Internet Engineering Task Force
Differentiated Services Working Group
Internet Draft
Expires in August, 2000
draft-ietf-diffserv-ba-def-01.txt

K. Nichols
Cisco Systems
Brian Carpenter
IBM

February, 2000

Definition of Differentiated Services Behavior Aggregates and Rules for their Specification

<draft-ietf-diffserv-ba-def-01.txt>

Status of this Memo

This document is an Internet-Draft and is in full conformance with all provisions of Section 10 of RFC2026. Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This document is a product of the Diffserv working group. Comments on this draft should be directed to the Diffserv mailing list <diffserv@ietf.org>. The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt. The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html.

Distribution of this memo is unlimited.

Copyright Notice

Abstract

The diffserv WG has defined the general architecture for differentiated services (RFC 2475) and has been focused on the definition and standardization of the "per-hop forwarding behaviors" (or PHBs) required in routers (RFCs 2474, 2597, and 2598). The differentiated services framework creates services within a network by applying rules at the edges in the creation of traffic aggregates (known as Behavior Aggregates) coupled with the forwarding path behavior. The WG has also discussed the behavior required at diffserv network edges or boundaries for conditioning the aggregates, elements such as policers and shapers [MODEL,

MIB]. A major feature of diffserv is that only the components applying the rules at the edge need to be changed in response to short-term changes in QoS goals in the network, rather than reconfiguring the interior behaviors.

The next step for the WG is to lay out how the forwarding path components (PHBs, classifiers, and traffic conditioners) can be used within the architectural framework to compose specific Behavior Aggregates. These BAs should have properties such that the transit of individual packets of a BA through a differentiated services network can be characterized by specific metrics. However, no microflow information should be required as packets transit a differentiated services network.

This document defines and discusses Behavior Aggregates in detail and lays out the format and required content for contributions to the Diffserv WG on BAs and the rules that will be applied for individual BA specifications to advance as WG products. This format is specified to expedite working group review of BA submissions.

A pdf version of this document is available at: ftp://ftp-eng.cisco.com/ftp/kmn-group/docs/BA_def.pdf.

Table of Contents

1.0 Introduction

Differentiated Services allows an approach to IP QoS that is modular, high performance, incrementally deployable, and scalable [RFC2475]. Although an ultimate goal is interdomain quality of service, there remain many untaken steps on the road to achieving this goal. One essential step, the evolution of the business models for interdomain QoS, will necessarily develop outside of the IETF. A goal of the diffserv WG is to provide the firm technical foundation that allows these business models to develop.

The Diffserv WG has finished the first phase of standardizing the behaviors required in the forwarding path of all network nodes, the per-hop forwarding behaviors or PHBs. The PHBs defined in RFCs 2474, 2597 and 2598 give a rich toolbox for differential packet handling. Although business models will have to evolve over time, there are technical issues in moving "beyond the box" that lead to QoS models within a single network, i.e., not crossing administrative domain boundaries. Providing QoS within a single network is useful in itself and will provide useful deployment experience for further IETF work as well as for the evolution of business models. This step is critical in the evolution of Diffserv QoS and should ultimately provide the technical input that will aid in the construction of business models. The ultimate goal of creating end to end QoS in the Internet imposes the requirement that we can create and quantify a behavior for a group of packets that is preserved when they are aggregated with other packets.

Once packets have crossed the DS boundary, adherence to the diffserv framework makes it possible to group packets solely according to the behavior they receive at each hop. This approach has well-known scaling advantages, both in the forwarding path and in the control plane. Less well recognized is that these scaling properties only result if the per-hop behavior definition gives rise to a particular type of invariance under aggregation. Since the per-hop behavior must be the same for every node in the domain while the set of packets marked for that PHB may be different at every node, a PHB should be defined such that its treatment of packets of a behavior aggregate doesn't change when other packets join or leave the BA. If the properties of a BA using a particular PHB hold regardless of how the aggregate mutates as it traverses the domain, then that BA scales. If there are limits to where the properties hold, that translates to a limit on the size or topology of a DS domain that can use that BA. Although useful

single-link BAs might exist, BAs that are invariant with network size or that have simple relationships with network size and whose properties can recovered by reapplying rules (that is, forming another diffserv boundary or edge to re-enforce the rules for the aggregate) are needed for building scalable end-to-end quality of service.

There is a clear distinction between the definition of a Behavior Aggregate in a DS domain and a service that might be specified in a Service Level Agreement. The BA definition is a technical building block that couples rules, specific PHBs, and configurations with specific observable characteristics. These definitions are intended to be useful tools in configuring DS domains, but the BA (or BAs) used by a provider are not expected to be visible to customers any more than the specific PHBs employed in the provider's network would be. QoS providers are expected to select their own measures to make customer-visible in contracts and these may be stated quite differently from the characteristics in a BA definition. Similarly, specific BAs are intended as tools for

ISPs to construct differentiated services offerings; each may choose different sets of tools, or even develop their own, in order to achieve particular externally observable metrics.

This document defines Differentiated Services Behavior Aggregates more precisely than past documents and specifies the format that must be used for submissions of particular Behavior Aggregates to the Diffserv WG.

2.0 Some Definitions from RFC 2474

The following definitions are stated in RFCs 2474 and 2475 and are repeated here for easy reference:

Behavior Aggregate: a collection of packets with the same codepoint crossing a link in a particular direction. The terms "aggregate" and "behavior aggregate" are used interchangeably in this document.

Differentiated Services Domain: a contiguous portion of the Internet over which a consistent set of differentiated services policies are administered in a coordinated fashion. A differentiated services domain can represent different administrative domains or autonomous systems, different trust regions, different network technologies (e.g., cell/frame), hosts and routers, etc. Also DS domain.

Differentiated Services Boundary: the edge of a DS domain, where classifiers and traffic conditioners are likely to be deployed. A diff-

erentiated services boundary can be further sub-divided into ingress and egress nodes, where the ingress/egress nodes are the downstream/ upstream nodes of a boundary link in a given traffic direction. A differentiated services boundary typically is found at the ingress to the first-hop differentiated services-compliant router (or network node) that a host's packets traverse, or at the egress of the last-hop differentiated services-compliant router or network node that packets traverse before arriving at a host. This is sometimes referred to as theboundary at a leaf router. A differentiated services boundary may be co-located with a host, subject to local policy. Also DS boundary.

3.0 The Value of Defining Edge-to-Edge BAs

Networks of DS domains can be connected to create end-to-end services, but where DS domains are independently administered, the evolution of the necessary business agreements and future signaling arrangements will take some time. Early deployments will be within a single administrative domain. The specification of the transit expectations of behavior aggregates across DS domains both assists in the deployment of that single-domain QoS and will help enable the composition of end-to-end, cross domain services to proceed. Putting aside the business issues, the same technical issues that arise in interconnecting DS domains with homogeneous administration will arise in interconnecting the autonomous systems (ASs) of the Internet.

Today's Internet is composed of multiple independently adminis-

tered domains or Autonomous Systems (ASs), represented by the circles in figure 1. To deploy ubiquitous end-to-end quality of service in the Internet, a business models must evolve that include issues of charging and reporting that are not in scope for the IETF. In the meantime, there are many possible uses of quality of service within an AS and the IETF can address the technical issues in creating an intradomain QoS within a Differentiated Services framework. In fact, this approach is quite amenable to incremental deployment strategies.

Figure 1: Interconnection of ASs and DS Domains

A single AS (for example, B in figure 1) may be composed of subnetworks and, as the definition allows, these can be separate DS domains. For a number of reasons, it might be useful to have multiple DS domains in an AS, most notable being to follow topological and/or technological boundaries and to separate the allocation of resources. If we confine ourselves to the DS bound-

aries between these "interior" DS domains, we avoid the non-technical problems of setting up a service and can address the issues of creating characterizable behavior aggregates.

The incentive structure for differentiated services is based on upstream domains ensuring their traffic conforms to agreed upon rules and downstream domains enforcing that conformance so that characteristics of behavior aggregates might sensibly be computed. The filled in boxes in figure 1 represent the conformance ensurers (e.g., shapers) and conformance enforcers (e.g., policers). Although we expect that policers and shapers will be required at the boundaries of ASs, they might appear anywhere, or nowhere, inside the AS. Thus, the boxes at the DS boundaries internal to the AS may or may not condition traffic. Understanding behavior under aggregation will result in guidelines for the placement of DS boundaries.

4.0 Understanding Diffserv Behavior Aggregates

4.1 Defining BAs

In this section we expand on the definition of Behavior Aggregates given in RFCs 2474 and 2475. Those RFCs define a Differentiated Services Behavior Aggregate as "a collection of packets with the same DS codepoint crossing a link in a particular direction" and further state that packets with the same DSCP get the same per-hop forwarding treatment (or PHB) everywhere inside a single DS domain. Note that even if multiple DSCPs map to the same PHB, this must hold for each DSCP individually.

Within a DS domain, BAs are formed by the application of rules to packets arriving at the DS boundary, through classification and traffic conditioning. Packets that conform to the rules are marked with the same DSCP (or a known set of DSCPs) within a domain. In the interior of a DS domain, where DSCPs should not be

remarked, as there are no rules being applied. Though a DS domain may be as small as a single node, more complex topologies are expected to be the norm, thus the BA's definition must hold as it is split and merged on the interior links of a DS domain. Packet flow in a network is not part of the BA definition; the application of rules as packets enter the DS domain and the consistent PHB through the DS domain must suffice. (Though limits can be put on the applicability of a specific BA.)

Associated with each BA are measurable, quantifiable, characteristics which can be used to describe what will happen to packets

of that BA as they cross the DS domain. These expectations derive from the rules that are enforced during the entry of packets into the DS domain (the creation of the BA) and the forwarding treatment (PHB) the BA gets inside the cloud. They may be absolute or statistical bounds and they may be parameterized by network properties.

4.2 Constructing BAs

Generally, the forwarding path of a DS domain is configured to meet the network operator's traffic engineering goals for the domain, independently of the performance goals for a particular flow of a BA. Once the interior is configured, the rules on allocating BAs come from meeting the desired performance goals subject to that configuration of link schedulers and bandwidth. The rules at the edge may be altered by provisioning or admission control but the decision about which to use and how to apply the rules comes from matching performance to goals.

For example, consider the diffserv domain of figure 1. A BA which specifies explicit bounds on loss must have rules at the edge to ensure that, on the average, no more packets are admitted than can emerge. As the network can contain queues, input traffic may not equal the output traffic over all timescales. However the averaging timescale should not exceed what might be expected for reasonably sized buffering inside the network. Thus if we allow bursts to arrive into the interior of the network, we must know there is enough capacity to ensure that losses don't exceed the BA's bound. Note that explicit bounds on the loss level can be particularly difficult as the exact way in which packets of a particular BA merge inside the network affect the aggregate burstiness and hence, loss.

PHBs give explicit expressions of what treatment a BA can expect from each hop. This behavior must continue to apply under aggregation of merging BA flows. Explicit expressions of what happens to this behavior under aggregation, possibly parameterized by node in-degrees or network diameters are required. This allows us to determine what to do at internal aggregation points. For example, do we reapply edge rules?

Characterizing a BA requires exploring what happens to a PHB

under aggregation. Rules must be recursively applied to result in a known behavior. As an example, since maximum burst sizes grow with the number of microflows or BA flows merged, a BA speci-

fication must address this. A clear advantage of constructing behaviors that aggregate is the ease of building up BAs that span interior DS domains and eventually farther. For example, a BA with known properties that crosses an interior DS domain of AS B in figure 1, can be merged with the same type of BA at the interior shaded routers. Using the same (or fewer) rules as were applied to create the BA at the entrance to AS B, there should be confidence that the BA can continue to be quantified by the expected behavior.

The specification of the transit expectations of behavior aggregates across domains both assists in the deployment of QoS within a DS domain and helps enable the composition of end-to-end, cross-domain services to proceed.

4.3 Forwarding path vs. control plane for BAs

The PHB and the edge rules that form and condition BAs are in the forwarding path and take place at line rates while the configuration of the DS domain edge to enforce rules on who goes into a BA and how the BA should behave temporally is done by the control plane on a very different time scale. For example, configuration of PHBs might only occur monthly or quarterly. The edge rules might be reconfigured at a few regular intervals during the day or might happen in response to signalling decisions thousands of times a day. Even at the shortest time scale, control plane actions are not expected to happen per-packet. Much of the control plane work is still evolving and is outside the charter of the Diffserv WG since how the configuration is done and at what time scale it is done should not affect the characteristics of the BA.

5.0 Format for Specification of Diffserv Behavior Aggregates

Behavior Aggregates arise from a particular relationship between edge and interior (which may be parameterized). The quantifiable characteristics of a BA MUST be independent of whether the network edge is configured statically or dynamically. The particular configuration of traffic conditioners at the DS domain edge is critical to how a BA performs, but the act(s) of configuring the edge is a control plane action which can be separated from the specification of the BA.

The following sections must be present in any specification of a Differentiated Services Behavior Aggregate. Of necessity, their length and content will vary greatly.

5.1 Applicability Statement

All BAs must have an applicability statement that outlines the

intended use of this BA and the limits to its use.

5.2 Rules

This section describes the rules to be followed in the creation of this BA. Rules should be distinguished with MAY, MUST, and SHOULD. The rules specify the edge behavior and configuration and the PHB (or PHBs) to be used and any additional requirements on their configuration beyond that contained in RFCs.

5.3 Characteristics

The characteristics of a BA tell how it behaves under ideal conditions if configured in a specified manner (where the specification may be parameterized). Characteristics of a BA might be drop rate, throughput, delay bounds measured over some time period. They may be absolute bounds or statistical bounds (e.g., "90% of all packets measured over intervals of at least 5 minutes will cross the DS domain in less than 5 milliseconds"). A wide variety of characteristics may be used but they MUST be explicit, quantifiable, and defensible. Where particular statistics are used, the document must be precise about how they are to be measured and about how the characteristics were derived.

Advice to a network operator would be to use these characteristics as guidelines in creating a service specification rather than use them directly. For example, a "loss-free" BA would probably not be sold as such, but rather as a service with a very small packet loss probability.

5.4 Parameters

The definition and characteristics of a BA MAY be parameterized by network-specific features; for example, maximum number of hops, minimum bandwidth, total number of entry/exit points of the BA to/from the diffserv network, maximum transit delay of network elements, minimum buffer size available for the BA at a network node, etc.

5.5 Assumptions

In most cases, BAs will be characterized assuming lossless links, no link failures, and relatively stable routing. This is reasonable since otherwise it would be very difficult to quantify behavior. However, these assumptions must be clearly stated. If additional restrictions, e.g., route pinning, are required, these must be stated. Some BAs may be developed without these assumptions, e.g., for

high loss rate links, and these must also be made explicit.

Further, if any assumptions are made about the allocation of resources within a diffserv network in the creation of the aggregate, these must be made explicit.

5.6 Example Uses

A BA specification must give example uses to motivate the understanding of ways in which a diffserv network could make use of the BA although these are not expected to be detailed. For example, "A bulk handling behavior aggregate may be used for all packets which should not take any resources from the network unless they would otherwise go unused. This might be useful for Netnews traffic or for traffic rejected from some other BA due to violation of that BA's rules."

5.7 Environmental Concerns (media, topology, etc.)

Note that it is not necessary for a provider to expose what Behavior Aggregate (if a commonly defined one) is being used nor is it necessary for a provider to specify the service by the BA's characteristics. For example, a service provider might use a BA with a "no queueing loss" characteristic in order to specify a "very low loss" service.

This section is to inject realism into the characteristics described above. Detail the assumptions made there and what constraints that puts on topology or type of physical media or allocation.

6.0 Structuring BAs to Meet Expectations

Associated with each BA is an expectation: measurable, quantifiable, characteristics which can be used to describe what will happen to packets of that BA as they cross the domain. These expectations result directly from the application of rules enforced during the creation of the BA and/or its entry into the domain and the forwarding treatment (PHB) packets of the BA get inside the domain. There are many ways in which traffic might be distributed, but creating a quantifiable, realizable service across the DS domain will limit the scenarios which can occur. There is a clear correlation between the strictness of the rules and the quality of the characterization of the BA.

There are two kinds of BA properties to consider. First are the properties over "long" time periods, or average behaviors. In a

description of a BA, these would be the rates or throughput seen over some specified time period. The second set of properties has to do with the "short" time behavior, usually expressed as the allowable burstiness in an aggregate. The short time behavior is important is understanding the buffering (and associated loss characteristics) and in quantifying how the BA aggregates, either within a DS domain or at the boundaries. For short-time behavior, we are interested primarily in two things: 1) how many back-to-back packets of this BA will we see at any point (this would be metered as a burst) and 2) how large a burst of packets of this BA can appear in a queue at once (gives queue overflow and loss).

Put simply, a BA specification should provide the answer to the

question: Under what conditions can we join the output of this domain to another under the same rules and expectations?

6.1 Considerations in specifying long-term or average BA characteristics

To make this more concrete, consider the DS domain of figure 2. First consider the average or long-term behavior that must be specified for a target BA which we designate as BAx. Can the DS domain handle the average traffic flow? Is that answer topology-dependent or are there some specific assumptions on routing which must hold for BAx to preserve its "adequately provisioned" capability? In other words, if the topology of D changes suddenly, will the properties of BAx change? Will the loss rate of BAx dramatically increase?

Figure 2: ISP and DS domain D connected in a ring and connected to DS domain E

Let figure 2 be an ISP ringing the U.S. with links of bandwidth B and with N tails to various metropolitan areas. If the link between the node connected to A and the node connected to Z were to go down, causing all the BAx traffic between the two to transit the entire ring, would the bounded behavior of BAx change? If some node of the ring now has a larger arrival rate to one of its links than the capacity of the link for BAx, clearly the loss rate would change dramatically. In that case, there were topological assumptions made about the path of the traffic from A to Z that affected the characteristics of BAx. Once these no longer hold, any assumptions on the loss rate of packets of BAx crossing the domain would change; for example, a characteristic such as "loss rate no greater than 1% over any interval larger than 10 minutes"

would no longer hold. A BA specification should spell out the assumptions made on preserving the characteristics.

6.2 Considerations in specifying short-term or bursty BA characteristics

Next, consider the short-time behavior of a BA, specifically whether permitting the maximum bursts to add in the same manner as the average rates will lead to properties that aggregate or under what rules this will lead to properties that aggregate. In our example, if domain D allows each of the uplinks to burst of p packets into BAx, they could accumulate as they transit the ring. For packets headed for link L, back-to-back BAx packets can come from both directions and arrive at the same time. If the bandwidth of link L is the same as the links of the ring, this probably does not present a buffering problem. If there are two input links that can send packets to queue for L, at worst, two packets can arrive simultaneously for L. If the bandwidth of link L equals or exceeds twice B, the packets won't accumulate. Further, if p is limited to one, and the bandwidth of L exceeds the rate of arrival (over the longer term) of BAx packets (required for bounding the

loss) then the queue of BAx packets for link L will empty before new packets arrive. If the bandwidth of L is equal to B, one packet of BAx must queue while the other is transmitted. This would result in N x p back-to-back packets of BAx arriving over L during the same time scale as the bursts of p were permitted on the uplinks. Link L should be configured to handle the sum of the rates that ingress to BAx, but that doesn't guarantee that it can handle the sum of the N bursts into BAx.

If the bandwidth of L is less than B, then the link must buffer Nxpx(B-L)/B BAx packets to avoid loss. If BAx is getting less than the full bandwidth L, this number is larger. For probabilistic bounds, a smaller buffer might do if the probability of exceeding it can be bounded.

More generally, for router indegree of d, bursts of BAx packets might arrive on each input. Then, in the absence of any additional rules, it is possible that dxpx(# of uplinks) back-to-back BAx packets can be sent across link L to domain E. Thus the DS domain E must permit these much larger bursts into BAx than domain D permits on the N uplinks or else the flow of BAx packets must be made to conform to the rules for entering E (e.g., by shaping).

What conditions should be imposed on a BA and on the PHBs which carry it in order to ensure BAs that can be interconnected as across the interior DS domains of figure 1? Edge rules for constructing a BA that has certain characteristics across a DS domain should apply independently of the origin of the packets. With reference to the example we've been exploring, the rules for a BA entering link L into domain E should not depend on the number of uplinks into domain D.

6.3 Example

Consider where all the uplinks have the same bandwidth B and link L has bandwidth L which is less than or equal to B. Flows of BAx packets from the N uplinks each have average rate R and are destined to cross L. If only a fraction a of link L is allocated to BAx, then R =axL/N fits the average rate constraint. If each of the N flows can have a burst of p packets and half the flows transit the ring in each direction, then 2xp packets can arrive at the BAx queue for link L in time it took to transmit p packets on the ring, p/B. Although the link scheduler for link L might allow the burst of packets to be transmitted at the line rate L, after the burst allotment has been exceeded, the queue should be expected to clear at only rate axL. Then consider the packets that can accumulate. It takes 2xp/(axL) to clear the queue of BAx packets. In that time, bursts of p packets from the other uplinks can arrive from the ring, so the packets do not even have to be back-to-back. Even if the packets do not arrive back-to-back, but are spaced by less time than it takes to clear the queue of BAx packets, either the required buffer size can become large or the burst size of BAx entering E

across L becomes large and is a function of N, the number of uplinks of domain D.

Let L = 1.5 Mbps, B = 45 Mbps, a = 1/3, N=10, p = 3. Suppose that the bursts from two streams of BAx arrive at the queue for link L very close together. Even if 3 of the packets are cleared at the line rate of 1.5 Mbps, there will be 3 packets remaining to be serviced at a 500 kbps rate. In the time allocated to send one of these, 9 packets can arrive on each of the inputs from the ring. If any non-zero number of these 18 packets are of BAx, the queue size will not reduce. If two more bursts (6 of the 18 packets) arrive, the queue increases to 8 packets. Thus, it's possible to build up quite a large queue, one likely to exceed the buffer allocated for BAx. The rate bound means that each of the uplinks will be idle for the time to send three packets at 50 kbps, possibly by policing at the ring egress, and thus the queue would eventually

decrease and clear, however, the queue at link L can still be very large. There may be BAs where the intention is to permit loss, but in that case, it should be constructed so as to provide a probabilistic bound for the queue size to exceed a reasonable buffer size of one or two bandwidth-delay products. Alternatively or additionally, rules can be used that bound the amount of BAx that queues by limiting the burst size at the ingress uplinks to one packet, resulting in a maximum queue of N or 10 or to impose additional rules on the creation of the aggregate, such as intermediate shaping.

7.0 Reference Behavior Aggregates

The intent of this section is to define one or a few "reference" BAs; certainly a Best Effort BA and perhaps others. This section is very preliminary at this time and meant to be the starting point for discussion rather than its end. These are BAs that have little in the way of rules or expectations.

7.1 Best Effort Behavior Aggregate

7.1.1 Applicability

A Best Effort (BE) BA is for sending "normal internet traffic" across a diffserv network. That is, the definition and use of this BA is to preserve, to a reasonable extent, the pre-diffserv delivery expectation for packets in a diffserv network that do not require any special differentiation.

7.1.2 Rules

There are no rules governing rate and bursts of packets beyond the limits imposed by the ingress link. At each network node in the interior of the network, packets marked for this BA are given the Default PHB (as defined in [RFC2474]).

7.1.3 Characteristics of this BA

"As much as possible as soon as possible".

Packets of this BA will not be completely starved and when resources are available, this BA should be configured to consume them.

Although some network operators may bound the delay and loss rate for this aggregate given knowledge about their network, such

characteristics are not required.

7.1.4 Parameters

None.

7.1.5 Assumptions

A properly functioning network.

7.1.6 Example uses

1. For the normal Internet traffic connection of an organization.

2. For the "non-critical" Internet traffic of an organization.

7.2 Bulk Handling Behavior Aggregate

7.2.1 Applicability

A Bulk Handling (BH) BA is for sending extremely non-critical traffic across a diffserv network. There should be an expectation that these packets may be delayed or dropped when other traffic is present.

7.2.2 Rules

There are no rules governing rate and bursts of packets beyond the limits imposed by the ingress link. At each network node in the interior of the network, packets marked for this BA are given a CS or AF PHB configured so that it may be starved when other traffic is present.

7.2.3 Characteristics of this BA

Packets are forwarded when there are idle resources.

7.2.4 Parameters

None.

7.2.5 Assumptions

A properly functioning network.

7.2.6 Example uses

1. For Netnews and other "bulk mail" of the Internet.

2. For "downgraded" traffic from some other BA.

8.0 Sketchy Examples of Creating and Using BAs

There is a clear interaction between the number and strictness of the rules and the number and strictness of quantifiable characteristics for a BA. Examples of more strictly defined BAs will be necessary to make this document's definitions clearer. This is being addressed in two ways. One, a companion document is in preparation defining a BA that uses the EF PHB and is related to the VLL described in [RFC2598]. In addition, this section includes two "sketchy" examples to motivate thinking and discussion on BAs. The following examples are illustrative rather than exhaustive or even complete.

The following should be looked at as "mythical" BAs that may never see the light of day and will likely not appear in future revisions of this document.

8.1 Loss Tolerant Provisioned

A loss-tolerant provisioned BA is useful for statistically provisioning a BA whose packets should have low delay, but are loss-tolerant. Rules for this aggregate are that entering composite BAs must not exceed a peak rate of Rp and may not burst more than two MTU packet-times at Rp. The BA uses CS3, selected by DSCP03 and configured so that its minimum share of all internal links is Smin (in bps), running active queue management with a low threshold (defined in time rather than packets) and a small maximum queue size (also in time). Characteristics of this BA:

Some 90th percentile bound on loss and delay.

Parameterized by Smin and Rp. The sum of all the Rp should be on the order of some over provisioning factor (larger than 1).

Assumptions on these characteristics are that the network is operating under ideal conditions.

8.2 Preferred

A Preferred BA is for provisioning traffic so as to give low-load performance across a DS domain. The rules governing it are that the packets of this BA arriving over any ingress to the domain are average rate-limited to Ra with a maximum burst size of Bmax. The BA uses CS4, selected by DSCP04 and configured so that its minimum share of all internal links is Smin (in bps) and the sum of all Ra < Smin. Characteristics of this BA:

Probabilistic bounds based on the sum of all allocated rates and the burst size.

Throughput measured over 5 minute intervals will be at least Ra.

Assumptions on these characteristics are that the network is operating under ideal conditions.

Example uses:

1. A voice service where customer is guaranteed a conformant packet loss rate of less than 0.5% and a latency bound of 20 ms, 99th percentile jitter less than 2 packet-times, median jitter of less than a packet-time across the domain.

2. A "leased line replacement" where the customer is guaranteed to receive throughput performance indistinguishable from a leased line at Rp with a per-packet delay of less than 20 msec through the cloud.

9.0 Procedure for submitting BAs to Diffserv WG

1. Following the guidelines of this document, write a draft and submit it as an Internet Draft and bring it to the attention of the WG mailing list.

2. Initial discussion on the WG should focus primarily on the merits of such a BA, though comments and questions on the claimed characteristics are reasonable.

3. Once consensus has been reached on a version of a draft that it is a useful BA and that the characteristics "appear" to be correct (i.e., not egregiously wrong) that version of the draft goes to a review panel the WG Co-chairs set up to audit and report on the characteristics. The review panel will be given a deadline for the review. The exact timing of the deadline will be set on a case-by-case basis by the co-chairs to reflect the complexity of the task and other constraints (IETF meetings, major holidays) but is expected to be in the 4-8 week range. During that time, the panel may correspond with the authors directly (cc'ing the WG co-chairs) to get clarifications. This process should result in a revised draft and/or a report to the WG from the panel that either endorses or disputes the claimed characteristics.

4. If/when endorsed by the panel, that draft goes to WG last call. If not endorsed, the author(s) can give a itemized response to the

panel's report and ask for a WG Last Call.

5. If/when passes Last Call, goes to ADs for publication as a WG Informational RFC in our "BA series".

10.0 Acknowledgements

The ideas in this document have been heavily influenced by the Diffserv WG and, in particular, by discussions with Van Jacobson, Dave Clark, Lixia Zhang, Geoff Huston, Scott Bradner, Randy Bush, Frank Kastenholz, Aaron Falk, and a host of other people who should be acknowledged for their useful input but not be held accountable for our mangling of it.

11.0 References

[RFC2474] RFC 2474, "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers", K.Nichols, S. Blake, F. Baker, D. Black, www.ietf.org/rfc/rfc2474.txt

[RFC2475] RFC 2475, "An Architecture for Differentiated Services", S. Blake, D. Black, M.Carlson,E.Davies,Z.Wang,W.Weiss, www.ietf.org/rfc/rfc2475.txt

[RFC2597] RFC 2597, "Assured Forwarding PHB Group", F. Baker, J. Heinanen, W. Weiss, J. Wroclawski, ftp://ftp.isi.edu/in-notes/rfc2597.txt

[RFC2598] RFC 2598, "An Expedited Forwarding PHB", V.Jacobson, K.Nichols, K.Poduri, ftp://ftp.isi.edu/in-notes/rfc2598.txt

[MODEL] "A Conceptual Model for Diffserv Routers", draft-ietf-diffserv-model-01.txt, Bernet et. al.

[MIB] "Management Information Base for the Differentiated Services Architecture", draft-ietf-diffserv-mib-01.txt, Baker et. al.

Authors' Addresses

Kathleen Nichols Brian E. Carpenter

Cisco Systems
170 West Tasman Drive
San Jose, CA 95134-1706

email: kmn@cisco.com

IBM
c/o iCAIR
Suite 150
1890 Maple Avenue
Evanston, IL 60201
USA
EMail: brian@icair.org