Research Article

Analysis of the M/M/N/N Queue with Two Types of Arrival Process: Applications to Future Mobile Radio Systems

**Peter J. Smith,[1] Abdulla Firag,[1]
Pawel A. Dmochowski,[2] and Mansoor Shafi[3]**

[1] *Department of Electrical and Computer Engineering, University of Canterbury,
Christchurch 8140, New Zealand*
[2] *School of Engineering and Computer Science, Victoria University of Wellington,
Wellington 6011, New Zealand*
[3] *Telecom New Zealand, Wellington 6140, New Zealand*

Correspondence should be addressed to Abdulla Firag, abdulla.firag@canterbury.ac.nz

The queueing system considered is essentially a M/M/N/N queue where two types of users compete for the N resources. The users may have different arrival and service rates and are denoted as primary or secondary users. The primary users have priority access to the resources, and three levels of priority are considered: perfect priority, partial priority, and no priority. This system models the recently developed cognitive radio concept, a methodology that has been proposed for future mobile radio systems. In this context, the primary users have certain rights to use the resources, whereas the secondary users must make opportunistic use of the resources without impacting too much on the performance of the primary users. For all priority settings, the mean number of primary and secondary users is derived as are the blocking probabilities for both users. When no priority is given to the primary user, the system collapses to a truncated form of two independent M/M/∞ queues. The product form solution for this special case is known, and, here, these results are given in a novel, compact form. In the case of nonzero priority, the dropping probability for the secondary users is also derived.

1. Introduction

The M/M/N/N queue or Erlang loss system [1] and its variations have a long history in queueing theory and communications applications [2]. In this paper, we focus on an M/M/N/N queue with two types of arrival process, one with a higher priority than the other and the ability to displace a lower priority item from service and replace it with a higher

priority item. Although previous work on priority issues [3–6] and multiple arrival streams [7] have appeared, this variation does not appear to be available.

The motivation for this problem arises from the concept of cognitive radio (CR), a potential radio architecture that is receiving enormous interest in the communications literature as a possible direction for future mobile radio systems [8]. In this context, an arrival process is a sequence of demands from users to access a wireless radio channel, and a service period is the length of the call made by the user. In CR, one set of users are the so-called primary users (PUs), and they have certain rights to use the radio channel. For example, they might be cell-phone users using a service operator that has purchased the cell-phone spectrum. In contrast, the second set of users are called secondary users (SUs), and they do not have rights to use the channel. They must access the channel only when they do not impair the performance of the PUs beyond some allowed limit. In order to evaluate the possible coexistence of PUs and SUs, we consider the following performance metrics for both users: the mean number of users, blocking probabilities, and dropping probabilities. Although the paper is motivated by the CR example, it has applications to other queueing problems with the same type of competition between two sets of users.

Many applications of queueing theory have appeared in the CR literature. For example, the performance analysis of SUs in CR systems has been widely studied [9–16]. In [9], Simeone et al. perform SU-throughput analysis for a system with single SU and PU pairs, while [10] considers the general case with multiple secondary users, deriving throughput and transmission delay. In [11], Tang and Mark model the CR system by a multidimensional Markov process and derive the SU blocking probabilities and wait times in a system with a SU preemption queue. Preemptive priority queueing is also considered by [13], where system dwelling time, blocking, and forced termination probabilities are derived. In [14], the authors analyse the distribution of the SU queue length in a CR system with multiple PUs employing slotted transmission and myopic SU sensing. Primary user spectrum occupancy has been analysed in [17], where queueing theory is used to obtain a model of the occupancy of N subbands, with wireless channel measurements used to derive model parameters.

Finally, queueing analysis has also been utilised in developing scheduling policies and optimum resource allocation for CR systems, for example, [18, 19]. In [18], the authors propose an opportunistic scheduling which maximizes the SU throughput subject to maximum collision constraints with the PUs. In [19], the authors consider the problem of scheduling the sensing activities in a CR network.

The novel contributions of this work are as follows. By considering a two-dimensional M/M/N/N queue to model a CR system with N resources, we derive the mean number of PUs and SUs, their respective blocking probabilities, and the SU dropping probability. Unlike [13], which also investigates the latter two metrics, we consider a model where all N resources are accessible to the PUs as opposed to each being assigned (licensed) to a specific PU. We believe this to be more representative of typical communication system models, such as a cellular system. Furthermore, in contrast to previous work in [11, 13], we consider varying levels of access priority. In addition to the case of instantaneous PU priority, we also investigate the performance for equal and imperfect access priority. This models the scenario of imperfect PU detection by SUs, or of SUs operating outside the rules. We show that, in the case of an equal priority mode, the system simplifies to a known, truncated form of two independent M/M/∞ queues. We derive new, closed form expressions for the steady-state probabilities, and thus the mean number of users and their blocking probabilities. All analytical results are verified via Monte Carlo simulations.

2. Cognitive Radio: A Future Mobile Radio System

Consider a mobile radio system with N resources. Traditionally, these resources might be frequency channels, but they could also be codes, antennas, and so forth, [20]. We assume that the PUs have priority access to these resources. In the context of CR systems [8], these users belong to a system which may have purchased the rights to use this part of the radio spectrum. In contrast, the SUs are allowed to make some kind of opportunistic use of the resources when they are not allocated to PUs. In the absence of any SUs, the PUs are assumed to occupy resources following the classical Erlang loss model or M/M/N/N queue [1]. This model assumes that the PUs instigate calls with an independent and identically distributed (iid) exponential length which arrive in a Poisson process at the base station (BS). It also assumes that only N calls can be handled concurrently, and blocked calls are cleared. For the PUs, we assume a call arrival rate of λ_p and a mean call length of $1/\mu_p$. This gives a traffic level of $\rho_p = \lambda_p/\mu_p$. The SUs also arrive as a Poisson process with rate λ_s have iid exponential call lengths of mean $1/\mu_s$ and a traffic level of $\rho_s = \lambda_s/\mu_s$. There are three modes of operation considered here for the SUs.

2.1. Secondary Users Cleared (SUC)

In this mode, a SU call is immediately dropped if all N resources are in use, at least one resource is taken by a SU and a PU call arrives. Here, PU calls are given instantaneous priority. This corresponds to the most likely implementation of a CR system where SUs are not allowed access to a resource whenever the PU is using it.

2.2. Secondary User Equality (SUE)

In this mode, the PU and SU calls have equal priority. Once a resource is taken by either a PU or a SU, then the resource is kept until the completion of the call. This may model a CR system where the SU arrival rate is low enough to be accepted into the system with an allowable loss of performance to the PUs. Alternatively, it can model the worst case scenario where the SUs are supposed to make room for PUs but fail to do so, either because they fail to recognize the arrival of the PU or because they are operating outside the rules. In this scenario, it is straight forward to identify the queue as a truncated form of two independent M/M/∞ queues. Hence, a simple product form solution exists.

2.3. Secondary Users Partially Cleared (SUPC)

In this mode, a SU call is dropped with some probability if all N resources are in use; at least, one resource is taken by a SU and a PU call arrives. This is the intermediate scenario between SUC and SUE, where PU calls are given imperfect priority. We assume that the SUs are supposed to make room for PUs but fail to do so, either because they fail to recognize the arrival of the PU or because they are operating outside the rules. The first possibility is of more interest, since the SUs must monitor the wireless channel to recognize the arrival of a PU. If the channel is poor, the PU may not be recognized. Hence, there is some probability that a SU will not detect the PU call.

The range of models, from SUC to SUE, correspond to a variety of actual implementations. SUC models the "perfect case" where a SU never uses a resource if the PU is already
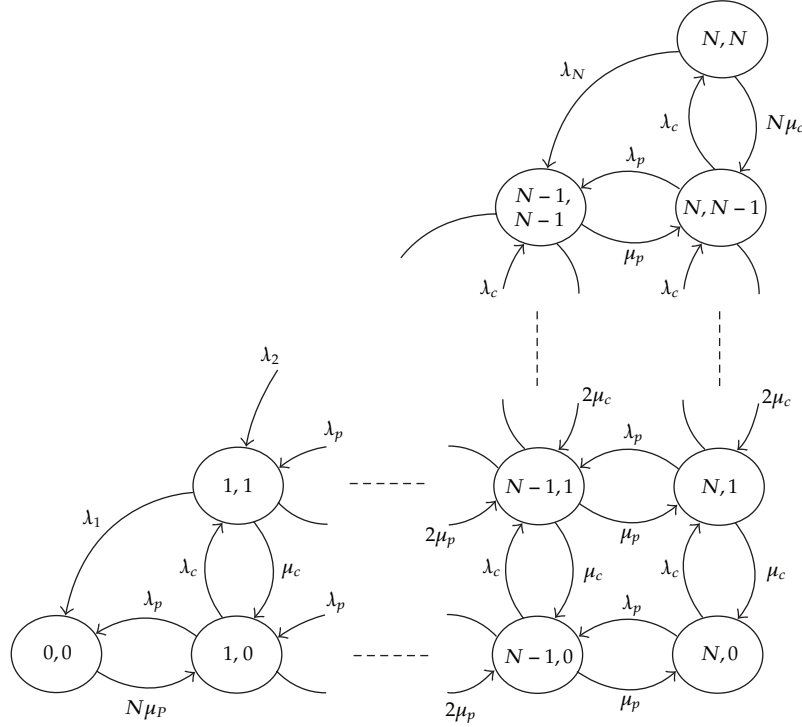
Figure 1: Two-dimensional queue for the CR system. Circles labelled i, j represent states, and possible transitions are given by arrows with the flow rates also marked.

using it. Here, there is instantaneous communication between PU and SU. In SUE mode, we have the opposite case where there is no communication, or the SU does not obey the rules. The intermediate SUPC cases model imperfect communication or imperfect spectrum sensing by the SU. Any real implementation or protocol will fall somewhere in this range.

3. State Space Model and Performance Metrics

To cater for the two types of user in the system described in Section 2, the state space of the traditional M/M/N/N queue, given by $\{0, 1, \ldots, N\}$, is enlarged to $\{(i, j) \mid 0 \leqslant i \leqslant j \leqslant N\}$ [7]. State (i, j) means that j resources are used by the SU, and the PU is using $N - i$ resources. This two-dimensional queue is illustrated in Figure 1. The main features of the queue are described below.

In Figure 1, horizontal flows to the right and left represent departures of PUs and arrivals of Pus, respectively. Vertical flows up and down represent the arrival and departure of SUs, respectively. The important region of the queue is in the upper diagonal, the (i, i) states, which represents a full system. It is the behaviour of these (i, i) states that models the relationship between the SUs and the PUs. In particular, consider what happens when a PU arrives and the system is in state (i, i), $i > 0$. In SUC mode, the arrival of the PU will result in the dropping of a SU call, and state (i, i) moves to state $(i - 1, i - 1)$ at rate λ_p. Hence, $\lambda_i = \lambda_p$. In SUE mode, the SUs always fail to relinquish their resources, and the PU is blocked. Here, $\lambda_i = 0$. Intermediate cases (SUPC mode) where the SU has imperfect ability to drop the

resource are handled by setting λ_i in the region $(0, \lambda_p)$. We propose the flow rate $\lambda_i = (1 - f^i)\lambda_p$ for this scenario. This models the situation where the i SUs all have probability f of failing to recognize the PU arrival, and all SUs operate independently. Although this model is used in the numerical results shown in Section 5, the analysis is general and holds for any $\lambda_i \in (0, \lambda_p)$. The general case is defined by $\lambda_i = (1 - q_i)\lambda_p$, where q_i represents the probability that all i SUs fail to relinquish a resource. Note that all three modes of operation can be defined in terms of the value of q_i. In SUC mode, $q_i = 0$ for $i = 1, 2, \ldots, N$. In SUE mode, $q_i = 1$ for $i = 1, 2, \ldots, N$ and in SUPC mode, $0 < q_i < 1$ for $i = 1, 2, \ldots, N$.

3.1. The Steady-State Probabilities

The queue in Figure 1 can be solved in the usual way by solving the flow balance equations [1]. Let $p_{i,j}$ denote the steady-state probability of state (i, j). The analysis proceeds by solving "flow in = flow out" for each of the $(N+1)(N+2)/2$ states. Although there are $(N+1)(N+2)/2$ states, there are only seven types of flow balance equation. These are identified directly from the state space diagram in Figure 1 and are enumerated below.

For state $(0, 0)$, the flow balance equation is

$$p_{0,0} N \mu_p = p_{1,1} \lambda_1 + p_{1,0} \lambda_p. \tag{3.1}$$

For states $(i, 0)$, where $1 \leqslant i \leqslant N - 1$, the flow balance equations are

$$p_{i,0}\big((N - i)\mu_p + \lambda_p + \lambda_s\big) = p_{i-1,0}(N - i + 1)\mu_p + p_{i,1}\mu_s + p_{i+1,0}\lambda_p. \tag{3.2}$$

For state $(N, 0)$, the flow balance equation is

$$p_{N,0}\big(\lambda_p + \lambda_s\big) = p_{N,1}\mu_s + p_{N-1,0}\mu_p. \tag{3.3}$$

For states (i, j), where $1 \leqslant j \leqslant N - 2$ and $j + 1 \leqslant i \leqslant N - 1$, the flow balance equations are

$$p_{i,j}\big((N - i)\mu_p + j\mu_s + \lambda_p + \lambda_s\big) = p_{i-1,j}(N - i + 1)\mu_p \\ + p_{i,j+1}(j + 1)\mu_s + p_{i+1,j}\lambda_p + p_{i,j-1}\lambda_s. \tag{3.4}$$

For states (i, i), where $1 \leqslant i \leqslant N - 1$, the flow balance equations are

$$p_{i,i}\big((N - i)\mu_p + i\mu_s + \lambda_i\big) = p_{i+1,i}\lambda_p + p_{i,i-1}\lambda_s + p_{i+1,i+1}\lambda_{i+1}. \tag{3.5}$$

For states (N, j), where $1 \leqslant j \leqslant N - 1$, the flow balance equations are

$$p_{N,j}\big(j\mu_s + \lambda_p + \lambda_s\big) = p_{N-1,j}\mu_p + p_{N,j+1}(j + 1)\mu_s + p_{N,j-1}\lambda_s. \tag{3.6}$$

For state (N, N), the flow balance equation is

$$p_{N,N} N \mu_s = p_{N,N-1}\lambda_s + p_{N-1,N-1}\lambda_N. \tag{3.7}$$

The steady-state probabilities sum to unity so that we also have

$$\sum_{i=0}^{N}\sum_{j=0}^{i}p_{i,j} = 1. \tag{3.8}$$

The flow balance equations in (3.1)–(3.7) and the normalization equation, (3.8), are all conveniently written in vector form. For example, (3.1) can be written as

$$\begin{pmatrix} N\mu_p & -\lambda_p & -\lambda_1 & 0 & \cdots & 0 \end{pmatrix}\begin{pmatrix} p_{0,0} \\ p_{1,0} \\ p_{1,1} \\ p_{2,0} \\ \vdots \\ p_{N,N} \end{pmatrix} = 0. \tag{3.9}$$

Combining all these vector equations, except (3.7), the result can be given as

$$\mathbf{Mp} = \mathbf{1}_0, \tag{3.10}$$

where $\mathbf{1}_0$ is a $(N+1)(N+2)/2 \times 1$ vector given as $\mathbf{1}_0 = (1,0,\ldots,0)^T$, $(\cdot)^T$ represents transpose, $\mathbf{p}$ is a $(N+1)(N+2)/2 \times 1$ vector of state probabilities, and $\mathbf{M}$ is a $(N+1)(N+2)/2 \times (N+1)(N+2)/2$ matrix, where the rows contain the coefficients of the steady-state probabilities in (3.8) and (3.1)–(3.6). Note that the first row contains the coefficients $1,1,\ldots,1$ from (3.8). The steady-state probabilities can now be computed using (3.10) as

$$\mathbf{p} = \mathbf{M}^{-1}\mathbf{1}_0. \tag{3.11}$$

Using the result in (3.11), the desired performance metrics can be calculated as below.

3.2. Mean Numbers of Users

The mean number of PU calls, $E(N_p)$, in the system can be calculated as

$$E(N_p) = \sum_{i=0}^{N}\sum_{j=0}^{i}(N-i)p_{i,j}, \tag{3.12}$$

where $(N-i)$ is the number of PU calls in the system if the system is in the state (i,j). The mean number of SU calls, $E(N_s)$, in the system can be given as

$$E(N_s) = \sum_{i=0}^{N}\sum_{j=0}^{i}jp_{i,j}, \tag{3.13}$$

where j is the number of SU calls in the system if the system is in state (i,j).

3.3. Blocking Probabilities

A single expression for the probability of blocking for a PU is given by

$$P(\text{PU blocking}) = p_{0,0} + \sum_{i=1}^{N} p_{i,i} q_i. \tag{3.14}$$

In SUC mode, $q_i = 0$ for $i = 1, 2, \ldots, N$, so the blocking probability is $p_{0,0}$ as a PU is only blocked when N PUs are already present. In SUE mode, $q_i = 1$ for $i = 1, 2, \ldots, N$, so the blocking probability is $\sum_{i=0}^{N} p_{i,i}$, as a PU is blocked whenever the system is full. Finally, in SUPC mode, the blocking probability is given by (3.14) since blocking occurs when N PUs are present or when the system is full and all the SUs in operation fail to sense the PU arrival.

For a SU, blocking always occurs when the system is full. Hence,

$$P(\text{SU blocking}) = \sum_{i=0}^{N} p_{i,i}. \tag{3.15}$$

3.4. Dropping Probabilities

A PU call is never dropped, but a SU call can be dropped in SUC or SUPC modes if the system is full, at least one SU is in operation and a PU call arrives. The probability that a SU call will be dropped can be calculated as

$$
\begin{aligned}
P(\text{SU dropping}) &= \lim_{T \to \infty} \left(\frac{\text{\# SU calls dropped in time } T}{\text{\# SU calls accepted into system in time } T} \right) \\
&= \lim_{T \to \infty} \left(\frac{(1/T)(\text{\# SU calls dropped in time } T)}{(1/T)(\text{\# SU calls accepted into system in time } T)} \right) \\
&= \frac{\text{SU call dropping rate}}{\text{SU call acceptance rate}} \\
&= \frac{\sum_{i=1}^{N} p_{i,i} \lambda_i}{\lambda_s (1 - P(\text{SU blocking}))} \\
&= \frac{\sum_{i=1}^{N} p_{i,i} \lambda_i}{\lambda_s \left(1 - \sum_{i=0}^{N} p_{i,i}\right)}.
\end{aligned} \tag{3.16}
$$

3.5. Model Issues and Extensions

The state-space model in Section 3 employs exponential call lengths and interarrival times for both SU and PU. Of course, there are real systems where these assumptions will not lead to accurate models. However, measured data from an urban cellular network, collected by the authors in 2010, suggests that these simple models are surprisingly accurate when the PU is a cellular system. In addition, many more complicated traffic models result in simulation

studies and do not allow analytic solutions. Furthermore, such complex models will often lose generality and potential application in other areas. Although data can be measured for existing systems, superimposing SUs on such a system still requires a SU traffic model since this data is not currently available. Hence, modeling PU-SU coexistence is an important issue.

The models used in Section 3 can be extended in many ways. Two important extensions for the future concern resource sharing and varying call types. In resource sharing, SUs and PUs may occasionally be able to use the same resource without causing interference. For example, the PU and SU may be separated by a building or a hill which blocks the interference. This can be modeled by an extra dimension in the state-space model. Hence, instead of state (i, j) denoting j SUs, we would have (i, j_1, j_2), where j_1 is the number of SUs using unused resources and j_2 SUs are sharing a resource with the PU. Similarly, varying call types can be handled by increasing the system dimension. If both PU and SU use two types of call, type A and type B, then state (i, j_1, j_2) can be generalized to the six-dimensional state $(i^A, j_1^A, j_2^A, i^B, j_1^B, j_2^B)$. Such extensions are beyond the scope of the current paper.

4. Closed form Analysis for the SUE Mode

A closed-form expression for (3.11) is difficult to obtain for SUC and SUPC. Hence, for these cases, the performance metrics are computed numerically by solving (3.11) and substituting the steady-state probabilities in (3.12)–(3.16). However, in SUE mode, a closed form expression for the steady-state probabilities is well known and is a direct application of Theorem 6.1 in [7, page 152]. The system in SUE mode behaves like two independent M/M/∞ queues truncated so that the total number of calls does not exceed N. For an M/M/∞ queue, the steady-state probability is $p_n = (1 - \rho)\rho^n/n!$ for $n = 0, 1, \ldots$, where ρ is the traffic level. Hence, in SUE mode, applying the product form solution gives

$$p_{i,j} = \frac{1}{D_1} \left(\frac{(1 - \rho_p)\rho_p^{N-i}}{(N - i)!} \times \frac{(1 - \rho_s)\rho_s^j}{j!} \right), \tag{4.1}$$

where $D_1 = \sum_{k=0}^{N} \sum_{l=0}^{k} ((1 - \rho_p)(1 - \rho_s)\rho_p^{N-k}\rho_s^l)/(N - k)!l!$. Simplifying (4.1) gives

$$p_{i,j} = \frac{1}{D} \frac{\rho_p^{N-i}}{(N - i)!} \frac{\rho_s^j}{j!}, \tag{4.2}$$

where

$$D = \sum_{k=0}^{N} \sum_{l=0}^{k} \frac{\rho_p^{N-k}}{(N - k)!} \frac{\rho_s^l}{l!}. \tag{4.3}$$

The state space probabilities in (4.2) can be simplified further by rewriting D in (4.3) as

$$D = \sum_{r=0}^{N} \sum_{t=0}^{N-r} \frac{\rho_p^{N-r-t}}{(N - r - t)!} \frac{\rho_s^t}{t!}, \tag{4.4}$$

using the change of index, $t = l$ and $r = k - t$. Rewriting (4.4) gives

$$
\begin{aligned}
D &= \sum_{r=0}^{N} \frac{\rho_p^{N-r}}{(N-r)!} \sum_{t=0}^{N-r} \binom{N-r}{t} \left(\frac{\rho_s}{\rho_p}\right)^t \\
&= \sum_{r=0}^{N} \frac{\rho_p^{N-r}}{(N-r)!} \left(1 + \frac{\rho_s}{\rho_p}\right)^{N-r} \\
&= \frac{e^\nu}{N!} \Gamma(N+1, \nu),
\end{aligned}
\tag{4.5}
$$

where $\nu = \rho_p + \rho_s$ and $\Gamma(\alpha, z) = \int_z^\infty e^{-t} t^{\alpha-1} dt$ is the upper incomplete gamma function also defined by the series expansion $\Gamma(\alpha, z) = (\alpha - 1)! e^{-z} \sum_{m=0}^{\alpha-1} (z^m / m!)$ [21, page 899, equation 8.352] for integer values of $\alpha > 0$. Now, using this simpler expression for D, the average number of PU and SU calls in the system and the blocking probabilities of the PUs and the SUs can also be expressed in a compact closed form.

The mean number of PU calls, $E(N_p)$, is given by

$$
\begin{aligned}
E(N_p) &= \sum_{i=0}^{N} \sum_{j=0}^{i} (N - i) p_{i,j} \\
&= \frac{1}{D} \sum_{i=0}^{N} \sum_{j=0}^{i} \frac{(N - i)}{(N - i)! j!} \rho_p^{N-i} \rho_s^j.
\end{aligned}
\tag{4.6}
$$

Using the same change of indices as in (4.4), $t = j$ and $r = i - t$, the numerator of (4.6) can be rewritten as

$$
\sum_{r=0}^{N-1} \frac{1}{(N-1-r)!} \rho_p^{N-r} \left(1 + \rho_p^{-1} \rho_s\right)^{N-1-r} = \frac{\rho_p e^\nu \, \Gamma(N, \nu)}{(N-1)!},
\tag{4.7}
$$

so that (4.6) becomes

$$
E(N_p) = \frac{N \rho_p \Gamma(N, \nu)}{\Gamma(N+1, \nu)}.
\tag{4.8}
$$

Similarly, the mean number of SU calls, $E(N_s)$, is given by

$$
\begin{aligned}
E(N_s) &= \sum_{i=0}^{N} \sum_{j=0}^{i} j p_{i,j} \\
&= \frac{1}{D} \sum_{i=0}^{N} \sum_{j=1}^{i} \frac{\rho_p^{N-i} \rho_s^j}{(N-i)!(j-1)!}.
\end{aligned}
\tag{4.9}
$$

Again, the numerator of (4.9) can be rewritten using a change of indices as

$$\sum_{r=1}^{N}\sum_{t=0}^{N-r} \frac{\rho_p^{N-r-t}\rho_s^{t+1}}{(N-r-t)!t!} = \sum_{r=1}^{N} \frac{\rho_p^{N-r}\rho_s}{(N-r)!}\left(1+\rho_p^{-1}\rho_s\right)^{N-r}$$

$$= \frac{\rho_s}{N!}\left(e^\nu\, \Gamma(N+1,\nu) - \nu^N\right), \tag{4.10}$$

so that (4.9) can be simplified to give

$$E(N_s) = \rho_s - \frac{\rho_s\nu^N}{e^\nu\, \Gamma(N+1,\nu)}. \tag{4.11}$$

In SUE mode, a PU or a SU will be blocked from the system if the user arrives when the system is full, that is, when the system is in one of the states (i, i) for $i = 0, 1, \ldots, N$. In this case, the blocking probabilities of the PUs and the SUs are equal and can be calculated as

$$P(\text{PU blocking}) = P(\text{SU blocking})$$

$$= \sum_{i=0}^{N} p_{i,i} = \frac{1}{D}\sum_{i=0}^{N}\frac{\rho_p^{N-i}\rho_s^i}{(N-i)!i!} \tag{4.12}$$

$$= \frac{\nu^N}{e^\nu\, \Gamma(N+1,\nu)}.$$

Finally, we recall that in SUE mode the system will not drop any SU calls and $P(\text{SU dropping}) = 0$.

5. Results

In this section, the results produced in this paper are verified, and the impact of the system parameters on the performance metrics is explored. For all the results given here, we use $N = 20$, $\mu_p = \mu_s$ and $\lambda_p = 1$. Note that we deliberately use $\rho_p = 20$ in the results which represents a very high loading by the PU. With $\rho_p = 20$, we have $\mu_p = 1/20$, and so the PU arrival rate, $\lambda_p = 1$, is equal to the total service rate, $N \times \mu_p = 1$. The results, therefore, represent a PU system that is already heavily utilized and this is a good test of the ability of the SUs to coexist with the PUs. Figure 2 gives the expected number of users in the system operating in SUC mode when $\rho_p = 20$. The expected number of PU calls remains fixed as expected since the SU calls do not effect the PU performance. As the arrival rate of the SUs increases, the system tends to become full, and $E(N_s) + E(N_p)$ becomes close to N. Figure 2 also shows that the simulations are in good agreement with the theoretical results obtained using (3.12) and (3.13).

Figure 3 gives the dropping and blocking probabilities of the SU and PU users in SUC mode when $\rho_p = 20$. The figure shows that all the simulations are in good agreement with the theoretical results obtained using (3.14)–(3.16). The blocking probability of the PU remains fixed since, as in Figure 2, the SU calls do not affect the PU performance. The figure also shows the expected increase in blocking and dropping of the SU calls as
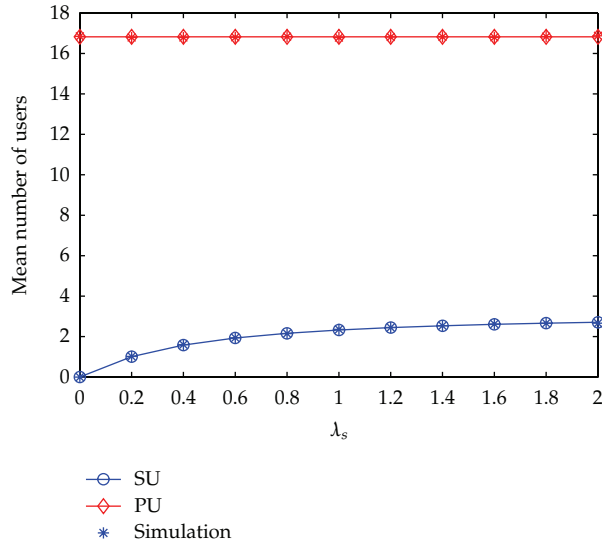
Figure 2: Expected number of users in SUC mode versus λ_s. The system parameters are $N = 20$, $\mu_p = \mu_s$ and $\lambda_p = 1$.
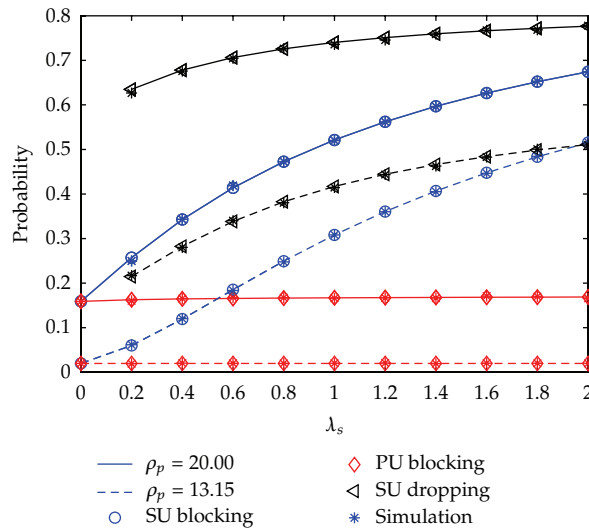

Figure 3: PU and SU blocking probabilities and SU dropping probability versus λ_s in SUC mode. The system parameters are $N = 20$, $\mu_p = \mu_s$ and $\lambda_p = 1$.

the arrival rate of the SU increases. Also shown for comparison is the same system with $\rho_p = 13.15$ giving a 2% blocking probability for the PU. This is a standard value considered in the telecommunications industry, and the reduced PU traffic has the effect of lowering the SU blocking and dropping rates. Also, the SU dropping probability overlaps the blocking probability. This is a general trend, and as the SU performance improves, the dropping probability drops faster than the blocking probability until it is smaller over a wide range of λ_s values.
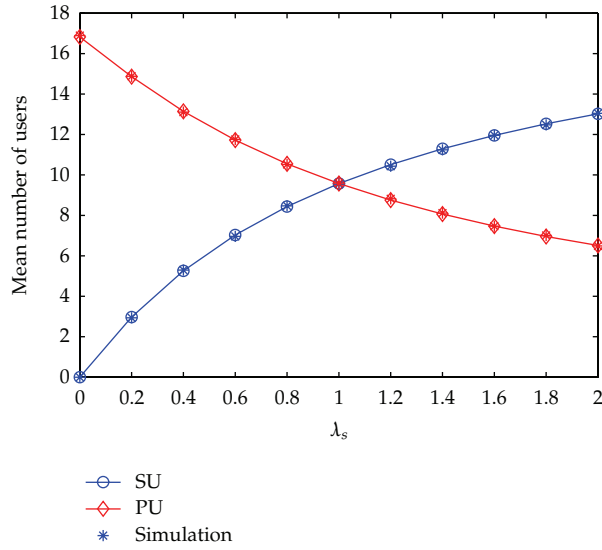
Figure 4: Expected number of users in SUE mode versus λ_s. The system parameters are $N = 20$, $\mu_p = \mu_s$ and $\lambda_p = 1$.
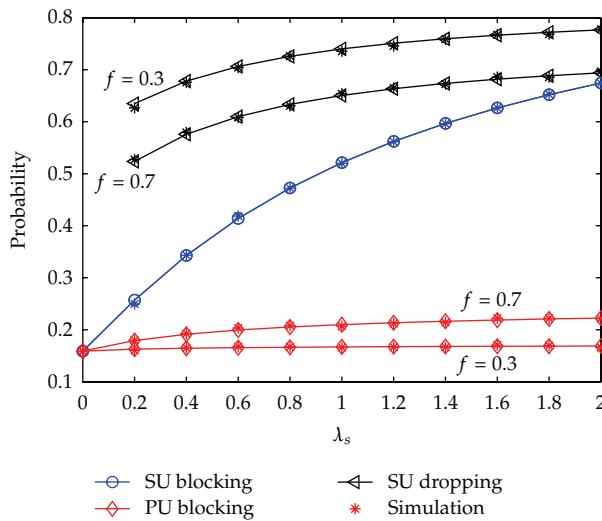

Figure 5: Blocking and dropping probabilities versus λ_s in SUPC mode. The systems parameters are $N = 20$, $\mu_p = \mu_s$, $\lambda_p = 1$ and $f = 0.3, 0.7$.

Next, we consider the SUE mode and verify the results obtained in Section 4. Figure 4 gives the expected number of calls/users in the system when $\rho_p = 20$. As expected, the number of SU calls in the system increases, and the number of PU users decreases as the arrival rate of the SU calls increases. The figure also shows that the simulations are in good agreement with the theoretical results, thus, validating the closed form expressions in (4.8) and (4.11). Figure 5 gives the blocking probabilities for the SU and PU users when $\rho_p = 20$. The figure shows that both blocking and dropping probabilities increase with the arrival rate

of the SU as expected. Comparing the PU blocking probabilities with Figure 3, we see that increasing f from zero in Figure 3 to 0.3 and 0.7 in Figure 5 has a relatively small effect. Hence, the PU may be able to accept SUs with only limited ability to relinquish a resource, say only 70% of the time ($f = 0.3$). The effect on the SU dropping rate is more marked, with smaller values of f boosting the dropping rate. The SU blocking probability in Figure 5 is identical to the equivalent curve in Figure 3 for SUC mode. This equivalence is due to the special case of the parameter values which satisfy $\mu_p = \mu_s$. For this scenario, SU blocking is the same for SUC, SUE, and SUPC since time to completion of any call in a (i, i) state is statistically identical as $\mu_p = \mu_s$. Hence, P(SU blocking) is independent of f, and there is only one curve for both $f = 0.3$ and $f = 0.7$. Figure 5 suggests that the PU may be able to accept the presence of SUs but the SUs themselves suffer a large performance degradation as λ_s increases.

6. Conclusions

This paper has considered the M/M/N/N queue with two types of arrivals process and various levels of priority. For all priority settings, we have derived the mean number of primary and secondary users and the blocking probabilities for both users. When no priority is given to the primary user, results for the steady-state probabilities are available. These were simplified and given in a novel, compact form which also leads to simple results for the performance metrics. To the best of our knowledge, these results are new. In the case where priority is given to the primary users, then the secondary user can be forced to drop a resource and the dropping probability was also derived. These results have important applications to CR systems, an architecture that has been proposed for future mobile radio systems. For these radio systems, the primary user will only accept a certain performance degradation due to the presence of the secondary users. This work allows this degradation to be assessed in terms of the mean number of primary users and their blocking probability. It also evaluates the performance of the secondary users. The closed form results for the no-priority case are of interest in their own right and also act as a lower bound to the primary user performance in the presence of CRs.

References

[1] D. Gross and C. M. Harris, *Fundamentals of Queueing Theory*, Wiley Series in Probability and Statistics: Texts and References Section, John Wiley & Sons, New York, NY, USA, 3rd edition, 1998.

[2] A. K. Erlang, "Solution of some problems in the theory of probabilities of significance in automatic telephone exchanges," *The Post Office Electrical Engineers' Journal*, vol. 10, pp. 189–197, 1918.

[3] A. Cobham, "Priority assignment in waiting line problems," *Journal of the Operations Research Society of America*, vol. 2, no. 1, pp. 70–76, 1954.

[4] H. R. Gail, S. L. Hantler, and B. A. Taylor, "Analysis of a nonpreemptive priority multiserver queue," *Advances in Applied Probability*, vol. 20, no. 4, pp. 852–879, 1988.

[5] D. McMillan, "Delay analysis of a cellular mobile priority queueing system," *IEEE/ACM Transactions on Networking*, vol. 3, no. 3, pp. 310–319, 1995.

[6] P. Zhu, J. Li, and X. Wang, "Scheduling model for cognitive radio," in *Proceedings of the 3rd International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CrownCom '08)*, Singapore, May 2008.

[7] J. H. Dshalalow, Ed., *Frontiers in Queueing: Models and Applications in Science and Engineering*, Probability and Stochastics Series, CRC Press, Boca Raton, Fla, USA, 1st edition, 1997.

[8] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 201–220, 2005.

[9] O. Simeone, Y. Bar-Ness, and U. Spagnolini, "Stable throughput of cognitive radios with and without relaying capability," *IEEE Transactions on Communications*, vol. 55, no. 12, pp. 2351–2360, 2007.

[10] F. Borgonovo, M. Cesana, and L. Fratta, "Throughput and delay bounds for cognitive transmissions," in *IFIP International Federation for Information Processing*, vol. 265 of *Advances in Ad Hoc Networking*, pp. 179–190, 2008.

[11] S. Tang and B. L. Mark, "Analysis of opportunistic spectrum sharing with markovian arrivals and phase-type service," *IEEE Transactions on Wireless Communications*, vol. 8, no. 6, Article ID 5089995, pp. 3142–3150, 2009.

[12] H. P. Shiang and M. van der Schaar, "Queuing-based dynamic channel selection for heterogeneous multimedia applications over cognitive radio networks," *IEEE Transactions on Multimedia*, vol. 10, no. 5, Article ID 4542759, pp. 896–909, 2008.

[13] J. Heo, J. Shin, J. Nam, Y. Lee, J. Park, and H. S. Cho, "Mathematical analysis of secondary user traffic in cognitive radio system," in *Proceedings of the 68th IEEE Vehicular Technology Conference (VTC '08-Fall)*, pp. 1–5, Calgary, Canada, September 2008.

[14] A. Laourine, S. Chen, and L. Tong, "Queuing analysis in multichannel cognitive spectrum access: a large deviation approach," in *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM '10)*, pp. 1–9, San Diego, Calif, USA, March 2010.

[15] M. M. Rashid, M. J. Hossain, E. Hossain, and V. K. Bhargava, "Opportunistic spectrum access in cognitive radio networks: a queueing analytic model and admission controller design," in *Proceedings of the 50th IEEE Global Telecommunications Conference (GLOBECOM '07)*, pp. 4647–4652, Washington, DC, USA, November 2007.

[16] M. M. Rashid, M. J. Hossain, E. Hossain, and V. K. Bhargava, "Opportunistic spectrum scheduling for multiuser cognitive radio: a queueing analysis," *IEEE Transactions on Wireless Communications*, vol. 8, no. 10, Article ID 5288960, pp. 5259–5269, 2009.

[17] C. Ghosh, S. Pagadarai, D. P. Agrawal, and A. M. Wyglinski, "Queueing theory representation and modeling of spectrum occupancy employing radio frequency measurements," in *Proceedings of the IEEE 70th Vehicular Technology Conference Fall (VTC '09)*, pp. 1–5, Anchorage, Alaska, USA, September 2009.

[18] R. Urgaonkar and M. J. Neely, "Opportunistic scheduling with reliability guarantees in cognitive radio networks," *IEEE Transactions on Mobile Computing*, vol. 8, no. 6, Article ID 4782965, pp. 766–777, 2009.

[19] A. T. Hoang and Y. C. Liang, "Adaptive scheduling of spectrum sensing periods in cognitive radio networks," in *Proceedings of the 50th IEEE Global Telecommunications Conference (GLOBECOM '07)*, pp. 3128–3132, Washington, DC, USA, November 2007.

[20] A. Goldsmith, *Wireless Communications*, Cambridge University Press, 2005.

[21] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, Academic Press, San Diego, Calif, USA, 7th edition, 2007.