PRAM: Penalized Resource Allocation Method for Video Services

F. Nikoonezhad and M. Ghanbari*, Life *Fellow, IEEE*ORCID: 0000-0002-5482-8378

***Abstract*—The human visual system response to picture quality degradation due to packet loss is very different from the responses of objective quality measures. While video quality due to packet loss may be impaired by at most for one Group of Pictures (GOP), its subjective quality degradation may last for several GOPs. This has a great impact on resource allocation strategies, which normally make decisions on instantaneous conditions of multiplexing buffer. This is because, when the perceptual impact of degraded video quality is much longer than its objective degradation period, any assigned resources to the degraded flow is wasted. This paper, through both simulations and analysis shows that, during resource allocation, if the quality of a video stream is significantly degraded, it is better to penalize this degraded flow from getting its full bandwidth share and instead assign the remaining share to other flows preventing them from undergoing quality degradation.**

***Index Terms*—Multiplexing, Resource allocation, Routing, Video quality measurement**

I. Introduction

WITH the advent of multimedia services over packet networks, such as video over internet, guaranteeing the desired quality of service for each media sharing the network has always been a problem. This is particularly true for video services, where apart from their higher sensitivity to channel constraints, they consume a larger portion of the internet traffic. More importantly, with the current situation of almost worldwide lockdown due to Covid-19, today use of video services can even be greater than the 82% of total traffic Cisco had predicted for the year 2022 [1].

At the early stage of video over the internet, the Internet Engineering Task Force (IETF) proposed two different architectures for maintaining satisfactory quality of service (QoS) [2]. They introduced Integrated services (IntServ) implemented on the edge nodes, to provide resource reservation and admission control per flow to guarantee the end-to-end delay. However, this approach is impractical for large volume of flows and does not scale well [3]. To solve this problem, IETF in 1997 introduced DiffServ, where services are differentiated into two groups, and the quality of service (QoS) for one group is guaranteed. Such a scheme which is implemented in the core network, does not guarantee QoS of individual services. This issue becomes harder for video services, due to their larger numbers and higher bandwidths than other services. Perhaps, the problem can be eased, when layered/scalable coded video is used, since the base-layer of these codecs comprise a small fraction of the whole video bit-rate. Moreover, normally base-layer has a nearly constant rate, and provision of guaranteed bandwidth for sum of base-layers becomes easier [4]. Sum of enhancement layers comprise the other group. For this and a variety of reasons, these days almost all standard video codecs are made scalable, and even older non-scalable codecs, such as H.261 can be easily converted to a layered/scalable codec [5]. However, scalable codecs due to their higher overhead over single layer codecs, are less attractive, and instead today multi-rate adaptive http streaming is preferred [22]

If resource allocation is carried out on routers in the network, where the number of flows per router is limited, then there is no need for DiffServ, as QoS within the router range can be easily maintained. In the past few decades, numerous resource allocation algorithms have been devised for distributing the available resources of the outgoing channels of the routers among the flows. These resource allocation algorithms are aiming to be *fair* to the flows, where for real-time services, fairness can be on short-term basis, or they may look at long-term fairness for better overall QoS [7]. Due to the heterogeneity of the Internet, it appears rate allocation with regard to end-to-end QoS gives the best performance. This is easily carried out in a centralized quality control system, such as Software Defined Network (SDN) [8]. For instance, Tiwari *et al* have shown the video quality of all users can be improved by centralizing the bit rate allocation algorithm [9]. This of course requires a central controller to supervise resource allocation and making sure the individual flows can get a fair share.

• Corresponding author
Authors are with the School of Electrical and Computer Engineering, College of Engineering, University of Tehran, Tehran, Iran 1439957131. (e-mails: fatemeh.nikoonezhad@gmail.com, ghan@ut.ac.ir.)

M. Ghanbari is also an Emeritus Professor in the School of Computer Science and Electronic Engineering, University of Essex, Colchester, UK, CO4 3SQ. (e-mail: ghan@essex.ac.uk.)

In non-centralized resource allocation, overlay multicast grouping is a good method for improving video quality. Boudko *et al* examine how available bandwidth can be allocated in an overlay system with multiple sources and multiple paths [10]. Zhu *et al* propose a bit rate allocation scheme in an ad-hoc network to improve the received video quality based on simple models of rate-distortion and rate-congestion [11]. This paper uses the gradient descent method to determine the bitrate sent in each frame on different paths. In a similar work [12], a bit rate assignment algorithm in a sub-gradient ad-hoc network was proposed that minimizes the overall distortions of all the video flows. The proposed method only requires link price update, and based on local observations, this method can reach optimal bitrate.

Some works have considered resource allocation based on subjective video quality, the so-called user Quality of Experience (QoE) [13, 14, 15]. For instance, in an overlay network, an optimal rate allocation algorithm based on the optimization of the overall video quality perceived within the multicast group has been proposed [16]. In this solution, each overlay node of a given overlay tree allocates the rates for its children in a way that the overall video quality perceived in the multicast group is optimized. At work [17], a cross-layer optimization framework has been introduced where the total QoE weight (the weighting parameters are selected based on the importance of each video source) of the video stream in a wireless environment is improved. Bideh *et al* propose a mesh-based peer-to-peer (p2p) streaming system of live video, where a packet scheduler has a major role in the overall quality of receiving video [18]. In the paper, the chunk-based scheduling scheme has two parts. Initially, the recipients declare their level of participation in each frame of video .These declared video frames are considered as a higher priority for video frames. In the second stage, the scheduler tries to request video frames within the highest priority of the peers that can deliver them in a shorter time. Yang *et al* assign bit rate allocation to users in conditions where the existing bandwidth is fluctuating [19]. It also considers that minimizing objective distortion not necessarily improves subjective quality. A utility-based pricing method is designed to optimize the subjective quality of every user's video. Experimental results show that subjective outcomes are increasingly improving with increasing bandwidth rates or bandwidth fluctuation in contrast to rate allocation solutions. Chakareski and Frossard specifically focus on a specific example of scheduling multiple video sequences in a wireless LAN scenario [20]. Each sender individually allocates a portion of the available bandwidth to the appropriate video stream according to network constraints and to maximize the quality of all video streams. A comprehensive review of resource allocation management based on QoE is given in [21].

Although in the above resource allocation algorithms both cases of QoS and QoE optimizations are considered, what is missing from these algorithms, is ignoring the short and long term human visual system response to the QoS and QoE variations. In fact, the existing methods no matter how best allocate resources to a source in a router, when it comes to serve the video flow at its given rate, it never considers what happened in the previous round of serving the sources. The fact is, if the video quality is temporarily degraded, the impact of quality degradation due to short term memory and recency effect can stay for a much longer time [24, 25]. In other words, when the video quality after the degradation is improved, observers are reluctant to accept the video quality is improved, and still regard it as lower quality. This in terms of resource allocation means, if due to lack of resources, video quality is temporarily degraded, then increasing its bandwidth afterwards, does not help the subjective quality and the allocated resources to such video flows are wasted. In this case, it is better to give a fraction of the required bits to such source and distribute the remaining parts among the other sources, preventing them from quality degradation. This is in fact the main contribution of this paper that we try to analyze and show through simulations that such *penalized* resource allocation can lead to a better overall video quality for all sources sharing a channel than the conventionally used *fair* resource allocation.

The rest of the paper is organized in the following order. Section II looks at the response of the human visual system (HVS) to video quality changes and its implications to channel rate allocation. Section III analyses the HVS response to video quality degradation due to packet loss. The proposed method of resource allocation algorithm is described in Section IV and experimental results are given in Section V. Section VI concludes the paper.

II. RESPONSE OF THE HUMAN VISUAL SYSTEM (HVS) TO VIDEO QUALITY VARIATION

Before dealing with the impact of Quality of Experience (QoE) on channel rate allocation, let us see how QoE in video is measured. Unlike assessing image quality through Double-Stimulus Continuous Quality-Scale (DSCQS), where both impaired and reference pictures are compared by viewers for a short period of 8-10 seconds, such a short period is not very meaningful in video quality assessment. Video becomes meaningful when it is viewed for a much longer time. However, giving a single score to a video clip, like DSCQS of images, suffers from a phenomenon called recency effect [24, 25]. This effectively means, subjects are biased towards the quality of a portion of the video they see just before they vote. For this reason, ITU under recommendation R-500 has recommended Single-Stimulus Continuous Quality Evaluation (SSCQE) method for assessing video quality [26, 27]. In this method subjects continuously record their votes through an electronic recording handset connected to a computer [28]. By sampling the handset output at regular intervals (normally 0.5 second), the assessor's opinion is recorded by the computer. It is worth noting that in the early era of digital TV (1990-2000), numerous works had been devoted to the human reactions to video quality variations, but their implications on video networking to some extents have been over-looked. This is one of the reasons that most referenced papers in this article, although look old, but they are fundamental references to our work.

Fig. 1 shows a trace of 2 minutes MPEG-2 coded video quality of 720×576 pixels resolution at 7.5 Mbps for the first 45 seconds, 1 Mbps for the next 45 seconds, and 4 Mbps for the

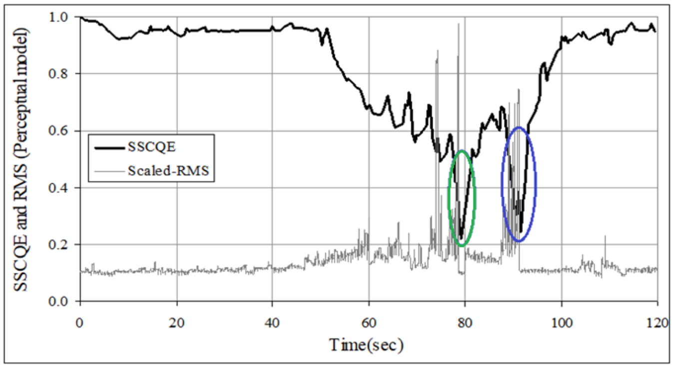


Fig. 1 An average subjective video quality of a 120 Sec video coded at 7.5, 1 and 4 Mbps measured through SSCQE [29]

last 30 seconds, measured through SSCQE. Tan *et al*, have used this trace to devise an objective video quality meter [29, 30]. In the figure the darker trace is the average of the SSCQE responses of 15 non-expert human observers. The figure also includes a lighter trace which is the corresponding per-frame distortion (root mean squared (RMS) of coding distortion), weighted to account for spatial masking. The RMS data have been scaled down to fit in the graph.

Comparing the subjective quality of SSCQE and the objective RMS quality can have some implications in resource allocation. These are:

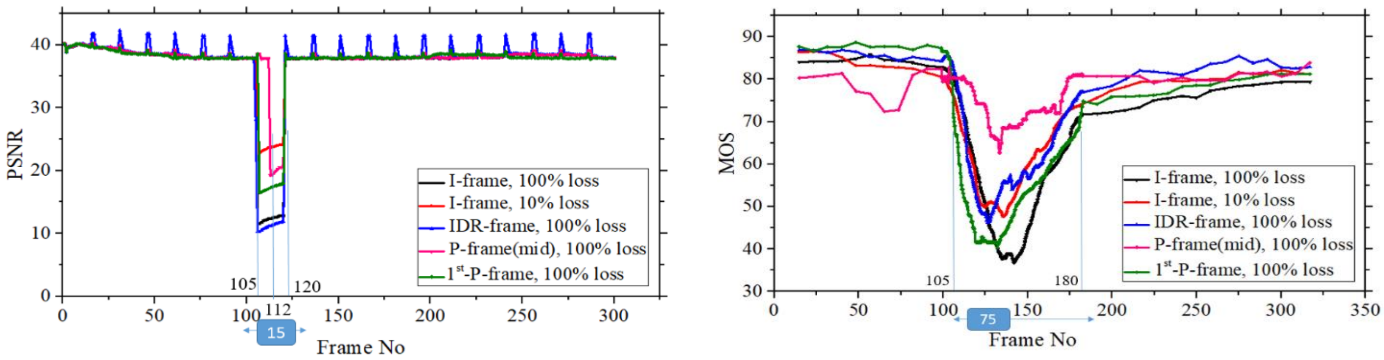
1. While the RMS data fluctuates rapidly from frame to frame, the SSCQE rating is much smoother. This is not just due to the subsampling of the SSCQE apparatus, but it is due to HVS smoothing effect [29, 30].

   In particular, during 1 Mbps interval, very high fluctuations of RMS are made much smoother and are mainly perceived as low-quality video. This indicates that viewers are biased towards one of the bad or good qualities, and the other one is ignored. In fact, humans are better able to remember unpleasant experiences than pleasant moments [31]. The implication of this phenomenon is that, when subjects view low quality video, such fluctuations just prolong the duration of poorly perceived video, which can be several seconds.
2. While in switching from 7.5 Mbps to 1 Mbps at time 45 Sec, the RMS distortion is not significant, but the subjective quality is very poor. This simply means that the viewers respond quickly to degradation in picture quality. The implication of this phenomenon in rate allocation is that sufficient rate should be assigned to video flows not to starve them to get into poor quality.
3. In switching from 1 Mbps to 4 Mbps at time 90 sec, while abrupt improvement in objective quality (RMS) is noticeable, SSCQE subjective quality improvement is gradual. This simply indicates, after very poor quality, if the channel rate is even increased by 4 times, it does not immediately improve subjective quality, and such an allocated rate is wasted.
4. Combining the outcomes of items 1, 2 and 3 means: In channel rate allocation, after a long period of poor quality video, it is better to increase channel rate gradually, and the extra bits should be distributed among the other flows preventing them from going into poor quality state. This is perhaps the most important message of subjective quality variation of video in channel rate allocation, which is the main focus of this paper.
5. Fig. 1 can also be regarded as a kind of bit rate switching, which today under the subject of http adaptive multi-rate HAS/DASH [22][23] stream switching is becoming state of the art video transmission. The implication of the above findings in HAS/DASH is: in switching to lower or higher rate streams, they should be switched to their immediately below or above rates. This will prevent the former from getting into severe degradation and the latter not wasting the channel capacity.

In the proposed resource allocation algorithm to be followed, we will explore the above findings and show how preventing bandwidth wastage can improve the overall quality of video flows sharing the channel.

III. The Impact of Packet Loss on QoE Variation

In the previous section, the relation between subjective quality and bit rate in multi-rate switching was studied. However, most video services are coded at a single bit stream and quality variation at the receiver is mainly due to its packet loss. In the followings the impact of packet loss in video quality is investigated and the goal is to see if the observations made in Section II are still valid for packet loss! To investigate this issue, 300 frames of the Park_joy test video sequence at resolution of 1280×720 pixels at 30 frames/sec with a GOP size of the 15 frames and structure of IPPPP were coded at a constant bit rate (CBR) of 6.5 Mbps. Each frame was divided into 74 slices. No B-frame was included in the GOP, as its loss has no impact on subjective quality [6]. As shown in Fig. 2a five different frame-type losses were introduced to the bit stream, starting at frame number 105 (beginning of a GOP). Lost packets were: (i) all packets of an I-frame at volume of almost 40 KBytes, (ii) all packets of an IDR frame with volume of almost 50 KBytes, (iii) 7 slices of an I-frame with a volume of 400 bytes (almost 10% of I-frame), (iv) all packets of a P-frame



(a) (b)

Fig. 2 (a) Objective video quality (PSNR) and (b) Subjective video quality (SSCQE) due to loss of various picture types

almost in the middle of the GOP) with 103 bytes and (v) all packets of the first P-frame in the GOP, were lost.

While Fig. 2a shows the PSNR quality of lossy video flows, their subjective quality through SSCQE averaged over 15 observers are presented in Fig. 2b. Comparing the PSNR objective quality and SSCQE subjective quality of Figs 2a and 2b respectively, the following observations can be made that can be exploited in resource allocation:

1. While losses are confined in only one video frame and at most disturbs the objective quality over only one GOP (15 frames), their impacts on subjective quality at least lasts for more than 77 frames, equal to five GOPs (frames 105-182). This is similar to the observations made under items 1 and 2 of Section II, that the subjective degradation spreads over several GOPs. The implication of this on chunk-based video streaming, like HAS/DASH is that, if long size chunks are used, their sizes should be in the order of 5 GOPs or longer.
2. While PSNR of seven lost slices in an I-frame, is better than loss of full I-frame, their subjective impacts are equally bad. This implies a severe loss causes deep and long duration of the bad subjective quality.
3. While the loss of full P-frame at the start of the GOP is subjectively annoying, loss of a P-frame in the middle of the GOP, in particular with low texture and motion is tolerable and can be ignored. This is called forgiveness effect [32]. This implies two points: first, the side effect of P-frame losses, depend on the position of P-frame in the GOP, and second, as long as loss of P-frame does not create a deep distortion, it will be treated as normal.

IV. The Proposed Penalized Resource Allocation Method (PRAM)

Studies in Sections II and III indicate that, as long as video distortion is moderate, the subjective and objective measures follow each other well. This means that, under no-severe loss, channel rate allocation in a router can be carried out in usual way. For instance, in Fig 3, we have chosen two video flows of almost equal characteristics, thus fair rate allocation is to give each, half of the channel rate. A kind of fair rate allocation is a Round Robin algorithm [33, 34]. The most popular version of RR is the well-known Deficit Round Robin (DRR) algorithm, used in the NS2 simulator [35]. In this algorithm, after knowing the allocated channel rate for each channel, flows are served at that rate one after the other. However, as Section III had shown, when severe distortion is experienced in a flow, allocation of the pre-assigned quota to that flow is just waste of resources provided the subjective video quality stays at low levels. Unfortunately, this period can be as large as several GOPs.

Thus, the severe lossy condition of a video flow should influence the channel rate distribution among the lossy and non-lossy flows. If there are no severe losses, or moderate losses, then the available channel rate can be equally or proportionally allocated to the flows, through say DRR algorithm. In this paper, this method is named as *fair* method. But, when any flow experiences a severe loss that may create a long period of bad quality, then that flow may be penalized from getting its fair share. In this case, the lossy flow is given a much lower rate and its remaining cake of channel rate is distributed among the other flows as extra resources, preventing them from getting into the bad quality mode. In this paper, resource allocation based on this condition is named *penalized* resource allocation method (PRAM). When the lossy condition is over, the resource allocation algorithm gets back to its normal *Fair* resource allocation mode.

In practice PRAM can be implemented in a variety of ways. For instance, as the objective quality in Fig. 2a indicates, when a severe loss occurs at the beginning of a GOP, be it an I-frame, or some slices of I-frame or from P-frames, the objective as well as the subjective video quality for one GOP (15 frames) are very poor. In this case, rather than displaying a GOP of poor video quality, this GOP can be ignored at the receiver and the video is frozen for one GOP. Huynh-Thu and Ghanbari have shown video freezing at this rate only marginally reduces the perceptual video quality [36]. After the first GOP, despite the objective quality in Fig. 2a is restored back to its original quality, but as Fig. 2b shows, the subjective video quality is gradually improving, and assignment of full channel rate does not improve it to the corresponding level of objective quality. One suggestion for practical implementation would be; since the duration of video quality recovery after the first GOP can take several GOPs (e.g., 5-10 GOPs), and as Fig 2b shows, it gradually improves, then the assigned channel rate in this period should be gradually increased. For instance, after the first GOP, 1-2 video frames, starting from an I-frame is sent, in the following GOP, this rate can be increased to 4 frames, and in the following one to 6 frames, till all the 15 frames are served. Of course, if in these periods the allocated rates to other flows are adequate, then for the Penalized flow in each GOP more fames can be sent. After this period, resource allocation is reset to its initial Fair mode.

Please note that, identifying packet frame type in a router is quite feasible [37]. In fact, recently we have shown, in a router, it is possible to identify I-frames with 100% precision and P-frames with better than 97% accuracy, as well as position of each frame in the GOP with such precisions [38].

It is assumed each router has a multiplexing buffer, where the incoming video flows are accepted to the buffer to be transmitted in due course. When the buffer overflows, or reached a certain threshold, the incoming video traffic of any desired flow into the buffer can be ceased. Also, assuming the frame-type packets and GOP structure of video flows of the multiplexing buffer are known, then switching from Fair resource allocation to Penalized one and vice versa in a router is easy. One way is explained in the following:

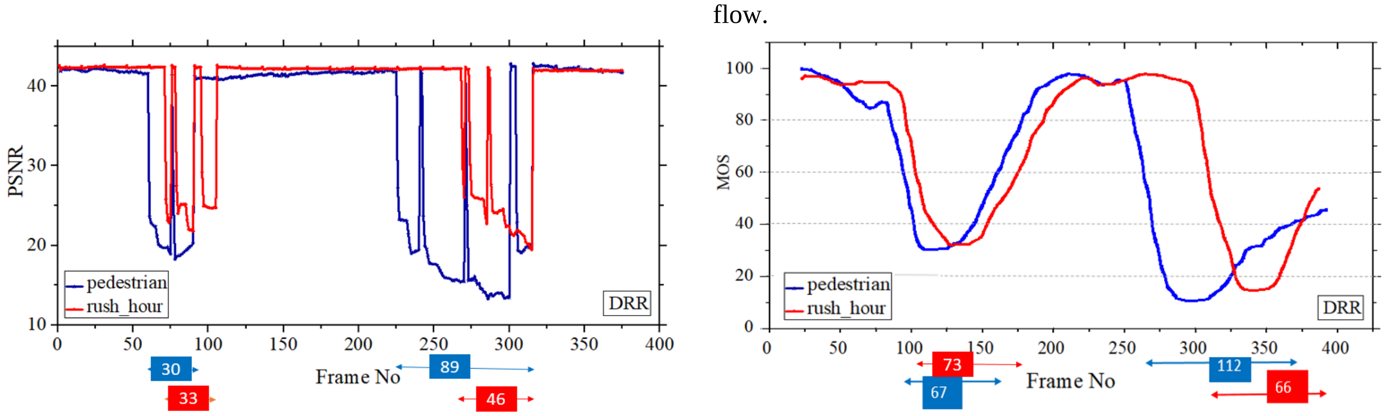
At the start, the channel rate is fairly distributed among all video flows. When packets from a flow are not accepted to the buffer, as if that packet is deemed to create a severe video quality degradation (e.g., packets are lost in I-frame or early P-frames of the GOP), then the remaining packets from that flow up to the end of the GOP (named this GOP as GOP0) are seized and the Penalized resource allocation mode is activated, otherwise the Fair resource allocation is continued. In the lossy video flow, its transmission frames at the following GOPs can be, for example:

GOP1: Send/accept frame I
GOP2: Send/accept frames I, P1, P2
GOP3: Send/accept frames I, P1, P2, P3, P4,
GOP4: Send/accept frames I, P1, P2, P3, P4, P5, P6
GOP5: Send/accept frames I, P1, P2, P3, P4, P5, P6, P7, P8
GOP6: Send/accept frames I, P1, P2, P3, P4, P5, P6, P7, P8, P9, P10
GOP7: Send/accept frames I, P1, P2, P3, P4, P5, P6, P7, P8, P9, P10, P11, P12
GOP8: Send/accept frames I, P1, P2, P3, P4, P5, P6, P7, P8, P9, P10, P11, P12, P13, P14
The following GOPs, like GOP 8: Send/accept all frames in the GOP.

In the above resource allocation algorithm, in addition to GOP0, 8 additional GOPs, create a delay of 9 GOPs, equal to 4.5 Sec, before the lossy video flow recovers back to its normal condition. This is well within the range of 2-10 Sec delay imposed in http adaptive streaming (HAS) and MPEG DASH [22], [23], where quality of each video chunk can change within 2-10 Sec intervals. If more or less delays are desired, then the number of extra frames in each transmitted GOP period can be altered.

Also, if the assigned rate for the non-lossy flow is more than its required rate, then more frames from the lossy flow at any GOP can be sent. Moreover, if the GOP structure includes any B-frames, the criterion for deciding between Fair and PRAM resource allocation can change to: if an error occurs in any I-, or P-frame, use PRAM resource allocation, otherwise carry on with Fair resource allocation.

After the lossy period is over, the resource allocation algorithm reverts back to the Fair mode. Note, if during the Penalized resource allocation, any other flow, as well as the Penalized flow experiences a severe packet loss, the above Penalized resource allocation algorithm is also executed in that flow.



(a) (b)

Fig. 3 (a) Objective (PSNR) and (b) subjective (SSCQE)quality in Fair (DRR) rate allocation mode

It is worth noting that, if video flows have different resolutions, frame rates, etc. the Fair resource allocation should assign them their proper fair rates. Also, fair resource allocation can be based on QoE, meaning that to each source, resources are assigned such that their end-to-end QoE is optimized. However, when such QoE-based Fair resource allocation is changed to Penalized mode, the assigned rate to a lossy source under the fair allocation is now changed from a minimum rate and gradually increased to its full rate.

Algorithm1 shows an algorithmic method for switching between DRR and PRAM resource allocations. Also, a constant array(PIVOT_POINT) is defined, which contains the threshold for severe loss. Each element of this array is a frame number, if the frame number of the lost packet of the stream is less than the element value, the penalized resource allocation will be applied. The Algorithm has two modes: 1- the stream state remains the same (Fair or Penalized) 2- the stream state should change.

Algorithm 1:

Function ResourceAllocation(*PacketNumber, PacketType, PacketSize, PacketLoss, State, CurrentGOPAllocation, Bandwidth, GOPSize, StreamNumber***);**

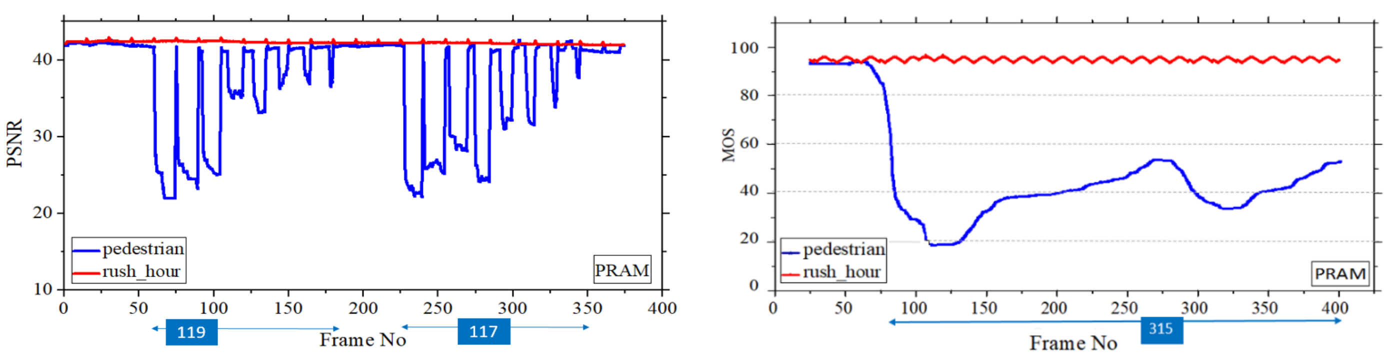
Input:the packets number in GOP, the packet type, the packet size, a value that indicates a packet is lost or not (based on sequence number), a state array which specifies each stream resource allocation state, an array that specifies how many frames should be kept in the current GOP for each stream, the maximum bandwidth that each stream can use, the GOP size for each stream and the stream number.

Output:A TRUE/FALSE value that specifies the current packet should be kept or skipped.

```
if PacketType == 'I' then
    Bandwidth[StreamNumber] = DEFAULT_BANDWIDTH
    CurrentGOPAllocation[StreamNumber] += 2
    if CurrentGOPAllocation[StreamNumber]>= GOPSize[StreamNumber] then
        State[StreamNumber] = 'Fair'
    end
end
if PacketLoss == True then
    if (PacketNumber - 1) < (PIVOT_POINT[StreamNumber]) then
        State[StreamNumber] = 'Penalized'
        CurrentGOPAllocation[StreamNumber] = 0
        return False
    else
        if Bandwidth[StreamNumber] > 0 then
            Bandwidth[StreamNumber] -= PacketSize
            return True
        else
            return False
        end
    end
else
    if State[StreamNumber] == 'Fair' then
        if Bandwidth[StreamNumber] > 0 then
            Bandwidth[StreamNumber] -= PacketSize
            return True
        else
            return False
        end
    elseif State[StreamNumber] == 'Penalized' then
        if PacketNumber<CurrentGOPAllocation[StreamNumber] then
            Bandwidth[StreamNumber] -=PacketSize
            return True
        else
            return False
        end
    end
end
```

V. Simulation Results

To test if our strategy (e.g., PRAM=un-Fair instead of the Fair (DRR) resource allocation) can lead to a better video quality, the following simulations were carried out. First, two HD test video sequences of Pedestrian and Rush_hour, each with 375 frames long and resolution of 1920×1080 pixels were CBR coded with JM19.0 software of H.264 standard video codec [42]. The two video sequences were carefully selected to have almost equal complexity, such that in Fair (DRR) resource allocation, the available channel rate is equally divided between them. In the case of PRAM resource allocation, while the lossy flow gets a small fraction of channel rate, the remaining channel rate is allocated to the non-lossy video frames. Thus, after the severe loss, the lossy video flow gradually increases its share of channel rate, from as low as an I-frame rate and at the end to its full rate of half the channel rate.



(a) (b)

Fig. 4 (a) Objective (PSNR) and (b) subjective (SSCQE) quality in Penalized(PRAM) resource allocation mode

The non-lossy flow, on the contrary, starts with a high channel rate as large as twice its normal capacity, gradually reducing it to just its own original rate.

The GOP length was set to 15 frames with IPPPP structure (without B frame) and the target bit rate for each video was 8.3 Mbps. The codec was set to slice mode, where each video frame was partitioned into 74 slices, each slice was coded at 1024 bytes. The two sequences, each with 8.3Mbps and a bursty background traffic of 1Mbps were passed through a link with a bandwidth of 17.6 Mbps and the buffer size was set to 500 slices (packets).

For simulation purposes, each slice was packetized into an RTP packet and the error concealment of the decoder was enabled. The simulator was of type ns2 EvalVid to simulate multiplexed video at the router [43]. As mentioned above, the resource allocation algorithm starts with DRR fair bandwidth algorithm between the streams. When severe loss is experienced, the resource allocation algorithm is switched to Penalized mode (PRAM). To understand how resource allocation can affect the video quality, we have focused only on one router activity, to trace and analyze each stream carefully.

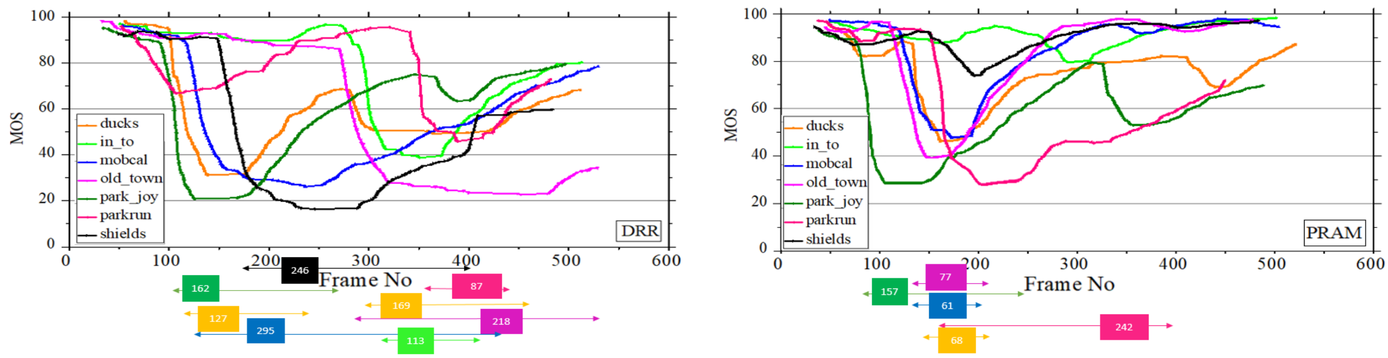
On subjective test, using SSCQE, 20 graduate students, on the screen of their laptops watch a series of video sequences and through mouse create a continuous trace of video quality. They were instructed to trace the quality such that within five quality intervals, the trace indicates: excellent (80-100), good (60-79), fair (40-59), poor (20-39) and bad (0-19). Please note that, currently most researchers instead of running subjective tests, they use the Netflix video quality meter (VMAF) to get quick results [41]. However, we could not use VMAF for this test, as it is not clear if the long duration of low subjective quality of lossy frame, shown in Fig 2b is included in VMAF. Also, a buffer size of 500 slices with a threshold of 300 slices is assumed in the router to manage the buffer [39, 40]. This threshold makes sure, there is still enough space in the buffer, such that packets of the other flow are not lost. The resource allocation starts with the DRR mode until the first packet loss from one of the flows occurs. In the DRR resource allocation mode as shown in Fig. 3a, the Pedestrian sequence starts to drop packets from its 61st frame, but at the next GOP the I-frame clears the error accumulation and PSNR peaks again. At frame 77, within a duration of 30 frames, the same error pattern occurs. Looking at the subjective curve of this resource allocation method in Fig 3b, it is shown that the degraded duration in Pedestrian sequence is much longer than just a GOP (in fact it is 8 GOPs). This means that the assigned resource during this 8 GOPs is wasted and does not immediately improve the lossy flow quality. Moreover, it has not been exploited by the Rush_hour sequence, since the Rush_hour sequence also drops packets starting from frame 72 and then on frames 79, 96 etc. Its subjective quality in Fig 3b, shows that shortly after the Pedestrian sequence, its quality is subjectively degraded. Both sequences from frame 106 to frame 225, have no losses and stay at high quality. The second batch of frame drops starts at frames 226 to 315, where again they have on-off frame losses for durations of 89 and 46 frames for Pedestrian and Rush-hour sequences, respectively. While quality degradation on the objective graph of Fig. 3a is fluctuating, that of subjective quality in Fig. 3b is smoother, as expected, due to HVS behavior. However, both sequences have almost a similar video quality variation, even from bad quality (under 20%) to excellent quality (above 80%), with almost similar durations. Hence it shows that the DRR is a truly fair resource allocation algorithm, causing bad and good quality equally.

In the Penalized resource allocation mode (PRAM), the situation is very different. First, as shown in Figs 4a, and 4b, both sequences start with DRR resource allocation, each getting their 100% channel rate share. When at frame 61, Pedestrian sequence losses packets, in the next GOP, it sends only an I-frame (see previous example). If we assume an I-frame rate is twice the P-frame rate, then in a GOP of one I and 14 P, I-frame rate is almost 12.5% of the initial channel rate, which is now allocated to Pedestrian, and the remaining rate is added to the share of Rush_hour which now goes up from its initial value of 100% to 187.8%. In the next GOP, Pedestrian sends another 2 P-frames, and now its rate becomes 25%, but the rate of Rush_hour now becomes 175%, and so on. As we see, unlike the DRR method, no channel rate is wasted, and the unused channel rate is given to the other video flow to be served and release its packets from the buffer.

This strategy has two implications, as follows: (i) The non-lossy, Rush_hour sequence with such a high-speed serving rate, sends almost entirely its whole packets. Hence it has no packet loss. Even at the end of the sequence, which in the DRR mode, there was a second batch of packet loss, this does not occur for Rush_hour. (ii) The second implication is on the lossy Pedestrian sequence itself. Since its channel rate is reduced, then its packets cannot be served on time and they accumulate in the buffer. Hence, even due to small burstiness in the background traffic, it will immediately start losing packets again, thus its serving rate is reverted back to 12.5% mode and the extra channel rate is again given back to Rush_hour. What happens, is this fact that, Pedestrian never recovers and its lossy duration gets longer, as seen from Figs 4a, and 4b. These extra channel rates are of course given to the Rush_hour sequence, making sure its quality stays high.

Streaming video with two very different qualities by PRAM, is of course unrealistic and this is because multiplexing two video sources, where in the event of loss, the other channel nearly gets the whole channel rate is not logical. If more video sources were multiplexed, then the extra channel rate will be divided among more flows, and their increased channel rates become smaller.

In the following experiments, seven video sources are used. All sequences were high resolution full HD of 1920×1080 pixels resolution with duration of 500 frames and 50 fps (now GOP duration becomes 300ms, making recovery time faster), 4:2:0 format with a GOP size of 15 frames and no B frames. The sequences were CBR coded at very high quality of 19.02



(a) (b)

Fig. 5 Subjective quality of 7 video sources with(a) Fair (DRR) and (b) Penalized (PRAM) resource allocation strategy

Mbps. To create packet loss, 1Mbps bursty background traffic was added to the multiplexer, and the total link capacity was set to 134.14 Mbps and buffer size of 50,000 packets. The experiments were carried out separately for both DRR and PRAM modes, and the video quality of all sequences under both modes was subjectively rated (Absolute Category Rating (ACR), using sliding apparatus, e.g., SSCQE). Due to space limitation, objective quality graphs are not shown and Figs 5a and 5b show the subjective quality of both DRR and PRAM modes respectively within 0-100 scale.

Although these figures are too crowded for stream-to-stream quality comparisons between the two quality graphs, but it is easy to see that durations of the times that video flows are below 60% (good quality) under DRR are much longer than those of PRAM. This means more video sources have poorer quality under DRR than under PRAM. Two sequences in Fig 5b (in_to_tree in light green and Shields_ter in black) never go below 60% (good quality). The reason for the better quality of PRAM in Fig 5b, is due to using the leftover channel rate in lossy frame, which is not exploited in DRR of Fig 5a. In DRR, like Fig 3b, the quality is fairly distributed among the video flows, but in PRAM, although all flows start with equal share of channel rate (100% of their own share), but when the first flow, Park_joy (Dark green) faces severe loss, in the next GOP, its share is reduced to 12.5%. The remaining 87.5% is now divided among the other 6 video flows, each getting 114.5% of their original share. In the next GOP, these numbers become respectively 25% and 112.5%, and so on. These given extra channel rates, of course, unlike Fig 4b is not that much to make them free of loss (but still two of them are above 60%, good quality), but help them to improve their quality and not to come under lower quality scales.

Perhaps, a better way to compare these two figures is to calculate the percentages of the time durations each stream stays in certain quality range. By magnifying Figs 5a and 5b, for each sequence, these intervals can be visually measured. Table 1 shows the percentages of durations each stream remains under the DRR and PRAM resource allocation algorithms in a 5-grade quality scale. First, in all three columns of Excellent (>80%), above good (>60%) and above fair (>40%), except Parkrun_ter sequence (Park_joy has equal performance), all the other 5 sequences under PRAM are transmitted at better quality than under DRR. Second, which is the most important quality comparison, is the quality of the streams at the 40% border (the fair/poor quality border which is almost acceptable). The fair quality border line (<40%) of the Table shows that under PRAM mode, 4 sequences are always above the fair quality. Two of them only for 4.4% and 15% of the time go below fair quality and Pakrun_ter has poor quality for 22% of the time. For the DRR resource allocation mode, only the Parkrun_ter sequence never goes below 40% (fair quality), and quality of other 6 sequences for 8%-44% of the time is not acceptable.

Please note that in Table 1, Parkrun_ter and in some quality rows Park_joy gain more in quality under DRR than PRAM. This depends on video characteristics as well as which video stream during congestion first experiences packet losses. Also, please note that, since allocated resources under PRAM are

TABLE I
PERCENTAGES OF DURATIONS OF VIDEO QUALITY IN A 5-GRADE QUALITY SCALE FOR BOTH FAIR (DRR) AND PENALIZED (PRAM) MODE RESOURCE ALLOCATION ALGORITHMS.

Sequence name	>80%		>60%		>40%		<40%		<20%	
	PRAM	DRR	PRAM	DRR	PRAM	DRR	PRAM	DRR	PRAM	DRR
shields_ter	91	30	100	32	100	55.8	0	44.2	0	19.6
mobcal_ter	76	24.4	87	40	100	65	0	35	0	0
old_town_cross	74	54	85	56	95.6	60	4.4	40	0	0
park_joy	16	19	54	63	85	79	15	21	0	0
In_to_tree	98.31	59	100	77	100	92	0	8	0	0
ducks_take_off	50	20	86	39	100	87.61	0	12.39	0	0
Parkrun_ter	31.4	47	51	88	77.96	99.98	22.04	0	0	0

different from DRR, then, each source leaving the router may face different loss pattern under the two schemes. Although on average video flows under PRAM face less packet losses (having more channel rates), but since packets of video flows under the influence of bursty traffic randomly interact with each other, then there might be some sources under DRR to face less losses, as seen in Table 1.

VI. CONCLUSION

Perceptual impression of video quality under packet loss is significantly different from the simple mean-squared-error objective measure. While the loss of compressed video packets can impair the objective quality for almost one group of pictures (GOP), its perceptual impact can last for several GOPs. This can severely damage the performance of conventional resource allocation algorithms, which mainly look at the instantaneous frame rates of the streams. The paper has shown that, when video quality due to packet loss is damaged, full rate allocation of resources to such stream will be a waste of resources. It is better to penalize such streams from getting their full shares and instead allocate their remaining shares to other streams, preventing them from quality degradation.

The subjective implication of packet loss can also be extended to multi-rate http adaptive stream switching. It appears, in stream switching, the switched stream rates should be close to each other, and switching between streams of larger gaps in their rates should be avoided.

Please note that in the experiments simple DRR is chosen to show how prevention of bandwidth wastage can lead to quality improvement of video flows sharing a channel. If a sophisticated resource allocation algorithm was used instead of DRR, adding PRAM methodology to that algorithm would also improve its distribution ability. Finally, we have used a variety of video resolutions from SD of 720×576 pixels to full HD of 1920×1080 pixels, as well as different codecs of MPEG-2 and H.264/AVC, and the results show the HVS reaction to perceived video quality due to packet loss is independent of picture resolution and type of video codec. These can be validated as a future work.

REFERENCES

[1] Cisco. 2016. Cisco Visual Networking Index: Forecast and Methodology, 2016–2021.

[2] F. Zhang, J. Macnicol, MR. Pickering, MR. Frater, JF. Arnold, “Efficient Streaming Packet Video over Differentiated Services Networks.” *IEEE Transactions on Multimedia* vol 8, no 5, pp. 1005–9, 2006.

[3] R. Braden, D. Clark, and S. Shenker, “Integrated Services in the Internet Architecture: An Overview” Jun. 1994, RFC1633.

[4] S. Przylucki and D. Czerwinski, “Priority-Aware Packet Pre-Marking for DiffServ Architecture Based on H.264/SVC Video Stream Structure.” *Wireless Personal Communications* vol 96, no 4, pp. 5391–5408, 2017.

[5] M. Ghanbari and V. Seferidis, “Efficient H. 261-based two-layer video codecs for ATM networks”, *IEEE Transaction on Circuits and Systems for Video Technology* vol 5, no 2, pp. 171-175, 1995.

[6] Mohammad Ghanbari, “Video Quality Measurement” *US Patent* 7869517, 2011.

[7] H. SHI, R. V. Prasad, E. Onur and I. G. M. M. Niemegeers, "Fairness in Wireless Networks: Issues, Measures and Challenges," *IEEE Communications Surveys & Tutorials,* vol 16, no 1, pp. 5-24, 2014.

[8] F. Z. Yousaf, M. Bredel, S. Schaller, and F. Schneider, “NFV and SDN Key Technology Enablers for 5G Networks,” *IEEE J. on Selected Areas in Comms*, vol 35, pp. 2468 –2478, Nov 2017.

[9] M. Tiwari, T. Groves and P. Cosman, "Bit-Rate Allocation for Multiple Video Streams Using a Pricing-Based Mechanism," *IEEE Transactions on Image Processing*, vol 20, no 11, pp. 3219-3230, Nov. 2011.

[10] S. Boudko, W. Leister, C. Griwodz and P. Halvorsen, "Maximizing video quality for several unicast streams in a multipath overlay network," 2010 *IEEE 4th International Conference on Internet Multimedia Services Architecture and Application*, Bangalore, pp. 1-5, 2010

[11] X. Zhu, S. Han and B. Girod, "Congestion-aware rate allocation for multipath video streaming over ad hoc wireless networks," 2004 *International Conference on Image Processing, 2004. ICIP '04*., Singapore, vol 4, pp. 2547-2550, 2004

[12] X. Zhu and B. Girod, “Distributed rate allocation for multi-stream video transmission over ad hoc networks”, *Int. Conference on Processing*, ICIP 2005. vol 2, 2005

[13] L. Skorin-Kapov, M. Varela, T. Hoßfeld, and K.-T. Chen, “A Survey of Emerging Concepts and Challenges for QoE Management of Multimedia Services,” *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol 14, pp. 2 –28, 2018.

[14] S. Petrangeli, J. V. D. Hooft, T. Wauters, and F. D. Turck, “Quality of Experience-Centric Management of Adaptive Video Streaming Services: Status and Challenges,” *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol 14, pp. 2 –28, 2018.

[15] A. Martin, J. Egaña, J. Flórez, J. Montalbán, I. G. Olaizola, M. Quartulli, R. Viola, and M. Zorrilla, “Network Resource Allocation System for QoE-Aware Delivery of Media Services in 5G Networks,” *IEEE Transactions on Broadcasting*, vol 64, pp. 561–574, 2018.

[16] B. Akbari, HR. Rabiee, M. Ghanbari, ‘An optimal discrete rate allocation for overlay video multicasting’, *Computer Communications*, vol 31, no 25, PP. 551-562, 2008.

[17] P. Goudarzi, MR. Nezami Ranjbar, “Bandwidth Allocation for Video Transmission with Differentiated Quality of Experience over Wireless Networks” *Computers and Electrical Engineering* vol 37, no 1, pp.75–90, 2011.

[18] M.K. Bideh, B. Akbari, and A.G. Sheshjavani “Adaptive Content-and-Deadline Aware Chunk Scheduling in Mesh-Based P2P Video Streaming.” *Peer-to-Peer Networking and Applications* vol 9, no 2, pp. 436–48, 2016.

[19] M. Yang, T. Groves, N. Zheng and P. Cosman, "Iterative Pricing-Based Rate Allocation for Video Streams With Fluctuating Bandwidth Availability," *IEEE Transactions on Multimedia*, vol 16, no 7, pp. 1849-1862, 2014.

[20] J. Chakareski and P. Frossard, "Rate-distortion optimized distributed packet scheduling of multiple video streams over shared communication resources," *IEEE Transactions on Multimedi*a, vol 8, no 2, pp. 207-218, 2006.

[21] A.A. Barakabitze et al., "QoE Management of Multimedia Streaming Services in Future Networks: A Tutorial and Survey," *IEEE Communications Surveys & Tutorials*, vol 22, no 1, pp. 526-565, 2020.

[22] H Amirpour, E Çetinkaya, C Timmerer, M Ghanbari, “Fast Multi-Rate Encoding for Adaptive HTTP Streaming”, *2020 Data Compression Conference (DCC)*, Snowbird, UT, USA, pp. 358-358, 2020.

[23] T. Zhang et al, “Modeling and analyzing the influence of chunk size variation on bitrate adaptation in DASH”, *IEEE INFOCOM 2017 - IEEE Conference on Computer Communications*, Atlanta, GA, pp. 1-9, 2017.

[24] R. Aldridge, J. Davidoff, M. Ghanbari, D. Hands, and D.E. Pearson, “Measurement of Scene-Dependent Quality Variations in Digitally Coded Television Pictures.” *IEE Proceedings - Vision, Image, and Signal Processing* vol 142, no 3, pp.149-154, 1995.

[25] R. Aldridge, J. Davidoff, M. Ghanbari, D. Hands, and D.E. Pearson, “Recency Effect in the Subjective Assessment of Digitally-Coded Television Pictures.” *Fifth Int. Conference on Image Processing and Its Applications*, 4-6 July, Edinburgh, pp. 336–339, 1995.

[26] T. Alpert and J. Evain, "SSCQE (Single Stimulus Continuous Quality Evaluation): A New Subjective Assessment Method Introduced in ITU-R Recommendation 500-7: Presentation and Results," *Int. Workshop on HDTV '96*, Los Angeles, CA, USA, pp. 417-427, 1996.

[27] ITU-R, "Methodology for the Subjective Assessment of the Quality of Television Pictures", *ITU-R Recommendation BT*. 500-7, 1995.

[28] H. de Ridder, and R. Hamberg, "Continuous Assessment of Image Quality", *J. SMPTE*, February, pp. 123-128, 1997.

[29] K.T. Tan, "Objective Picture Quality Measurement for MPEG-2 Coded Video", PhD Thesis, Department of Electronic Systems Engineering, University of Essex, Colchester, United Kingdom, 2000.

[30] K.T. Tan, M. Ghanbari and D.E. Pearson, 'An objective measurement tool for MPEG video quality', *Signal Processing*, vol 70, pp. 279-294, 1998.

[31] D. Kahneman and A. Tversky, "Prospect Theory: An analysis of decisions under risk", *Econometrica*, vol 47, pp 263-291, 1979.

[32] V Seferidis, M Ghanbari, DE Pearson,'Forgiveness effect in subjective assessment of packet video', *Electronics Letters* vol 28, no 21, pp. 2013-2014, 1992.

[33] K. Maruta and H. Furukava, "Highly Efficient Multi Channel Packet Forwarding with Round Robin Intermittent Periodic Transmit for Multihop Wireless Backhaul Networks", *Sensors*, vol 17, no 11 , Nov 2017, doi: 10.3390/s17112609

[34] J. Prassana and N. Venkataraman, "Threshold Based Multi-Objective Memetic Optimized Round Robin Scheduling for Resource Efficient Load Balancing in Cloud", *Mobile Networks and Applications*, vol 24, no 4, August 2019, doi.org/10.1007/s11036-019-01259-x

[35] M. Shreedhar and G. Varghese, "Efficient fair queuing using deficit round-robin," *IEEE/ACM Transactions on Networking*, vol 4, no 3, pp. 375-385, June 1996.

[36] Q. Huynh-Thu and M. Ghanbari, "Impact of jitter and jerkiness on perceived video quality", *Proc. 2nd International Workshop on Video Processing and Quality Metrics (VPQM)*, 2006

[37] A. Huszák, and S. Imre, "Analysing GOP Structure and Packet Loss Effects on Error Propagation in MPEG-4 Video Streams." *Final Program and Abstract Book - 4th International Symposium on Communications, Control, and Signal Processing, ISCCSP 2010* (March), pp.:3–5, 2010.

[38] A. Moharrami, M. Ghasempour and M. Ghanbari, "An in-router identification scheme for selective discard of video packets" (2021) http://arxiv.org/abs/2104.13013

[39] Fang, Yechang and Kang Yen. 2010. "Buffer Management Algorithm Design and Implementation Based on Network Processors." *(IJCSIS) International Journal of Computer Science and Information Security*, vol 8, no 1, 2010.

[40] P.A.A. Assuncao and M. Ghanbari, 'Buffer analysis and control in CBR video transcoding', *IEEE Transactions on circuits and systems for video technology*, vol 10, no 1, pp 83-92, 2000.

[41] VMAF: Perceptual video quality assessment based on multi-method fusion, Netflix, Inc., 2017-07-14, retrieved 2017-07-15

[42] JM-19.0, H.264/AVC reference softwarehttp://iphome.hhi.de/suehring/

[43] Evalvid; https://networksimulator2.com/evalvid/evalvid-download/

Fatemeh Nikoonezhad graduated from Shahrood University of Technology with a BSc degree in Information Technology Engineering, and MSc in Information Technology Engineering from University of Tehran. Her research was on Video Quality Assessment on Networks. Currently, she is a Ph.D. student at University of Tehran.

Mohammad Ghanbari (IEEE M'78–SM'97–F'01–LF'14) is a Professor at the School of Electrical and Computer Engineering, University of Tehran, as well as an Emeritus Professor at the School of Computer Science and Electronic Engineering, University of Essex, U.K. He has authored or co-authored eight books and has registered for 13 international patents and authored over 770 technical papers on various aspects of video networking, many of which have had fundamental influences in this field. These include video/image compression, layered/scalable video coding, video over networks, video transcoding, motion estimation, and video quality metrics. He is internationally best known for pioneering work on layered video coding, for which he was honored to IEEE Fellowship in 2001 and was promoted to IEEE Life Fellow in 2014. His book Video Coding: An Introduction to Standard Codecs (IET Press, 1999) received the Rayleigh Prize as the best book of the year 2000 by IET. He has also received several prizes, such as Reeves Prize for best paper award in 1995 and 14th Khwarizmi international award for work on video networking in 2001. He was one of the founding Associate Editors to IEEE Trans on Multimedia from 1999-2002.