

A Practical Approach for Providing QoS in the Internet Backbone

XiPeng Xiao, Redback Networks Inc.

Thomas Telkamp, Global Crossing Ltd.

Victoria Fineberg, Independent Consultant

Cheng Chen, NEC Corp.

Lionel M. Ni, Hong Kong University of Science and Technology

ABSTRACT

This article describes a practical approach for providing quality of service in the Internet backbone. The approach considers both technical and economic factors. We first present network service provider (NSP) billing models and how NSPs provision their networks. We then analyze causes of QoS-related problems, and describe a practical approach for providing QoS. This approach makes use of good network design, differentiated services, traffic protection, traffic engineering, and traffic management techniques. The relative importance of these techniques is pointed out. Although this approach largely focuses on issues within a single NSP domain, if multiple NSPs adopt such an approach (or a similar approach), interdomain QoS can also be provided.

INTRODUCTION

In this section we describe network service provider (NSP) billing models, network provisioning, and quality of service (QoS)-related problems in NSP networks.

NSP BILLING MODEL

In order to understand real-world QoS issues, one must first understand the economic model of the Internet.

Today, there are three basic billing models used by NSPs:

- Flat rate
- Bandwidth-based
- Data-based

With flat-rate billing, a customer is charged a fixed amount of money every month. The billing NSP may implicitly or explicitly set an upper bound on the amount of bandwidth a customer can use. This billing model is usually applied to residential users and some corporate users who own low-speed links (e.g., T1/E1).

With bandwidth-based billing, a customer is charged monthly at a base rate for a certain amount of bandwidth and at a premium rate for any extra bandwidth usage. For example, a customer collocated in an NSP point of presence (POP) may own a Fast Ethernet (FE) link to the NSP. Although link capacity is 100 Mb/s, the customer may only commit to pay for 10 Mb/s at a base rate of \$200/Mb/s. Even if the customer uses less than 10 Mb/s, it still pays $\$200 \times 10 = \2000 . If the customer uses more than 10 Mb/s, it pays \$300 for each extra 1 Mb/s. Because the base rate is lower than the premium rate, it is in the customer's best interest to commit to as much bandwidth as it needs. Sometimes, an upper bound on bandwidth usage may also be specified and any traffic exceeding that bound will be discarded. This billing model applies to customers collocated in POPs and customers with high-speed access lines (e.g., DS-3).

The instant bandwidth values of a link vary in time. The value used for billing is usually the 95th percentile value. Assuming that totally N bandwidth samples are collected via Simple Network Management Protocol (SNMP) in a month and sorted in ascending order, the $N \times 95$ th percent value is the 95th-percentile value.

The data-based billing approach is similar to the bandwidth-based approach, except that customers are charged per megabyte rather than per megabit per second.

Note that billing is applied to traffic in both directions, sent and received. An NSP may use all three billing approaches, with different customers billed differently. In addition, billing is generally insensitive to traffic destinations.

NSP NETWORK PROVISIONING

Many NSPs claim that they overprovision their networks. But users often hear that NSP networks are oversubscribed. What is the reality? This in fact depends on which billing model is used.

With flat-rate billing, NSPs do not get more revenue from sending/receiving more customer traffic. Therefore, they tend to oversubscribe the uplinks (i.e., links carrying customer traffic toward the backbone).

With bandwidth-based or data-based billing, NSPs get more revenue for sending/receiving more customer traffic. Therefore, they tend to overprovision the uplinks to accommodate customer traffic growth. Coexistence of oversubscription and overprovisioning is illustrated in Fig. 1: 100 customers are aggregated by a switch. Each customer owns an FE link. Theoretically the total traffic from the customers can be as high as $100 \times 100 \text{ Mb/s} = 10 \text{ Gb/s}$. Therefore, the NSP needs at least 10 Gb/s of uplink capacity from the switch to the routers. In Fig. 1, since the NSP provides only 2 Gb/s of uplink capacity, it is 5:1 oversubscribed. However, not every customer is using 100 Mb/s all the time. In fact, a customer may only commit to pay for 10 Mb/s although it owns an FE link. So the actual total traffic may be only 1 Gb/s. Based on that, the NSP uses two GigE uplinks to connect the switch to the routers and maintains each uplink's utilization at about 50 percent. Therefore, the NSP is theoretically 5:1 oversubscribed but practically 2:1 overprovisioned. The NSP will continuously monitor utilization of the uplinks. For example, if customer traffic increased to 1.5 Gb/s and caused utilization of the two GigE uplinks to increase to 75 percent, the NSP would add a third GigE link to maintain the link utilization at about 50 percent, an "ideal" utilization for many NSPs.

QOS-RELATED PROBLEMS

Given the background information, we are now ready to analyze causes of QoS problems. They can generally be divided into two categories: non-network-related and network-related causes.

Non-network-related causes include:

Overloaded servers (e.g., Web or email) users are trying to access: In this case, common ways to improve QoS are to upgrade the servers, or to add servers and use a better load-balancing scheme among them.

Network operation errors: Configuring routers/switches is a complex and error-prone process. For example, duplicate IP addresses can be mistakenly configured and cause routing problems.

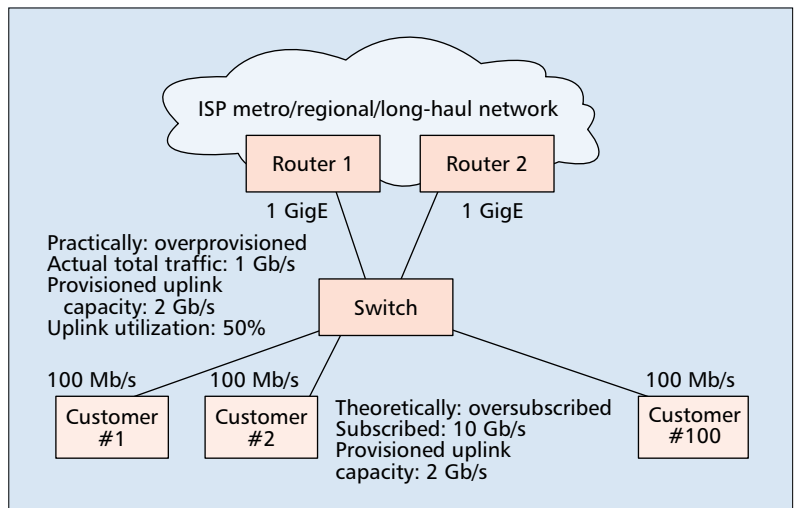
Dealing with non-network-related problems is outside the scope of this article. It is worth pointing out that providing operations support systems (OSSs) for IP networks is essential to reduce operation errors.

Network-related causes include:

Equipment problems: Routers/switches are complex systems with sophisticated software and hardware that are required to process millions of packets per second. Equipment vendors are compelled to deliver products as early as possible. Therefore, it is not uncommon that routers/switches have hardware and software problems.

Lack of access capacity: For economic reasons, there are always customers with slow access links (e.g., dialup modems) or oversubscribed uplinks. The technical solution for this kind of problem is clear:

- Adding capacity



■ Figure 1. Network provisioning for collocated customers.

- Classifying traffic and marking it differently for subsequent treatment using policing, shaping, and so on. This will be further discussed in our approach.

However, it should be pointed out that providing QoS may not make economic sense here if users are not willing to pay for it.

Uneven traffic distribution that causes some links to be congested: This is the most common cause of network-related QoS problems in the backbone. Even though the average link utilization of a network can be low, say 30 percent during the peak hour, a small number of links can still have very high link utilization (close to 100 percent). Such congested links will cause long packet delay and jitter or packet loss. The causes of such hot spots in the network can be:

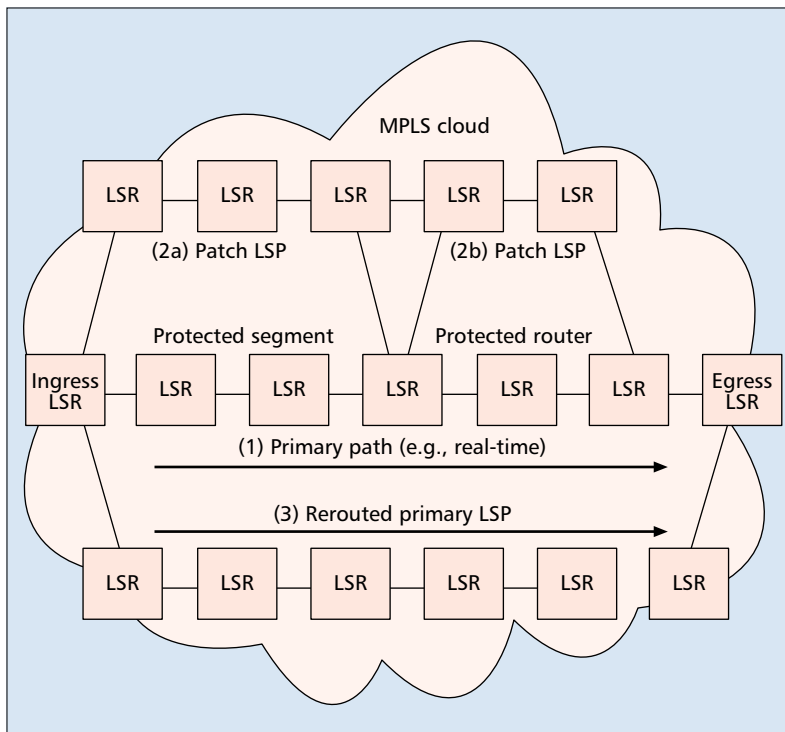
- Unexpected events such as fiber cut or equipment failure.
- Traffic pattern shift while the network topology and capacity cannot be changed as rapidly. In the backbone, new capacity may not always be available at the right place and the right time. For example, the sudden success of a Web site or an unplanned broadcast of multimedia traffic can cause some links to become congested.

A PRACTICAL APPROACH FOR PROVIDING QOS

In this section we describe an approach to solve the remaining problems described previously. This approach is discussed step by step. The steps are listed in order of decreasing importance. NSPs should start from the first step and add additional steps if needed.

STEP 1: CLEAN UP THE NETWORK

Networks are generally well designed and provisioned in the beginning. But over time, problems caused by quick-and-dirty fixes will accumulate. Therefore, regular cleanups should be performed. Single points of failure and bottlenecks should be removed. Capacity should be added at appropriate places so that even the most critical router/link failure will not cause traffic conges-



■ Figure 2. Fast reroute.

tion. Interior Gateway Protocol (IGP) metrics, Border Gateway Protocol (BGP), and peering policies should be evaluated and readjusted. Logs should be examined and security measures checked. Audits should be performed to correct configuration mistakes. NSPs should also educate their customers to tighten up security and upgrade congested last mile circuits. These are by far the most important and useful things to do for providing QoS in the Internet.

In our QoS approach, this step serves to prevent QoS-related problems from happening.

STEP 2: DIVIDE TRAFFIC INTO MULTIPLE CLASSES

In our approach three classes of services are proposed:

- Premium
- Assured
- Best effort

Premium service provides reliable, low-delay, low-jitter service. Real-time traffic (e.g., videoconferencing) and mission-critical traffic (e.g., financial or network control traffic) can benefit from such a service. Assured service provides reliable and predictable service. Non-real-time virtual private network (VPN) traffic can benefit from it. Best effort service is traditional Internet service.

Some people may think that more than three traffic classes are needed. Two questions can be used to help decide how many classes are needed. First, what are the target applications for each traffic class? If there is no target application, that class should be eliminated. Second, how should we distinguish different traffic classes to end users? If traffic class A is more expensive than B, but an NSP cannot clearly show that class A has better service, A should be eliminat-

ed because (eventually) nobody is going to pay for it. The second question is particularly important because the answer to the first question can be somewhat arbitrary.

In our approach, the first question has been answered in this section. The second question will be answered in subsequent sections.

STEP 3: PROTECTION FOR PREMIUM TRAFFIC AND TRAFFIC ENGINEERING

Multiprotocol label switching (MPLS) [1] is used for traffic protection and traffic engineering in our approach. MPLS is an advanced forwarding scheme. It extends IP routing with respect to signaling and path controlling. It has been deployed in many tier 1 NSPs including AT&T and Global Crossing [2, 3].

Traffic Protection — First, MPLS label switched paths (LSPs) are configured in the network. Each ingress router will have two LSPs to the egress. One LSP is for premium traffic, and the other is shared by assured and best effort traffic. The premium LSP will have fast reroute [2] enabled. The basic idea of fast reroute is to have a patch LSP preconfigured for a link, a router, or a segment of the path consisting of multiple links and routers. This link or router or path segment is called a *protected segment*. When there is a failure in a protected segment, the router immediately upstream of the protected segment (called a *protecting router*) will detect the failure from layer 2 notification. The patch LSP will then be used to get around the failure. This protection can take effect within 50–100 ms. During fast reroute, the path taken by the LSP can be suboptimal. To correct that, the protecting router will send a message to the ingress router of the LSP, which will then compute a new path for the LSP and switch traffic to the new LSP. This process is illustrated in Fig. 2.

Fast reroute is essential for applications that cannot tolerate packet loss. However, fast reroute adds considerable complexity to the network. In the future, if IGP can converge faster (e.g., subsecond) [4], the need for fast reroute will be reduced.

In our approach, traffic protection serves to provide high availability for premium traffic.

Traffic Engineering — Because network topology and capacity cannot and should not be changed rapidly, uneven traffic distribution can cause congestion in some part of a network, even when total capacity of the network is greater than total demand.

In our approach, each ingress router will have two LSPs to an egress. One LSP is for premium traffic to that egress, and the other is shared by assured and best effort traffic. Traffic from customers (including other NSPs) is classified at the ingress router based on the incoming interface and put into the appropriate LSPs. Network operators can also provide multifield (source and destination IP addresses, port numbers, and protocol ID, etc.) classification as a value-added service. In addition, the experimental (EXP) fields of packets are set accordingly.

In order to avoid concentration of premium

traffic at any link, an upper limit is set for each link regarding how much bandwidth can be reserved by premium traffic. When that portion of bandwidth is not used, it can be used by other traffic classes if that is desirable. The percentage of the premium traffic should be determined by NSP policy and premium service demand. DiffServ-aware traffic engineering is done for these two sets of LSPs to avoid congestion at each link [5].

In our approach, traffic engineering serves two purposes:

- To prevent (as much as possible) congestion caused by uneven traffic distribution from happening.
- If congestion does happen, to relieve it quickly.

By doing traffic engineering in a DiffServ-aware manner, a third purpose is also served:

- To make the percentage of premium traffic reasonably low in each link so that:
 - Delay and jitter of premium traffic are low.
 - If necessary, premium traffic can preempt resources of low-priority traffic (which is not possible if all traffic is of high priority).

Compared to the traffic management schemes (described later) such as policing, shaping, and buffer management, traffic engineering can control traffic and network performance on a much larger scale. It can be considered as macro control mechanism.

STEP 4: CLASS-BASED QUEUING AND SCHEDULING

Based on the EXP field of the MPLS header, packets of the different classes are put into different queues. How to configure output rate and size for these queues is a difficult task. One possible approach is described below.

The arrival rate of each queue at an interface can be obtained by summing up rates of all transiting LSPs in that queue. The rate of these LSPs can be obtained via SNMP. Depending on the relative importance (e.g., monetary value) of each class, different weight can be introduced for them. For example, the weight for premium, assured, and best effort traffic can be set to 6, 3, and 1, respectively. Output rate of each queue can then be computed as follows:

$$o(q) = bw * \{ [w(q)*i(q)] / [6*i(PQ) + 3*i(AQ) + 1*i(BQ)] \},$$

where

- q = queue of interest. For the premium queue $q = PQ$; for the assured queue $q = AQ$; for the best effort queue $q = BQ$
- bw = bandwidth of the interface (e.g., 2.5 Gb/s for an OC-48c interface)
- $o(q)$ = output rate of queue q
- $w(q)$ = weight of queue q (e.g., $w(PQ) = 6$, $w(AQ) = 3$, $w(BQ) = 1$)
- $i(q)$ = input rate of queue q , where q can be PQ or AQ or BQ

For example, for the premium queue, the formula becomes

$$o(PQ) = bw \times \{ [6*i(PQ)] / [6 \times i(PQ) + 3 \times i(AQ) + 1 \times i(BQ)] \},$$

Again, note that all bandwidth values are 95th percentile values. Also, MPLS simplifies

the above scheme, because per-(LSP, EXP) statistics can be obtained.

Because traffic grows, $i(PQ)$, $i(AQ)$, and $i(BQ)$ change over time. Therefore, the output rates of these queues should be adjusted periodically (e.g., weekly) [6]. But note that adjustment of queue rates only affects performance of traffic going through a specific interface. Compared to traffic engineering, queue rate adjustment (and other traffic management schemes) can be considered as micro control.

From the traffic contract in the SLA, the output rate of a queue, and other requirements such as maximum allowable queuing delay, the size of the queue can then be determined as follows:

$$\text{queue size} = \text{output rate of the queue} \times \text{maximum allowable queuing delay.}$$

Using the approach described above, the overprovisioning factor for PQ, that is, the ratio of (output rate/input rate), is usually far greater than 1.0 (it also depends on the relative amount of premium, assured, and best effort traffics). This is enough to guarantee that PQ is empty or very short most of the time. Therefore, the delay and jitter of the premium traffic will be sufficiently low. Simulations have been done to study the effect of the overprovisioning factor on delay and jitter [7, Appendix]. They confirmed the validity of our approach. Network performance monitoring also confirms it.

Another alternative is to use a priority queue for premium traffic. That is, premium traffic will always be sent before other traffic. In fact, this can be more effective in distinguishing premium service from other services. However, care must be taken to ensure that premium traffic will not starve other traffic.

In our approach, queuing and scheduling are the actual mechanisms to ensure that high-priority traffic is treated preferably. It is important to prevent congestion of low-priority traffic, if any, from affecting performance of high-priority traffic. This is useful when network capacity becomes insufficient in meeting demand because of fiber cut or other equipment failure.

STEP 5: DEPLOY OTHER TRAFFIC MANAGEMENT SCHEMES

In this section we discuss applicability of policing, shaping, and Random Early Detection (RED) [8].

Policing and Shaping — When a customer signs up for network service, it will have a service level agreement (SLA) with its NSP. The SLA specifies the amount of traffic (in each class if applicable) the customer can send/receive. Many people think this means that NSPs will always do policing or shaping. But whether this is true actually depends on how a network is provisioned, which is in turn determined by the billing model.

In the access network where the flat-rate billing model is applied, the network is generally oversubscribed. Therefore, policing and shaping are useful to ensure that no customer can consume more bandwidth than it signed up for. The policing/shaping parameters are usually fairly

In our approach, queuing and scheduling are the actual mechanisms to ensure that high-priority traffic is treated preferably. It is important to prevent congestion of low-priority traffic, if any, from affecting performance of high-priority traffic.

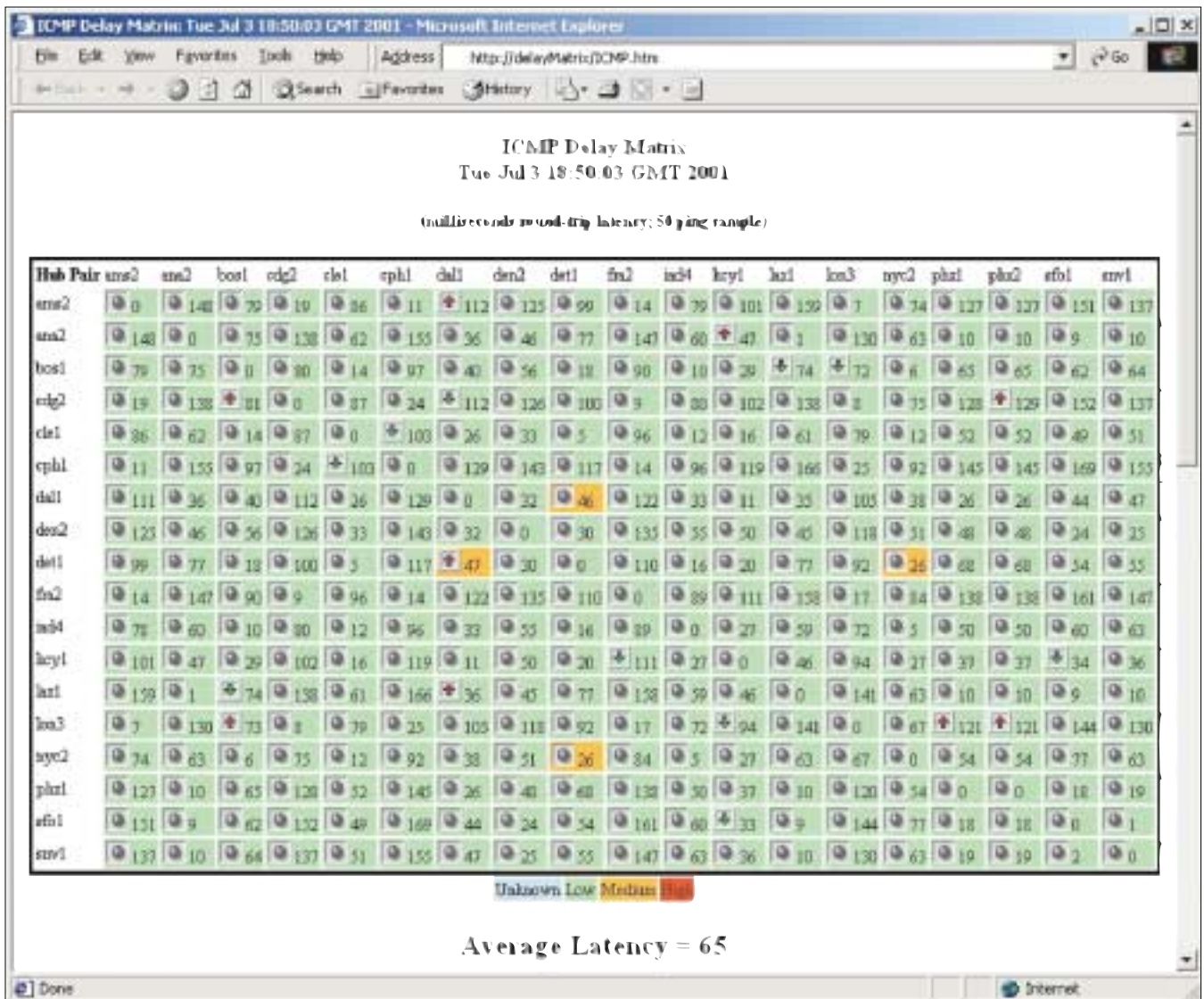


Figure 3. Round-trip delay matrix of Global Crossing's worldwide backbone.

static in such case. However, policing and shaping may affect performance of the access device. In that case, an alternative is to aggregate traffic from many customers and police/shape it collectively. The individual customer's traffic is only policed/shaped when that customer is causing trouble for others.

With bandwidth-based or data-based billing, because the NSP can increase revenue when the customer sends/receives more traffic than the SLA specifies, there is no need for the NSP to police or shape a customer's traffic (but accounting will always be done), unless the customer is causing problems for others. From Fig. 1 we can see that unused uplink capacity is usually significantly greater than the amount a single customer can send or receive (1 Gb/s vs. 100 Mb/s in Fig. 1); therefore, it is quite unlikely that a customer can cause problems for others. In the relatively rare case when a customer is causing problems for others or is being attacked (e.g., more than 100 Mb/s of traffic was sent to a customer connected by an FE link), traffic from/to that particular customer can be policed/shaped. Excess traffic can be marked

with high drop probability or dropped immediately. With these two billing models, the amount of traffic specified in an SLA is mainly for network planning purpose.

In some cases, a customer may request an upper bound on the amount of money it pays for bandwidth usage. In these cases, the NSP may need to do policing/shaping.

RED/WRED — RED is a buffer management scheme designed to prevent tail drop caused by traffic burst. Backbone routers can generally buffer traffic for up to 100 ms per port. If traffic bursts over output line rate for longer than that, the buffer will become full and subsequent packets will be dropped. Tail drop causes many TCP flows to decrease, and later increase, their rate simultaneously, and can cause oscillating network utilization. By preventing tail drop, RED is widely believed to be useful for enhancing network performance.

Weighted RED (WRED) is a more advanced RED scheme. It depends on other mechanism(s) such as classification/marketing/policing to mark packets with different drop priorities. WRED will

then drop them with different probabilities (which also depends on the average queue length).

RED/WRED is useful to prevent transient bursts from causing tail drop. However, it should be noted that it is quite difficult to set the (W)RED parameters (e.g., different drop probabilities at different queue lengths) in a scientific way. Guidelines are yet to be developed on how to set such parameters. If a backbone's link utilization (time-averaged over a period of 1–5 min) can be maintained at 50 percent or lower, there should be sufficient capacity to accommodate transient bursts to avoid tail drop. The need for (W)RED can be reduced.

In general, traffic engineering is more effective in controlling traffic distribution in a network and has a bigger impact on network performance than traffic management schemes such as policing, shaping, and WRED. It should be invoked before traffic management schemes.

In our approach, policing, shaping and WRED are the enforcing mechanisms. They are used only when all other techniques such as traffic engineering fail to prevent congestion from happening. In that case, these traffic management schemes make sure that high-priority traffic is treated preferably compared to low-priority traffic.

EFFECTIVENESS OF OUR APPROACH

We examine the effectiveness of our approach with respect to:

- Distinguishing different classes of traffic
- Meeting the delay and jitter requirements of applications

Distinguishing Different Classes of Traffic

— When a link or router fails, it takes IGP, MPLS, and BGP from seconds to minutes to reconverge. During this period, packets will experience long delay or be dropped. MPLS fast reroute can protect premium traffic during the reconverge period. Therefore, network availability is better for premium traffic than for assured traffic. Besides, a larger ratio of (output rate/input rate) for the premium queue enables premium traffic to have lower delay and jitter.

One-way delay	Characterization of quality
0–150 ms	“Acceptable for most user applications”
150–400 ms	“May impact some applications”
Above 400 ms	“Unacceptable for general network planning purposes”

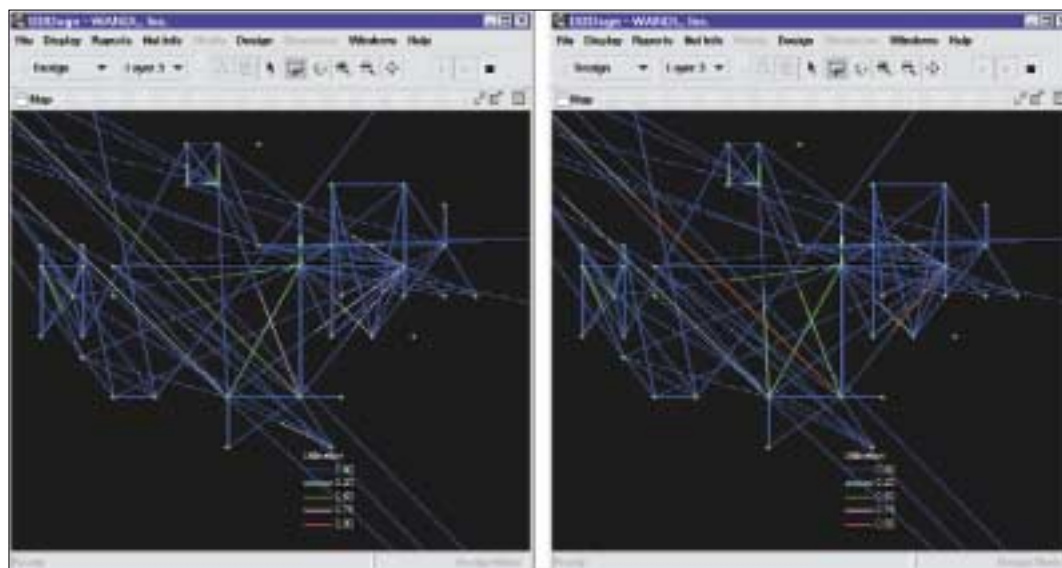
■ **Table 1.** ITU G.114 delay recommendations.

Because assured traffic can use three (or any other configured value) times more resources than best effort traffic, its performance will be better than best effort traffic, especially when there is failure and link utilization becomes high. Practically, an NSP that plans to offer QoS can begin with just the premium and best effort classes. The assured class can be added later when the need arises.

Meeting the Delay and Jitter Requirements of Applications

— The described approach has been almost fully implemented in Global Crossing's global IP backbone. MPLS traffic engineering has been deployed since the second quarter of 1999. It turned out to be very effective in meeting the delay and jitter requirements of applications. Generally, coast-to-coast round-trip delay in the United States is below 80 ms and jitter is below 2 ms. A snapshot of the network delay matrix is showed in Fig. 3. (Hubs *ams2*, *cdg2*, *cph1*, *fra2*, and *lon3* are located in Europe and the rest are in the United States.) This is a very good network performance that far exceeds ITU G.114 Recommendations on delay for applications (Table 1). Note that Fig. 3 lists round-trip delay, while Table 1 lists one-way delay.

In case of a node or link failure, traffic engineering will automatically reroute traffic and avoid any congestion. This may increase delay slightly for some traffic because a longer path is taken, but will prevent packet loss and maintain low jitter after the network reconverges. Fig. 4 compares network condition with and without traffic engineering:



■ **Figure 4.** Network condition with TE (left) and without TE.

Good network design plus a certain degree of over-provisioning not only makes a network more robust against failure, but also prevents many QoS-related problems from happening and eliminates the need for complex mechanisms designed to solve those problems.

- The simulator used here is the NPAT package from WANDL.
- The picture shows part of Global Crossing's network in the California Bay Area.
- Each green dot represents a router.
- Each line between two dots represents a link between two routers. Each line has two colors, showing link utilization in each direction. Red color (utilization 90 percent or higher) implies congestion. Note that there is no congestion in the left picture, and there are two instances of congestion in the right picture.

CONCLUSIONS

Today the Internet is not perceived as reliable enough for critical missions. But this is not because of lack of advanced mechanisms such as per-flow shaping/policing/shaping. Instead, the challenge lies in how to maintain a clean network and make the right trade-off between simplicity and more control.

Good network design, simplicity, high availability, and protection are the keys for providing QoS in the Internet backbone. Good network design plus a certain degree of over-provisioning not only makes a network more robust against failure, but also prevents many QoS-related problems from happening and eliminates the need for complex mechanisms designed to solve those problems. This keeps the network simple and increases its availability. Three traffic classes (premium, assured, and best effort) are sufficient to meet foreseeable customer need. Different traffic classes will be handled differently, especially under adverse network conditions. MPLS fast reroute or other protection schemes can be used to protect premium traffic during router or link failure. When failure happens in one part of the network, traffic engineering should be used to move traffic to other part of the network. DiffServ-aware traffic engineering can be used to prevent concentration of high priority traffic at any link so that high priority traffic will have low delay and jitter, and can be treated preferably at the expense of other classes of traffic if necessary. In the backbone, traffic management schemes such as policing and shaping should be treated as micro control and be used when traffic engineering is insufficient. Traffic management schemes are more appropriate for the access network before the last mile circuits and the congested uplinks are upgraded.

REFERENCES

- [1] E. Rosen *et al.*, "Multiprotocol Label Switching Architecture," RFC 3031, Jan. 2001.

- [2] X. Xiao and L. Ni, "Internet QoS: A Big Picture," *IEEE Network*, Mar./Apr. 1999.
- [3] X. Xiao *et al.*, "Traffic Engineering with MPLS in the Internet," *IEEE Network*, Mar. 2000.
- [4] C. Alaettinoglu *et al.*, "Towards Millisecond IGP Convergence," Internet draft, Nov. 2000.
- [5] D. Awduche *et al.*, "Overview and Principles of Internet Traffic Engineering," RFC 3272, May 2002.
- [6] X. Xiao, "Providing QoS in the Internet," Ph.D. thesis, Dept. of Computer Science, Michigan State Univ., May 2000, <http://www.cse.msu.edu/~xiaoxipe>.
- [7] V. Jacobson *et al.*, "An Expedited Forwarding PHB," RFC 2598, June 1999.
- [8] S. Floyd and V. Jacobson, "Random Early Detection gateways for Congestion Avoidance," *IEEE/ACM Trans. Net.*, vol. 1, no. 4, Aug. 1993.

ADDITIONAL READING

- [1] S. Blake *et al.*, "An Architecture for Differentiated Services," RFC 2475, Dec. 1998.

BIOGRAPHIES

XIPENG XIAO (xipeng@redback.com) is director of product management at Redback Networks, Inc. He works with network service providers and defines product requirements for Redback. Prior to Redback, he was Sr. Manager of Advanced Technology at Global Crossing where he engaged in network design, operations and equipment evaluation. He received his Ph.D. degree in Computer Science from Michigan State University, and his M.S. and B.S. degrees from Zhejiang University, China. (<http://www.cse.msu.edu/~xiaoxipe>).

THOMAS TELKAMP (telkamp@gblx.net) is director of network architecture at Global Crossing Ltd., responsible for the planning and architecture of Global Crossing's MPLS backbone, Internet services, and VPNs. Before joining Global Crossing, he was at AT&T-Unisource Communications Services (now Infonet Europe) and SURFnet, and consulted for several other companies.

VICTORIA FINEBERG [SM] (fineberg@illinoisalumni.org) is a professional engineer. After graduating with a Master's degree from the University of Illinois at Urbana-Champaign in 1989, she joined Bell Laboratories at AT&T and then Lucent. Her professional interests include QoS, MPLS, and VoIP. Presently she is an independent consultant.

CHENG CHEN (CChen@necam.com) received his Ph.D. from Florida State University in 1981. He was a faculty member at the University of South Carolina in 1980 and at Temple University from 1981 through 1982. He has worked for AT&T Bell Laboratories from 1982 through 1989, NEC America's Advanced Switching Laboratories from 1989 through 1994, MCI from 1994 through 1997, and again NEC America from 1997 to the present. His research areas include IP/MPLS, QoS routing, traffic engineering, ATM, and network reliability engineering.

LIONEL M. NI [F] (ni@cs.ust.hk) received his Ph.D. degree in E.E.C.S. from Purdue University in 1980. He was a professor in the Computer Science and Engineering Department at Michigan State University from 1981 to 2002. In 1994 he became a Distinguished Professor. In July 2002 he joined the Department of Computer Science at Hong Kong University of Science and Technology. He has published over 200 articles in refereed journals and co-authored the book *Interconnection Networks: An Engineering Approach*. He has chaired many professional conferences and has received many awards. His paper (with his former student Chris Glass) "The Turn Model for Adaptive Routing" was selected as one of the 41 most significant papers in the last 25 years in computer architecture in 1998.