                   Deterministic Forwarding PHB
              draft-svshah-tsvwg-deterministic-forwarding-03

Abstract

   This document defines a Differentiated Services Per-Hop-Behavior
   (PHB) Group called Deterministic Forwarding (DF). The document
   describes the purpose and semantics of this PHB. It also describes
   creation and forwarding treatment of the service class. The document
   also describes how the code-point can be mapped into one of the
   aggregated Diffserv service classes [RFC5127].

Status of this Memo

   This Internet-Draft is submitted in full conformance with the
   provisions of BCP 78 and BCP 79.

   Internet-Drafts are working documents of the Internet Engineering
   Task Force (IETF). Note that other groups may also distribute
   working documents as Internet-Drafts. The list of current Internet-
   Drafts is at http://datatracker.ietf.org/drafts/current/.

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

   This Internet-Draft will expire on September 2, 2015.

Table of Contents

[1](#). Introduction

There is a demand to provide deterministic forwarding to certain type
of traffic in machine to machine networks over IP. With an
introduction of machine to machine networks over IP, a new set of IP
applications are emerging. Traffic types from such applications/
networks are some-what different from the traditional traffic types.
Though most traffic types have characteristics similar to that of
traditional ones [[LLN-DIFF](#)], certain control signals for some of the
applications are extremely sensitive to latency and jitter and thus
timely scheduled delivery. End to end deterministic path for such
traffic may be through one or more administered inter-connected
machine to machine networks.

Deterministic Forwarding (DF) PHB group is a means for each node in
machine to machine networks, in an end to end path, to deliver
required deterministic behavior. DF class in each DS node is
allocated with a scheduled transmission time. DS node may be
allocated one scheduled time for the whole aggregate DF traffic, or
may be allocated with different schedule time for each micro-flow or
set of micro-flows in a DF class. IP packets that wish to use
deterministic service are assigned DF code-point, typically at the
originator of such traffic.

In a DS node, the level of forwarding determinism of an IP packet
depends on scheduled time, at which packet then serviced independent
of existence and load of any other type of traffic. For example when
a DF packet arrives at the DS node, it is queued until its
provisioned scheduled time of service. At the trigger of that
scheduled time, service to all other traffic is pre-empted to service
DF traffic.

This document describes the DF PHB group. DF capability is not a
required function for a DS-compliant node, but a DS-compliant node
that implements DF PHB group MUST conform to the specification in
this document.

Typically for an application where end to end deterministic service
is important, relevant traffic can be serviced through DF PHB at
every hop in the path. However, in cases where intermediate hops (or
DS domains) either do not support DF PHB or supports only aggregated
service classes described in [RFC5127](#), DF traffic in those DS domains
MUST be mapped to Real Time Treatment class (EF PHB) defined in
[RFC5127](#). Traffic in such scenario MUST be conditioned at the Edge
before entering and after exiting such DS domains. This is described
further in later section.

2. Terminology

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.

3. DF Per Hop Behavior

The DF PHB is to implement deterministic scheduled, deterministic in
terms of a time, forwarding treatment. DF traffic MUST be serviced
in a manner to meet configurable scheduled time. A DS node may pre-
allocate dedicated resources available at configured scheduled time
for optionally configurable maximum size of data. Every conforming
packet, belonging to DF class, gets deterministic service
irrespective of traffic in other Diffserv class and current load on
the system. A DS node MAY allow though its dedicated scheduled
resources available to other Differentiated service classes when DF
class does not have any packets to serve during its service time.

A DS node MAY be configured with the same parameters for the entire
DF traffic class or different parameters for each micro-flow in a DF
class. For the later case, DF traffic MUST be serviced in a manner
to meet scheduled service time for each individual micro-flow. Per-
flow DF parameters may be provisioned dynamically thru a signaling
protocol. Use of any signaling protocol is agnostic to the DF PHB
and thus out of scope of this specification [An example of such
signaling protocol referred in 6TisCH]. What signaling protocol
requires to convey, at minimum to each DS node in the end to end
path, is request for DF service along with the flow classification
and associated DF parameters, parameters like intended time of a
service and size of data to be transmitted during time of service.
In a packet path, packet first is classified if it belongs to DF PHB.
Once classified for DF PHB, it gets deterministic treatment if
provisioned for per-flow DF parameters or else gets aggregate DF
treatment.

4. Traffic Conditioning

A DF supported DS Domain MAY condition traffic at the ingress Edge of
the domain. Traffic conditioning MUST discard any incoming packet
that does not conform to the configured DF service. As per PHB
definition, packets are required to be scheduled and delivered at a
precise absolute or relative time interval. Any packet that has
missed the window of its service time MUST be discarded. For example
if a DF queue is provisioned to serve a packet with less than x ms of
jitter and for an arrived packet, if next scheduled time for a packet

results in more than x ms of jitter then such packet MUST be
discarded. The packet MUST also be checked against the size of the
data. If size of the packet is bigger than max size of the data a
scheduled time is provisioned to service then such packet MUST be
discarded. In addition to DS node at the ingress Edge of the domain,
other DS nodes in the path MAY implement Traffic Conditioning.

5. Diffserv behavior through non-DF DS domain

In deployments if two DF domains are connected through a domain that
does not support DF PHB, traffic from such intermediate domain MUST
be forwarded with low latency. DF traffic at the egress Edge of the
sender DF domain MUST be mapped to EF PHB aggregate service, defined
as Real Time Service aggregation in RFC5127. Such traffic when
entered in the receiving DF domain MUST be conditioned, as described
in earlier section, at the ingress Edge of that receiving domain.

6. Potential implementation of DF scheduling

Following are examples of potential implementations. They are not
any form of guidelines or recommendations but simply a reference to
potential implementations.

 There are at least two ways to implement scheduling for DF traffic
 class.

 1) One queue to buffer and schedule all DF traffic (from all flows),

 2) Multiple sub-queues for DF traffic class, one queue for each DF
    provisioned flow

Any chosen DF scheduling implementation MUST run traffic conditioning
at enqueue to decide if packets to be enqueued or discarded.
Discussed more in later section.

   1) Single queue to buffer all DF traffic

This single queue maintains, possibly a circular, indexed buffer
list. Each index logically maps to each scheduled time service. If
conditioning results in not to discard a packet, packet gets en-
queued at a relevant index in the buffer list that maps to a relevant
scheduled time slot. If there is no packet(s) received for a
specific scheduled time service then buffer index for that scheduled
service remains empty. This also means that during dequeue, at a

schedule time service, an empty index results in no dequeued packets
from the DF queue and thus nothing to be transmitted from the DF
queue at that point in time. Queuing system may de-queue packets
from non-DF queues when an index in DF buffer list found to be an
empty during a specific scheduled time service.

```
                                        .
                                        |`.
      EF (Low latency) ----------------||---->   `.
                                  High  |      `.
                       .                |          `.
      rate queues  |`.                  |            `.
      AF1 ----||--->   `.               |              `.
                  |      `.             |                `.
      AF3 ----||--->       '------------------>          '------>
                  |     .'         Low   |            .'
      BE   ----||--->   .'               |          .'
                  | .'                   |        .'
                  .'                     |      .'
                           Deterministic|    .'
      DF                  ---------------||---->  .'
       (scheduled time/interrupt driven de-queue)|.'
```


      2) multiple queues to buffer each DF traffic flows

   If conditioning results in not to discard a packet, packet gets
   enqueued in the relevant DF queue designated for that flow. At a
   scheduled time slot, scheduler dequeues a packet from the respective
   queue for that flow. Every scheduled time service interrupt is
   mapped to a flow specific DF queue to dequeue a packet from.

```
                                         .
                                        |`.
        EF (Low latency) ---------------||----->  `.
                                  High   |      `.
                    .                    |        `.
        rate queues  |`.                 |         `.
        AF1 ----||--->  `.               |          `.
                 |       `.              |           `.
        AF3 ----||--->      '------------------->         '------>
                 |      .'            Low  |          .'
        BE  ----||--->   .'                |        .'
                 | .'                      |      .'
                .'                         |    .'
        (DF queues)             Deterministic|    .'
        DF (at interval 1, 6, 11 ..) ----||----->  .'
        DF (at interval 3, 8, 13 ..) ----||----->.'
      (scheduled time/interrupt driven de-queue)|
```
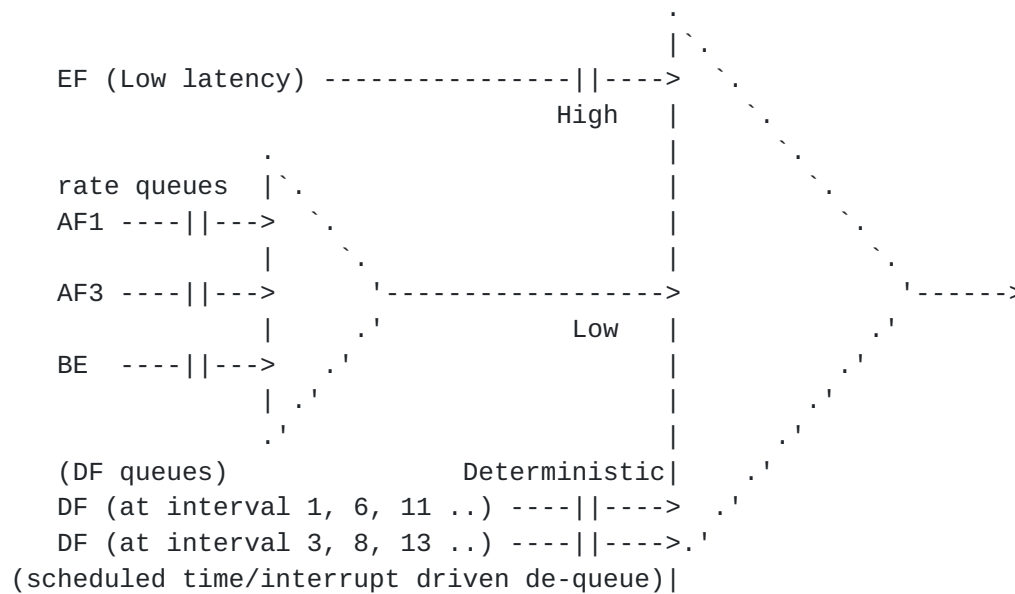
7. Updates to [RFC4594](#) and [RFC5127](#)

   This specification updates [RFC4594](#) with an addition of a new Diffserv
   Class. It also updates [RFC5127](#) to aggregate DF class of traffic to
   Real Time Aggregation Class.

8. IANA Considerations

   This document defines a new DSCP code-point DF. IANA maintains the
   list of existing DSCPs. Proposal is to allocate a new one for the DF
   code-point.

9. Security Considerations

   There is no security considerations required besides ones already
   understood in the context of Differentiated services architecture

10. Acknowledgements

   Fred Baker, Norm Finn, David Black, Brian Carpenter for their
   comments and feed-back.

11. References

11.1. Normative 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", RFC 2475, December 1998.

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

   [RFC4594]   Babiarz, J., Chan, K., and F. Baker, "Configuration
               Guidelines for DiffServ Service Classes", RFC 4594,
               August 2006.

   [RFC5127]   Chan, K., Babiarz, J., and F. Baker, "Aggregation of
               Diffserv Service Classes", RFC 5127, February 2008.

11.2. Informative References

   [TiSCH]   Thubert, P., Watteyne, T., and R. Assimiti, "An
               Architecture for IPv6 over the TSCH mode of IEEE
               802.15.4e, I-D.draft-ietf-6tisch-architecture", Nov 2013.

   [LLN-DIFF]
               Shah, S. and P. Thubert, "Differentiated Service Class
               Recommendations for LLN Traffic,
               I-D.draft-svshah-tsvwg-lln-diffserv-recommendations",
               Aug 2013.

Authors' Addresses

   Shitanshu Shah
   Cisco Systems
   170 W. Tasman Drive
   San Jose, CA 95134
   US

   Email: svshah@cisco.com

Pascal Thubert
Cisco Systems
Village d'Entreprises Green Side
400, Avenue de Roumanille
Batiment T3
Biot - Sophia Antipolis 06410
FRANCE

Email: pthubert@cisco.com