Yoram Bernet, Microsoft
James Binder, 3-Com
Steven Blake, Torrent Networking Technologies
Mark Carlson, Redcape Software
Srinivasan Keshav, Cornell University
Elwyn Davies, Nortel UK
Borje Ohlman, Ericsson
Dinesh Verma, IBM
Zheng Wang, Bell Labs Lucent Technologies
Walter Weiss, Lucent Technologies

A Framework for Differentiated Services
<draft-ietf-diffserv-framework-01.txt>

Status of this Memo

1. Abstract

This document provides a general description of issues related to the definition, configuration and management of services enabled by the differentiated services architecture [DSARCH]. This document should be read along with its companion documents, the differentiated services architecture [DSARCH] and the definition of the DS field [DSHEAD]. A glossary of specialist terms used may be found in [DSARCH].

2. Structure of this Draft

Section 3 defines Differentiated Services and explains the motivation behind its deployment. Section 4 defines the concept of a service and the components that comprise a service. Section 5 discusses several service examples. Section 6 examines intra-domain provisioning, configuration and management issues. Section 7 examines inter-domain provisioning, configuration and management. Section 8 addresses interoperability with Integrated Services and RSVP. Section 9 discusses the interaction of differentiated services with multicast and tunnelling. Section 10 addresses security concerns.

3. Differentiated Services - Motivation and Definition

Traditionally, network service providers (both enterprise and traditional ISPs) provide all customers with the same level of performance (best-effort service). Most service differentiation has been in the pricing structure (individual vs. business rates) or the connectivity type (dial-up access vs. leased line, etc.). However, in recent years, increased usage of the Internet has resulted in scarcity of network capacity, compromising performance of traditional, mission critical applications. At the same time, new applications have emerged which demand much improved service quality. As a result, service providers are finding it necessary to offer their customers alternative levels of service. As well as meeting new customer expectations, this allows service providers to improve their revenues through premium pricing and competitive differentiation of service offerings, which in turn can fund the necessary expansion of the network.

The Differentiated Services architecture offers a framework within which service providers can offer each customer a range of network services which are differentiated on the basis of performance in addition to pricing tiers used in the past. Customers request a specific performance level on a packet by packet basis, by marking the DS field of each packet with a specific value (see [DSHEAD] for more details). This value specifies the Per-hop Behavior (PHB) to be allotted to the packet within the provider’s network. Typically, the customer and provider negotiate a profile (policing profile) describing the rate at which traffic can be submitted at each service level. Packets submitted in excess of this profile may not be allotted the service level requested. A salient feature of differentiated services is its scalability, which allows it to be deployed in very large networks. This scalability is achieved by forcing as much complexity out of the core of the network into edge devices which process lower volumes of traffic and lesser numbers of flows, and offering services for aggregated traffic rather than on a per-micro-flow basis.

4. Services

[DSARCH] defines a Service as "the overall treatment of a defined subset of a customer's traffic within a DS-domain, or end-to-end". Although PHBs are at the heart of the differentiated services architecture, it is the service obtained as a result of marking traffic for a specific PHB, which is of value to the customer. PHBs are merely building blocks for services. Service providers combine PHB implementations with traffic conditioners, provisioning strategies and billing models which enable them to offer services to their customers. Providers and customers negotiate agreements with respect to the services to be provided at each customer/provider boundary. These take the form of Service Level Agreements (SLAs).

Bear in mind when considering the services that are offered in a DS-domain that:

* DS services are all for unidirectional traffic only
* DS services are for traffic aggregates, not individual micro-flows

4.1 Customer/Provider Boundaries

Present day network traffic generally traverses a concatenation of networks which may include hosts, home or office networks, campus or corporate networks and several large transit networks. Home and office networks are typically customers of campus or corporate networks, which are in turn customers of large transit networks. While one would expect the initial deployment of differentiated services to be within large transit networks, its deployment may also be extended to the smaller campus and corporate networks and in special cases, all the way to individual hosts. As such, there may be numerous customer/provider boundaries at which the concept of a 'service' applies.

4.2 SLAs and TCAs

At each differentiated service customer/provider boundary, the service provided is defined in the form of an SLA. The SLA is a contract which specifies the overall features and performance which can be expected by the customer. Because DS services are unidirectional the two directions of flow across the boundary will need to be considered separately. An important subset of the SLA is the traffic conditioning agreement, or TCA. The TCA specifies detailed service parameters for each service level. Such parameters include:

1. Detailed service performance parameters such as expected throughput, drop probability, latency.
2. Traffic profiles which must be adhered to for the requested service to be provided, such as token bucket parameters.

3. Disposition of traffic submitted in excess of the specified profile.
4. Marking services provided.
5. Shaping services provided.

In addition to the details in the TCA, the SLA may specify more general service characteristics such as:
1. Availability/Reliability, which may include behavior in the event of failures resulting in rerouting of traffic
2. Encryption services
3. Routing constraints
4. Authentication mechanisms
5. Mechanisms for monitoring and auditing the service
6. Responsibilities such as location of the equipment and functionality, action if the contract is broken, support capabilities
7. Pricing and billing mechanisms

These additional characteristics are important, and in the case of pricing and billing, fundamental to the service offering but they do not affect the service itself and are not considered further here.

4.3 Service Taxonomy: Quantitative through Qualitative and alternatives

The Differentiated Services architecture can support a broad spectrum of different kinds of service. Categorizing these services provides some constraints on the corresponding SLAs that can be offered for the service.

Some services can be clearly categorized as qualitative or quantitative depending on the type of performance parameters offered. Examples of qualitative services are as follows:
1. Traffic offered at service level A will be delivered with low latency.
2. Traffic offered at service level B will be delivered with low loss.

The assurances offered in examples 1 and 2 are relative and can only be verified by comparison.

Examples of quantitative services are as follows:
3. 90% of in profile traffic delivered at service level C will experience no more than 50 msec latency.
4. 95% of in profile traffic delivered at service level D will be delivered.

Examples 3 and 4 both provide concrete guarantees that could be verified by suitable measurements on the example service irrespective of any other services offered in parallel with it.

There are also services which are not readily categorized as qualitative or quantitative as in the following examples:

5. Traffic offered at service level E will be allotted twice the bandwidth of traffic delivered at service level F.
6. Traffic with drop precedence AF12 has a lower probability of delivery than traffic with drop precedence AF13.

In example 5, the provider is quantifying the relative benefit of submitting traffic at service level E vs. service level F, but the customer cannot expect any particular quantifiable throughput. This can be described as a 'Relative Quantification Service'.

In general, when a provider offers a quantitative service, it will be necessary to specify quantitative policing profiles. In many cases, quantitative policing profiles will be specified even for services that do not offer quantitative performance.

Determining how to monitor and audit the delivery of a qualitative or relative quantification service in such a way as to convince the user that he has received fair measure requires careful attention. It will be up to the customer to determine if the advantage offered is sufficient to make the service worthwhile. The SLA must clearly avoid making quantitative commitments for these services.

4.4 The Scope of a Service

The scope of a service refers to the topological extent over which the service is offered. For example, assume that a provider offers a service to a customer which connects to their network at ingress point A. The service may apply to:

1. all traffic from ingress point A to any egress point
2. all traffic between ingress point A and egress point B
3. all traffic from ingress point A to a set of egress points

Egress points may be in the same domain as the ingress point or may be in other domains which are either directly or indirectly connected to the ingress domain. If the egress point is in another domain, it will be necessary for the ingress provider to negotiate SLAs with the relevant peer domains, which will recursively negotiate with their peers to ensure that the service offered at ingress point A can indeed be extended to the egress points specified. The scope of a service is part of the SLA governing ingress point A.

In general, providers will be able to offer quantitative services most efficiently when a specific set of egress points is specified. Quantitative services which span multiple domains also require tighter coupling between the SLA offered to the customer at ingress point A and the SLAs negotiated with intermediate domains. Qualitative services can more readily be offered to arbitrary sets of egress points and require looser coupling between the SLA at ingress point A and SLAs at intermediate domain boundaries.

4.4.1 Services Governing Received Traffic

A special case of service scope is a service that governs all traffic between any ingress point and egress point B. The SLA that defines this service would be at egress point B and would effectively allow the customer to control the mix of traffic received from the provider. While such a service is theoretically possible, it conflicts with the more traditional use of diffserv which governs the quality with which traffic is sent, rather than received.

A number of concerns would have to be addressed by such a service, including:

Traffic going to point B from an ingress point A under the terms of the SLA of this service may also be governed by an SLA for traffic submitted at point A. The SLAs may conflict and it will not, in general, be possible to resolve all such conflicts across all the ingress points.

- Establishing a traffic profile for this service at every possible ingress which prevents overload of the receiver can be more complex than for other service scopes: Static profiles are likely to be either inefficient (e.g. dividing the egress profile into fixed proportions) or risky (e.g. allowing every ingress to send the whole profile) whilst dynamic profiles require processes and communication mechanisms to coordinate the settings.
- Without effective ingress profiles for the service, denial of service attacks will be a serious problem.

Some of the characteristics of receiver oriented services can be provided by local policies and the SLA for the domain to which traffic is sent via the egress point as described in Sec. 4.6.4.

4.5 Dynamic vs. Static SLAs

SLAs may be static or dynamic. Static SLAs are the norm at the present time. These are instantiated as a result of negotiation between human agents representing provider and customer. A static SLA is first instantiated at the agreed upon service start date and may periodically be renegotiated (on the order of days or weeks or months). The SLA may specify that service levels change at certain times of day or certain days of the week, but the agreement itself remains static.

Dynamic SLAs, on the other hand, may change frequently. Such changes may result for example, from variations in offered traffic load relative to preset thresholds or from changes in pricing offered by the provider as the traffic load fluctuates. Dynamic SLAs change without human intervention and thus require an automated agent and protocol, in effect, a bandwidth broker to represent the differentiated service provider's domain (such as suggested in [BB]).

Dynamic SLAs also present challenging problems to both end users and network providers:

- Network providers have to balance frequently changing loads on different routes within the provider network. This requires the provider to adopt dynamic, automated resource provisioning mechanisms rather than relying on static provisioning.
- Customer equipment will have to adapt to dynamic SLAs in order to make the most out of the changing SLA.
- End user applications may have to adapt their behavior during a session to make the most of (or even, cope with) dynamic SLAs.

It is worth reiterating that the SLAs in Differentiated Services apply to aggregates of traffic and not individual flows. For scalability, it is undesirable to envisage modifying an SLA every time a new micro-flow is added or removed from an aggregate.

4.6 Provisioning Traffic Conditioners in Boundary Devices to Provide Services

Once an SLA has been negotiated, the service provider (and optionally the customer) will configure traffic conditioning components at the boundary between the two networks. The service provider does so with the goal of protecting the provider's network such that the resources granted to the customer meet but do not exceed the terms of the TCA. The customer does so with the goal of making the best use of the service purchased from the provider. In this section, we will briefly describe configuration of traffic conditioners in boundary devices.

Note that the provider's self interests require only that the provider identify
- for which service level specific traffic is submitted,
- by which customer it is submitted, and
- for traffic with double-ended SLAs (i.e. SLA scope is type 2 or 3 of Sec. 4.4) only, the destination address to which the traffic is directed.

Customer traffic may be authenticated either by the physical connection on which it arrives or by some sophisticated cryptographic means which is beyond the scope of this draft. The provider need not be concerned with the customer's individual micro-flows in delivering basic Differentiated Services (see Sec. 4.6.3 for additional services).

[DSARCH] identifies four traffic conditioning components:
1. Meters
2. Markers
3. Shapers
4. Droppers

The combination and interaction of the traffic conditioning components is selected on a packet-by-packet basis by the DS codepoint. The configuration parameters for the components at each codepoint are determined by the policies and profiles applied, so that the conditioner polices the traffic in the BA specified by the codepoint. Meters measure submitted traffic for conformance to a profile, providing control input for the other components which implement the policing:
- Shapers police by delaying submitted traffic such that it does not exceed the traffic rate specified in a profile.
- Droppers police by dropping traffic that is submitted at a rate exceeding that specified in a profile.
- Markers police by re-marking the traffic with a different codepoint either
  - to demote out-of-profile traffic to a different PHB,
  - as a result of an SLA which specifies codepoint mutation, or
  - to ensure that only valid codepoints are used within the domain.

In addition to these four components, traffic classifiers are required in order to separate submitted traffic into different classes. Classifiers may separate traffic based only on the DS-field of submitted packets (BA classifiers) or may do so based on multiple fields within the packet header and even the packet payload (MF classifiers). MF classifiers may be used at boundaries to provide certain per-micro-flow services to customers. Examples of such services include per-flow marking or shaping. Typically, traffic will arrive at the boundary of a DS domain pre-marked and pre-shaped. However, at interfaces with some non-DS customer

networks, it is possible that traffic will require marking and shaping.

Even if a customer has pre-marked and pre-shaped, the service provider will wish to police the traffic at the ingress boundary to meet the domain's self-interests. This may result in traffic being re-marked or dropped.

Traffic conditioning components (in particular, meters) will also be the primary source of billing information for a differentiated Services network.

4.6.1 Minimal Functionality at Provider's Ingress

At the very least, the service provider must limit traffic carried on behalf of the customer to the constraints specified in the TCA. A simplified TCA can be represented in the form of a table wherein each row has the format:

DS-Mark : Profile : Disposition of non-conforming traffic

This row indicates that the provider commits to carry traffic marked with 'DS-Mark' at the corresponding service level, provided that it conforms to the 'Profile'. Traffic that is submitted with 'DS-Mark' and which does not conform to the 'Profile', is subjected to 'Disposition of non-conforming traffic'. This is generally a policing action such as re-marking to a lower service level, delaying in a shaper, or dropping. Alternatively, it may be carried at the requested service level, but subjected to a surcharge. The SLA for this type of service would normally be expected to be of type 1 of Sec. 4.4.1, where the traffic destination can take it through any egress point of the domain.

To provide this minimal functionality, the provider must configure a BA classifier to separate traffic into the different service level requested, based on DS-Mark. Following the BA classifier, each class must be metered for conformance to the corresponding profile. Following the profiler, either a dropper, shaper or re-marker is likely to be employed. The Better than Best Efforts service described in Sec. 5.1 is an example of a service for which this functionality is sufficient.

4.6.2 Functionality at Provider's Ingress for Double-ended SLAs

If quantitative or other services needing double-ended SLAs (types 2 and 3 of Sec. 4.4.1) are implemented in a DS Domain, these services specify the possible egress port(s) for traffic conforming to the SLA. The traffic conditioner needs to consider the destination address of the packet as additional input to the policing process, so that traffic is not accepted for egress ports for which an SLA does not exist. The Virtual Leased Line service described in

Sec. 5.2 is an example of a service that would require this functionality.

A QoS VPN can be constructed by provisioning multiple instances

o

f

services of type 2, building in effect, a mesh of point to point QoS links.

Services of type 3 are most likely to be used for multicast applications (see Sec. 9).

4.6.3 Added Value Functionality at Provider's Ingress

The functionality described in Secs. 4.6.1 and 4.6.2 serves only to protect the provider's network resources in line with the terms of the TCA. It provides no assistance to the customer. The burden of marking packets and shaping traffic falls entirely on the customer. In some cases, the SLA may call for the provider to provide additional services to the customer. Such services may include:

1. Marking traffic from specific micro-flows to a specific behaviour aggregate (marking the DS-field).
2. Policing traffic from specific micro-flows or sets of micro-flows, either in the form of dropping or shaping.

In order to provide such services, the provider must generally employ an MF classifier in addition to the BA classifier. The need for an MF classifier arises only when the customer requires the provider to provide some form of traffic separation or authentication on behalf of the customer. The provider may charge dearly for these services depending on the degree of granularity and the amount of work required. For example, shaping thousands of customer micro-flows might consume considerable resources in the provider's edge device. On the other hand marking based on source subnet addresses would consume considerably fewer resources.

4.6.4 Functionality at Customer's Egress

Strictly speaking, the customer need not apply any specific traffic conditioning. In this case, the customer relies on the provider to mark as per negotiated MF classification criteria. In many cases it is preferable for the customer to mark. Customer marking may be necessary when customer packets are encrypted (as in the case of end-to-end IPSec). Customer marking enables the customer to direct specific traffic from specific users or applications to specific service classes. This may be difficult or impossible for a provider to do on behalf of a customer when, for example, applications use volatile ports and users are assigned IP addresses based on DHCP.

In addition to marking, it is in the customer's best interest to at least shape per service level, at the customer network's egress point. Otherwise, customer traffic may be policed by the service

provider with undesirable consequences (e.g. dropped packets). Shaping per service level does not however, provide for micro-flow traffic separation. As a consequence, a renegade traffic source may cause the profile to be exceeded for a specific service level, negatively impacting all customer flows which are marked for that service level. Therefore, it is often in the customer's interest to shape or at least to police, with micro-flow granularity.

4.6.5 Functionality at Provider's Egress

At the egress from a provider's domain there may be an SLA in place with a peer DS domain, which might be either another provider or an end user domain. As in Sec. 4.6.4, it is in the provider's best interests to shape the traffic leaving the domain.

Depending on the SLA, the egress may be required to remark and/or police or shape the traffic. Note that the forwarding treatment applied to the packet in the egress node of the domain would be that selected by the codepoint before it was remarked (otherwise, the egress node has to support multiple codepoint to PHB mappings).

The provider may also wish to offer additional services to a customer by policing egress traffic with micro-flow granularity if the customer might expect to receive excessive traffic in a single BA and wishes to apply greater control than could be achieved by normal policing of the aggregate. This would be specified via an SLA in the usual way.

4.7 Internal Provisioning

The provider must provision internal nodes in the provider network to meet the assurances offered by SLAs negotiated at the boundaries of the network. To do so, the provider may use similar traffic conditioning mechanisms to those used at the network boundaries. However, providers are unlikely to apply MF classification within the interior of the network. The provider may police periodically within the network, by reshaping, remarking or discarding traffic. Service providers are experienced in provisioning large networks which offer uniform service, assisted by predictive tools, traffic modeling tools and real time measurements. Current techniques will likely be applied to differentiated services networks, although, the complexity of provisioning will increase significantly. In a differentiated service network, the provider must ensure that resources granted to traffic of one service level does not inappropriately compromise assurances regarding traffic at other service levels (for example, in example service 6, traffic in AF13 can legitimately compromise traffic in AF11 if an increase in AF13 traffic causes more AF11 traffic to be dropped). As mentioned previously, internal provisioning in the case of dynamic SLAs will likely require dynamic resource allocation protocols.

4.8 End-to-End Service Construction

The Differentiated Services architecture proposes that an end-to-end service can be constructed by the concatenation of domain services and their associated customer-provider SLAs for each of the domains which the service traffic has to cross.

Clearly, not all PHBs and services can be meaningfully concatenated, and the definition of suitable services and their associated PHBs will be a major focus of future Differentiated Services development. This is discussed at greater length in Sect. 7.

5. Service Examples

In this section, we describe service examples and show how they can be supported by specific PHBs. We base these examples on PHBs which are defined in [AF]and [EF]. These examples are intended to be illustrative of the wide range of services that can be employed using the differentiated services model, and are not intended to be an exhaustive list.

5.1 Better than Best-Effort (BBE) Service

This is a qualitative service which promises to carry specific web server traffic at a higher priority than competing best-effort traffic. Such a service offers relatively loose (not quantifiable) performance from a given ingress point to any egress point. Such a service is suitable for example for businesses offering access to web based content. The BBE service enables the web content provider to provide content at a generally higher rate than other content providers are able to, in so reducing the latency experienced by consumers of the web site.

5.1.1 Service Implementation

In this example, we assume that there is an SLA which defines the service at the customer’s ingress point. This is the point at which the customer injects web server responses into the differentiated services network. The information in the TCA can be represented in the following form [AF]:

AF13 Mark : 1 Mbps : Any egress point : Excess traffic handled by marking with AF11 mark.

Packets submitted for the BBE service should be marked with the DS-field codepoint corresponding to the AF13 PHB. The provider is promising to carry up to 1 Mbps of traffic from the ingress point to any egress point at a higher priority than best-effort traffic. A lesser class of service corresponding to the AF11 PHB will be applied to traffic submitted for the AF13 PHB, in excess of 1 Mbps.

The provider will provision a policer at the ingress point. Traffic submitted up to the 1 Mbps limit will be directed to the AF13 PHB. Traffic submitted in excess of 1 Mbps will be remarked for the AF11 PHB. Note that the scheme will preserve ordering of packets since AF13 and AF11 use a single queue..

In order to provide this service, the provider will have to implement the AF13 and AF11 PHBs in core network equipment. The AF13 and AF11 PHBs can be implemented for example, using a single RIO queue. The provider will also have to provision equipment within the core of the provider's network to provide the AF13/AF11 service. By provisioning the appropriate RED parameters, for example, the provider is able to control the priority of AF13 traffic relative to AF11 traffic at each network node. Since there are no quantitative guarantees, the provider can be quite liberal in its provisioning strategy and may realize significant statistical multiplexing gains. Also, the absence of quantitative guarantees makes it easy to provide this type of service across multiple DS provider domains. This is because is not necessary to negotiate, then provision and enforce quantitative guarantees at multiple boundaries.

5.2 Leased Line Emulation Service

This is a quantitative service which emulates traditional leased line service. As such, it promises to deliver customer traffic with very low latency and very low drop probability, up to a negotiated rate. Above this rate, traffic is dropped. Such a service is typically offered between two specific points. It is suitable for many customer applications. However, due to the high quality guarantees, it is likely to be priced higher than alternate services and therefore, to be used only for applications which really require this type of service. An example of such an application is IP telephony. A corporate customer might purchase leased line emulation service between each pair of a number of corporate network sites.

5.2.1 Service Implementation

In this example, we consider a customer with three geographically dispersed networks interconnected via a single provider network. Customer attachment points are represented as A, B and C. At each attachment point, an SLA describes the leased line service to be provided to the other two points. The table below represents the information required in the TCA at attachment point A:

EF-Mark : 100 Kbps : Egress point B : Discard non-conforming traffic

EF-Mark : 50 Kbps : Egress point C : Discard non-conforming traffic

Packets submitted for leased line service should be marked with the DS-field codepoint corresponding to the EF PHB [EF]. From the ingress point A, to the egress point B, the provider is promising to

carry up to 100 Kbps of traffic. Excess traffic will be discarded. From ingress point A, to egress point C, the provider promises to carry 50 Kbps of traffic. Of course, there is some tolerance required in policing the traffic and thus, there may be a specification of tolerated jitter or burst size. However, for a leased line service, the primary traffic profile parameter would be the sustained traffic rate.

The provider will provision a policer at ingress point A to limit traffic destined for egress point B, to 100 Kbps. Similarly, a policer will be configured to limit traffic destined for egress point C, to 50 Kbps. These policers will require classification based on the DS-Mark and the destination address in each packet.

In order to provide this service, the provider will have to implement the EF PHB in core network equipment. The EF PHB can be implemented using strict priority queuing or alternatively, by assigning EF marked packets to a heavily weighted queue in a WFQ scheme. The provider will have to provision equipment within the core of the provider's network. For example, routers carrying traffic between point A and points B and/or C will have to be provisioned considering the resources committed by the TCA at point A. This means that a router which is both in the path from A to B and from A to C, will have to be considered to have committed 150 Kbps of bandwidth as a result of the TCA in place at A. A router that is only in the path from A to B, will have to be considered to have committed 100 Kbps as a result of this TCA, and so on. Of course, routing is subject to change and so, failover paths may have to be provisioned as well. These may be provisioned to provide some fraction of the service in the case of failover or alternatively, the SLA at point A might reflect appropriate service availability parameters. To enhance the assurances offered by EF service, providers may employ route pinning mechanisms or QoS routing mechanisms.

5.3 Quantitative Assured Media Playback Service

This service offers looser assurances than the leased line service described above, but is still considered a quantitative service. In particular, it promises to deliver traffic with a high degree of reliability and with variable but bounded latency, up to a negotiated rate. Above this rate, traffic is subject to significant delay or drop. Such a service is typically offered between a specific set of points. It is suitable for many customer applications. It would likely be priced lower than a leased line service, due to the latency variability. However, due to the latency bound and high degree of delivery, it is likely to be priced higher than alternate services. This service is particularly suitable for video or audio playback, in which considerable bandwidth is required on a continual basis, but the non-interactive nature of the traffic makes it somewhat delay tolerant.

5.3.1 Service Implementation

In this example, we again consider a customer with three geographically dispersed networks interconnected via a single provider network. The table below represents the information required in the TCA at attachment point A:

AF13-Mark : 100 Kbps sustained, 100 Kb bursts tolerated at up to 200 Kbps : Egress point B : Excess burst traffic over sustained rate marked with AF12-mark : Non-conforming traffic marked with AF11-mark : Max latency = 1 second

AF13-Mark : 50 Kbps sustained, 100 Kb bursts tolerated at up to 100 Kbps : Egress point C : Excess burst traffic over sustained rate marked with AF12-mark : Non-conforming traffic marked with AF11-mark : Max latency = 2 seconds

Packets submitted for the assured playback service should be marked with the DS-field codepoint corresponding to the AF13 PHB. From the ingress point A, to the egress point B, the provider is promising to carry up to 100 Kbps of sustained traffic with bursts of 100 Kb in size at a peak rate of 200 Kbps. Excess burst traffic will be marked with the codepoint for AF12 and out of profile traffic will be carried but with the AF13 codepoint. So long as these conditions are met, latency will be limited to 1 second. Note that for this service, the traffic profile is described using a full set of token bucket parameters. Since the latency bounds for such a service are less strict than those required for the leased line service, a certain degree of traffic burstiness can be tolerated.

The provider must support the AF11, AF12 and AF13 PHBs in core network routers. These PHBs might be provided, for example, by assigning AF11, AF12 and AF13 marked traffic to a single RIO queue with high drop thresholds. The policers at the edge would limit competing traffic in line with the TCA, in order to assure that the latency bounds can be met. In addition, the service provider will have to provision devices in the core of the network. The provisioning considerations discussed in the context of the leased line service apply here as well, however, in general, the service provider has the liberty of being less conservative in provisioning and realizing better statistical gains.

5.4 Superposition of Quantitative and Qualitative Services in the Same Network

A compelling model would provide both quantitative and qualitative services in the same differentiated service network(s) as follows. A number of corporate campus networks would be interconnected by a differentiated service network providing quantitative services between the sites. For example, a mesh of leased line services would

enable IP telephony between the sites. A mesh of media playback service using the AF11 PHB would enable audio/video playback between the sites. In addition, each corporate site would be allotted some level of BBE service to arbitrary destinations. In this model, the differentiated service network is effectively providing a mesh of quantitative services between fixed locations (similar to a VPN). This mesh is superimposed on a cloud supporting BBE service.

6. Provisioning and Configuration

The provision of differentiated services requires careful network wide provisioning and configuration. Provisioning refers to the determination and allocation of the resources needed at various points in the network. Provisioning may dictate the addition or removal of physical resources at various points (physical provisioning). Provisioning may also dictate the modification of operating parameters within existing physical network equipment to alter the relative share of the equipment's resources which are allotted to one or another class of traffic (logical provisioning). Configuration refers to the distribution of the appropriate operating parameters to network equipment to realize the provisioning objectives.

In Secs. 4.6 and 4.7, we briefly discussed provisioning and configuration requirements both at the network boundaries and in the network interior. In this section we will focus primarily on the coordination of provisioning and configuration throughout the network, such that end-to-end services can be provided reliably. We will discuss the roles of protocols such as SNMP, CLI, RSVP, COPS and LDAP in the provisioning process.

6.1 Boundary vs. Interior Provisioning and Configuration

For the sake of brevity, consider the term 'provisioning' to refer both to provisioning and configuration, except where otherwise noted. It is helpful to consider provisioning at the network boundaries, separately from provisioning of the interior. Since the differentiated service provider is selling a contract (SLA) at the network boundary, we can consider the boundary provisioning which supports SLAs, to drive the interior provisioning. The two are not entirely separable in that each affects the other. For example, a network operator cannot offer an SLA which cannot be met by the resources available in the interior of the network. In general, the overall provisioning process iterates between the boundaries and the interior. From here on we will refer to provisioning with respect to the TCA rather than the SLA, since the TCA is the component of the SLA that defines detailed traffic handling parameters.

6.1.1 Boundary Provisioning

Boundary provisioning was considered briefly in Sec. 2.6. We discussed the minimal provisioning that a provider must implement to enforce a TCA. We also discussed additional configuration which a provider may use to provide additional (especially per-flow) services to a customer. The latter is not actually related to the provisioning of resources within the differentiated services network, but rather assists the customer by determining which subsets of the customer's traffic make use of the resources provisioned within the differentiated services network. As such, it is out of the scope of this section. Here, we consider only the minimal provisioning required at the boundary.

At a minimum, the provider must assure that sufficient physical resources are provisioned at the boundary to meet the requirements of the TCA. For example, if the sum of the profiles supported at a particular ingress point would allow 10 Mbps of traffic to be supported, it is unacceptable to provision a T-1 access link. A T-3 however, would be sufficient. Once the physical provisioning is implemented, it is necessary to apply the appropriate logical provisioning. This is achieved by configuring policers which limit the amount of traffic accepted from the T-3 access link, at each service level. It may also be necessary to set up the amount of buffering available for the queues used for the service. Similar provisioning is also appropriate at each egress point if the aggregate of profiles provisioned to the egress exceeds the capacity of the output link.

6.1.2 Interior Provisioning

For the purpose of provisioning the interior of the network, it is desirable to understand or to control the volume of traffic of each class which traverses each network node. The greater this understanding, the more efficiently the network can be provisioned while still meeting the requirements of the TCAs. It is feasible to understand the volume of traffic traversing each node if this traffic is admitted according to a TCA which dictates egress point as well as ingress point. (This case generally applies to quantitative services and was discussed in the context of the EF PHB and the leased line service in Sec. 3.2.1). While traffic volumes cannot be anticipated with 100% accuracy, it is possible to approximate them quite well, especially with the help of route pinning mechanisms. It is therefore possible to provision the network reasonably accurately for traffic submitted for quantitative services. Although such provisioning may be quite difficult in a large network, it is nonetheless a tractable problem. We will refer to this component of the provisioning problem as quantitative provisioning.

On the other hand, many (if not most) of the services offered by differentiated service networks will not specify egress points (as is the case for qualitative services) and will not restrict submitted traffic to specific egress points, let alone specific routes. Thus, interior nodes will have to be provisioned without an a-priori understanding of the volume of traffic submitted for qualitative services which will arrive at each node. It is necessary to be able to provision differentiated service networks to support both quantitative services with specific egress points as well as qualitative services, which do not have specific egress points on the same physical resources. To this end, it is necessary to isolate the impact of qualitative traffic on the resources reserved for quantitative traffic. This can only be achieved if the former is treated with lower priority than the latter. Thus, in general, resources will have to be provisioned first for quantitative traffic, using quantitative provisioning mechanisms. Then, qualitative provisioning can be used to allocate remaining resources to qualitative traffic. Qualitative provisioning can also be applied to services which offer a relative quantification of traffic volumes.

The impact of the two types of traffic will have to be isolated by ensuring that they do not share PHB codepoints. PHBs used for quantitative services will always have higher priority access to resources than those used for qualitative services. As a result, it is necessary to carefully police traffic submitted for quantitative PHBs. Failure to do so can result in the starvation of lower priority traffic. In general it can be expected that only a small fraction of the resources at each node will be provisioned for quantitative traffic.

Similarly, a significant fraction of the traffic capacity should remain for best-efforts service to provide a ’soft’ target for traffic dropping if congestion occurs or it is necessary to redirect non-best efforts traffic in the event of failure.

6.1.2.1 Quantitative Provisioning of the Interior

As discussed previously, quantitative provisioning is a difficult but tractable problem. With knowledge of the network routing topology and the TCAs at the boundaries, it is possible to compute the resources required at each interior node to carry the quantitative traffic offered at the edges. Based on the results of this computation, interior nodes must be provisioned and configured with sufficient capacity to accommodate the quantitative traffic which will arrive at the node, while leaving sufficient capacity remaining to accommodate some amount of qualitative traffic.

The provisioning mechanism described assumes a top-down approach, in which the network administrator studies the network topology and traffic routing and computes the provisioning requirements. An

alternative approach uses signalling to automate the process [MPLS]. For example, RSVP messages could be launched along the paths that will be followed by submitted quantitative traffic. If a TCA calls for 100 Kbps of leased-line service from ingress point A to egress point B, an RSVP message could be transmitted from point A towards point B, with a flowspec specifying 100 Kbps. This message would traverse each node at which resources would have to be committed. Conventional RSVP routers would install a reservation. In a differentiated service network, RSVP could be adapted to provision the resources required per the differentiated services model. In a network which offers a number of static TCAs, such RSVP messages could be launched from the TCA ingress point at the time the TCA is initially instantiated with the effect of instantiating the appropriate cumulative provisioning in routers along the various routes. The advantage of this approach is that it does not require explicit knowledge of the network topology. We will revisit these two approaches to quantitative provisioning of the interior in a later section.

Once the resources required for quantitative traffic at each node have been determined, provisioning of the node consists of installing or configuring interfaces of the appropriate capacity to easily accommodate the quantitative traffic that will traverse the node. Note that we do not state the precise meaning of 'to easily accommodate'. A number of factors must be considered when determining the appropriate capacity, given a certain volume of predicted quantitative traffic. These include:

1. Margin of error
2. Statistical gain desired
3. Capacity remaining for qualitative (including best efforts) traffic

The first, margin of error, accommodates mistakes in computation, effects of transient route changes which are not otherwise accounted for, effects of traffic clustering as it moves through the network and so on. The statistical gain desired refers to the degree to which a provider is willing to gamble that not all sources of quantitative traffic will be simultaneously active at the limit dictated by the TCAs at the ingress points (vs. the penalty the provider would be willing to pay in terms of refunded charges or lost customers). Finally, the provider must determine how much capacity will be reserved for qualitative traffic at each node. Thus, if it is determined that 1 Mbps of quantitative traffic might traverse a specific node in a specific direction, the provider might install a 10 Mbps interface in the node, to serve the corresponding traffic direction. This would leave 9 Mbps of capacity quite safely for qualitative traffic. In this case, the provider would be assuming that statistical gains which might be realized will be used to offset the margin of error which would compromise the resources available.

In addition to installing or configuring the appropriate capacity at each interface, it may be desirable to configure policers to assure that the resources actually consumed by the higher priority quantitative traffic do not exceed expectations. This is especially important if the provider is attempting to achieve a high degree of statistical gain or has not allowed for a reasonable margin of error. Policers need not be configured at each interior node, but should probably be configured at certain key nodes. It may also be necessary to configure the internal resources of the router (queues and buffers) to deliver the services required.

6.1.2.2 Qualitative Provisioning of the Interior

As explained previously, it is necessary first to determine the resources which must be provisioned at each node for quantitative traffic. Once these have been determined, interfaces must be installed or provisioned to accommodate the required resources while leaving sufficient capacity for qualitative traffic. In order to do so, it is necessary to determine the resources required at the node for qualitative traffic. Since qualitative traffic cannot be assumed to follow specific routes with the same degree of predictability as quantitative traffic, this provisioning problem is far more difficult and provisioning parameters must be estimated based on heuristics, experience and possibly on real time measurement.

Once physical interfaces have been selected to accommodate the resources required by the computed quantitative traffic load and the estimated qualitative traffic load, additional configuration is required to support qualitative traffic. Such configuration amounts to the selection of relative weights for queues for different service levels (in a WFQ scheme), or the selection of RIO or RED thresholds or alternate logical resource provisioning parameters. It is assumed that if quantitative traffic is accommodated via similar queuing mechanisms (as opposed to strict priority queuing), that the weighting parameters chosen for quantitative traffic isolate it effectively from the effects of qualitative traffic. However, the configuration parameters which differentiate the various qualitative services may not provide such a degree of isolation among the qualitative services. Thus, it may be necessary to attempt to estimate the relative traffic arriving for each qualitative service and to anticipate the interaction between traffic of different qualitative services. It may be impossible to both efficiently and conservatively provision a network for certain combinations of qualitative services. To aid in the provisioning of a network for qualitative services, it may be useful to configure policers to control the volume of traffic arriving at a given node. However, such policing might have to be restricted to shaping (rather than discarding) in order to avoid violating TCAs in place at the network boundaries.

6.2. Static vs. Dynamic Provisioning

So far, we have considered static provisioning techniques. Even the example of RSVP usage for provisioning assumed that the RSVP messages were launched at the time a TCA was instantiated as opposed to dynamically. In the case that TCAs are static, static provisioning is adequate for quantitative traffic. However, since qualitative traffic [e.b.1] offers less predictable patterns, it is likely to cause traffic volumes at different nodes in the network to change dynamically, even when the TCA is static. For this reason, dynamic provisioning techniques are desirable and may assist the service provider in making better use of network resources. In addition, dynamic provisioning may enable the service provider to provision more liberally for quantitative services, realizing statistical gains. If we consider further, that it may be desirable to provide dynamically changing TCAs, then the appeal of dynamic provisioning techniques is even stronger.

Dynamic provisioning may be signalling based, measurement based or both. For example, a conventional RSVP router supports signalling based dynamic provisioning. Hosts signal the router to request more or less resources and the router adjusts accordingly. The host may or may not actually submit traffic at the rate at which it signalled it would, but regardless, the resources are committed in case it does. Measurement based provisioning would adjust the resources committed in response to the traffic loads actually measured at the device. While differentiated services does not specify any form of signalled or measurement based provisioning, both may be useful.

6.3 Distributing Configuration Information

The process of physical provisioning is by necessity relatively static and cannot be automated since it requires installation of physical equipment. However, logical provisioning and configuration can and should be automated to the degree possible. In this section, we look at techniques for distributing configuration information.

6.3.1 Top Down Distribution of Configuration Information

In the simplest case, TCAs are static and both the boundaries and interior of the network are provisioned statically by ’pushing’ configuration information down to the appropriate network nodes. Configuration of boundary nodes requires primarily the pushing of policing information to enforce the TCAs in place. (Additional fine grain information may be pushed to provide traffic separation services on behalf of the customer, but these are not addressed in this context). Configuration information for boundary nodes is determined at the time the TCA is negotiated. At this time, the nodes are configured by the provider. The network administrator may use one of several protocols to do so, including for example SNMP or CLI.

In order to accommodate the traffic submitted by the provisioning of new TCAs, it is necessary to provision the interior of the network. As discussed previously, it is possible to compute the resources required for quantitative traffic. Assuming that sufficient physical capacity has been provisioned, configuration amounts to logically provisioning sufficient capacity at each interior interface and to configuring policers for the quantitative traffic at various interior nodes. In addition, qualitative provisioning requires the configuration of queues, WFQ weights and/or RIO parameters at various interior nodes, and may also include the configuration of some number of policers. In the case, of static, top down configuration, interior configuration information is also pushed down via a configuration protocol such as SNMP or CLI.

The difficulty of such top down provisioning is that it requires the network administrator to coordinate the provisioning of each network node, at boundaries as well as in the interior, such that the network is provisioned end-to-end in a consistent manner and is able to efficiently deliver the services promised by the TCAs. In order to assist the network administrator in this task, it is useful to consider a database which holds both current topology information as well as the current TCAs instantiated at the network boundaries. This information is stored in a format dictated by a standard schema as suggested in [Elleson].

Of course, the database is ideally maintained in a way which is logically centralized (for ease of programming and modifying) but is physically distributed (for the sake of robustness and fault tolerance). Policy servers may be used to extract information from the database and to convert it to configuration information which is pushed down to individual nodes. In this scenario, policy servers would likely use a directory access protocol such as LDAP to retrieve information from the directory and would use a configuration protocol such as SNMP or CLI to push the configuration information down to the network nodes. Note that in this example, the policy servers and the directory schemas are in effect fulfilling the role of bandwidth broker [BB]. In particular, the policy servers use an awareness of the network topology to provision interior nodes such that certain end-to-end QoS routes can be constructed and assurances implied by the TCAs at the boundaries can be delivered.

6.3.2 Distribution of Configuration Information via Signalling

An alternate mechanism of distributing configuration information is via signalling messages transmitted between boundary nodes of the same differentiated service domain (intra-domain signalling). It is also interesting to consider inter-domain signalling, but this will be addressed separately. An example of such signalling was described previously, in the usage of a modified form of RSVP. Such signalling

is particularly useful for the purpose of installing configuration information for quantitative services which affect specific paths and is somewhat less useful (though not useless) for the purpose of configuring qualitative services. It is likely that such a signalling approach would be used in conjunction with top down provisioning. For example, the directory schema might dictate the amount of resources to be available for high priority quantitative services at each node. These limits might be pushed down to individual nodes a- priori. Signalling from the network boundaries, at TCA instantiation time, would then be used to claim resources from the pool of quantitative resources available at each node. Alternatively, nodes might consult policy servers as the signalling resource requests arrive at each node. The latter model is similar to the use of per- flow RSVP signalling and PEP/PDP policy usage in traditional RSVP networks. Qualitative configuration information would still be pushed in a top down manner. The advantage of the latter model is that policy servers would be dynamically updated with information regarding the current usage of network resources. In this model, it is likely that a variant of COPS would be used to communicate between network nodes and the policy servers. Note that COPS may be used for distribution of top down configuration information as well, though it is not specifically designed for this purpose.

One of the advantages of configuration via signalling, is that it facilitates the support of dynamic TCAs. TCAs could be dynamically renegotiated using inter-domain signalling. Such renegotiation would require dynamically modifying the provisioning within the affected domain, a process which requires some automated signalling protocol such as an aggregated form of RSVP signalling between boundary nodes in a provider’s domain. This protocol would in effect, represent a distributed bandwidth broker [BB] for the domain.

6.3.3 Modification of Configuration Information Based on Real-Time Measurement

A third mechanism for the configuration of interior nodes would be based on measurement of current traffic loads at key network nodes. Measurement based configuration is less necessary for quantitative provisioning, since quantitative traffic patterns are relatively predictable. However, it can significantly enhance the efficiency with which qualitative provisioning can be achieved. For example, network nodes may feed policy servers with current qualitative traffic load measurements. In response, bandwidth brokers and policy servers might recompute the relative weights for different service queues in a WFQ node and push the new configuration information to the routers. It is likely that measurement based configuration for qualitative services would be used in conjunction with signalling based configuration for quantitative services.

7. Inter-Domain Considerations and End-to-End Services

So far we have considered differentiated service primarily in the context of a single DS domain providing service to a single customer. The ultimate customers of the differentiated service network are hosts and end users residing on peripheral stub networks. In general, these are interconnected by multiple domains and require service which spans these domains. Therefore, it is important to consider the interaction of services provided by a concatenation of differentiated service domains and the peripheral stub networks, rather than the service provided by a single domain. In this section, we discus inter-domain issues related to TCAs, provisioning and service and PHB mapping.

7.1 TCAs

Each service provider is expected to negotiate bilateral agreements at each boundary node at which it connects to an adjacent provider's network. Such agreements are captured in the form of two TCAs, one governing the services provided to provider A's traffic by provider B and the other governing the services provided to provider B's traffic by provider A. Note that provider A serves as a provider to provider B with respect to traffic flowing from provider B to provider A. On the other hand provider A is a customer of provider B with respect to traffic flowing from provider A to provider B. The two TCAs can be considered separately.

In general, the TCAs negotiated by a provider at any boundary will be dictated by TCAs negotiated at other boundaries. For example, assume that provider A offers leased line service to a customer with an ingress point in provider A's domain, but an egress point in provider B's domain. In this case, it is necessary that the TCA between provider A and provider B be sufficient to accommodate the assurance made by provider A to its leased line service customer. Provider A may serve a number of customers with leased line services terminating at various boundary points in provider B's network. Thus, the TCA between provider A and provider B must represent the aggregate requirements of the TCAs of all of provider A's customers.

7.2 Inter-Domain Provisioning

The inter-domain provisioning problem is not unlike the intra-domain provisioning problem. The provider would generally begin by evaluating the TCAs it has negotiated with its customers, and then computing the impact of each of these TCAs on the TCAs it has negotiated with its providers. For quantitative services, the provider can compute the quantitative requirements of TCAs at each of its provider's boundary nodes, as described above in the context of the leased line service. For qualitative services, the process of determining the requirements from its providers is fuzzier, since

the volume of qualitative traffic expected to be carried through any boundary is less deterministic.

In the simplest case, provisioning is based on static TCAs. In this case, provisioning is an iterative process in which providers negotiate TCAs with each of their customers, then apply the appropriate internal provisioning techniques to meet these requirements. In the process of internal provisioning, a provider might determine that a particular TCA cannot be met due to internal resource constraints. The provider would then either have to add internal resources or renegotiate one or more customer TCAs. Although the process may be somewhat iterative, it is relatively static in that changes in boundary TCAs and internal provisioning occur relatively infrequently (on the order of hours, days or months) and require human intervention.

Internal provisioning to meet the requirements of TCAs relies on provisioning techniques described previously. As TCAs are negotiated, the provider must check that the existing internal provisioning is sufficient to meet the requirements of the new TCA, or must alter the internal provisioning. Recall that internal provisioning might be pushed in a top down manner, from a domain's logically centralized point of administration, or alternatively might be distributed from the edges via signalling. In the former case, some form of a bandwidth broker would be directly consulted or notified regarding changes in TCAs negotiated at the domain boundaries. In the case that signalling is used, provisioning messages (such as described previously) would be launched from the boundary at which the new TCA is negotiated. These would claim a share of existing provisioned resources, or would notify the bandwidth broker in the case that additional resources are required.

A more sophisticated model would allow TCAs to be renegotiated dynamically. In this case, the process would be automatic, and would not require human intervention. Each domain would in effect, represent a bandwidth broker, via one protocol or another. A specific inter-domain protocol might be used to communicate between centralized bandwidth broker agents, or alternatively, an inter-domain variant of RSVP might be used. In the latter case, there is no direct interaction with a bandwidth broker per-se. However, the collection of network nodes, policy servers and directory behave collectively as a bandwidth broker which communicates using RSVP. In either case, TCA renegotiations would be triggered by load measurements at boundary nodes. These could be in the form of changes in actual measured traffic volume, or alternatively, based on explicit fine grain RSVP resource requests from hosts at the periphery. Domains would approve renegotiations based both on resource constraints as well as predetermined policy constraints.

7.3 Service, PHB and Codepoint Mapping

In order to provide end-to-end service to customers, it must be possible to extend services across multiple domains. Several complexities may arise at inter-domain boundaries, as follows:

1. The services provided by a certain domain may not be compatible with the services provided by a neighbour domain.
2. The services provided by a certain domain may be compatible with those provided by the neighbour domain, but the PHB used to obtain the service might be different.
3. The PHB might be the same, but the codepoint used to request the PHB might be different.
4. The PHB and codepoint are the same but differences in provisioning and charging models results in different services.

Resolution of these complexities requires determination of the compatible services and negotiation of the PHB codepoints which will be used to request the services. This process is greatly simplified by the provision of a set of universal services using universally recognized codepoints. The leased line service and EF codepoint is likely to be one such example. Generally, extension of quantitative services across multiple domains will require more uniformity in the nature of the services provided. Qualitative services on the other hand, may be extended end-to-end by a concatenation of services which vary from domain to domain. For example, one domain may base a qualitative service on a WFQ scheme with RED while another may use priority queuing with RIO. Since the assurances provided by qualitative services tend to be looser, it is possible that a meaningful service can be provided end-to-end by concatenating these two service types.

7.4 Host-Domain Boundaries

In certain cases, a host may be directly attached to a differentiated service domain. This is likely both in the case of campus networks that provide differentiated services within the network or in the case of dial-up users connecting to a differentiated service provider. In these cases, the host can be considered the customer of the differentiated service network. Legacy hosts are unlikely to mark their own packets for the appropriate DS-field and are also unlikely to shape or police their traffic. In the case of legacy hosts, the differentiated service provider will have to provide these services on behalf of the customer. In the case of campus networks, some network wide policy would likely be used to configure these services in the DS boundary devices. In the case of dial-up hosts, marking, shaping and resources provided would likely be negotiated at the time the customer signs up with the provider.

Newer hosts may be capable both of marking and of traffic shaping. In this case, the overall per-host resource constraints are still

likely to be somewhat static. However, the manner in which the host shares these resources among its various traffic flows is determined by the host. Of course, the provider will have to configure policers to assure that the host does not seize more than its share of resources in the differentiated service network.

8. Inter-operability with RSVP/Integrated Services

In this section, we discuss alternatives for inter-operability between differentiated services and RSVP/Integrated services.

8.1 RSVP/Integrated Services Over Differentiated Services

This scenario is discussed in detail in [E2EQOS]. It assumes a model in which peripheral stub networks are RSVP and Intserv aware. These are interconnected by differentiated service networks. In this model, the scalability of differentiated service networks helps to extend the reach of RSVP/Integrated service (Intserv)networks. Intervening differentiated service networks appear as a single RSVP hop to the RSVP/Intserv networks. Hosts attached to the peripheral RSVP/Intserv networks signal to each other for per-flow resource requests across the differentiated service networks. Standard RSVP/Intserv processing is applied within the RSVP/Intserv peripheral networks. RSVP signalling messages are carried transparently through the differentiated service networks. Devices at the boundaries between the RSVP/Intserv networks and the differentiated service networks process the RSVP messages and provide admission control based on the availability of appropriate resources within the differentiated service network.

This model is predicated on the availability of services within the differentiated service network which can extend the reach of intserv type services. For example, the leased line service can extend the intserv guaranteed service across a differentiated service network. Multiple guaranteed service micro-flows which exist in peripheral networks are aggregated into the EF behaviour aggregate at the edge of the diffserv network. When an RSVP request for guaranteed service arrives at the edge of a differentiated service network, RSVP style admission control is applied based on the amount of resources requested in the intserv flowspec and the availability of differentiated services at the corresponding service level (per the TCA). If admission control succeeds, the originating host (or its agent) marks traffic on the signalled microflow, for the appropriate differentiated service level.

The RSVP/Intserv over differentiated service model is especially suitable for providing quantitative end-to-end services. The use of differentiated services eliminates the scalability concerns of RSVP/Intserv networks. The use of RSVP signalling provides admission control to the differentiated service network, based on resource availability and policy decisions. It also greatly simplifies the

configuration of differentiated service classifiers, policers and other traffic conditioning components.

Variations on this theme would enable some number of nodes within the differentiated service networks to process the per-flow RSVP messages passing through. These could be used to aid in dynamic provisioning without necessarily requiring any per-flow state or processing within the differentiated service network. In yet another model, the transition of per-flow RSVP messages through the differentiated service network might trigger aggregated RSVP signalling between differentiated service domain edges, for the purpose of renegotiating TCAs and adjusting provisioning dynamically [GBH97, CLASSY].

8.2 Parallel Operation

Another alternative for the interoperation of differentiated service and RSVP/Intserv networks is simple parallel operation. In this mode, each node within the differentiated service network may also be an RSVP capable node. Some strategy would have to be selected for determining which packets are handled using RSVP and which are handled using differentiated services. For example, those that classify to an RSVP installed filter might be handled using RSVP, while those not classifying to specific RSVP filters would be handled according to the DS-field using differentiated service mechanisms. Such a model is likely to be deployed in smaller networks (since the RSVP/Intserv component is less suited for large networks). In particular, the stub networks cited in [E2EQOS] would likely provide differentiated services for those qualitative applications which do not signal, while providing RSVP/Intserv services for those quantitative applications which do signal.

9. Multicast Services

Because the Differentiated Services Architecture deals only with unidirectional flows, a ’multicast’ service in a DS network will in fact offer a point-to-multipoint unidirectional service. Each source of traffic that wishes to send to the multicast group using this service needs a separate SLA which applies at the ingress point where the traffic enters the network.

The network resources that must be provisioned for a multicast service will be affected by the mechanisms used by the routers to provide the service. Depending on the capabilities of the routers and the multicast routing protocol employed, sub-optimal replication of a packet may result in multiple copies travelling over the same link.

If receivers can be added dynamically to a multicast group whilst a flow is in progress, the complexity of provisioning grows considerably: The amount of network resources that will be consumed

by multicast traffic originating from a particular upstream network may be difficult to forecast in advance. Consequently, it may not be possible to offer quantitative services where dynamic addition of receivers adds to the paths through the network already used by the flow.

9.1 Codepoints and PHBs for Multicast Services

To achieve resource isolation of multicast traffic from unicast traffic, it may be necessary to use separate codepoints and separate instances of a PHB or different PHBs for the multicast and unicast services. If the multicast traffic is not adequately isolated, dynamic addition of new members of the multicast group can adversely affect existing unicast traffic.

Because a multicast service traffic flow can exit from a domain to several peer domains, care must be taken to use a codepoint and PHB that is compatible with the peering SLAs at the egress points. This may be a more stringent requirement than for a unicast service where a flow need only be compatible with a single egress point SLA.

9.2 Provisioning Multicast Services

The scope of a multicast service would normally be either case 1 (any egress point) or case 3 (a pre-defined set of egress points) of Sec. 4.4.

For a quantitative service the scope will, in general, need to be case 3. The service can be provisioned in a similar way to corresponding unicast services with the same volume of traffic along each of the paths from ingress to egress, but taking into account that all paths will be used simultaneously and allowing for multiple copies of traffic if necessary. If the multicast routing protocol used can generate different multicast trees depending on the order in which members join the group, provisioning may not be possible. Solving this problem may require pinning of the multicast tree branch points; the solution of this problem is outside the scope of this framework.

For a qualitative service, provisioning is essentially the same as the unicast case, but statistical multiplexing gains are likely to be less because all paths may be used at once.

The traffic conditioning mechanisms for multicast services are not significantly different from those for the unicast services but multiple shapers may be required where traffic exits from several interfaces on a single router or multiple replicas exit from one interface.

10. Security and Tunnelling Considerations

The security and tunnelling implications for the actual data transport of the services of the Differentiated Services Architecture have been extensively discussed in {DSARCH] and [DSHEAD] to which the reader is referred.

Additional security considerations arise from the services overlaid on the data transport:

1. The services are the subject of differential charging. Accordingly, the service users have to be authenticated and authorised, and the accounting data needed must be secured.
2. The mechanisms used to create and distribute the policy and resource allocations must be secured.
3. Statistical data needed to audit service delivery must be secured.

The mechanisms used to provide this security are outside the scope of this framework, but are under consideration by the AAA working group.

The use of tunnels in general and IPsec tunnels in particular impedes the work of MF Classifiers by concealing the fields used by L4 and higher layer classifiers. Thus traffic conditioners within the area where IPsec encryption is used will need to rely only on IP header fields, including the DS field (BA Classifiers will work normally). If more sophisticated Mf classification is required it will have to take place before the tunnel ingress and the application of IPsec encryption. If IPsec encryption is used end-to-end, then Differentiated Services may require host marking.

If a tunnel carries multiple flows with different traffic types, they may be marked with different DS codepoints so that they are subjected to appropriate behaviors in the network interior. This may be considered to be a security breach as it allows traffic patterns to become visible. If just one codepoint is used for all traffic it should be selected carefully to be appropriate for all the traffic in the tunnel.

11. Acknowledgements

The authors would like to acknowledge the helpful comments and suggestions of the following individuals: Kathleen Nichols, Brian Carpenter, David Black, Konstantinos Dovrolis, Shivkumar Kalyana, Wu-chang Feng, Marty Borden, and Ronald Bonica.

12. References

[BB] K. Nichols, V. Jacobson, and L. Zhang, "A Two-bit Differentiated Services Architecture for the Internet", Internet Draft

[CLARK] D. Clark and J. Wroclawski, "An Approach to Service Allocation in the Internet", Internet Draft

[CLASSY] S. Berson and S. Vincent, "Aggregation of Internet

Integrated Services State", Internet Draft, November 1997.

[COPS] J. Boyle, R. Cohen, D. Durham, S. Herzog, R. Rajan, and A. Sastry, "COPS (Common Open Policy Service) Protocol", March 1998.

[DSARCH] D. Black, S. Blake, M. Carlson, E. Davies, Z. Wang, and W. Weiss, "An Architecture for Differentiated Services", Internet Draft, May 1998.

[DSHEAD] K. Nichols and S. Blake, "Definition of the Differentiated Services Field (DS Byte) in the IPv4 and IPv6 Headers", Internet Draft, May 1998.

[AF] J.Heinanen, _Assured Forwarding PHB Group_Internet Draft, August 1998.

[EF] V.Jacobson, _Expedited Fowarding Per Hop Behavior_, Internet Draft, August 1998.

[Ellesson] E. Ellesson and S. Blake, "A Proposal for the Format and Semantics of the TOS Byte and Traffic Class Byte in IPv4 and IPv6", Internet Draft, November 1997.

[E2EQOS] Y. Bernet, R. Yavatkar, P. Ford, F. Baker, and L. Zhang, "A Framework for End-to-End QoS Combining RSVP/Intserv and Differentiated Services", Internet Draft, March 1998.

[GBH97] R. Guerin, S. Blake, and S. Herzog, "Aggregating RSVP- based QoS Requests", Internet Draft, November 1997.

[IntServ] R. Braden, D. Clark, and S. Shenker, "Integrated Services in the Internet Architecture: An Overview", Internet RFC 1633, July 1994.

[RSVP] B. Braden et. al., "Resource ReSerVation Protocol (RSVP) -- Version 1 Functional Specification", Internet RFC 2205, September 1997.

11. Author's Addresses

Bernet, Yoram
Microsoft
One Microsoft Way
Redmond, WA 98052
Phone: +1 (425) 936-9568
Email: yoramb@microsoft.com

Binder, James
3Com Corp.
5400 Bayfront Plaza
Santa Clara, CA 95052
Phone: +1 (408) 326-6051
Email: james_binder@3com.com

Blake, Steven
Torrent Networking Technologies
3000 Aerial Center, Suite 140
Morrisville, NC 27560
Phone: +1-919-468-8466 x232
Fax: +1-919-468-0174
Email: slblake@torrentnet.com

Carlson, Mark
RedCape Software Inc.
2990 Center Green Court South
Boulder, CO 80301
Phone: +1 (303) 448-0048 x115
Email: mac@redcape.com

Davies, Elwyn
Nortel UK
London Road
Harlow, Essex CM17 9NA, UA
Phone: +44-1279-405498
Email: elwynd@nortel.co.uk

Ohlman, Borje
Ericsson Radio
Dialoggatan 1 (Kungens Kurva)
S-126 25 Stockholm
Sweden
Phone: +46-8-719 3187
Email: Borje.Ohlman@ericsson.com

Srinivasan Keshav
4107B Uspon Hall
Cornell University
Ithaca, NY 14853
Phone: +607-255-5395
Email: skeshav@cs.cornell.edu

Dinesh Verma IBM T. J. Watson Research Center
P.O. Box 704
Yorktown Heights, NY 10598
Phone: +1 (914) 784-7466
Email: dverma@watson.ibm.com

Zheng Wang
Bell Labs Lucent Tech
101 Crawfords Corner Road
Holmdel, NJ 07733
Email: zhwang@bell-labs.com

Walter Weiss
Lucent Technologies
300 Baker Avenue, Suite 100 Concord, MA 01742-2168
Email: wweiss@lucent.com