Available online at www.sciencedirect.com

ScienceDirect

Control Engineering Practice 15 (2007) 1169–1178

CONTROL ENGINEERING PRACTICE

www.elsevier.com/locate/conengprac

On tuning and complexity of an adaptive model predictive control scheduler

Mehran Mahramian[a,*], Hassan Taheri[a], Mohammad Haeri[b]

[a]Department of Electrical Engineering, Amirkabir University of Technology, Tehran, Iran
[b]Department of Electrical Engineering, Sharif University of Technology, Tehran, Iran

Received 19 January 2006; accepted 9 February 2007
Available online 30 March 2007

Abstract

In this paper, adaptive model predictive control is applied to schedule differentiated buffers in routers. The proposed algorithm, adaptive model predictive control scheduler (AMPCS), dynamically regulates the service rates of aggregated traffic classes. This algorithm guarantees some required constraints on proportional or absolute delay. The control parameters and the way they are adjusted as well as the problems of implementing the controller at high data rates are investigated. Theoretical analysis and numerical simulations demonstrate stability of AMPCS and its acceptable quality of service differentiations at core routers while maintaining end to end delay constraints.

Keywords: Quality of service; Differentiated services; Proportional delay; Scheduling; Model predictive control

1. Introduction

In the past few decades, the Internet has moved from a limited low-bandwidth network to a sophisticated infrastructure supporting many new applications. These applications require many different services in terms of delay, loss, bandwidth and jitter (delay variations). If the network provides an environment in which each user is satisfied with his/her received service based on a service level agreement, it is said that the network provides quality of service (QoS). Several architectures such as Differentiated Services (DiffServ) (Blake et al., 1998; Nichols, Blake, Baker, & Black, 1998) and Integrated Services (IntServ) (Braden, Clark, & Shenker, 1994; Braden, Zhang, Berson, Herzog, & Jamin, 1997) have been proposed to improve the QoS characteristics. IntServ is an architecture in which the required bandwidth is reserved for each flow between source and destination in all network nodes. Unfortunately, this is not a scalable scheme and may not be used in large networks. The other approach is DiffServ, which aggregates the traffic with similar quality requirements into the same traffic class. So, the network nodes could reserve resources and provide QoS for a limited number of aggregated traffic classes.

During the past decade, much effort was spent on providing priority for some classes in DiffServ architecture. The service is defined qualitatively, i.e. higher priority classes receive better QoS. DiffServ has sacrificed accuracy to improve scalability of IntServ architecture. In many recent works the DiffServ accuracy is strengthened besides reduction in the complexity of the network especially at core routers (Dovrolis, 2000; Mahramian, Taheri, & Haeri, 2005; Stoika & Zhang, 1998; Striegel & Manimaran, 2002). They assume nothing about traffic shaping and minimize or avoid on line negotiations about traffic conditions between routers. Scheduling at the routers is the main concern of these algorithms to satisfy specific delay constraints for some aggregated traffic classes (Fig. 1).

An idea to increase QoS accuracy while preventing complexity was proposed by Dovrolis (2000). He assumed

---

*Corresponding author. Informatics Services Corp., Shangarf Alley, Mirdamad Blvd., Tehran, Iran. Tel.: +98 862 4222226, +98 912 3861450; fax: +98 21 29982855.

E-mail addresses: m_mahramian@isc.iranet.net (M. Mahramian), htaheri@aut.ac.ir (H. Taheri), haeri@sina.sharif.edu (M. Haeri).

0967-0661/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.conengprac.2007.02.002

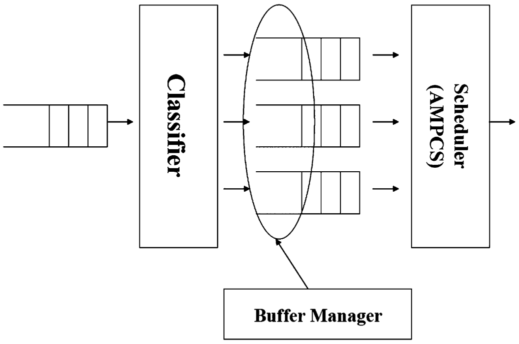

Fig. 1. Basic parts of a router in DiffServ architecture.

proportional delay and loss guarantees for different traffic classes instead of an absolute delay or loss for each class. This means one can guarantee that delay of the higher priority service is not more than a constant fraction of the delay of the lower priority one in addition to absolute boundaries for delay. This is done using efficient scheduling algorithms. The proportional average delay (PAD), waiting time priority (WTP) and hybrid proportional delay (HPD) scheduler are the proposed scheduling algorithms in Dovrolis (2000) that guarantee proportional delay, but they are not accurate enough for strong QoS.

The scheduling algorithms can be generally placed in two groups. In the first group, service rate is dynamically changed according to the number of waiting packets in each queue. Algorithms of this group are basically the advanced versions of generalized processor sharing (GPS) (Parekh & Gallagher, 1993). In the second group, priorities of queues are adaptively modified to satisfy the limits. WTP (Dovrolis, 2000) and mean delay proportional (MDP) (Abdelzaher & Lu, 2000) are some examples.

Robustness and accuracy of algorithms in communication networks and throughput of the communication links have been the main problems of the computer networks. Recently, control theoretic approach has been widely utilized as a powerful tool to solve these problems (Chisci, Pecorella, & Fantacci, 2006; Kim, Shin, & Kwon, 2004; Priscoli & Isidori, 2005). Most researchers use classic and linear control tools for these purposes (Gu, Wang, Hong, & Bushnell, 2001). In this paper, the model predictive controller (MPC), that uses a model to predict the system behavior, is employed. Using the adaptive scheme, the model and the controller are improved as time goes on. The high dynamism of the system of the DiffServ queues corresponds to the system behavior changes regarding traffic burstiness, traffic average, backlog (number of packets in the queue) variations and some events like congestion and packet loss. In a normal network, parameter changes are not so fast in the short durations in which the traffic is stable and there is no special event like congestion. In some instants, the parameters change quickly due to some of the above-mentioned events or traffic change. So, the system parameters change slowly in short durations and the adaptive model is applicable, but there are also some rapid changes of parameters in long term so that adaptive modeling can also cope with them. Besides, the controller is model predictive control, which tolerates some errors in model parameters. The joint use of adaptive model and model predictive control can make the controller perform more appropriately in the above situation. The other advantage of the adaptive control is that it omits the need for higher order models (which are usually assumed complex in the parameter estimation). In other words, the controller copes with the system variations using only local models with lower order. The proposed algorithm, adaptive model predictive control scheduler (AMPCS), uses indirect adaptive MPC to update model parameters in each sample time and determines control signal accordingly.

Although MPC is one of the most popular modern control methods having significant impact on industrial processes, its application in computer network algorithms has rarely been reported. The reason might be its complexity from the network designers' point of view as well as implementation problems. Search results show that no previous research has used model predictive control to solve the scheduling problem. Many previous works in networking show that the algorithms with prediction provide better results. This was an idea to apply MPC to the scheduling problem to see if the delay and its variations can be controlled precisely.

The rest of the paper is organized as follows. Together with a brief explanation of the MPC, systemic approach to the given problem is discussed in Section 2. The AMPCS algorithm is presented in Section 3 and its computational complexity is investigated in Section 4. Simulation results are presented in Section 5 and, finally, Section 6 concludes the paper.

2. System model

Modeling is an important step in using the control theoretical approach. A good model can bring about proper results, while an inaccurate model may result in an unacceptable or unstable controller. In this section, the proposed model for DiffServ queues is described.

2.1. System description

Consider a router in which packets are arriving from input ports and departing from output ports. There are some blocks in a router such as the packet classifier, the IP lookup, the buffer manager and the scheduler (Fig. 1). Packets entering the router have labels indicating the class of traffic they belong to. The classifier places the packets in the related queues. The packets must wait in the queues until the scheduler decides to serve them. The system contains Q queues; packets in high-priority queues should encounter less delay than those in the low-priority ones. Note that while the actual input and output of the queue