Optimal Control of Two-Stage Discrete Event Systems with Real-Time Constraints

Jianfeng Mao · Christos G. Cassandras

Received: 20 October 2006 / Accepted: 13 July 2007 / Published online: 24 August 2007 © Springer Science + Business Media, LLC 2007

Abstract We consider discrete event systems (DES) involving tasks with real-time constraints and seek to control processing times so as to minimize a cost function subject to each task meeting its own constraint. When tasks are processed over a single stage, it has been shown that there are structural properties of the optimal sample path that lead to very efficient solutions of such problems. When tasks are processed over multiple stages and are subject to end-to-end real-time constraints, these properties no longer hold and no obvious extensions are known. We consider a two-stage problem with homogeneous cost functions over all tasks at each stage and derive several new optimality properties. These properties lead to the idea of introducing “virtual” deadlines at the first stage, thus partially decoupling the stages so that the known efficient solutions for single-stage problems can be used. We prove that the solution obtained by an iterative virtual deadline algorithm (VDA) converges to the global optimal solution of the two-stage problem and illustrate the efficiency of the VDA through numerical examples.

Keywords Discrete event system · Optimal control · Real-time constraint

The authors’ work is supported in part by the National Science Foundation under Grant DMI-0330171, by AFOSR under grants FA9550-04-1-0133 and FA9550-04-1-0208, by DOE under grant DE-FG52-06NA27490, and by ARO under grant DAAD19-01-0610.

J. Mao (✉) · C. G. Cassandras
Dept. of Manufacturing Engineering and Center for Information and Systems Engineering, Boston University, Brookline, MA 02446, USA
e-mail: jfmao@bu.edu

C. G. Cassandras
e-mail: egc@bu.edu
1 Introduction

A large class of discrete event systems (DES) involves the control of resources allocated to tasks according to certain operating specifications (e.g., tasks may have real-time constraints associated with them). The basic modeling block for such DES is a single-server queueing system, whose dynamics are given by the well-known max-plus equation

\[ x_i = \max(x_{i-1}, a_i) + s_i(u_i) \]  

(1)

where \( a_i \) is the arrival time of task \( i = 1, 2, \ldots \), \( x_i \) is the time when task \( i \) completes service, and \( s_i(u_i) \) is its service time which may be controllable through \( u_i \). Examples arise in manufacturing systems, where the operating speed of a machine can be controlled to trade off between energy costs and requirements on timely job completion (Pepyne and Cassandras 2000); in computer systems, where the CPU speed can be controlled to ensure that certain tasks meet specified execution deadlines (Liu 2000); and in wireless networks where severe battery limitations call for new techniques aimed at maximizing the lifetime of such a network (Miao and Cassandras 2006). A particularly interesting class of problems arises when such systems are subject to real-time constraints, i.e., \( x_i \leq d_i \) for each task \( i \) with a given “deadline” \( d_i \). In order to meet such constraints, one typically has to incur a higher cost associated with control \( u_i \). Thus, in a broader context, we are interested in studying optimization problems of the form

\[
\min_{u_1, \ldots, u_N} \left\{ \sum_{i=1}^{N} \theta_i(u_i) \right\} \\
\text{s.t. } x_i = \max(x_{i-1}, a_i) + s_i(u_i) \leq d_i, \quad s_i(u_i) \geq s_{\text{min}, i}, \quad i = 1, \ldots, N
\]  

(2)

where \( \theta_i(u_i) \) is a given cost function assumed to be monotonically increasing in \( u_i \), \( s_i(u_i) \) is assumed to be monotonically decreasing in \( u_i \), \( s_{\text{min}, i} > 0 \) is a lower bound on the service time of task \( i \), and all \( a_i \), \( d_i \) are known. In general, this is a hard nonlinear optimization problem, complicated by the inequality constraints \( x_i \leq d_i \) and the nondifferentiable \( \max \) operator involved. Nonetheless, it was shown in Mao et al. (2004) that when \( \theta_i(u_i) \) is convex and differentiable the solution to such problems is characterized by attractive structural properties leading to a highly efficient algorithm termed critical task decomposition algorithm (CTDA). The CTDA does not require any numerical optimization problem solver, but only identifying a set of “critical” tasks in \( \{1, \ldots, N\} \). The efficiency of the CTDA is crucial for applications where small, inexpensive devices are required to perform on-line computations with minimal on-board resources.

Extending the problem in Eq. 2 to a network environment, where each node in the network is characterized by dynamics of the max-plus form in Eq. 1 coupled to those of other nodes, presents many challenges. As a first step, we consider in this paper a two-stage DES where tasks at the first stage satisfy

\[ x_{i,1} = \max(x_{i-1,1}, a_i) + s_{i,1}(u_{i,1}) \]  

(3)

and at the second stage:

\[ x_{i,2} = \max(x_{i-1,2}, x_{i,1}) + s_{i,2}(u_{i,2}) \leq d_i \]  

(4)