Network Working Group
Request for Comments: 3086
Category: Informational

K. Nichols
Packet Design
B. Carpenter
IBM

April 2001

Definition of Differentiated Services Per Domain Behaviors and Rules for their Specification

Status of this Memo

This memo provides information for the Internet community. It does not specify an Internet standard of any kind. Distribution of this memo is unlimited.

Copyright Notice

Copyright (C) The Internet Society (2001). All Rights Reserved.

Abstract

The differentiated services framework enables quality-of-service provisioning within a network domain by applying rules at the edges to create traffic aggregates and coupling each of these with a specific forwarding path treatment in the domain through use of a codepoint in the IP header. The diffserv WG has defined the general architecture for differentiated services and has focused on the forwarding path behavior required in routers, known as "per-hop forwarding behaviors" (or PHBs). The WG has also discussed functionality required at diffserv (DS) domain edges to select (classifiers) and condition (e.g., policing and shaping) traffic according to the rules. Short-term changes in the QoS goals for a DS domain are implemented by changing only the configuration of these edge behaviors without necessarily reconfiguring the behavior of interior network nodes.

The next step is to formulate examples of how forwarding path components (PHBs, classifiers, and traffic conditioners) can be used to compose traffic aggregates whose packets experience specific forwarding characteristics as they transit a differentiated services domain. The WG has decided to use the term per-domain behavior, or PDB, to describe the behavior experienced by a particular set of packets as they cross a DS domain. A PDB is characterized by specific metrics that quantify the treatment a set of packets with a particular DSCP (or set of DSCPs) will receive as it crosses a DS domain. A PDB specifies a forwarding path treatment for a traffic aggregate and, due to the role that particular choices of edge and

PHB configuration play in its resulting attributes, it is where the forwarding path and the control plane interact. The measurable parameters of a PDB should be suitable for use in Service Level Specifications at the network edge.

This document defines and discusses Per-Domain Behaviors in detail and lays out the format and required content for contributions to the Diffserv WG on PDBs and the procedure that will be applied for individual PDB specifications to advance as WG products. This format is specified to expedite working group review of PDB submissions.

Table of Contents

1 Introduction

Differentiated Services allows an approach to IP Quality of Service that is modular, incrementally deployable, and scalable while introducing minimal per-node complexity [RFC2475]. From the end user's point of view, QoS should be supported end-to-end between any pair of hosts. However, this goal is not immediately attainable. It will require interdomain QoS support, and many untaken steps remain on the road to achieving this. 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 first major step will be to support edge-to-edge or intradomain QoS between the ingress and egress of a single network, i.e., a DS Domain in the terminology of RFC 2474. The intention is that this edge-to-edge QoS should be composable, in a purely technical sense, to a quantifiable QoS across a DS Region composed of multiple DS domains.

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 by individual boxes. The general architectural model for diffserv has been documented in RFC 2475. An informal router model [MODEL] describes a model of traffic conditioning and other forwarding behaviors. However, technical issues remain in moving "beyond the box" to intradomain QoS models.

The ultimate goal of creating scalable end-to-end QoS in the Internet requires that we can identify and quantify behavior for a group of packets that is preserved when they are aggregated with other packets as they traverse the Internet. The step of specifying forwarding path attributes on a per-domain basis for a set of packets distinguished only by the mark in the DS field of individual packets is critical in the evolution of Diffserv QoS and should provide the technical input that will aid in the construction of business models. This document defines and specifies the term "Per-Domain Behavior" or PDB to describe QoS attributes across a DS domain.

Diffserv classification and traffic conditioning are applied to packets arriving at the boundary of a DS domain to impose restrictions on the composition of the resultant traffic aggregates, as distinguished by the DSCP marking , inside the domain. The classifiers and traffic conditioners are set to reflect the policy and traffic goals for that domain and may be specified in a TCA (Traffic Conditioning Agreement). Once packets have crossed the DS boundary, adherence to diffserv principles makes it possible to group packets solely according to the behavior they receive at each hop (as selected by the DSCP). 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 equivalent for every node in the domain, while the set of packets marked for that PHB may be different at every node, PHBs should be defined such that their characteristics do not depend on the traffic volume of the associated BA on a router’s ingress link nor on a particular path through the DS domain taken by the packets. Specifically, different streams of traffic that belong to the same traffic aggregate merge and split as they traverse the network. If the properties of a PDB using a particular PHB hold regardless of how the temporal characteristics of the marked traffic aggregate change as it traverses the domain, then that PDB scales. (Clearly this assumes that numerical parameters such as bandwidth allocated to the particular PDB may be different at different points in the network, and may be adjusted dynamically as traffic volume varies.) 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 PDB. Although useful single-link DS domains might exist, PDBs that are invariant with network size or that have simple relationships with network size and whose properties can be recovered by reapplying rules (that is, forming another diffserv boundary or edge to re-enforce the rules for the traffic aggregate) are needed for building scalable end-to-end quality of service.

There is a clear distinction between the definition of a Per-Domain Behavior in a DS domain and a service that might be specified in a Service Level Agreement. The PDB definition is a technical building block that permits the coupling of classifiers, traffic conditioners, specific PHBs, and particular configurations with a resulting set of specific observable attributes which may be characterized in a variety of ways. These definitions are intended to be useful tools in configuring DS domains, but the PDB (or PDBs) used by a provider is not expected to be visible to customers any more than the specific PHBs employed in the provider's network would be. Network providers are expected to select their own measures to make customer-visible in contracts and these may be stated quite differently from the technical attributes specified in a PDB definition, though the configuration of a PDB might be taken from a Service Level Specification (SLS). Similarly, specific PDBs 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. Nevertheless, the measurable parameters of a PDB are expected to be among the parameters cited directly or indirectly in the Service Level Specification component of a corresponding SLA.

This document defines Differentiated Services Per-Domain Behaviors and specifies the format that must be used for submissions of particular PDBs to the Diffserv WG.

2 Definitions

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.

" 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 differentiated 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 the boundary at a leaf router. A differentiated services boundary may be co-located with a host, subject to local policy. Also DS boundary.

To these we add:

" Traffic Aggregate: a collection of packets with a codepoint that maps to the same PHB, usually in a DS domain or some subset of a DS domain. A traffic aggregate marked for the foo PHB is referred to as the "foo traffic aggregate" or "foo aggregate" interchangeably. This generalizes the concept of Behavior Aggregate from a link to a network.

" Per-Domain Behavior: the expected treatment that an identifiable or target group of packets will receive from "edge-to-edge" of a DS domain. (Also PDB.) A particular PHB (or, if applicable, list of PHBs) and traffic conditioning requirements are associated with each PDB.

" A Service Level Specification (SLS) is a set of parameters and their values which together define the service offered to a traffic stream by a DS domain. It is expected to include specific values or bounds for PDB parameters.

3 The Value of Defining Edge-to-Edge Behavior

As defined in section 2, a PDB describes the edge-to-edge behavior across a DS domain's "cloud." Specification of the transit expectations of packets matching a target for a particular diffserv behavior across a DS domain will both assist in the deployment of single-domain QoS and will help enable the composition of end-to-end, cross-domain services. Networks of DS domains can be connected to create end-to-end services by building on the PDB characteristics without regard to the particular PHBs used. This level of

abstraction makes it easier to compose cross-domain services as well as making it possible to hide details of a network's internals while exposing information sufficient to enable QoS.

Today's Internet is composed of multiple independently administered domains or Autonomous Systems (ASs), represented by the "clouds" in figure 1. To deploy ubiquitous end-to-end quality of service in the Internet, 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.

Where DS domains are independently administered, the evolution of the necessary business agreements and future signaling arrangements will take some time, thus, early deployments will be within a single administrative domain. 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.

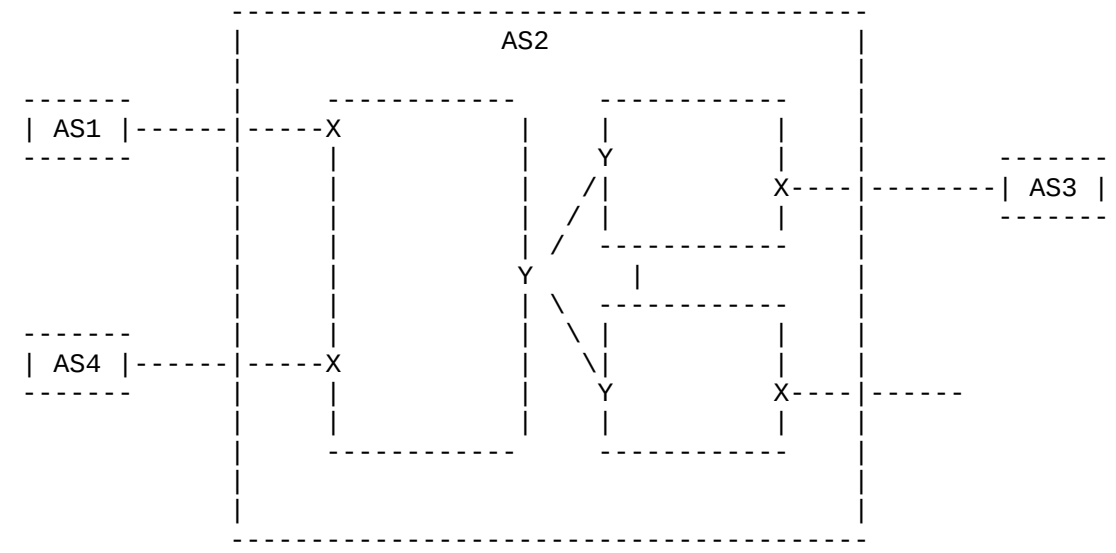

Figure 1: Interconnection of ASs and DS Domains

A single AS (e.g., AS2 in figure 1) may be composed of subnetworks and, as the definition allows, these can be separate DS domains. An AS might have multiple DS domains for a number of reasons, most notable being to follow topological and/or technological boundaries

and to separate the allocation of resources. If we confine ourselves to the DS boundaries between these "interior" DS domains, we avoid the non-technical problems of setting up a service and can address the issues of creating characterizable PDBs.

The incentive structure for differentiated services is based on upstream domains ensuring their traffic conforms to the Traffic Conditioning Agreements (TCAs) with downstream domains and downstream domains enforcing that TCA, thus metrics associated with PDBs can be sensibly computed. The letters "X" and "Y" in figure 1 represent the DS boundary routers containing traffic conditioners that ensure and enforce conformance (e.g., shapers and policers). Although policers and shapers are expected at the DS boundaries of ASs (the "X" boxes), they might appear anywhere, or nowhere, inside the AS. Specifically, the boxes at the DS boundaries internal to the AS (the "Y" boxes) may or may not condition traffic. Technical guidelines for the placement and configuration of DS boundaries should derive from the attributes of a particular PDB under aggregation and multiple hops.

This definition of PDB continues the separation of forwarding path and control plane described in RFC 2474. The forwarding path characteristics are addressed by considering how the behavior at every hop of a packet’s path is affected by the merging and branching of traffic aggregates through multiple hops. Per-hop behaviors in nodes are configured infrequently, representing a change in network infrastructure. More frequent quality-of-service changes come from employing control plane functions in the configuration of the DS boundaries. A PDB provides a link between the DS domain level at which control is exercised to form traffic aggregates with quality-of-service goals across the domain and the per-hop and per-link treatments packets receive that results in meeting the quality-of-service goals.

4 Understanding PDBs

4.1 Defining PDBs

RFCs 2474 and 2475 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. In section 2 of this document, we introduced a more general definition of a traffic aggregate in the diffserv sense so that we might easily refer to the packets which are mapped to the same PHB everywhere within a DS domain. Section 2 also presented a short definition of PDBs which we expand upon in this section:

Per-Domain Behavior: the expected treatment that an identifiable or target group of packets will receive from "edge to edge" of a DS domain. A particular PHB (or, if applicable, list of PHBs) and traffic conditioning requirements are associated with each PDB.

Each PDB has measurable, quantifiable, attributes that can be used to describe what happens to its packets as they enter and cross the DS domain. These derive from the characteristics of the traffic aggregate that results from application of classification and traffic conditioning during the entry of packets into the DS domain and the forwarding treatment (PHB) the packets get inside the domain, but can also depend on the entering traffic loads and the domain's topology. PDB attributes may be absolute or statistical and they may be parameterized by network properties. For example, a loss attribute might be expressed as "no more than 0.1% of packets will be dropped when measured over any time period larger than T", a delay attribute might be expressed as "50% of delivered packets will see less than a delay of d milliseconds, 30% will see a delay less than 2d ms, 20% will see a delay of less than 3d ms." A wide range of metrics is possible. In general they will be expressed as bounds or percentiles rather than as absolute values.

A PDB is applied to a target group of packets arriving at the edge of the DS domain. The target group is distinguished from all arriving packets by use of packet classifiers [RFC2475] (where the classifier may be "null"). The action of the PDB on the target group has two parts. The first part is the the use of traffic conditioning to create a traffic aggregate. During traffic conditioning, conformant packets are marked with a DSCP for the PHB associated with the PDB (see figure 2). The second part is the treatment experienced by packets from the same traffic aggregate transiting the interior of a DS domain, between and inside of DS domain boundaries. The following subsections further discuss these two effects on the target group that arrives at the DS domain boundary.

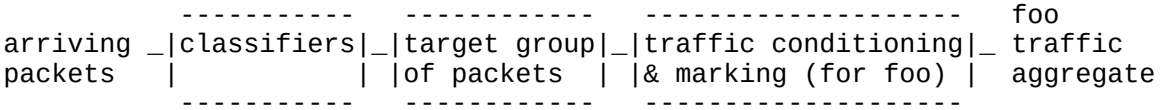

Figure 2: Relationship of the traffic aggregate associated with a PDB to arriving packets

4.1.1 Crossing the DS edge: the effects of traffic conditioning on the target group

This effect is quantified by the relationship of the emerging traffic aggregate to the entering target group. That relationship can depend on the arriving traffic pattern as well as the configuration of the traffic conditioners. For example, if the EF PHB [RFC2598] and a strict policer of rate R are associated with the foo PDB, then the first part of characterizing the foo PDB is to write the relationship between the arriving target packets and the departing foo traffic aggregate. In this case, "the rate of the emerging foo traffic aggregate is less than or equal to the smaller of R and the arrival rate of the target group of packets" and additional temporal characteristics of the packets (e.g., burst) may be specified as desired. Thus, there is a "loss rate" on the arriving target group that results from sending too much traffic or the traffic with the wrong temporal characteristics. This loss rate should be entirely preventable (or controllable) by the upstream sender conforming to the traffic conditioning associated with the PDB specification.

The issue of "who is in control" of the loss (or demotion) rate helps to clearly delineate this component of PDB performance from that associated with transiting the domain. The latter is completely under control of the operator of the DS domain and the former is used to ensure that the entering traffic aggregate conforms to the traffic profile to which the operator has provisioned the network. Further, the effects of traffic conditioning on the target group can usually be expressed more simply than the effects of transiting the DS domain on the traffic aggregate’s traffic profile.

A PDB may also apply traffic conditioning at DS domain egress. The effect of this conditioning on the overall PDB attributes would be treated similarly to the ingress characteristics (the authors may develop more text on this in the future, but it does not materially affect the ideas presented in this document.)

4.1.2 Crossing the DS domain: transit effects

The second component of PDB performance is the metrics that characterize the transit of a packet of the PDB’s traffic aggregate between any two edges of the DS domain boundary shown in figure 3. Note that the DS domain boundary runs through the DS boundary routers since the traffic aggregate is generally formed in the boundary router before the packets are queued and scheduled for output. (In most cases, this distinction is expected to be important.)

DSCPs should not change in the interior of a DS domain as there is no traffic conditioning being applied. If it is necessary to reapply the kind of traffic conditioning that could result in remarking, there should be a DS domain boundary at that point, though such an "interior" boundary can have "lighter weight" rules in its TCA. Thus, when measuring attributes between locations as indicated in figure 3, the DSCP at the egress side can be assumed to have held throughout the domain.

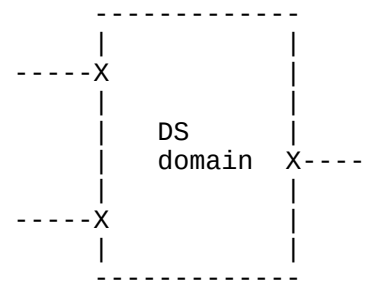

Figure 3: Range of applicability of attributes of a traffic aggregate associated with a PDB (is between the points marked "X")

Though a DS domain may be as small as a single node, more complex topologies are expected to be the norm, thus the PDB definition must hold as its traffic aggregate is split and merged on the interior links of a DS domain. Packet flow in a network is not part of the PDB definition; the application of traffic conditioning as packets enter the DS domain and the consistent PHB through the DS domain must suffice. A PDB's definition does not have to hold for arbitrary topologies of networks, but the limits on the range of applicability for a specific PDB must be clearly specified.

In general, a PDB operates between N ingress points and M egress points at the DS domain boundary. Even in the degenerate case where N=M=1, PDB attributes are more complex than the definition of PHB attributes since the concatenation of the behavior of intermediate nodes affects the former. A complex case with N and M both greater than one involves splits and merges in the traffic path and is non-trivial to analyze. Analytic, simulation, and experimental work will all be necessary to understand even the simplest PDBs.

4.2 Constructing PDBs

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 traffic aggregate. Once the

interior routers are configured for the number of distinct traffic aggregates that the network will handle, each PDB's allocation at the edge comes from meeting the desired performance goals for the PDB's traffic aggregate subject to that configuration of packet schedulers and bandwidth capacity. The configuration of traffic conditioners at the edge may be altered by provisioning or admission control but the decision about which PDB to use and how to apply classification and traffic conditioning comes from matching performance to goals.

For example, consider the DS domain of figure 3. A PDB with an explicit bound on loss must apply traffic conditioning at the boundary to ensure that on the average no more packets are admitted than can emerge. Though, queueing internal to the network may result in a difference between input and output traffic over some timescales, the averaging timescale should not exceed what might be expected for reasonably sized buffering inside the network. Thus if bursts are allowed to arrive into the interior of the network, there must be enough capacity to ensure that losses don't exceed the bound. Note that explicit bounds on the loss level can be particularly difficult as the exact way in which packets merge inside the network affects the burstiness of the PDB's traffic aggregate and hence, loss.

PHBs give explicit expressions of the treatment a traffic aggregate can expect at each hop. For a PDB, this behavior must apply to merging and diverging traffic aggregates, thus characterizing a PDB requires understanding what happens to a PHB under aggregation. That is, PHBs recursively applied must result in a known behavior. As an example, since maximum burst sizes grow with the number of microflows or traffic aggregate streams merged, a PDB specification must address this. A clear advantage of constructing behaviors that aggregate is the ease of concatenating PDBs so that the associated traffic aggregate has known attributes that span interior DS domains and, eventually, farther. For example, in figure 1 assume that we have configured the foo PDB on the interior DS domains of AS2. Then traffic aggregates associated with the foo PDB in each interior DS domain of AS2 can be merged at the shaded interior boundary routers. If the same (or fewer) traffic conditioners as applied at the entrance to AS2 are applied at these interior boundaries, the attributes of the foo PDB should continue to be used to quantify the expected behavior. Explicit expressions of what happens to the behavior under aggregation, possibly parameterized by node in-degrees or network diameters, are necessary to determine what to do at the internal aggregation points. One approach might be to completely reapply the traffic conditioning at these points; another might employ some limited rate-based remarking only.

Multiple PDBs may use the same PHB. The specification of a PDB can contain a list of PHBs and their required configuration, all of which would result in the same PDB. In operation, it is expected that a single domain will use a single PHB to implement a particular PDB, though different domains may select different PHBs. Recall that in the diffserv definition [RFC2474], a single PHB might be selected within a domain by a list of DSCPs. Multiple PDBs might use the same PHB in which case the transit performance of traffic aggregates of these PDBs will, of necessity, be the same. Yet, the particular characteristics that the PDB designer wishes to claim as attributes may vary, so two PDBs that use the same PHB might not be specified with the same list of attributes.

The specification of the transit expectations of PDBs 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 by making it possible to hide details of a domain's internals while exposing characteristics necessary for QoS.

4.3 PDBs using PHB Groups

The use of PHB groups to construct PDBs can be done in several ways. A single PHB member of a PHB group might be used to construct a single PDB. For example, a PDB could be defined using just one of the Class Selector Compliant PHBs [RFC2474]. The traffic conditioning for that PDB and the required configuration of the particular PHB would be defined in such a way that there was no dependence or relationship with the manner in which other PHBs of the group are used or, indeed, whether they are used in that DS domain. In this case, the reasonable approach would be to specify this PDB alone in a document which expressly called out the conditions and configuration of the Class Selector PHB required.

A single PDB can be constructed using more than one PHB from the same PHB group. For example, the traffic conditioner described in RFC 2698 might be used to mark a particular entering traffic aggregate for one of the three AF1x PHBs [RFC2597] while the transit performance of the resultant PDB is specified, statistically, across all the packets marked with one of those PHBs.

A set of related PDBs might be defined using a PHB group. In this case, the related PDBs should be defined in the same document. This is appropriate when the traffic conditioners that create the traffic aggregates associated with each PDB have some relationships and interdependencies such that the traffic aggregates for these PDBs should be described and characterized together. The transit attributes will depend on the PHB associated with the PDB and will not be the same for all PHBs in the group, though there may be some

parameterized interrelationship between the attributes of each of these PDBs. In this case, each PDB should have a clearly separate description of its transit attributes (delineated in a separate subsection) within the document. For example, the traffic conditioner described in RFC 2698 might be used to mark arriving packets for three different AF1x PHBs, each of which is to be treated as a separate traffic aggregate in terms of transit properties. Then a single document could be used to define and quantify the relationship between the arriving packets and the emerging traffic aggregates as they relate to one another. The transit characteristics of packets of each separate AF1x traffic aggregate should be described separately within the document.

Another way in which a PHB group might be used to create one PDB per PHB might have decoupled traffic conditioners, but some relationship between the PHBs of the group. For example, a set of PDBs might be defined using Class Selector Compliant PHBs [RFC2474] in such a way that the traffic conditioners that create the traffic aggregates are not related, but the transit performance of each traffic aggregate has some parametric relationship to the other. If it makes sense to specify them in the same document, then the author(s) should do so.

4.4 Forwarding path vs. control plane

A PDB's associated PHB, classifiers, and traffic conditioners are all in the packet forwarding path and operate at line rates. PHBs, classifiers, and traffic conditioners are configured in response to control plane activity which takes place across a range of time scales, but, even at the shortest time scale, control plane actions are not expected to happen per-packet. Classifiers and traffic conditioners at the DS domain boundary are used to enforce who gets to use the PDB and how the PDB should behave temporally. Reconfiguration of PHBs might occur monthly, quarterly, or only when the network is upgraded. Classifiers and traffic conditioners might be reconfigured at a few regular intervals during the day or might happen in response to signalling decisions thousands of times a day. Much of the control plane work is still evolving and is outside the charter of the Diffserv WG. We note that this is quite appropriate since the manner in which the configuration is done and the time scale at which it is done should not affect the PDB attributes.

5 Format for Specification of Diffserv Per-Domain Behaviors

PDBs arise from a particular relationship between edge and interior (which may be parameterized). The quantifiable characteristics of a PDB 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 PDB performs, but the act(s) of configuring the edge is a control plane action which can be separated from the specification of the PDB.

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

5.1 Applicability Statement

All PDB specs must have an applicability statement that outlines the intended use of this PDB and the limits to its use.

5.2 Technical specification

This section specifies the rules or guidelines to create this PDB, each distinguished with "may", "must" and "should." The technical specification must list the classification and traffic conditioning required (if any) and the PHB (or PHBs) to be used with any additional requirements on their configuration beyond that contained in RFCs. Classification can reflect the results of an admission control process. Traffic conditioning may include marking, traffic shaping, and policing. A Service Provisioning Policy might be used to describe the technical specification of a particular PDB.

5.3 Attributes

A PDB's attributes tell how it behaves under ideal conditions if configured in a specified manner (where the specification may be parameterized). These might include drop rate, throughput, delay bounds measured over some time period. They may be bounds, statistical bounds, or percentiles (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 as guidelines in creating a service specification rather than use them directly. For example, a "loss-free" PDB 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 PDB may be parameterized by network-specific features; for example, maximum number of hops, minimum bandwidth, total number of entry/exit points of the PDB to/from the diffserv network, maximum transit delay of network elements, minimum buffer size available for the PDB at a network node, etc.

5.5 Assumptions

In most cases, PDBs will be specified assuming lossless links, no link failures, and relatively stable routing. This is reasonable since otherwise it would be very difficult to quantify behavior and this is the operating conditions for which most operators strive. However, these assumptions must be clearly stated. Since PDBs with specific bandwidth parameters require that bandwidth to be available, the assumptions to be stated may include standby capacity. Some PDBs may be specifically targeted for cases where these assumptions do not hold, e.g., for high loss rate links, and such targeting must also be made explicit. If additional restrictions, especially specific traffic engineering measures, are required, these must be stated.

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

5.6 Example Uses

A PDB specification must give example uses to motivate the understanding of ways in which a diffserv network could make use of the PDB although these are not expected to be detailed. For example, "A bulk handling PDB 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 PDB by traffic policers."

5.7 Environmental Concerns (media, topology, etc.)

Note that it is not necessary for a provider to expose which PDB (if a commonly defined one) is being used nor is it necessary for a provider to specify a service by the PDB’s attributes. For example, a service provider might use a PDB 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.

5.8 Security Considerations for each PDB

This section should include any security considerations that are specific to the PDB. Is it subject to any unusual theft-of-service or denial-of-service attacks? Are any unusual security precautions needed?

It is not necessary to repeat the general security discussions in [RFC2474] and [RFC2475], but a reference should be included. Also refer to any special security considerations for the PHB or PHBs used.

6 On PDB Attributes

As discussed in section 4, measurable, quantifiable attributes associated with each PDB can be used to describe what will happen to packets using that PDB as they cross the domain. In its role as a building block for the construction of interdomain quality-of-service, a PDB specification should provide the answer to the question: Under what conditions can we join the output of this domain to another under the same traffic conditioning and expectations? Although there are many ways in which traffic might be distributed, creating quantifiable, realizable PDBs that can be concatenated into multi-domain services limits the realistic scenarios. A PDB's attributes with a clear statement of the conditions under which the attributes hold is critical to the composition of multi-domain services.

There is a clear correlation between the strictness of the traffic conditioning and the quality of the PDB's attributes. As indicated earlier, numerical bounds are likely to be statistical or expressed as a percentile. Parameters expressed as strict bounds will require very precise mathematical analysis, while those expressed statistically can to some extent rely on experiment. Section 7 gives the example of a PDB without strict traffic conditioning and concurrent work on a PDB with strict traffic conditioning and attributes is also in front of the WG [VW]. This section gives some general considerations for characterizing PDB attributes.

There are two ways to characterize PDBs with respect to time. First are properties over "long" time periods, or average behaviors. A PDB specification should report these as the rates or throughput seen over some specified time period. In addition, there are properties of "short" time behavior, usually expressed as the allowable burstiness in a traffic aggregate. The short time behavior is important in understanding buffering requirements (and associated loss characteristics) and for metering and conditioning considerations at DS boundaries. For short-time behavior, we are

interested primarily in two things: 1) how many back-to-back packets of the PDB's traffic aggregate will we see at any point (this would be metered as a burst) and 2) how large a burst of packets of this PDB's traffic aggregate can appear in a queue at once (gives queue overflow and loss). If other PDBs are using the same PHB within the domain, that must be taken into account.

6.1 Considerations in specifying long-term or average PDB attributes

To characterize the average or long-term behavior for the foo PDB we must explore a number of questions, for instance: Can the DS domain handle the average foo traffic flow? Is that answer topology dependent or are there some specific assumptions on routing which must hold for the foo PDB to preserve its "adequately provisioned" capability? In other words, if the topology of D changes suddenly, will the foo PDB's attributes change? Will its loss rate dramatically increase?

Let domain D in figure 4 be an ISP ringing the U.S. with links of bandwidth B and with N tails to various metropolitan areas. Inside D, if the link between the node connected to A and the node connected to Z goes down, all the foo traffic aggregate between the two nodes must transit the entire ring: Would the bounded behavior of the foo PDB change? If this outage results in some node of the ring now having a larger arrival rate to one of its links than the capacity of the link for foo's traffic aggregate, clearly the loss rate would change dramatically. In this case, topological assumptions were made about the path of the traffic from A to Z that affected the characteristics of the foo PDB. If these topological assumptions no longer hold, the loss rate of packets of the foo traffic aggregate transiting the domain could change; for example, a characteristic such as "loss rate no greater than 1% over any interval larger than 10 minutes." A PDB specification should spell out the assumptions made on preserving the attributes.

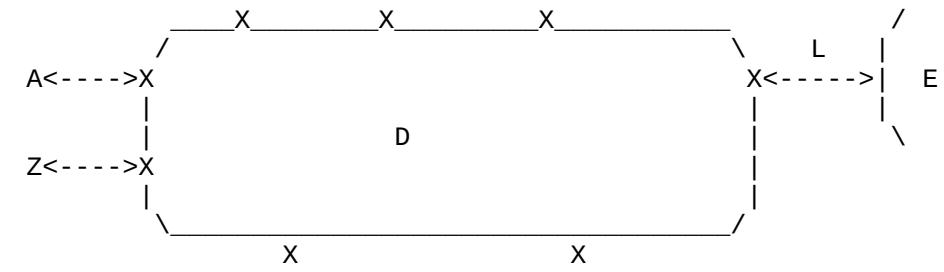

Figure 4: ISP and DS domain D connected in a ring and connected to DS domain E

6.2 Considerations in specifying short-term or bursty PDB attributes

Next, consider the short-time behavior of the traffic aggregate associated with a PDB, 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 conditions this will lead to properties that aggregate. In our example, if domain D allows each of the uplinks to burst p packets into the foo traffic aggregate, the bursts could accumulate as they transit the ring. Packets headed for link L can come from both directions of the ring and back-to-back packets from foo's traffic aggregate can 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 foo packets (required for bounding the loss) then the queue of foo packets for link L will empty before new packets arrive. If the bandwidth of L is equal to B, one foo packet must queue while the other is transmitted. This would result in N x p back-to- back packets of this traffic aggregate arriving over L during the same time scale as the bursts of p were permitted on the uplinks. Thus, configuring the PDB so that link L can handle the sum of the rates that ingress to the foo PDB doesn't guarantee that L can handle the sum of the N bursts into the foo PDB.

If the bandwidth of L is less than B, then the link must buffer Nxpx(B-L)/B foo packets to avoid loss. If the PDB 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 foo packets might arrive on each input. Then, in the absence of any additional traffic conditioning, it is possible that dxpx(# of uplinks) back-to-back foo packets can be sent across link L to domain E. Thus the DS domain E must permit these much larger bursts into the foo PDB than domain D permits on the N uplinks or else the foo traffic aggregate must be made to conform to the TCA for entering E (e.g., by shaping).

What conditions should be imposed on a PDB and on the associated PHB in order to ensure PDBs can be concatenated, as across the interior DS domains of figure 1? Traffic conditioning for constructing a PDB that has certain attributes across a DS domain should apply independently of the origin of the packets. With reference to the

example we've been exploring, the TCA for the PDB's traffic aggregate entering link L into domain E should not depend on the number of uplinks into domain D.

6.3 Remarks

This section has been provided as motivational food for thought for PDB specifiers. It is by no means an exhaustive catalog of possible PDB attributes or what kind of analysis must be done. We expect this to be an interesting and evolutionary part of the work of understanding and deploying differentiated services in the Internet. There is a potential for much interesting research work. However, in submitting a PDB specification to the Diffserv WG, a PDB must also meet the test of being useful and relevant by a deployment experience, described in section 8.

7 A Reference Per-Domain Behavior

The intent of this section is to define as a reference a Best Effort PDB, a PDB that has little in the way of rules or expectations.

7.1 Best Effort PDB

7.1.1 Applicability

A Best Effort (BE) PDB is for sending "normal internet traffic" across a diffserv network. That is, the definition and use of this PDB 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. Although the PDB itself does not include bounds on availability, latency, and packet loss, this does not preclude Service Providers from engineering their networks so as to result in commercially viable bounds on services that utilize the BE PDB. This would be analogous to the Service Level Guarantees that are provided in today's single-service Internet.

In the present single-service commercial Internet, Service Level Guarantees for availability, latency, and packet delivery can be found on the web sites of ISPs [WCG, PSI, UU]. For example, a typical North American round-trip latency bound is 85 milliseconds, with each service provider's site information specifying the method of measurement of the bounds and the terms associated with these bounds contractually.

7.1.2 TCS and PHB configurations

There are no restrictions governing rate and bursts of packets beyond the limits imposed by the ingress link. The network edge ensures that packets using the PDB are marked for the Default PHB (as defined in [RFC2474]), but no other traffic conditioning is required. Interior network nodes apply the Default PHB on these packets.

7.1.3 Attributes of this PDB

"As much as possible as soon as possible".

Packets of this PDB will not be completely starved and when resources are available (i.e., not required by packets from any other traffic aggregate), network elements should be configured to permit packets of this PDB to consume them.

Network operators may bound the delay and loss rate for services constructed from this PDB given knowledge about their network, but such attributes are not part of the definition.

7.1.4 Parameters

None.

7.1.5 Assumptions

A properly functioning network, i.e., packets may be delivered from any ingress to any egress.

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.

3. For standard domestic consumer connections

7.1.7 Environmental Concerns

There are no environmental concerns specific to this PDB.

7.1.8 Security Considerations for BE PDB

There are no specific security exposures for this PDB. See the general security considerations in [RFC2474] and [RFC2475].

8 Guidelines for writing PDB specifications

G1. Following the format given in this document, write a draft and submit it as an Internet Draft. The document should have "diffserv" as some part of the name. Either as an appendix to the draft, or in a separate document, provide details of deployment experience with measured results on a network of non-trivial size carrying realistic traffic and/or convincing simulation results (simulation of a range of modern traffic patterns and network topologies as applicable). The document should be brought to the attention of the diffserv WG mailing list, if active.

G2. Initial discussion should focus primarily on the merits of the PDB, though comments and questions on the claimed attributes are reasonable. This is in line with the Differentiated Services goal to put relevance before academic interest in the specification of PDBs. Academically interesting PDBs are encouraged, but would be more appropriate for technical publications and conferences, not for submission to the IETF. (An "academically interesting" PDB might become a PDB of interest for deployment over time.)

The implementation of the following guidelines varies, depending on whether there is an active diffserv working group or not.

Active Diffserv Working Group path:

G3. Once consensus has been reached on a version of a draft that it is a useful PDB 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.

G4. If/when endorsed by the panel, that draft goes to WG last call. If not endorsed, the author(s) can give an itemized response to the panel's report and ask for a WG Last Call.

G5. If/when passes Last Call, goes to ADs for publication as a WG Informational RFC in our "PDB series".

If no active Diffserv Working Group exists:

G3. Following discussion on relevant mailing lists, the authors should revise the Internet Draft and contact the IESG for "Expert Review" as defined in section 2 of RFC 2434 [RFC2434].

G4. Subsequent to the review, the IESG may recommend publication of the Draft as an RFC, request revisions, or decline to publish as an Informational RFC in the "PDB series".

9 Security Considerations

The general security considerations of [RFC2474] and [RFC2475] apply to all PDBs. Individual PDB definitions may require additional security considerations.

10 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. Grenville Armitage coined "per domain behavior (PDB)" though some have suggested similar terms prior to that. Dan Grossman, Bob Enger, Jung-Bong Suk, and John Dullaert reviewed the document and commented so as to improve its form.

References

[RFC2474] Nichols, K., Blake, S. Baker, F. and D. Black, "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers", RFC 2474, December 1998.

[RFC2475] Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z. and W. Weiss, "An Architecture for Differentiated Services", December 1998.

[RFC2597] Heinanen, J., Baker, F., Weiss, W. and J. Wroclawski, "Assured Forwarding PHB Group", RFC 2597, June 1999.

[RFC2598] Jacobson, V., Nichols, K. and K. Poduri, "An Expedited Forwarding PHB", RFC 2598, June 1999.

[RFC2698] Heinanen, J. and R. Geurin, "A Two Rate Three Color Marker", RFC 2698, June 1999.

[MODEL] Bernet, Y., Blake, S., Grossman, D. and A. Smith, "An Informal Management Model for Diffserv Routers", Work in Progress.

[MIB] Baker, F., Chan, K. and A. Smith, "Management Information Base for the Differentiated Services Architecture", Work in Progress.

[VW] Jacobson, V., Nichols, K. and K. Poduri, "The 'Virtual Wire' Per-Domain Behavior", Work in Progress.

[WCG] Worldcom, "Internet Service Level Guarantee", http://www.worldcom.com/terms/service_level_guarantee/t_sla_terms.phtml

[PSI] PSINet, "Service Level Agreements", http://www.psinet.com/sla/

[UU] UUNET USA Web site, "Service Level Agreements", http://www.us.uu.net/support/sla/

[RFC2434] Alvestrand, H. and T. Narten, "Guidelines for IANA Considerations", BCP 26, RFC 2434, October 1998.

Authors' Addresses

Kathleen Nichols
Packet Design, LLC
2465 Latham Street, Third Floor
Mountain View, CA 94040
USA

EMail: nichols@packetdesign.com

Brian Carpenter
IBM
c/o iCAIR
Suite 150
1890 Maple Avenue
Evanston, IL 60201
USA

EMail: brian@icair.org

Full Copyright Statement

Copyright (C) The Internet Society (2001). All Rights Reserved.

This document and translations of it may be copied and furnished to others, and derivative works that comment on or otherwise explain it or assist in its implementation may be prepared, copied, published and distributed, in whole or in part, without restriction of any kind, provided that the above copyright notice and this paragraph are included on all such copies and derivative works. However, this document itself may not be modified in any way, such as by removing the copyright notice or references to the Internet Society or other Internet organizations, except as needed for the purpose of developing Internet standards in which case the procedures for copyrights defined in the Internet Standards process must be followed, or as required to translate it into languages other than English.

The limited permissions granted above are perpetual and will not be revoked by the Internet Society or its successors or assigns.

This document and the information contained herein is provided on an "AS IS" basis and THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIMS ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Acknowledgement

Funding for the RFC Editor function is currently provided by the Internet Society.