A New Routing Architecture for DiffServ Domains

Shiping Chen
Network Center
Univ. of Shanghai for Science and Technology
Shanghai, China
chensp@usst.edu.cn

Yibei Ling
Applied Research Laboratories
Telcordia Technologies
New Jersey, USA
lingy@research.telcordia.com

Shigang Chen Yong Tang
Dept. of Com. & info. Sci. & Tech.
Univ. of Florida
Gainesville, Florida, USA
{sgchen,yt1}@cise.ufl.edu

ABSTRACT

This paper [1] proposes a new DiffServ routing architecture (PAP) that integrates the admission control signaling and the QoS routing. It differs from traditional routing in its ability to route most Expedite Forwarding (EF) traffic along the shortest paths while making use of alternative paths to absorb transient overload. Once an EF data flow is admitted, its performance is assured. The overhead for storing alterative path information is minimal since only one routing entry at a branching point is needed for each alternative path. The route map of Cisco IOS provides a mechanism for implementing PAP.

KEY WORDS

Communications Protocol, Quality of Service, Routing, Admission Control

1 Introduction

As the effort of converging voice and data into a single network accelerates, a wide range of multimedia applications emerges. The requirement for timely delivery of digitized audio-visual information raises new challenges for the broadband networks. Research on providing quality-of-service (QoS) in data networks have continued to be a hot area in the past decade.

It is a challenging problem to commit network resources in a scalable way so that the delay or throughput sensitive traffic is appropriately treated as they are routed across the network. The solution to this problem requires different system functions to cooperate with each other. Admission control ensures that the total traffic in the network does not overwhelm the available resources. Traditional routing protocols make sure that the packets get to their destinations, while QoS routing protocols [1] make sure that the QoS traffic is well spread on different paths to increase the utilization of the network's capacity. Packet scheduling and resource management at the routers allow differentiated treatment for packet streams with varied service requirements.

The work on QoS support roughly falls in two broad categories: Integrated Services (IntServ) and Differentiated Services (DiffServ). At the heart of IntServ is RSVP [2]. In RSVP, the required resources are reserved at every router along the path of a traffic stream (flow), and hence the performance of the traffic stream is guaranteed. Such a fine level of per-stream resource reservation is very flexible in supporting various QoS requirements, e.g., guaranteed bandwidth, bounded end-to-end delay, and bounded delay jitter. However, this approach is not scalable as per-stream information is stored at the core routers, which may take thousands or even millions of traffic streams simultaneously.

DiffServ solves this problem by making a different tradeoff. It pushes the complexity to the edge of the network, where the data traffic is classified into different service classes by setting a codepoint in the IP header of every packet. At the interior of a DiffServ domain, the packets are treated according to the service classes they belong to. In this way, the per-stream information is eliminated inside the domain. The resource management (e.g., packet scheduling) is conducted at a coarse level based on service classes.

Most work on DiffServ [3, 4, 5, 6] has focused on defining service classes, their individual and relative properties, per-hop behaviors, and implementation-related issues. The routing support for DiffServ is still an open problem. Although the design of DiffServ is independent from routing, the routing function has significant impact on the actual effectiveness of some DiffServ service classes such as the Expedite Forwarding [4]. In fact, the routing function directly affects the admission control, which determines the traffic volume that a DiffServ network can accommodate for each service class without violating the service contract.

[1]This research is funded by the Development Foundation of Shanghai Education Committee (Grant No. 01F10).

The traditional shortst-path-only routing is good for the best effort traffic but too restrictive for QoS traffic, because there may exist plenty of other paths that can support the required QoS when the shortest path cannot. In the recent years, there was a large body of research on QoS routing [7, 1, 8], which was designed primarily for a RSVP-like environment. Before a traffic stream is delivered, the sender or the receiver activates the QoS routing protocol to establish a routing path that has the required resources. A routing entry is inserted to the routing table at every router on the path. These entries are deleted when the traffic stream terminates. The above approach has the same problem of RSVP; it deposits per-stream information at the routers. An alternative is for each packet to carry the routing path in the source-routing IP option, which however causes excessive overhead on the routers.

In this paper, we propose a new routing architecture for DiffServ, particularly for the Expedite Forwarding Service. The idea behind is simple: We rely mainly on the destination-indexed shortest path, called the *primary path*. Only when the primary path is saturated and can not support more traffic without QoS degradation, temporary alternative paths are established on demand. The traffic following an alternative path will switch back to the primary path whenever the path becomes able to support the traffic. In this routing architecture, no per-stream routing information is needed under normal conditions; in transient overload conditions, alternative paths are used to absorb the overload portion of the traffic. Remarkably, it needs only one extra routing entry at a branching point to store an alternative path. The admission control signaling is an integrated component in the new architecture, and the existing QoS routing protocols find their place in the DiffServ world.

2 Network Service Model

The network consist of a set of autonomous DiffServ domains. There are three types of routers in a DiffServ domain: edge routers, interior routers, and ingress/egress routers. An edge router is at the boundary of a DiffServ domain. It negotiates and enforces a service level agreement (SLA) with a customer. The edge routers implement packet classification, marking, monitoring, traffic shaping, policing, and other functions. Between two domains, the router from which the traffic leaves a domain is called an egress router, and the router from which the traffic enters a domain is called an ingress router. The ingress router (or together with the egress router) is responsible for ensuring that the incoming traffic conforms to the SLA between the two domains.

In addition to the best-effort traffic, various DiffServ PHB groups were defined, such as Expedited Forwarding (EF) [5] and Assured Forwarding (AS) [9]. The traffic in these groups are called the *EF traffic* and the *AS traffic*, respectively. The EF PHB has also been described as Premium Service in [6]. It is a service with low loss, low latency, low jitter, and guaranteed bandwidth. The idea is to always keep the total EF traffic passing through any link in the network under a limit, which is set to be smaller than the link bandwidth. A simple priority queue is then used to schedule EF packets before packets from the other service classes. Since the receiving rate of EF traffic is always smaller than the sending rate at every router, the EF traffic is guaranteed for minimized delay and assured bandwidth. The difficult problem is how to make sure the EF traffic never exceeds a limit (particularly, the link bandwidth). Admission control and routing are essential to solving this problem.

The AS PHB group defines the dropping precedence among different classes of AS traffic when network congestion occurs. Queue management rather than admission control and routing is essential to the implementation of the AS service.

This paper focuses on EF because it provides guaranteed QoS if admission control and routing are properly done. The host that sends out an EF traffic stream is called the *source host*. The edge router that the source host connects to is called the *source router*. The host that receives the EF traffic stream is called the *destination host*. The edge router that the destination host connects to is called the *destination router*. An EF traffic stream is one-way traffic. Two-way traffic is modelled by two EF streams in opposite directions. Two-way traffic is admitted into the network only after its two EF streams are both accepted by the admission control. Without the lose of generality, we will focus on a single EF traffic stream in the rest of the paper.

3 A New Routing Architecture

We first give an overview of the new routing architecture, and then present the details of various components, including (1) how an EF packet is routed by using two routing tables (TRT and QRT), (2) how the admission control signaling operates, and (3) how the QoS routing protocol is triggered to find an alternative path. Finally we study a QoS routing protocol and use it as an example to show where it fits in the picture.

3.1 Overview

One common assumption is to route the EF traffic by the traditional routing table (TRT) in the same way the best-effort traffic is routed. Typically, the TRT provides a single routing path between each pair of nodes. If the path is overloaded by the EF traffic, the TRT approach lacks the flexibility of using alternative paths.

We proposes a new routing architecture, which primarily uses TRT to route the EF traffic but relies on alternative paths to handle the overload condition. The alternative paths are stored in the *QoS routing tables* (QRTs), which are constructed on demand by the QoS routing protocols. The QoS routing protocols are invoked only when the EF traffic overloads the primary routing path. Under normal conditions, the system does not see the existence of the QoS routing protocols.

When an EF traffic stream arrives, the source host issues a request to the source router. The source router initiates the admission control signaling between the source router and the destination router. A REQUEST message is sent towards the destination router to check the bandwidth availability along the path. In the proposed routing architecture, all control messages and non-EF data packets are routed along the primary paths by TRT in the same way the current IP networks route packets.

As a router on the primary path receives the REQUEST, it performs a simple acceptance test to check if it has enough bandwidth for the EF traffic. If it does, the REQUEST is forwarded to the next hop on the primary path. If the acceptance tests of all intermediate routers are passed and the REQUEST successfully arrives at the destination router, it means that the primary path can support the new EF traffic. The destination router sends an ACCEPT message to the source router, which in turns notifies the source host to start sending data packets. The data packets will be routed by TRTs and follow the primary path to the destination host.

On the other hand, if the acceptance test fails at an intermediate router, which means the primary path can not support the traffic, then a QoS routing protocol is triggered to find an alternative path. If an alternative path that supports the traffic is found, an ACCEPT message is sent to the source router and the data packets of the EF stream will follow the alternative path. If an alternative path cannot be found, a REJECT message is sent to the source router, which will either reject the EF traffic or retry the admission control at a later time.

The alternative paths are stored in QRTs. In order to keep the size of the QRTs small, traffic using an alternative path merges back to the primary path when there is sufficient bandwidth freed up on the primary path.

3.2 TRT and QRT

We discuss the difference between TRT and QRT and study how data packets are forwarded by these routing tables.

Each router has one TRT maintained by the traditional routing protocols such as RIP, OSPF, IGRP, and/or BGP. In addition, it has a QRT for each network interface. The QRTs are maintained by the QoS routing protocols. The reason to use multiple QRTs instead of one for the entire router is to reduce the size of each QRT. As it will become clear shortly that each incoming packet will be matched againt one QRT, smaller table size results in faster processing.

TRT is indexed by destination IP addresses. Each TRT table entry consists of a destination IP address, a next-hop IP address, and other information. The outgoing interface to which a packet is forwarded can be determined from the next-hop IP address. QRT is indexed by EF traffic identifiers. An EF traffic identifier is composed of a source IP address, a destination IP address, a protocol identifier, a source port number, and a destination port number. Each QRT table entry consists of an EF traffic identifier, a next-hop IP address, and other information. The route map of Cisco IOS exactly matches the above description, which demonstrates the feasibility of our routing architecture from an implementation point of view. The missing link is that IOS's route map requires manual configuration, but in our architecture it needs to be dynamically updated by QoS routing protocols in order to support EF.

For all non-EF packets, only TRT is looked up to find the next hop, which is exactly what the current IP networks do. Figure 1 illustrates how an EF packet is forwarded at a router. After the packet arrives at the incoming interface, the QRT at that interface is looked up. If there is a matched table entry, the packet bypasses TRT and proceeds directly to the outgoing interface which sends the packet to the next hop. If there is not a match in the QRT, the TRT is looked up and the default shortest-path is used.

Our objective is to minimize the size of QRT. We keep as much EF traffic as possible in the primary path. If the network is in normal conditions without any congestion, all EF traffic will travel along the primary path and the size of QRT will be zero. In this case, only TRT is looked up for all packets. On the other hand, when the aggregated EF traffic on a primary path reaches the maximum allowed quota for EF traffic, a QoS routing protocol will be triggered to find alternative paths for new EF streams. New table entries are inserted into QRT for the duration of the traffic streams. It should be stressed that only the overload portion of the EF traffic is routed via QRT along the alternative paths, and this portion of the traffic will switch back to the primary path whenever possible.

3.3 Admission Control

The admission control is done only for the EF traffic, which receives assured bandwidth and fast forwarding under our routing architecture. The signaling process starts from the source router and follows the primary path towards the destination router. The signaling message, REQUEST, carries (1) the traffic identifier (source IP address, destination IP address, protocol identifier, source port, and destination port), (2) the service class identifier which is EF, and (3) the bandwidth requirement B.

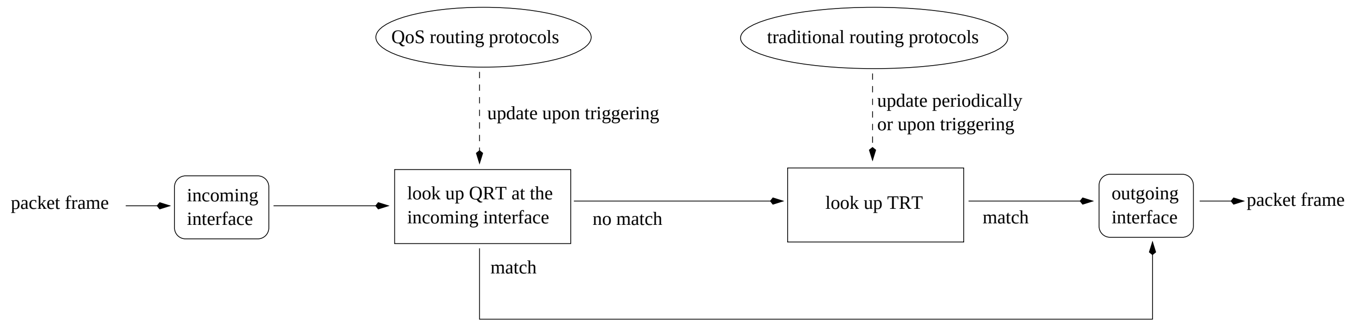

Figure 1. TRT and QRT: When an EF packet arrives, the QRT at the incoming interface is first looked up. If there is a match, the packet is forwarded to the next hop and the TRT is not checked; otherwise, the TRT is looked up.

In order to assist the admission control, each router keeps track of its resource availability. Two variables are maintained for each outgoing interface: EF_{max} and EF_{agg}. EF_{max} is the maximum sustainable bandwidth allowed to be used for sending EF traffic from this interface. EF_{agg} is the aggregated bandwidth currently used by all EF traffic on this interface. EF_{max} is set at the configuration time. EF_{agg} is measured at the run time.

When a router receives a REQUEST, it uses TRT to find the outgoing interface to the next hop on the primary path. A simple acceptance test is performed at the outgoing interface to see if there is enough bandwidth for the new traffic. If $B \leq EF_{max} - B_{agg}$, the REQUEST is sent to the next hop. If $B > EF_{max} - B_{agg}$, the traffic can not be admitted by using this outgoing interface. A QoS routing protocol is triggered to find an alternative routing path. If an alternative path is not found, a REJECT message is sent to the source router that rejects the traffic or retry the admission control after certain delay. On the other hand, if the REQUEST successfully arrives at the destination router or the QoS routing protocol finds an alternative path, an ACCEPT message will be sent to the source router. The source router will notify the source host to send data traffic.

With the help of admission control and QoS routing, the volume of EF traffic at every router is limited so that once admitted the EF traffic receives reliable, guaranteed service of fast forwarding. More specifically, our new architecture makes sure that (1) the primary paths are never overloaded by EF traffic, which prevents performance degradation of the EF traffic; (2) additional EF traffic can be admitted by using alternative paths, which absorbs temporary surges of EF traffic.

Every router in a DiffServ domain is expected to have a bounded delay for the EF PHB. Each REQUEST accumulates the end-to-end delay as it travels to the destination. The ACCEPT then carries this delay value back to the source router, which determines if the end-to-end delay is acceptable. If not, either the traffic is rejected or a delay-constrained QoS routing protocol is invoked to find a delay-constrained alternative path.

3.4 QoS Routing

If the primary path has the bandwidth to support the new EF traffic stream, the QoS routing protocol is not triggered by the admission control, as illustrated in Figure 2 (a). All data packets will follow the primary paths directed by TRT because there is no matched table entry in QRT.

Under congestion conditions, however, an intermediate router may not have the required bandwidth. In Figure 2 (b), suppose i fails the acceptance test at the outgoing interface connecting to j. The corresponding link (i, j) is called an *infeasible link*, which is represented by a dotted line. A link that passes the acceptance test is called a *feasible link*.

The admission control signaling can not proceed along link (i, j), and a QoS routing protocol is activated. The proposed routing architecture is independent of any particular QoS routing protocol. For the purpose of completeness, we present one QoS routing protocol to show how it fits in the architecture. One big advantage of the protocol is that it relies only on the local state stored at each router, which makes it scalable and easy to implement.

The basic idea is as follows: When an infeasible link is encountered, it is considered as an indication of local congestion. The QoS routing protocol tries to find an alternative path by detouring around the infeasible link. Without the knowledge about the extent of the congestion, the protocol branches out towards multiple directions and searches multiple paths for one that can support the new EF traffic.

In Figure 2 (c), i sends out ROUTING messages along all adjacent feasible links. i is called the *branching point*. Apparently, a ROUTING message should not be sent to the link from which the REQUEST was received, and it will not be sent to (i, j), which is an infeasible link. A ROUTING message carries two IP addresses: *bpAddr*, which is the address of the branching point, and *nbAddr*, which is the address of the neighbor that receives the message. It also accumulates the delay of the path it traverses.

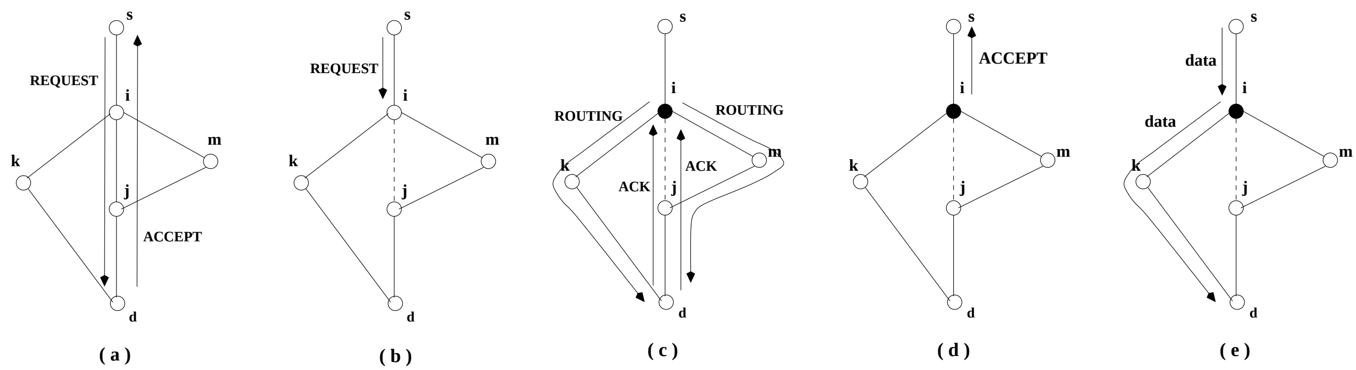

Figure 2. QoS Routing

After a ROUTING messages arrive at the neighbor node k (or m), it is routed by TRTs from there on. Hence, the ROUTING message follows the primary paths from k (or m) to the destination router. This has a very important implication: in order to store an alternative path, it is sufficient to add a single table entry in the QRT at i to direct traffic to k (or m). From k (or m) on, TRTs will be used.

Similar to REQUEST, a ROUTING message causes an acceptance test to be performed at every router it traverses. The ROUTING message is sent to the next hop only if the acceptance test is passed. Therefore, if the ROUTING message successfully reaches the destination router, an alternative path for the new EF traffic is found, which is the path that the message has just traversed. In this case, an ACK is sent back to the branching point i, whose IP address is *bpAddr* that was carried in the ROUTING message. The ACK copies *nbAddr* and the accumulated path delay from the ROUTING message. Upon receipt of the ACK, i inserts a table entry in the QRT at the incoming interface from which the REQUEST was received. The next hop in the entry is set to be *nbAddr*. Also stored in the entry is the delay carried back by ACK, which is the total expected delay for EF traffic on the alternative path. In addition, i sends an ACCEPT message to the source router to admit the traffic (Figure 2 (d)). The source router will notify the source host. As shown in Figure 2 (e), the data packets follow the primary path until it reaches the branching point i, where the QRT at the incoming interface directs the packets away from the primary path. Once the packets reach the neighbor router k, they are again routed by TRTs.

If multiple ROUTING messages arrive at the destination router, then multiple alternative paths are found and multiple ACKs are sent back to the branching point. When the branching point receives an ACK and finds that there is already a table entry in the QRT for the EF traffic stream, it checks if the delay in the ACK is smaller than the delay in the entry. If it is smaller, the next hop in the entry is replaced by the *nbAddr* value in the ACK. Therefore, the best found alternative path will be used.[2]

When a router receives a ROUTING message, if the acceptance test fails, it sends a NACK message back to the branching point. If the branching point receives a NACK from every ACK sent out, it concludes that the QoS routing protocol fails in finding an alternative path. A REJECT message is sent back to the source router, indicating that the traffic can not be admitted at this moment.

The above routing protocol allows only one branching point to deviate from the primary path. More sophisticated design may allow multiple branching points. When a ROUTING message reaches a router that fails the acceptance test, the router can branch again and send ROUTING messages to the neighbors other than the one pointed by TRT. In such a design, every branching point needs a table entry in QRT in order to store the alternative path.

In order to cancel the alternative path and switch back to the primary path whenever possible, the branching point periodically sends CHECK messages down the primary path to see if the primary path can now support the EF traffic stream. It is a mini-version of admission control. If the CHECK message passes the acceptance test at all intermediate routers and successfully reaches the destination router, the source router will be instructed to not send KEEPALIVE. The EF traffic will automatically follow the primary path when the alternative path times out.

4 Simulation

Some preliminary simulation results are presented in this section. The simulation setup is described below. For each data point, we randomly generate ten topologies (each having 600 nodes) based on the Power-Law model [10]. Each link is randomly assigned a *link suc-*

[2] An ACK for an alternative path with smaller EF delay may arrive later because the control messages are not EF traffic and hence may experience a different delay.

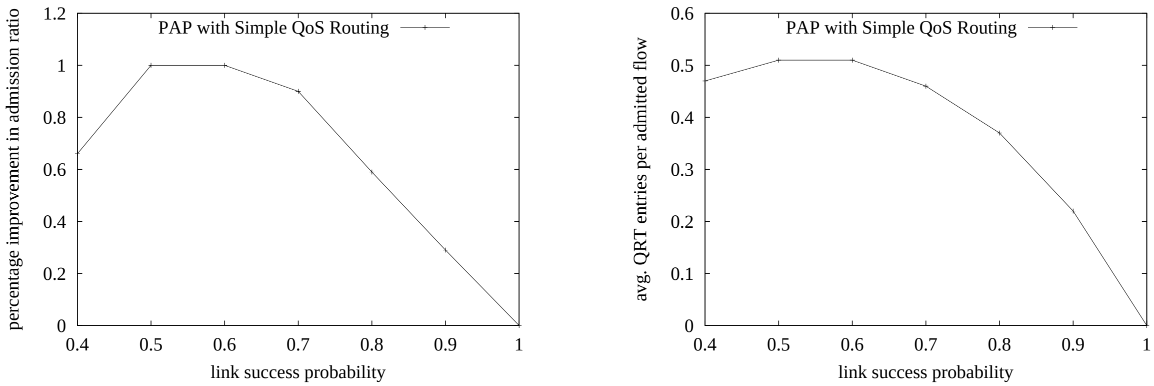

Figure 3. Admission Ratios and QRT entries per admitted flow

cess probability, which is the probability for the link to pass an acceptance test. A large (small) probability corresponds to a light (heavy) load condition. On each topology, 6000 requests are generated with the source and the destination randomly selected from the topology. Two scenarios are compared. One is to perform the admission control on the shortest paths only, and the other is to use the proposed PAP. The QoS routing algorithm used in PAP is the one described in Section 3. The average result of the 60000 requests gives a data point.

Figure 3 shows the simulation results. Comparing with traditional routing, PAP improves the admission ratio (percentage of requests that are admitted) significantly, up to 100%. It achieves such improvement at a small cost — in the worst case, depositing 0.51 QRT entry per admitted flow on average. The reason for the less-than-one average is that, for requests that can be supported by the primary paths, PAP is equivalent to traditional routing (TRT only), and no QRT entry is required.

5 Conclusion

In this paper, we proposed a new routing architecture for DiffServ domains. It integrates the traditional routing, QoS routing, packet forwarding, and admission control. The admission control and QoS routing ensure that the EF traffic admitted to the network is limited under the maximum allowed quota so that the EF traffic always receives assured bandwidth and low delay. Much care has been taken to reduce the size of the QoS routing tables. In particular, the proposed QoS routing protocol requires only one table entry at a branching point to store an alternative routing path.

References

[1] S. Chen and K. Nahrstedt, “An Overview of Quality-of-Service Routing for the Next Generation High-Speed Networks: Problems and Solutions,” *IEEE Network, Special Issue on Transmission and Distribution of Digital Video*, Nov./Dec. 1998.

[2] L. Zhang, S. Deering, D. Estrin, S. Shenker, and D. Zappala, “RSVP: A New Resource ReSerVation Protocol,” *IEEE Network*, September 1993.

[3] D. Black, S. Blake, M. Carlson, E. Davies, Z. Wang, and W. Weiss, “An Architecture for Differentiated Services,” *IETF Network Working Group RFC 2475*, December 1998.

[4] K. Nichols, S. Blake, F. Baker, and D. Black, “Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers,” *IETF Network Working Group RFC 2474*, December 1998.

[5] V. Jacobson, K. Nichols, and K. Paduri, “An Expedited Forwarding PHB,” *IETF Network Working Group RFC 2598*, June 1999.

[6] K. Nichols, V. Jacobson, and L. Zhang, “A Two-bit Differentiated Services Architecture for the Internet,” *IETF Network Working Group RFC 2638*, July 1999.

[7] R. Guerin and A. Orda, “QoS-based Routing in Networks with Inaccurate Information: Theory and Algorithms,” *Infocom’97, Japan*, April 1997.

[8] S. Chen, K. Nahrstedt, and Y. Shavitt, “A QoS-Aware Multicast Routing Protocol,” *IEEE INFOCOM 2000*, May 2000, Also in JSAC special issue on QoS in the Internet, December 2000.

[9] J. Heinanen, F. Baker, W. Weiss, and J. Wroclawski, “Assured Forwarding PHB Group,” *IETF Network Working Group RFC 2598*, June 1999.

[10] Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos, “On Power-Law Relationships of the Internet Topology,” *ACM Proceedings of SIGCOMM ’99*, 1999.