Future Priorities for Internet2 QoS

Ben Teitelbaum*

October 2, 2001

Abstract

Internet2's QoS program is nearly four years old and the value story is far from clear. This memo takes stock of our successes, failures, and frustrations to date and presents a four-part QoS plan for the coming year that accelerates a strategic shift already underway. The shift is towards lightweight QoS approaches that deploy incrementally and cheaply and that offer immediate value from and experience with QoS mechanisms present in IP routers today. A second facet of the proposed QoS program is a renewed emphasis on applications and their QoS needs. This memo originated as an internal UCAID staff memo and has been adapted for publication as an informational document of the Internet2 QoS Working Group.

Contents

1 Executive Summary

"Quality of service" (QoS) refers to a variety of technologies that allow the network to offer differing service qualities to different users, applications, or traffic classes. The Internet2 QoS program began nearly four years ago to assure that resource-hungry

*Ben Teitelbaum is with Internet2 and Advanced Network & Services. E-mail: ben@internet2.edu.

advanced applications would get the network performance they need. Even at the project's inception, it was clear that advances in optical networking would push link speeds exponentially higher and per-bit bandwidth costs lower at an equally dramatic pace. What was not clear was whether internet bandwidth would become scarce. Simultaneous with the much-heralded bandwidth explosion, Internet2 undertook an extensive set of efforts to spark explosive growth of new applications. In this light, Internet2's QoS program may be seen as an effort to put in place a "safety belt" that could avert the success catastrophe of an inversion of bandwidth supply and demand, combined with increasing dependence on performance sensitive applications.

Today, nearly four years later, use of new resource intensive applications and bandwidth demand have not grown as quickly as expected and there has been no need for a QoS safety belt. The reasons for this are complex and not fully understood. One set of factors that is fairly well understood, however, has come to be known as the "end-to-end performance problem".

Paradoxically, most new applications can be made to run well on today's over-provisioned Internet2 networks, while at the same time, most Internet2 hosts rarely achieve the network speeds they should. Applications ("advanced" or otherwise) are often hamstrung by mundane faults in the end-to-end path. Most commonly, these faults relate to poor application design, improperly configured end-system network stacks, mis-configured LANs, and faulty cabling. Rarely is network congestion the limiting factor. Internet2 has recognized this as a serious problem and recently launched a $1.5 million/year initiative to address it.

Going forward, the safety belt rationale for QoS is still sound. As the Internet2 End-to-End Performance Initiative progresses, an ever-increasing number of hosts will be able to achieve very high network speeds and a high-bandwidth Internet2 "killer app" may yet emerge[1].

Although the safety belt rationale is theoretically still sound, it is a hard sell. Internet2's QoS program has made a considerable effort and investment in designing, testing, and promoting an IP circuit emulation service known as the "QBone Premium Service" (QPS). The deployment of QPS has been frustrated not only by lack of demand, but also by lack of support from router vendors and a design that requires all-or-nothing network upgrades and dramatic operational and economic paradigm shifts. Although the world will change and QPS may yet become viable, the membership deserves to reap value from and begin to gain experience with QoS before then.

Learning from the Premium deployment experience, we recommend that Internet2 efforts to promote IP-layer service differentiation focus on "non-elevated" ser-

[1]Note that in relation to many commercial backbone networks Abilene is woefully underprovisioned from the perspective of host access speeds; Abilene's interior circuits could be saturated by only 25 hosts each connected to the backbone on a path no slower than 100 Mbps; moreover there are thousands, if not tens of thousands of such hosts in Internet2 today.

vices. "Non-elevated" services include those that are "lower" than best-efforts, like the QBone Scavenger Service (QBSS), as well as those that split best-efforts into multiple "different but equal" service classes. Non-elevated services require no policing and little operational change. Moreover, they may be deployed incrementally and cheaply since they require deployment only at interfaces where congestion is a potential problem. Although these approaches offer far weaker service assurances than QPS, they deploy much more easily and have the potential to enable delay sensitive applications across congested links.

A second area of new emphasis is applications. There is an acute lack of understanding of the network performance that applications really require. Because of its very simple virtual wire service model, Premium service largely begged the issue of precisely describing application utility as a function of network performance. After all, if the application couldn't run on its own wire, then clearly there was something wrong with the application. Now that virtual wire services have been found to be undeployable, a closer scrutiny of application needs is in order. We have begun and will continue a program to study and document application QoS needs. We hope to raise awareness not only of what applications need, but also of design tradeoffs that exist between adding complexity to the network versus adding it to the application.

Work in the coming year will focus on four areas: non-elevated services, elevated services, applications, and a catch-all area labeled "research". Detailed descriptions of the goals and proposed work programs for each of these four areas may be found below in section 2. By way of background, the reader may find a short history of Internet2 QoS in the appendix.

2 Proposed Program

Overall, the QoS program proposed here is less overtly ambitious, more incrementalist, and more pragmatic. In some respects this program is a retreat. It recommends suspending QPS and a "back to the drawing board" reassessment of application QoS needs. In other respects this program is boldly ambitious and practical, seeking to promote forms of QoS that can creep incrementally into networks and applications.

It is important to explain why the QBone Premium Service has not succeeded and to emphasize that work on QPS has not been for naught. Ultimately, QPS was a grand experiment, having great potential, but so far yielding only negative results. As with nuclear fusion research on tokamaks, there is an abundance of valuable data even in results that fail the ultimate metric of success—ignition. Consequently, although we are abandoning the deployment and promotion of QPS for the time being, we will carefully document our experiences, attempting to shed light on why QPS is so hard. The QBone Architecture Design Team is working to capture the best current thinking about how an IP circuit-emulation service like QPS *could* be made to work and is beginning to assess whether there are *any* practical approaches to deploying elevated interdomain services.

In the coming year, work by UCAID staff and the QoS working group should focus on four areas: non-elevated services, elevated services, applications, and a catch-all area labeled "research". The following subsections detail goals and specific activities for each of the proposed work areas.

It is important to note that the QoS working group may undertake work that is outside the scope of this work program. The working group has been restructured in the image of an IETF working group, in which participation is open and sets of participants are free to author and propose drafts to the working group as a whole. As chair, I seek to exercise leadership and advance an agenda that I see as best for the community. The working group is, however, ultimately democratic and free to go off in other directions if it so chooses.

Non-Elevated Services

Goals

- Demonstrate, evangelize, catalyze, and monitor the deployment of the QBone Scavenger Service (QBSS)
- Demonstrate the value of QBSS at Supercomputing 2001 (SC2001) with a near-gigabit TCP flow scavenging bandwidth throughput the course of the conference
- Design a trial deployment of a service like Alternative Best-Effort (ABE) or Best-Effort Differentiated Service (BEDS) that creates an unpoliced, low-delay, best-effort class
- If successful, evangelize an ABE-like approach to router vendor partners

This will be the most visible work area in terms of active deployment and field testing. QBSS was launched at the Spring 2001 Internet2 Member Meeting and has steadily progressed. At least four schools are marking dormitory subnets for QBSS, high-energy physics groups at SLAC, IU, CERN, and in the UK have been developing QBSS trials, and patches have been developed for Apache, WU-FTP, and SLAC's BBFTP to mark bulk transfers for QBSS. QBSS has been tested across Advanced's OC3 circuit to Abilene and experiments are underway to evaluate QBSS across TRANSPAC, as well as between SLAC and IN2P3 (Lyon, France), CERN, and/or Daresbury Labs (UK). And, we are in discussions with John Crowcroft and others in the UK about modifying JANET's metered pricing regime to exempt QBSS traffic.

We will continue to evangelize QBSS, promoting traffic marking by applications and leaf domains and the configuration of scavenger queues for QBSS at bottleneck interfaces. We also hope to show the value of QBSS at SC2001 as part of a double-whammy demonstration in which we scavenge unused bandwidth with a QBSS-marked TCP flow across the WAN, achieving near-gigabit throughputs in the absence of competing best-effort traffic, but yielding immediately in its presence. We aim for a press release that would read something like: "X Terabits of Data Moved to SC2001 Show Floor, Not a Single Best-Effort Packet Delayed". To achieve this will require not only further development and testing of the FreeBSD GE test machines we have built at Internet2, but also careful engineering across the end-to-end path to honor QBSS markings at potential bottlenecks.

Abilene has a minimal role to play in advancing QBSS. It is, however, ideally situated to monitor its use and growth. We are currently using NetFlow monitoring to track QBSS and hope to operationalize MRTG-style historic traffic plots.

The final piece of work in this area relates to a new family of QoS approaches that divide today's monolithic best-effort service class into multiple "different but equal" best-effort service classes that trade off loss for delay. Two such approaches are the Alternative Best-Effort service (ABE) and the Best-Effort Differentiated Service (BEDS). Both require support from routers that does not exist today, but have the very attractive property that they can offer bounded delay to delay-intolerant applications without the need for policing, settlement, reservation signaling, admissions control, or other complexity typically associated with QoS approaches that bound delay. We are currently working with the inventors of ABE and BEDS to understand implementation options and how we might facilitate an experimental deployment.

Elevated Services

Goals

- Complete the design of the QBone Premium Service, revising, repairing, and refining the QBone architecture and documenting the factors that have frustrated deployment
- Complete NC I-TEC lab testing of new Cisco GSR QoS functionality
- Suspend all further testing and deployment of elevated services, including the Abilene Premium Service (APS) test program
- Study alternative approaches to engineering deployable elevated services, emphasizing statistically provisioned approaches
- Participate in the evolution of Abilene, studying the implementability of QPS under various design options

There are really two main problem with QPS. First, there are remaining gaps in the theoretical understanding of how to design and build such a service. Second, quite independent theoretical design issues, essential components of QPS have been found to be undeployable. Essential building blocks are missing from most routers and switches (even in the newest and most advanced gear) and QPS fundamentally requires all-or-nothing network upgrades and dramatic operational and economic paradigm shifts for operators that are hard to justify in the absence of clear demand.

Much of the work in this area will be performed by the QBone Architecture Design Team. The QBone architecture that has been recommended, even evangelized, to the Internet2 membership is out-of-date and unrealistic. In the absence of "running code" for the QBone Premium Service (QPS), the Architecture DT must restructure the architecture into a genuine recommendation for Internet2 network operators. At this point, the architecture should probably be scaled back to a loose framework for QoS measurement and monitoring, combined with the formal specification and recommendation of a single QBone service—the QBone Scavenger Service (QBSS). Having removed QPS from the architecture, the design team will undertake two additional efforts: 1) a post-mortem exercise to document the obstacles that frustrated QPS deployment; 2) a concerted effort to understand whether the theoretical problems in the service design can be fixed and, if not, study alternative approaches to engineering a deployable elevated service.

The Abilene Premium Service (APS) test program—an effort to deploy QPS edge-to-edge across Abilene—will need to be suspended. Even if the next generation of Abilene edge cards (so-called "engine 3" cards) perform brilliantly in the lab, they would need to be deployed at all connector interfaces, a prohibitively expensive undertaking. Before configuring priority queueing for QPS in the core, all connectors (including those not participating in APS) must be prevented from injecting more than their contracted profile of EF-marked traffic. The edge, which is now very porous, must be crisped. Simple crisping techniques that clear the DSCP are unacceptable as they would destroy QBSS transparency.

Although we are recommending that the APS test program be canceled, we recommend completing the work that is underway at the North Carolina I-TEC to test the policing, forwarding, and shaping performance of Cisco's "engine 3" GSR line cards. Additionally, UCAID QoS staff will remain engaged in planning for the evolution of the Abilene backbone network, evaluating engineering alternatives from the perspective of QPS implementability. Ability to support QPS in Abilene should not be made a design requirement. However, should a "magic bullet" design emerge that would make a QPS deployment trivial, we will be there to recognize it and may recommend returning to the "lead from the backbone" approach that motivated APS in the first place.

"Magic bullet" routers that support increased sophistication and per-packet complexity in the forwarding plane are unlikely. Guy Almes has observed that the trend towards dramatically lower circuit costs has led to a situation where routers and interface cards are responsible for an increasing share of overall network capital costs. This is in turn putting pressure on router vendors to keep per-interface costs (and hence complexity) as low as possible.

Applications

Goals

- Successfully fill Internet2 Fellowship on Application QoS Needs and work to complete the systematic survey of applications QoS needs
- Study the potential value of an application-specific service initiative (*e.g.* an Internet2 Voice-over-IP (VoIP) service)

Under the leadership of Amela Sadagic, an effort is already underway to survey and document the relationship between application utility and network performance for several key applications. Amela is chairing the Applications QoS Needs Design Team, which has solicited applicants for a fellowship in this area. The design team is currently reviewing more than thirty applications and will select a recipient shortly. The design team will work to supervise the recipient to successful completion of the survey.

The team has already made considerable effort to impart structure, focus, and clarity to this exercise. Several key applications will be selected with distinction made between "usage specific" applications (*e.g.* exchange of high-resolution radiological images) and "implementation specific" applications (*e.g.* a particular piece of software, implementing particular codec, buffering, and file transfer techniques).

The resulting survey is expected to include:

Taxonomy of advanced Internet2 applications This set will be larger than the set chosen for detailed analysis by the fellow. Ultimately, it is hoped that the survey will be expanded to include more applications than are analyzed initially.

Application behavior Several key applications will be described in terms of the demands they place on the network. Some applications provide a constant flow of data and therefore will pose a demand for constant provision of network resources, while others pose occasional but high-pitched demands. Still others are capable of sophisticated adaptation to network performance.

Quantitative measure of application needs The document should contain a review of existing studies that have treated application needs for network services in quantitative terms. Metrics such as loss, delay and jitter must be related to tolerances for different sensory and control data (*e.g.* video, audio, haptic, tracking, database and event transactions, simulation, remote rendering, control). For many applications the lack of adequate response times causes a break in "presence" and "copresence", which has a direct impact on perceived quality and application utility.

Tools and experimentation procedures Finally, the document should identify the tools and experimental methods that were used or are being developed to quantify how changes in network-layer performance translate to changes in application quality and user satisfaction.

A second major activity in the applications space *could* be an application-specific service initiative. The working group should explore options in this area. One possibility would be to form a partnership with the Internet2 VoIP initiative. There is considerable mythology about the stringency of VoIP's QoS needs and, consequently, there could be great value in demonstrating how VoIP service requirements could be met with various simple QoS approaches. In addition to highlighting VoIP's service requirements, there would also be considerable value in nurturing a large user community that values low network delay.

Research

Goals

- Measure and analyze the global Internet2 network infrastructure to improve understanding of application performance, monitor the use and deployment of new services, and catalog and track "congestion trouble tickets"
- Develop an open implementation of the One-Way Delay Protocol (OWDP) that has been proposed in the IETF IPPM working group
- Perform experiments and demonstrations of new services
- Support networking researchers with measurement data and with special access to infrastructure for experimentation
- Stay abreast of emerging QoS research and provide researchers with operators' perspectives on their ideas
- Study ways that Internet2 might facilitate exploration of new pricing models

This is really a catch-all work area for a number of measurement and research activities to be performed by UCAID QoS staff and interested QoS working group members. The relationship between QoS and network measurement runs deep. Network measurement has the ability to provide users, operators, and researchers with invaluable information about the macroscopic (as well as the microscopic) performance of networked applications. A recent CRA panel comprised of a balance of networking "insiders" and "outsiders" criticized the networking research community for not capturing and studying comprehensive network measurements from a large scale internet. The grand challenge issued by the panel was to capture all packets in the Internet for a single day. Networking insiders, of course, complained that this was ridiculous and impossible. With its spirit of openness and cooperation, however, Internet2 is uniquely situated to make a large contribution towards filling this void. Building a thorough and openly instrumented Internet2 may well be the single largest contribution that Internet2 can make towards the advancement of networking technology.

In addition to ongoing passive monitoring of QBSS use, two new QoS-specific measurement activities are proposed.

First, UCAID engineering staff have proposed (in the IETF) a standard protocol for the measurement of one-way delay (OWDP). With the increasing availability of good time sources, accurately measuring one-way delay is becoming possible for an increasingly large number of hosts. The goal of this protocol is to allow interoperable one-way delay measurements between different active measurement devices (*e.g.* Surveyors, AMPs, arbitrary hosts near good time sources). Developing an open-source software implementation of OWDP would help its advancement through the standards process and help one-way delay measurement generally scale to become as commonplace as measurement of round-trip delay is today with ping. The success of OWDP would allow for one-way delay measurements of new and emerging network service classes along a wide variety of network paths. A software engineer is being recruited to assist with this development.

Secondly, we propose to undertake a survey of congestion incidents in the greater Internet2 networking environment and their resolution. There is a great deal of hand-waving and exchange of anecdotes by those who advocate the need for QoS, as well as by those who dismiss it. There is a need for hard data in this debate. We propose an effort to document and track congestion-based performance problems. Since the Internet2 End-to-End Performance Initiative is already evolving a systematic approach to tracking performance problems, it would be natural to leverage this effort to track systematically that minority of end-to-end performance problems whose cause is congestion.

Appendix: A Short History of Internet2 QoS

Since its inception in Fall 1997, the Internet2 QoS Working Group has contributed significantly towards focusing Internet2 QoS activities. In a series of early requirements gathering workshops, the WG studied the QoS requirements of Internet2 applications, obstacles to deploying QoS in campus, gigaPoP, and backbone environments, and evaluated the best current thinking and the state of IETF standards activities in the area.

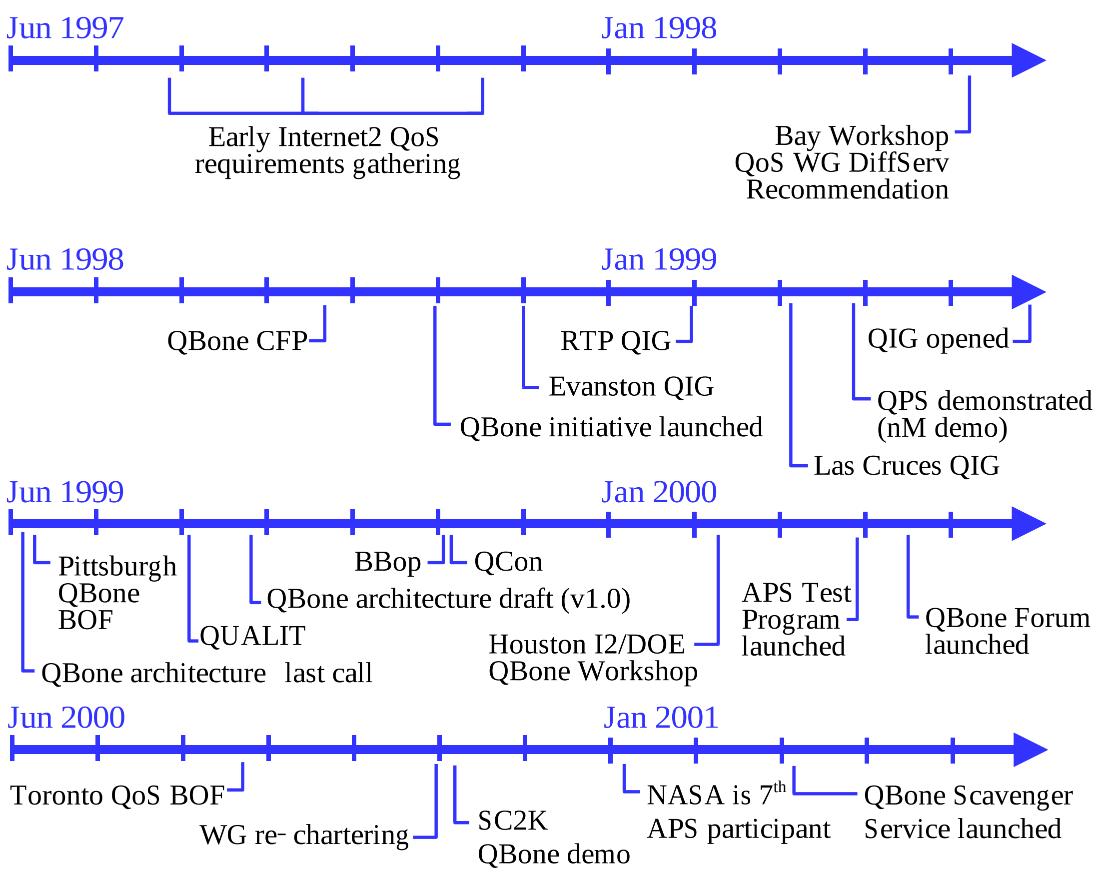

Figure 1: Internet2 QoS Timeline (through March, 2001)

At the first major Internet2 QoS workshop at Bay Networks in May 1998, the WG recommended that Internet2 deploy services that leverage the then newly-emergent differentiated services (DiffServ) architecture for IP QoS. The DiffServ model of QoS was (and remains) attractive for its simplicity and scalability. DiffServ relies on marking IP packets to indicate the forwarding treatment to be applied and emphasizes pushing complexity to the edge of the network (where there are slower link speeds and more processing power) to achieve scalable differentiated treatment in the core. Edge devices police connectors or hosts to agreed-upon traffic profiles, while interior routers simply respect packet markings and queue traffic aggregates accordingly. The IETF has standardized a small number of per-hop behaviors (PHBs) for which traffic may be marked. PHBs are designed to be implementable by various scheduling disciplines commonly available in modern routers.

In May 1998, the WG further recommended that Internet2 focus on deploying the "Premium" service described by Van Jacobson. Premium is defined by a service model that has been described as a "virtual leased line" or "virtual wire" and was judged by the WG to offer the best hope of supporting new classes of real-time interactive applications that require stringent bounds on loss and jitter. Applications with less stringent QoS requirements were judged by the WG as likely to thrive in an over-provisioned best effort environment. Specifically, work on forms of QoS that would offer bounded on TCP bulk transfer transaction times was judged to be premature given the level of over-provisioning and the commonness with which TCP throughput problems are linked to mis-configurations of end-systems and LANs, rather than a lack of available bandwidth.

A second major contribution of the WG has been the creation of the QBone initiative and architecture. The QBone initiative has brought together a number of leading Internet2 networkers and application developers to specify an architecture for an interdomain DiffServ testbed and to facilitate its deployment through a series of workshops and exchanges of deployment experiences.

The QBone architecture specifies the requirements for an Internet2 network to be a *"QBone domain"*, specifies the QBone Premium Service (QPS), defines the notion of a QPS reservation, defines a set of measurement collection and reporting standards to support end-to-end debugging and auditing of QPS reservations, and specifies a common set of operational practices and procedures to be followed by network operators when requesting and responding to reservation requests.

The WG originally felt that the QBone would be most successful if kept to a small, tightly knit group of participating networks. A call for participation was issued in September 1998 and one month later a group of roughly 15 networks assembled into the QBone Interoperability Group (QIG). The QIG was tasked with coordinating actual deployment (as opposed to specification) of the QBone architecture and had a series of three very energetic meetings through March 1999. These meetings focused on sharing information about early DiffServ lab and field trial experiences, discussing implications and likely problems that would-be deployers faced, and, although it was not the intended purpose of the QIG, contributing to the specification of the QBone architecture.

By late May 1999, however, it was clear that while talk was cheap and exciting, actual deployment was neither. The QIG was effectively dissolved (er, "opened") and a new approach to fostering deployment adopted. The key attributes of this new approach were:

Open Participation Model Participation in the QBone would be self-declared; any Internet2 network that implements the architecture and peers with another QBone domain could call itself a QBone domain;

Aggressive Backbone Deployment Greatly increased attention to making Abilene a model QBone participant and using Abilene's centrality as an Internet2 backbone to catalyze QBone participation by Abilene connectors; additionally, it was hoped that UCAID, the Abilene partners, and the I-TECs could muster enough resources to develop tools and configuration recipes that would be transfered to the broader Internet2 community, enabling campuses and gigaPoPs to participate who might otherwise not have sufficient human resources to engineer a QBone domain from scratch; I call this approach *"leading from the backbone"*.

Work on the QBone architecture continued through Spring and Summer of 1999 and, in August 1999, the QBone Architecture Draft was frozen at v1.0, awaiting further deployment experience and input on an automated signaling design from the "QBone Bandwidth Broker Advisory Committee". During this time, UCAID submitted the QBone University and (Department of Energy) Lab Interconnect Testbed (QUALIT) proposal, which was funded accepted by DOE and funded in August 1999 at $0.5 million/year. (Because DOE's NGI funding was subsequently terminated by congress, QUALIT's funding lasted only one year.)

With the exception of the organization of one major workshop (First Joint Internet2/DOE QoS Workshop), a bake-off for proto-bandwidth broker implementations (BBop), a DiffServ conformance testing event (QCon), and ongoing work by some working group members in the design of an experimental signaling protocol for possible inclusion in the QBone architecture (SIBBS), the WG was mostly idle between Summer 1999 and January 2001. This was largely due to the extent to which progress on the QBone architecture had out-paced progress on deployment.

In January 2001, the WG was re-chartered, opened, enlarged, and re-energized. Consistent with the operating guidelines for Internet2 working groups developed by Russ Hobby, the bulk of the actual "work" of the QoS Working Group is now executed by "design team" sub-groups. With rough consensus from the Internet2 QoS Working Group as a whole, the deliverables of the design teams (*e.g.* documents, workshops, recommendations) will become official outputs of the working group. There are currently four active design teams: QBone Architecture, QBone Signaling, Scavenger Service, Application QoS Needs.

Perhaps the most important achievement of 2001 was the recognition of the magnitude of the difficulty of deploying the QBone Premium Service (QPS) and the redirection of Internet2's QoS efforts towards incrementally deployable QoS solutions and towards a careful reassessment of application QoS needs. Other recent accomplishments include:

Testing and deployment of Abilene Premium Service Though Abilene's QBone efforts have been hamstrung by factors described elsewhere in this document, we were able to provide some value to the membership through the Abilene Premium Service test program. Bob Dixon and Paul Schopis (Ohio State University) completed field testing of H.323 video sensitivity to QPS policing. Additionally, Roch Guerin (University of Pennsylvania), his graduate students, and researchers at Northwestern University's iCAIR center completed research on the impact of policing and rate guarantees on streaming media applications; this work was published at SIGCOMM 2001.

Testing of Cisco QoS capabilities Working closely with Cisco, we have successfully completed an early field trial (EFT) of Cisco's MPLS DiffServ aware traffic engineering functionality at the NC-ITEC (these results were reported to the IETF Traffic Engineering WG at the March, 2001 IETF meeting); we have also continued to work extensively with Cisco to critique new QoS capabilities planned for the GSR with respect to their ability to implement a "virtual-wire" Premium service.

Successful demonstration of QPS at SC2000 The successful engineering of the Abilene / DOE Science GRID QoS peering was highlighted at Supercomputing 2000 (SC200) in a demonstration entitled: "QoS-enabled Audio Teleportation". This demonstration used IP Premium service to provide CD-quality interactive audio under artificially induced congestion conditions along two wide-area paths. QoS-enabled audio was demonstrated between the SC200 show floor in Dallas, TX and Lawrence-Berkeley National Labs and between the SC2000 show floor and the Computer Music Center at Stanford University (CCRMA). The path to LBNL traversed Abilene, the DOE Science GRID, and the LBNL QoS testbed, while the path to Stanford traversed Abilene, CalREN2, and a portion of the Stanford campus LAN. Edge routers were signalled to mark application traffic for QoS using the GARA client from the Globus project. To our knowledge, this is the first demonstration of an interdomain IP Premium service. In recognition of the engineering challenges overcome for this demonstration, it was awarded the SC2000 Network Challenge award for "Best-Tuned and Most Captivating Application".

Successful launch of QBone Scavenger Service (QBSS) initiative At least four schools are marking dormitory subnets, patches have been developed for Apache, WU-FTP, and BBFTP to mark bulk transfers for QBSS, and a number of high-energy physics groups are in trials to use QBSS.

Completion of the QUALIT project UCAID's QUALIT project, tasked with advancing interoperable QoS along paths between Internet2 universities and DOE laboratories, concluded July 31, 2001.