Network Working Group
Internet-Draft
Intended status: Informational
Expires: January 7, 2014

C. Lai
W. Wang
S. Yang
T. Eckert
F. Yip
Cisco Systems

July 6, 2013

Normalization Marker for AF PHB Group in DiffServ
draft-lai-tsvwg-normalizer-02

Abstract

In DiffServ, preferential dropping of packets in AF PHB groups has long been considered beneficial, typically for video flows with discardable packets. Unfortunately, the ecosystem of bandwidth contention at congestion is very likely to discourage those video endpoints from generating packets with lower precedence markings, i.e. they would lose more packets if doing so. Thus, to offer an incentive for more collaborative and mutually beneficial behaviors of video endpoints in AF PHB groups, we propose a Normalization Marker (NM) for traffic conditioning at network edges. Deployment of NM will encourage the video endpoints to generate finer layers of intra-flow precedence (IFP) with discardable packets in more balanced distributions.

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 January 7, 2014.

Copyright Notice

Copyright (c) 2013 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (http://trustee.ietf.org/license-info) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction

Assured Forwarding (AF) Per-Hop Behavior (PHB) groups are described in [RFC2597] (with terminology clarified in [RFC3260]) for DiffServ (DS) multimedia service classes such as realtime video conferencing and on-demand streaming. Four AF PHB groups have been defined in [RFC4594] with DS codepoint (DSCP): AF1x, AF2x, AF3x and AF4x where x=1, 2 or 3 for drop precedence in each independent AF PHB group. The DS nodes that support an AF PHB group must set configuration of Active Queue Management (AQM) properly w.r.t. those DSCP markings. For example, for AF4x PHB group which includes AF41, AF42 and AF43 markings, an AQM implementation by Weighted Random Early Detection (WRED) should be configured with some drop probabilities and queue thresholds such that the packet loss rate of AF41 <= AF42 <= AF43 on congestion of the queue.

For an AF PHB group, a DS boundary node or host in the DS domain should use a marking algorithm that properly assigns AF markings of drop precedence to all packets w.r.t. the traffic profiles and Service Level Agreements (SLA). For example, [RFC2697] and [RFC2698] use a token-bucket mechanism for metering each stream of packets and respectively define "srTCM" and "trTCM" markers, to mark packets by the data rate and burst size limit in traffic profiles. Those rate-control markers can be useful at DS boundary nodes for traffic conditioning [RFC2475] and to support IntServ/RSVP traffic over DS regions [RFC2998]. Multiple markers may be applied to the same stream, either on the same or multiple DS nodes along the path. For example, srTCM and trTCM can operate in a so-called "color-aware" mode such that for each incoming packet that already carries an AF marking, the local srTCM/trTCM either keeps the same or lowers the drop precedence of that packet by metering.

However, modern video codec technologies are being advanced not only in coding efficiency (i.e. better compression ratio) but also in two key areas for transport on IP networks: (1) encoder rate-control and dynamic adaptation; (2) ability to generate discardable packets in multiple layers to tolerate packet losses in the network without significant degradation of video quality observed at the decoder. For (1), the encoder dynamically limits its output rate of packets into the AF PHB group, i.e., the encoder’s host is the first DS node equiped with srTCM/trTCM if it marks packets in that behavior. The next DS node is the first-hop router which may add extra srTCM/trTCM to enforce the traffic conditioning or policing from the network’s perspective. Thus, we consider this an incentive for (1) because an encoder using a self rate-control is less likely to see packet losses by the network. Unfortunately, an incentive for (2) is arguably missing today.

To see the missing incentive for (2), consider the following example where 2 video flows A and B with rate control are sent in AF4x PHB group. Each sends 5Mbps on average with some burstiness, but still complies with the rate and burst limit in its traffic profile. However, A and B generate packets with AF4x markings in different distributions of precentage:

Flow A

80% or 4Mbps in AF41

20% or 1Mbps in AF42

0% or 0Mbps in AF43

Flow B

40% or 2Mbps in AF41

40% or 2Mbps in AF42

20% or 1Mbps in AF43

Flow B at above is likely using a more advanced video technology to generate multiple layers of discardable video packets, and thus, its distribution of AF4x markings looks finer and more balanced. That is, flow B acts more friendly to other flows in this AF4x PHB group.

Thus, we argue that the ecosystem in practical deployment should offer an incentive for flows to behave similarly to what flow B is doing above, i.e., on congestion, the AF4x PHB group should try to drop packets in the same amount from each flow, while a flow with finer layers of discardable packets and/or in a more balanced distribution should be able to benefit from its own efforts and see good results in video quality preservation.

Unfortunately, this incentive is still missing today. Suppose that congestion occurs in the AF4x WRED queue where A and B compete for bandwidth and there is no other flow, for simplicity. B's packet loss rate is very likely to become higher than A's, despite B's effort of acting friendly:

o If the queue drops 1Mbps in total,

A sees 0% or 0Mbps loss;

B sees 20% or 1Mbps loss (all its AF43 are lost).

o If the queue drops 4Mbps in total,

A sees 20% or 1Mbps loss (all its AF42 are lost);

B sees 60% or 3Mbps loss (all its AF42 and AF43 are lost).

Thus, to create the missing incentive at above, we propose a new "Normalization Marker" (NM) and describe it in this memo. NM can be deployed on DS boundary nodes for traffic conditioning in practical deployment with AF PHB groups for multimedia service classes. In summary, if NM is applied to a DS boundary node for an AF PHB group, it re-assigns the AF markings of all packets per flow such that the distributions of the AF markings are similar in all flows, i.e., it "normalizes" the distributions of AF markings in all flows. It also attempts to maintain the original orders of the intra-flow drop precedence carried by the input AF markings, as linearly as possible. After the AF-marking distributions are normalized, all those flows should see very similar packet loss rates at AQM for this AF PHB group on congestion of the queue. Then, a codec implementation may have better video quality preservation on network congestion if it employs a more advanced video technology to generate discardable packets with finer markings of drop precedence in a more balanced distribution.

```
                            +------------+
                            |  Runtime   |
                            | Statistics |
                            |            V
                        +-------+    +--------+
                        |       |    |        |
      AF-Marked         | Distr.|    |  Norm  |        AF-Marked
    Packet Stream ===>| Meter |===>| Marker |===> Packet Stream
                        |       |    |        |
                        +-------+    +--------+
```

Normalization Marker (NM) with AF PHB Group

Figure 1

Note that the use of NM is not necessarily limited to video service classes, but could be extended to wherever AF PHB groups can be used, or to any other PHB groups that require a similar incentive NM can provide.

2. Background

2.1. Video Packets in Structure

Modern video codec technologies such as ITU-T H.264/MPEG-4 AVC [H264] typically generate a stream of encoded video packets with internal structure of data dependency for decoding. This has been designed for at least 3 fundamental reasons:

- Coding Efficiency: An encoder improves its coding efficiency typically by reducing spatial and temporal redundancy of the input. For video, spatial redundancy is reduced by intra-frame motion prediction and compensation, while temporal redundancy refers to inter-frame since a video stream is composed of a sequence of frames or pictures in the temporal order. With motion prediction, a frame can be encoded by referencing some pixels of the picture data that will be decoded earlier either in the same (intra) or another (inter) frame so that it can use significantly fewer bits to encode this frame. The frame where the pixels are referenced by any other frame is thus called a referenced frame in the video stream; for example, Instantaneous Decoding Refresh (IDR) in H.264 or Intra (I) frames are typically referenced by subsequential frames, while Predictive (P) frames may be referenced at the encoder's choice, by the Group of Picture (GOP) profile, and/or by some proprietary algorithm in the codec implementation.

- Lossy Network: To use network transport that may lose packets, the encoder may choose to generate a stream with two or more layers each of which the packets are marked with some layer identifier (ID). The network can simply use the layer ID to determine the drop precedence of each packet in the video stream.

  * Layers in Hierarchy of Dependency: If these layers are coded in hierarchy of dependency, the packets in an "enhancement" layer will depend on 1 or more "base" layers to get decoded without errors, while packets in a base layer without dependency can be independently decoded without errors.

    + If some enhancement layer packets are lost, the decoding errors in that picture frame will not stay or cascade to other frames given that no others depend on those lost data. This nice property allows the network to safely drop packets in some enhancement layers, if needed, without badly impacting the video quality at decoder.

    + If some base layer packets are lost, the impact can be severe since these decoding errors will stay in buffer and cascade to all other picture pixels that depend on the lost data to decode in the current and/or a later frame. This impact can last tens of seconds as the video quality continues getting worse, resulting in unpleasant user experiences, until the decoder receives the next IDR or I frame, either on-demand or periodically, to remove those errors.

For example, H.264 Annex G defines Scalable Video Coding (SVC) using a 3-dimensional (i.e. spatical, temporal and quality) hierarchy of layer dependency at the encoder’s choice, but for simplicity, it also defines a scalar number called Priority ID (PID) in its header so the network could instead use PID, if set by the encoder, to determine drop precedence in the stream.

* Layers NOT in Hierarchy of Dependency: Sometimes the encoder will generate multiple layers without any dependency between those layers. These mechanisms usually enlarge the amount of encoded video data for vairous purposes. For example,

  + Forward Error Correction (FEC) may be used at the encoder to generate extra FEC packets, so that the decoder can tolerate certain amounts of packet losses.

  + Simulcast (i.e. simultaneous multicast) by an encoder will actually generate multiple layers each of which can be transmitted and decoded independently, in parallel by IP or application multicast. Each layer carries video in a different resolution and/or quality. The decoder can choose 1 or more of those layers to receive according to the required, available or detected bandwidth, packet losses, delays, jitter etc. in its network service.

  With FEC and/or Simulcast, the encoder can still mark the packets with different drop precedence in those layers to better protect the more important data for video quality at decoding when congestion occurs.

o In-Band Signaling: An encoded video stream usually carries in-band control messages that are most critical for adequate encoder and decoder behaviors. For example,

  * H.264 Annex D defines Supplemental Enhancement Information (SEI), which could also carry proprietary codec parameters. These in-band control signals should be given the highest drop precedence.

* Real Time Control Protocol (RTCP) carries in-band control messages for Real Time Protocol (RTP) [RFC3550], which is mostly used for realtime multimedia transmission on IP networks. RTCP messages are defined as RTP packets with special payload types in the RTP stream. RTCP packets should be given the highest drop precedence but should receive the same delay/jitter as regular RTP packets in the same stream.

2.2. Intra-Flow Precedence (IFP)

For abstraction, we define "Intra-Flow Precedence" (IFP) to represent the drop precedence in one individual flow that may carry a video stream of IP packets in multimedia networks. Here is a summary of IFP characteristics:

- o IFPs are drop precedence levels that are only significant within each individual flow.

- o IFPs are integer numbers that can be numerically compared if needed. 0 represents the highest precedence. The larger numerical value an IFP is, the lower precedence it represents.

- o The number of IFP levels in each flow is not necessarily the same.

- o IFPs between any 2 flows should NOT be compared to determine drop precedence between their packets in a queue.

- o IFPs may be assigned by the original encoder of the stream and carried in some bits field of all packets in the stream.

- o IFPs may be assigned or re-assigned by a middle box or router if it is capable of understanding the stream packet format and codec symantics.

For example, an H.264 AVC flow may have the following IFP assignments at the video encoder's choice.

IFP = 0 for in-band signals

IFP = 1 for IDR frames

IFP = 2 for referenced P (rP) frames

IFP = 3 for non-referenced P (nrP) frames and others

IFP assignements as well as their distribution can vary a lot among different encoder implementations and codec profiles. For example, some encoders may generate both long-term and short-term referenced P frames, where a long-term referenced P frame should have higher drop precedence. In case of H.264 SVC, the IFP assignments could simply be the same as the PID assignments if set by the encoder properly, or be calculated based on the SVC layer ID that has 3 tuples for the spatial, temporal and quality dimensions, respectively.

2.3. Mapping IFP to AF Markings

When a flow is sent in an AF PHB group, the number of its IFP levels is not necessarily equal to the number of the AF markings. In fact, since each of the currently defined AF PHB groups has only 3 AF markings, it is likely that an encoder or DS node needs to apply an n-to-1 mapping from IFPs to AF markings in practice.

The mapping decision is made usually by the encoder, but can also be made by another DS node if necessary and if the DS node is able to understand the encoded video packets, which may require Deep Packet Inspection (DPI), e.g. to read in RTP payload and parse the H.264 headers [RFC6184], or in a proprietary bits field in the IP payload, to retrieve or calculate the IFP of each packet in a flow before locally mapping the IFP to an AF marking.

This n-to-1 mapping can be arbitrary but should be appropriate. Consider 2 IFPs, say x and y, where x and y are mapped to AF markings AF(x) and AF(y), respectively. Then, the mapping should ideally obey the following criteria to keep linearity from IFPs to AF markings.

If x < y, AF(x) <= AF(y);

If x > y, AF(x) >= AF(y).

Although the above two do NOT imply that if x = y, AF(x) = AF(y), it is usually so in practical implementation as it is straightforward. Then, if the encoder algorithm generates a lot of packets with the same IFP, all those packets will be assigned the same AF marking, possibly resulting in an unbalanced distribution of AF markings in the AF PHB group. Thus, an encoder with advanced technologies should make good efforts to generate packets with a finer and more balanced IFP distribution in the first place.

For example, if AF4x PHB group is used to send an H.264 AVC flow with the IFP assignments in the example of Section 2.2, one possible IFP-to-AF4x mapping is:

AF(0) = AF41

AF(1) = AF41

AF(2) = AF42

AF(3) = AF43

This mapping actually results in the following AF markings:

AF41 for in-band signals and IDR frames

AF42 for referenced P (rP) frames

AF43 for non-referenced P (nrP) frames and others

Now, consider two encoders that generate flow A and B, respectively, both using this mapping, but with different IFP distributions as follows.

Flow A

5% in IFP=0 for in-band signals

75% in IFP=1 for IDR frames

20% in IFP=2 for rP frames

Flow B

5% in IFP=0 for in-band signals

35% in IFP=1 for IDR frames

40% in IFP=2 for rP frames

20% in IFP=3 for nrP frames

Thus,

Flow A

80% in AF41

20% in AF42

0% in AF43

Flow B

40% in AF41

40% in AF42

20% in AF43

This results in exactly the two AF marking distribtions that we have previously used in Section 1.

Note that in terms of encoded data size, an IDR frame is typically 10 times larger than a P frame on average. Assume that flow B’s coding efficiency has rP twice as large as nrP. Then, flow A and B might be sending frames periodically in patterns by Group of Picture (GOP) as follows:

Flow A: IDR, rP, rP, rP

Flow B: IDR, rP, nrP, rP, nrP, rP, nrP, rP, nrP

If so, it shows that flow B’s encoder is making efforts to generate discardable packets with more layers in a more balanced distribution, which is desirable.

3. Normalization Marker (NM)

Referring to Figure 2, NM has 3 major components: IFP reconstructor, IFP distribution meter, and normalizer. NM may operate in either "color-aware" (CA) or "color-blind" (CB) mode.

```
        +---------------+    +--------------+    +------------+
        |               |    |              |    |            |
        |      IFP      |    |     IFP      |    |            |
    ===>| Reconstructor |===>| Distribution |===>| Normalizer |===>
        |  in CA or CB  |    |    Meter     |    |            |
        |               |    |              |    |            |
        +---------------+    +--------------+    +------------+
```

Normalization Marker (NM) Architecture

Figure 2

The packets arrive at the IFP reconstructor which determines the IFP of each packet depending on whether NM is in CA or CB mode. This is fed into the IFP distribution meter that keeps a runtime statistics. Then, by the runtime statistics and the IFP of the very packet, the normalizer writes a proper AF-marking in that packet.

3.1. Color-Aware vs. Color-Blind Mode

When NM operates in "color-aware" (CA) mode, it reads the incoming AF-markings that are carried in the packets as the drop precedence. This CA mode should be supported in all NM implementations.

When NM operates in "color-blind" (CB) mode, which is optionally supported, it reads certain bits field(s) other than the AF-markings in the packets to determine the actual drop precedence of that packet. This implies that NM may need DPI in the packets, e.g. parsing into H.264 AVC header in each RTP packets, or alternatively use some method where the drop precedence is carried from the encoder in a customized bits field other than the AF-marking in each packet.

In comparison, CB is more complex than CA in implementation. However, CB could probabily produce better normalization results because the AF-markings are actually outcomes of an n-to-1 mapping from IFPs, as previoulsy mentioned in Section 2.3, which can reduce granularity, e.g. for IFPs x and y, if x > y at encoder, it is possible that AF(x) = AF(y) when NM sees those packets in CA mode. On the contrary, NM in CB mode may reconstruct IFPs x > y for those packets by local DPI.

Note that NM in CB mode may fail to determine the IFP of a packet for various reasons at runtime. If so, NM should randomly assign an IFP to each of those packets with an even distribution over the IFPs. The failure could be due to payload encryption that prevents DPI. Another reason may be that the NM does not support the codec used for encoding those packets in the flow. For example, an NM might only support H.264 AVC but is unable to parse packets in H.264 Annex G (SVC), so it fails to determine the IFPs of packets in an H.264 SVC flow.

3.2. Distribution Meter

The IFP distribution meter keeps a runtime statistics of the IFPs per flow so that the normalizer will be able to assign a proper AF-marking for each packet. The types of statistics to collect at runtime depend on the NM algorithm in the implementation.

For example, an NM implementation may keep a counter of packets per IFP in a flow since the beginning of the flow’s lifetime. Another implementation may choose to keep only the running average of the packet counter per IFP. An even simpler implementation may choose to keep only the running average of IFPs of all packets per flow.

3.3. Normalizer

The normalizer should reference the runtime statistics kept by the IFP distribution meter, and adaptively map the IFP of the very packet to an AF marking, such that the resulting AF-marking distributions for all flows are similar or even identical to a target distribution.

The target distribution of an NM can be simply an even distribution over all possible AF-markings in the AF PHB group. However, in a more complex NM implementation, it may allow configuration for other target distributions as appropriate with the AQM configuration.

4. Acknowledgements

The authors would like to thank many colleagues for comments and supports, and thank Shuai Dai for testing NM with actual H.264 video endpoints.

5. IANA Considerations

This memo includes no request to IANA.

6. Security Considerations

This memo has no security consideration at the time of writing.

7. References

7.1. Normative References

[RFC2475] Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z., and W. Weiss, "An Architecture for Differentiated Services", RFC 2475, December 1998.

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

[RFC2697] Heinanen, J. and R. Guerin, "A Single Rate Three Color Marker", RFC 2697, September 1999.

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

[RFC2998] Bernet, Y., Ford, P., Yavatkar, R., Baker, F., Zhang, L., Speer, M., Braden, R., Davie, B., Wroclawski, J., and E. Felstaine, "A Framework for Integrated Services Operation over Diffserv Networks", RFC 2998, November 2000.

[RFC3260] Grossman, D., "New Terminology and Clarifications for Diffserv", RFC 3260, April 2002.

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

7.2. Informative References

[H264] ITU-T Recommendation H.264, "Advanced video coding for generic audiovisual services", March 2010.

[RFC3550] Schulzrinne, H., Casner, S., Frederick, R., and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications", STD 64, RFC 3550, July 2003.

[RFC6184] Wang, Y., Even, R., Kristensen, T., and R. Jesup, "RTP Payload Format for H.264 Video", RFC 6184, May 2011.

Authors' Addresses

Cheng-Jia Lai
Cisco Systems
170 West Tasman Drive
San Jose, CA 95134
US

Email: chelai@cisco.com

Wenyi Wang
Cisco Systems
170 West Tasman Drive
San Jose, CA 95134
US

Email: wenywang@cisco.com

Stan Yang
Cisco Systems
170 West Tasman Drive
San Jose, CA 95134
US

Email: stanyang@cisco.com

Toerless Eckert
Cisco Systems
170 West Tasman Drive
San Jose, CA 95134
US

Email: eckert@cisco.com

Fred Yip
Cisco Systems
San Diego, CA
US

Email: fyip@cisco.com