Computers & Operations Research 43 (2014) 181–190

Contents lists available at ScienceDirect

Computers & Operations Research

journal homepage: www.elsevier.com/locate/caor

Analysis and optimization of Guard Channel Policy in cellular mobile networks with account of retrials

Chesoong Kim [a], V.I. Klimenok [b], A.N. Dudin [b,*]

[a] Department of Industrial Engineering, Sangji University, Wonju, Kangwon 220-702, South Korea
[b] Belarusian State University, 4, Nezavisimosti Ave., Minsk 220030, Belarus

ARTICLE INFO

Available online 17 September 2013

Keywords:
Mobile cellular network
Retrial queue
Handover calls
Priority
Channels reservation
Optimization

ABSTRACT

A multi-server retrial queue with two types of calls (handover and new calls) is analyzed. This queue models the operation of a cell of a mobile communication network. Calls of two types arrive at the system according to the Marked Markovian Arrival Process. Service times of both types of the calls are exponentially distributed with different service rates. Handover calls have priority over new calls. Priority is provided by means of reservation of several servers of the system exclusively for service of handover calls. A handover call is dropped and leaves the system if all servers are busy at the arrival epoch. A new call is blocked if all servers available to new calls are busy. Such a call has options to balk (to leave the system without getting the service) or to retry later on. The behavior of the system is described by the four-dimensional Markov chain belonging to the class of the asymptotically quasi-Toeplitz Markov chains (*AQTMC*). In the paper, a constructive ergodicity condition for this chain is derived and the effective algorithm for computing the stationary distribution is presented. Based on this distribution, formulas for various performance measures of the system are obtained. Results of numerical experiments illustrating the behavior of key performance measures of the system depending on the number of the reserved servers under the different shares of the handover and the new calls are presented. An optimization problem is considered and high positive effect of server's reservation is demonstrated.

1. Introduction

The importance of the problem of optimal handling both handover and new calls in a cell of a wireless network is well-recognized. The limited bandwidth of a target cell and competition between the calls may create essential problems, especially for the moving users. When an active mobile user enters the target cell moving from the adjoining cell, his/her communication can be terminated due to lack of free channels. The requests of such on-going (handover) calls compete with the requests of the calls originated in the target cell (new calls).

From user's perspective, it is more intolerable to drop an on-going service than to block a service that has yet to be established. Therefore, with limited bandwidth in a cell, satisfying requests of on-going (handover) calls is more important than satisfying requests of new calls generating in the cell.

To provide some kind of priority to handover calls over the new ones, different policies are elaborated. The well-known policy is a so-called Guard Channel Policy, see, e.g., [11]. This policy assumes the reservation of some part of servers (channels) exclusively for the service of handover calls. Let the total number of channels in the target cell be equal to K. Under the Guard Channel Policy, some number $R, R<K$, of channels is reserved for the service of handover calls only. A new call is accepted for service only if the number of busy servers at the arrival moment is less than $N=K-R$. Naturally, the problem of optimal choice of the number R of the reserved channels arises. As a criterion for the quality of optimization, a cost function can be considered. This function should account for probabilities of the handover call loss (sometimes referred to as a dropping probability), the new call loss (blocking probability), and the utilization of bandwidth of the cell. As a result, under the fixed values of charges for the loss of calls and for under-utilization of bandwidth, the optimal value R^* of the number of servers reserved exclusively for the service of handover calls should be found.

There is a lot of works where such type of optimization problem is considered. Some of them are cited in [16] and in the papers mentioned therein. Advantage of the work [16] over the previous papers is the exact mathematical analysis of the model with taking into consideration different service rates for handover and new calls. Shortcomings of the model considered in [16] are as follows:

(i) handover and new calls arrive in the independent stationary Poisson flows. This assumption contradicts the fact that traffic in

* Corresponding author. Tel.: +375 172095486; fax: +375 172265548.
E-mail addresses: dowoo@sangji.ac.kr (C. Kim), vklimenok@yandex.ru (V.I. Klimenok), dudin@bsu.by, dudin_alex@yahoo.com (A.N. Dudin).

0305-0548/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cor.2013.09.005

the modern wireless mobile communication networks exhibits correlation and high variability of inter-arrival times;

(ii) a blocked new call does not repeat the attempts to get service and leaves the system forever. Under such assumption, the effect of retrials is ignored.

In the model under study, the correlation and the variability of inter-arrival times in the input flow of the heterogeneous calls are taken into account via the consideration of the *Marked Markovian Arrival Process* (*MMAP*). We also take into account the retrial phenomenon assuming that a blocked new call enters the orbit of an infinite size and repeats its attempts to reach a server in a random amount of time independent of the other orbital calls. Thus, in this paper we consider a multi-server retrial queue with two types of calls arriving according to a *MMAP*. Note that the *MMAP* is a generalization of the well-known, see, e.g., [5,17], *Markovian Arrival Process* (*MAP*) to the case of heterogeneous calls. State of the art in analysis of the retrial queues with homogeneous calls arriving according to *MAP*s and *BMAP*s – *Batch Markovian Arrival Processes* can be found in [2,10].

The closest to our model are the models considered in papers [1,6–8,20].

In [6], a multi-server queueing model with independent *MAP*s of handover and new calls is considered. In case when all servers are busy at an arrival time, the handover call joins a buffer of an infinite capacity while the new call moves to the orbit of a finite capacity and retries for service later on. No reservation of servers for handover calls is suggested in [6].

In the paper [1], the authors consider quite general model. They assume that service and inter-retrial times have a so-called phase type (*PH*) distribution which is much more general compared to exponential distributions suggested in our paper. Disadvantage of the paper [1] consists of the following. From the mathematical point of view, the main difficulty of the analysis of a retrial queue is due to infinite orbit size and state inhomogeneous behavior of the multi-dimensional Markov chain describing the operation of the system. The authors of [1] overcome this difficulty by cutting capacity of the orbit. Namely this suggestion allows the authors of [1] to incorporate an assumption about the *PH* distribution of inter-retrial time. Although this suggestion greatly simplifies the mathematical analysis of the model, it reduces adequacy of the model to real life. Due to the mobility of the users, it is hardly possible to cut the possible number of users in a given cell by a finite number, which should be fairly small to guarantee the feasibility of the computational algorithm offered in [1].

A very close to the model analyzed in [1] is the retrial queueing system considered in the recent paper [20]. But, in order to simplify the mathematical analysis of the model, the authors of this paper assume constant retrial rate from the orbit. We assume that the retrial rate depends on the number of calls in the orbit that is obviously true in real life.

The paper [7] deals with a queueing system, which is a special case of our queue, under the following assumptions: (a) arrival flows of handover and new calls are defined as stationary Poisson ones; (b) service times of handover and new calls are identically distributed. It is clear that both assumptions simplify the model to the detriment of its adequacy. The system has been analyzed by means of the approximate phase merging algorithm.

In the paper [8], the author analyzes a model similar to the model studied in [7]. The inessential difference is that the author that a dropped handover call will retry later on while in [7] it is assumed that such a call is lost. The significant disadvantage of the model presented in [8] compared to the model in [7] is that it is assumed there, as well as in [20], that the flow of retrials from the orbit has a constant rate independent of the number of calls in the orbit. As it has already been noted above, this assumption facilitates the mathematical analysis while it is hardly practically motivated.

The advantages of our analysis compared to the results given in paper [7] are two-fold: we consider more general queueing model (more general arrival process and different service rates of handover and new calls) and provide the exact mathematical analysis.

Notably, none of the papers on the subject present ergodicity condition for the model with infinite capacity of the orbit and independent retrials of the calls from the orbit. We derive such a condition. This condition has an analytically tractable form, which, however, was not apparent at first sight.

It should be mentioned that our analysis is implemented under suggestion that the parameters of the arrival process and service times are known. Sure, an application of the obtained results for investigation of some practical system can be done only after evaluation of these parameters in this concrete system. The problem of fitting real life flows by the *MMAP* is well known in literature, see, e.g., [4]. Concerning evaluation of the service time parameters, see, e.g., paper [1] where some way for defining service times of handover and new calls via the holding and residence times (and their residuals) is presented.

Note that extension of the presented analysis to the case of the *PH* type distribution of service times is possible by analogy with [3] or [13]. However, we restrict ourselves in this paper by the exponential distribution taking into account both difficulty in fitting the parameters and essential reduction of the dimension of the process under study in exponential case what is very important on the stage of computer implementation.

The rest of the paper is organized as follows. In Section 2, the queueing system under consideration is defined. In Section 3, the multi-dimensional continuous time Markov chain, which describes the behavior of this system, is constructed. The generator of this chain is presented and the fact that this Markov chain belongs to the class of the asymptotically quasi-Toeplitz Markov chains is proved. In Section 4, the stability condition of this Markov chain is derived and justified intuitively. The algorithm for computing the stationary probabilities is outlined. In Section 5, formulas for some key performance measures of the system are derived and an optimization problem is formulated. Numerical examples are presented in Section 6. Section 7 concludes the paper.

2. Mathematical model

The structure of the system under study is presented in Fig. 1.

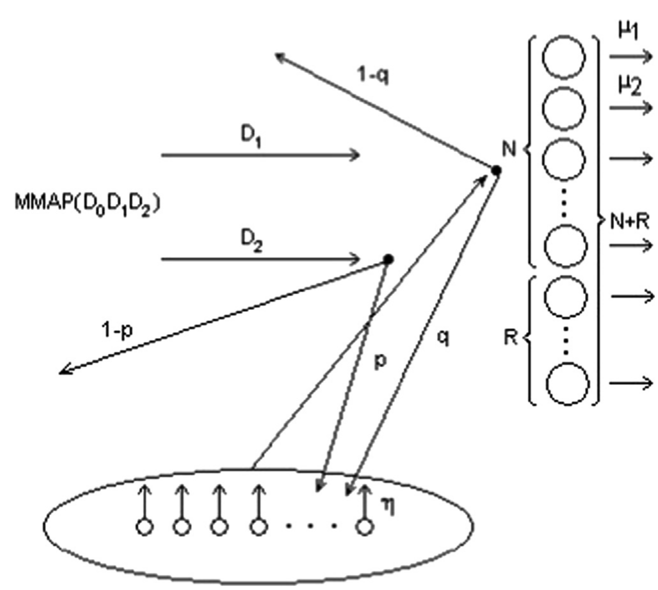

Fig. 1. Structure of the system.