IP QoS Support in the Internet Backbone

OCTOBER 2000

Technical Paper

ALCATEL
ARCHITECTS OF AN INTERNET WORLD
IP QoS Support in the Internet Backbone

Quality of Service in the Internet  1

IP QoS Architectures  1
Integrated Services Architecture  1
Differentiated Services Architecture  2

Alcatel’s Vision for IP QoS in the Internet  3

Alcatel’s Solution for IP QoS in the Internet Backbone  6
IP QoS Requirements for High Speed Edge and Core Network Routers  6
IP QoS Implementation and the Alcatel 7770 RCP’s LSR Functionality  9

Conclusions  14

Acronyms  15

References  15
Quality of Service in the Internet

Today’s Internet applies best effort (BE) IP forwarding. The network attempts to deliver all traffic as soon as possible within the limits of its abilities, but without any guarantees related to throughput, delay, delay variation and packet loss. It is left up to the end systems to cope with network transport impairments.

The BE forwarding paradigm has been suitable because most applications running over IP are low priority and low bandwidth data applications with high tolerance for delay and delay variation (jitter), but this scenario is changing rapidly. The convergence of other networks (telephony, audio and video) on the Internet is well underway. New, multimedia Internet applications require significant bandwidth and/or strict timing. Furthermore, the number of Internet users is increasing tremendously, resulting in greater congestion that causes network delays and packet loss.

Simply throwing bandwidth at the problem is not enough. It is true that the uniform BE network service would remain adequate if the capacity of the network elements and links could be increased to the level that all delays are eliminated. Adding bigger pipes is indeed part of the solution. However, it is not a sufficient response, as temporary overload and congestion can never be avoided. The network must be able to offer different levels of guarantees to certain applications or customers. This is what IP quality of service (QoS) is about. IP QoS functionality allows network operators to offer a range of network services beyond the traditional BE service, differentiated on the basis of performance, possibly encompassing throughput, delay, jitter and packet loss.

Although best effort will remain adequate for the majority of users, QoS support at the IP layer will be required if the Internet is to establish itself as the de facto universal multiservice infrastructure.

IP QoS profoundly affects the design of network elements. In this paper, we discuss the implementation of IP QoS support in core and high speed edge routers. These network elements will form the backbone of the future multiservice public network.

IP QoS Architectures

Integrated Services Architecture

The integrated services (IntServ) architecture defines a set of extensions to the traditional BE service model of the Internet to provide application-dependent QoS to individual sessions [1]. Applications explicitly request from the network a network service providing a set of well-defined QoS guarantees. Signaling is used to create and maintain the required flow-specific states in network elements (for example, routers) allowing them to provide the requested services. This is fundamentally different from the current BE paradigm, in which all flows are handled in a uniform manner.

The IntServ architecture provides a set of service definitions. Currently, two service classes are defined in addition to the best-effort network service, namely controlled load [2] and guaranteed service [3]. Controlled-load service (CLS) aims at offering a low average delay and limited loss. It is intended to offer roughly the same QoS independent of the network load, that is, if a flow is accepted for the CLS, then the routers make a commitment to offer that flow a service level equivalent to that seen by a BE flow on an unloaded network. The CLS aims at supporting those applications that can tolerate a small amount of delay and loss (for example, adaptive real time services). Guaranteed service (GS) offers a quantifiable bounded queuing delay and no loss, and is intended for real time applications with stringent timing requirements.

Figure 1 illustrates how IntServ affects the design of routers.

A module for the processing of IntServ signaling is required. IntServ uses resource reservation protocol (RSVP) to signal per flow QoS requirements and to reserve resources in the network elements along the route between source and destination [4, 5]. The intermediate nodes apply IntServ admission control. An algorithm decides whether a new flow’s QoS requests can be supported without affecting the QoS guarantees required by already admitted flows. RSVP is a receiver initiated protocol. It establishes “soft” state; RSVP sends periodic refresh messages to maintain the state along the reserved path. In the absence of refresh messages, the state automatically times out and is deleted.
In addition, traffic control mechanisms in each network element are required to ensure that each admitted flow receives the service requested in strict isolation from other traffic. To this end, RSVP signaling configures microflow packet classifiers in IntServ-capable routers along the path of the traffic flow. Each admitted flow is properly policed. Misbehaving users cannot steal bandwidth in CL or GS traffic classes because nonconforming packets are treated as BE traffic.

IntServ nodes offer the required performance guarantees through the implementation of multiple queues per output port in combination with scheduling and active buffer management. Typical mechanisms are weighted fair queuing (WFQ) scheduling and a random early detection (RED) variant for congestion avoidance. IntServ/RSVP aims at providing absolute service guarantees to GS on a per flow basis. This typically requires a separate queue per GS flow.

There are a number of weaknesses that impede the deployment of IntServ/RSVP at large and, especially, in high speed backbone networks. The most important of these is scalability.

IntServ/RSVP requires a substantial amount of per flow processing in each node. This includes end-to-end signaling and information to identify the flow, track and bill the resource consumption, police traffic for compliance to the agreed traffic profile, and schedule the traffic according to the requested performances. The resources that a network element needs for RSVP processing and storage increase proportionally with the number of IntServ flows.

In view of the above, the IETF started working on the differentiated services (DiffServ) architecture for IP QoS. The premise of DiffServ networks is that routers give different treatment to packets according to the aggregate traffic stream to which they belong. This significantly simplifies processing and reduces storage requirements. In addition, no explicit signaling is required.

### Differentiated Services Architecture

In contrast to the per flow behavior of IntServ, DiffServ tries to deliver a particular service to aggregated flows. In the DiffServ architecture, packets are classified into a small number (maximum 64) of aggregated flows or classes. A DiffServ class is referred to as a behavior aggregate (BA).

Membership to a class is indicated by the DiffServ CodePoint (DSCP) in the packet’s IP header [7]. The DSCP is defined in the differentiated services (DS) field of an IP header. The DS field redefines the IPv4 type of service (ToS) byte (Figure 2) and the IPv6 traffic class byte.

At each DiffServ router, packets are subjected to a per hop behavior (PHB) according to the class they belong to. A PHB defines a DiffServ router’s externally observable forwarding behavior related to a BA. A set of BAs can be combined to form a so-called ordered aggregate (OA). In that case, IP packets belonging to that set of BAs will not be reordered, regardless of the DSCP of the packets. The set of PHBs applied to such an OA is then referred to as a PHB scheduling class (PSC).

PHBs enable operators to construct the intra-domain services they believe will meet their customers’ needs. It is the operator’s responsibility to decide what services to support and to assign the corresponding DSCP. Different operators can have different mappings between DSCPs and PHBs. Therefore an appropriate translation of DSCPs must take place at their borders.

![Figure 1: IntServ router](image)
The primary benefit of DiffServ is its scalability, since it eliminates the need for per flow state and processing, at least in the core. Since service is allocated at the granularity of a class, the amount of state information is only proportional to the number of classes.

**Alcatel’s Vision for IP QoS in the Internet**

In the above section we gave an overview of the integrated services and differentiated services architectures for IP QoS. The merits and limits of both are reflected in the trade-off between scalability and level of QoS performance assurance. These IP QoS architectures should be seen as complementary tools that should coexist and effectively interoperate to provide end-to-end QoS.

This complementarity is reflected in Alcatel's IP QoS strategy, which positions DiffServ in the backbone network while IntServ/RSVP can potentially be deployed in the access part of the network at the scale of corporate and campus networks.

Before we elaborate further on this QoS strategy, it is useful to introduce a reference network model. Figure 3 shows a typical network configuration for an Internet service provider (ISP). The core network routes traffic between a number of points-of-presence (PoP). A PoP typically consists of numerous relatively low speed access routers, distribution equipment and a few high speed edge routers. Users connect to the PoP through access routers. The distribution equipment in the PoP aggregates the user traffic onto higher speed links toward the edge routers. The edge routers connect the PoP to the core network. Besides individual users and large corporations, an ISP can also have other ISPs as customers. Today, peering with other ISPs is mostly done at the level of the access routers or edge routers in the PoP. This requires routers with border functionality.
A sample mapping of Alcatel’s IP QoS strategy on this network model is represented in Figure 4. Such a mapping is generally accepted to be the most scalable solution for implementing service differentiation in the Internet.

IntServ and all related state information is pushed out of the core as far as possible towards the end user premises. Typically, per flow packet processing, such as per flow policing and shaping and DSCP marking, is only done in customer premises equipment (CPE) or in the access part of the network, providing these microflow classification and conditioning operations are most appropriately performed close to the customer’s premises where they scale better due to the lower speeds and the limited number of flows.

This eliminates the need to keep any per flow state information elsewhere in the network. Packets are classified and marked to receive a particular forwarding behavior in the DiffServ domain interconnecting the customers. Within the DiffServ domain, packets are forwarded according to the per hop forwarding behavior associated with the DSCP.

Note that this is a simplified network model, which assumes that the DiffServ domain extends from access to access. In reality, some users might prefer to outsource part of the required per flow packet processing to their ISP. In that case, the access routers (or edge routers) in the PoP will be responsible for mapping IntServ flows on DiffServ BAs.
In this IP QoS framework it is important to make a distinction between the boundary nodes and the interior nodes of a DiffServ domain. DiffServ boundary nodes are located at the edges of the DiffServ domain. In our reference model, the DiffServ boundary nodes are the access routers and edge routers in the PoP that directly peer with customers (either individual users or peering ISPs).

Providers and customers negotiate agreements with respect to the services to be provided: service level agreements (SLAs) define the conditions for delivering services at each customer/provider boundary. The administration of a DiffServ domain must assure that sufficient resources are provisioned to support the SLAs committed by the domain. Moreover, the DiffServ boundary nodes are required to perform traffic classification and conditioning to enforce the negotiated SLAs. Traffic classification and conditioning at DiffServ boundary nodes ensure that traffic entering the DiffServ domain complies with the agreed SLA. In our reference model, these traffic classification and conditioning functions operate at the granularity of DiffServ BAs.

However, finer grained SLAs need to be enforced when part of the per flow processing (for example, per flow classification and DSCP marking) is outsourced to the ISP.

In general, interior nodes within a DiffServ domain do not have to perform any complex traffic classification and conditioning operations.

This end-to-end QoS framework is built around the vision that using RSVP/IntServ in the customer premises and in the access part of the network (and possibly in parts of the PoP), in conjunction with DiffServ in the core and (parts of) the PoP, allows providers to support applications with tight QoS requirements (for example, voice over IP, voice on demand, etc.) while simplifying the management of end-to-end QoS. This requires, of course, an adequate IntServ to DiffServ mapping at the customer’s premises or in the PoP, combined with suitable admission control and resource provisioning in the DiffServ domain. With suitable admission control and resource provisioning, an ISP will be able to offer services that effectively extend integrated services across a DiffServ backbone.
Alcatel’s Solution for IP QoS in the Internet Backbone

In this section we provide a detailed overview of the functional requirements for a router that could be positioned as a core or high speed edge DiffServ router. Afterwards, we introduce the Alcatel 7770 Routing Core Platform (RCP), Alcatel’s answer to these requirements.

IP QoS Requirements for High Speed Edge and Core Network Routers

In order to effectively extend integrated services across a DiffServ backbone, the backbone needs to provide a set of services that offer distinct throughput, loss, delay and jitter performance guarantees as well as an appropriate mapping of IntServ flows on DiffServ aggregates.

DiffServ-capable routers implement different forwarding behaviors (PHBs) for distinct traffic aggregates. Services built on different PHBs will differ in performance guarantees in terms of throughput, loss, delay and jitter. However, these performance guarantees will be met only if incoming traffic is kept in profile through proper traffic classification and conditioning functions at the boundaries of the DiffServ domain. In addition, distinct handling of traffic aggregates must be complemented with a resource reservation scheme and an admission control (AC) function for the DiffServ domain.

In this section we review the functionality required to offer service differentiation and QoS performance guarantees in a DiffServ backbone. This results in a separate set of requirements for interior (inner core and core edge) and boundary (high speed edge) nodes.

Traffic Classification and Conditioning

Providers and customers negotiate SLAs with respect to the services to be provided at each customer/provider boundary. The service level specification (SLS), a subset of the SLA, describes the technical aspects of the service’s QoS performance and includes a traffic conditioning specification (TCS) to be used by the traffic classification and conditioning functions at the boundary between the provider and the customer.

The TCS includes detailed service parameters for each service level, encompassing service performance parameters, the scope of the service, traffic profiles (for example, token bucket parameters), actions to be taken upon noncompliant traffic, and requirements related to marking and shaping.

The scope of the service puts constraints on the egress points to which the service is provided. Based on the scope of the service, SLAs can be classified into two categories: committed information rate (CIR) and committed access rate (CAR). CIR SLAs, or so-called ‘pipe’ model SLAs, describe traffic sent from a specific ingress point to a fixed egress point. CAR SLAs, or ‘hose’ model SLAs, describe user traffic in terms of characteristics of the traffic that will be entering the core, without specifying the egress point(s) of the traffic. CAR-type SLAs are standard in the current Internet, characterized by any-to-any connectivity.

In addition to the TCS, the SLS may specify service characteristics such as availability and reliability, authentication mechanism, pricing and billing mechanisms, etc.

The granularity of traffic classification and subsequent conditioning at the DiffServ boundary depends on the specifics of the service offering. The SLA between the customer and provider domains specifies which domain is responsible for traffic marking into DiffServ aggregates and policing of the aggregated traffic.

In the case of hose SLAs, if the nodes within the customer domain perform appropriate traffic classification and conditioning, including traffic marking, then authentication may be based on the physical link on which the traffic was received, and classification may be based on one packet header field, namely the DSCP (so-called BA classification).

In some cases, more advanced classification means may be needed at the boundary of the DiffServ domain. This is accommodated by multi-field (MF) classification in which the classification key consists of a number of packet header fields. MF classification allows a service provider to offer value-added services such as marking, policing, or shaping of specific microflows. Customer domains may want to outsource these functions to their service provider. This functionality is, for instance, required for peering to a non-DiffServ-capable domain.
In general, the boundary routers of a DiffServ backbone network need not be concerned with individual microflows. Packets arriving at the boundary of a DiffServ backbone are likely to be premarked. Classification is best done close to the traffic source or in low speed routers in the PoP. In any case, within the interior of a DiffServ domain, BA classification suffices. Traffic arriving at an interior node has been premarked and preshaped either by the customer or by a boundary router.

Although conditioning in the customer domain may ensure that the aggregated traffic forwarded to the provider’s DiffServ domain conforms to the negotiated TCS, DiffServ boundary nodes should also monitor and enforce conformance to the TCS. In general, a DiffServ boundary node has to assume that incoming traffic may not conform to the TCS. At the very least, the service provider must limit traffic carried on behalf of a customer to the constraints specified in the TCS. Traffic conditioning at the network boundaries is essential for the creation of services on the PHBs implemented in the DiffServ domain. Traffic conditioning protects the network and consequently, other flows sharing the resources of the network from misbehaving users.

Because DiffServ services are unidirectional, a separate treatment is required for the two directions of flow across the boundary.
The required classification and conditioning functions at the DiffServ boundary can be summarized as a set of traffic components including classifiers, meters, (re)markers, shapers and droppers. The relationship between these traffic components is shown in Figure 6.

Classifiers sort received packets into a set of output streams by means of filters. Meters measure the temporal properties of aggregated traffic streams selected by a classifier against a preestablished traffic profile. They determine whether a particular packet is in-profile (compliant) or out-of-profile, and provide control inputs for the other components that implement policing. They may also be used for accounting and measurement purposes. Different traffic conditioning actions may be applied to in-profile packets and out-of-profile packets. Different accounting actions may be triggered for compliant and noncompliant traffic. Shapers police traffic by delaying submitted traffic within the boundaries of the traffic profile, while droppers simply discard nonconforming traffic. Markers/remarkers can also be used as a policing tool. The SLS could specify that out-of-profile traffic should be remarked to an inferior BA. Markers can also be used to offer a network marking functionality to specific microflows or for CodePoint translation.

These functions may also be required at the egress of the provider’s DiffServ domain. Depending on the SLA at that boundary, the provider may want to condition the traffic leaving its domain.

Complex traffic conditioning functions are only needed in the boundary nodes (ingress and egress) of a DiffServ domain, although they can be used in interior nodes.

**Per Hop Behaviors**
DiffServ routers should implement an extensive set of PHBs. They are essential for the creation of differentiated services in the backbone.

DiffServ nodes provide multiple PHBs by implementing multiple queues per output port in combination with scheduling and active buffer management. Similar to the IntServ case, typical mechanisms are WFQ scheduling and a RED variant for congestion avoidance, but in this case applied to the level of traffic aggregates.

A number of PHBs and PHB groups have been specified for standardization within the IETF as additions to the default best-effort PHB. These are the expedited forwarding PHB [8], the assured forwarding PHB group [9] and the class selector PHB group [7]. The difference in forwarding behavior for these PHBs can be expressed in terms of throughput, experienced loss, delay and jitter.

The expedited forwarding (EF) PHB offers low loss, low latency and low jitter for an aggregated flow. It is particularly useful for Internet telephony, video conferencing and virtual leased line services. EF traffic should encounter queues that are almost empty everywhere along the end-to-end path. This requires that EF traffic be treated as high priority traffic and that sufficient resources be provisioned for it within the network.

The assured forwarding (AF) PHB group is intended for customers that need predictable services, even in times of congestion. It specifies four independent forwarding classes (AF1 to AF4), each with three levels of drop precedence (or ‘colors’). The drop precedence defines the relative importance of a packet within a particular class in the event of congestion. Each forwarding class within the AF PHB group constitutes an OA, that is, packets belonging to a particular AF class may not be reordered by a DiffServ node, regardless of the drop precedence of the packets. Consequently, each AF class in the AF PHB group corresponds with one PSC. Each of the four AF PSCs consists of three PHBs.
The AF PHB group controls congestion through the implementation of a congestion avoidance mechanism (for example, RED). The AF PHB group could, for instance, be used to offer so-called ‘Olympic’ services (Gold, Silver and Bronze), characterized by different probabilities of timely forwarding. Within each Olympic service, one can further offer loss differentiation based on the three levels of drop precedence.

The class selector (CS) PHB group assures backward compatibility with non-DiffServ networks that implement IPv4 precedence classification and forwarding. Each CS PHB forms a separate PSC as packets marked with different Class Selector CodePoints may be reordered.

For BE traffic, no assurance on any performance parameters is given besides the fact that this BA may not be starved.

The above listed PHBs and PHB groups form the minimal set of PHBs that any DiffServ router should support.

**Resource Provisioning and Admission Control**

Traffic classification into a set of BAs at the boundary of the DiffServ domain, and subjecting these aggregated flows to distinct PHBs inside the DiffServ domain, are necessary but insufficient operations to provide any service guarantees. They have to be complemented with resource provisioning and admission control functions.

Resource provisioning in a DiffServ network is fundamental for offering services that effectively extend integrated services across a DiffServ domain. In every node, the network administrator must allocate a sufficient amount of resources in terms of bandwidth and buffer space to each service class to support the SLAs committed by the domain. The required resource reservations can be calculated from network topology information and from the characteristics and performance requirements of the aggregated traffic forwarded to the provider’s DiffServ domain (as described in SLAs).

If one wants to guarantee the service quality, one must also control the amount of traffic admitted for each service class. Traffic conditioning at the boundaries of the DiffServ domain is one essential part of the solution. However, it has to be extended with an admission control function for the DiffServ domain that can evaluate whether or not additional SLAs (or changes to existing SLAs) can be supported without impacting the already committed SLAs.

**IP QoS Implementation and the Alcatel 7770 RCP’s LSR Functionality**

**Alcatel 7770 Routing Core Platform (RCP)**

The Alcatel 7770 Routing Core Platform (RCP), Alcatel’s core router with label switch router (LSR) functionality, is a terabit-ready, wire-speed DiffServ switch router. The Alcatel 7770 RCP targets the inner core, core edge and high speed edge (OC-48) of the future multiservice public network.

In this section we focus on the Alcatel 7770 RCP’s IP QoS implementation.

The Alcatel 7770 RCP uses state-of-the-art IP QoS technology to offer DiffServ support for native IP packets as well as labeled multiprotocol label switching (MPLS) packets. This DiffServ implementation is based on the network communication processor developed by IBM in co-operation with Alcatel.

The Alcatel 7770 RCP incorporates a wire-speed forwarding engine and supports advanced classification, traffic conditioning (shaping, policing), buffer management (congestion avoidance) and packet scheduling functions — in short, everything that is needed to create an extensive set of PHBs in the Alcatel 7770 RCP, as detailed below.

**Per Hop Behavior and Service Differentiation in the Alcatel 7770 RCP**

The supported PHBs form the basis for offering value-added services in the IP backbone network. Services are realized by the use of particular packet classification and traffic conditioning mechanisms at boundaries, coupled with the concatenation of well-provisioned per hop forwarding behaviors along the transit path of the traffic.

By default, the Alcatel 7770 RCP supports all PHBs and PHB groups currently specified for standardization within the IETF.

The EF PHB is defined as a forwarding treatment for a particular DiffServ aggregate where the departure rate of the aggregate’s packets from any DiffServ node must be equal to a configurable rate. Besides an assured bandwidth, services built on the EF PHB also offer low loss, low latency and low jitter, provided that sufficient resources are reserved for EF traffic on every link in the network.
Typically an aggregate peak rate value is reserved for EF traffic. The EF queue is given priority up to the configured rate: on every egress port, EF traffic is shaped at the configured departure rate. A head-of-line priority system can be created by choosing the departure rate equal to the link rate. Note that the EF PHB is only one of two required parts to implement EF-based services. The EF PHB must be complemented by network boundary traffic conditioners, which condition the aggregate via policing and shaping.

The AF PHB group provides delivery of IP packets in four independently forwarded AF classes. Within each AF class, an IP packet can be assigned one of three different levels of drop precedence. A minimum amount of forwarding resources (buffer space and bandwidth) can be allocated to each AF class. Packet forwarding within one AF class is independent from traffic belonging to other AF classes: each AF class is queued separately. The AF implementation responds to long term congestion within each class by means of RED-based congestion avoidance mechanisms. The level of forwarding assurance of an IP packet belonging to one of the AF classes depends on how many forwarding resources have been allocated to the corresponding AF class, the current load of the AF class and, in case of congestion within the class, the drop precedence of the packet. Based on the AF PHB, a service can be created for which IP packets are forwarded with high probability as long as the aggregate traffic does not exceed a subscribed information rate (profile). However, traffic may exceed the subscribed profile with the understanding that the excess traffic will not be delivered with as high probability as the traffic that fits the profile. Offering this type of service again requires an adequate provisioning of AF resources within the network elements, complemented by network boundary traffic conditioners for AF traffic. AF traffic conditioning actions at the ingress of the provider’s DiffServ domain make sure that an AF class in the DiffServ nodes is only moderately loaded by packets with the lowest drop precedence value and is not overloaded by packets with the two lowest drop precedence values. Under these conditions, the AF class can offer a high level of forwarding assurance for packets that are within the subscribed profile (that is, marked with the lowest drop precedence value) and offer up to two lower levels of forwarding assurance for the excess traffic.

The class selector (CS) PHB group implemented in the Alcatel 7770 RCP assures backward compatibility with non-DiffServ networks that implement IPv4 precedence classification and forwarding. By default, these CS PHBs are mapped on the EF and AF PHBs. CS traffic with the highest priority is mapped on the EF class. Other CS traffic is mapped on the AF classes. An operator has the option to change this mapping.

The BE PHB implements a queuing discipline that sends packets of this aggregate whenever the output link is not required to satisfy another PHB. Proper provisioning ensures that this aggregate is not starved.

In addition to these ‘standardized’ services, the Alcatel 7770 RCP offers a large amount of flexibility to service providers to create proprietary DiffServ services. By default, the Alcatel 7770 RCP is configured to support all of the PHBs and PHB groups discussed earlier. However, a service

| DiffServ and Non-DiffServ Traffic Classes |
|-------------------------------|------------------|----------------|---------------|---------------|
| **Traffic Classes** | **Throughput** | **Loss** | **Delay** | **Delay Variation** |
| EF                | Guaranteed       | Minimized     | Minimized     | Minimized     |
| AF 1–4, Pipe      | Minimum Guaranteed | Minimized, depending on drop precedence | Statistical | Statistical |
| AF 1–4, Hose      | Statistical      | Statistical, depending on drop precedence | Statistical | Statistical |
| CS PHB group (Pipe or Hose) | Guaranteed or Statistical (*) | Minimized or Statistical (*) | Minimized or Statistical (*) | Minimized or Statistical (*) |
| BE (Pipe or Hose) | Statistical (not starved) | Statistical (not starved) | No guarantee | No guarantee |
| Better Best Effort (Pipe) | Guaranteed | Minimized | Statistical | Statistical |

* Example configuration where the CS PHBs are mapped on the EF and AF PHBs.
provider is free to create additional PHBs. For instance, it is possible to create an additional PHB (for example, AF5) if it seems useful to offer five independent levels of forwarding for AF services in the backbone.

The Alcatel 7770 RCP can also be deployed in a non-DiffServ domain. In that case, BE and better best-effort (BBE) PHBs are supported. BBE is an extension to the traditional best-effort service model. The BBE service allows carrying specific traffic flows at a higher priority than competing best-effort traffic. Bandwidth can be reserved for these flows. The BBE service is used for label switched paths (LSPs) and for network control and internetwork control traffic (as indicated by the precedence bits in the ToS byte).

**DiffServ Support for Label Switched Paths**

The Alcatel 7770 RCP is a terabit-ready switch router. As such, it supports DiffServ not only for native IP packets but also for MPLS packets. The Alcatel 7770 RCP’s MPLS implementation enables hardware label swapping when used as a transit LSR, and line rate classification and labeling when used as an ingress LSR. Backup LSPs and fast swapping at the ingress LSR are supported. Although MPLS labeling is primarily intended to determine a packet’s next hop, functionality has recently been added to offer service differentiation to MPLS packets [13].

Both EXP-inferred PHB scheduling class (PSC) LSPs (E-LSP) and label-inferred PSC LSPs (L-LSP) are supported, with or without bandwidth reservation. E-LSPs and L-LSPs differ in how they encode a packet’s PSC and drop precedence.

With an E-LSP, the packet’s DSCP value is translated into the EXP field of the MPLS shim header [15] at the boundary of the MPLS-DiffServ domain. Consequently, the EXP field encodes both the PSC and the drop precedence of the packet (Figure 7). The PSC can thus be inferred from the experimental bits in the MPLS header.

Consequently, one E-LSP established for a given forwarding equivalence class (FEC), as determined by the MPLS label, may be used for transport of up to eight BAs of that FEC. Of course, different BAs transported over the same E-LSP are given a different forwarding treatment by the DiffServ LSR. The mapping between EXP and PHB is a local matter, defined by the network administrator and configured in each LSR.

![Figure 7: MPLS label stack entry](image)
With L-LSPs, the PSC applied to a labeled packet is inferred from the MPLS label value, while the EXP field only encodes the drop precedence to be applied by the LSR. So, all packets belonging to a single OA (PSC) and the same FEC are sent in a single LSP. Different BAs (PHBs) belonging to the same OA are characterized by different drop precedences. An LSR determines the PHB to be applied to a packet from the MPLS label and the EXP field.

L-LSPs have some advantages over E-LSPs. First, the number of supported BAs should not be limited to eight. Moreover, traffic engineering can be done at a finer granularity with L-LSPs. L-LSPs are established per (FEC, OA) pair, instead of per FEC. As different OAs are supported via separate LSPs, they can be routed separately. Finally, L-LSPs give the flexibility to apply LSP protection at the level of (FEC, OA) pairs. This gives the network operator the flexibility to protect certain OAs with a backup LSP while others may be supported without backup LSP.

The use of L-LSPs also has some drawbacks. Using L-LSPs results in a higher volume of LSP signaling, since each (FEC, OA) pair requires label setup/teardown signaling. Moreover, the amount of forwarding state information to be maintained in LSRs increases significantly with L-LSPs, as a separate entry is needed per OA, whereas E-LSPs allow up to eight BAs per label. Finally, using L-LSPs also results in a faster consumption of the MPLS label space and limits the number of FECs that can be supported.

During the setup of an L-LSP, the PSC should be explicitly signaled. The extensions to RSVP defined for the establishment of LSPs in MPLS networks can be found in LDP as specified in [16]. [13] specifies extensions to LDP for the support of differentiated services by means of RSVP in MPLS networks. [13] also specifies extensions to LDP for the support of E-LSPs and L-LSPs.

The Alcatel 7770 RCP supports both E-LSPs and L-LSPs. This gives the network operator flexibility in selecting the granularity supported for traffic engineering and LSP protection in the domain. The operator can choose to apply multiple LSPs for a given FEC (that is, L-LSPs) only when this adds value. Otherwise, E-LSPs can be used to conserve label space and to reduce LSP signaling overhead in the network. Both RSVP and LDP are supported for the setup of LSPs.

Resource Management and Admission Control in the DiffServ Core: Alcatel’s Bandwidth Broker

The Alcatel 7770 RCP provides the necessary hooks to operate with centralized or distributed resource management and admission control. The Alcatel 7770 RCP supports access to a bandwidth broker. Moreover, the Alcatel 7770 RCP also supports resource reservation for DiffServ pipes and LSPs through signaling, and is provided with a per node admission control function.

In this section, we focus on the case where resource provisioning and admission control is the responsibility of a central controller. Alcatel’s bandwidth broker is briefly introduced.

A bandwidth broker is a central controller that enables SLA negotiations by granting or refusing new SLAs or changes to existing SLAs. The bandwidth broker is an off-line ‘oracle’ that has a complete view of network topology and resource availability to make SLA admission control decisions.

The bandwidth broker makes decisions based on the network topology and the network traffic characteristics.

- The network topology consists of a description of all available network resources: nodes, links, link metrics, physical link capacities, allocatable link capacity, resource class (Gold links, links only to be used for premium customers), etc. The topology information is derived directly from the nodes if possible (via CLI or SNMP), but can be manually adjusted.

- The network traffic characteristics are expressed as a set of traffic trunks, which mainly express a bandwidth requirement between core edge nodes. This information is supplied by the policy manager, which stores committed SLTs.

Resource provisioning for CIR SLTs is relatively straightforward because the bandwidth broker knows the network topology and the routing protocol. Hence, for each CIR SLT, the route taken between the fixed ingress and egress point can be calculated and the amount of traffic on the associated links can be accounted for. Quantitative guarantees can be given in CIR SLTs.
It is harder to account for CAR SLS-based traffic. In this case, the egress points are not known in advance. For hose model SLSs, resources are reserved using a prediction of the traffic distribution based on historical patterns. Monitoring of the actual traffic flows enables the reservation to be refined over time and the level of overprovisioning needed to offer the promised statistical (qualitative) guarantees to be dropped. To achieve this, the bandwidth broker has access to the network performance monitoring data.

Depending on the CIR service offered, it can be useful to provision resources first for pipe-based traffic, because provisioning for CIR services can be deterministic. For CAR-based traffic, provisioning can only be done statistically. As a result, hard deterministic QoS can be granted to CIR-based traffic, while CAR-based traffic receives statistical QoS. The uncertainty related to hose-based traffic makes it necessary to use different DSCPs for CIR-based and CAR-based traffic so that both types of flows can be isolated from each other within the DiffServ core.

On the basis of the topology and network traffic characteristics, the bandwidth broker can calculate different indicators, such as link loads. This information can be used to determine whether a new SLS can be accepted. The position of the bandwidth broker is depicted in Figure 8.

Note that if a lack of resources is detected in some node(s) in the core, traffic engineering (TE) might be triggered to provide an optimization of the resource utilization by rerouting some traffic around the bottleneck nodes. The problem is that the used interior gateway protocols (IGP), such as OSPF and IS-IS, select routes based on static parameters (for example, the path with minimal accumulated

![Figure 8: Bandwidth broker](image-url)
link weights) and do not take into account the available bandwidth. Depending on the topology of the network, IGP routing might lead to a load distribution that is far from optimal, causing recurrent points of congestion and artificial bottlenecks. This leads to a few overutilized links, while other subsets of the networks remain underutilized. TE provides techniques to route paths around bottlenecks by deviating traffic from the path selected by the IGP protocol onto less congested links. In other words, it helps to optimize the network resources.

Alcatel’s bandwidth broker includes an advanced TE tool that optimizes resource utilization in the backbone. A detailed discussion on traffic engineering and Alcatel’s TE tool can be found in [10].

QoS Performance Monitoring
An operator must be able to verify whether QoS performance guarantees committed in SLAs are in fact being met in the network. This requires an in-service verification of throughput, delay and packet loss characteristics. In order to facilitate a continuous monitoring of the QoS performance offered to customers, the Alcatel 7770 RCP collects statistics on throughput, buffer occupancy, packet loss and delay on a per flow queue basis.

Conclusions
The convergence of other networks on the Internet and the exponential growth of the number of users on the Internet push for the introduction of QoS functionality at the IP layer. IP QoS is a prerequisite for the Internet to become the de facto universal platform for global communication.

In this paper we briefly reviewed the two IP QoS architectures considered by the IETF, namely IntServ and DiffServ. The merits and limits of both are reflected in the trade-off between scalability and level of QoS performance assurance. Both IP QoS architectures are complementary, a characteristic that is reflected in Alcatel’s IP QoS strategy, which positions DiffServ in the backbone, while IntServ can potentially be deployed at the customer premises and in the access part of the network at the scale of corporate or campus networks.

The implementation of DiffServ capabilities in core and high speed edge routers was discussed, including a detailed overview of the DiffServ capabilities of Alcatel’s core switch router. The Alcatel 7770 RCP, a terabit-ready, line-speed DiffServ switch router, supports advanced classification, traffic conditioning, buffer management and packet scheduling functions. They form the basis for an extensive set of supported PHBs. The Alcatel 7770 RCP also supports DiffServ for LSPs. Both E-LSP and L-LSPs are supported.

The implementation of a different forwarding behavior for distinct traffic aggregates in itself is not sufficient to offer any QoS performance guarantees. Distinct handling of traffic aggregates must be combined with a dynamic resource reservation scheme and an admission control function. Alcatel’s bandwidth broker performs this function and can be coupled with traffic engineering to optimize resource utilization in the backbone.
Acronyms

AC admission control
AF assured forwarding
BA behavior aggregate
BBE better best effort
BE best effort
BGP border gateway protocol
CAC connection admission control
CAR committed access rate
CBS committed burst size
CIR committed information rate
CL controlled load
CPE customer premises equipment
CS class selector
DiffServ differentiated services
DSCP differentiated services
EBS excess burst size
EF expedited forwarding
E-LSP EXP-inferred PSC label
EXP experimental bits
FEC forwarding equivalence class
GS guaranteed service
IETF Internet Engineering Task Force
IGP interior gateway protocol
IntServ integrated services
IP Internet protocol
IS-IS intermediate system-to-
intermediate system
ISP Internet service provider
L-LSP label-inferred PSC label
LSP label switched path
MF multi-field
MPLS multiprotocol label switching
MPPS million packets per second
OA ordered aggregate
OSPF open shortest path first
P1 Priority 1
P2 Priority 2
PBS peak burst size
PHB per hop behavior
PIR peak information rate
PSC PHB scheduling class
RED random early detection
RSVP resource reservation protocol
QoS quality of service
SLA service level agreement
SLS service level specification
SRED shock-absorber RED
SrTCM single rate three color marker
TCM three color marker
TCP transmission control protocol
TCS traffic conditioning specification
TDM time division multiplexing
TE traffic engineering
TP transmission port
TrTCM two rate three color marker
TTL time to live
UDP user datagram protocol
VoD video on demand
VoIP voice over IP
WFQ weighted fair queuing
wRED weighted RED

References
