

Workgroup: Independent Submission  
Internet-Draft:  
draft-dong-priority-rtp-packet-02  
Published: 10 July 2022  
Intended Status: Informational  
Expires: 11 January 2023  
Authors: L. Dong  
Futurewei Technologies Inc.  
R. Li  
Futurewei Technologies Inc.  
S. Clayman M. Sayit  
University College London Ege University  
**Discarding Priority of RTP Video Packets**

## Abstract

This document illustrates that significance difference or discarding priority might exist among RTP packets which encapsulate video streaming data with the existing modern video codecs, i.e., H.264/AVC, SVC, H.265/HEVC and H.266/VVC.

The document overviews the RTP NALU header format for the existing modern video codecs. Each contains at least one field that indicates the RTP packet's relative significance within the video stream. With the dominance of video traffic in the Internet, selectively dropping RTP packets from competing video streams according to their significances or discarding priorities could be a complementary mechanism when dealing with network congestion. The document proposes the Differentiated Services Code Point (DSCP) value mapping to the RTP packet discarding priority carried in the RTP NALU header. The document also proposes a new Hop-by-Hop Extension Header (HbH-EH) with a value that is copied from the discarding priority of the RTP packet, if the 6-bit DSCP value is not long enough for the mapping.

## 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 <https://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 11 January 2023.

## Copyright Notice

Copyright (c) 2022 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 (<https://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 Revised BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Revised BSD License.

## Table of Contents

- [1. Introduction](#)
- [2. Terms and Abbreviations](#)
- [3. Packet Level Priority](#)
  - [3.1. Packet Level Priority Difference in H.264 RTP Packets](#)
  - [3.2. Packet Level Priority Difference in SVC RTP Packets](#)
  - [3.3. Packet Level Priority Difference in H.265 RTP Packets](#)
  - [3.4. Packet Level Priority Difference in H.266 RTP Packets](#)
- [4. Implementation of Priority-Based Discarding of RTP Video Packets](#)
- [5. IANA Considerations](#)
- [6. Security Considerations](#)
- [7. Acknowledgements](#)
- [8. Informative References](#)
- [Authors' Addresses](#)

## 1. Introduction

The modern video codecs, e.g., H.264/AVC [[H.264](#)], SVC [[H.264](#)], H.265/HEVC [[H.265](#)], and H.266/VVC [[ISO23090-3](#)] [[VVC](#)] use the NAL-unit-based syntax structure. The NAL unit structure provides convenient packetization/framing of video data to be transmitted in packet-based systems using transport protocols such as RTP [[RFC3550](#)]. The transport layer can identify the boundaries among adjacent NAL units without use of start code. Therefore, the overhead for these start codes can be eliminated. Depending on the characteristics of the NAL unit(s) encapsulated in a RTP packet, the priority/importance of RTP packets from the same video streaming flow could differ from each other. In the following, we firstly overview how the priority information is carried in RTP packets for H.264/AVC, SVC, H.265/HEVC, and H.266/VVC by referring to [[RFC6184](#)] [[RFC6190](#)] [[RFC7798](#)] [[RTP.VVC](#)] respectively. Next we discuss how to make the network

layer aware of and utilize such priority information for selective packet dropping when network congestion happens and outgoing buffer overflows.

## **2. Terms and Abbreviations**

The terms and abbreviations used in this document are listed below.

- \*AF: Assured Forwarding
- \*AP: Aggregation Packet
- \*AVC: Advanced Video Coding
- \*DF: Default Forwarding
- \*DSCP: Differentiated Services Code Point
- \*EF: Expedited Forwarding
- \*HDTV: High Definition Television
- \*HEVC: High Efficiency Video Coding
- \*HbH-EH: Hop-by-Hop Extension Header
- \*IDR: Instantaneous Decoding Refresh
- \*FU: Fragmentation Unit
- \*MANE: Media Aware Network Element
- \*MTAP: Multi-Time Aggregation Packet
- \*NAL: Network Abstract Layer
- \*PACI: PAYload Content Information
- \*PHB: Per Hop Behavior
- \*QoE: Quality of Experience
- \*QoS: Quality of Service
- \*RTP: Real Time Protocol
- \*STAP: Signal-Time Aggregation Packet
- \*SNR: Signal-to-Noise Ratio
- \*SVC: Scalable Video Coding

\*VCL: Video Coding Layer

The above terminology is defined in greater details in the remainder of this document.

### 3. Packet Level Priority

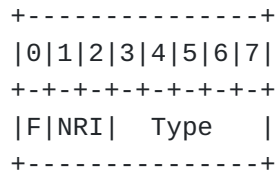
For different versions of video encoding schemes, the RTP packet payload format has been and is being standardized. Within a video flow, the importance or discarding priority can differ among different RTP packets, depending on the NAL unit(s) encapsulated in the RTP packets. In the following, we give a brief overview of such property, which is shown in different versions of video encoders.

#### 3.1. Packet Level Priority Difference in H.264 RTP Packets

The H.264 video codec [H.264] has a very broad application range that covers all forms of digital compressed video, from low bitrate Internet streaming applications to HDTV broadcast and digital cinema applications with nearly lossless coding. The coded video data is organized into NAL units, each of which contains an integer number of bytes. The H.264/AVC specification adopts a byte stream format. Each NAL unit has a prefix of a specific pattern of three bytes, which is called a start code prefix. The boundaries of the NAL unit can then easily be detected by searching the coded data for this unique start code prefix pattern. A set of NAL units in a specified form comprises as an access unit. The decoding of each access unit results in one decoded picture.

The syntax and semantics of the NAL unit type octet are specified in [H.264], includes the essential properties of the NAL unit type octet in the NAL unit header. The RTP packet for H.264 video [RFC6184] inherits the same NAL unit header. As shown in Figure 1, the 2 bits NRI field (i.e., nal\_ref\_idc) indicates the relative importance/transport priority of the NRI unit determined by the encoder. A value of 00 indicates that the content of the NAL unit is not used to reconstruct reference pictures for inter picture prediction. Such NAL units can be discarded without risking the integrity of the reference pictures. Values greater than 00 indicate that the decoding of the NAL unit is required to maintain the integrity of the reference pictures. The H.264 specification requires that the value of NRI SHALL be equal to 0 for all NAL units having nal\_unit\_type equal to 6, 9, 10, 11, or 12. For NAL units having nal\_unit\_type equal to 7 or 8 (indicating a sequence parameter set or a picture parameter set, respectively), an H.264 encoder should set the value of NRI to '11'. For coded slice NAL units of a primary coded picture having nal\_unit\_type equal to 5 (indicating a coded slice belonging to an IDR picture), an H.264 encoder sets the value of NRI to '11'. Non-IDR coded slice is

specified with '10' NRI value, coded slice data partition A has '10' NRI value, while partition B and C have '01' NRI value.



The Structure of the H.264 NAL Unit Header.

Figure 1

The 'Type' field indicates the payload format with three different basic payload structures:

\*Single NAL Unit Packet: Contains only a single NAL unit in the payload. The NRI field is associated with this single NAL unit.

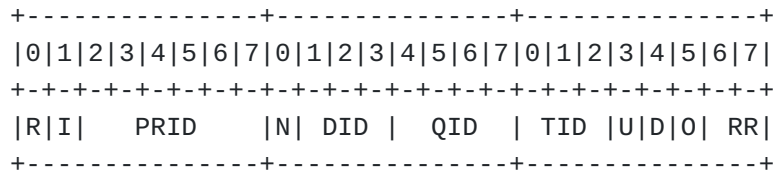
\*Aggregation Packet (AP): Packet type used to aggregate multiple NAL units into a single RTP payload. This packet exists in four versions, the Single-Time Aggregation Packet type A (STAP-A), the Single-Time Aggregation Packet type B (STAP-B), Multi-Time Aggregation Packet (MTAP) with 16-bit offset (MTAP16), and Multi-Time Aggregation Packet (MTAP) with 24-bit offset (MTAP24). A NAL unit header is followed by one or more NAL units in aggregation packets. The value of NRI is the maximum of all the NAL units carried in the aggregation packet.

\*Fragmentation Unit (FU): Used to fragment a single NAL unit over multiple RTP packets. It exists with two versions, FU-A and FU-B respectively. Each FU packet has a FU indicator which has the same format as above. The value of the NRI field is set according to the value of the NRI field in the fragmented NAL unit, which means all the FU packets belong to the same NAL unit have the same NRI value.

### 3.2. Packet Level Priority Difference in SVC RTP Packets

Scalable Video Coding (SVC) extension of the H.264/AVC video coding standard is specified in Amendment 3 to ISO/IEC 14496 Part 10 [[ISO IEC14496-10](#)] and equivalently in Annex G of ITU-T Rec. H.264 [[H.264](#)]. SVC defines a coded video representation in which a given bitstream offers representations of the source material at different levels of scalability: spatial (picture size), quality (or Signal-to-Noise Ratio (SNR)), and temporal (pictures per second). Bitstream components associated with a given level of spatial, quality, and temporal fidelity are identified using corresponding parameters in the bitstream: `dependency_id`, `quality_id`, and `temporal_id`. There are

three additional octets in the NAL unit header of SVC RTP packets [RFC6190], which are shown in [Figure 2](#).



Additional Octets in the SVC NAL Unit Header.

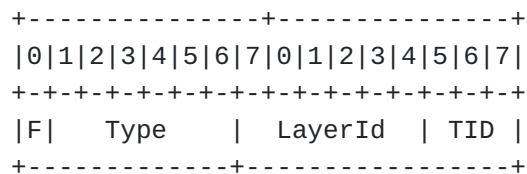
Figure 2

The priority of a NAL unit in SVC video stream can be further specified by the priority\_id field (PRID), which has 6 bits. A lower value of PRID indicates a higher priority.

### 3.3. Packet Level Priority Difference in H.265 RTP Packets

The H.265/HEVC [H.265] significantly improves coding efficiency over H.264. Similarly, H.265 also includes a Video Coding Layer (VCL), which is often used to refer to the coding-tool features, and a Network Abstraction Layer (NAL), which is often used to refer to the systems and transport interface aspects of the codecs. HEVC includes an improved support of temporal scalability over H.264, by inclusion of the signaling of TemporalId in the NAL unit header. HEVC maintains the NAL unit concept of H.264 with modifications. The RTP packet for H.265/HEVC video [RFC7798] uses a two-byte NAL unit header as shown in [Figure 3](#).

The 3 bits field TID specifies the temporal identifier of the NAL unit plus 1. The value of TemporalId is equal to TID minus 1. The TID value indicates (among other things) the relative importance of an RTP packet. For example, because NAL units belonging to higher temporal sub-layers are not used for the decoding of lower temporal sub-layers. A lower value of TID indicates a higher importance. More-important NAL units might need to be better protected against transmission loss or packet dropping than less-important NAL units.



The Structure of the HEVC NAL Unit Header.

Figure 3

The type field indicates the different types of RTP packet payload structures.

\*Single NAL Unit Packet: Contains only a single NAL unit in the payload. The TID field is associated with this single NAL unit.

\*Aggregation Packet (AP): Packet type used to aggregate multiple NAL units into a single RTP payload. A payload header is followed by one or more NAL units in aggregation packets. The value of TID is set as the lowest value of TID of all the aggregated NAL units.

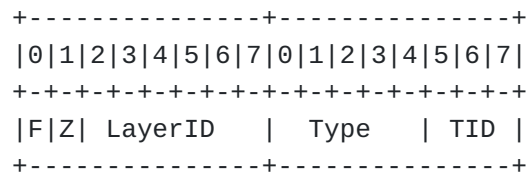
\*Fragmentation Unit (FU): Used to fragment a single NAL unit over multiple RTP packets. Each FU packet has a FU payload header which has the same format as above. The value of the TID field is set according to the value of the TID field in the fragmented NAL unit, which means all the FU packets belong to the same NAL unit have the same TID value.

\*Payload Content Information (PACI): Used to increase the payload header efficiency. The value of TID is a copy of the TID field of the PACI payload NAL unit or NAL-unit-like structure.

### 3.4. Packet Level Priority Difference in H.266 RTP Packets

Versatile Video Coding (VVC) is formally published as both ITU-T Recommendation H.266 [[VVC](#)] and ISO/IEC International Standard 23090-3 [[ISO23090-3](#)]. VVC is reported to provide significant coding efficiency gains over H.265/HEVC, and other earlier video codecs. The RTP payload format for H.266/VVC [[RTP.VVC](#)] allows for packetization of one or more Network Abstraction Layer (NAL) units in each RTP packet payload as well as fragmentation of a NAL unit into multiple RTP packets.

VVC maintains the NAL unit concept of HEVC with modifications. VVC uses a two-byte NAL unit header, as shown in [Figure 4](#). The payload of a NAL unit refers to the NAL unit excluding the NAL unit header.



The Structure of the VVC NAL Unit Header.

Figure 4

Similar to H.265, the TID value indicates (among other things) the relative importance of an RTP packet, for example, because NAL units belonging to higher temporal sublayers are not used for the decoding of lower temporal sublayers. A lower value of TID indicates a higher importance. More-important NAL units might need to be better protected against transmission loss or packet dropping than less-important NAL units.

The LayerID field is used to identify the layer a NAL unit belongs to, wherein a layer may be, e.g., a spatial scalable layer, a quality scalable layer, a layer containing a different view, etc. The LayerID has integer values, where higher values designate components that are higher in the hierarchy. Decoding of a particular component requires the availability of all the components it depends upon, either directly, or indirectly. So the NAL unit with lower LayerID would be likely be used to predict the NAL units with higher LayerID, therefore likely to be more important.

The type field indicates the different types of RTP packet payload structures.

\*Single NAL Unit Packet: Contains only a single NAL unit in the payload. The TID field is associated with this single NAL unit.

\*Aggregation Packet (AP): Packet type used to aggregate multiple NAL units into a single RTP payload. A payload header is followed by one or more NAL units in aggregation packets. The value of TID is set as the lowest value of TID of all the aggregated NAL units.

\*Fragmentation Unit (FU): Used to fragment a single NAL unit over multiple RTP packets. Each FU packet has a FU payload header which has the same format as above. The value of the TID field is set according to the value of the TID field in the fragmented NAL unit, which means all the FU packets belong to the same NAL unit have the same TID value.

#### **4. Implementation of Priority-Based Discarding of RTP Video Packets**

Due to the explicit layering in the protocol stack, the upper layer data or headers are transparent to the network layer. The priority or importance associated with the NAL units encapsulated in RTP packets is invisible to intermediate routers. The concept of media-aware network element (MANE) was introduced in [\[RFC6184\]](#), which is a network element, such as a middlebox or application layer gateway that is capable of parsing certain aspects of the RTP payload headers or the RTP payload and reacting to the contents. The concept of a MANE goes beyond normal routers or gateways in that a MANE has to be aware of the signaling (e.g., to learn about the payload type



mappings of the media streams) and that it has to be trusted when working with Secure Real-time Transport Protocol (SRTP) [[RFC3711](#)]. The advantage of using MANEs is that they allow packets to be dropped according to the needs of the media coding. For example, if a MANE has to drop packets due to congestion on a certain link, it can identify and remove those packets whose elimination produces the least adverse effect on the user experience.

MANEs can access the field that indicates the importance of the NAL unit, which was overviewed in the previous section. In summary:

- \*The two bits NRI field in H.264 and SVC NAL unit header.
- \*The 3 bits TID field in H.265 and H.266 NAL unit header.
- \*The 6 bits PRID field in SVC NAL unit extension header, which provides even finer granularity of priority differentiation for NAL units in SVC.
- \*The 6 bits LayerID field in H.266 NAL unit payload header, which provides even finer granularity of priority differentiation for NAL units in VVC.

MANE is an overlay network element that might be co-located with a few routers, e.g., at network edge. So when network congestion happens in other routers that is not deployed with MANE, the packet dropping is subject to DiffServ classification [[RFC2475](#)]. DiffServ uses a 6-bit differentiated services code point (DSCP) in the 8-bit differentiated services field (DS field) in the IP header for packet classification purposes. In theory, a network could have up to 64 different traffic classes by using the 64 available DSCP values. However, the commonly defined per-hop behaviors only include 4 categories:

- \*Default Forwarding (DF) PHB, which is typically best-effort traffic.
- \*Expedited Forwarding (EF) PHB, which is dedicated to low-loss, low-latency traffic.
- \*Assured Forwarding (AF) PHB, which gives assurance of delivery under prescribed conditions
- \*Class Selector PHBs, which maintain backward compatibility with the IP precedence field.

We consider the two video types: interactive video and non-interactive video. The video stream from both types could be encoded according to H.264, SVC, H.265, H.266. For H.264 and SVC, the NAL units have the NRI field to indicate the discarding priority of the

RTP packets. For H.265 and H.266, the NAL units have the TID field to indicate the discarding priority of the RTP packets. The NRI field is of 2 bits, and the TID field is of 3 bits, thus the DSCP value can be mapped according to either the NRI value or the TID value, as well as the video types. In general, the NAL units with the same NRI value or the TID value in interactive video has higher priority than in non-interactive video. The recommended DSCP values for RTP packets according to NRI value and video type are shown in [Table 1](#). The recommended DSCP values for RTP packets according to TID value and video type are shown in [Table 2](#). These values are based on the framework and recommended values in [[RFC4594](#)].

<b>NRI Value</b>	<b>Interactive Video</b>	<b>Non-Interactive Video</b>
11	AF41	AF42
10	AF42	AF43
01	AF31	AF32
00	AF32	AF33

Table 1: Recommended DSCP Values for RTP Packets According to NRI Value and Video Type (with H.264 or SVC Encoder)

<b>TID Value</b>	<b>Interactive Video</b>	<b>Non-Interactive Video</b>
001	AF41	AF42
010	AF42	AF43
011	AF31	AF32
100	AF32	AF33
101	AF21	AF22
110	AF22	AF23
111	AF11	AF12

Table 2: Recommended DSCP Values for RTP Packets According to TID Value and Video Type (with H.265 or H.266 Encoder)

Either the video host or the MANE at the DiffServ domain edge can do the mapping and set up the DSCP value for each RTP packet. The discarding precedence of the RTP packets can be determined when link congestion happens.

Compared to H.265, SVC and H.266 employ additional scalability other than the temporal scalability, namely spatial scalability and quality scalability. Thus in the NAL extension header for SVC, there is an additional field (i.e., PRID) used to indicate the importance of the RTP packet at finer granularity. The PRID field occupies 6 bits additionally. In the NAL unit header for h.266, the LayerID is used to identify the layer a NAL unit belongs to, wherein a layer may be, e.g., a spatial scalable layer, a quality scalable layer, a layer containing a different view, etc. The LayerID field provides

the importance information of the RTP packet at finer granularity as well. The LayerID field occupies 6 bits additionally.

It is not feasible to use the DSCP mapping to indicate the additional discarding precedence provided by the 6 bits PRID, and the 6 bits LayerID. Thus, other solutions need to be explored in the future if discarding precedence at finer granularity is considered to be supported.

## **5. IANA Considerations**

This document requires no actions from IANA.

## **6. Security Considerations**

This document introduces no new security issues.

## **7. Acknowledgements**

## **8. Informative References**

- [H.264] ITU-T, "H.264 : Advanced Video Coding for Generic Audiovisual Services", 2019, <<https://www.itu.int/rec/T-REC-H.264-201906-I/en>>.
- [H.265] "High efficiency video coding, ITU-T Recommendation H.265", 2019, <<http://handle.itu.int/11.1002/1000/14107>>.
- [ISO23090-3] ISO/IEC 23090-3, "Information technology - Coded representation of immersive media Part 3 Versatile Video Coding", 2021, <<https://www.iso.org/standard/73022.html>>.
- [ISO\_IEC14496-10] "ISO/IEC International Standard 14496-10", 2015.
- [RFC2475] Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z., and W. Weiss, "An Architecture for Differentiated Services", RFC 2475, December 1998, <<https://datatracker.ietf.org/doc/html/rfc2475>>.
- [RFC3550] Schulzrinne, H., Casner, S., Frederick, R., and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications", STD 64, RFC 3550, DOI 10.17487/RFC3550, July 2003, <<https://www.rfc-editor.org/info/rfc3550>>.
- [RFC3711] Baugher, M., McGrew, D., Naslund, M., Carrara, E., and K. Norrman, "The Secure Real-time Transport Protocol

(SRTP)", RFC 3711, DOI 10.17487/RFC3711, March 2004,  
<<https://www.rfc-editor.org/info/rfc3711>>.

- [RFC4594] Babiarz, J., Chan, K., and F. Baker, "TConfiguration Guidelines for DiffServ Service Classes", RFC 4594, DOI 10.17487/RFC4594, August 2006, <<https://www.rfc-editor.org/info/rfc4594>>.
- [RFC6184] Wang, Y.-K., Even, R., Kristensen, T., and R. Jesup, "RTP Payload Format for H.264 Video", RFC 6184, DOI 10.17487/RFC6184, May 2011, <<https://www.rfc-editor.org/info/rfc6184>>.
- [RFC6190] Wenger, S., Wang, Y.-K., Schierl, T., and A. Eleftheriadis, "RTP Payload Format for Scalable Video Coding", RFC 6190, DOI 10.17487/RFC6190, May 2011, <<https://www.rfc-editor.org/info/rfc6190>>.
- [RFC7798] Wang, Y.-K., Sanchez, Y., Schierl, T., Wenger, S., and M. Hannuksela, "RTP Payload Format for High Efficiency Video Coding (HEVC)", RFC 7798, DOI 10.17487/RFC7798, March 2016, <<https://www.rfc-editor.org/info/rfc7798>>.
- [RTP.VVC] Zhao, S., Black, D., Wnger, S., Sanchez, Y., and Y. Wang, "RTP Payload Format for Versatile Video Coding (VVC)", <<https://www.ietf.org/archive/id/draft-ietf-avtcore-rtp-vvc-14.html>>.
- [VVC] "Versatile Video Coding, ITU-T Recommendation H.266", 2020, <<http://www.itu.int/rec/T-REC-H.266>>.

#### Authors' Addresses

Lijun Dong  
Futurewei Technologies Inc.  
United States of America

Email: [lijun.dong@futurewei.com](mailto:lijun.dong@futurewei.com)

Richard Li  
Futurewei Technologies Inc.  
United States of America

Email: [richard.li@futurewei.com](mailto:richard.li@futurewei.com)

Stuart Clayman  
University College London  
United Kingdom

Email: [s.clayman@ucl.ac.uk](mailto:s.clayman@ucl.ac.uk)

Muge Sayit  
Ege University  
Turkey

Email: [mugefesci@gmail.com](mailto:mugefesci@gmail.com)