Network Working Group
Request for Comments: 2474
Obsoletes: 1455, 1349
Category: Standards Track

K. Nichols
Cisco Systems
S. Blake
Torrent Networking Technologies
F. Baker
Cisco Systems
D. Black
EMC Corporation

December 1998

Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers

Status of this Memo

Copyright Notice

Abstract

Differentiated services enhancements to the Internet protocol are intended to enable scalable service discrimination in the Internet without the need for per-flow state and signaling at every hop. A variety of services may be built from a small, well-defined set of building blocks which are deployed in network nodes. The services may be either end-to-end or intra-domain; they include both those that can satisfy quantitative performance requirements (e.g., peak bandwidth) and those based on relative performance (e.g., "class" differentiation). Services can be constructed by a combination of:

- setting bits in an IP header field at network boundaries (autonomous system boundaries, internal administrative boundaries, or hosts),
- using those bits to determine how packets are forwarded by the nodes inside the network, and
- conditioning the marked packets at network boundaries in accordance with the requirements or rules of each service.

The requirements or rules of each service must be set through administrative policy mechanisms which are outside the scope of this document. A differentiated services-compliant network node includes a classifier that selects packets based on the value of the DS field, along with buffer management and packet scheduling mechanisms capable of delivering the specific packet forwarding treatment indicated by the DS field value. Setting of the DS field and conditioning of the temporal behavior of marked packets need only be performed at network boundaries and may vary in complexity.

This document defines the IP header field, called the DS (for differentiated services) field. In IPv4, it defines the layout of the TOS octet; in IPv6, the Traffic Class octet. In addition, a base set of packet forwarding treatments, or per-hop behaviors, is defined.

For a more complete understanding of differentiated services, see also the differentiated services architecture [ARCH].

Table of Contents

1. Introduction

Differentiated services are intended to provide a framework and building blocks to enable deployment of scalable service discrimination in the Internet. The differentiated services approach aims to speed deployment by separating the architecture into two major components, one of which is fairly well-understood and the other of which is just beginning to be understood. In this, we are guided by the original design of the Internet where the decision was made to separate the forwarding and routing components. Packet forwarding is the relatively simple task that needs to be performed on a per-packet basis as quickly as possible. Forwarding uses the packet header to find an entry in a routing table that determines the packet's output interface. Routing sets the entries in that table and may need to reflect a range of transit and other policies as well as to keep track of route failures. Routing tables are maintained as a background process to the forwarding task. Further, routing is the more complex task and it has continued to evolve over the past 20 years.

Analogously, the differentiated services architecture contains two main components. One is the fairly well-understood behavior in the forwarding path and the other is the more complex and still emerging background policy and allocation component that configures parameters used in the forwarding path. The forwarding path behaviors include the differential treatment an individual packet receives, as implemented by queue service disciplines and/or queue management disciplines. These per-hop behaviors are useful and required in network nodes to deliver differentiated treatment of packets no matter how we construct end-to-end or intra-domain services. Our focus is on the general semantics of the behaviors rather than the specific mechanisms used to implement them since these behaviors will evolve less rapidly than the mechanisms.

Per-hop behaviors and mechanisms to select them on a per-packet basis can be deployed in network nodes today and it is this aspect of the differentiated services architecture that is being addressed first. In addition, the forwarding path may require that some monitoring, policing, and shaping be done on the network traffic designated for "special" treatment in order to enforce requirements associated with the delivery of the special treatment. Mechanisms for this kind of traffic conditioning are also fairly well-understood. The wide deployment of such traffic conditioners is also important to enable the construction of services, though their actual use in constructing services may evolve over time.

The configuration of network elements with respect to which packets get special treatment and what kinds of rules are to be applied to the use of resources is much less well-understood. Nevertheless, it is possible to deploy useful differentiated services in networks by using simple policies and static configurations. As described in [ARCH], there are a number of ways to compose per-hop behaviors and traffic conditioners to create services. In the process, additional experience is gained that will guide more complex policies and allocations. The basic behaviors in the forwarding path can remain the same while this component of the architecture evolves. Experiences with the construction of such services will continue for some time, thus we avoid standardizing this construction as it is premature. Further, much of the details of service construction are covered by legal agreements between different business entities and we avoid this as it is very much outside the scope of the IETF.

This document concentrates on the forwarding path component. In the packet forwarding path, differentiated services are realized by mapping the codepoint contained in a field in the IP packet header to a particular forwarding treatment, or per-hop behavior (PHB), at each network node along its path. The codepoints may be chosen from a set of mandatory values defined later in this document, from a set of recommended values to be defined in future documents, or may have purely local meaning. PHBs are expected to be implemented by employing a range of queue service and/or queue management disciplines on a network node's output interface queue: for example weighted round-robin (WRR) queue servicing or drop-preference queue management.

Marking is performed by traffic conditioners at network boundaries, including the edges of the network (first-hop router or source host) and administrative boundaries. Traffic conditioners may include the primitives of marking, metering, policing and shaping (these mechanisms are described in [ARCH]). Services are realized by the use of particular packet classification and traffic conditioning mechanisms at boundaries coupled with the concatenation of per-hop behaviors along the transit path of the traffic. A goal of the differentiated services architecture is to specify these building blocks for future extensibility, both of the number and type of the building blocks and of the services built from them.

Terminology used in this memo is defined in Sec. 2. The differentiated services field definition (DS field) is given in Sec. 3. In Sec. 4, we discuss the desire for partial backwards compatibility with current use of the IPv4 Precedence field. As a solution, we introduce Class Selector Codepoints and Class Selector

Compliant PHBs. Sec. 5 presents guidelines for per-hop behavior standardization. Sec. 6 discusses guidelines for allocation of codepoints. Sec. 7 covers security considerations.

This document is a concise description of the DS field and its uses. It is intended to be read along with the differentiated services architecture [ARCH].

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

2. Terminology Used in This Document

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.

Classifier: an entity which selects packets based on the content of packet headers according to defined rules.

Class Selector Codepoint: any of the eight codepoints in the range 'xxx000' (where 'x' may equal '0' or '1'). Class Selector Codepoints are discussed in Sec. 4.2.2.

Class Selector Compliant PHB: a per-hop behavior satisfying the Class Selector PHB Requirements specified in Sec. 4.2.2.2.

Codepoint: a specific value of the DSCP portion of the DS field. Recommended codepoints SHOULD map to specific, standardized PHBs. Multiple codepoints MAY map to the same PHB.

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.

Differentiated Services-Compliant: in compliance with the requirements specified in this document. Also DS-compliant.

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 Field: the IPv4 header TOS octet or the IPv6 Traffic Class octet when interpreted in conformance with the definition given in this document. Also DS field.

Mechanism: The implementation of one or more per-hop behaviors according to a particular algorithm.

Microflow: a single instance of an application-to-application flow of packets which is identified by source address, destination address, protocol id, and source port, destination port (where applicable).

Per-hop Behavior (PHB): a description of the externally observable forwarding treatment applied at a differentiated services-compliant node to a behavior aggregate. The description of a PHB SHOULD be sufficiently detailed to allow the construction of predictable services, as documented in [ARCH].

Per-hop Behavior Group: a set of one or more PHBs that can only be meaningfully specified and implemented simultaneously, due to a common constraint applying to all PHBs in the set such as a queue servicing or queue management policy. Also PHB Group.

Traffic Conditioning: control functions that can be applied to a behavior aggregate, application flow, or other operationally useful subset of traffic, e.g., routing updates. These MAY include metering, policing, shaping, and packet marking. Traffic conditioning is used to enforce agreements between domains and to condition traffic to receive a differentiated service within a domain by marking packets with the appropriate codepoint in the DS field and by monitoring and altering the temporal characteristics of the aggregate where necessary. See [ARCH].

Traffic Conditioner: an entity that performs traffic conditioning functions and which MAY contain meters, policers, shapers, and markers. Traffic conditioners are typically deployed in DS boundary nodes (i.e., not in interior nodes of a DS domain).

Service: a description of the overall treatment of (a subset of) a customer’s traffic across a particular domain, across a set of interconnected DS domains, or end-to-end. Service descriptions are covered by administrative policy and services are constructed by

applying traffic conditioning to create behavior aggregates which experience a known PHB at each node within the DS domain. Multiple services can be supported by a single per-hop behavior used in concert with a range of traffic conditioners.

To summarize, classifiers and traffic conditioners are used to select which packets are to be added to behavior aggregates. Aggregates receive differentiated treatment in a DS domain and traffic conditioners MAY alter the temporal characteristics of the aggregate to conform to some requirements. A packet's DS field is used to designate the packet's behavior aggregate and is subsequently used to determine which forwarding treatment the packet receives. A behavior aggregate classifier which can select a PHB, for example a differential output queue servicing discipline, based on the codepoint in the DS field SHOULD be included in all network nodes in a DS domain. The classifiers and traffic conditioners at DS boundaries are configured in accordance with some service specification, a matter of administrative policy outside the scope of this document.

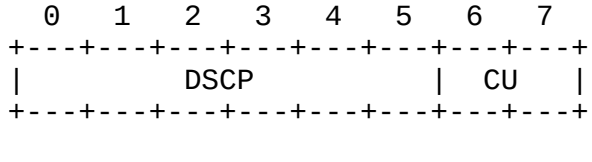
Additional differentiated services definitions are given in [ARCH].

3. Differentiated Services Field Definition

A replacement header field, called the DS field, is defined, which is intended to supersede the existing definitions of the IPv4 TOS octet [RFC791] and the IPv6 Traffic Class octet [IPv6].

Six bits of the DS field are used as a codepoint (DSCP) to select the PHB a packet experiences at each node. A two-bit currently unused (CU) field is reserved and its definition and interpretation are outside the scope of this document. The value of the CU bits are ignored by differentiated services-compliant nodes when determining the per-hop behavior to apply to a received packet.

The DS field structure is presented below:



DSCP: differentiated services codepoint
CU: currently unused

In a DSCP value notation 'xxxxxx' (where 'x' may equal '0' or '1') used in this document, the left-most bit signifies bit 0 of the DS field (as shown above), and the right-most bit signifies bit 5.

Implementors should note that the DSCP field is six bits wide. DS-compliant nodes MUST select PHBs by matching against the entire 6-bit DSCP field, e.g., by treating the value of the field as a table index which is used to select a particular packet handling mechanism which has been implemented in that device. The value of the CU field MUST be ignored by PHB selection. The DSCP field is defined as an unstructured field to facilitate the definition of future per-hop behaviors.

With some exceptions noted below, the mapping of codepoints to PHBs MUST be configurable. A DS-compliant node MUST support the logical equivalent of a configurable mapping table from codepoints to PHBs. PHB specifications MUST include a recommended default codepoint, which MUST be unique for codepoints in the standard space (see Sec. 6). Implementations should support the recommended codepoint-to-PHB mappings in their default configuration. Operators may choose to use different codepoints for a PHB, either in addition to or in place of the recommended default. Note that if operators do so choose, re-marking of DS fields may be necessary at administrative boundaries even if the same PHBs are implemented on both sides of the boundary.

See [ARCH] for further discussion of re-marking.

The exceptions to general configurability are for codepoints 'xxx000' and are noted in Secs. 4.2.2 and 4.3.

Packets received with an unrecognized codepoint SHOULD be forwarded as if they were marked for the Default behavior (see Sec. 4), and their codepoints should not be changed. Such packets MUST NOT cause the network node to malfunction.

The structure of the DS field shown above is incompatible with the existing definition of the IPv4 TOS octet in [RFC791]. The presumption is that DS domains protect themselves by deploying re-marking boundary nodes, as should networks using the RFC 791 Precedence designations. Correct operational procedure SHOULD follow [RFC791], which states: "If the actual use of these precedence designations is of concern to a particular network, it is the responsibility of that network to control the access to, and use of, those precedence designations." Validating the value of the DS field at DS boundaries is sensible in any case since an upstream node can easily set it to any arbitrary value. DS domains that are not isolated by suitably configured boundary nodes may deliver unpredictable service.

Nodes MAY rewrite the DS field as needed to provide a desired local or end-to-end service. Specifications of DS field translations at DS boundaries are the subject of service level agreements between providers and users, and are outside the scope of this document. Standardized PHBs allow providers to build their services from a well-known set of packet forwarding treatments that can be expected to be present in the equipment of many vendors.

4. Historical Codepoint Definitions and PHB Requirements

The DS field will have a limited backwards compatibility with current practice, as described in this section. Backwards compatibility is addressed in two ways. First, there are per-hop behaviors that are already in widespread use (e.g., those satisfying the IPv4 Precedence queueing requirements specified in [RFC1812]), and we wish to permit their continued use in DS-compliant nodes. In addition, there are some codepoints that correspond to historical use of the IP Precedence field and we reserve these codepoints to map to PHBs that meet the general requirements specified in Sec. 4.2.2.2, though the specific differentiated services PHBs mapped to by those codepoints MAY have additional specifications.

No attempt is made to maintain backwards compatibility with the "DTR" or TOS bits of the IPv4 TOS octet, as defined in [RFC791].

4.1 A Default PHB

A "default" PHB MUST be available in a DS-compliant node. This is the common, best-effort forwarding behavior available in existing routers as standardized in [RFC1812]. When no other agreements are in place, it is assumed that packets belong to this aggregate. Such packets MAY be sent into a network without adhering to any particular rules and the network will deliver as many of these packets as possible and as soon as possible, subject to other resource policy constraints. A reasonable implementation of this PHB would be a queueing discipline that sends packets of this aggregate whenever the output link is not required to satisfy another PHB. A reasonable policy for constructing services would ensure that the aggregate was not "starved". This could be enforced by a mechanism in each node that reserves some minimal resources (e.g, buffers, bandwidth) for Default behavior aggregates. This permits senders that are not differentiated services-aware to continue to use the network in the same manner as today. The impact of the introduction of differentiated services into a domain on the service expectations of its customers and peers is a complex matter involving policy decisions by the domain and is outside the scope of this document. The RECOMMENDED codepoint for the Default PHB is the bit pattern ’000000’; the value ’000000’ MUST map to a PHB that meets these

specifications. The codepoint chosen for Default behavior is compatible with existing practice [RFC791]. Where a codepoint is not mapped to a standardized or local use PHB, it SHOULD be mapped to the Default PHB.

A packet initially marked for the Default behavior MAY be re-marked with another codepoint as it passes a boundary into a DS domain so that it will be forwarded using a different PHB within that domain, possibly subject to some negotiated agreement between the peering domains.

4.2 Once and Future IP Precedence Field Use

We wish to maintain some form of backward compatibility with present uses of the IP Precedence Field: bits 0-2 of the IPv4 TOS octet. Routers exist that use the IP Precedence field to select different per-hop forwarding treatments in a similar way to the use proposed here for the DSCP field. Thus, a simple prototype differentiated services architecture can be quickly deployed by appropriately configuring these routers. Further, IP systems today understand the location of the IP Precedence field, and thus if these bits are used in a similar manner as DS-compliant equipment is deployed, significant failures are not likely during early deployment. In other words, strict DS-compliance need not be ubiquitous even within a single service provider's network if bits 0-2 of the DSCP field are employed in a manner similar to, or subsuming, the deployed uses of the IP Precedence field.

4.2.1 IP Precedence History and Evolution in Brief

The IP Precedence field is something of a forerunner of the DS field. IP Precedence, and the IP Precedence Field, were first defined in [RFC791]. The values that the three-bit IP Precedence Field might take were assigned to various uses, including network control traffic, routing traffic, and various levels of privilege. The least level of privilege was deemed "routine traffic". In [RFC791], the notion of Precedence was defined broadly as "An independent measure of the importance of this datagram." Not all values of the IP Precedence field were assumed to have meaning across boundaries, for instance "The Network Control precedence designation is intended to be used within a network only. The actual use and control of that designation is up to each network." [RFC791]

Although early BBN IMPs implemented the Precedence feature, early commercial routers and UNIX IP forwarding code generally did not. As networks became more complex and customer requirements grew, commercial router vendors developed ways to implement various kinds of queueing services including priority queueing, which were

generally based on policies encoded in filters in the routers, which examined IP addresses, IP protocol numbers, TCP or UDP ports, and other header fields. IP Precedence was and is among the options such filters can examine.

In short, IP Precedence is widely deployed and widely used, if not in exactly the manner intended in [RFC791]. This was recognized in [RFC1122], which states that while the use of the IP Precedence field is valid, the specific assignment of the priorities in [RFC791] were merely historical.

4.2.2 Subsuming IP Precedence into Class Selector Codepoints

A specification of the packet forwarding treatments selected by the IP Precedence field today would have to be quite general; probably not specific enough to build predictable services from in the differentiated services framework. To preserve partial backwards compatibility with known current uses of the IP Precedence field without sacrificing future flexibility, we have taken the approach of describing minimum requirements on a set of PHBs that are compatible with most of the deployed forwarding treatments selected by the IP Precedence field. In addition, we give a set of codepoints that MUST map to PHBs meeting these minimum requirements. The PHBs mapped to by these codepoints MAY have a more detailed list of specifications in addition to the required ones stated here. Other codepoints MAY map to these same PHBs. We refer to this set of codepoints as the Class Selector Codepoints, and the minimum requirements for PHBs that these codepoints may map to are called the Class Selector PHB Requirements.

4.2.2.1 The Class Selector Codepoints

A specification of the packet forwarding treatments selected by the The DS field values of ’xxx000|xx’, or DSCP = ’xxx000’ and CU subfield unspecified, are reserved as a set of Class Selector Codepoints. PHBs which are mapped to by these codepoints MUST satisfy the Class Selector PHB requirements in addition to preserving the Default PHB requirement on codepoint ’000000’ (Sec. 4.1).

4.2.2.2 The Class Selector PHB Requirements

We refer to a Class Selector Codepoint with a larger numerical value than another Class Selector Codepoint as having a higher relative order while a Class Selector Codepoint with a smaller numerical value than another Class Selector Codepoint is said to have a lower relative order. The set of PHBs mapped to by the eight Class Selector Codepoints MUST yield at least two independently forwarded classes of traffic, and PHBs selected by a Class Selector Codepoint

SHOULD give packets a probability of timely forwarding that is not lower than that given to packets marked with a Class Selector codepoint of lower relative order, under reasonable operating conditions and traffic loads. A discarded packet is considered to be an extreme case of untimely forwarding. In addition, PHBs selected by codepoints '11x000' MUST give packets a preferential forwarding treatment by comparison to the PHB selected by codepoint '000000' to preserve the common usage of IP Precedence values '110' and '111' for routing traffic.

Further, PHBs selected by distinct Class Selector Codepoints SHOULD be independently forwarded; that is, packets marked with different Class Selector Codepoints MAY be re-ordered. A network node MAY enforce limits on the amount of the node's resources that can be utilized by each of these PHBs.

PHB groups whose specification satisfy these requirements are referred to as Class Selector Compliant PHBs.

The Class Selector PHB Requirements on codepoint '000000' are compatible with those listed for the Default PHB in Sec. 4.1.

4.2.2.3 Using the Class Selector PHB Requirements for IP Precedence Compatibility

A DS-compliant network node can be deployed with a set of one or more Class Selector Compliant PHB groups. This document states that the set of codepoints 'xxx000' MUST map to such a set of PHBs. As it is also possible to map multiple codepoints to the same PHB, the vendor or the network administrator MAY configure the network node to map codepoints to PHBs irrespective of bits 3-5 of the DSCP field to yield a network that is compatible with historical IP Precedence use. Thus, for example, codepoint '011010' would map to the same PHB as codepoint '011000'.

4.2.2.4 Example Mechanisms for Implementing Class Selector Compliant PHB Groups

Class Selector Compliant PHBs can be realized by a variety of mechanisms, including strict priority queueing, weighted fair queueing (WFQ), WRR, or variants [RPS, HPFQA, DRR], or Class-Based Queuing [CBQ]. The distinction between PHBs and mechanisms is described in more detail in Sec. 5.

It is important to note that these mechanisms might be available through other PHBs (standardized or not) that are available in a particular vendor's equipment. For example, future documents may standardize a Strict Priority Queueing PHB group for a set of

recommended codepoints. A network administrator might configure those routers to select the Strict Priority Queueing PHBs with codepoints 'xxx000' in conformance with the requirements of this document.

As a further example, another vendor might employ a CBQ mechanism in its routers. The CBQ mechanism could be used to implement the Strict Priority Queueing PHBs as well as a set of Class Selector Compliant PHBs with a wider range of features than would be available in a set of PHBs that did no more than meet the minimum Class Selector PHB requirements.

4.3 Summary

This document defines codepoints 'xxx000' as the Class Selector codepoints, where PHBs selected by these codepoints MUST meet the Class Selector PHB Requirements described in Sec. 4.2.2.2. This is done to preserve a useful level of backward compatibility with current uses of the IP Precedence field in the Internet without unduly limiting future flexibility. In addition, codepoint '000000' is used as the Default PHB value for the Internet and, as such, is not configurable. The remaining seven non-zero Class Selector codepoints are configurable only to the extent that they map to PHBs that meet the requirements in Sec. 4.2.2.2.

5. Per-Hop Behavior Standardization Guidelines

The behavioral characteristics of a PHB are to be standardized, and not the particular algorithms or the mechanisms used to implement them. A node may have a (possibly large) set of parameters that can be used to control how packets are scheduled onto an output interface (e.g., N separate queues with settable priorities, queue lengths, round-robin weights, drop algorithm, drop preference weights and thresholds, etc). To illustrate the distinction between a PHB and a mechanism, we point out that Class Selector Compliant PHBs might be implemented by several mechanisms, including: strict priority queueing, WFQ, WRR, or variants [HPFQA, RPS, DRR], or CBQ [CBQ], in isolation or in combination.

PHBs may be specified individually, or as a group (a single PHB is a special case of a PHB group). A PHB group usually consists of a set of two or more PHBs that can only be meaningfully specified and implemented simultaneously, due to a common constraint applying to each PHB within the group, such as a queue servicing or queue management policy. A PHB group specification SHOULD describe conditions under which a packet might be re-marked to select another PHB within the group. It is RECOMMENDED that PHB implementations do not introduce any packet re-ordering within a microflow. PHB group

specifications MUST identify any possible packet re-ordering implications which may occur for each individual PHB, and which may occur if different packets within a microflow are marked for different PHBs within the group.

Only those per-hop behaviors that are not described by an existing PHB standard, and have been implemented, deployed, and shown to be useful, SHOULD be standardized. Since current experience with differentiated services is quite limited, it is premature to hypothesize the exact specification of these per-hop behaviors.

Each standardized PHB MUST have an associated RECOMMENDED codepoint, allocated out of a space of 32 codepoints (see Sec. 6). This specification has left room in the codepoint space to allow for evolution, thus the defined space ('xxx000') is intentionally sparse.

Network equipment vendors are free to offer whatever parameters and capabilities are deemed useful or marketable. When a particular, standardized PHB is implemented in a node, a vendor MAY use any algorithm that satisfies the definition of the PHB according to the standard. The node's capabilities and its particular configuration determine the different ways that packets can be treated.

Service providers are not required to use the same node mechanisms or configurations to enable service differentiation within their networks, and are free to configure the node parameters in whatever way that is appropriate for their service offerings and traffic engineering objectives. Over time certain common per-hop behaviors are likely to evolve (i.e., ones that are particularly useful for implementing end-to-end services) and these MAY be associated with particular EXP/LU PHB codepoints in the DS field, allowing use across domain boundaries (see Sec. 6). These PHBs are candidates for future standardization.

It is RECOMMENDED that standardized PHBs be specified in accordance with the guidelines set out in [ARCH].

6. IANA Considerations

The DSCP field within the DS field is capable of conveying 64 distinct codepoints. The codepoint space is divided into three pools for the purpose of codepoint assignment and management: a pool of 32 RECOMMENDED codepoints (Pool 1) to be assigned by Standards Action as defined in [CONS], a pool of 16 codepoints (Pool 2) to be reserved for experimental or Local Use (EXP/LU) as defined in [CONS], and a pool of 16 codepoints (Pool 3) which are initially available for experimental or local use, but which should be preferentially

utilized for standardized assignments if Pool 1 is ever exhausted. The pools are defined in the following table (where 'x' refers to either '0' or '1'):

Pool	Codepoint space	Assignment Policy
1	xxxxx0	Standards Action
2	xxxx11	EXP/LU
3	xxxx01	EXP/LU (*)

(*) may be utilized for future Standards Action allocations as necessary

This document assigns eight RECOMMENDED codepoints ('xxx000') which are drawn from Pool 1 above. These codepoints MUST be mapped, not to specific PHBs, but to PHBs that meet "at least" the requirements set forth in Sec. 4.2.2.2 to provide a minimal level of backwards compatibility with IP Precedence as defined in [RFC791] and as deployed in some current equipment.

7. Security Considerations

This section considers security issues raised by the introduction of differentiated services, primarily the potential for denial-of-service attacks, and the related potential for theft of service by unauthorized traffic (Section 7.1). Section 7.2 addresses the operation of differentiated services in the presence of IPsec including its interaction with IPsec tunnel mode and other tunnelling protocols. See [ARCH] for more extensive treatment of the security concerns raised by the overall differentiated services architecture.

7.1 Theft and Denial of Service

The primary goal of differentiated services is to allow different levels of service to be provided for traffic streams on a common network infrastructure. A variety of techniques may be used to achieve this, but the end result will be that some packets receive different (e.g., better) service than others. The mapping of network traffic to the specific behaviors that result in different (e.g., better or worse) service is indicated primarily by the DS codepoint, and hence an adversary may be able to obtain better service by modifying the codepoint to values indicating behaviors used for enhanced services or by injecting packets with such codepoint values. Taken to its limits, such theft-of-service becomes a denial-of-service attack when the modified or injected traffic depletes the resources available to forward it and other traffic streams.

The defense against this class of theft- and denial-of-service attacks consists of the combination of traffic conditioning at DS domain boundaries with security and integrity of the network infrastructure within a DS domain. DS domain boundary nodes MUST ensure that all traffic entering the domain is marked with codepoint values appropriate to the traffic and the domain, remarking the traffic with new codepoint values if necessary. These DS boundary nodes are the primary line of defense against theft- and denial-of-service attacks based on modified codepoints, as success of any such attack indicates that the codepoints used by the attacking traffic were inappropriate. An important instance of a boundary node is that any traffic-originating node within a DS domain is the initial boundary node for that traffic. Interior nodes in a DS domain rely on DS codepoints to associate traffic with the forwarding PHBs, and are NOT REQUIRED to check codepoint values before using them. As a result, the interior nodes depend on the correct operation of the DS domain boundary nodes to prevent the arrival of traffic with inappropriate codepoints or in excess of provisioned levels that would disrupt operation of the domain.

7.2 IPsec and Tunnelling Interactions

The IPsec protocol, as defined in [ESP, AH], does not include the IP header's DS field in any of its cryptographic calculations (in the case of tunnel mode, it is the outer IP header's DS field that is not included). Hence modification of the DS field by a network node has no effect on IPsec's end-to-end security, because it cannot cause any IPsec integrity check to fail. As a consequence, IPsec does not provide any defense against an adversary's modification of the DS field (i.e., a man-in-the-middle attack), as the adversary's modification will also have no effect on IPsec's end-to-end security.

IPsec's tunnel mode provides security for the encapsulated IP header's DS field. A tunnel mode IPsec packet contains two IP headers: an outer header supplied by the tunnel ingress node and an encapsulated inner header supplied by the original source of the packet. When an IPsec tunnel is hosted (in whole or in part) on a differentiated services network, the intermediate network nodes operate on the DS field in the outer header. At the tunnel egress node, IPsec processing includes removing the outer header and forwarding the packet (if required) using the inner header. The IPsec protocol REQUIRES that the inner header's DS field not be changed by this decapsulation processing to ensure that modifications to the DS field cannot be used to launch theft- or denial-of-service attacks across an IPsec tunnel endpoint. This document makes no change to that requirement. If the inner IP header has not been processed by a DS boundary node for the tunnel egress node's DS

domain, the tunnel egress node is the boundary node for traffic exiting the tunnel, and hence MUST ensure that the resulting traffic has appropriate DS codepoints.

When IPsec tunnel egress decapsulation processing includes a sufficiently strong cryptographic integrity check of the encapsulated packet (where sufficiency is determined by local security policy), the tunnel egress node can safely assume that the DS field in the inner header has the same value as it had at the tunnel ingress node. An important consequence is that otherwise insecure links within a DS domain can be secured by a sufficiently strong IPsec tunnel. This analysis and its implications apply to any tunnelling protocol that performs integrity checks, but the level of assurance of the inner header's DS field depends on the strength of the integrity check performed by the tunnelling protocol. In the absence of sufficient assurance for a tunnel that may transit nodes outside the current DS domain (or is otherwise vulnerable), the encapsulated packet MUST be treated as if it had arrived at a boundary from outside the DS domain.

8. Acknowledgements

The authors would like to acknowledge the Differentiated Services Working Group for discussions which helped shape this document.

9. References

[AH] Kent, S. and R. Atkinson, "IP Authentication Header", RFC 2402, November 1998.

[ARCH] Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z. and W. Weiss, "An Architecture for Differentiated Services", RFC 2475, December 1998.

[CBQ] S. Floyd and V. Jacobson, "Link-sharing and Resource Management Models for Packet Networks", IEEE/ACM Transactions on Networking, Vol. 3 no. 4, pp. 365-386, August 1995.

[CONS] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", RFC 2434, October 1998.

[DRR] M. Shreedhar and G. Varghese, Efficient Fair Queueing using Deficit Round Robin", Proc. ACM SIGCOMM 95, 1995.

[ESP] Kent, S. and R. Atkinson, "IP Encapsulating Security Payload (ESP)", RFC 2406, November 1998.

[HPFQA] J. Bennett and Hui Zhang, "Hierarchical Packet Fair Queueing Algorithms", Proc. ACM SIGCOMM 96, August 1996.

[IPv6] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", RFC 2460, December 1998.

[RFC791] Postel, J., Editor, "Internet Protocol", STD 5, RFC 791, September 1981.

[RFC1122] Braden, R., "Requirements for Internet hosts - communication layers", STD 3, RFC 1122, October 1989.

[RFC1812] Baker, F., Editor, "Requirements for IP Version 4 Routers", RFC 1812, June 1995.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RPS] D. Stiliadis and A. Varma, "Rate-Proportional Servers: A Design Methodology for Fair Queueing Algorithms", IEEE/ ACM Trans. on Networking, April 1998.

Authors' Addresses

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

Phone: +1-408-525-4857
EMail: kmn@cisco.com

Steven Blake
Torrent Networking Technologies
3000 Aerial Center, Suite 140
Morrisville, NC 27560

Phone: +1-919-468-8466 x232
EMail: slblake@torrentnet.com

Fred Baker
Cisco Systems
519 Lado Drive
Santa Barbara, CA 93111

Phone: +1-408-526-4257
EMail: fred@cisco.com

David L. Black
EMC Corporation
35 Parkwood Drive
Hopkinton, MA 01748

Phone: +1-508-435-1000 x76140
EMail: black_david@emc.com

Full Copyright Statement