Available online at www.sciencedirect.com
SCIENCE DIRECT®
Computer Networks 48 (2005) 137–154
Computer Networks
www.elsevier.com/locate/comnet

Statistical admission control for real-time services under earliest deadline first scheduling

Zhi Quan, Jong-Moon Chung *

School of Electrical and Computer Engineering, Oklahoma State University, 202 Engineering South, Stillwater, OK 74078, United States

Received 18 July 2003; received in revised form 3 March 2004; accepted 20 October 2004
Available online 12 January 2005

Responsible Editor: E. Knightly

Abstract

The design of call admission control (CAC) mechanisms has been a critical issue in providing quality-of-service (QoS) guarantees for heterogeneous traffic flows over integrated service (IntServ) and/or differentiated service (DiffServ) networks. Earliest deadline first (EDF) is an ideal scheduler for real-time services because of its optimal admissible region and delay bound properties. The major difficulty in developing an effective and efficient CAC algorithm for statistical services is the analysis of per-class deadline violation (loss) probabilities with respect to the delay bounds. In this paper, we provide an analytical approach to evaluate the aggregate and per-class deadline violation probabilities of an EDF scheduler. Based on these theoretical foundations, we derive the admission control conditions and then propose a CAC algorithm for statistic services under EDF scheduling. In addition, we show that the QoS metrics that an EDF scheduler actually guarantees have an asymptotic ordering property, which provides an important insight into the design and control of EDF networks. The effectiveness and performance of our proposed algorithm have been validated by trace-driven simulation experiments using MPEG and H.263 encoded video sources.

Keywords: Call admission control; Quality-of-service; Earliest deadline first; Deadline violation probability; Dominant time scale

1. Introduction

Future high speed packet-switching networks are expected to provide heterogeneous QoS guarantees for a variety of applications. Call admission control plays a critical role in achieving this purpose, which is an integration of the traffic models, scheduling disciplines, and QoS specifications. Its major task is to decide whether a new connection can be granted while the QoS requirements of this new connection and all the existing connections can still be guaranteed. This becomes much more

* Corresponding author. Tel.: +1 405 744 9924; fax: +1 405 744 9198.
E-mail address: jmc29@alumni.psu.edu (J.-M. Chung).

1389-1286/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.comnet.2004.10.017

complicated and challenging for the networks supporting various applications with heterogeneous QoS requirements.

The scheduler is an indispensable part in the design of call admission control mechanisms. Among the existing scheduling disciplines in the literature, static priority (SP), generalized processor sharing (GPS) [1], also known as weighted fair queueing (WFQ), and EDF [2] are probably the most promising to provide QoS guarantees for heterogeneous traffic. Although GPS has been considered an ideal scheduling discipline in terms of its combined delay and fairness properties, it has been shown to be sub-optimal in terms of the schedulable [1] region by the studies in [3]. On the other hand, EDF has been shown in [4–8] to be able to offer a substantial performance gain over GPS under both deterministic and statistic contexts. In particular, the authors of [8] have proved that GPS inherently requires dynamic weight re-synchronization in order to realize the maximum schedulable region. In this paper, we will focus on the CAC issues for a network node deploying the EDF scheduling discipline.

Scheduling disciplines for call admission control were originally studied in the deterministic setting (see [1,3,9,10]). These deterministic frameworks are intrinsically conservative, since they have to consider the worst cases that might occur in the schedulers with very low probabilities. As a result, the network utilization achievable will be very low for bursty traffic. Therefore, they are limited to only applications requiring deterministic services (without packet loss), and are not appropriate for adaptive statistical services that can both tolerate and adapt to certain amounts of loss and end-to-end delay. Due to this, the scheduler behaviors under the context of statistical multiplexing have received a lot of attention in various aspects. In [11], Elwalid et al. analyzed the first-in-first-out (FIFO) scheduling using the effective bandwidth with Chernoff bounds, where the traffic patterns studied were independent and periodic ON–OFF sources. The results were then extended for GPS scheduling by the authors of [12,13], and for the EDF discipline in [7]. On the other hand, some researchers have dedicated to extend the deterministic models to the statistical setting. The authors of [14] introduced the concept of effective envelopes and devised the admission control tests for a set of scheduling algorithms (i.e., FIFO, SP, and EDF). Qiu and Knightly [15] proposed using adaptive and measurement-based maximal rate envelopes to characterize the traffic processes for multi-class networks. Other results on statistical multiplexing include [16,17] for SP, [18,19] for GPS, and [20] for EDF.

Rigorous investigation of statistical performance metrics for EDF scheduling dates back to [8] and [20]. By assuming that all delay bounds are close together (i.e., about the same order of magnitude) and considering traffic models with Markovian sources, the authors of [20] derived the aggregate deadline violation probabilities. However, the assumptions in [8] and [20] limit the extension to various delay requirements of different flows. Also, the long-range dependent or self-similar characteristic [21] of the real network traffic has made the previous analysis of EDF scheduling face a fundamentally different set of problems. To develop an efficient statistical CAC scheme for the EDF scheduler, the analysis of per-flow loss probabilities with respect to the delay bounds becomes necessary. It is well known that the analysis of a stochastic system such as an EDF (or GPS) scheduler, where several flows with different QoS requirements share a server and a queue, is generally very difficult. The reason behind this is that the amount of service received by a flow at any moment depends not only on its arrival process and content in the queue at that moment, but also on the arrival processes of other flows and their contents in the queue [22]. To circumvent this difficulty, the authors of [8] developed a mechanism to obtain per-flow metrics by exploiting a fair packet discard scheme, which makes the implementation more costly.

Recently, the studies on the statistical analysis of EDF scheduling [23] have present an efficient and accurate framework, within which the asymptotic solutions for the aggregate and per-class metrics have been derived via defining a *virtual*

[1] Throughout this paper, we will use the notations of schedulable and admissible interchangeably.