Available online at www.sciencedirect.com
ScienceDirect
Computer Networks 51 (2007) 2015–2032
Computer Networks
www.elsevier.com/locate/comnet

# Consistent proportional delay differentiation: A fuzzy control approach

Jianbin Wei [a,*], Cheng-Zhong Xu [b]

[a] *Mathematics and Computer Science Department, South Dakota School of Mines and Technology, Rapid City, SD 57701, United States*
[b] *Department of Electrical and Computer Engineering, Wayne State University, 650 Merrick Street, Detroit, MI 48202, United States*

Received 20 November 2004; received in revised form 15 June 2006; accepted 16 October 2006
Available online 9 November 2006

Responsible Editor: M. Buddhikot

**Abstract**

Proportional delay differentiation (PDD) is an important service model for providing relative differentiated services on the Internet. It aims to maintain pre-specified packet queueing-delay ratios between different classes of traffic at each hop. Existing rate-allocation approaches for PDD services assume the average queueing delay of a class is inversely proportional to its service rate. This assumption is not necessarily valid when the system is not heavily loaded. To provide *consistent* PDD services under various load conditions, in this paper, we propose a novel rate-allocation approach that applies fuzzy control theory to capture the nonlinear relationship between the queueing delay and the service rate. In the approach, a class's service rate is adjusted according to a set of fuzzy control rules defined over its error (the difference between the target delay ratio and the achieved one), the change of error, and the change of service rate. We prove that the fuzzy control system is stable and the service rate of a class converges to its equilibrium point at steady state. Simulation results demonstrate that, in comparison with other rate-allocation approaches, the fuzzy control approach is able to provide consistent PDD services under wide range load conditions. It is also shown robust under various system conditions, including with multiple classes, changing target delay ratios, changing load conditions, and different traffic patterns.


*Keywords:* Quality of service; Proportional delay differentiation; Fuzzy control; Rate allocation

## 1. Introduction

The past decade has seen an increasing demand for provisioning of different levels of quality of service (QoS) to various network applications and customers. Differentiated services (DiffServ) [2] is a major service architecture in order to meet this requirement. In the architecture, two service models have been proposed: absolute DiffServ, which is to guarantee the end-to-end QoS [12,14]; and relative DiffServ, which is to quantify the QoS differences between different classes. Although absolute

* Corresponding author. Tel.: +1 605 394 2470.
*E-mail addresses:* Jianbin.Wei@sdsmt.edu (J. Wei), czxu@wayne.edu (C.-Z. Xu).

1389-1286/$ - see front matter 
doi:10.1016/j.comnet.2006.10.005

DiffServ is desirable for Internet services (e.g., Internet telephony) that have hard time constraints, relative DiffServ is sufficient for soft real-time applications, such as e-commerce transactions and multimedia applications, which may last from several seconds to tens minutes. In relative DiffServ service model, the network traffic is divided into a number of classes. A higher ranked class should receive better or at least no worse services than a lower ranked class in terms of local service metrics, such as the queueing delay and the loss rate [7]. In particular, proportional delay differentiation (PDD) service model is to maintain pre-specified queueing delay ratios between different classes [8].

Existing approaches for PDD services can be classified into two categories: dynamic-priority approaches and rate-allocation approaches. In dynamic-priority approaches, a class's priority is adjusted according to its current measured states, such as its head-of-line delay [8,16], the average delay of its departed packets [9], or the delay of all arrived packets [19,24]. Although most of the approaches are capable of providing consistent PDD services under various load conditions, they introduce significant overhead because of their needs for calculating the priorities of all backlogged classes to determine the one with the highest priority upon *every* packet departure. Moreover, the dynamic-priority approaches are incompatible with those current-in-use routers because they have different queueing disciplines, such as class based weighted fair queueing and modified deficit round robin (MDRR) implemented by Cisco [5]. This incompatibility makes it difficult to deploy these approaches in practice, and necessitate the design of simpler approaches.

In rate-allocation approaches, a class's service rate is adjusted periodically based on current system states [4,8,17]. For example, in backlog-proportional rate (BPR), a class's service rate is adjusted according to its backlogged queue length [8]; in joint buffer management and scheduling (JoBS), the service rate is set based on delay predictions of its backlogged traffic [17]. In [4], the authors proposed a linear feedback control approach (referred to as LFB in the rest of this paper), in which a class's service rate is adjusted according to the difference between its normalized head-of-line delay and the average of all backlogged classes.

Rate-allocation approaches incur smaller overhead than dynamic-priority approaches because they adjust service rates only when needed. Moreover, they are implemented with the same queueing disciplines as those used in current-in-use routers and thus can be deployed easily.

Existing rate-allocation approaches, however, are able to deliver PDD services only when the system is heavily loaded. For example, BPR showed good PDD services only when the system load becomes higher than 90% [8]; performance studies of JoBS and LFB assumed the system load to be as high as 99% [4,17]. We tested the approaches in non-heavily loaded systems and found that none of them was able to achieve the goal of PDD services consistently under light and medium load conditions. It is because they assume that the average queueing delay of a class is inversely proportional to its service rate. This assumption is valid during busy periods of the system. It is not always the case if there exist idle periods in the queueing system.

We note that consistent PDD services under various load conditions are desirable. It is known that the load of a system (e.g., the link utilization) in practice is often light or moderate [3]. Without guarantee of the consistency, it would become less incentive for high priority customers to pay more for the same level of QoS most of the time. Dynamic-priority approaches achieved the goal of consistency at a cost of frequent priority changes.

For consistent PDD services, a rate-allocation approach should be able to capture the nonlinear relationship between the queueing delay and the service rate. Although lots of work exist for network traffic characterization (see [21,27] for examples), there is still lack of accurate and solvable mathematical models for the relationship. Classical control methods approximate nonlinear systems by linear ones with a compromise of accuracy, as we push the performance envelope of the controller.

In this paper, we propose a novel rate-allocation approach based on fuzzy control. Fuzzy control theory provides a formal methodology for representing, manipulating, and implementing a human's heuristic knowledge about how to control a system. In the fuzzy control approach, we quantify heuristic control knowledge developed from system behavior analysis into a set of control rules. The rule-base states the relationship between the error (the difference between the target delay ratio and the achieved one), the change of error, and the change of service rate. A class's service rate is adjusted according to the rules.

We prove that the fuzzy control system is stable and the service rate of a class converges to its equilib-