Deterministic Forwarding PHB
draft-svshah-tsvwg-deterministic-forwarding-00

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 July 12, 2014.

Copyright Notice

Table of Contents

1. Introduction

   IP Networks typically implement Diffserv to provide differentiated
   forwarding behavior to different class of traffic. Networks that
   implement Diffserv relies on DSCP code-point in the IP header of a
   packet to select PHB as a specific forwarding treatment for that
   packet [RFC2474, RFC2475]. This document describes a particular PHB
   called Deterministic Forwarding (DF). The proposed new code-point
   defines a service class for the purpose of forwarding treatment of a
   packet at determined/fixed scheduled time providing no jitter service
   to the class of traffic (updates RFC4594 with the addition of a new
   Service Class).

   DF PHB can be used for the network services that require the
   capability to ensure a predictable interaction between networked
   systems and guarantee a very strict time scheduled services.
   Applications of such networks may be able to absorb a loss but are
   very sensitive to end to end latency and jitter. Examples of such
   networks include Machine to Machine (M2M) control and monitoring
   deployment with IP over varieties of Layer 2 networks.

   The definition of Expedited Forwarding (EF) [RFC2598] PHB is low
   latency and thus one can envision use of EF code-point for such
   service. However, even though EF defines low latency and low jitter,
   it does not guarantee deterministic/fixed scheduled time service.
   Depending on co-existence of the other traffic in the network, EF
   traffic may have more or less variance on jitter and thus not
   suitable for the deterministic service. DF PHB thus is more suitable
   for deterministic time sensitive traffic.

   Typically for an application where end to end deterministic service
   is important, relevant traffic should be provisioned through DF PHB
   at every hop in that end to end 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.

1.1. Use-cases

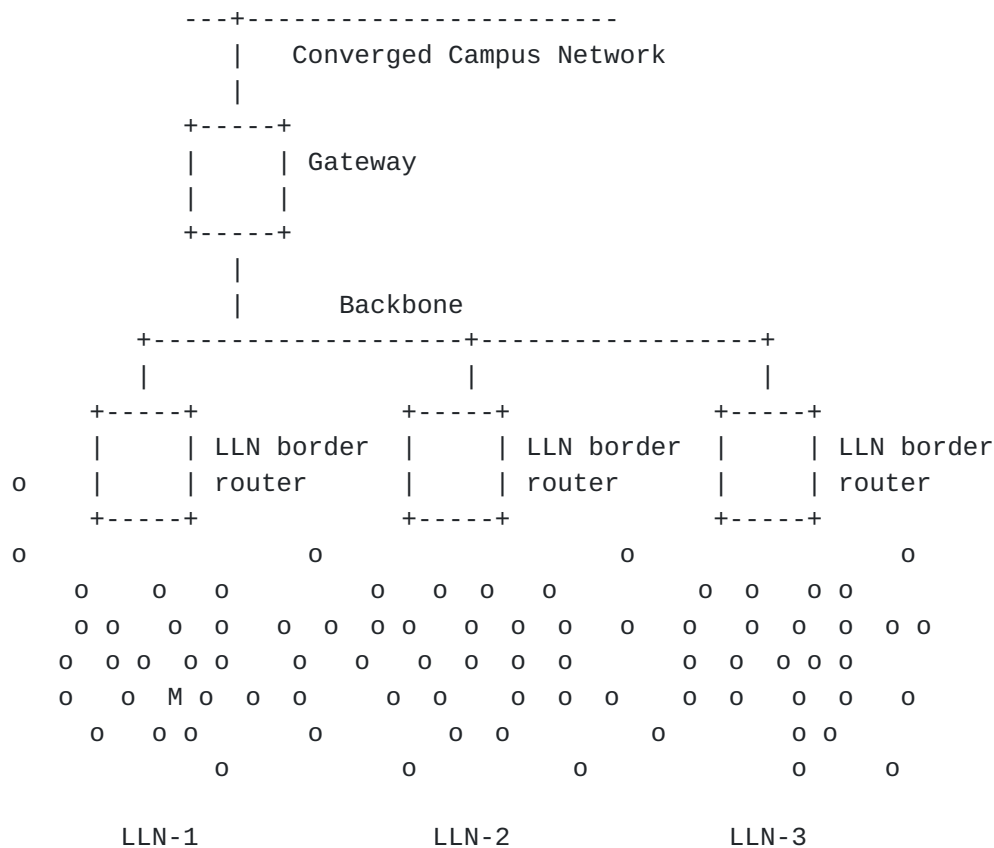   With an introduction of machine to machine networks over IP, a new
   set of applications are emerging. Traffic types from such
   applications/networks are some-what different from the traditional
   traffic types. Though most traffic types have characteristics
   similar to that of traditional ones [LLN-DIFF], certain control
   signals for some of the applications are extremely sensitive to

   latency and jitter. Such control signals demand much stricter
   latency and jitter, at pretty much decisive time scheduled delivery,
   end to end. Industrial automation, Smart cities and automobiles/
   planes/trains built around such networks are examples of such use-
   cases.

   Machine to machine networks may be implemented on varieties of Layer
   2 protocols. 802.3 and 802.15e [TiSCH] are examples of layer 2 that
   are enhancing their capabilities to allow time scheduled delivery of
   packets.

   In a wireless sensor networks, that are implemented over IP, multiple
   LLN (Low power and Lossy Networks) may be connected through Backbone.

```
              ---+------------------------
                 |    Converged Campus Network
                 |
              +-----+
              |     | Gateway
              |     |
              +-----+
                 |
                 |        Backbone
          +--------------------+-----------------+
          |                    |                 |
        +-----+              +-----+           +-----+
        |     | LLN border   |     | LLN border |     | LLN border
    o   |     | router       |     | router     |     | router
        +-----+              +-----+           +-----+
      o              o                  o                    o
        o    o   o       o   o  o   o        o  o   o o
         o o    o o    o o  o o o  o  o  o   o   o  o  o  o o
         o  o o  o o    o   o    o  o  o  o      o  o  o o o
         o    o M o  o  o      o  o    o  o  o   o  o   o  o    o
          o    o o       o       o  o         o       o o
              o               o          o            o    o

           LLN-1              LLN-2              LLN-3
```

   As shown in the diagram, multiple LL Networks are connected to each
   other via Backbone through LLN Border routers. Each LL Network
   consist of many nodes. There are different types of traffic
   forwarded through each LL node and from one LL Network to another.
   Most LLN traffic types have characteristics similar to that of
   traditional ones and thus can be supported through existing Diffserv
   classes except time sensitive control signals. Without segregating

such control signals to a specific Diffserv class would require
Intserv support for LLN traffic in such networks. All traffic would
be subject to flow classification to differentiate time sensitive
control signals which can be a big scale concern. Supporting time
sensitive control signals via newly proposed DF Diffserv class allows
implementation of Diffserv in LLN Networks.

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 code-point Behavior

The DF PHB is to implement time scheduled forwarding treatment.
Provisioning of such a service has two parts,

    1) Provisioning of the fixed/relative time for scheduling of such
       service
    2) Provisioning of the max size of the data to be transmitted at
       each scheduled time

Provisioned scheduled time may be absolute or relative. For example,
a DF class may be provisioned to schedule packets (or bytes) at every
fixed time. Fixed time can be time of a day or any other absolute
definition. In a multi hop forwarding of DF traffic, absolute time
service provisioning at each hop may require to be dependent on the
clock synchronization (clock synchronization is not in the scope of
this specification). In relative time scheduling, packets to be
scheduled at every specific interval or it could be relative to any
other specific event/trigger. The definition of the time interval or
any other event is relevant to that specific provisioned node only.

The size of the data, to be transmitted at each scheduled time
service, provisioned can be in the unit of bytes or time. Once DF
PHB is provisioned and enabled, forwarding treatment MUST service
packets (bytes) from this class at the scheduled time for max
allowable data. Scheduling MUST pre-empt any other service,
including EF, during the schedule time service for the DF class. In
order to avoid incurred latency to EF class of traffic, it is
expected to carefully provision DF class to limit scheduled time
service to as minimal data transmission that would prevent larger
than expected delays to EF class of traffic.

Provisioning can be done via any of multiple possible methods. It
could be via command interface, or could be via external provisioning
agents, or could be via some sort of signaling that may dynamically
pre-negotiate time window of transmission at each node in a network
path.

3.1. Potential implementation of DF scheduling

Following are examples of potential implementations. They are not
any form of guidelines or recommendations.

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

Flow here represents macro definition, it does not have to be only
5-tuple.

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) One data-plane queue to buffer all DF traffic

This one queue maintains, possibly a circular, indexed buffer list.
Every time scheduled slot is an index in the buffer list. If enqueue
conditioning decides not to discard a packet, packet gets en-queued
at the relevant index in the buffer list in such a way that relevant
index pointer, and thus buffered relevant packets for that index, at
the head of the list is ready to be de-queued at next scheduled time.
Subsequent buffer index is scheduled for the subsequent scheduled
time slot and so forth. If a specific flow has not received any
packet for a scheduled time then buffer index for that flow remains
empty. A packet from other flows do not get buffered at that empty
index. That means during dequeue, at a schedule time service, an
empty index results in no packets to dequeue and thus nothing to be
transmitted from the DF queue at that point in time.

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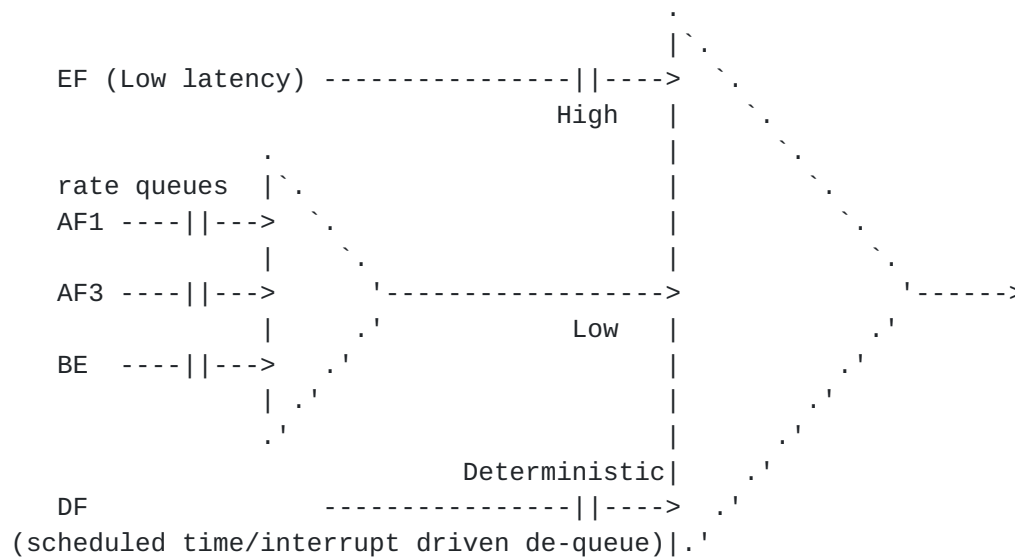
2) multiple sub-queues for each DF flows

If enqueue conditioning decides not to discard a packet, packet gets
enqueued in the relevant DF sub-queue designated for that flow. At a
scheduled time slot, scheduler dequeues a packet from the respective
sub-queue. Every scheduled time service interrupt is mapped to a
specific DF sub-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 ..) ----||---->.'
```
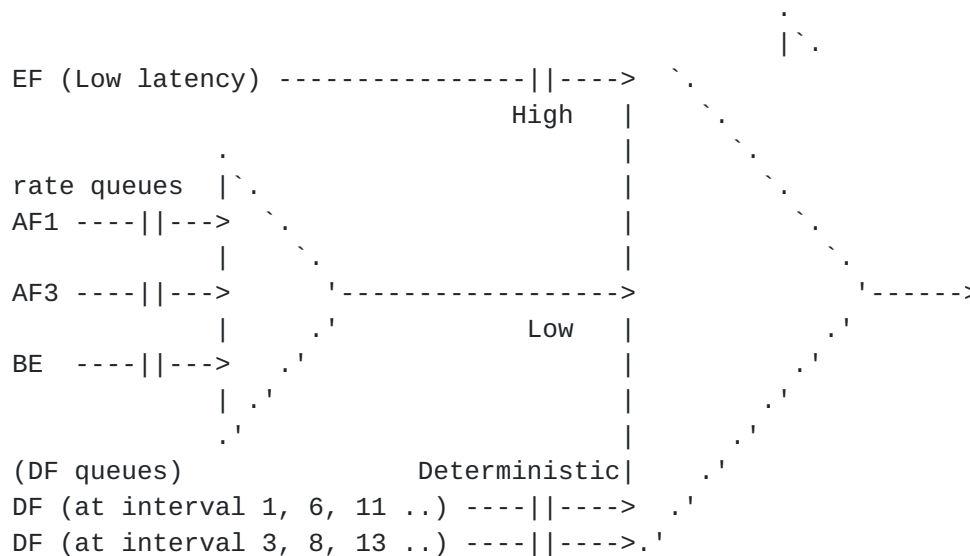
[3.2]. **Conditioning DF traffic at Enqueue**

   DF traffic MUST be conditioned at the enqueue. 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. That would
   also mean any packet coming from the previous hop MUST be conditioned
   at the enqueue for validity of its scheduled service. 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 enqueued packet MUST also be checked against the size
   of the data. If size of the data to be enqueued in a DF queue is
   bigger than what scheduled time slot is provisioned for then such
   packet MUST be discarded.

[4]. **Diffserv behavior through non-DF DS domains**

   In cases where DF traffic is forwarded through multiple DS domains,
   DS domains close to the source and receiver understand application's
   deterministic service requirement well and so MUST be provisioned for
   the precise time scheduled forwarding treatment. Intermediate DS
   domains MAY support DF PHB. Intermediate domains that can not
   support DF PHB, DF traffic from such domains SHOULD get EF treatment,
   as defined in [RFC5127] for Real Time Service aggregation. Sender and
   Receiver DS domains, in such cases, MUST condition DF traffic at the
   respective Edge. If EF service through intermediate DS domains can
   have a predictable upper bound, receiver DS domain Edge can add a
   correction to an incurred latency/jitter with its own defined time
   interval for DF service.

[5]. **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.

[6]. **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.

7. **Security Considerations**

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

8. **Acknowledgements**

   Fred Baker and Norm Finn.

9. **References**

9.1. **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.

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

9.2. **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