P

Pricing Quality of Service in Diffserv IP Networks

Vladanka S. Acimovic-Raspopovic
University of Belgrade, Serbia

Mirjana D. Stojanovic
University of Belgrade, Serbia

INTRODUCTION

In order to reduce costs and make it easier to integrate disparate systems, networked and virtual organizations (NVOs) should adopt open technology standards throughout the entire organization—standard computing architectures, standard networks, and standard application interfaces. The Internet protocol (IP) technology has been foreseen as a basic networking infrastructure that supports communication requirements of NVOs. With the growing demand for the integration of heterogeneous telecommunication services (e.g., voice, data, video and multimedia), there is a strong need for deploying quality of service (QoS) in IP-based networks. Under such circumstances, the flat pricing models that have been satisfying in traditional best-effort Internet so far do not encourage users to make reasonable use of resources. QoS differentiation introduces a clear need for incentives to be offered to users to encourage them to choose the service that is most appropriate for their needs. In commercial networks, this can be most effectively achieved through pricing.

Falkner, Devetsikiotis, and Lambadaris (2000) and Da Silva (2000) supplied comprehensive reviews and evaluation of pricing schemes developed during the nineties and mainly related with per-flow IP QoS approaches, such as the integrated services (IntServ) specified by Braden, Clark, and Shenker (1994). Proliferation of the differentiated services framework, since the late 90s posed a number of new issues and resulted in novel proposals for pricing IP QoS and network resources.

THE DIFFERENTIATED SERVICES FRAMEWORK

The differentiated services (DiffServ) approach to QoS provisioning comprises standard-compliant (Blake et al., 1998) or proprietary architectures which have a common property that packets belonging to different traffic flows, but with similar QoS requirements, may be associated to the same *service class* and processed at the network nodes in the same manner. The basic idea is that IP packet headers include a field called DiffServ code point (DSCP). The DSCP allows packets to be classified and identifies the specific per-hop behavior (PHB) that determines packet scheduling and drop precedence at a particular network node. Edge routers are configured with a large number of per-flow conditioning policies, which include metering, marking, shaping and dropping of the ingress traffic. Core routers are configured to perform fast and simple queuing and scheduling operations, with a few forwarding classes, thus avoiding resource-exhaustive handling of per-flow states.

In the DiffServ network, QoS is negotiated between the provider and the user (end user or another IP domain) for each traffic flow and this process results in the contract called service level agreement (SLA). The technical part of the SLA contains parameters that describe the particular traffic profile and traffic conditioning rules, as well as service class in the sense of QoS guarantees, performance metrics, reliability, availability, service scheduling, and so forth. SLA also includes information about measurement methods, tariff and billing principles, as well as penalties for both user and provider in the case of contract violation.

The DiffServ approach is today widely accepted as a basis for providing QoS in large networks because of its simplicity, scalability and capability to operate over the existing IP-based infrastructure. It is also foreseen as a basic QoS model of networking infrastructure for *grid computing* (Sander, 2004; Zeng et al., 2004). However, providing end-to-end QoS in the IP-based networks requires implementation at the network level a variety of additional mechanisms and algorithms dealing with the call and packet handling, as well as network resource management. As a consequence, enhanced QoS architectures based on the DiffServ concept have been investigated and tested in research networks (Giordano, Salsano, Van den Berghe et al., 2003).

A SURVEY AND COMPARISON OF PRICING SCHEMES IN DIFFSERV-BASED NETWORKS

Accounting, pricing, charging and billing are basic processes that constitute *accounting management* in IP-based networks (Acimovic-Raspopovic & Stojanovic, 2005; Pras, Van Beijnum, Sprenkels, & Parhonoyi, 2001). The accounting process involves monitoring, measuring and collecting of resource usage information, related to a single customer's service utilization. It provides the accounting data, which represent the collection of resource consumption data, for the purposes of pricing, charging and provisioning information to users. Pricing is the process of allocating tariff, that is cost per unit of resource. It is based on particular pricing scheme and controlled by a pricing *policy*. The pricing process provides pricing data as input to charging process, which then determines the amount to be charged for use of a resource, based on tariff and the accounted data. The billing process produces an invoice on the basis of charging data, while the payment results in the actual transfer of money.

Although different classifications of pricing schemes are possible (Yang, 2004), we will further address two main categories of schemes, according to their dependency on the actual state of network resources and delivered QoS: static pricing and dynamic pricing. We will also discuss hybrid schemes that represent combination of static and dynamic models.

Static Pricing Schemes

Static pricing assumes that price is independent on actual network resource utilization (Da Silva, 2000). Such definition is broad enough to encompass changes of tariffs according to time of day or week.

The simplest static scheme is Paris-metro pricing (PMP), which has been proposed by Odlyzko (1999). The PMP assumes dividing of the total bandwidth capacity into a set of several (typically three or four) differently priced logical channels. Within each channel, flat pricing is applied, that is user is charged a fixed amount of money per time-unit. In the DiffServ network, logical channels correspond to service classes. For each traffic flow, prices are specified through the SLA, according to service class to which the flow corresponds.

In the broader sense, static approach corresponds to the expected capacity pricing, introduced by Clark (1997). Expected capacity represents the user's expectations in the sense of one or more performance metrics, when the network is congested. It can be specified in many ways, including minimum required capacity, maximum delay or an effective bandwidth (Chao & Guo, 2002; Courcoubetis, Kelly, Siris, & Weber, 2000). In the DiffServ environment, the expected capacity should be specified in the corresponding SLA and the user should be charged according to it.

The main advantage of static pricing refers to its simplicity — a long-term SLA should be used to anticipate resource utilization; hence no measurements of actual resource usage are required. For that reason, charging and billing of each user can be easily performed at the network edge. Studies on static pricing (Cocchi, Estrin, Shenker, & Zhang, 1993; Da Silva, Petr & Akar, 2000) have mainly investigated the efficiency of different policies to network performance, customer's choice, provider's revenue, and so forth.

Deficiencies of static approach in the DiffServ network mainly concern lack of support for *congestion management* and problems with charging of users who either under-utilize or over-utilize resources that correspond to their contracted traffic profiles. In the former case, a user would be charged for more resources than it actually consumes. In the latter case, there is a risk of exhaustion of resources, and consequently deterioration of QoS delivered to legitimate customers.

Dynamic Pricing Schemes

Dynamic pricing assumes that price is determined as a cost per unit of resource consumption (on a packet level, on a flow level or on an aggregate level) and according to level of QoS guarantees provided for the observed service class (Bouras & Sevasti, 2004). Well-known dynamic pricing schemes encompass different variants of auction-based mechanisms and responsive pricing.

Auction is a pricing approach which "consists of submitted bids by clients who specify the desired amount of resources and the price they are willing to

Related Content