Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Computer Communications xx (2005) 1–17
computer communications
www.elsevier.com/locate/comcom

Graceful degradation of loss-tolerant QoS using (*m*,*k*)-firm constraints in guaranteed rate networks☆

Anis Koubâa*, Ye-Qiong Song

LORIA-TRIO-INPL, Computer Science, 615, rue du jardin botanique, Villers Les Nancy 54602, France

Received 12 March 2004; revised 24 January 2005; accepted 28 January 2005

Abstract

The Weighted Fair Queueing (WFQ) scheduling algorithm and its variants can be used to provide real-time guarantees by making bandwidth reservation. However, while hard guarantees are based on a peak workload model that leads to underutilize network resources, soft guarantees, based on average workload model, are not always sufficient to maintain acceptable Quality-of-Service (QoS) since consecutive packet losses or deadline misses may occur, which are not suitable for real-time applications. In this paper, we propose a trade-off between hard and soft real-time guarantees to maintain an acceptable QoS in overload conditions and efficiently maximize the utilization of network resources. The key to our solution is based on the fact that many real-time applications, such as voice and video, are loss-tolerant. The loss profile must be well defined, and such a profile can be easily specified using the (*m*,*k*)-firm model. Therefore, we propose the (*m*,*k*)-WFQ algorithm to take into account (*m*,*k*)-firm timing constraints to provide delay guarantees of at least m packets out of any *k* consecutive packets without violating bandwidth fairness or misusing network resources. Using the Network Calculus theory, an analytic study gives the deterministic delay bound provided by the (*m*,*k*)-WFQ algorithm for upper bounded arrival curve traffic. We extend our analytic results for guaranteed-rate networks, such as the IntServ QoS model and ATM networks, and the DiffServ QoS model. Analytic results and simulations show a noticeable improvement in delay guarantee made by (*m*,*k*)-WFQ compared to WFQ without much degradation of bandwidth fairness.

Keywords: Fair queueing; (*m*,*k*)-Firm; Delay guarantee; Bandwidth fairness; QoS; Network Calculus

1. Introduction

Bandwidth guarantee has been widely used in Internet Quality of Service (QoS) architecture as well as Asynchronous Transfer Mode (ATM) networks to make deterministic real-time guarantees for time-sensitive applications. However, hard guarantee is not without cost since it requires deterministic predictability on network delays and thus underutilizes network resources. Soft real-time guarantee, on the other hand, maximizes resource utilization but has less stringent guarantee and QoS may degrade drastically in overload conditions.

For real-time applications, the important QoS metrics are delay and loss. To ensure short delay and loss-free (hard) guarantee, the peak-rate bandwidth reservation is commonly used, especially for Variable Bit-Rate (VBR) flows whose carried burst size is quite large. This kind of reservation overstates the bandwidth requirement and reduces network utilization. A second approach consists in making an average-rate bandwidth reservation to maximize resource utilization but the guaranteed delay may be longer than the application requirement when the burst size is quite large, resulting in many deadline misses.

Weighted Fair Queueing (WFQ) [10] and its variants [11–13] are the basic algorithms used to make bandwidth guarantee. Several studies have dealt with improving the delay guarantee provided by WFQ for real-time flows with high burst size. Wang et al. [2] proposed a technique called Priority-based WFQ (PWFQ), which combines fixed-priority assignment and WFQ scheduling algorithm in order to better manage the delay bounds for various flow sessions. The problem was how to guarantee short delays for flows with low service share. The approach consists in

☆ Part of this work is accepted for presentation at the IEEE Real-Time and Embedded Technology and Applications Symposium (RTAS'2004).

* Corresponding author. Tel.: +33 3 83 58 17 66; fax: +33 3 83 59 56 62.

E-mail address: akoubaa@loria.fr (A. Koubâa).

0140-3664/$ - see front matter
doi:10.1016/j.comcom.2005.01.019