

Independent Submission
Internet-Draft
Intended status: Informational
Expires: 23 April 2022

L. Dong
K. Makhijani
R. Li
Futurewei Technologies Inc.
20 October 2021

A Use Case of Packets' Significance Difference with Media Scalability
draft-dong-usecase-packet-significance-diff-01

Abstract

This document introduces a use case of packets' significance difference embedded with media scalability. With the dominance of video traffic on the Internet, selectively dropping packets or parts of packets from competing media streams becomes a complementary mechanism when dealing with network congestion.

The document describes the characteristics of media scalability, some limitations of existing end-to-end congestion control mechanisms through rate control and adaptation, explains why current ways of entire packet dropping at the traffic class level using in-network active queue management are not most appropriate to meet end users' Quality of Service expectations. The document identifies that there exists "significance difference" among packets or even among parts of the packets within a flow, and brings out a new set of requirements for application and network to support packet significance difference to improve the Quality of Experience of end users.

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 23 April 2022.

Internet-Draft [draft-dong-packet-significance-diff](#)

October 2021

Copyright Notice

Copyright (c) 2021 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 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	2
2.	Terms and Abbreviations	3
3.	Media Scalability and Congestion Control	4
4.	Packet Dropping	5
5.	Significance Difference Among Packets and Within Packets . .	6
6.	New Requirements	7
7.	IANA Considerations	8
8.	Security Considerations	8
9.	Acknowledgements	8
10.	Informative References	8
	Authors' Addresses	11

[1.](#) Introduction

Recent studies [[CiscoNetworkingIndex](#)] show that IP video traffic will be 82 percent of all consumer Internet traffic by 2021 in a global scale, up from 73 percent in 2016. Live video has grown 15-fold from 2016 to 2021, accounts for 13 percent of Internet video traffic by 2021. VR (Virtual Reality) and AR (Augmented Reality) traffic has increased 20-fold between 2016 and 2021, at a CAGR (Compound Annual Growth Rate) of 82 percent. With the rapid growth of multimedia streaming traffic, it is increasingly likely that multiple streaming flows share a bottleneck link, which would inevitably cause network congestion. Today's transport protocols and Internet protocols are oblivious to multimedia streaming applications or end users' QoE (Quality of Experience) expectations. From the perspective of user experience and user expectation, the following two observations could

be made.

- * It is very likely that a user may prefer to acquire the media content in a somewhat degraded quality that is above the tolerance threshold rather than getting nothing at all for a few seconds.

- * A user may be particularly interested in certain group of blocks belonging to the interested objects in the media content (i.e., Region of Interest, RoI). It is necessary to prevent the RoI blocks from being lost during transmission.

At the beginning of this document, the different types of scalability are discussed in current video codecs, facilitating the rate control and adaptation mechanisms carried out in video segments when dealing with network congestion during the media streaming. It is acknowledged that such mechanisms have efficiently improved users' QoE. However, the packets on the wire cannot avoid the possibility of being entirely dropped when the bottleneck network nodes cannot retain them due to buffer overflowing during congestion. Thanks to the scalability characteristics designed to the video codecs, it is not hard to find out that the importance or significance of different packets within a media streaming flow or even different parts of the single packet could vary for their usefulness in decoding and recovering the media content to meet receiver's expectation. The document highlights the requirements of making the user's preference and application context aware to the network to help further improve the QoE of media streaming. Accordingly, the network could treat the packets or different parts of the packets according to the characteristics of the packets and end users' preferences.

[2.](#) Terms and Abbreviations

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

- * AR: Augmented Reality
- * CAGR: Compound Annual Growth Rate
- * DASH: Dynamic Adaptive Streaming over HTTP
- * GOP: Group of Picture

- * HAS: HTTP Adaptive Stream
- * HTTP: Hypertext Transfer Protocol
- * QoE: Quality of Experience
- * QoS: Quality of Service
- * SNR: Signal-to-Noise Ratio
- * SVC: Scalable Video Coding

- * VR: Virtual Reality

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

[3.](#) Media Scalability and Congestion Control

A visual scene is represented in digital form by sampling the real scene spatially on a rectangular grid in the video image plane and sampling temporally at regular time intervals as a sequence of still frames. Correspondingly, modern media codec [[Conklin2001](#)] [[Kim2001](#)] incorporates three types of "Scalability": i.e., temporal scalability, spatial scalability, and quality scalability, which adapt the media bitstream by adding or removing some portions to/from it in order to match the different needs or preferences of end users as well as to the network conditions.

Temporal scalability refers to scalability designed to allow the frame rate of the video bitstream to be varied using interlayer prediction. Spatial scalability represents the spatial resolution variations with respect to the original image frame. The lower layer provides the basic spatial resolution. The enhancement layer employs the spatially interpolated lower layers and constructs the source video in its full spatial resolution. Quality scalability is also commonly referred to as fidelity or SNR (Signal-to-Noise Ratio) scalability. Each spatial layer could have many quality layers. For example, SVC (Scalable Video Coding)[[SVC](#)] is an H.264 [[H.264](#)] extension that divides a single video bitstream into multiple representations or layers. This hierarchical layered structure

comprises a base layer and two enhancement layers. The media may be scaled up by adding the enhancement layer(s) or scaled down by dropping the enhancement layer(s). The levels of scalability included in the media stream affect the quality of media presented to the end users' devices.

Bursty loss and longer-than-expected delay have catastrophic effect on QoE to end-users in media streaming. They are usually caused by network congestion. Despite all kinds of congestion control mechanisms developed in the community over the decades [[Saadi2019](#)] [[Adams2013](#)], they often target different goals, e.g., link utilization improvement, loss reduction, fairness enhancement. By leveraging the flexibility and variety of media qualities provided by different types of media scalability, for media streaming, minimizing the possibility of network congestion can often be achieved by rate control and media adaptation methods.

Existing rate control and adaptation methods [[Bentaleb2019](#)] [[Wu2001](#)] can be at source-side and receiver-side, which are carried at end devices and servers, respectively.

- * In source-based schemes [[Wu2000](#)] , source regulates the sending rate to maintain the packet loss ratio below a threshold by employing the feedback from probing experiments, or source determines the sending rate through a TCP-friendly model. However, some constraints exist, media codecs can usually only adjust their output rates in a much more coarse-grained fashion than, for example, TCP. Users' QoE would also suffer if encoding rates are switched too frequently.
- * HTTP (Hypertext Transfer Protocol)-based dynamic video adaptation methods [[Kua2017](#)] could be driven by source. The server collects the feedback from the network and client (e.g., dynamic variation of network bandwidth and receiving buffer capacity of the client), and accordingly, the video quality will be adapted and streamed. On the other hand, adaptation techniques are also proposed at receiver-side, which mainly use DASH (Dynamic Adaptive Streaming over HTTP) [[MPEG-DASH-SAND](#)] [[MPEG-DASH](#)] and HAS (HTTP Adaptive Stream) for streaming adapted video data.

- * The receiver-based rate control [[McCanne1996](#)] is typically used in multicasting scalable media content, which is split into multiple layers, with each layer corresponding to one channel in the multicast tree. Receivers could regulate their own receiving rates by adding/dropping channels. Thus receiver-based rate has its limited usage in unicasting. All these techniques consider full quality while streaming from sender to receivers; hence, they consume more resources in the network.

[4.](#) Packet Dropping

Acknowledging the benefits offered by various congestion control and congestion avoidance mechanisms, we would like to point out that the feedback and rate adaption might not be prompt enough to cope with the dropping of packets on the wire.

In the current Internet, a packet is treated as the minimal, independent, and self-sufficient unit that gets classified, forwarded, or dropped completely by a network node, according to the local configuration and congestion condition. Although congestion discard can be mitigated by a mixture of ingress traffic shaping and active queue management mechanisms [[Thiruchelvi2008](#)] [[Adams2013](#)] to avoid any network resource overdrawn, it is not feasible to be deployed on a large scale, meanwhile wastes network resources preparing for the worst possible scenario.

DiffServ [[RFC2475](#)] is used to manage resources such as bandwidth and queuing buffers on a per-hop basis between different classes of traffic. The Internet traffic may be separated into different classes with differentiated priorities. This allows preferential treatment for latency or loss sensitive traffic over more tolerant applications, for example those that can afford retransmission. However, with video traffic dominating Internet traffic, flows of media streaming applications with the same class still compete for network resources when encountering bottleneck links and fighting network congestion, preference decided on traffic class would not be effective to eliminate the possibility of degraded service levels or packet drops due to collisions with each other.

The routers treat every bit/byte in the packet payload equally, which means every bit/byte has the same significance to the routers. Each

to-be-dropped packet is discarded completely. If the transport layer protocol is TCP, after timeout or duplicate acknowledgements received at the sender, the sender may re-try to send the dropped packet before the maximum number of re-transmissions reaches.

Retransmission of packets wastes network resources, reduces the overall throughput of the connection and causes longer latency for the packet delivery. The study [[RFC8836](#)] has shown that a loss rate of 1% is tolerable to users while a loss rate of 3% is intolerable to most users who found the quality to be annoying (or worse), according to the subjective opinions of the effects of packet loss on media quality. Therefore, the current way of handling network congestion by discarding the packet entirely and retransmitting the packets in a blind-of-application-context manner is not very suitable for media streaming.

5. Significance Difference Among Packets and Within Packets

With the various scalability implemented in the media codec, some bits of an encoded media stream are more important than others. Bits belonging to base layer usually are more significant to the decoder than bits belonging to enhancement layers. For example, I-frames hold complete picture data [[Orosz2015](#)] and is frequently referenced by the subsequent frames. It is inserted by the encoder when the scene changes. Losing the first I-frame in the GOP (Group of Pictures) would cause video picture even missing for few seconds, because P- and B-frames referencing to the I-frame would not be decoded nor displayed either. Thus, I-frames are most essential in the media stream, which have the most effect on perceived video quality, and such effect can last through the whole GOP. P- and B-frames are inserted at appropriate places to reduce the video size or bitrate and are tuned to maintain a certain video quality level. P-frame stands for Predicted Frame and allows macroblocks to be compressed using temporal prediction in addition to spatial

prediction. A P-frame might be referenced by a P frame after it, or a B frame before or after it. B-frame stands for bi-directional frame, which can be predicted using backward prediction and forward prediction. A B-frame can act as a reference, and if so, it is termed as a reference B-frame. If a B-frame is not to be used as a reference, it is called a non-reference B-frame. Video scenes with a low level of movement are less sensitive to both B-frame and P-frame packet loss, alternatively video scenes with a high level of movement

are more sensitive to both B-frame and P-frame packet loss. A lost P-frame can impact the remaining part of the GOP. A lost B-frame has only local effects in a slowly moving content or with large static background. In a scene of a dynamically moving content, losing B-frame has more dramatic impact and its scale can be as far-reaching as a P-frame loss.

As another example, macroblocks that are identified to represent the objects in RoI are likely more important than other macroblocks of non-RoI regions. For packets carrying RoI macroblocks in the media stream need to have higher priority to be retained compared to other packets carrying non-RoI macroblocks.

According to the characteristics of frames contained in the video packet payload, namely: frame type, whether the frames are referenced by other frames, movement level of the pictures, whether the picture contained in the packet belongs to RoI or not, etc., significance difference could present among packets for the video decoding at the receiver side and the QoE improvement of end users. The dropping priority is possibly implemented at packet level in the network.

On the other hand, let's say that the end-users can reveal their preferences to the network, e.g., degree of tolerance to the decoded media content' quality degradation, which might reflect visually such as resolution reduction, missing objects in non-RoI regions, the network could selectively drop packets in a differentiated manner according to such information. This avoids retransmission or delay of those packets with higher significance, reduce the experienced end-to-end latency of end users, and maintain the continuous streaming of the media. This is achieved at the cost of dropping lower-significance packets.

6. New Requirements

We have discussed in the previous sections that due to the various types of scalability implemented in the media codecs, "significance difference" exists among packets or even among parts of the packets. In other words, some packets containing the more important macroblocks (e.g., RoI macroblocks, base layer macroblocks) show higher significance than other packets for the media decoding at the

the network be able to treat the packets of media streams in a differentiated manner and at finer granularity than DiffServ, the application shall reveal some information to the network to enable selective packet dropping or partial packet dropping. For example, an API could be implemented to input such information or metadata from the application. which might be mapped to IPv6 extension header, IPv4 options or a dedicated metadata field in the IP header. Some examples of such information or metadata are listed below:

- * Receiving end user's preference on media quality, e.g. tolerable quality degradation regarding for example resolution.
- * Characteristics of media content contained in the packets, e.g., frame type, whether the packet contains frames that are referenced by other frames, movement level of the video sample contained in the packet.
- * Labeling of the packets or some parts of the packets that correspond to receiver's interested objects as RoI.

Correspondingly, the network shall be able to leverage the above information revealed by the application, and selectively drop packets or parts of the packets from competing media streaming flows with precedence order when network congestion happens. The retransmission could be maximumly eliminated. The receiving end user is able to consume the delivered packets as many as possible in-time with acceptable quality.

7. IANA Considerations

This document requires no actions from IANA.

8. Security Considerations

This document introduces no new security issues.

9. Acknowledgements

10. Informative References

[Adams2013]

Adams, R., "Active Queue Management: A Survey", IEEE Communications Surveys and Tutorials, vol. 15, no. 3, pp. 1425-1476, 2013, <<https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6329367>>.

[Bentaleb2019]

Bentaleb, A., Taani, B., Begen, A. C., Timmerer, C., and R. Zimmermann, "A Survey on Bitrate Adaptation Schemes for Streaming Media Over HTTP", IEEE Communications Surveys and Tutorials, vol. 21, no. 1, pp. 562–585, 2019, <<https://ieeexplore.ieee.org/document/8424813>>.

[CiscoNetworkingIndex]

Cisco, "Cisco Visual Networking Index: Forecast and Methodology, 2016 to 2021", June 2017, <<https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html>>.

[Conklin2001]

Conklin, G. J., Greenbaum, G. S., Lillevold, K. O., Lippman, A. F., and Y. A. Reznik, "Video Coding for Streaming Media Delivery on the Internet", IEEE Transactions on Circuits and Systems for Video Technology, vol. 11, no. 3, pp. 269–281, 2001, <<https://ieeexplore.ieee.org/document/911155>>.

[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>>.

[Kim2001]

Kim, T., "Scalable video Streaming Over Internet", Ph.D. Thesis, School of Electrical and Computer Engineering, Georgia Institute of Technology, January 2005, <<https://smartech.gatech.edu/handle/1853/6829>>.

[Kua2017]

Kua, J., Armitage, G., and P. Branch, "A Survey of Rate Adaptation Techniques for Dynamic Adaptive Streaming Over HTTP", IEEE Communications Surveys and Tutorials, vol. 19, no. 3, pp. 1842–1866, 2017, <<https://ieeexplore.ieee.org/document/7884970>>.

[McCanne1996]

McCanne, S., Jacobson, V., and M. Vetterli, "Receiver-Driven Layered Multicast", ACM Sigcomm, pp. 117–130, 1996, <<http://www.cs.toronto.edu/syslab/courses/csc2209/06au/papers/recmc.pdf>>.

[MPEG-DASH]

ISO/IEC, "23009-1:2019, Dynamic Adaptive Streaming over HTTP (DASH) – Part 1: Media Presentation Description and

Segment Formats", 2019,
<<https://www.iso.org/standard/79329.html>>.

[MPEG-DASH-SAND]

ISO/IEC, "23009-5:2017, Dynamic Adaptive Streaming over HTTP (DASH) – Part 5: Server and Network Assisted DASH (SAND)", February 2017,
<<https://www.iso.org/standard/69079.html>>.

[Orosz2015]

Orosz, P., Skopko, T., and P. Varga, "Towards Estimating Video QoE Based on Frame Loss Statistics of the Video Streams", DOI: 10.1109/INM.2015.7140482, IFIP/IEEE International Symposium on Integrated Network Management (IM), pp. 1282-1285, 2015,
<<https://ieeexplore.ieee.org/document/7140482>>.

[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>>.

[RFC8836] Jesup, R. and Z. Sarker, "Congestion Control Requirements for Interactive Real-Time Media", [RFC 8836](#), January 2001,
<<https://datatracker.ietf.org/doc/html/rfc8836>>.

[Saadi2019]

Al-Saadi, R., Armitage, G., But, J., and P. Branch, "A Survey of Delay-Based and Hybrid TCP Congestion Control Algorithms", IEEE Communications Surveys and Tutorials, vol. 21, no. 4, pp. 3609-3638, 2019,
<<https://ieeexplore.ieee.org/document/8668433>>.

[SVC]

Schwarz, H., Marpe, D., and T. Wiegand, "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard", IEEE Transactions on Circuits and Systems for Video Technology, vol. 17, no. 9, 1103-1120, 2007,
<<https://ieeexplore.ieee.org/document/4317636>>.

[Thiruchelvi2008]

Thiruchelvi, G. and J. Raja, "A Survey On Active Queue Management Mechanisms", International Journal of Computer

Science and Network Security, vol. 8, 2008,
<https://www.researchgate.net/publication/310468829_A_Survey_on_Active_Queue_Management_Techniques>.

- [Wu2000] Wu, D., Hou, Y., and Y. Zhang, "Transporting Real-Time Video Over the Internet: Challenges and approaches", Proceedings of the IEEE, vol. 88, no. 12, 1855-1875, 2000, <http://www.wu.ece.ufl.edu/mypapers/ProcIEEE_camera.pdf>.

Dong, et al.

Expires 23 April 2022

[Page 10]

Internet-Draft

[draft-dong-packet-significance-diff](#)

October 2021

- [Wu2001] Wu, D., Hou, Y., Zhu, W., Zhang, Y., and J. Peha, "Streaming Video Over the Internet: Approaches and Directions", IEEE Transactions on Circuits and Systems for Video Technology, vol. 11, no. 3, pp. 282-300, 2001, <<https://ieeexplore.ieee.org/document/911156>>.

Authors' Addresses

Lijun Dong
Futurewei Technologies Inc.

Email: lijun.dong@futurewei.com

Kiran Makhijani
Futurewei Technologies Inc.

Email: kiran.ietf@gmail.com

Richard Li
Futurewei Technologies Inc.

Email: richard.li@futurewei.com

