ARTICLE IN PRESS

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

Contents lists available at ScienceDirect

Performance Evaluation

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

Optimal steady-state and transient trajectories of multi-queue switching servers with a fixed service order of queues

D.A.J. van Zwieten [a,*], E. Lefeber [a,*], I.J.B.F. Adan [b,a]

[a] Department of Mechanical Engineering, Eindhoven University of Technology, PO Box 513, 5600MB Eindhoven, Netherlands

[b] Department of Industrial Engineering and Innovation Sciences, Eindhoven University of Technology, PO Box 513, 5600MB Eindhoven, Netherlands

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Switching server
Optimization
Quadratic programming
Steady-state trajectory
Transient trajectory

ABSTRACT

The optimal scheduling problem of a system with two fluid queues attended by a switching server is addressed from two angles, the optimal steady-state and the optimal transient problem. The considered system includes features, such as setup times, setup costs, backlog and constraints on queue contents, cycle times and service times. First, the steady-state problem is formulated as a quadratic problem (QP), given a fixed cycle time. Evaluation of the QP problem over a range of cycle times results in the optimal steady-state trajectory, minimizing the total cycle costs or time average costs. Second, given initial conditions, we derive the optimal transient trajectory that leads to the optimal steady-state trajectory in a finite amount of time at minimal costs. For systems with backlog, we introduce additional costs on the number of cycles required to reach the steady-state trajectory in order to simplify the transient trajectory. The transient switching behavior and optimal initial modes are also addressed. Furthermore, we show by means of an example that the method can be extended to multi-queue switching servers.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

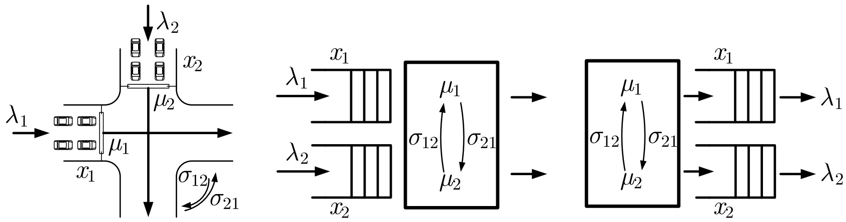
Optimal scheduling of systems with switching behavior is a problem of great importance. This problem, for even the most simple system, i.e., a server attending two queues, has been investigated by many researchers, see for example [1–9], and references therein. We follow the general framework introduced by [10] and model the production flow as continuous rather than discrete. The system considered in this paper is a server that can attend two fluid queues, where only a single queue can be attended at a time. Switching service to another queue might require a setup process, which might take time or involves switching costs. These systems arise in numerous contexts, such as manufacturing systems, signalized traffic intersections, computer communication networks and hospital rooms. Usage of optimal schedules can reduce time or costs. For instance, optimal schedules can reduce costs for manufacturing systems via lowering the required storage capacity and shortening lead times, or, for traffic signals at signalized intersections, optimal schedules can reduce congestion and thereby improve mobility and reduce the amount of environmentally harmful emissions.

In the aforementioned literature, the two queue switching servers are restricted in the sense that either setup times, setup costs, backlog or limited queue contents are required, omitted or not allowed. Also, most studies assume the simplifying

* Corresponding authors.
E-mail addresses: dirkvzwieten@gmail.com (D.A.J. van Zwieten), A.A.J.Lefeber@tue.nl (E. Lefeber), I.J.B.F.Adan@tue.nl (I.J.B.F. Adan).

http://dx.doi.org/10.1016/j.peva.2015.11.003
0166-5316/© 2015 Elsevier B.V. All rights reserved.

Please cite this article in press as: D.A.J. van Zwieten, et al., Optimal steady-state and transient trajectories of multi-queue switching servers with a fixed service order of queues, Performance Evaluation (2015), http://dx.doi.org/10.1016/j.peva.2015.11.003



(a) Signalized intersection with 2 flows.
(b) Switching server with 2 queues.
(c) Manufacturing system with 2 jobs.

Fig. 1. Different two queue switching server layouts.

condition that the system is symmetric, see [1,2,6,8]. In the current work, a two queue switching server is considered without restrictions on any parameters and with the flexibility of allowing setup times, setup costs and backlog, as well as constraints on cycle time, service time and queue contents.

In this paper, we divide the optimal scheduling problem into two subproblems: the derivation of optimal steady-state trajectories and the derivation of optimal transient trajectories. The current study is an extension of the work in [11]. Similar to [12,11], we formulate both subproblems as Quadratic Programming (QP) problems, with the addition of backlog and setup costs. Once the optimal steady-state trajectory is known, we study the best way of reaching it from any initial state, i.e., with minimal costs. This is a transient optimization problem, occurring for instance in case of a machine which is failure prone, or in case of a traffic intersection which gives priority to buses [13]. In these cases, we assume that deviations from the steady-state trajectory rarely occur, allowing the system to recover to the steady-state situation after each interruption. For systems without backlog and without capacity constraints, the policy for optimal transient behavior is presented. For systems with backlog, we introduce additional costs on the number of cycles required to reach the steady-state trajectory in order to simplify the transient trajectory.

Furthermore, we show by means of an illustration that the proposed methods can be extended to multi-queue switching servers, given the order of service of queues.

The remainder of this paper is organized as follows. Section 2 describes the system and presents the constraints. The optimal steady-state problem is addressed in Section 3 and examples of optimal trajectories are presented. In Section 4, the optimal transient problem is addressed. An illustration of optimal trajectories for a multi-queue switching server is presented in Section 5. Conclusions are provided in Section 6.

2. System description

We consider a system of two queues served by a single switching server. Fluid arrives at each queue $i = 1, 2$ with arrival rate λ_i. The content of queue i at time t is denoted by $x_i(t)$. The server is limited to serve only one queue at a time. If the server serves queue i, the service rate is given by $r_i \in [0, \mu_i]$. Three examples of the system under consideration are presented in Fig. 1, a signalized traffic intersection with two flows in Fig. 1(a), a 2-queue switching server in Fig. 1(b) and a 2-product manufacturing system in Fig. 1(c). The latter system has constant demands λ_i instead of constant arrivals.

Typically, switching service between different queues implies a setup process, either a *setup time* $\sigma_{i,j} \geq 0$ for switching from queue i to queue j, *setup costs* $s_{i,j}$ or a combination of these. For instance, a setup time can be reserved for vehicles to leave the intersection after the queue has received a red light (end of service), thereby preventing collisions, or for a machine to adjust configurations or to perform cleaning. In the latter case, also switching costs might be involved. A *cycle* consists of the setup and service of both queues. The *total setup time* in a cycle is denoted by $\sigma = \sigma_{1,2} + \sigma_{2,1}$ and the *total setup costs* in a cycle by $s = s_{1,2} + s_{2,1}$.

Given the setup times and cyclic behavior, we assume that the system can operate in four modes, denoted by $m \in \{1, 2, 3, 4\}$. Without loss of generality, the first mode, $m = 1$, indicates a setup to serve queue 1 and possible idling of the server, $m = 2$ indicates serving queue 1, $m = 3$ indicates a setup to serve queue 2 and possible idling of the server and $m = 4$ indicates serving queue 2. Note that for a system without setup times, i.e., $\sigma = 0$, modes 1 and 3 can have a duration of zero time units. The state x of the system not only consist of queue levels x_1 and x_2, but also of the *remaining idle time* x_0 (including setup times) and mode m, i.e., $x(t) = \begin{bmatrix} x_0(t), & x_1(t), & x_2(t), & m(t) \end{bmatrix}$.

A service time is defined as the uninterrupted interval during which the queue is served. The duration of a service time for queue n is nonnegative and is denoted by τ_n. Once the server is allocated to serve queue n, the server requires an *idle period* τ_n^0, which consists of the setup time and a possible idle time, i.e.,

$$\sigma_{j,n} \leq \tau_n^0, \quad n, j = 1, 2, \; n \neq j. \tag{1}$$

In this paper, we distinguish between a system with and a system without backlog. In case of backlog, denoting an accumulation over time of work waiting to be done or orders to be fulfilled, the contents of queue n can be negative and are therefore divided into an *inventory level* $x_n^+(t) = \max(x_n(t), 0)$ and a *backlog level* $x_n^-(t) = \min(x_n(t), 0)$. Hence, $x_n(t) = x_n^+(t) + x_n^-(t)$ and at each time instance either the backlog or inventory level is zero, i.e., $x_n^+(t)x_n^-(t) = 0, \forall t$. For

Please cite this article in press as: D.A.J. van Zwieten, et al., Optimal steady-state and transient trajectories of multi-queue switching servers with a fixed service order of queues, Performance Evaluation (2015), http://dx.doi.org/10.1016/j.peva.2015.11.003

Download English Version:

https://daneshyari.com/en/article/6888538

Download Persian Version:

https://daneshyari.com/article/6888538

Daneshyari.com