Measuring and Modeling the Impact of Buffering and Interrupts on Streaming Video Quality of Experience

Josh Allard, Andrew Roskuski, and Mark Claypool
claypool@cs.wpi.edu
Worcester Polytechnic Insitute
Worcester, MA, USA

ABSTRACT

Streaming video clients use playout buffers to smooth out variations in network bitrates, especially important for mobile devices. The playout buffer sizes trade-off: the time *delay* before a video starts playing with *interrupts* when a video stops playing in the middle. The best buffer size a client should choose for the video and network conditions depends upon the relative impacts of buffering delays and interrupts on the user quality of experience (QoE). We design user study experiments that isolate buffering delays and interrupts, allowing for direct, quantitative comparisons of the impact on QoE for buffering delays versus interrupts. In our user study, over thirty users watched and rated 17 videos with a broad range of content, encoded with varying amounts of buffering delays and interrupts. Analysis of the data reveals interrupts more costly to QoE than the corresponding amount of buffering by a factor of about 2 to 1. The data is used to construct an analytic model of QoE which incorporates the impacts of buffering delays and interrupts, a model that can be a tool for assessing and improving how streaming video clients pick buffer sizes to maximize user QoE.

CCS CONCEPTS

• **Human-centered computing** → *Empirical studies in HCI*; • **Information systems → Multimedia streaming**;

KEYWORDS

streaming video, QoE, playout buffer

ACM Reference Format:
Josh Allard, Andrew Roskuski, and Mark Claypool. 2020. Measuring and Modeling the Impact of Buffering and Interrupts on Streaming Video Quality of Experience. In *The 18th International Conference on Advances in Mobile Computing and Multimedia (MoMM '20), November 30-December 2, 2020, Chiang Mai, Thailand.* ACM, New York, NY, USA, 8 pages.
https://doi.org/10.1145/3428690.3429173

MoMM '20, November 30-December 2, 2020, Chiang Mai, Thailand

ACM ISBN 978-1-4503-8924-2/20/11…$15.00
https://doi.org/10.1145/3428690.3429173

1 INTRODUCTION

The increasing power of today's computers coupled with the rise in network capacities has fueled the growth in streaming video. Improvements to wireless networking and mobile phones have especially increased video consumption on mobile devices. Over 60 percent of U.S. survey respondents watched video online via a mobile device in 2018, more than any other platform [6]. World-wide, over 2.7 billion people are predicted to watch video on their mobile phones by 2023, up from from 2.2 billion in 2019 [21]. Cisco's annual Internet report predicts that by 2022, video will make up 82% of all Internet traffic and 79% of all mobile traffic [4].

Despite network improvements, video quality can still suffer during periods of congestion when networks do not have enough capacity to meet current demands. Congestion for streaming video means the rate the video is received by a client player is lower than the rate the video is played out. When this happens, the video must inevitably stop playing until enough video has arrived to resume playout. Even when the overall average receive rate is greater than the average playout rate, fluctuations in the rate over the life of the video playout can cause periods where the instantaneous receive rate is insufficient to meet playout demands.

Playout buffering, depicted in Figure 1, is a technique that can reduce stops in video playout due to fluctuations in the incoming network bitrates. With playout buffering, the client buffers received video for some time before playing it out at the normal playout rate. Over the life of the video playout, the actual amount buffered varies, increasing when the incoming rate (A) is greater than the video playout rate (B) and decreasing when the incoming rate is less than the video playout rate. However, as long as the buffer does not drain completely, the video playout can proceed smoothly, without interruption. This suggests the client should use a large buffer to prevent the buffer from draining completely and interrupting playout. However, the larger the buffer, the longer the user must wait for playout to begin. There is a fundamental trade-off between the degradation to the viewer's quality of experience (QoE) caused by the delay when waiting for a video to buffer and the degradation to the viewer's quality of experience caused by playout interrupts. Thus, the size of the buffer is an important parameter the video player chooses in order to balance the delay from buffering and the interrupts when there is no remaining buffer.

Commercial systems have quantified the amount of buffering and/or interrupts observed (e.g., 20% of streaming videos have at least one interrupt and 20% have long buffer times [7]), or have surveyed users' frustration with buffering [3], but do not measure, much less report, how each affects user QoE. Other work has analyzed methods to compensate for interrupts by buffering, quantifying the exact trade-off between buffers and interrupts [15], but

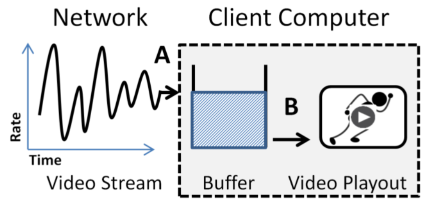

Figure 1: Playout buffer as part of a streaming video system.

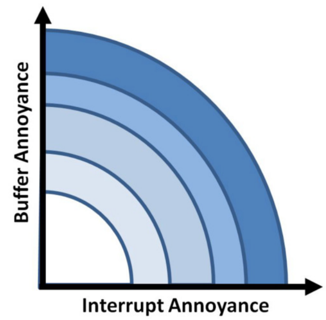

Figure 2: Video quality of experience for buffering and interrupts. The best quality is at the origin. Lines represent regions of equivalent quality.

has not quantitatively compared the impact to QoE for each. Additional research has attempted to infer video QoE by how long users watched a streaming video [19], but without directly measuring or knowing the relationship between buffer times and interrupts. Another approach [2] examined various video metrics in an effort to find components relevant to QoE, but only inferred quality by viewing time. Research that did measure and model the impact of interrupts and buffering delay on QoE [10], but fails to quantify interrupts.

Unfortunately, choosing the best buffer size can only be done if the comparative impacts on QoE for buffering delay and interrupts are quantified. In other words, the amount that delay impacts viewing QoE compared to the amount that interrupts impact viewing QoE must be clearly established. Streaming approaches that use buffer occupancy [1, 11, 12] or throughput [18] in order to pick streaming rates, and models that assess the probability of interrupts for different startup delays [20] would benefit from the explicit "cost" to the viewing experience from interrupts versus delay. In the absence of this information, current video players must resort to heuristics to choose buffer sizes, hoping these result in the best viewing experiences. Instead, if the impacts of buffer delays and number of interrupts on QoE were quantified, then video players could make informed decisions about the sizes of their buffers, effectively finding the balance between buffering delay and interrupts that maximizes viewer QoEs.

For an illustration of a possible representation of QoE for buffering and interrupts, consider Figure 2. The x-axis represents the user annoyance from interrupts and the y-axis the user annoyance from buffering. The best quality is at the origin, with no buffering and no interrupts. However, as noted above, during congestion, some amount of buffering is necessary or there will be interrupts. The axes are normalized so that a unit degradation in QoE due to interrupts is the same as a unit degradation in quality due to buffering. For video streams with both buffering and interrupts, QoE could be computed by the Euclidean distance from the origin, providing regions of equivalent quality around this origin. A video player could then choose the best buffer size so as to maximize quality based on these dimensions.

Our approach is to take a step towards ascertaining the best video buffer sizes by measuring and then modeling the impacts on QoE for buffering delay versus playout interrupts through a user study designed to provide a direct comparison between buffering delays and interrupts. A set of short videos with a range of content were selected and re-encoded with different amounts of buffering delays and interrupts. The videos were embedded into an interface that allowed users to watch and evaluate the videos in a controlled laboratory setting. Volunteers were solicited to participate in the study, each watching and rating 17 videos. Over 30 users participated, providing about 600 quantifiable QoE data points on buffering delays and interrupts.

Analysis of the results shows that based on total time, playout interrupts have a larger degradation to QoE – since our method allows for a quantifiable comparison, about twice as large – as buffering delay. As a guideline, buffer times of about 30 seconds degrade QoE the same as do 2 one-second interrupts in playout. QoE degradations for both buffering delays and interrupts are well modeled by logarithmic functions confirming earlier work [9]. Viewing these models together enables analysis of equivalent regions of quality. Moreover, the models can be combined to provide an overall model for QoE which can be used by streaming video clients to determine the best buffer sizes given observed network conditions.

The rest of this paper is organized as follows: Section 2 summarizes research related to this paper; Section 3 describes our user study to measure the impact of buffering delay and interrupts; Section 4 analyzes the resulting data and derives our QoE models; and Section 5 summarizes our conclusions and presents possible future work.

2 RELATED WORK

Egger et al. [9] discuss the psycho-physics basis for a model of quality of experience and human time perception (e.g., delay when buffering a video during initial playout). The authors describe a set of studies that lay out the basis for a logarithmic relationship for waiting time and user satisfaction ratings. This relationship forms the basis of a "WQL" model, which assumes the impact of waiting time (W) on the Quality of Experience (Q) for video is logarithmic (L). We use this WQL relationship in modeling our QoE experiences.

Modlovan and Hoßfeld [17] study the impact of network bitrate variance (and, hence, video variance) on the QoE of video. The authors use a previously established QoE model to map network parameters to video quality, modeling QoE as an exponential with the number of interrupts and buffering length. However, their explorations do not consider the initial buffer delay, but only consider delays that occur once playout has started (i.e., interrupts). Work such as theirs could benefit from a model of QoE that includes initial buffer delays, too.

Mok et al. [16] investigate the relationship among network quality of service (QoS) and video quality of experience (QoE). Their user study results show the frequency of rebuffering events is the main factor in QoE variation. They produce an analytic model for QoE with interrupts that is coarsely parameterized by 3 levels. Our work is based on a 3x larger user study and quantifies the relationship between buffering and interrupts more precisely. Our study confirms the authors' results that interrupts degrade QoE more than do the initial buffer delays.

Dobrian et al. [8] analyze a real-world dataset obtained from client-side measurements streaming from popular video content providers. They quantify user engagement, finding that the percentage of time spent buffering (effectively, they sum the total time spent during interrupts) has the largest impact on the user engagement. Our work is complementary in that we also analyze the impact of buffer time relative to interrupts, thus confirming these results, but we also develop models that allow buffering and interrupts to be directly compared.

Li et al. [14] and Kim et al. [13] investigate video playout buffer requirements for a desired video quality given network and video bitrate characteristics. They develop and experimentally validate a video streaming model and derive an analytic expression of the minimum playout buffer required. Our work is complementary to such approaches in that a more accurate QoE model based on initial buffer delays and number of interrupts can be used by lower-level analytic models to determine appropriate system-level buffers.

Research by Hoßfeld et al. [10] is most closely related to ours, measuring the impact on QoE from delays and interrupts through a user study. In fact, as a bonus, the authors produce a model of QoE for initial delay that is directly comparable to our model. Our work reproduces aspects of Hoßfeld et al.'s for confirmation, and extends their work by computing effect sizes, directly comparing buffering and interrupts on QoE, and modeling the impact of the number of interrupts instead of just the total length of the interrupts.

3 METHODOLOGY

Our method to measure degradation to streaming video quality of experience (QoE) for number of interrupts versus buffering delay is as follows: 1) select and re-encode videos (Section 3.1); 2) develop an interface for users to watch and evaluate videos (Section 3.2); 3) install videos in a dedicated user laboratory (Section 3.3); 4) solicit users to watch and rate videos (Section 3.4); and 5) analyze the results (Section 4).

3.1 Videos

Videos were selected from the Internet, primarily YouTube. A broad range of content was chosen, from online comedy (e.g., the Daily Show[1]) to music videos (e.g., Bruno Mars[2]), with the intent to have videos with general appeal so as not to let specific content unduly influence opinions on QoE. All the content was live action (as opposed to animation). In all, 17 videos were selected, one for each of the experimental conditions (see below). The full list of videos with their descriptions can be found on our Web site.[3]

All videos had an original source of 1280x720 pixels at 30 frames per second, encoded in MPEG-4. From their original source, the videos were clipped to 30 seconds long, where each clip contained a single scene with no transitions. Each video was then re-encoded with artificial buffering delay and interrupts using Microsoft Windows Movie Maker. The buffering and interrupts were depicted in the videos as the YouTube buffering animation on a black background.

Each video was re-encoded in eight ways – with four different buffer delays and with four different numbers of interrupts. This preparation supported a between-subjects study, where different users watch the same content, but enabled evaluation of video degradations in multiple ways.

Buffer delays were 2, 4, 8 and 16 seconds. Interrupts were each one second, uniformly placed in the video for 2, 4, 8, and 16 interrupts.

The pre-encoded videos were stored locally during the user study in order to avoid any additional artifacts (e.g., buffering delays and/or interrupts) from streaming that were not controlled.

3.2 Interface

A graphical user interface was created to allow users to watch and rate the locally-stored videos. Figure 3 shows a screenshot of the user study interface. The top line shows what number video the user is watching and how many remain until the study is complete as a progress indicator. The main graphic shows the video as it plays. The user presses the play button and watches the video until completion. After the video playout has stopped, the user rates their annoyance using a slider on a 5 point scale and provides an opinion on the content, also with a slider on a 5 point scale.

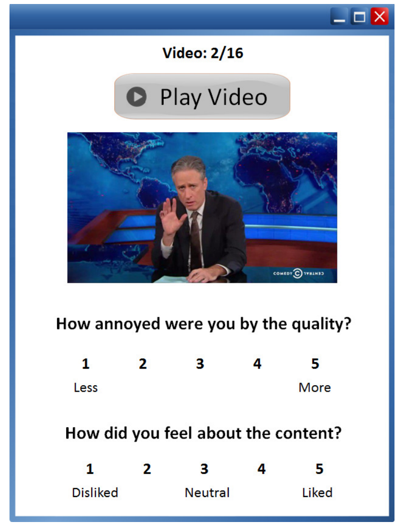

Figure 3: User study interface.

[1]https://en.wikipedia.org/wiki/The_Daily_Show
[2]https://en.wikipedia.org/wiki/Bruno_Mars
[3]https://web.cs.wpi.edu/~claypool/papers/buff-int/

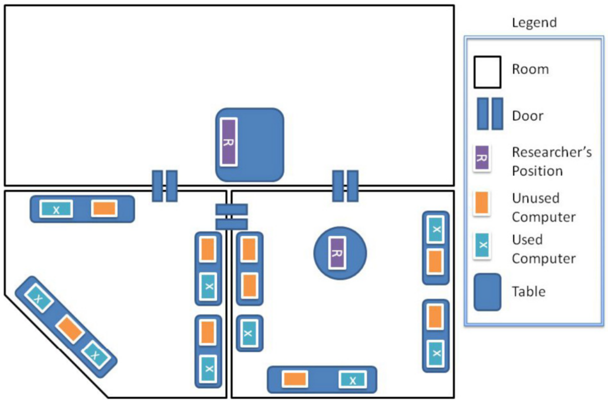

Figure 4: Lab for user study.

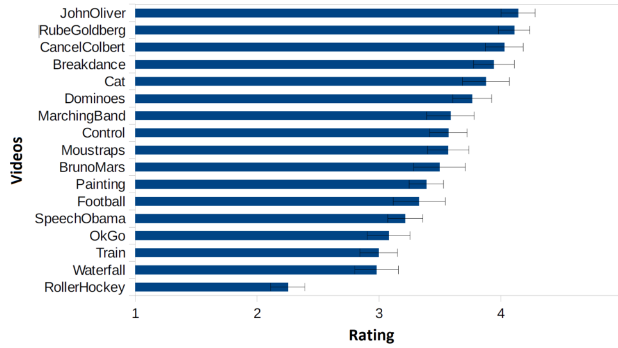

Figure 5: Content ratings. Horizontal bars are mean ratings across all users, shown with standard error bars.

3.3 Lab

Our user study was conducted in a computer lab dedicated to the study, lit with bright, fluorescent lighting, the layout shown in Figure 4. The researcher would stage participants for consent information before leading participants to a computer workstation. The computers were Dell PCs with Intel i7-3770 processors and 12 GB of RAM, running Microsoft Windows 7. The monitors were 24" Dell U2412M LCDs with a native resolution of 1920x1200 pixels and a refresh rate of 59p Hz.

3.4 Procedure

Participants were solicited through advertising via University email lists and the University Social Science research participant pool.[4] Incentives included a raffle for a $25 gift card for participating.

For each user, the study proceeded as follows:

(1) The user heard a scripted brief about the purpose of the study and signed an Institute Review Board (IRB) consent form.
(2) The user sat at a lab computer, was told to make themself comfortable by adjusting chair height and monitor angle/tilt and to put on headphones and adjust the volume.[5]
(3) The user filled out a demographic information, (e.g., gender, age, major, and experience with streaming media), coded using the Qualtrics survey tool.[6]
(4) The user watched a video with no buffer delay and no interrupts, being told that this video had the best conditions to use as a reference.
(5) The user watched videos with only buffering delay or playout interrupts. After each video, the user rated QoE (annoyance) and content.

The total time to complete the user study for one user, including filling out the initial forms and then watching and rating all 17 videos, was about 15 minutes.

[4]https://wpi.sona-systems.com/
[5]Headphones were provided for use, but participants were also welcome to use their own.
[6]https://www.qualtrics.com/

4 ANALYSIS

This section presents the demographics of the users that participated in the study (Section 4.1), graphical analysis of the results (Section 4.2), and regression models and comparison of quality with buffer delays and interrupts (Section 4.3).

4.1 Demographics

Thirty-seven users participated in the study. Twenty-two identified as male and 15 as female, about the same gender breakdown as our University as a whole. Most (about 35%) were students majoring in Computer Science or Game Development, with the next most popular majors other science and/or engineering disciplines. Ages ranged from 17-22 years with a mean and median of 20. The mean self-rating for how often users watched streaming video, 1 (rarely) to 5 (often), was 4.7, with a standard deviation of 0.9.

4.2 Results

Figure 5 shows the distribution of the content ratings, ordered high to low, each horizontal bar the mean rating across all users and shown with a standard error bar of spread. Content ratings for the 5 point scale averaged a low of 2.25 for Roller Hockey to a high of 4.1 for John Oliver. Most videos averaged above a 3, indicating they were more liked than disliked. Subsequent analysis showed no correlation between content and QoE for any levels of buffer delay or interrupts, so we do not consider content differences further in this paper. Similarly, subsequent analysis showed no correlation between levels of temporal motion or spatial scene complexity [5] in the video content and buffer delay or interrupts, so we do not consider these elements further.

Figure 6 depicts cumulative distribution functions (CDFs) of the annoyance levels for different amounts of buffer delay. The x-axis is the annoyance level and the y-axis is the cumulative distribution. There are four trend lines, one for each buffer delay in the user study: 2, 4, 8 and 16 seconds. There is a separation of trend lines, where the higher the buffer size the lower and to the right the line (i.e., more annoyed). An ANOVA test shows there was a significant effect of buffer delay on annoyance at the 0.05 significance level [$F(1, 286) = 10.91$, $p = 0.001$]. However, the fact that the trend lines cover the same range horizontally indicates that a

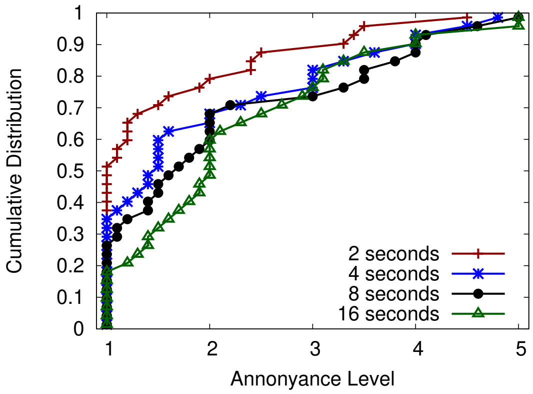

Figure 6: Cumulative distribution of annoyance for buffer delay.

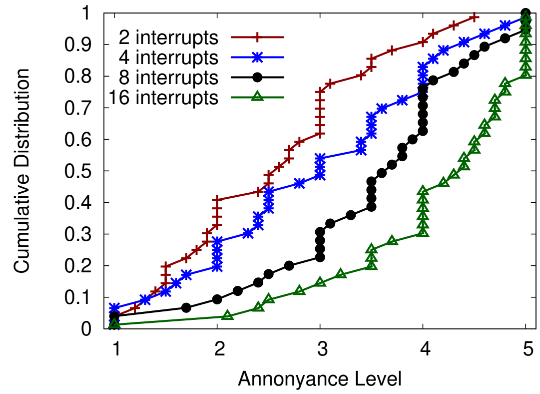

Figure 7: Cumulative distribution of annoyance for number of interrupts.

lower buffer delay is not always perceived as less annoying than a higher buffer delay for all videos for all users.

We compare the mean annoyance values for buffer delay – 2, 4, 8, and 16 seconds – for each pair by doing independent, 2-tailed t tests ($\alpha = 0.05$) with a Bonferroni correction, as well as compute the effect sizes. The Cohen's d effect size quantifies the differences in means in relation to the standard deviation, providing a quantitative measure of the magnitude of the difference in annoyance for different buffer delays. Generally small effect sizes are anything under 0.2, medium is 0.2 to 0.5, large 0.5 to 0.8, and very large is above 0.8. The t test and effect size results are shown in Table 1. From the table, the difference between 2 seconds and 16 seconds of buffer delay is significant with a large effect size, and the effect sizes for all other pairs except of 8 and 16 seconds are medium.

Table 1: Buffer delay annoyance t tests and effect sizes.

Buffers		t(72)	p	Effect Size
2	4	1.56	0.120	0.36
2	8	2.18	0.033	0.51
2	16	2.74	0.007	0.64
4	8	0.98	0.332	0.14
4	16	1.02	0.308	0.24
8	16	0.37	0.780	0.09

Figure 7 depicts a CDF of the annoyance levels for different numbers of interrupts, with the same axes as for Figure 6. Here, the four trend lines are for the number of interrupts in the user study: 2, 4, 8 and 16. There is a clear separation of trend lines, where the more interrupts the lower and to the right the line (i.e., more annoyed). An ANOVA test shows there was a significant effect of number of interrupts on annoyance at the 0.05 significance level [$F(1, 301) = 99.80$, $p < 0.0001$]. As for buffer delays, there is horizontal overlap in the trend lines, indicating that fewer interrupts is not always perceived as less annoying than more interrupts for all videos for all users.

Similar to Table 1, we compare interrupts via t tests and compute effects sizes, shown in Table 2, From the table, the differences between all pairs are significant, except for 2 and 4 interrupts (with medium effect size), and all other effect sizes are large. Generally, the effect sizes for interrupts are larger than those for buffer delay.

Table 2: Interrupt annoyance t tests and effect sizes.

Interrupts		t(72)	p	Effect Size
2	4	1.64	0.1030	0.38
2	8	3.99	0.0002	0.88
2	16	7.38	0.0001	1.67
4	8	2.31	0.0184	0.53
4	16	5.06	0.0001	1.15
8	16	2.89	0.0065	0.64

Comparing Figure 7 to Figure 6 shows the distributions shifted to the right for interrupts, suggesting annoyance levels are higher for the number of interrupts tested compared to the amount of buffer delays tested. This relationship becomes clearer when comparing the mean values.

Figure 8 shows a graph of the mean annoyance levels versus the buffer delays and number of interrupts. The bottom x-axis is the buffer delay in seconds and the top x-axis is the number of interrupts. The y-axis is the annoyance level. Each point is the mean annoyance level across all users, shown bounded by a 95% confidence interval. There is a noticeable increase in the mean annoyance level as both the buffer delay and the number of interrupts increase, with a more sharp increase in annoyance on the left side of the trendlines compared to the right. The trend in QoE with buffer delay independently confirms results reported previously [10].

Since each interrupt is 1 second, there is some merit in comparing, for example, 4 interrupts to 4 seconds of buffer delay in that they both have the user wait for the same amount of time. Comparing the interrupts trendline to the buffer delay trendline shows the means are higher for interrupts, suggesting annoyance levels are

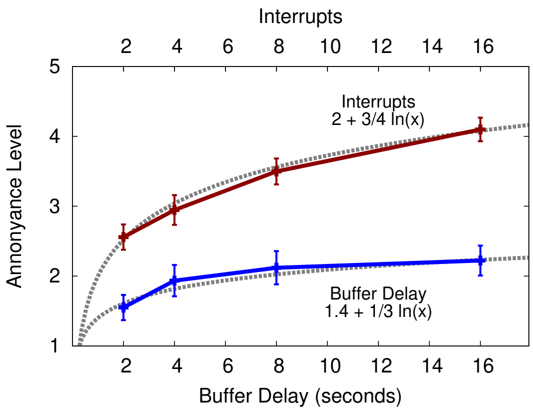

Figure 8: Mean annoyance for buffer delay (bottom axis and trendline) and interrupts (top axis and trendline).

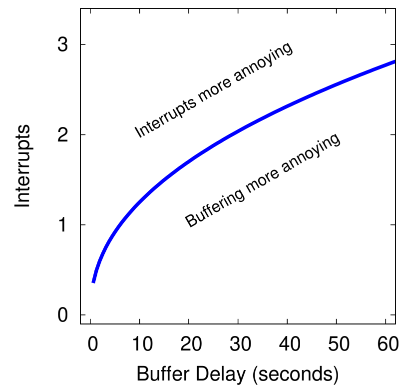

Figure 9: QoE for interrupts versus buffer delay. The curved line depicts where annoyance from buffer delay (x-axis) equals the annoyance from number of interrupts (y-axis).

higher for the same amount of waiting when, as is true for interrupts, the waiting is divided into pieces and occurs in the middle of the video. While this result has been suggested before (e.g., [10]) and may even make intuitive sense, to the best of our knowledge, Figure 8 is the first graph to quantitatively compare buffering and interrupts. Moreover, the data allows trading-off interrupts and buffer delay and computing Quality of Experience (QoE) with an analytic model (next section).

4.3 Model

Based on Egger et al. [9], since the expectation is for annoyance to increase sharply then more gradually as the number of interrupts and amount of buffer delay increase, we fit[7] a logarithmic regression model to the annoyance level for both buffer delay and interrupts. For buffer delay, the resulting model for annoyance level A_b (from 1 (low) to 5 (high)) is:

$$A_b = \begin{cases} 1.4 + \frac{1}{3} ln(b) & b \geq 0.6 \\ 1.25 & otherwise \end{cases} \tag{1}$$

where b is the buffer delay, in seconds. The model is shown as the lower dashed curve in Figure 8. The adjusted R^2 is 0.88.

For interrupts, the resulting model for annoyance level for interrupts A_i (from 1–5) is:

$$A_i = \begin{cases} 2 + \frac{3}{4} ln(i) & i \geq 0.4 \\ 1.25 & otherwise \end{cases} \tag{2}$$

where i is the number of interrupts. The model is shown as the upper dashed curve in Figure 8. The adjusted R^2 is 0.99.

Note, given the natural logarithms in Equations 1 and 2, A_b and A_i go asymptotically to $-\infty$ as buffer delay and interrupts go to zero, respectively. Since the average annoyance level is 1.25 for the control video (no interrupts, no buffer delay), that is used as a minimum value for both functions.

The user study by Mok et al. [16] suggests there is no interaction between interrupts and initial buffering in determining QoE. Thus, assuming the annoyance due to buffer delays is independent of the annoyance due to number of interrupts, the total annoyance level A_t is then:

$$A_t = A_b + A_i \tag{3}$$

Equation 3 allows for a direct comparison of the degradation in QoE (increased annoyance, in our case) due to buffer delays (A_b) and number of interrupts (A_i), respectively. Since the client video player has direct control of the playout buffer size (the buffer delay), it is useful to relate the number of interrupts to the playout buffer delay. Solving for where the annoyance from the number of interrupts (i) is equivalent to the annoyance from the seconds of buffer delay (b) ($A_i = A_b$) yields:

$$i = e^{\frac{4ln(b)-7.2}{9}} \tag{4}$$

Figure 9 depicts the equal number of interrupts-buffer delay relationship by graphing Equation 4. The x-axis is the buffer delay chosen by the client and the y-axis is the resulting number of interrupts for that buffer size. The blue line is the curve from Equation 4. For example, a streaming video with 30 seconds of buffer delay has about the same level of annoyance as 2 interrupts. The area below this curve is where the user is more annoyed by the buffer delay than by the interrupts. The area above this curve is where the user is more annoyed by the interrupts than by the buffer delay.

The blue curve in Figure 9 provides guidance for choosing the buffer size (hence the buffer delay). When streaming, if the buffer size chosen and resulting number of interrupts lies below the curve, the buffer should be decreased until the buffer delay & number of interrupts intersects the curve. Similarly, if the buffer size chosen and resulting number of interrupts lies above the curve, the buffer size should be increased until the buffer delay & number of interrupts intersects the curve.

For example, consider a hypothetical example in Figure 10. The x-axis is the buffer delay chosen by the client and the y-axis is the resulting number of interrupts. The red zig-zag line represents the

[7] Using R, https://www.r-project.org/

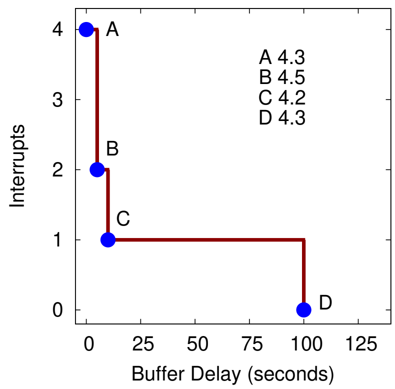

Figure 10: Choosing Buffer Size. Red zig-zag line represents video quality for different buffer sizes. QoEs for indicated points are shown in upper right.

set of QoEs that could be achieved for the playout of this particular video under this specific set of network conditions. The client video player could choose a playout buffer anywhere along the x-axis and, for this particular situation, the corresponding number of interrupts would lie on the given red line. For example, if the client chose a 50 second buffer, there would be 1 interrupt during playout. Increasing the buffer size where the red line is horizontal only increases the buffer delay but does not decrease the number of interrupts. For example, increasing the buffer from 50 to 60 seconds still has 1 interrupt. Thus, the best QoE is at one of the points: A, B, C or D.

Using Equation 3, the QoEs for points A, B, C and D are computed and shown in the upper right corner of Figure 10 (lower is better). The best quality is achieved at point C, where the QoE is 4.2. Point C is better than points A or B because a modest increase in the buffer size from 0 to 10 seconds decreases the number of interrupts from 4 to 1. Decreasing the interrupts from 1 to 0, however, going from point C to point D, requires 100 seconds of buffering which is not worth the cost to QoE.

5 CONCLUSION

Video players accommodate the variations in bitrates inherent in today's Internet using playout buffering, especially important for wireless video over mobile devices where bitrates can vary considerably. Playout buffers hold arriving video for a short period of time before playing it out, thus providing smooth playout even when the network bitrate momentarily drops below the playout rate. In fact, if large enough, playout buffers can completely eliminate all mid-stream video interruptions. However, a larger buffer means a longer waiting time for the user to start the video. Thus, there is a trade-off between the delay in buffering and the interrupts in video playout. Having a quantifiable impact of delay buffering and number of interrupts on Quality of Experience (QoE) would allow for models that could help determine the buffer size that maximizes QoE for given video and network conditions. While past work has studied aspects of the problem, choosing buffer sizes based on heuristics or has studied the effects of video parameters, including playout interrupts, on QoE, to the best of our knowledge, there have not been measurements directly comparing the impact to QoE from buffering delays versus number of interrupts.

This paper presents the derivation of a model that can ascertain the best streaming video buffer size based on the impact on QoE from buffering delay and number of interrupts. A wide range of short videos were selected and encoded with 4 different buffer times and 4 different numbers of interrupts. A user study with 37 participants had users watch the videos in a controlled setting, providing quantifiable ratings comparing QoE under the different conditions. Analysis of the data shows interrupts have a much greater impact on QoE, about 2x greater, than does buffering for the same amount of time. An analytic model derived from the data allows computation of the exact QoE for a video with a given buffer delay and interrupt count, providing a tool for analyzing buffer sizing in video clients to determining buffer sizes that maximize QoE.

While the model is a promising step towards choosing and evaluating playout buffer sizes, there are additional areas of future work that can be considered. The QoE model, Equation 3, assumes the impact of buffer delay and number of interrupts on QoE are independent. While this assumption is supported by other research [16], a user study that explicitly measures user annoyance with videos with both interrupts and delay can verify if this assumption holds and, if appropriate, derive a needed interaction term. Our QoE model was developed from a set of specific video parameters – $1280x720pixels$, 30 f/s and 30 seconds long and interrupt lengths were fixed at 1 second, evenly spaced. Additional research can investigate if the results hold for a wider range of video lengths and encoding parameters, buffering lengths and interrupt lengths and distributions – Internet bitrate decreases can be uneven. The interaction between accompanying audio may also be worth studying as there was some indication that interrupts during speech or music was more annoying than interrupts during silence. Lastly, the results are taken in a laboratory setting which may influence user tolerance for buffer delay and interrupts compared to, say, a living room couch. Future work could repeat the experiments in non-laboratory settings and tease out "real-world" differences that impact user QoE tolerances.

REFERENCES

[1] Lamine Amour, Sami Souihi, Abdelhamid Mellouk, and S.M. Mushtaq. 2020. Q2ABR: QoE-aware Adaptive Video Bit Rate Solution. *International Journal of Communication Systems* 33, 10 (July 2020).

[2] A. Balachandran, V. Sekar, A. Akella, S. Seshan, I. Stoica, and H. Zhang. 2012. A Quest for an Internet Video Quality-of-Experience Metric. In *Proceedings of HotNets*. Seattle, WA, USA.

[3] Ross Benes. 2018. Buffering Is the Streaming Snafu that Won't Go Away. Online: https://tinyurl.com/y3yeghr2.

[4] Cisco. 2020. Cisco Annual Internet Report (2018–2023). Online: https://tinyurl.com/y9lfom7e.

[5] Mark Claypool. 2009. Motion and Scene Complexity for Streaming Video Games. In *Proceedings of the 4th ACM International Conference on the Foundations of Digital Games (FDG)*. Florida, USA.

[6] J. Clement. 2019. Devices Used to Watch Online Video in the U.S. Statista. Online: https://www.statista.com/statistics/784383/online-video-devices-in-the-us/.

[7] Conviva. 2012. Viewer Experience Report. http://www.conviva.com/vxr [Accessed 1-March-2014].

[8] F. Dobrian, V. Sekar, A. Awan, I. Stoica, D. Joseph, A. Ganjam, J. Zhan, and H. Zhang. 2011. Understanding the Impact of Video Quality on User Engagement. *ACM SIGCOMM CCR* 31, 4 (2011).

[9] S. Egger, P. Reichl, T. Hoßfeld, and R. Schatz. 2012. Time is Bandwidth? Narrowing the Gap between Subjective Time Perception and Quality of Experience. In *IEEE ICC ICC CQRM*. Ottawa, Canada.

[10] T. Hoßfeld, S. Egger, R. Schatz, M. Fiedler, K. Masuch, and C. Lorentzen. 2012. Initial Delay vs. Interruptions: Between the Devil and the Deep Blue Sea. In *Proceedings of IEEE QoMEX*.

[11] Weiwei Huang, Yipeng Zhou, Xueyan Xie, Di Wu, Min Chen, and Edith Ngai. 2018. Buffer State is Enough: Simplifying the Design of QoE-Aware HTTP Adaptive Video Streaming. *IEEE Transactions on Broadcasting* 64, 2 (June 2018).

[12] Nabin Kumar Karn, Hongli Zhang, Feng Jiang, Rahul Yadav, and Asif Ali Laghari. 2020. Measuring Bandwidth and Buffer Occupancy to Improve the QoE of HTTP Adaptive Streaming. In *Proceedings of NetSoft*. Ghent, Belgium.

[13] T. Kim, N. Avadhanam, and S. Subramanian. 2006. Dimensioning Receiver Buffer Requirement for Unidirectional VBR Video Streaming over TCP. In *Proceedings of IEEE Image Processing*.

[14] Mingzhe Li, Mark Claypool, and Robert Kinicki. 2009. Playout Buffer and Rate Optimization for Streaming over IEEE 802.11 Wireless Networks. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP)* 5, 3 (Aug. 2009).

[15] G. Liang and B. Liang. 2007. Balancing Interruption Frequency and Buffering Penalties in VBR Video Streaming. In *Proceedings of IEEE Infocom*.

[16] R. Mok, E. Chan, and R. Chang. 2011. Measuring the Quality of Experience of HTTP Video Streaming. In *Proceedings of IFIP/IEEE IM*. Dublin, Ireland.

[17] Christian Moldovan and Tobias Hoßfeld. 2016. Impact of Variances on the QoE in Video Streaming. In *Proceedings of IEEE ITC*.

[18] Minh Nguyen, Christian Timmerer, and Hermann Hellwagner. 2020. H2BR: an HTTP/2-based Retransmission Technique to Improve the QoE of Adaptive Video Streaming. In *Proceedings of the 25th ACM Workshop on Packet Video*.

[19] F. Qiu and Y. Cui. 2011. A Quantitative Study of User Satisfaction in Online Video Streaming. In *Proceedings IEEE CCNC*. Las Vegas, NV, USA.

[20] Philipp Schulz, Henrik Klessig, Meryem Simsek, and Gerhard Fettweis. 2020. Modeling QoE for Buffered Video Streaming in Interference-Limited Cellular Networks. *IEEE Transactions on Multimedia* (April 2020).

[21] Paul Verna. 2019. APAC Dominates in Mobile Phone Video Viewership. eMarketer. Online: https://tinyurl.com/y5eqtsvb.