Performance Evaluation 98 (2016) 52–69

Contents lists available at ScienceDirect

Performance Evaluation

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

Analysis and optimization of vacation and polling models with retrials ☆

Murtuza Ali Abidini, Onno Boxma, Jacques Resing *

EURANDOM, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands
Department of Mathematics and Computer Science, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

ARTICLE INFO

Article history:
Available online 9 February 2016

Keywords:
Vacation queue
Polling model
Retrials

ABSTRACT

We study a vacation-type queueing model, and a single-server multi-queue polling model, with the special feature of retrials. Just before the server arrives at a station there is some deterministic glue period. Customers (both new arrivals and retrials) arriving at the station during this glue period will be served during the visit of the server. Customers arriving in any other period leave immediately and will retry after an exponentially distributed time. Our main focus is on queue length analysis, both at embedded time points (beginnings of glue periods, visit periods and switch- or vacation periods) and at arbitrary time points.

1. Introduction

Queueing systems with retrials are characterized by the fact that arriving customers, who find the server busy, do not wait in an ordinary queue. Instead of that they go into an orbit, retrying to obtain service after a random amount of time. These systems have received considerable attention in the literature, see e.g. the book by Falin and Templeton [1], and the more recent book by Artalejo and Gomez-Corral [2].

Polling systems are queueing models in which a single server, alternatingly, visits a number of queues in some prescribed order. Polling systems, too, have been extensively studied in the literature. For example, various different service disciplines (rules which describe the server's behavior while visiting a queue) and both models with and without switchover times have been considered. We refer to Takagi [3,4] and Vishnevskii and Semenova [5] for some literature reviews and to Boon, van der Mei and Winands [6], Levy and Sidi [7] and Takagi [8] for overviews of the applicability of polling systems.

In this paper, motivated by questions regarding the performance modeling of optical networks, we consider vacation and polling systems with retrials. Despite the enormous amount of literature on both types of models, there are hardly any papers having both the features of retrials of customers and of a single server polling a number of queues. In fact, the authors are only aware of a sequence of papers by Langaris [9–11] on this topic. In all these papers the author determines the mean number of retrial customers in the different stations. In [9] the author studies a model in which the server, upon polling a station, stays there for an exponential period of time and if a customer asks for service before this time expires, the customer is served and a new exponential stay period at the station begins. In [10] the author studies a model with two types of customers: primary customers and secondary customers. Primary customers are all customers present in the station at the instant the server polls the station. Secondary customers are customers who arrive during the sojourn time

☆ This is an invited, considerably extended version of Boxma and Resing (2014) [28]. The main additions are Sections 2.4 and 3. These present respectively the optimal behavior of a single queue system, and the performance analysis for a general number of queues.

* Corresponding author.
E-mail addresses: m.a.abidini@tue.nl (M.A. Abidini), o.j.boxma@tue.nl (O. Boxma), j.a.c.resing@tue.nl (J. Resing).

http://dx.doi.org/10.1016/j.peva.2016.02.001
0166-5316/

of the server in the station. The server, upon polling a station, first serves all the primary customers present and after that stays an exponential period of time to wait for and serve secondary customers. Finally, in [11] the author considers a model with Markovian routing and stations that could be either of the type of [9] or of the type of [10].

In this paper we consider a polling station with retrials and so-called *glue* periods. Just before the server arrives at a station there is some deterministic glue period. Customers (both new arrivals and retrials) arriving at the station during this glue period "stick" and will be served during the visit of the server. Customers arriving in any other period leave immediately and will retry after an exponentially distributed time.

Our study of queueing systems with retrials and glue periods was at first instance motivated by questions regarding the performance modeling and analysis of optical networks. Optical fiber offers some big advantages for communication w.r.t. copper cables: huge bandwidth, ultra-low losses, and an extra dimension—the wavelength of light. Performance analysis of optical networks is a challenging topic (see e.g. Maier [12] and Rogiest [13]). In a telecommunication network, packets must be routed from source to destination, passing through a series of routers and switches. In copper-based transmission links, packets from different sources are time-multiplexed. This is often modeled by a single server polling system. In optical switches, too, one has the need for a protocol to decide which packet may be transmitted. One might again use a cyclic polling strategy, cyclic meaning that there is a fixed pattern for giving service to particular ports/stations. However, unlike electronics, buffering of optical packets is not easy, as photons cannot be stopped. Whenever there is a need to buffer photons, they are made to move locally in fiber loops. These fiber loops or fiber delay lines (FDL) originate and end at the head of a switch. When a photon arrives at the switch at a time it cannot be served, it is sent into an FDL, thereby incurring a small delay to its time of arrival without getting lost or displaced. Depending on the availability, requirement, traffic, size of photon and other such factors, the length (delay produced) of these FDLs can differ. Hence we assume that these FDLs delay the photons by a random amount of time. Also, if a packet does not receive service after a cycle through an FDL, then depending on the model it can go into either the same or a longer or a shorter or randomly to any of the available FDLs. Hence we assume that two consecutive retrials are independent of each other. This FDL feature can be modeled by a retrial queue.

A sophisticated technology that one might try to add to this is varying the speed of light by changing the refractive index of the fiber loop, cf. [14]. By increasing the refractive index in a small part of the loop we can achieve 'slow light', which implies slowing the packets. When a port 'knows' that it will soon be served, it may start the process of increasing the refractive index at FDLs and at the end of fibers of incoming packets. By doing this, it slows down the packets which arrive at the station just before the visit period of the station begins. This feature is, in our model, incorporated as glue periods immediately before the visit period of the corresponding station. Packets arriving in this glue period can be served in that subsequent visit period.

The concept of glue period is, to the best of our knowledge, new in polling systems. It may also be interpreted as a *reservation* period. We view a reservation period as a period in which customers can make a reservation at a station for service in the subsequent visit period of that station. In our case, such a reservation period occurs immediately before the visit period, and could be viewed as the last part, G_i, of a switchover period of length $S_i + G_i$. Ordinary gated polling could be viewed as a service discipline in which it is always possible to make a reservation for the following visit period.

The main contributions of the paper are the following. (i) For the case of a single queue with vacations and glue periods, we provide a detailed queue-length analysis at particular embedded epochs and at an arbitrary epoch. We also show how to choose the length of the glue period that minimizes the mean number of customers in the system. (ii) We also provide a detailed queue-length analysis for the N-queue polling case—again at particular embedded epochs and at an arbitrary epoch.

The paper is organized as follows. In Section 2 we consider the case of a single queue with vacations and retrials; arrivals and retrials only "stick" during a glue period. We study this case separately because (i) it is of interest in its own right, (ii) it allows us to explain the analytic approach as well as the probabilistic meaning of the main components in considerable detail, (iii) it makes the analysis of the multi-queue case more accessible, and (iv) results for the one-queue case may serve as a first-order approximation for the behavior of a particular queue in the N-queue case, switchover periods now also representing glue and visit periods at other queues. In Section 3 the N-queue case is analyzed. Section 4 presents some conclusions and suggestions for future research.

2. Queue length analysis for the single-queue case

2.1. Model description

In this section we consider a single queue Q in isolation. Customers arrive at Q according to a Poisson process with rate λ. The service times of successive customers are independent, identically distributed (i.i.d.) random variables (r.v.), with distribution $B(\cdot)$ and Laplace–Stieltjes transform (LST) $\tilde{B}(\cdot)$. A generic service time is denoted by B. After a visit period of the server at Q it takes a vacation. Successive vacation lengths are i.i.d. r.v., with S a generic vacation length, with distribution $S(\cdot)$ and LST $\tilde{S}(\cdot)$. We make all the usual independence assumptions about interarrival times, service times and vacation lengths at the queues. After the server's vacation, a *glue* period of deterministic (i.e., constant) length begins. Its significance stems from the following assumption. Customers who arrive at Q do not receive service immediately. When customers arrive at Q during a glue period G, they stick, joining the queue of Q. When they arrive in any other period, they immediately leave