Characteristics Aware Video Packet Dropping and Packet Wash

Lijun Dong and Richard Li

Futurewei Technologies Inc.
2220 Central Expressway, Santa Clara, CA, USA
email: {lijun.dong,richard.li}@futurewei.com

***Abstract*—In the current implementation, the network layer lacks awareness of the specific characteristics of data within packet payloads. This limitation prevents the network from recognizing the importance of individual packets during their transmission through the network. However, with the increasing prevalence of video streaming as the primary form of Internet traffic, there is a need to address this issue. To tackle this problem, the paper presents an example using MPEG video and proposes mechanisms that allow the network to identify video packets of higher significance. By doing so, these critical packets can be prioritized and prevented from being discarded during instances of network congestion. Furthermore, the proposed mechanisms can be expanded to encompass packets containing multiple video frames with varying priorities. This approach leverages the concept of Qualitative Communication and a corresponding packet wash operation. As a result, video frames that are more important to the decoder can better withstand the risk of being dropped during transmission, especially when facing network congestion. Ultimately, these improvements in packet prioritization and transmission contribute to enhancing the Quality of Experience (QoE) for video receivers.**

***Index Terms*—Packet dropping, network congestion, packet wash, Qualitative Communication, video characteristics, frame type, reference frame, movement level, region of interest.**

I. Introduction

Cisco's newest Visual Networking Index [1] shows that video traffic accounts for 82 percent of all Internet traffic in 2022 in a global scale, up from 75 percent in 2017. Global IP video traffic grows four-fold from 2017 to 2022, with a Compound Annual Growth Rate (CAGR) of 29 percent. Internet video traffic has grown fourfold from 2017 to 2022, with a CAGR of 33 percent. Live Internet video accounts for 17 percent of Internet video traffic by 2022, which has increased 15-fold from 2017 to 2022. Virtual Reality (VR) and Augmented Reality (AR) traffic has increased 12-fold between 2017 and 2022 globally, with a CAGR of 65 percent. Consumer Video-on-Demand (VoD) traffic nearly doubles by 2022. The amount of VoD traffic by 2022 was predicted to be equivalent to 10 billion DVDs per month. 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.

Bursty loss and longer-than-expected delay have catastrophic effects on the Quality of Experience (QoE) to end-users in video streaming, which are usually caused by network congestion. Despite all kinds of congestion control mechanisms developed in the community over the decades [2]–[4], congestion control mechanisms often target different goals, e.g. link utilization improvement, loss reduction, fairness enhancement. For media streaming, minimizing the possibility of network congestion can often be achieved by rate control and video adaptation methods. Acknowledging the benefits offered by various congestion control and congestion avoidance mechanisms, the existing solutions offering feedback and rate adaption might not be prompt enough to cope with the dropping of packets on the wire.

Differentiated services (or DiffServ) [5], [6] is a type of Quality of Service (QoS) mechanism which classifies and manages network traffic on a per-hop basis between different classes of traffic. Internet traffic might be separated into different classes with differentiated priority (usually at flow level). When encountering bottleneck links and fighting network congestion, utilizing a packet dropping priority decided on traffic class is not effective, because the video traffic that shares the same class still compete for network resources. Other than the traffic

979-8-3503-7099-7/24/$31.00 ©2024 IEEE

class included in the IP packet header, the semantics associated with the packet and its payload is invisible to the network. Every packet is treated (e.g., classified, forwarded or dropped) in its entirety as the minimal processing unit. The network protocols always make sure that the destination receives the packet with each bit matching to the original data. In this paper, we raise the questions: (1) Is it possible to disclose certain semantics of the packet payload to the network layer in order to prevent blind packet dropping and allow for consideration of the traffic class associated with the flow? Within a flow, some packets could be more important than other packets. For example, bits belonging to base layer usually are more significant to the decoder than bits belonging to enhancement layers. In this paper, we will try to lay out some major characteristics that are associated with the video packet data. As as result, the packet dropping granularity would be reduced from the flow level to the packet level. (2) Can we further reduce the packet dropping granularity to be within the packet? Qualitative Communication [7]–[9] has been proposed in the literature that considers the packet payload as a combination of multiple chunks, each of which could be treated independently in the network.

In this paper, we focus on the MPEG video streams and propose the methods of providing finer granularity of packet dropping and partial packet dropping in MPEG video flows when network congestion happens. The rest of the paper is arranged as following: in the related work section II, we briefly review the New IP framework and its major component: Qualitative Communication. Section III proposes the methods to enable the video data characteristics awareness and finer granularity of prioritized packet treatment in the network. Section IV presents the performance evaluation. Section V concludes the paper.

II. New IP and Qualitative Communication

The New IP framework [10], [11] is an extension, optimization and evolution of IP with new functions (capabilities, features), and is being designed to be inter-operable with IPv4/v6 and many others. New IP is harnessed by the Contract component. A contract describes a formal service specification that is visible to the network layer (e.g., routers), which includes clauses to describe the type of network service capability, event, condition, actions, and accounting information. Each contract clause could optionally include the **Metadata** associated with the parties involved in the Contract. The contract metadata could include the semantics associated with the packet payload, as well as end user's context and packet delivery requirement.

Qualitative Communication [7], [8] is enabled by the New IP payload, which is divided into multiple chunks and becomes individual unit processed by the routers. The packetization scheme could be significance based [9], and random linear network coding based [12], [13]. Each chunk might have its own corresponding semantics such as: its relative significance level compared to other chunks, its relationship with other chunks, its boundary to adjacent chunks, etc. With New IP contract metadata, such semantics of each chunk can be made aware to the network. The semantics-aware chunk-level payload dropping could be performed to remove the less important chunks form the payload, and reduce the packet size for different types of purposes (e.g., congestion mitigation, in-time guarantee for packet delivery [14]), which is called 'packet wash'. The data received does not need to match bit-to-bit to the original data, yet will be still useful to the end user. Qualitative Communication intends to largely reduce the number of packet re-transmissions or even avoid the re-transmission.

III. Video Data Characteristics Awareness

In the following, we use MPEG as the example to introduce the major characteristics associated with coded video data. A video can be viewed as a sequence of images stacked in the spatial and temporal dimension. MPEG exploits the spatial and temporal redundancy inherent in video images and sequences. The temporal sequence of MPEG frames consists of three types, namely intracoded I-frames, intercoded P-frames and B-frames (P-frames are unidirectionally interpredicted, while B frames are bidirectionally interpredicted). These frame types are designed to strike a balance between random access of frames, compression efficiency, and maintaining decent video quality. MPEG video sequences are composed of Groups of Pictures (GOPs), each includes a limited number of coded frames, including one I frame and one or more P and B frames.

I frames are key frames that provide checkpoints for re-synchronization to provide support of trick modes (e.g., pause, fast forward, rewind) and error recovery. These frames are independently coded and do not refer to any other frames. With I frames being the the first frame in a closed GOP, and is the only frame within in the current GOP which is directly or indirectly referenced by other subsequent frames, the impact of error propagation caused by corrupted or lost frames can be restricted within the GOP, providing certain degree of error resiliency.

P-frames are temporally encoded using motion estimation and compensation techniques. P-frames are initially partitioned into blocks before motion-compensated prediction is applied. The prediction refers to the I-frame, which is the first frame in a GoP before the P-frame. As a result, P-frames are forward predicted or extrapolated and the prediction is unidirectional. An Motion Vector (MV) is calculated between a block in the current P-frame and the previously coded I-frame to determine the value and direction of the prediction for each block.

B-frames are temporally encoded using bidirectional motion-compensated predictions from a forward reference frame and a backward reference frame. The I-frames and P-frames usually serve as reference frames for the B-frames. Due to the bidirectional predictions, B-frames incur higher overheads than P-frames in order to carry twice number of the MVs. B-frames take a longer time to encode compared to I-frames and P-frames. Live video content should employ B-frames minimally. Newer video coding standards allow B-frames to be used as reference frames.

Intracoding a block in a P-frame or B-frame might happen if the prediction residual turns out to be large and the motion compensation is less effective, which means that the P-frame or B-frame might use an I-frame outside of its own GOP (i.e., Open GOP).

A. Important Characteristics of MPEG Video Packets

In this sub-section, we focus on discussing the major characteristics of payload data contained in the MPEG video packets, which could indicate the significance of the packets in the video flow, considering its importance for decoding and maintaining the QoE (Quality of Experience) of end users.

1) Frame Type: As described above, there are three types of frames:

- I frame: An I-frame consists only use intra-prediction. If the I frame of a GOP is corrupted, all other frames within the GOP cannot be decoded.
- P-frame: An P-frame allows macroblocks to be compressed using temporal prediction in addition to spatial prediction.
- B-frame: An B-frame is a frame that can refer to frames that occur both before and after it.

Consequently, we may use two bits to indicate the frame type (e.g., ’00” corresponds to I-frame, ”01” corresponds to P-frame, ”10” corresponds to B-frame).

2) Being Referenced or Not: Another characteristics that shall be considered to determine the significance of a video packet is whether the frame(s) contained in the packet payload is referenced by other frames.

- I frame: I-frame does not refer to any other frame, is at least referenced by a P frame after it. Losing the first I frame in the GOP could jeopardize the video picture quality or even cause the video picture even missing for few seconds, because both P-frames and B-frames referencing to the I frame directly or directly would not be decoded nor displayed either.
- P frame: A P-frame refers to a picture in the past, might be referenced by a P frame after it, or a B frame before or after it. If any P-frame is corrupted, the error may propagate to other P-frames or B-frames.
- B frame: A B-frame can act as a reference in the recent codecs, 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.

Consequently, we may use one bit to indicate whether a frame or frame(s) contained a packet payload is referenced by other frames or not.

3) Movement level and Nature of the Movement: Data loss in the P-frame and B-frame could be concealed with inter-frame interpolation by content from the reference frame or from the MV of the neighbouring blocks depending on the motion of the video. 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. The spatial inconsistency [15] could happen and is proportional to the movement level of the video content if there are artifacts in the B-frames and P-frames.

Consequently, we may use a few bits to indicate the movement level and the degree of the movement regulation. In this paper, we set the number of bits to be 3.

4) Region of Interest: Region of Interest (RoI) is defined as the region in the video samples that the viewer/receiver pays particular attention to. The frames that contain the RoI are consequently more important and shall be treated with higher significance than other non-RoI frames in the video stream. For example, in video surveillance application, a particular object of interest in the video sequence needs to be identified in the video sequence. Such detection can be performed at the sender side while monitoring the activity of the pre-specified person or the pre-defined object. All the video frames containing the single Object of Interest (OoI) would have higher dropping precedence than the remaining frames when encountering network congestion during packet transmission. We may use one bit to identify whether the frame contained in the packet payload belongs to a pre-defined RoI or OoI.

In summary, the following information relevant to the video sample contained in the packet payload can be revealed to the network layer through New IP Metadata, IPv6 extension header, or IPv4 Options field.

- PT: it indicates the packet type. If we only differentiate the packets into two types: video type, and non-video type, then one bit is used for PT. FT: it indicates the type of the frame that is contained in the packet payload. R: it denotes whether the frame is being referenced by other frames. ML: it describes the movement level of the video scene contained in the packet payload data. RoI: it indicates whether the packet payload contains RoI or OoI related frames.

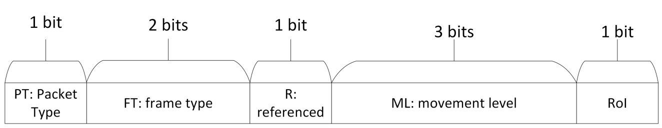

Fig. 1. Metadata related to video sample contained in the packet payload

B. Packet Level Dropping Precedence

When network congestion happens to a router, the router could drop packets based on their priorities. If we consider packets carrying video data packets, the following algorithm could be adopted by a router:

- If $FT = 1$ (I frame), $R = 1$, $RoI = 1$, then $priority = 0$ (0 is regarded as the highest priority.).
- Otherwise, the priority of the video packet is defined as: $priority = FT \times (2-R) \times (8-ML) \times (2-RoI)$, the lower value of $priority$, the higher dropping precedence the packet is.
  - The lower value of FT is, the more important the packet could be.
  - If R is true ($R = 1$), the more important the packet could be.
  - The larger value of ML is, the more important the packet could be.
  - If RoI is true ($RoI = 1$), the more important the packet could be.

The above is a simple and exemplary algorithm. With metadata carried in the packet, a router could have its own priority calculation algorithm. However, independent priority calculation and determination would increase the computation overhead as well as prolong the packet processing and forwarding latency incurred from it. Alternatively, the priority value could also be calculated by the video source, if the same algorithm is standardized and adopted by all video encoders. As a result, only the priority value is needed to be carried in the packet.

C. Significance-Based Packet Wash

A packet payload may incorporate multiple frames of a video stream, according to the following two observations:

- A frame's size might be very small, for example, the motion vector prediction process for B-frames and P-frames might result in a decorrelated residual signal that will mainly comprise small values, if the differences from the reference frame are tiny.
- Jumbo-sized packet [16] is preferable to be adopted due to the following reasons: The signaling and packet header overhead would be reduced carrying the same amount of data in jumbo-sized packets. On the other hand, by adopting jumbo packet, it would reduce the number of smaller packets being processed and forwarded by routers, which in turn decreases the energy consumption of the routers and achieves global carbon emission reduction.

Qualitative Communication introduces the concept and mechanisms of splitting/dividing the payload in a packet itself into smaller units called chunks such that its chunks are treated independently. A delimiter (a special sequence) is used to separate the multiple frames in the payload. The packet wash operation associated with Qualitative Communication can be applied to a video packet if multiple frames are contained in the packet' payload. According to the priority either calculated by the video source or by the router, the packet wash operation could be applied to a video packet. By reducing the packets' sizes, the network congestion could likely be relieved.

The number of frames included in the payload needs to be revealed (denoted by N). The priority information of each frame encapsulated in the packet is included in the metadata (denoted by $Priority$). A flag indicating whether the frame has been removed due to the packet wash operation is included additionally (denoted by $Washed$). For the multiple frames scenario, if the multiple frames are included in the payload from head to tail following the original order, the metadata form is shown in Fig. 2.

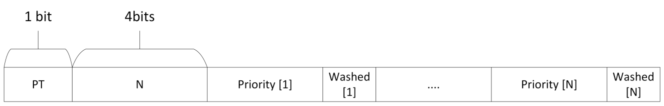

Fig. 2. Metadata form 1—multiple frames scenario

Alternatively, the multiple frames could be included in the payload from head to tail following the highest to least priority of the frames. The frame in the tail always has the least priority, and the packet wash operation is performed from the payload tail as well. Consequently, the original order of the frames needs to be disclosed (denoted by $order$), and the $Washed$ flag would follow the same order to indicate whether a frame from the tail has been removed or not, as shown in Fig. 3. For example, it is assumed that there are 4 frames and the original order is 4, 3, 2, 1. The priority of the frames from highest to least is determined as 3, 1, 4, 2, then the 4 bits in $Washed$ field will indicate whether the frame has been removed for frame 3, 1, 4, 2 correspondingly.

Fig. 3. Metadata form 2—multiple frames scenario

IV. Performance Evaluation

In this section, we present the performance simulation results. We simulated a network scenario, in which there are multiple video streaming flows between different sender and receiver pairs as shown in 4. The intermediate routers might have some probability of getting congested during the configured simulations.

In the first set of simulations, we assume each packet only contains one frame or one portion of a frame. When congestion happens at a router, packet level dropping is applied. We compared the two packet dropping scenarios: (1) `MPEGAware`, in which a router drops packets based on the dropping precedence as proposed in Section III-B ; (2) `Tail`, in which the packets are dropped from the tail of the outgoing queue without considering the significance of the packets. The following parameters are subject to be configured for the simulator:

- Congestion level (CL). It indicates the level of congestion, which is represented by the ratio between the total number of packets that need to be dropped and the total number of packets that arrive at the router's input port.
- Number of congested routers (N). It indicates the number of intermediate routers on the path from a sender to a receiver that are congested during the simulation. It can be represented by the ratio of congested routers times the total number of intermediate routers between the source and destination.

In the first set of simulations, we set the following parameters to be fixed:

- The average number of intermediate routers between the source and destination to be 10.
- The number of frames in a GoP is configured to be 15, and follows the reference relationships as shown in Fig. 5.

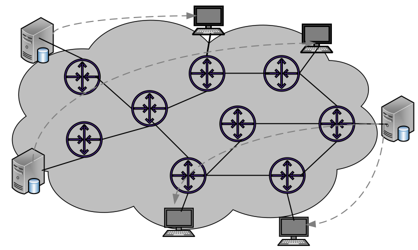

Fig. 4. Network scenario

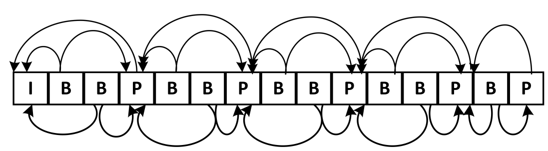

Fig. 5. A sample GoP

Fig. 6, Fig. 7, Fig. 8 show the number of received I frames, P frames, and B frames respectively. The light grey bar shows the number of I frames, P frames and B frames sent from the source. The grey bar shows the number of I frames, P frames and B frames received by the receiver if the tail dropping (`Tail`)is adopted when network congestion happens. The black bar show the number of I frames, P frames and B frames received by the receiver if the proposed MPEG aware mechanism (`MPEGAware`) is adopted when network congestion happens. We can observe that with `MPEGAware`, when CL is less than 0.3, all the I frames, which are the most important frames in a GoP could reach the receiver without any loss. And even though after CL gets higher, `MPEGAware` starts to lose I frames to fight the network congestion, `MPEGAware` still can retain as many I frames as possible. But with `Tail`, the congested router is not able to understand which packet is more significant and needs to be prevented from being dropped. The I frames could be dropped when there is even slight congestion in the network. The above observations also apply to the P-type frames as shown in Fig. 7. Since the B frames are likely to be the least important frames compared to other frames, they are dropped the earliest in `MPEGAware`. But it is also possible that some of the B frames and P frames have higher priority than I frames, because there are other factors that could affect the dropping precedence/priority of a frame, e.g., ML, RoI as discussed in Section III-B. As a result, when CL is configured to 0.4, a small number of B frames that have high ML and are in the RoI pictures could still reach the receivers.

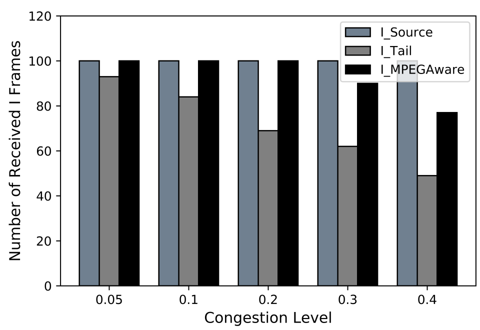

Fig. 6. Number of received I frames

The following Fig. 9, Fig. 10 show the number of received frames with different priorities when congestion level is low and high respectively. When the network congestion level is low, `MPEGAware` only drops the packets with the least priority, which has the priority value of 72, while retaining all other frames with higher priorities. In `Tail`, on the contrary, congested routers have no visibility to the priorities of the frames carried in the packet payloads, the high priority packets are not able to reach the receivers, thus resulting poor QoE (Quality of Experience) for the receivers. Fig. 10 again verifies that the packet dropping order in `MPEGAware` always follows the least priority first highest priority last principle. When network congestion is high, the impact on the receivers' QoE is minimized. In contrast, `Tail` cannot guarantee the receivers' QoE, as it drops packets with the highest priorities indiscriminately.

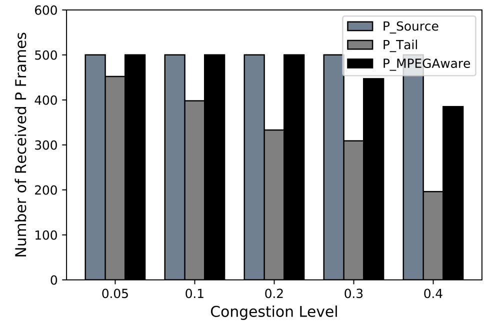

Fig. 7. Number of received P frame

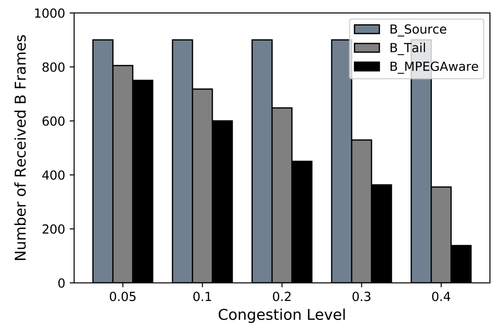

Fig. 8. Number of received B frame

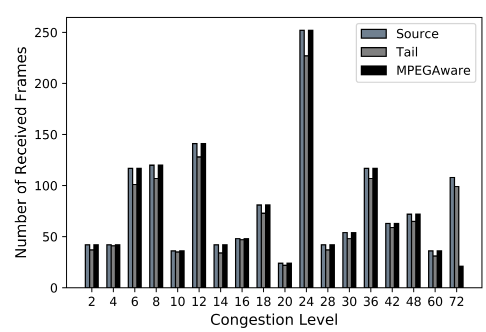

Fig. 9. Number of received frames with different priorities when congestion level is low

We also simulate the scenarios in which multiple frames with different priority levels could be encapsulated in one packet. `MPEGAware` operates by washing

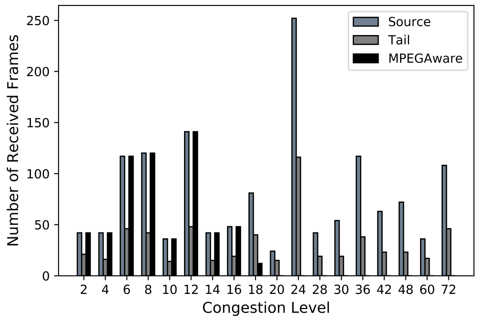

Fig. 10. Number of received frames with different priorities when congestion level is high

the packet from the tail to the head and discards the lowest priority frames first, while prioritizing the retention of higher priority frames. On the other hand, `Tail` drops the entire packet without taking into account the priority of the frames contained within. Similar results are observed by comparing `Tail` and `MPEGAware`. Although some packets are washed during transit with `MPEGAware`, they still reach the receivers with partial payload, thereby significantly improving the QoE of the receivers compared to `Tail`.

V. Conclusion

The major characteristics of a video frame contained in a packet's payload determine how important the packet is to the decoder in order to decode the video properly and improve viewer's QoE. The major characteristics information discussed in the paper includes frame type, being referenced or not, movement level and nature of the movement,inside of RoI or not. With those information being revealed through New IP metadata, IPv6 extension header, or IPv4 options to the network layer, a router calculates the priority of the packets from different video flows, drops the packets with lower priority when encountering congestion. A packet which contains multiple frames with different priority levels is subject to packet wash or partial packet dropping based on the priority. The performance evaluations verify that making MPEG video frames' characteristics aware to the routers would help routers prioritize the packet forwarding and dropping when network congestion happens, which in turn dramatically improves the end users' QoE.

References

[1] Cisco, "Cisco visual networking index: forecast and trends, 2017–2022," White Paper, 2018.

[2] A. Afanasyev, N. Tilley, P. Reiher, and L. Kleinrock, "Host-to-host congestion control for tcp," *IEEE Communications Surveys and Tutorials*, vol. 12, no. 3, pp. 304–342, 2010.

[3] R. Adams, "Active queue management: A survey," *IEEE Communications Surveys and Tutorials*, vol. 15, no. 3, pp. 1425–1476, 2013.

[4] R. Al-Saadi, G. Armitage, J. But, and P. Branch, "A survey of delay-based and hybrid tcp congestion control algorithms," *IEEE Communications Surveys Tutorials*, vol. 21, no. 4, pp. 3609–3638, 2019.

[5] D. L. Black, Z. Wang, M. A. Carlson, W. Weiss, E. B. Davies, and S. L. Blake, "An Architecture for Differentiated Services," RFC 2475, 1998. [Online]. Available: https://rfc-editor.org/rfc/rfc2475.txt

[6] D. L. Black and P. Jones, "Differentiated services (Diffserv) and real-time communication," RFC 7657, 2015. [Online]. Available: https://rfc-editor.org/rfc/rfc7657.txt

[7] R. Li, L. Dong, C. Westphal, and K. Makhijani, "Qualitative Communication for Emerging Network Applications with New IP: Invited Paper," in *IEEE MSN*, 2021.

[8] L. Dong and R. Li, "Improve Multiple-Camera Assisted Remote Driving by Qualitative Communication and New IP," in *IEEE/IFIP Network Operations and Management Symposium*, 2022.

[9] R. Li, K. Makhijani, H. Yousefi, C. Westphal, L. Dong, T. Wauters, and F. De Turck, "A framework for Qualitative Communications using Big Packet Protocol," in *Proceedings Of The 2019 ACM Sigcomm Workshop On Networking For Emerging Applications And Technologies*, 2019.

[10] R. Li, K. Makhijani, and L. Dong, "New IP: A data packet framework to evolve the Internet : Invited Paper," in *2020 IEEE 21st International Conference on High Performance Switching and Routing (HPSR)*, 2020.

[11] R. Li, A. Clemm, U. Chunduri, L. Dong, and K. Makhijani, "A new framework and protocol for future networking applications," *ACM Sigcomm Workshop on Networking for Emerging Applications and Technologies (NEAT 2018)*, May 2018.

[12] L. Dong and R. Li, "In-packet network coding for effective packet wash and packet enrichment," in *IEEE Globecom Workshops*, 2019.

[13] L. Dong, K. Makhijani, and R. Li, "Qualitative communication via network coding and New IP : invited paper," in *International Conference on High Performance Switching and Routing (HPSR)*, 2020.

[14] L. Dong and A. Clemm, "High-precision end-to-end latency guarantees using packet wash," in *2021 IFIP/IEEE International Symposium on Integrated Network Management (IM)*, 2021, pp. 259–267.

[15] M. Mu, R. Gostner, A. Mauthe, G. Tyson, and F. Garcia, "Visibility of individual packet loss on h.264 encoded video stream – a user study on the impact of packet loss on perceived video quality," 01 2009.

[16] D. A. Borman, D. S. E. Deering, and B. Hinden, "IPv6 Jumbograms," RFC 2675, Aug. 1999. [Online]. Available: https://www.rfc-editor.org/info/rfc2675