Assured Forwarding PHB Group

Status of this Memo

Copyright Notice

Abstract

   This document defines a general use Differentiated Services (DS)
   [Blake] Per-Hop-Behavior (PHB) Group called Assured Forwarding (AF).
   The AF PHB group provides delivery of IP packets in four
   independently forwarded AF classes. Within each AF class, an IP
   packet can be assigned one of three different levels of drop
   precedence. A DS node does not reorder IP packets of the same
   microflow if they belong to the same AF class.

1. Purpose and Overview

   There is a demand to provide assured forwarding of IP packets over
   the Internet. In a typical application, a company uses the Internet
   to interconnect its geographically distributed sites and wants an
   assurance that IP packets within this intranet are forwarded with
   high probability as long as the aggregate traffic from each site does
   not exceed the subscribed information rate (profile). It is
   desirable that a site may exceed the subscribed profile with the
   understanding that the excess traffic is not delivered with as high

probability as the traffic that is within the profile. It is also

important that the network does not reorder packets that belong to the same microflow, as defined in [Nichols], no matter if they are in or out of the profile.

Assured Forwarding (AF) PHB group is a means for a provider DS domain to offer different levels of forwarding assurances for IP packets received from a customer DS domain. Four AF classes are defined, where each AF class is in each DS node allocated a certain amount of forwarding resources (buffer space and bandwidth). IP packets that wish to use the services provided by the AF PHB group are assigned by the customer or the provider DS domain into one or more of these AF classes according to the services that the customer has subscribed to. Further background about this capability and some ways to use it may be found in [Clark].

Within each AF class IP packets are marked (again by the customer or the provider DS domain) with one of three possible drop precedence values. In case of congestion, the drop precedence of a packet determines the relative importance of the packet within the AF class. A congested DS node tries to protect packets with a lower drop precedence value from being lost by preferably discarding packets with a higher drop precedence value.

In a DS node, the level of forwarding assurance of an IP packet thus depends on (1) how much forwarding resources has been allocated to the AF class that the packet belongs to, (2) what is the current load of the AF class, and, in case of congestion within the class, (3) what is the drop precedence of the packet.

For example, if traffic conditioning actions at the ingress of the provider DS domain make sure that an AF class in the DS nodes is only moderately loaded by packets with the lowest drop precedence value and is not overloaded by packets with the two lowest drop precedence values, then the AF class can offer a high level of forwarding assurance for packets that are within the subscribed profile (i.e., marked with the lowest drop precedence value) and offer up to two lower levels of forwarding assurance for the excess traffic.

This document describes the AF PHB group. An otherwise DS-compliant node is not required to implement this PHB group in order to be considered DS-compliant, but when a DS-compliant node is said to implement an AF PHB group, it must conform to the specification in this document.

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 [Bradner].

2. The AF PHB Group

   Assured Forwarding (AF) PHB group provides forwarding of IP packets
   in N independent AF classes. Within each AF class, an IP packet is
   assigned one of M different levels of drop precedence. An IP packet
   that belongs to an AF class i and has drop precedence j is marked
   with the AF codepoint AFij, where $1 <= i <= N$ and $1 <= j <= M$.
   Currently, four classes (N=4) with three levels of drop precedence in
   each class (M=3) are defined for general use. More AF classes or
   levels of drop precedence MAY be defined for local use.

   A DS node SHOULD implement all four general use AF classes. Packets
   in one AF class MUST be forwarded independently from packets in
   another AF class, i.e., a DS node MUST NOT aggregate two or more AF
   classes together.

   A DS node MUST allocate a configurable, minimum amount of forwarding
   resources (buffer space and bandwidth) to each implemented AF class.
   Each class SHOULD be serviced in a manner to achieve the configured
   service rate (bandwidth) over both small and large time scales.

   An AF class MAY also be configurable to receive more forwarding
   resources than the minimum when excess resources are available either
   from other AF classes or from other PHB groups. This memo does not
   specify how the excess resources should be allocated, but
   implementations MUST specify what algorithms are actually supported
   and how they can be parameterized.

   Within an AF class, a DS node MUST NOT forward an IP packet with
   smaller probability if it contains a drop precedence value p than if
   it contains a drop precedence value q when $p < q$. Note that this
   requirement can be fulfilled without needing to dequeue and discard
   already-queued packets.

   Within each AF class, a DS node MUST accept all three drop precedence
   codepoints and they MUST yield at least two different levels of loss
   probability. In some networks, particularly in enterprise networks,
   where transient congestion is a rare and brief occurrence, it may be
   reasonable for a DS node to implement only two different levels of
   loss probability per AF class. While this may suffice for some
   networks, three different levels of loss probability SHOULD be
   supported in DS domains where congestion is a common occurrence.

   If a DS node only implements two different levels of loss probability
   for an AF class x, the codepoint AFx1 MUST yield the lower loss
   probability and the codepoints AFx2 and AFx3 MUST yield the higher

loss probability.

A DS node MUST NOT reorder AF packets of the same microflow when they
belong to the same AF class regardless of their drop precedence.
There are no quantifiable timing requirements (delay or delay
variation) associated with the forwarding of AF packets.

The relationship between AF classes and other PHBs is described in
Section 7 of this memo.

The AF PHB group MAY be used to implement both end-to-end and domain
edge-to-domain edge services.

3. Traffic Conditioning Actions

A DS domain MAY at the edge of a domain control the amount of AF
traffic that enters or exits the domain at various levels of drop
precedence. Such traffic conditioning actions MAY include traffic
shaping, discarding of packets, increasing or decreasing the drop
precedence of packets, and reassigning of packets to other AF
classes. However, the traffic conditioning actions MUST NOT cause
reordering of packets of the same microflow.

4. Queueing and Discard Behavior

This section defines the queueing and discard behavior of the AF PHB
group. Other aspects of the PHB group's behavior are defined in
Section 2.

An AF implementation MUST attempt to minimize long-term congestion
within each class, while allowing short-term congestion resulting
from bursts. This requires an active queue management algorithm. An
example of such an algorithm is Random Early Drop (RED) [Floyd].
This memo does not specify the use of a particular algorithm, but
does require that several properties hold.

An AF implementation MUST detect and respond to long-term congestion
within each class by dropping packets, while handling short-term
congestion (packet bursts) by queueing packets. This implies the
presence of a smoothing or filtering function that monitors the
instantaneous congestion level and computes a smoothed congestion
level. The dropping algorithm uses this smoothed congestion level to
determine when packets should be discarded.

The dropping algorithm MUST be insensitive to the short-term traffic
characteristics of the microflows using an AF class. That is, flows
with different short-term burst shapes but identical longer-term
packet rates should have packets discarded with essentially equal

probability. One way to achieve this is to use randomness within the dropping function.

The dropping algorithm MUST treat all packets within a single class and precedence level identically. This implies that for any given smoothed congestion level, the discard rate of a particular microflow's packets within a single precedence level will be proportional to that flow's percentage of the total amount of traffic passing through that precedence level.

The congestion indication feedback to the end nodes, and thus the level of packet discard at each drop precedence in relation to congestion, MUST be gradual rather than abrupt, to allow the overall system to reach a stable operating point. One way to do this (RED) uses two (configurable) smoothed congestion level thresholds. When the smoothed congestion level is below the first threshold, no packets of the relevant precedence are discarded. When the smoothed congestion level is between the first and the second threshold, packets are discarded with linearly increasing probability, ranging from zero to a configurable value reached just prior to the second threshold. When the smoothed congestion level is above the second threshold, packets of the relevant precedence are discarded with 100% probability.

To allow the AF PHB to be used in many different operating environments, the dropping algorithm control parameters MUST be independently configurable for each packet drop precedence and for each AF class.

Within the limits above, this specification allows for a range of packet discard behaviors. Inconsistent discard behaviors lead to inconsistent end-to-end service semantics and limit the range of possible uses of the AF PHB in a multi-vendor environment. As experience is gained, future versions of this document may more tightly define specific aspects of the desirable behavior.

5. Tunneling

When AF packets are tunneled, the PHB of the tunneling packet MUST NOT reduce the forwarding assurance of the tunneled AF packet nor cause reordering of AF packets belonging to the same microflow.

6. Recommended Codepoints

  Recommended codepoints for the four general use AF classes are given
  below. These codepoints do not overlap with any other general use PHB
  groups.

  The RECOMMENDED values of the AF codepoints are as follows: AF11 = '
  001010', AF12 = '001100', AF13 = '001110', AF21 = '010010', AF22 = '
  010100', AF23 = '010110', AF31 = '011010', AF32 = '011100', AF33 = '
  011110', AF41 = '100010', AF42 = '100100', and AF43 = '100110'. The
  table below summarizes the recommended AF codepoint values.

```
                  Class 1    Class 2   Class 3   Class 4
                +----------+----------+---------+----------+
  Low Drop Prec   | 001010 | 010010 | 011010 | 100010 |
  Medium Drop Prec| 001100 | 010100 | 011100 | 100100 |
  High Drop Prec  | 001110 | 010110 | 011110 | 100110 |
                +----------+----------+---------+----------+
```

7. Interactions with Other PHB Groups

  The AF codepoint mappings recommended above do not interfere with the
  local use spaces nor the Class Selector codepoints recommended in
  [Nichols]. The PHBs selected by those Class Selector codepoints may
  thus coexist with the AF PHB group and retain the forwarding behavior
  and relationships that was defined for them. In particular, the
  Default PHB codepoint of '000000' may remain to be used for
  conventional best effort traffic. Similarly, the codepoints '11x000'
  may remain to be used for network control traffic.

  The AF PHB group, in conjunction with edge traffic conditioning
  actions that limit the amount of traffic in each AF class to a
  (generally different) percentage of the class's allocated resources,
  can be used to obtain the overall behavior implied by the Class
  Selector PHBs. In this case it may be appropriate within a DS domain
  to use some or all of the Class Selector codepoints as aliases of AF
  codepoints.

  In addition to the Class Selector PHBs, any other PHB groups may co-
  exist with the AF PHB group within the same DS domain. However, any
  AF PHB group implementation should document the following:

  (a) Which, if any, other PHB groups may preempt the forwarding
  resources specifically allocated to each AF PHB class. This
  preemption MUST NOT happen in normal network operation, but may be
  appropriate in certain unusual situations - for example, the '11x000'

codepoint may preempt AF forwarding resources, to give precedence to
unexpectedly high levels of network control traffic when required.

(b) How "excess" resources are allocated between the AF PHB group and other implemented PHB groups. For example, once the minimum allocations are given to each AF class, any remaining resources could be allocated evenly between the AF classes and the Default PHB. In an alternative example, any remaining resources could be allocated to forwarding excess AF traffic, with resources devoted to the Default PHB only when all AF demand is met.

This memo does not specify that any particular relationship hold between AF PHB groups and other implemented PHB groups; it requires only that whatever relationship is chosen be documented. Implementations MAY allow either or both of these relationships to be configurable. It is expected that this level of configuration flexibility will prove valuable to many network administrators.

8. Security Implications

In order to protect itself against denial of service attacks, a provider DS domain SHOULD limit the traffic entering the domain to the subscribed profiles. Also, in order to protect a link to a customer DS domain from denial of service attacks, the provider DS domain SHOULD allow the customer DS domain to specify how the resources of the link are allocated to AF packets. If a service offering requires that traffic marked with an AF codepoint be limited by such attributes as source or destination address, it is the responsibility of the ingress node in a network to verify validity of such attributes.

Other security considerations are covered in [Blake] and [Nichols].

9. Intellectual Property Rights

The IETF has been notified of intellectual property rights claimed in regard to some or all of the specification contained in this document. For more information, consult the online list of claimed rights.

10. IANA Considerations

This document allocates twelve codepoints, listed in section 6, in Pool 1 of the code space defined by [Nichols].

Appendix: Example Services

  The AF PHB group could be used to implement, for example, the so-
  called Olympic service, which consists of three service classes:
  bronze, silver, and gold. Packets are assigned to these three
  classes so that packets in the gold class experience lighter load
  (and thus have greater probability for timely forwarding) than
  packets assigned to the silver class. Same kind of relationship
  exists between the silver class and the bronze class. If desired,
  packets within each class may be further separated by giving them
  either low, medium, or high drop precedence.

  The bronze, silver, and gold service classes could in the network be
  mapped to the AF classes 1, 2, and 3. Similarly, low, medium, and
  high drop precedence may be mapped to AF drop precedence levels 1, 2,
  or 3.

  The drop precedence level of a packet could be assigned, for example,
  by using a leaky bucket traffic policer, which has as its parameters
  a rate and a size, which is the sum of two burst values: a committed
  burst size and an excess burst size. A packet is assigned low drop
  precedence if the number of tokens in the bucket is greater than the
  excess burst size, medium drop precedence if the number of tokens in
  the bucket is greater than zero, but at most the excess burst size,
  and high drop precedence if the bucket is empty. It may also be
  necessary to set an upper limit to the amount of high drop precedence
  traffic from a customer DS domain in order to avoid the situation
  where an avalanche of undeliverable high drop precedence packets from
  one customer DS domain can deny service to possibly deliverable high
  drop precedence packets from other domains.

  Another way to assign the drop precedence level of a packet could be
  to limit the user traffic of an Olympic service class to a given peak
  rate and distribute it evenly across each level of drop precedence.
  This would yield a proportional bandwidth service, which equally
  apportions available capacity during times of congestion under the
  assumption that customers with high bandwidth microflows have
  subscribed to higher peak rates than customers with low bandwidth
  microflows.

  The AF PHB group could also be used to implement a loss and low
  latency service using an over provisioned AF class, if the maximum
  arrival rate to that class is known a priori in each DS node.
  Specification of the required admission control services, however, is
  beyond the scope of this document. If low loss is not an objective,
  a low latency service could be implemented without over provisioning

by setting a low maximum limit to the buffer space available for an
AF class.

References

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

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

[Clark]   Clark, D. and Fang, W., Explicit Allocation of Best Effort
          Packet Delivery Service. IEEE/ACM Transactions on
          Networking, Volume 6, Number 4, August 1998, pp. 362-373.

[Floyd]   Floyd, S., and Jacobson, V., Random Early Detection
          gateways for Congestion Avoidance. IEEE/ACM Transactions on
          Networking, Volume 1, Number 4, August 1993, pp. 397-413.

[Nichols] 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.

Authors' Addresses

  Juha Heinanen
  Telia Finland
  Myyrmaentie 2
  01600 Vantaa, Finland

  EMail: jh@telia.fi

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

  EMail: fred@cisco.com

  Walter Weiss
  Lucent Technologies
  300 Baker Avenue, Suite 100,
  Concord, MA 01742-2168

  EMail: wweiss@lucent.com

  John Wroclawski
  MIT Laboratory for Computer Science
  545 Technology Sq.
  Cambridge, MA 02139

  EMail: jtw@lcs.mit.edu

Full Copyright Statement

Acknowledgement