Wireless Networks 8, 597–605, 2002

A Homogeneous PCS network with Markov Call Arrival Process and Phase Type Cell Residence Time

ATTAHIRU SULE ALFA *
Department of Industrial and Manufacturing Systems Engineering, University of Windsor, Windsor, ON, N9B 3P4, Canada

WEI LI
Department of Electrical Engineering and Computer Engineering, University of Toledo, Toledo, OH 43606-3390, USA

Abstract. In this paper, the arrival of calls (i.e., new and handoff calls) in a personal communications services (PCS) network is modeled by a Markov arrival process (MAP) in which we allow correlation of the interarrival times among new calls, among handoff calls, as well as between these two kinds of calls. The PCS network consists of homogeneous cells and each cell consists of a finite number of channels. Under the conditions that both cell's residence time and the requested call holding time possess the general phase type (PH) distribution, we obtain the distribution of the channel holding times, the new call blocking probability and the handoff call failure probability. Furthermore, we prove that the cell residence time is PH distribution if and only if

- *the new call channel holding time* is PH distribution; or
- *the handoff call channel holding time* is PH distribution; or
- *the call channel holding time* is PH distribution;

provided that the requested call holding time is a PH distribution and the total call arrival process is a MAP. Also, we prove that the actual call holding time of a non-blocked new call is a mixture of PH distributions. We then developed the Markov process for describing the system and found the complexity of this Markov process. Finally, two interesting measures for the network users, i.e., the duration of new call blocking period and the duration of handoff call blocking period, are introduced; their distributions and the expectations are then obtained explicitly.

Keywords: personal communications services (PCS) network model, Markovian arrival process (MAP), phase type (PH) distribution, duration of call blocking period

1. Introduction

Personal communications services (PCS) network is a set of capabalities that allows some combination of terminal mobility, personal mobility and service profile management. The PCS service area is populated with base stations. The coverage area of a base station is called a cell. The development of analytically tractable models to compute teletraffic performance characteristics of mobile wireless cellular networks has been the thrust of recent works (see [5,8,31,32] etc.). For most of the existing cellular systems, the wireless calls are charged based on the call holding times, and these systems can be appropriately modeled with the exponential call holding time distribution [23,24]. However, for future PCS systems (especially the low-power PCS systems such as CT-2 [39], Dect [4], PACS [30]), flat-rate billing programs have been proposed. Thus, it is very important that we follow the wireline telephone network engineering approach, that is, use a more general distribution to present the call holding time distribution [6]. Not only does call holding time vary with the new applications, also the time a mobile user stays or dwells in a cell, which is called the *cell residence time*, will depend on the mobility of the customer, the geographic situation, and the hand-off scheme used, and therefore needs to be modeled as a random variable with more general distribution. All of the interesting performance measures in the PCS networks depends on the user's mobility, which in turn can be characterized by the cell residence time. Thus, in order to appropriately characterize the performance measures, such as new call blocking probabilities, handoff call blocking probability and channel holding times [8] etc., it is necessary to have a good mobility model for the cell residence time. By assuming that a cell has specific shape and combining specific distributions of speed and movement direction of a mobile user it then becomes possible to directly model the cell residence time as a random variable with more general probability distribution to capture the overall effects of the cellular shape and the user's mobility patterns. This approach has been adopted in the past by a few researchers. The appropriate probability distributions include the negative exponential distribution for cell residence time as reported in [11,12,16,33,34], the sum of negative exponential random variables [34], the sum of hyper-exponentials [31,32], or the hyper-Erlang distribution [5]. The authors in [6–8] considered the general distribution and particularly found the handoff rate formula [8] based on the input parameters of the new call arrival process, new call blocking probability and the handoff failure probability. The gamma

* Corresponding author.

distribution [40] was also used to model cell residence time. However, it does not have the required memoryless property and cannot be used in the Markov framework. Some other studies model mobility as a random walk [15]. An extension of this to characterizing user movement as a diffusion process is given in [35]. An overview of other mobility modeling techniques is given in [19]. One other relevant and recent studies reported in [17], in which the authors have simulated a cellular communication network using a variety of probability density functions as cell residence times.

In this paper, we model both the *cell residence time* and *requested call holding time* as general phase type (PH) distributions [29], respectively. The PH distributions have been widely used in queueing system to model service times. The special cases of the PH distribution include negative exponential distribution, the sum of negative exponential distribution, the sum of hyperexponentials and the hyper-Erlang distribution. We also consider a more general correlated arrival process for new calls and handoff calls. The arrival of calls, i.e., new and handoff calls, in the network are modeled by a continuous Markov arrival process MAP. This process allows correlation of the interarrival times among each new calls and each handoff calls, as well as between these two kinds of calls. The continuous time MAP was introduced by Neuts [28] as a versatile point process. The discrete analog of the MAP can be found in Alfa and Neuts [1]. Both the continuous and discrete time versions are Markov renewal processes. In [21], the authors considered a PCS network with Markovian call arrival process, split channel scheme and the exponential assumptions for other input random parameters. One special case of the MAP is the *Markov Modulated Poisson Process* (MMPP), which is a doubly stochastic Poisson process where the rate process is determined by the state of a continuous-time Markov chain and which has been extensively used to describe superposition of packet streams whose interarrival times are known to be correlated [14,38]. Our contribution is organized as follows. A detailed description of the parameters of the PCS network is given in the next section. In section 3, we discuss the channel holding times and prove the fact that the cell residence time is PH distribution if and only if *the new call channel holding time* is PH distribution; or *the handoff call channel holding time* is PH distribution; or *the call channel holding time* is PH distribution provided that the total call arrival process is a MAP and the requested call holding time is PH distribution. By transforming the situation of a cell into a Markov process and finding the explicit expression of the corresponding transition rate matrix in section 4 of the paper, we obtain the expression for the two key performance measures, i.e., new call blocking probability and the handoff call failure probability. In section 5 we prove that the actual call holding time, under the condition that the the new call is not blocked, is a SPH distribution introduced in [37]. Finally, two interesting measures for the network users, i.e., the duration of new call blocking period and the duration of handoff call blocking period, are introduced; their distributions and the expectations are then obtained explicitly.

2. Model description

In this paper, we consider a PCS network with priority given to the handoff calls and with general phase type cell residence time and general phase type requested call holding time (a valuable method to obtain the probability distribution for both of the above measures is given in [36]). Here, the system is assumed to be homogeneous [33], i.e., the underlying processes and parameters for all cells within the PCS networks are the same, so that all cells are statistically identical. Each cell has M channels assigned to it and can therefore support at most M calls simultaneously. A *cutoff priority scheme* [31] is used to give handoff calls priority over new calls. For this purpose, m $(0 \leqslant m < M)$ channels in each cell are reserved for calls that arrive to the cell as handoffs. Individual channels are not reserved, just the number. The overall effect is that new calls that arise in a cell will be blocked if more than $M - m$ channels are in use, while handoff calls will always be served if there are idle channels which are not in use. Some other detailed assumptions are as follows:

1. The vehicular mobility is characterized by the cell residence time of a vehicle in a cell, which is a random variable, $\mathbf{R}$, and is assumed to have a general phase type distribution with representation $(\boldsymbol{\alpha}, T)$ and with dimension r, i.e., $P(\mathbf{R} \leqslant x) = 1 - \boldsymbol{\alpha}\exp(Tx)\mathbf{e}$, where $\boldsymbol{\alpha}\mathbf{e} + \alpha_{r+1} = 1$, and $T\mathbf{e} + \mathbf{T}^0 = 0$ (see [29] for details).

2. The requested call holding time, say $\mathbf{H}$, of a new call (which is the duration of the requested new call connection to a PCS network for a new call and is also referred to as the uncumbered call holding time [6] or the unencumbered session time [31]) are assumed to be independent and identically distributed (i.i.d.) and possess general phase type distribution with representation $(\boldsymbol{\beta}, S)$ and with dimension h, i.e., $P(\mathbf{H} \leqslant x) = 1 - \boldsymbol{\beta}\exp(Sx)\mathbf{e}$, where $\boldsymbol{\beta}\mathbf{e} + \beta_{h+1} = 1$, and $S\mathbf{e} + \mathbf{S}^0 = 0$.

3. The call's total arrival process (including new call arrival process and handoff arrival process) to the PCS network are modeled by a continuous Markov arrival process (MAP) in which we allow correlation of the interarrival times among each new calls and each handoff calls, as well as among these two kinds of calls. More precisely, let C be the irreducible infinitesimal generator of a K-state Markov process, the sojourn time in state i is exponentially distributed with parameter λ_i, $i = 1, 2, \ldots, K$. At the end of the sojourn in the state i, three kinds of transitions can take place:

   (a) with probability $p_0(i,k)$, $k = 1, 2, \ldots, K, k \neq i$, a transition occurs to the state k, without any kinds of call arrivals;

   (b) with probability $p_{\mathrm{N}}(i,k)$, $k = 1, 2, \ldots, K$, a transition occurs to the state k, with a new call arrival;

   (c) with probability $p_{\mathrm{H}}(i,k)$, $k = 1, 2, \ldots, K$, a transition occurs to the state k, with a handoff call arrival;

where

$$\sum_{k\neq i}^{K} p_0(i,k) + \sum_{k=1}^{K} p_{\mathrm{N}}(i,k) + \sum_{k=1}^{K} p_{\mathrm{H}}(i,k) = 1,$$
$$1 \leqslant i \leqslant K.$$

Let $(C_0)_{i,i} = -\lambda_i$, $(C_0)_{i,k} = \lambda_i p_0(i,k)$, $(C_{\mathrm{N}})_{i,j} = \lambda_i p_{\mathrm{N}}(i,j)$ and $(C_{\mathrm{H}})_{i,j} = \lambda_i p_{\mathrm{H}}(i,j)$ for $i \neq k$ and $i, j, k = 1, 2, \ldots, K$, then $C = C_0 + C_{\mathrm{N}} + C_{\mathrm{H}}$. Denote by $\boldsymbol{\pi}$ the stationary probability vector of the generator C, i.e., $\boldsymbol{\pi}$ satisfies $\boldsymbol{\pi} C = 0$ and $\boldsymbol{\pi}\mathbf{e} = 1$, and $\mathbf{e}$ is a column vector of 1's, then the *new call arrival rate*, say λ_{N}, of the MAP is $\lambda_{\mathrm{N}} = \boldsymbol{\pi} C_{\mathrm{N}}\mathbf{e}$, and that the *handoff call arrival rate*, say λ_{H}, of the MAP is $\lambda_{\mathrm{H}} = \boldsymbol{\pi} C_{\mathrm{H}}\mathbf{e}$. For the special case when $K = 1$, the MAP is a Poisson process with rate λ_1 and consists of two independent Poisson processes with new call rate $\lambda_{\mathrm{N}} = \lambda_1 p_{\mathrm{N}}(1,1)$ and handoff call rate $\lambda_{\mathrm{H}} = \lambda_1 p_{\mathrm{H}}(1,1)$. When both matrices C_{N} and C_{H} are diagonal matrices, the MAP is an MMPP, which is a particularly useful class of non-renewal process. There are many specific examples of MAP in [25,27] and the references therein. For some related additional literatures see [1,13,26].

3. Channel holding times

Channel holding time is defined as the amount of time that a call occupies a channel in a particular cell (see [8,11,31]) and is an important quantity in teletraffic analysis of PCS networks and depends on the mobility of users, which could be characterized by the cell residence times. This quantity is needed to derive key network design parameters such as the new call blocking probability and the handoff call failure probability [16]. Bolotin [3] studied common-channel signaling systems and found that channel throughput drops significantly more under the actual measured call holding time distribution model than under an exponential requested call holding time distribution. Orlik et al. [31] studied the channel holding times with the sum of hyperexponential distribution for the cell residence time. Fang et al. [8] and Fang et al. [5] investigated the channel holding times with the hyper-Erlang distribution for the cell residence time and first gave some necessary and sufficient conditions for the channel holding time to be exponentially distributed. It is easy to show that both the sum of hyperexponential distribution and the hyper-Erlang distribution are the special cases of the phase type distribution. Also, from the discussion below, one will find that the use of phase type (PH) distribution has an advantage because of its matrix form. Secondly, there have been major advances in fitting phase type distributions to real data (see [2,20]).

3.1. Distribution of the channel holding times

Based on the assumption of the system and [29, theorem 2.2.3], we know that the residual cell residence time, say $\overline{\mathbf{R}}$, of a vehicle is a phase type with representation $(\overline{\boldsymbol{\alpha}}, T)$ and with dimension r, where $\overline{\boldsymbol{\alpha}} = \overline{\boldsymbol{\alpha}}(T + \mathbf{T}^0\boldsymbol{\alpha})$ and $\overline{\boldsymbol{\alpha}}\mathbf{e} = 1$; and similarly, the residual requested call holding time, say $\overline{\mathbf{H}}$, is a phase type with representation $(\overline{\boldsymbol{\beta}}, S)$ and with dimension h, where $\overline{\boldsymbol{\beta}} = \overline{\boldsymbol{\beta}}(S + \mathbf{S}^0\boldsymbol{\beta})$ and $\overline{\boldsymbol{\beta}}\mathbf{e} = 1$.

Define the *channel holding time of a new call* as the time period that a non-blocked new call spends in a cell and denote it as $\mathbf{S}_{\mathrm{N}}$. It is clear that $\mathbf{S}_{\mathrm{N}}$ is independent of the arrival process and it is the minimum of the residual cell residence time and the requested call holding time, i.e.,

$$\mathbf{S}_{\mathrm{N}} = \min\{\overline{\mathbf{R}}, \mathbf{H}\}. \tag{3.1}$$

From this expression and the result of [29, theorem 2.2.9], we find that $\mathbf{S}_{\mathrm{N}}$ is also a phase type distribution with the representation $(\boldsymbol{\delta}_{\mathrm{N}}, L_{\mathrm{N}}) = [\overline{\boldsymbol{\alpha}} \otimes \boldsymbol{\beta}, T \oplus S]$ with dimension hr, i.e., the distribution function of the $\mathbf{S}_{\mathrm{N}}$ is given by

$$P\{\mathbf{S}_{\mathrm{N}} \leqslant x\} = 1 - \boldsymbol{\delta}_{\mathrm{N}} \exp(L_{\mathrm{N}} x)\mathbf{e}, \tag{3.2}$$

where $A \otimes B \equiv (A_{i,j} B)$ is the Kronecker product and $A \oplus B \equiv A \otimes I + I \otimes B$ is the Kronecker sum of the matrices A and B. Similarly, if we define the *channel holding time of a handoff call* as the time period that a non-blocked handoff call stays in a cell as $\mathbf{S}_{\mathrm{H}}$, then we know that it is independent of the arrival process and is the minimum of the cell residence time and the residual requested call holding time, i.e.,

$$\mathbf{S}_{\mathrm{H}} = \min\{\mathbf{R}, \overline{\mathbf{H}}\}. \tag{3.3}$$

Thus, $\mathbf{S}_{\mathrm{H}}$ is also a phase type distribution with the representation $(\boldsymbol{\delta}_{\mathrm{H}}, L_{\mathrm{H}}) = [\boldsymbol{\alpha} \otimes \overline{\boldsymbol{\beta}}, T \oplus S]$ with dimension hr, i.e., the distribution function of the $\mathbf{S}_{\mathrm{H}}$ is given by

$$P\{\mathbf{S}_{\mathrm{H}} \leqslant x\} = 1 - \boldsymbol{\delta}_{\mathrm{H}} \exp(L_{\mathrm{H}} x)\mathbf{e}. \tag{3.4}$$

Sometimes, we are more interested in *the channel holding time*, say $\mathbf{S}$, for any non-blocked call (either non-blocked new call or non-blocked handoff call), i.e., the channel holding time for the merged traffic of new calls and handoff calls, as used in the current literature. In this case, based on the Markov call arrival process, it is intuitive that

$$P\{\text{an arrived call is a new call}\} = \frac{\boldsymbol{\pi} C_{\mathrm{N}}\mathbf{e}}{\boldsymbol{\pi} C_{\mathrm{N}}\mathbf{e} + \boldsymbol{\pi} C_{\mathrm{H}}\mathbf{e}},$$
$$P\{\text{an arrived call is a handoff call}\} = \frac{\boldsymbol{\pi} C_{\mathrm{H}}\mathbf{e}}{\boldsymbol{\pi} C_{\mathrm{N}}\mathbf{e} + \boldsymbol{\pi} C_{\mathrm{H}}\mathbf{e}}.$$

Based on this, we will find that the distribution of the channel holding time is

$$\begin{aligned} P\{\mathbf{S} \leqslant x\} &= \frac{\boldsymbol{\pi} C_{\mathrm{N}}\mathbf{e}}{\boldsymbol{\pi} C_{\mathrm{N}}\mathbf{e} + \boldsymbol{\pi} C_{\mathrm{H}}\mathbf{e}} P\{\mathbf{S}_{\mathrm{N}} \leqslant x\} \\ &\quad + \frac{\boldsymbol{\pi} C_{\mathrm{H}}\mathbf{e}}{\boldsymbol{\pi} C_{\mathrm{N}}\mathbf{e} + \boldsymbol{\pi} C_{\mathrm{H}}\mathbf{e}} P\{\mathbf{S}_{\mathrm{H}} \leqslant x\} \\ &= 1 - \left[\frac{(\boldsymbol{\pi} C_{\mathrm{N}}\mathbf{e})[\boldsymbol{\delta}_{\mathrm{N}} \exp(L_{\mathrm{N}} x)\mathbf{e}]}{\boldsymbol{\pi} C_{\mathrm{N}}\mathbf{e} + \boldsymbol{\pi} C_{\mathrm{H}}\mathbf{e}} \right. \\ &\quad \left. + \frac{(\boldsymbol{\pi} C_{\mathrm{H}}\mathbf{e})[\boldsymbol{\delta}_{\mathrm{H}} \exp(L_{\mathrm{H}} x)\mathbf{e}]}{\boldsymbol{\pi} C_{\mathrm{N}}\mathbf{e} + \boldsymbol{\pi} C_{\mathrm{H}}\mathbf{e}} \right], \end{aligned} \tag{3.5}$$

i.e., S is also a phase type distribution with representation $(\boldsymbol{\delta}, L)$, where

$$\boldsymbol{\delta} = \left(\frac{(\boldsymbol{\pi} C_{\mathrm{N}}\mathbf{e})\boldsymbol{\delta}_{\mathrm{N}}}{\boldsymbol{\pi} C_{\mathrm{N}}\mathbf{e} + \boldsymbol{\pi} C_{\mathrm{H}}\mathbf{e}}, \frac{(\boldsymbol{\pi} C_{\mathrm{H}}\mathbf{e})\boldsymbol{\delta}_{\mathrm{H}}}{\boldsymbol{\pi} C_{\mathrm{N}}\mathbf{e} + \boldsymbol{\pi} C_{\mathrm{H}}\mathbf{e}}\right) \quad \text{and} \quad L = \begin{bmatrix} L_{\mathrm{N}} & 0 \\ 0 & L_{\mathrm{H}} \end{bmatrix}. \tag{3.6}$$

3.2. Necessary and sufficient conditions

For the special case that when both of the new call arrival process and the handoff call arrival process are independently Poissonian and the requested call holding time is exponentially distributed, Fang et al. [8] proved that the cell residence times of the portable are exponentially distributed if and only if

- *the new call channel holding time* is exponentially distributed; or
- *the handoff call channel holding time* is exponentially distributed; or
- *the call channel holding time* is exponentially distributed.

Based on our scheme, we can easily generalize these results to the more general phase type distribution and the Markov call arrival process. More precisely, if the call arrival process is a Markov call arrival process and the requested call holding time is phase type distribution with representation $(\boldsymbol{\beta}, S)$, then the cell residence time of a portable is phase type distribution with representation $(\boldsymbol{\alpha}, T)$ if and only if

- *the new call channel holding time* is phase type distribution with representation $(\overline{\boldsymbol{\alpha}} \otimes \boldsymbol{\beta}, T \oplus S)$; or
- *the handoff call channel holding time* is phase type distribution with representation $(\boldsymbol{\alpha} \otimes \overline{\boldsymbol{\beta}}, T \oplus S)$; or
- *the call channel holding time* is phase type distribution with representation $(\boldsymbol{\delta}, L)$, where $\boldsymbol{\delta}$ and L are given in equation (3.6).

We only give a proof of the first result and the others are similar. In fact, we have proved in equation (3.2) that *the new call channel holding time* is phase type distribution if we know that both the cell residence times and requested call holding times are of phase type distributions. On the other hand, since $\exp(T \oplus S) = [\exp(T)] \otimes [\exp(S)]$ (see [10]), we will know that

$$(\overline{\boldsymbol{\alpha}} \otimes \boldsymbol{\beta}) \exp\bigl[(T \oplus S)x\bigr]\mathbf{e} = \bigl\{\bigl[\overline{\boldsymbol{\alpha}} \exp(Tx)\bigr]\mathbf{e}\bigr\}\bigl\{\bigl[\boldsymbol{\beta} \exp(Sx)\bigr]\mathbf{e}\bigr\}.$$

However, by noting that $\mathbf{S}_{\mathrm{N}} = \min\{\overline{\mathbf{R}}, \mathbf{H}\}$ implies

$$P(\mathbf{S}_{\mathrm{N}} > x) = P\bigl(\overline{\mathbf{R}} > x\bigr)P(\mathbf{H} > x),$$

i.e.,

$$(\overline{\boldsymbol{\alpha}} \otimes \boldsymbol{\beta}) \exp(T \oplus Sx)\mathbf{e} = P\bigl(\overline{\mathbf{R}} > x\bigr)\bigl\{\bigl[\boldsymbol{\beta} \exp(Sx)\bigr]\mathbf{e}\bigr\},$$

we will directly know that $P(\overline{\mathbf{R}} > x) = \overline{\boldsymbol{\alpha}} \exp(Tx)\mathbf{e}$, i.e., residual cell residence time $\overline{\mathbf{R}}$ is a phase type distribution with representation $(\overline{\boldsymbol{\alpha}}, T)$. By the relationship between cell residence time and residual cell residence time, we know that the cell residence time must be a PH distribution with representation $(\boldsymbol{\alpha}, T)$.

By the results of the channel holding time for new calls and handoff calls, we can now consider some other performance measures. In next section, we will show how to obtain the blocking probability of new calls and handoff calls.

4. Blocking probabilities

Blocking probabilities, e.g., new call's blocking probability and handoff call failure probability are two important measures in design and optimization of the updated PCS communications. By using the general sum of hyperexponentials (SOHYP) for the cell residence times to characterize the mobilities of the portables, Orlik and Rappaport in [31,32] proposed useful analytical models for mixed users and mixed services. They described the driving processes and the transitions in detail. The case of mixed users and/or mixed services is so complicated that they have to limit the work to scalar case. Even then, they cannot find the explicit expressions of the interesting performance measures. Here, we consider the general phase type cell residence time and use the matrix analytic method. In terms of the results in the previous section, by transforming the cell's state into a multi-dimensional Markov process, we find the explict iterative expressions of the steady-state probabilities first and then those of the new call blocking probability and handoff call failure probability. In fact, we could find more interesting performance measures of the networks after the steady-state probabilities are obtained.

4.1. Multi-dimensional Markov process

Since the PCS network is a homogeneous network, every cell in the network can be considered to be statistically identical and independent of each other [18]. Thus, by analyzing the performance of one single cell, the performance of the whole network can be characterized. Let $\{X(t)\colon t \geqslant 0\}$ be a stochastic process in a given cell on the following state space: $\Delta = \{(0, u) \cup (i, i_{\mathrm{N}}, u, \mathbf{s}_{i_{\mathrm{N}},\mathrm{N}}, \mathbf{s}_{i-i_{\mathrm{N}},\mathrm{H}});\ 0 \leqslant i_{\mathrm{N}} \leqslant \min\{i, M - m\};\ 1 \leqslant i \leqslant M;\ 1 \leqslant u \leqslant K\}$, where $\mathbf{s}_{i_{\mathrm{N}},\mathrm{N}} = (s_1^{\mathrm{N}}, s_2^{\mathrm{N}}, \ldots, s_{i_{\mathrm{N}}}^{\mathrm{N}})$ $(1 \leqslant s_v^{\mathrm{N}} \leqslant hr;\ \ 1 \leqslant v \leqslant i_{\mathrm{N}})$ and $\mathbf{s}_{i-i_{\mathrm{N}},\mathrm{H}} = (s_1^{\mathrm{H}}, s_2^{\mathrm{H}}, \ldots, s_{i-i_{\mathrm{N}}}^{\mathrm{H}})$ $(1 \leqslant s_v^{\mathrm{H}} \leqslant hr;\ \ 1 \leqslant v \leqslant i - i_{\mathrm{N}})$ and

- the state $(0, u)$ represents the state with no call in the system and the arrival process is in phase u;
- the state $(i, i_{\mathrm{N}}, u, \mathbf{s}_{i_{\mathrm{N}},\mathrm{N}}, \mathbf{s}_{i-i_{\mathrm{N}},\mathrm{H}})$ represents the state with
  - i calls (new calls and the handoff calls) in service in the given cell and within these i calls there are i_{N} new calls, which were originated in the given cell;
  - the arrival process is in phase u;
  - the vth $(v = 1, 2, \ldots, i_{\mathrm{N}})$ new call among these i_{N} new calls is being served in phase s_v^{N};
  - the vth $(v = 1, 2, \ldots, i - i_{\mathrm{N}})$ handoff call among those $i - i_{\mathrm{N}}$ handoff calls is being served in phase s_v^{H}.

By the assumptions of the phase-type distributions and the MAP call arrival process, it is easy to know that the stochastic process $\{X(t)\colon t \geqslant 0\}$ is indeed a Markov process with continuous time and discrete state and the generator matrix, Q, describing this Markov chain is given as

$$Q = \begin{bmatrix} B_0 & U_0 & & & & \\ D_1 & B_1 & U_1 & & & \\ & D_2 & B_2 & U_2 & & \\ & & D_3 & B_3 & U_3 & \\ & & & \ddots & \ddots & \ddots \\ & & & & D_M & B_M \end{bmatrix}, \tag{4.1}$$

where

- B_i $(i = 0, 1, \ldots, M)$ refers to no change in the number of calls in the cell when there are i calls in the system receiving service;
- U_i $(i = 0, 1, \ldots, M-1)$ refers to an arrival of a call when there are i calls in the system receiving service;
- D_i $(i = 1, \ldots, M)$ refers to a departure of a call when there are i calls in the system receiving service.

In order to determine the block matrices B_i, U_i and D_i in detail, we need to introduce the following notations first:

- I_w is an identity matrix of dimension w and $I(w, s) = \underbrace{I_w \otimes \cdots \otimes I_w}_{s}$ $(w, s = 1, 2, \ldots)$;
- $W_{\mathrm{N}}(i) = \underbrace{L_{\mathrm{N}} \oplus \cdots \oplus L_{\mathrm{N}}}_{i}$, which refers to service stages of i new calls that are still on-going;
- $W_{\mathrm{H}}(i) = \underbrace{L_{\mathrm{H}} \oplus \cdots \oplus L_{\mathrm{H}}}_{i}$, which refers to service stages of i handoff calls that are still on-going;
- $V_{\mathrm{N}}(i) = \sum_{j=0}^{i-1} I(hr, j) \otimes \mathbf{L}_{\mathrm{N}}^0 \otimes I(hr, i-j-1)$, which refers to service completion of one of the i new calls, where we define $I(hr, 0) \otimes \mathbf{L}_{\mathrm{N}}^0 = \mathbf{L}_{\mathrm{N}}^0 \otimes I(hr, 0) = \mathbf{L}_{\mathrm{N}}^0$;
- $V_{\mathrm{H}}(i) = \sum_{j=0}^{i-1} I(hr, j) \otimes \mathbf{L}_{\mathrm{H}}^0 \otimes I(hr, i-j-1)$, which refers to service completion of one of the i handoff calls, where similarly as above we define $I(hr, 0) \otimes \mathbf{L}_{\mathrm{H}}^0 = \mathbf{L}_{\mathrm{H}}^0 \otimes I(hr, 0) = \mathbf{L}_{\mathrm{H}}^0$.

Note that $W_{\mathrm{N}}(i)$ and $W_{\mathrm{H}}(i)$ are the same and also $V_{\mathrm{N}}(i)$ and $V_{\mathrm{H}}(i)$ are the same simply because $L_{\mathrm{N}} = L_{\mathrm{H}}$. However, we prefer to maintain separate notations in this paper for clarity. By considering the detailed transition issues among the states, we find

1. $B_0 = C_0$ and

$$B_i = \begin{bmatrix} B_{i,0} & & & \\ & B_{i,1} & & \\ & & \ddots & \\ & & & B_{i,d_i} \end{bmatrix}, \quad 1 \leqslant i \leqslant M,$$

where $d_i = \min\{i, M-m\}$. B_i is a $(d_i+1) \times (d_i+1)$ block matrix and

$$B_{i,j} = \begin{cases} C_0 \oplus W_{\mathrm{N}}(j) \oplus W_{\mathrm{H}}(i-j), \\ \quad \text{if } 1 \leqslant i \leqslant M-m-1; \\ (C_0 + C_{\mathrm{N}}) \oplus W_{\mathrm{N}}(j) \oplus W_{\mathrm{H}}(i-j), \\ \quad \text{if } M-m \leqslant i \leqslant M-1; \\ (C_0 + C_{\mathrm{N}} + C_{\mathrm{H}}) \oplus W_{\mathrm{N}}(j) \oplus W_{\mathrm{H}}(M-j), \\ \quad \text{if } i = M, \end{cases}$$

which refers to no change in the state of the system when there are i callers in the system with j of those callers being new calls.

2. $U_0 = [C_{\mathrm{H}} \otimes \boldsymbol{\delta}_{\mathrm{H}},\ C_{\mathrm{N}} \otimes \boldsymbol{\delta}_{\mathrm{N}}]$,

$$U_i = \begin{bmatrix} U_{i,0,\mathrm{H}} & U_{i,0,\mathrm{N}} & & & \\ & U_{i,1,\mathrm{H}} & U_{i,1,\mathrm{N}} & & \\ & & \ddots & \ddots & \\ & & & U_{i,i,\mathrm{H}} & U_{i,i,\mathrm{N}} \end{bmatrix},$$

if $1 \leqslant i \leqslant M-m-1$, is an $(i+1) \times (i+2)$ block matrix and

$$U_i = \begin{bmatrix} U_{i,0,\mathrm{H}} & U_{i,0,\mathrm{N}} & & & \\ & U_{i,1,\mathrm{H}} & U_{i,1,\mathrm{N}} & & \\ & & \ddots & \ddots & \\ & & & U_{i,M-m-1,\mathrm{H}} & U_{i,M-m-1,\mathrm{N}} \\ & & & 0 & U_{i,M-m,\mathrm{H}} \end{bmatrix},$$

if $M-m \leqslant i \leqslant M-1$, is an $(M-m+1) \times (M-m+1)$ square block matrix, where the element $U_{i,j,N} = C_{\mathrm{N}} \otimes I(hr, j) \otimes \boldsymbol{\delta}_{\mathrm{N}} \otimes I(hr, i-j)$ refers to an arrival of new call when there are i calls, with j new calls, receiving service; and that $U_{i,j,H} \equiv C_{\mathrm{H}} \otimes I(hr, j) \otimes I(hr, i-j) \otimes \boldsymbol{\delta}_{\mathrm{H}}$ refers to an arrival of handoff call when there are i calls, with j new calls, receiving service.

3.

$$D_1 = \begin{bmatrix} I_K \otimes \mathbf{L}_{\mathrm{H}}^0 \\ I_K \otimes \mathbf{L}_{\mathrm{N}}^0 \end{bmatrix},$$

$$D_i = \begin{bmatrix} D_{i,0,\mathrm{H}} & 0 & & & \\ D_{i,1,\mathrm{N}} & D_{i,1,\mathrm{H}} & & & \\ & D_{i,2,\mathrm{N}} & D_{i,2,\mathrm{H}} & & \\ & & \ddots & \ddots & \\ & & & D_{i,i-1,\mathrm{N}} & D_{i,i-1,\mathrm{H}} \\ & & & & D_{i,i,\mathrm{N}} \end{bmatrix},$$

$2 \leqslant i \leqslant M-m-1$, is an $(i+1) \times i$ block matrix and

$$D_i = \begin{bmatrix} D_{i,0,\mathrm{H}} & 0 & & & & \\ D_{i,1,\mathrm{N}} & D_{i,1,\mathrm{H}} & & & & \\ & D_{i,2,\mathrm{N}} & D_{i,2,\mathrm{H}} & & & \\ & & \ddots & \ddots & & \\ & & & D_{i,M-m-1,\mathrm{N}} & D_{i,M-m-1,\mathrm{H}} & 0 \\ & & & & D_{i,M-m,\mathrm{N}} & D_{i,M-m,\mathrm{H}} \end{bmatrix},$$

$M-m \leqslant i \leqslant M$, is an $(M-m+1) \times (M-m+1)$ square block matrix, where the element $D_{i,j,\mathrm{N}} = I_K \otimes V_{\mathrm{N}}(j) \otimes I(hr, i-j)$ refers to the departure (completion of service) of a new call when there are i calls receiving service with j of those callers being new callers; and that

$D_{i,j,\mathrm{H}} \equiv I_K \otimes I(hr, j) \otimes V_{\mathrm{H}}(i-j)$ refers to the departure (completion of service) of a handoff call when there are i calls receiving service with j of those callers being new callers.

4.2. Interesting probabilities

In this subsection, we will consider the stationary probability of the Markov chain and then obtain the blocking probability of new calls and handoff calls, respectively. Denote by $\overline{\pi}(0,u)$ and $\overline{\pi}(i, i_{\mathrm{N}}, u, \mathbf{s}_{i_{\mathrm{N}},\mathrm{N}}, \mathbf{s}_{i-i_{\mathrm{N}},\mathrm{H}})$ the steady-state probability of the system in the equilibrium case when the system is in state $(0,u)$ and $(i, i_{\mathrm{N}}, u, \mathbf{s}_{i_{\mathrm{N}},\mathrm{N}}, \mathbf{s}_{i-i_{\mathrm{N}},\mathrm{H}})$, respectively. Let $\overline{\boldsymbol{\pi}}_i$ $(i = 0, 1, \ldots, M)$ be the steady-state probability vector of the system in the equilibrium case when there are i calls in the given cell, the sequence of the elements in the vector $\overline{\boldsymbol{\pi}}_0$ ($\overline{\boldsymbol{\pi}}_i$) is ordered in the *lexicographic* order based on the probability $\overline{\pi}(0,u)$ ($\overline{\pi}(i, i_{\mathrm{N}}, u, \mathbf{s}_{i_{\mathrm{N}},\mathrm{N}}, \mathbf{s}_{i-i_{\mathrm{N}},\mathrm{H}})$). By using [9, lemma 3], if we let $\prod_{i=1}^n V_i = V_1 V_2 \cdots V_n$ for matrices $V_1, \ldots, V_n$, we will find from equation (4.1) the steady-state probability vector $\overline{\boldsymbol{\pi}}_i$ when there are i calls in a given cell is given by

$$\overline{\boldsymbol{\pi}}_n = \overline{\boldsymbol{\pi}}_0 \prod_{i=1}^{n} \big[U_{i-1}(-A_i)^{-1}\big], \quad n = 1, 2, \ldots, M, \tag{4.2}$$

where $\overline{\boldsymbol{\pi}}_0$ satisfies $\overline{\boldsymbol{\pi}}_0 A_0 = 0$ and

$$\overline{\boldsymbol{\pi}}_0 \left[I + \sum_{n=1}^{M} \prod_{i=1}^{n} \big[U_{i-1}(-A_i)^{-1}\big] \right] \mathbf{e} = 1,$$

here A_i can be recursively determined by $A_M = B_M$ and $A_n = B_n + U_n(-A_{n+1}^{-1}) D_{n+1}$ $(n = 0, 1, \ldots, M-1)$.

From these results, we will directly know that the new calls blocking probability p_{N} is

$$\begin{aligned} p_{\mathrm{N}} &= \sum_{i=M-m}^{M} \frac{1}{\lambda_{\mathrm{N}}} \overline{\boldsymbol{\pi}}_n D_{\mathrm{N}} \mathbf{e} \\ &= \frac{\overline{\boldsymbol{\pi}}_0}{\lambda_{\mathrm{N}}} \sum_{i=M-m}^{M} \prod_{s=1}^{i} \big[U_{s-1}(-A_s)^{-1}\big] D_{\mathrm{N}} \mathbf{e}; \end{aligned} \tag{4.3}$$

and the handoff calls failure probability p_{H} is

$$\begin{aligned} p_{\mathrm{H}} &= \frac{1}{\lambda_{\mathrm{H}}} \overline{\boldsymbol{\pi}}_M D_{\mathrm{H}} \mathbf{e} \\ &= \frac{\overline{\boldsymbol{\pi}}_0}{\lambda_{\mathrm{H}}} \prod_{s=1}^{M} \big[U_{s-1}(-A_s)^{-1}\big] D_{\mathrm{H}} \mathbf{e}. \end{aligned} \tag{4.4}$$

In terms of the results in equation (4.2), we could find more interesting measures. For example, the *total carried traffic* (TCT), which is defined as the average number of channel occupied by the calls [32], could be obtained as

$$\begin{aligned} \mathrm{TCT} &= \sum_{n=0}^{M} n \boldsymbol{\pi}_n \mathbf{e} \\ &= \boldsymbol{\pi}_0 \left[\sum_{n=1}^{M} n \prod_{i=1}^{n} \big[U_{i-1}(-A_i)^{-1}\big] \right] \mathbf{e}. \end{aligned}$$

By using the two key performance measures in equations (4.3) and (4.4), we are now able to consider the actual call holding time of a call in the next subsection.

5. Actual call holding time

The actual call holding time of a new call is one of the practical performance measures in wireless telecommunication systems. In [22], the authors obtained a general formula for distribution of the actual call holding time, say $\mathbf{X}_{\mathrm{g}}$, when the requested call holding time and cell residence time are general random variables, by

$$\begin{aligned} P(\mathbf{X}_{\mathrm{g}} \leqslant x) = 1 - (1 - p_{\mathrm{N}}) r p_{\mathrm{H}}^2 & \\ \times \sum_{k=1}^{\infty} (1 - p_{\mathrm{H}})^{k-1} \overline{H}(x) \int_x^{\infty} \overline{R}_k(t)\, \mathrm{d}t & \end{aligned} \tag{5.1}$$

where $\overline{H}(t) = 1 - H(t)$, and $H(t)$ is the distribution function of the the requested call holding time; and $\overline{R}_k(t) = 1 - R_k(t)$, and $R_k(t)$ is the distribution function of, say $\mathbf{R}_k$, the sum of k i.i.d. random variables with distribution of the cell residence time. Based on this result and the results in equations (4.3) and (4.4), we can prove that, when both requested call holding time and cell residence time are of phase type distributions as described in section 1, the actual call holding time of a new call, under the condition that the new call is not blocked, is an infinite mixture of PH distributions, which is referred to as the SPH distribution [37]. The SPH is a PH distribution with countably infinite number of states. In fact, from result of [29, theorem 2.2.2], we know that $\mathbf{R}_k$ is a phase type distribution with representation $(\boldsymbol{\theta}_k, \Theta_k)$, where $\boldsymbol{\theta}_k = (\boldsymbol{\alpha}, \alpha_{r+1}\boldsymbol{\alpha}, \alpha_{r+1}^2\boldsymbol{\alpha}, \ldots, \alpha_{r+1}^{k-1}\boldsymbol{\alpha})$ and

$$\Theta_k = \begin{bmatrix} T & \mathbf{T^0}\boldsymbol{\alpha} & \cdots & \alpha_{r+1}^{k-3}\mathbf{T^0}\boldsymbol{\alpha} & \alpha_{r+1}^{k-2}\mathbf{T^0}\boldsymbol{\alpha} \\ & T & \cdots & \alpha_{r+1}^{k-4}\mathbf{T^0}\boldsymbol{\alpha} & \alpha_{r+1}^{k-3}\mathbf{T^0}\boldsymbol{\alpha} \\ & & \ddots & \vdots & \vdots \\ & & & T & \mathbf{T^0}\boldsymbol{\alpha} \\ & & & & T \end{bmatrix}.$$

Since $(r/k)\int_0^x \overline{R}_k(t)\,\mathrm{d}t$ is the distribution of residual time, say $\overline{\mathbf{R}}_k$, of random variable $\mathbf{R}_k$, we know that $\overline{\mathbf{R}}_k$ is a phase type distribution with representation $(\overline{\boldsymbol{\theta}}_k, \Theta_k)$ where $\overline{\boldsymbol{\theta}}_k$ is the unique solution of the equation $\overline{\boldsymbol{\theta}}_k = \overline{\boldsymbol{\theta}}_k(\Theta_k + \mathbf{\Theta}_k^0 \boldsymbol{\theta}_k)$ and $\overline{\boldsymbol{\theta}}_k \mathbf{e} = 1$, in which $\mathbf{\Theta}_k^0 = -\Theta_k \mathbf{e}$. By the results of [37, theorem 2.2] that a countable mixture of PH-distribution is an SPH-distribution or the results of a finite version in [29, theorem 2.2.4], we know that the following probability distribution $F(x) \equiv r p_{\mathrm{H}}^2 \sum_{k=1}^{\infty} (1 - p_{\mathrm{H}})^{k-1} \int_0^x \overline{R}_k(t)\,\mathrm{d}t$ is a SPH distribution with representation $(\boldsymbol{\theta}, \Theta)$, where

$$\boldsymbol{\theta} = p_{\mathrm{H}}^2 \big(\overline{\boldsymbol{\theta}}_1, 2(1 - p_{\mathrm{H}})\overline{\boldsymbol{\theta}}_2, \ldots, k(1 - p_{\mathrm{H}})^{k-1}\overline{\boldsymbol{\theta}}_k, \ldots\big)$$

and

$$\Theta = \mathrm{diag}(\Theta_i).$$

Therefore, by the result in equation (5.1), we know that the distribution of the actual call holding time, under the condition that the new call is not blocked, is a phase type distribution with representation $(\boldsymbol{\beta} \otimes \boldsymbol{\theta}, S \oplus \Theta)$, i.e.,

$$\begin{aligned} &P\{\text{Actual call holding time} \leqslant x|\text{the new call is not blocked}\} \\ &\quad = 1 - \boldsymbol{\beta} \otimes \boldsymbol{\theta} \exp\big\{(S \oplus \Theta)x\big\}\mathbf{e}, \end{aligned}$$

and thus, the noncentral moments, say μ_k, of the actual call holding time, under the condition that the new call is not blocked, is $\mu_k = (-1)^k k!(\boldsymbol{\beta} \otimes \boldsymbol{\theta})(S \oplus \Theta)^{-k}\mathbf{e}$, $k = 0, 1, 2, \ldots$.

6. Duration of call blocking period

In this section, we introduce two new performance measures, i.e., the duration of new call blocking period and the duration of handoff call blocking period in equilibrium. The first measure is defined as the stationary time period starting from the epoch that the channels for receiving new calls are full for the first time to the first epoch that at least one of the channels within the target cell is available for a new call. While the second one is defined as the stationary time period starting from the epoch that all channels have just been occupied to the epoch that at least one of the channels is available for a handoff call for the first time. More precisely, the duration of new call blocking period is the time period starting from the epoch when the $M-m$ channels have just been occupied after a call arrived and connected with a channel in the target cell to the epoch that there are $m+1$ channels which are available for the first time; the duration of handoff call blocking period is the time period when the total M channels have been occupied after a handoff call arrived and connected with a channel in the target cell to the epoch that one of the channels for a handoff call is available for the first time. Intuitively, these two performance measures are interesting and are practical measures for the network users. Upon knowing these types of information, the new caller and/or handoff caller could make a decision whether to wait and try again. In the next two subsections, we find the distribution and then the expectation of these two measures.

6.1. Duration of new call blocking period

Denote by $\{X_{\mathrm{N}}(t)\colon\ t \geqslant 0\}$ a stochastic process in a given cell during the duration of new call blocking period on the following state space:

$$\begin{aligned} \Delta = \big\{ &* \cup (i, i_{\mathrm{N}}, u, \mathbf{s}_{i_{\mathrm{N}},\mathrm{N}}, \mathbf{s}_{i-i_{\mathrm{N}},\mathrm{H}}); 0 \leqslant i_{\mathrm{N}} \leqslant M - m; \\ &M - m \leqslant i \leqslant M; 1 \leqslant u \leqslant K\big\}, \end{aligned}$$

where $*$ is an absorbing state, which means that at least one of the channels begins to be available for the new call and the meaning of the other states remain the same as in section 4.1. By the assumption of the system, it is easy to prove that $\{X_{\mathrm{N}}(t)\colon\ t \geqslant 0\}$ is indeed a continuous time Markov process with the following infinitesimal generator matrix:

$$Q_{\mathrm{N}} = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ D_{M-m} & B_{M-m} & U_{M-m} & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & D_{M-1} & B_{M-1} & U_{M-1} \\ 0 & 0 & 0 & \cdots & 0 & D_M & B_M \end{bmatrix},$$

where D_i, B_i and U_i are the same as defined in section 4.1. The duration of new call blocking period, say DNCBP, is just the time to absorption into the absorbing state $*$, of Markov process $\{X_{\mathrm{N}}(t), t \geqslant 0\}$ starting from the initial state vector $\boldsymbol{\psi}_{\mathrm{N}} = (0, \boldsymbol{\theta}_{\mathrm{N}})$, where

$$\boldsymbol{\theta}_{\mathrm{N}} = \left(\frac{\boldsymbol{\pi}(C_{\mathrm{N}} + C_{\mathrm{H}}) \otimes I}{\lambda_{\mathrm{H}} + \lambda_{\mathrm{N}}}, 0, 0, \ldots, 0\right).$$

Based on these arguments, if we denote by

$$T_{\mathrm{N}} = \begin{bmatrix} B_{M-m} & U_{M-m} & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & D_{M-1} & B_{M-1} & U_{M-1} \\ 0 & 0 & \cdots & 0 & D_M & B_M, \end{bmatrix},$$

then from result [29, lemma 2.2.2], we know the distribution of the duration of new call blocking period is

$$P(\mathrm{DNCBP} \leqslant x) = 1 - \boldsymbol{\theta}_{\mathrm{N}} \exp(T_{\mathrm{N}} x)\mathbf{e} \quad \text{for } x \geqslant 0,$$

and the noncentral moments, $E(\mathrm{DNCBP})^k$, are given by

$$E(\mathrm{DNCBP})^k = (-1)^k k!(\boldsymbol{\theta}_{\mathrm{N}} T_{\mathrm{N}}^{-k}\mathbf{e}) \quad \text{for } k \geqslant 0.$$

6.2. Duration of handoff call blocking period

Similar to the above subsection, we denote by $\{X_{\mathrm{H}}(t)\colon\ t \geqslant 0\}$ a stochastic process of a given cell during the duration of handoff call blocking period on the following state space:

$$\begin{aligned} \Delta = \big\{ &* \cup (M, i_{\mathrm{N}}, u, \mathbf{s}_{i_{\mathrm{N}},\mathrm{N}}, \mathbf{s}_{i-i_{\mathrm{N}},\mathrm{H}}); \\ &0 \leqslant i_{\mathrm{N}} \leqslant M - m; 1 \leqslant u \leqslant K\big\}, \end{aligned}$$

where $*$ is an absorbing state, which means that at least one of the M channels begins to be available for a handoff call and the meaning of the other states remain the same as in section 4.1. It is easy to prove that $\{X_{\mathrm{H}}(t)\colon\ t \geqslant 0\}$ is indeed a continuous time Markov process with the following infinitesimal generator matrix

$$Q_{\mathrm{H}} = \begin{bmatrix} 0 & 0 \\ D_M & B_M \end{bmatrix},$$

where D_M and B_M are same as those in section 4.1. The duration of handoff call blocking period, say DHCBP, is just the time to absorption into the absorbing state $*$, of Markov process $\{X_{\mathrm{H}}(t), t \geqslant 0\}$ starting from the initial state vector $\boldsymbol{\psi}_{\mathrm{H}} = (0, \boldsymbol{\pi}(C_{\mathrm{H}} \otimes I)/\lambda_{\mathrm{H}})$. Based on these arguments and from the result [29, lemma 2.2.2], we know the distribution of the duration of handoff call blocking period is

$$P(\text{DHCBP} \leqslant x) = 1 - \frac{1}{\lambda_\text{H}}\big(\boldsymbol{\pi}(C_\text{H} \otimes I)\exp(B_M x)\mathbf{e}\big) \quad \text{for } x \geqslant 0,$$

and the noncentral moments, $E(\text{DHCBP})^k$, are given by

$$E(\text{DHCBP})^k = \frac{(-1)^k k!}{\lambda_\text{H}}\big(\boldsymbol{\pi}(C_\text{H} \otimes I)B_M^{-k}\mathbf{e}\big) \quad \text{for } k \geqslant 0.$$

7. Conclusions

In this paper, we considered a homogeneous PCS network with Markov call arrival process and general phase type cell residence time and general phase type requested call holding time. We generalized the necessary and sufficient conditions obtained by Fang et al. [8] to a general framework. By transforming the situation of a cell into a Markov process and finding the explicit expression of corresponding transition rate matrix, we obtained the expression for the two key performance measures, i.e., new call blocking probability and the handoff call failure probability. This is one of the main contributions of this paper, i.e., the development of the infinitesimal generator matrix of the Markov model for the system described. Efficient computational aspects are still under investigation. The complexity of the algorithm is $O((K(M-m)(hr)^M)^2)$. In addition to this, we proved that the actual call holding time, under the condition that the new call is not blocked, has an SPH distribution introduced in [37]. Finally, two practical and interesting measures for the network users, i.e., the duration of new call blocking period and the duration of handoff blocking period, were introduced; their distributions and the expectations were obtained explicitly.

Acknowledgements

The authors would like to express their sincere appreciation to the reviewers for their suggestions and comments which have greatly helped to improve the quality of the presentation of this paper. This work has been supported by Grant No. OGP0006584 from the Natural Sciences and Engineering Research Council of Canada to A.S. Alfa and was supported in part by Louisiana Board of Regents Support Fund under grant LEQSF(2000-03)-RD-A-40 to Wei Li.

References

[1] A.S. Alfa and M.F. Neuts, Modelling vehicular traffic using the discrete time Markovian arrival process, Transportation Science 29(2) (1995) 109–117.

[2] S. Asmussen, Phase-type distributions and related point processes: Fitting and recent advances, in: *Matrix-Analytic Methods in Stochastic Models*, eds. S.R. Chakravarthy and A.S. Alfa (Marcel Dekker, New York, 1996) pp. 137–149.

[3] V.A. Bolotin, Modeling call holding time distribution for CCS network design and performance analysis, IEEE Journal on Selected Areas in Communications 12(3) (1994) 433–438.

[4] ETSI, Digital European telecommunications services and facilities requirements specification, Technical report, ETSI, DI/RES 3002, European Telecommunications Standards Institute (1991).

[5] Y.G. Fang and I. Chlamtac, Teletraffic analysis and mobility modeling of PCS networks, IEEE Transactions on Communications 47(7) (1999) 1062–1072.

[6] Y.G. Fang, I. Chlamtac and Y.B. Lin, Call performance for a PCS network, IEEE Journal on Selected Areas in Communications 15(8) (1997) 1568–1581.

[7] Y.G. Fang, I. Chlamtac and Y.B. Lin, Modeling PCS networks under general call holding time and cell residence time distributions, IEEE/ACM Transactions on Networking 6(5) (1997) 893–906.

[8] Y.G. Fang, I. Chlamtac and Y.B. Lin, Channel occupancy times and handoff rate for mobile computing and PCS networks, IEEE Transactions on Computers 47(6) (1998) 679–692.

[9] D.P. Gaver, P.A. Jacobs and G. Latouche, Finite birth-and death models in randomly changing environments, Advances in Applied Probability 16 (1984) 715–731.

[10] A. Graham, *Kronecker Products and Matrix Calculus with Applications* (Ellis Horwood, Chichester, 1981).

[11] R. Guerin, Channel occupancy time distribution in a cellular radio system, IEEE Transactions on Vehicular Technology 36 (1987) 89–99.

[12] Guerin, Queueing-blocking system with two arrival streams and quard channels, IEEE Transactions on Communications 36(2) (1988) 153–163.

[13] Q.M. He and M.F. Neuts, Markov chains with marked transitions, Stochastic Processes and their Applications 74 (1998) 37–52.

[14] H. Heffes and D.M. Lucantoni, A Markov modulated characterization of packetized voice and date traffic and related statistical multiplexer performance, IEEE Journal on Selected Areas in Communications 4(6) (1986) 856–868.

[15] J.S.M. Ho and I.F. Akyildiz, Mobile user location update and paging under delay constraints, Wireless Networks 1 (1995) 413–425.

[16] D. Hong and S.S. Rappaport, Traffic model and performance analysis for cellular mobile radio telephone systems, IEEE Transactions on Communications 36(2) (1988) 153–163.

[17] F. Khan and D. Zeghlache, Effect of cell residence time distribution on the performance of cellular mobile networks, in: *Proc. IEEE VTC* (May 1997) pp. 949–953.

[18] M.D. Kulavaratharasah and A.H. Aghvami, Teletraffic performance evaluation of microcellular personal communication networks (PCN's) with prioritized handoff procedures, IEEE Transactions on Vehicular Technology 48(1) (1999) 137–152.

[19] D. Lam, D.C. Cox and J. Widom, Teletraffic modeling for personal communications services, IEEE Communication Magazine 35 (February 1997) pp. 79–87.

[20] A. Lang and J.L. Arthur, Parameter approximation for phase-type distributions, in: *Matrix-Analytic Methods in Stochastic Models*, eds. S.R. Chakravarthy and A.S. Alfa (Marcel Dekker, New York, 1996) pp. 151–206.

[21] W. Li and A.S. Alfa, A PCS network with correlated arrival process and splitted-rate channels, IEEE Journal on Selected Areas in Communications 17(7) (1999) 1318–1325.

[22] W. Li and A.S. Alfa, Channel reservation for handoff calls in a PCS network, IEEE Transactions on Vehicular Technology 49 (2000) 95–104.

[23] Y.B. Lin, R. Anthony and D. Harasty, The sub-rate channel assignment strategy for PCS handoffs, IEEE Transactions on Vehicular Technology 45(1) (1996) 122–130.

[24] Y.B. Lin and I. Chlamtac, Effective call holding times for a PCS network, IEEE Journal on Selected Areas on Communications (submitted).

[25] D.M. Lucantoni, New results on the single server queue with a batch Markovian arrival process, Communications in Statistics – Stochastic Models 7 (1991) 1–46.

[26] D.M. Lucantoni, G.L. Choudhury and W. Whitt, The transient BMAP/G/1 queue, Communications in Statistics – Stochastic Models 10(1) (1994) 145–182.

[27] D.M. Lucantoni, K.S. Meier-Hellstern and M.F. Neuts, A singer-server queue with server vacations and a class of non-renewal arrival processes, Advances in Applied Probability 22 (1990) 676–705.
[28] M.F. Neuts, A versatile Markovian point process, Journal of Applied Probability 16 (1979) 764–779.
[29] M.F. Neuts, *Matrix-Geometric Solutions in Stochastic Models* (The John Hopkins University Press, 1981).
[30] A.R. Noerpel, Y.B. Lin and H. Sherry, PACS: personal access communications system – A tutorial, IEEE Personal Communications 3 (1996) 32–43.
[31] P.V. Orlik and S.S. Rappaport, A model for teletraffic performance and channel holding time characterization in wireless cellular communication with general session and dwell time distributions, IEEE Journal on Selected Areas in Communications 16(5) (1998) 788–803.
[32] P.V. Orlik and S.S. Rappaport, Teletraffic performance and mobility modeling of cellular communications with mixed platforms and highly variable mobilities, Proceedings of the IEEE 86(7) (1998) 1464–1479.
[33] S.S. Rappaport, The multiple-call handoff problem in high-capacity cellular communications systems, IEEE Transactions on Vehicular Technology 40(3) (1991) 546–557.
[34] S.S. Rappaport, Blocking, handoff and traffic performance for cellular communications with mixed platforms, IEE Proceedings 140(5) (1993) 389–401.
[35] C. Rose and R. Yates, Location uncertainly in mobile networks: A theoretical framework, IEEE Communication Magazine 35 (February 1997) 94–101.
[36] G. Ruiz, T.L. Doumi and J.G. Gardiner, Teletraffic analysis and simulation for nongeostationary mobile satellite systems, IEEE Transactions on Vehicular Technology 47(1) (1998) 311–320.
[37] D.H. Shi, J. Guo and L. Liu, SPH-distributions and the rectangle-iterative algorithm, in: *Matrix-Analytic Methods in Stochastic Models*, eds. S.R. Chakravarthy and A.S. Alfa (Marcel Dekker, New York, 1996) pp. 207–224.
[38] K. Sriram and W. Whitt, Characterizing superposition arrival processes in an ATM transport network, IEEE Journal on Selected Areas in Communications 4(6) (1986) 359–367.
[39] R. Steedman, The common air interface MPT 1375, in: *Cordless Telecommunications in Europe*, ed. W.H. Tuttlebee (Springer-Verlag, New York, 1990).
[40] M.M. Zonoozi and P. Dassanayake, User mobility modeling and characterization of mobility patters, IEEE Journal on Selected Areas in Communications 15(7) (1997) 239–252.

Attahiru Sule Alfa received the B.Eng. degree from the Ahmadu Bello University, Nigeria, in 1971, the MSc. degree from the University of Manitoba, Winnipeg, Canada, in 1974, and the Ph.D. degree from the University of New South Wales, Australia, in 1980. He is a Professor of Industrial and Manufacturing Systems Engineering (with cross appointments in mathemetics and statistics, and also electrical and computer engineering) and the Associate Vice-President of Research at the University of Windsor. His research interest is in the area of operations research, with focus in the applications of queueing and network theories to telecommunication systems, especially mobile wireless communications. He also has interests in network restoration problems. He has contributed to the development and applications of the matrix-analytic method for stochastic models, especially in the area of discrete time queueing systems.
E-mail: alfa@uwindsor.ca

Wei Li (M'99) received his Ph.D. degree from Chinese Academy of Sciences, Beijing, China, in 1994, a M.Sc. degree from Hebei University of Technology, Tianjin, China, in 1987, and a B.Sc. degree from Shannxi Normal University, Xian, China, in 1982. He is currently an Associate Professor in the Department of Electrical Engineering and Computer Science, University of Toledo, Ohio, USA. He was a Bell South/BORSF Professor in Telecommunications from 2000 to 2002 and an Assistant Professor in the Department of Electrical and Computer Engineering, University of Louisiana at Lafayette, Louisiana, USA, from 1999 to 2002. He was engaged in research and teaching at the University of Manitoba and the University of Winnipeg in Canada from 1995 to 1999. He was once a senior researcher at the University of Newcastle in England and an associate researcher at the Chinese Academy of Sciences in China. His current research interests are in the evaluation, design and implementation of teletraffic models for personal communications services networks, wireless networks and mobile satellite communications networks as well as the applications in these areas of operations research, queueing networks and reliability networks.
E-mail: wli@eecs.utoledo.edu