SLAs for Application-Specific QoS for IP-based VPN Services
Open Group QoS Task Force Roundtable Discussion
April 30, 2001 / Burlington, MA

1. Situation

- IP-based networks provide Web, email and remote connectivity to millions of users. In addition, IP-based VPNs are increasingly being used as alternatives to leased line, Frame Relay, or ATM-based private networks. Multi-service IP networks are gaining favor over "siloed" application-specific networks that require separate circuits, specialized equipment, and system-specific metrics.

- End-users and service providers are loading networks with more data types (voice, video, e-mail), and with more service scenarios (e-commerce, enterprise applications, batch transfer, peer-to-peer, streaming) than ever before. Everyone wants his or her applications to work reliably, even at peak load.

- A growing community of end-users, service providers, and vendors sees IP Quality of Service (QoS) as the smart alternative to massive over-provisioning, layered networks, and brute force mixing of data. The QoS approach allows everyone to use the network more efficiently with router and switch-based QoS.

- Technology standards (DiffServ, MPLS) have emerged to enable delivery of QoS over IP networks. DiffServ provides a language for the creation of service level agreements (SLAs), and a clear method for IP billing. DiffServ also satisfies requirement for edge conditions by domain to customers, across peering points, and to the backbone.

2. Problem

- Two basic requirements for managing QoS delivery over IP networks are *monitoring* and *enforcement*. Yet while the *technology* of QoS on IP networks has advanced considerably in recent years, not as much progress has be made in how Service Level Agreements for QoS on IP networks should be *specified, structured, and/or enforced,* nor how end-user organization can *request, signal or manage* the QoS that they require over their service provider's IP networks.

- Policy-oriented work from the past, while specifying user group, security level, locations, devices and, occasionally, even time & date and network conditions, fails to address the always-on, one-through-many to one-through-many, and intense mobility access requirements faced by network-delivered services today. And now, given malicious attacks from around the world, security and accountability are often the critical factor in defining any service. QoS-enabled SLAs, in contrast, promise highly granular control over data types, which translate into application-specific contracts and tighter monitoring on both sides of the customer/service provider dialog.

3. Hypotheses

- Two levels of analysis make sense for today's discussion: i) classifying applications to the required QoS and ii) ensuring that IP QoS at the network level is mapped to underlying network behaviors. Each of these two steps continue to need further definition, but when accomplished can result in a clear basis for negotiation among end-users, service providers, and vendors.

- At the *application* level, leading candidates for defining QoS are uptime, throughput, responsiveness, and emergency provisioning. At the *network* level, leading candidates are guaranteed bandwidth, burst capability, loss tolerance, and delay.

4. Questions for Discussion

- Can we prioritize QoS attributes for end-users and service providers?
- What minimum set of QoS attributes need to be added to today's SLAs?
- Who flags an out-of-contract situation? How do we know when an attribute is out-of-contract?
- How do we think about service interruption in a QoS world? How do we think about availability, mean time to repair (MTTR), mean time between service outage (MTBSO)?
- How to manage admission control for various traffic types?
- Specific techniques for defining SLAs for security in the context of data encryption, NAT and firewalls?