Computers & Operations Research 37 (2010) 1228–1237

Contents lists available at ScienceDirect

Computers & Operations Research

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

The *BMAP/PH/N* retrial queueing system operating in Markovian random environment

Che Soong Kim [a,*], Valentina Klimenok [b], Vilena Mushko [b], Alexander Dudin [b]

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

ARTICLE INFO

Available online 26 September 2009

Keywords:
BMAP/PH/N retrial queueing model
Random environment
Stationary state distribution
Ergodicity
Approximation

ABSTRACT

We consider the *BMAP/PH/N* retrial queueing system operating in a finite state space Markovian random environment. The stationary distribution of the system states is computed. The main performance measures of the system are derived. Presented numerical examples illustrate a poor quality of the approximation of the main performance measures of the system by means of the simpler queueing models. An effect of smoothing the traffic and an impact of intensity of retrials are shown.

1. Introduction

Effect of repeated requests should be taken into account in most applications of queueing theory in telephone and telecommunication networks, local and metropolitan area networks and queues in daily life systems. This makes popular investigation of retrial queues. Bibliography on algorithmic analysis of retrial queues was recently published in [7]. Multi-server retrial queues are much more complicated subject for research comparing to single-server queues. Even the simplest model where the input flow is described by the stationary Poisson arrival process, service and retrial times are exponentially distributed, the model can be investigated only algorithmically. Due to the space inhomogeneity of the Markovian process, which describes behavior of the multi-server retrial queue, existing algorithms for calculation of the system states distribution suggest some kind of truncation of the state space. Different schemes of truncation for the $M/M/N$ retrial queue are discussed in the book [5]. Much more complicated retrial queueing model of the *BMAP/PH/N* type with different strategies of retrials was investigated in [1]. Here the *BMAP* stands for the Batch Markov Arrival Process while *PH* means the phase type service time distribution. In contrast to the conventional truncation of the equilibrium equations, approach offered in [1] assumes derivation and solution of an alternative system of equations for the stationary probabilities. This approach also requires a truncation of the sequence of low size matrices computed via the backward recursion, not truncation of the large system of equilibrium equations. So, it requires much less memory and computing time comparing to traditional approaches. In the paper [11], results of [1] are extended to the case when calls are not persistent, i.e., with fixed probability, the call leaves the system forever if its retrial was not successful. Various system performance measures were computed and the results of the extensive numerical study were presented.

Currently, the design of fast growing radio access networks of the third and fourth generation provides a major challenge to teletraffic engineering. Particularly, the network design of hot spots with very high sporadic demand caused by surges of suddenly arriving mobile subscribers requires adequate engineering tools. Such situations typically arise at access points related to mass transportation like airport, train or bus terminals, hotels or exhibition areas where large groups of subscribers arrive simultaneously and intend to satisfy their communication needs immediately with satisfactory service quality. In the paper [2], the model analyzed in [1] was applied for evaluation of performance measures of hot spots in telecommunication networks. Arrival of call was modeled by the *IPP* (interrupted Poisson process) that is a particular case of the *BMAP*, having zero intensity of arrivals during some intervals, e.g., intervals between aircrafts arrival to a given hot spot. Analysis presented in [2] is quite accurate if the arriving aircrafts have the same capacity. However, the aircrafts hosted in a hot spot may be essentially different, say Boeing-737 and Boeing-767 and the *BMAP/PH/N* type retrial queue, although it is very general and takes into account possible non-stationarity in the arrival process, does not give an adequate model of a situation. In general, this retrial queue also fails in situations when the arrival, service and retrial processes are simultaneously influenced by some external process which may cause sharp changes in the behavior of these processes.

* Corresponding author. Tel.: +82 33 730 0464; fax: +82 33 743 1115.
E-mail addresses: dowoo@sangji.ac.kr (C.S. Kim), klimenok@bsu.by (V. Klimenok), vmushko@tut.by (V. Mushko), dudin@bsu.by (A. Dudin).

0305-0548/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cor.2009.09.008

In such situations, analysis of more complicated models, the so-called models operating in the random environment (*RE*), is required. Speaking about a queue in an *RE*, we assume that there are a queueing system and an external finite state space stochastic process called the *RE*. Under the fixed state of the *RE*, the queueing system operates as a classic queueing system of the correspondent type. However, when the *RE* jumps into another state, the parameters of the queueing system (inter-arrival times distribution or arrival rate, service times distribution or service rate, number of servers, retrial rate, etc.) can immediately change their values. In this paper we assume that the *RE* is defined by a finite state continuous time Markov chain. It is well known that the *BMAP* arrival process and *PH* type service process, in turn, are governed by some underlying finite state continuous time Markov chains. So, the natural question may arise: whether or not the *BMAP*/*PH*/*N* type retrial queue operating in the *RE* can be reduced to the usual *BMAP*/*PH*/*N* type retrial queue by means of suitable extension of state spaces of the underlying processes of the *BMAP* and *PH*. The answer is definitely positive only in the case if the *RE* causes variation of behavior only of the underlying process of arrivals. If the *RE* changes the behavior of the service and (or) retrial processes, the answer is strongly negative because this behavior depends not only on the *RE*, but also on the state of the queue (there is no service when the server is idle and there is no retrial if the orbit is empty), see Section 6.

In the model of a hot spot in airport, the current state of the *RE* is defined by the type of the aircraft approaching to analyzed terminal (and probably by the duration of the finished flight) and this state defines not only the parameters of the *IPP* process of calls generated by the arriving passengers (intensity of calls and duration of an active period), but the total rate of retrials as well. This is why in this paper we allow arbitrary dependence of the total retrial rate on the number of retrying calls and assume that this dependence can be different from linear dependence, which is suggested in the overwhelming majority of the papers, and be various in different states of the *RE*. Note also that temporary restrictions of call duration can be imposed in periods corresponding to the states of the *RE* when call intensity is very high, what leads to the change of the mean service time in the corresponding queueing model. Thus, the queueing model cannot be reduced to the classical model with independent arrival and service processes.

One more real world example of the system that should be modeled by the *BMAP*/*PH*/*N* type retrial queue operating in the *RE* is the following. There is a wireless local area network which provides a service to residents of some group of buildings. Tariffs (dollars per gigabyte) for providing access to Internet are more or less high and are fixed in contracts of a provider and user. While tariffs for access to internal resources of the service provider (video, multi-media, music, computer games, cinema, etc.) are, as a rule, more flexible and are much lower. So, internal resources are usually very popular among users and their competition arises. Provider knows statistics about the arrival rate of requests from the users which shows that this rate essentially varies several times during the day and night time (we can interpret again this is the influence of the *RE*) what may be accomplished by a poor quality of service during the peak rate periods (denial of access or long downloading time). To improve situation, provider has several options: (i) to regulate tariffs in such a way as to make them higher during peak periods and lower during periods when the server (and wireless channels for access) is under-utilized (what should imply smoothing the traffic), (ii) to restrict duration of seances of access to the server, (iii) to monitor and restrict retrial rates from individual users. The *BMAP*/*PH*/*N* type retrial queue operating in the *RE* provides the valuable tool for modeling various scenarios of assigning tariffs, restricting duration of a continuous service and retrial rates in such a local area network with the purpose to provide better service to users and higher profit to provider.

In general, the nature of an *RE* in real world wireless systems can be versatile. Change of the service rates over time can be caused by the random degradation of the transmission thread due to technical reasons, weather conditions, temperature inversions in the atmosphere, frequency interference, noise caused by enterprises and transport, fluctuation of the base station, etc. The first system in the *RE* was investigated in [6] as the queue with partially broken server. Change of the arrival rates over time is explained by the users mobility and the cyclic change of their activity during a day or a night. It is worth to note that importance of investigation of the queues operating in an *RE* drastically increased in the last years due to the following reason. The flows of information in modern communication networks are heterogeneous. Some types of information are very sensitive with respect to a delay and an jitter but tolerant with respect to losses. Another ones are tolerant with respect to the delay but very sensitive with respect to the loss of the packets. So, different schemes of the dynamic bandwidth sharing among these types exist and are developing. They assume that, in the case of congestion, transmission of the delay tolerant flows is temporarily postponed to provide better conditions for transmission of the delay sensitive flows. Analysis of such schemes requires the probabilistic analysis of the multi-dimensional processes describing transmission process of the different flows. This analysis is often impossible due to the mathematical complexity. In such a case, it is reasonable to decompose a simultaneous consideration of all flows to separate analysis of the processes of transmission of the delay sensitive and the delay tolerant flows. To this end, one can model transmission of the delay sensitive flows in terms of the queues with the controlled service or (and) arrival rate where the rates can be changed depending on the queue length or the waiting time. Redistribution of a bandwidth to avoid congestion for the delay sensitive flows causes a variation, at random moments, of an available bandwidth for the delay tolerant flows. Correspondingly, the queues operating in an *RE* naturally arise as the mathematical model for the delay tolerant flows transmission.

Queueing systems operating in the *RE* are in focus of investigation for more than 40 years. Yearly research in this topic was done by B.V. Gnedenko, I.N. Kovalenko, U. Yechiali, P. Naor, U. Yadin, P. Purdue, M. Neuts and others. Short bibliography in the investigation of queueing systems operating in the *RE* was presented in the recent paper [8]. To the best of our knowledge, the variety of retrial queues operating in the *RE* is restricted to the set of papers [3,9,12] where the single server retrial queues are dealt with. So, the present paper is the first one devoted to the multi-server retrial queue operating in the *RE* and the results are novel even if we assume stationary Poisson arrival process and exponential service time distribution under the fixed value of the *RE*. The results are novel also if we assume that the arrival process and the service or retrial process do not depend on the state of the *RE*.

The rest of the paper is organized as follows. Mathematical model is formulated in Section 2. Section 3 contains derivation of stability condition and short description of the algorithm for computing the stationary distribution of the system states. Performance measures of the system are given in Section 4. Section 5 contains results of the numerical experiments and their analysis. Section 6 concludes the paper.

2. Mathematical model

We consider a multi-server retrial queueing system having N identical servers. The behavior of the system depends on the state