer, D.Y. "Queueing with impatient customers and ordered service", *Operational Research*, **5**, pp 650-656 (1957).
14. Boxma, O.J. and Wall, P.R. "Multiserver queues with impatient customers", *Proceedings of the 14th International Teletraffic Congress*, **14**, Antibes, France, pp 743-756 (1994).
15. Brandt, A. and Brandt, M. "On the M(n)/M(n)/s queue with impatient calls", *Performance Evaluation*, **35**, pp 1-18 (1999).
16. Brandt, A. and Brandt, M. "Asymptotic results and a Markovian approximation for the M(n)/M(n)/s+GI system", *Queuing Systems*, **41**, pp 73-94 (2002).
17. Daley, D.J. "General customer impatience in queue GI/G/1", *Journal of Applied Probability*, **2**, pp 186-205 (1965).
18. Movaghar, A. "On queuing with customer impatience until the beginning of service", *Queuing Systems*, **29**, pp 337-350 (1998).
19. Movaghar, A. "On queuing with customer impatience until the end of service", *Stochastic Models*, **22**, pp 149-173 (2006).
20. Palm, C. "Methods for judging the annoyance caused by congestion", *Tele*, **2**, pp 1-20 (1953).
21. Zhao, W. and Stankovic, J.A. "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 (1989).
22. Leulseged, A. and Nissanke, N. "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 (2004). (LNCS 2968).

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

24. Hong, J., Tan, X. and Towsley, D. "A performance analysis of minimum axity and earliest deadline scheduling in a real-time system", *IEEE Transactions on Computers*, **38**(12), pp 1736-1744 (1989).

25. Kargahi, M. and Movaghar, A. "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 (2004).

26. Kargahi, M. and Movaghar, A. "A method for performance analysis of earliest-deadline-first scheduling policy", *Journal of Supercomputing, Springer*, **37**(2), pp 197-222 (2006).

27. Kargahi, M. and Movaghar, A. "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 (2005).

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

29. Lehoczky, J.P. "Real-time queuing theory", *Proceedings of the 17th IEEE Real-Time Systems Symposium*, Washington, D.C., USA, pp 186-195 (1996).

30. Lehoczky, J.P. "Using real-time queuing theory to control lateness in real-time systems", *Performance Evaluation Review*, **25**(1), pp 158-168 (1996).

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

32. Brandt, A. and Brandt, M. "On the two-class $M/M/1$ system under preemptive resume and impatience of the prioritized customers", *Queuing Systems*, **47**(1-2), pp 147-168 (2004).

33. Choi, B.D., Kim, B. and Chung, J. "$M/M/1$ queue with impatient customers of higher priority", *Queuing Systems*, **38**, pp 49-66 (2001).

34. Kruk, L., Lehoczky, J., Shreve, S. and Yeung, S.N. "Multiple-input heavy-traffic real-time queues", *Annals of Applied Probability*, **13**(1), pp 54-99 (2003).

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

36. Jaiswal, N., *Priority Queues*, Academic Press, New York (1968).

37. Miller, R.G. "Priority queues", *Annals of Mathematical Statistics*, **31**, pp 86-103 (1960).