ARTICLE IN PRESS

Performance Evaluation ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Performance Evaluation

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

Transient analysis of cycle lengths in cyclic polling systems

P. Vis [a,b,*], R. Bekker [a], R.D. van der Mei [b,a]

[a] VU University Amsterdam, Amsterdam, Netherlands

[b] Centre for Mathematics and Computer Science, Amsterdam, Netherlands

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Polling models
Transient analysis
Cycle times
Correlations
Waiting times

ABSTRACT

We consider cyclic polling models with gated or globally gated service, and study the transient behavior of all cycle lengths. Our aim is to analyze the dependency structure between the different cycles, as this is an intrinsic property making polling models challenging to analyze. Moreover, the cycle structure is related to the output of a polling model and the current analysis may be useful to study networks of polling models. In addition, transient performance is of great interest in systems where disruptions or breakdowns may occur, leading to excessive cycle lengths. The time to recover from such events is a primary performance measure. For the analysis we assume that the distribution of the first cycle (globally gated) or N residence times (gated), where N is the number of queues, is known and that the arrivals are Poisson. The joint Laplace–Stieltjes transform (LST) of all x subsequent cycles (globally gated) or all $x > N$ subsequent residence times (gated) is expressed in terms of the LST of the first cycle. From this joint LST, we derive first and second moments and correlation coefficients between different cycles. Finally, a heavy-tailed first cycle length or the heavy-traffic regime provides additional insights into the time-dependent behavior.

1. Introduction

Polling systems are multi-queue systems in which a single server visits the queues in some order to serve the customers waiting at the queues, typically incurring some amount of switch-over time to proceed from one queue to the next. Polling models find a wide variety of applications in which processing power (e.g., CPU, bandwidth, manpower) is shared among different types of users. Typical application areas of polling models are computer–communication systems, logistics, flexible manufacturing systems, production systems and maintenance systems; the reader is referred to [1] for an extensive overview of the applicability of polling models. Over the past few decades the performance analysis of polling models has received much attention in the literature. We refer to Vishnevskii and Semenova [2] for an excellent overview of the available results on polling models.

In this paper, we study cyclic polling models with gated or globally gated service, and focus on the transient behavior of the successive cycle times. Our goal is to gain an understanding in the dependency structure between the different cycles. This study is motivated by our interest in systems where disruptions or breakdowns may occur, often leading to excessive cycle lengths. In this context, we are interested in the following questions:

1. If the system encounters an excessively long cycle time (e.g., due to a disruption or a breakdown), then how will that influence the durations of the subsequent cycle times? What is the time needed to recover from excessive cycle times?

* Corresponding author at: VU University Amsterdam, Amsterdam, Netherlands.
E-mail address: petra.vis@vu.nl (P. Vis).

http://dx.doi.org/10.1016/j.peva.2015.06.018
0166-5316/

Please cite this article in press as: P. Vis, et al., Transient analysis of cycle lengths in cyclic polling systems, Performance Evaluation (2015), http://dx.doi.org/10.1016/j.peva.2015.06.018

2. What is the dependency structure between various residence and cycle times? More specifically, what is the correlation between the successive cycle (and residence) times?

A primary motivation for the second question is that the dependency structure makes polling models challenging to analyze. Insights into the dependency between cycles and residence times might pave the way for approximation methods. For instance, for polling models in tandem, the output of some queues may feed into another queue. The output of a specific queue in a polling system is essentially driven by an on–off source with dependent on and off times ('on' representing visit times and 'off' representing intervisit times). Similar relations have also motivated the study of some vacation models, see e.g. [3–5]. Finally, we note that waiting-time and queue-length distributions can be expressed in terms of the marginal cycle-time distribution for polling models with (globally) gated and exhaustive service.

In this paper, we assume that the distribution of the first cycle (in case of globally gated service) or N residence times (in case of gated service), where N is the number of queues, is known and that the arrivals are Poisson. Using this, we show how the joint Laplace–Stieltjes transform (LST) of all x subsequent cycles (globally gated) can be expressed in terms of the LST of the first cycle. Moreover, for the case of gated service we show how all $x > N$ subsequent residence times can be expressed in terms of the LST of the first cycle. From these joint LST's, we derive the first two moments and correlation coefficients between different cycles. Lastly, we analyze a heavy-tailed first cycle length, due to disruptions or breakdown, or the heavy-traffic regime to provide new fundamental insights into the time-dependent behavior.

The remainder of this paper is organized as follows. In Section 2 the models are described and the method and goals of the paper are outlined. In Section 3 we study the case of globally-gated service, whereas we study the case of gated service in Section 4. Both sections contain asymptotic results, such as heavy-tailed initial cycle lengths and heavy traffic, and numerical illustrations.

2. Model, method and goals

2.1. Model description

We consider a system of $N \geq 2$ infinite-buffer queues, $Q_1, \ldots, Q_N$, and a single server that visits and serves the queues in cyclic order. Customers arrive at Q_i according to a Poisson process $\{N_i(t),\ t \in \mathbb{R}\}$ with rate λ_i. These customers are referred to as type-i customers. The total arrival rate is denoted by $\Lambda = \sum_{i=1}^{N} \lambda_i$. The service time of a type-i customer is a random variable B_i, with LST $B_i^*(\cdot)$, and kth moment $\mathbb{E}[B_i^k],\ k = 1, 2, \ldots$, when it is finite. The kth moment of the service time of an arbitrary customer is denoted by $\mathbb{E}[B^k] = \sum_{i=1}^{N} \lambda_i \mathbb{E}[B_i^k]/\Lambda,\ k = 1, 2, \ldots$. The load offered to Q_i is $\rho_i = \lambda_i \mathbb{E}[B_i]$ and the total load offered to the system is equal to $\rho = \sum_{i=1}^{N} \rho_i$. The switch-over time required by the server to proceed from Q_i to Q_{i+1} is a random variable S_i with mean $\mathbb{E}[S_i]$ and LST $S_i^*(\cdot)$. Let $S = \sum_{i=1}^{N} S_i$, with LST $S^*(\cdot)$, denote the total switch-over time in a cycle. We define $\delta_i(s) := \lambda_i(1 - B_i^*(s))$ and let $\mathbf{e}_i$ be a unit vector with 1 in the ith position and 0 in the other positions.

We consider the gated and globally gated service disciplines. When the service discipline is gated, a gate at Q_i closes when the server arrives at Q_i. Every customer standing in front of the gate is served, while customers arriving at Q_i during service of Q_i must wait for the next cycle, this holds for all $i = 1, \ldots, N$. When the service discipline is globally gated, a gate closes at all queues when the server arrives at Q_1. During the following cycle, every customer standing in front of the gate is served.

2.2. Method and goals

Throughout we assume that the distribution of the length of the first cycle is known. For the gated service discipline, this requires that the joint distribution of the first N residence times is known, where a residence time is a visit time plus the subsequent switch-over time. When the probabilistic behavior of the first cycle is known, the next residence time can be expressed in terms of the first cycle, as it consists of a visit time to serve all the work that arrived at the queue during the first cycle plus the switch-over time. For globally gated, this is true for every queue, as the gate closes at the start of a cycle. For gated, the length of a visit time is always determined by the work that arrived at the corresponding queue during the last N residence times. It can be seen that the second cycle is completely determined in terms of the first cycle. Consequently, the third cycle can be expressed in terms of the second cycle and so also in terms of the first cycle. As a result, every cycle can recursively be expressed in terms of the first cycle. We use this fact to derive the joint LST of x consecutive cycles or residence times in terms of the LST of the first cycle.

Let us first consider the globally gated case. Our goal is to determine the joint LST of x consecutive cycle times, denoted by $\gamma_x(\mathbf{z})$. The vector $\mathbf{z}$ of length x contains the variables $z_1, \ldots, z_x$, corresponding to cycles $1, \ldots, x$, with the LST of the first cycle, $\gamma_1(\mathbf{z})$, assumed to be given. Choosing the z_i in specific ways, enables us to calculate all kinds of useful performance measures. For example, when $z_i = z$ for all $i \in J \subseteq \{1, \ldots, x\}$ and 0 otherwise, we obtain the LST of the sum of cycles of set J. Such a choice is especially convenient to calculate moments, which are then obtained by differentiating with respect to z and taking $z = 0$.

Also, the covariance between cycle 1 and cycle x can be calculated using the following property of the covariance: if X_1 and X_2 are random variables, then $\mathrm{Var}(X_1 + X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + 2\mathrm{Cov}(X_1, X_2)$, with the variance of a random variable

Please cite this article in press as: P. Vis, et al., Transient analysis of cycle lengths in cyclic polling systems, Performance Evaluation (2015), http://dx.doi.org/10.1016/j.peva.2015.06.018