Are Incentive Schemes Needed for WebRTC based Distributed Streaming? A Crowdsourced Study on the Relation of User Motivation and Quality of Experience

Matthias Wichtlhuber
TU Darmstadt, Germany
mwichtlh@ps.tu-darmstadt.de

Nikola Aleksandrov
TU Darmstadt, Germany
naleks@ps.tu-darmstadt.de

Markus Franz
TU Darmstadt, Germany
franz@emarkets.tu-darmstadt.de

Oliver Hinz
TU Darmstadt, Germany
hinz@wi.tu-darmstadt.de

David Hausheer
TU Darmstadt, Germany
hausheer@ps.tu-darmstadt.de

ABSTRACT

Video traffic is the main driver of Internet traffic volume. Thus, content providers and Content Delivery Networks (CDNs) are searching for ways to provide reliable video transmission at a low cost. Hybrid CDN/Peer-to-Peer (P2P) deployments like Akamai NetSession have been shown to combine the high reliability of a CDN backbone and the low cost of P2P networks. In the near future, the biggest barrier for user adoption will fall: the installation of a dedicated P2P client software will be replaced by website embedded browser-to-browser communication logic. However, this requires the explicit consent of users, and, since users need to share their upload capacity, their willingness to participate in such a system. In this work, the efficiency of incentive mechanisms trading a higher Quality of Experience (QoE) of video transmission for user's consent to utilize their upload capacity are investigated. This is the first study to investigate the question of incentives in distributed, adaptive streaming systems from a user perspective using a crowd working approach. The work presents results from 192 test subjects. We identify three classes of users and show how behavioral economics can be utilized to increase the impact of an incentive scheme.

CCS Concepts

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

Keywords

Quality of Experience; Incentive; P2P

MMSys'16, May 10-13, 2016, Klagenfurt, Austria

DOI: http://dx.doi.org/10.1145/2910017.2910598

1. INTRODUCTION

Video traffic has seen a steep increase in volume during the last years [3]. Cisco projects that in "Western Europe, IP video will be 79% of all traffic in 2018, up from 59% in 2013". The trend is driven by the growing access bandwidths at home premises, new mobile devices and the rise of commercial and non-commercial content providers like YouTube, Netflix, or Twitch. However, from the perspective of CDN network providers responsible for delivering the data, the massively growing video traffic volume increases the pressure for reliable video content distribution at a low cost.

Consequently, the idea of combining a P2P network and a CDN backbone for content distribution is a compelling concept. On the one hand, a large share of the upstream traffic of CDN servers can be shifted to the end user, thus decreasing bandwidth costs for the CDN provider. On the other hand, the CDN provider can tightly monitor the system and compensates for the unreliability of peers by throwing in backbone resources whenever needed. Probably the most prominent example of this type of systems is Akamai NetSession, which is reported to consist of more than 32 million peers, offloads a share of up to 80 % of traffic from the CDN backend to end users, and delivers roughly the same transmission quality as the pure CDN service [35].

Currently, systems like NetSession rely on the installation of dedicated client software usually bundled with other software (e.g., video games). However, CDN/P2P hybrid content distribution is likely to gain much more momentum with the widespread adoption of inter browser communication. In particular, the Web Real Time Communication (WebRTC) framework [5] is currently in the standardization phase. It offers the possibility to exchange data as well as video streams among browsers using pure JavaScript.

Inter browser communication in the context of distributed streaming re-raises the long researched question of user motivation to participate in such a system. On the one hand, the barrier of participation is much lower than in traditional P2P systems, only requiring the user to click a button instead of installing a software. However, from a purely economic point of view, a user giving consent to providing upload capacity for the benefit of a content provider acts irrational [11]. Consequently, user motivation should increase when an appropriate incentive in terms of better QoE of the streaming session is provided.

To the best of the authors' knowledge, this work is the first to perform research on this question from a user perspective in a large-scale crowdsourcing based user study. In particular, this work addresses four research questions:

(a) How high is the fraction of altruistic users giving consent without further benefits?
(b) How high is the fraction of non-altruistic users that can be convinced to give their consent in exchange for a better QoE?
(c) Can incentive mechanisms based on findings from behavioral economics be utilized to increase consent?
(d) How sensitive are users to a utilization of upload capacity without consent?

The remainder of this work is organized as follows: Section 2 explains the necessary background information to understand the work and discusses related approaches. Section 3 discusses the user study design based thereupon and the selection and preparation of test video material. Section 4 presents evaluation results and Section 5 concludes the work and gives an outlook on future extensions.

2. BACKGROUND AND RELATED WORK

This work is based on a number of building blocks which are discussed in the following paragraphs. Related work performed in the area of the respective building block is discussed accordingly.

Crowdsourcing: Crowdsourcing became widely popular during the last decade, as it allows for a fast and cost-efficient execution of online micro tasks requiring human intelligence (Human Intelligence Tasks (HITs)). Platforms offering crowdsourcing as a service are virtual market places acting as a broker between online workers and employers. Employers place an offer for a HIT which workers can choose to work on for a predefined, usually small amount of money. In the area of multimedia systems, crowdsourcing has been widely used to perform user studies for QoE assessment due to fast turnaround times and low cost. However, unreliability of crowd workers and a less controllable environment are problems that have to be addressed through task design.

Keimel et al. [18] provide a software framework for crowd sourced video QoE assessment and provide valuable recommendations on experiment setups [17]. A similar concept is proposed and evaluated by [7], showing that with a well-defined user response filtering approach, the results do not deviate from parallel laboratory studies. The authors of [13] conduct a QoE assessment study to investigate the influence of playback disruptions for YouTube content. They are able to quantify the QoE deterioration by using mathematical models with a high goodness of fit. The authors show that the results do not deviate largely from a lab study conducted in parallel. Moreover, a valuable source on the conduction of QoE experiments and user response filtering is provided by the same authors [12]. Closely related to our work is the work from Sackl et al. [26], which evaluate the relation of charging schemes and video QoE in a crowdsourcing study. While presenting a number of interesting findings w.r.t. the monetization of QoE, Sackl et al. do not investigate the question of bandwidth sharing.

Adaptive Video Streaming: Adaptive video streaming refers to the adaptation of video quality in terms of spatial resolution, temporal resolution or quantization of the compression codec (i.e., an adaptation of the Signal-to-Noise Ratio (SNR)) during the transmission from server to client. The adaptation is done to reduce the bandwidth of the video and to prevent freezing playback of the video, which was shown to be highly negatively correlated to user's QoE [10]. With the advent of web standards like Dynamic Adaptive Streaming over HTTP (DASH) [27], the transmission of adaptive video streams is about to become widespread. As this work does not investigate adaptive streaming, but merely utilizes the adaptive streaming capabilities of modern browsers, we refer to [27] for an in-depth discussion.

Video Quality Assessment: Video Quality Assessment methods can be subdivided into subjective methods and objective methods [8]. This work uses both approaches. A subjective, crowdsourced study is performed to relate the value of upload capacity to an increase of video QoE. Moreover, we measure video quality differences with an objective, Full Reference (FR) video quality model in order to determine the visual degradation of a video compared to a high quality reference. In this work, Pinson et al.'s Video Quality Metric (VQM) is used, as VQM reaches a high Pearson Correlation Coefficient (PCC) for subjective test results ($>$ 0.9 [8]) and was standardized by the National Telecommunication and Information Administration (NTIA). Moreover, VQM is user validated for resolutions up to full HD video sequences [23, 32] and a mapping to the Mean Opinion Score (MOS) scale exists [36].

CDN/P2P hybrids and WebRTC based streaming: CDN/P2P hybrid systems exist as deployments and scientific research prototypes. As opposed to pure P2P systems, these systems rely on a stable CDN backbone and are tightly controlled by a central authority to reach a maximum offloading factor from the backend without sacrificing Quality of Service (QoS)/QoE. The most prominent example of a deployment is Akamai's NetSession system. Akamai reports 32 million participating active clients[1]. According to [35], NetSession offloads a share of up to 80 % (depending on the swarm size) from the backend and delivers comparable transmission quality to the pure CDN service. Among the list of research systems, [15] focuses on file sharing and live streaming, achieving a reduction of the backend traffic between 70% and 95%. The authors of [33] design LiveSky. LiveSky focuses on live streaming and reaches an offloading factor of 58%. The system was validated with 145000 users during the transmission of the Chinese National Congress. The authors of SmoothCache [25] do not report an offloading factor for the backend. The system is optimized for low delay during live streaming transmissions and reaches less than 5s between source and peers.

A number of related works investigate WebRTC based P2P streaming. As the WebRTC standard is still in its infancy, most academic works do not go beyond a proof of concept, such as [31, 24]. Nevertheless, a number of startups and open source projects exist. Bem.tv[2] and Peer5[3] provide working demos. The latter provides DASH and Digital Rights Management (DRM) for commercial users and was

[1] http://wwwnui.akamai.com/gnet/globe/index.html, last visited 11/01/2016.
[2] http://bem.tv/, last visited 11/01/2016.
[3] https://www.peer5.com/, last visited 11/01/2016.

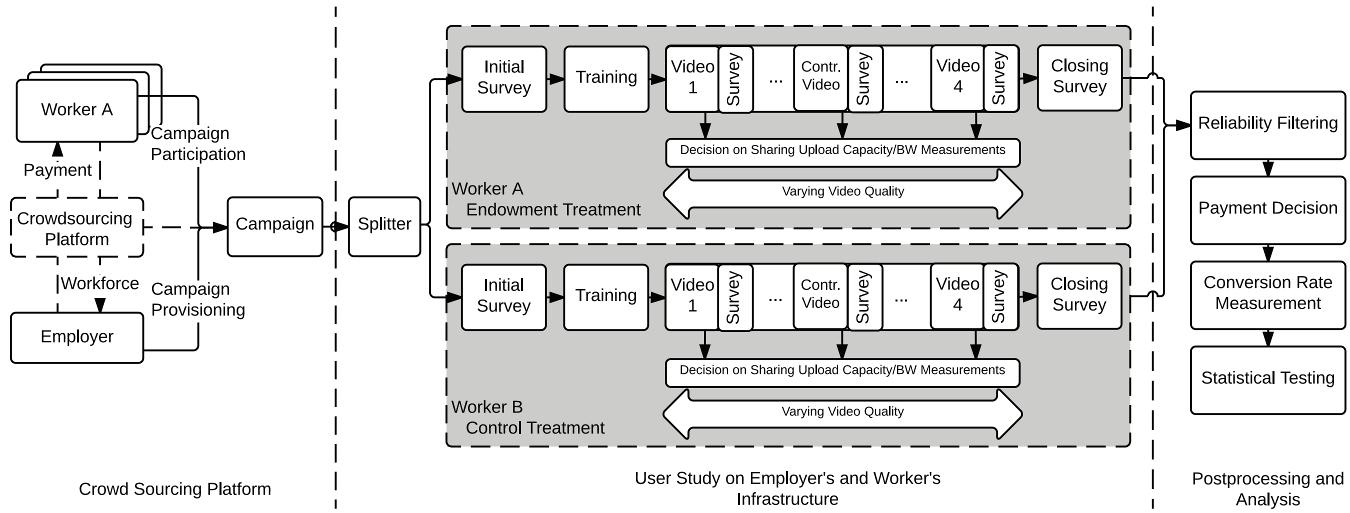


Figure 1: Overview on user study design and technical embedding.

acquired by Yahoo[4]. Moreover, different projects provide Video on Demand (VoD) streaming of content from BitTorrent swarms using WebRTC[5].

P2P Incentive Schemes for Adaptive Video Streaming: There is a number of works investigating the issue of fairness and incentive in distributed, adaptive streaming systems from a pure system centric perspective. LayerP2P is an approach proposing a bilateral stream trading scheme [20] using Scalable Video Coding (SVC) encoded video based on BitTorrent's Tit-for-Tat strategy: a peer A only sends data to a peer B for a reciprocal exchange of data from B to A (reciprocation). However, Tit-for-Tat is a poor choice for streaming. As opposed to file sharing, where each part of the file has the same weight in terms of importance, in the streaming case only the parts next to the playback deadline are important, thus limiting trading opportunities and downgrading performance [34].

Tian et al. [29] present a game theoretic approach for adaptive streaming. In this scheme, peers trade a virtual currency (tokens) for service from other peers. The higher the amount of tokens a peer can earn by providing service, the better for his own welfare as the tokens can be spent to buy service. Even though, the idea of freely tradeable tokens is compelling, a practical implementation is not simple and has to either use some block chain[6] architecture or has to be handled centrally by the CDN provider.

The works of Chu et al. [9] and Hu et al. [14] are based on taxation to reach the goal of making peers contributing proportionally to their bandwidth. However, the schemes are prone to subversion, as they are based on the assumption that users honestly report their available upload capacity.

All of the works discussed are based on the assumption of a purely rational user in the economic sense, thus denying the existence of altruistic users. Moreover, users are assumed to be aware of the technical details of Internet access. This work takes a different approach by investigating the question of incentives from a user centric perspective based on a crowd sourced experiment.

[4]http://finance.yahoo.com/news/peer5-live-streaming-cdn-leverages-160000342.html, last visited 11/01/2015.
[5]https://webtorrent.io/, last visited 11/01/2016.
[6]http://bitcoin.org/bitcoin.pdf, last visited 11/01/2015.

3. USER STUDY DESIGN

This section details the rationale of the user study design, starting with an overview and proceeding with a detailed survey and experiment design description.

3.1 Overview

An overview of the user study and technical embedding is provided in Figure 1. The acquisition of test subjects is based on workforce from a crowdsourcing platform. Subjects are hired from the Microworkers.com[7] platform. After the subject decides to participate in the campaign, he/she is referred to the Employer's HTTP Server hosting the study. In the following, the crowd workers are referred to as subjects.

As discussed initially, the user study aims at answering the four research questions regarding (a) the share of altruistic subjects, (b) whether non-altruistic subjects can be incentivized to give consent in exchange for higher QoE, (c) the utilization of behavioral economics, and (d) the general sensitivity of users with respect to a utilization of upload capacity without consent.

In order to answer question (c), one of two treatments is randomly assigned to users (Splitter). One treatment (Endowment Treatment) is utilizing an effect from behavioral economics to optimize the consent rates of users (for details, see Section 3.4). The other treatment is a Control Treatment serving as a control for comparison.

For both treatments, the subject first fills in a survey (Initial Survey) to measure the demographic characteristics of the sample. Afterwards, subjects are trained, i.e., prepared for the HIT. After the training (Training), the video experiments start. Successively, a total of four videos is shown to the user, where each video is playing with a varying visual quality ranging from a MOS value of "Bad" quality to "Good" quality as measured by the VQM. During the playback of each video, the subject can choose the option of increasing video quality for providing his or her upload capacity. If the

[7]https://microworkers.com/, last visited 11/01/2016

subject decides to accept, the quality is increased and a P2P workload fully utilizing the subject's uplink is emulated. If the subject denies the offer, the video is continued to be centrally served by the content provider but the quality is not increased.

The total of four video experiments allows to gather data for answering research questions (a) and (b). After completing the four video experiments, each user fills in a second survey (Closing Survey) in order to measure subject's sentiment to answer research question (d).

After finishing the treatments, the results of all subjects are filtered according to several reliability criteria. If the subject passes the filters, the payment of 0.4 US dollars is triggered. The filtered data is used to measure the conversion rate of subjects, i.e., the fraction of users accepting to share their upload capacity for better QoE under the different treatments.

3.2 Initial Survey

The initial survey serves three purposes: (a) collecting information on the sample characteristics, (b) identification (and exclusion) of unsuitable subjects, (c) preloading of the video content.

More precisely, we ask for age and whether the subject needs any optical aids like glasses or contact lenses. The latter serves as a criterion for excluding subjects that are likely to suffer from visual impairments. Nevertheless, the data analysis shows that the filtering criterion has no impact on the results. Moreover, we ask for habits with respect to the consumption of online video services. Subjects are asked to estimate the number of videos consumed online on average per day.

Besides measuring the demography of the sample, the time for filling in the survey is utilized to preload the video data for the experiments in the background. The subject cannot proceed with the task if the preloading is not finished. Preloading the video data is essential to control the QoE during the experiment. If video data is not preloaded, the buffer of the player may deplete due to a slow network connection. A depleting buffer leads to freezing playback of the video (stalling), which is well known to have an exponentially negative effect on QoE [13, 12, 10].

3.3 Training

The International Telecommunication Union (ITU) has defined a number of recommendations on how to conduct QoE experiments (ITU-BT.500, [2]). As the standard aims at pure QoE assessment of video streams in a controlled lab environment, the methodology has to be adapted to this work. However, we follow the standard if possible. The standard recommends to train users, since "subjects could misunderstand their task" [2]. Consequently, we include two training steps before the actual videos are shown.

Quality Adaptation Training: The first training step familiarizes users with the type of quality changes occurring during the experiment. For this purpose, an introductory text and a video player is shown to the subject. The text states that when starting the playback, a switch in video quality will happen. After the subject has started playback, a 15 second video with an initially good quality is shown. After 7 seconds the quality drops clearly perceivable to a low video quality level to give the subject a reference of a quality change.

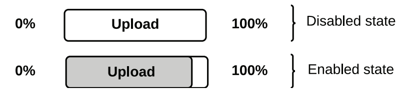


Figure 2: Gauge bars indicating whether sharing upload capacity is enabled or disabled.

After the subject has seen a quality change during playback, the step is repeated. However, this time the subject is asked to indicate when the change in quality happens by clicking a button. If the button is pressed before the change has happened, the subject is notified about his or her mistake. The number of attempts as well as the time needed to press the button are recorded for reliability filtering.

Capacity Sharing Training: The second training step is intended to introduce users to upload capacity sharing in a P2P network. As the concepts behind computer networks are non-trivial to understand for subjects without a technical background, the consequences of sharing upload capacity to the performance of the application level are explained. We explain the consequences for three frequently used applications (web browsing, email, Voice over IP (VoIP)) on desktop PCs in a text staying as non-technical as possible. Likewise, we explain the consequence of congestion on the uplink to other users sharing the same Internet access. Afterwards, subjects are asked to answer five binary questions (Yes/No) with respect to their understanding of the consequences of sharing upload capacity. Users can only proceed after all questions have been answered correctly. The survey is designed in a way such that a purely random choice of the answers has a probability of less than 4% to be correct. Moreover, users are only informed on the number of incorrect answers, but there is no indication, which answers are wrong. The attempts to answer the questions correctly are recorded, which serves as a filtering criterion for unreliable users or users with a sub-standard understanding.

Likewise, users are trained to the visualization used for displaying the upload capacity sharing status during the video experiments. Enabling upload capacity sharing is indicated by a neutral grey bar as indicated in Figure 2.

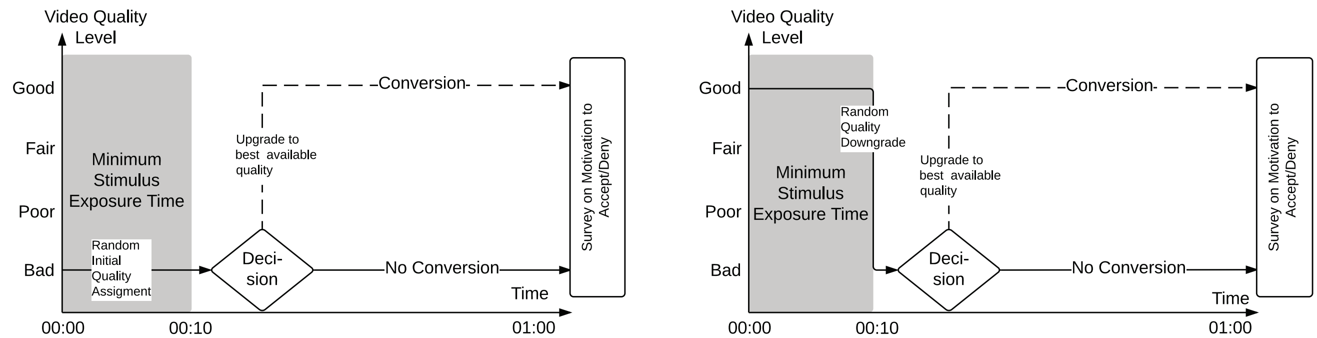
3.4 Video Experiments

After the training, the video experiments are started. The video experiments are intended to answer research question (a) regarding the altruism of users, question (b) regarding the incentivation of non-altruistic users, and question (c) regarding the utilization of effects from behavioral economics.

Questions (a) and (b) are investigated by showing videos with varying visual quality and asking for consent to share upload capacity in exchange for a better or equal visual quality. The first case allows to measure differences in conversion rates for different QoE levels for answering question (b). The latter case is used to answer the question for altruism (a), as users in this case give their consent for sharing upload capacity without any benefit.

We aim at answering question (c) by utilizing the Endowment Effect as an incentive mechanism. More precisely, we investigate whether exposing users to an Endowment Effect can incentivize the users' readiness for sharing their upload capacity in return for a higher QoE.

The term Endowment Effect was first coined by Thaler et al. [28] based on the work of [16]. It describes the bias of



(a) Control Treatment. (b) Endowment Treatment.

Figure 3: Overview of different treatments.

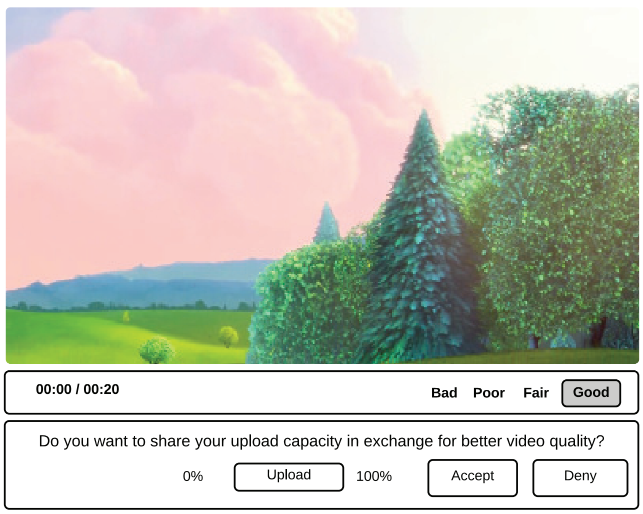


Figure 4: User interface design of the video player showing a frame from the Big Buck Bunny (BBB) test video. The design is kept deliberately simple to prevent distraction of subjects.

subjects to estimate the value of a good already possessed higher than the willingness to pay for the same good without possessing it. In other words: people value a good higher merely because the good is owned. The Endowment Effect can be explained by loss aversion, i.e., the aversion of individuals for experiencing the loss of an owned object [6]. Following this theory, we assume that giving subjects a glimpse of the achievable quality in the P2P mode before showing the lower, pure CDN quality level can be utilized to simulate a loss experience [30]. Consequently, the conversion rates should increase.

Based on these considerations, we design the Control Treatment and the Endowment Treatment as depicted in Figure 3a and Figure 3b. For the Control Treatment (Figure 3a), subjects are assigned to one of four random initial quality levels. The video starts playing immediately and the subject is exposed to the stimulus for a minimum of 10 seconds as recommended by ITU-BT.500 [2]. After the minimum stimulus exposure time, the subject is asked whether he or she wants to share his or her upload capacity in exchange for the best possible quality. The best possible quality is chosen for an upgrade because an upper limit for the effectiveness of the incentive is intended to be measured.

The Endowment Treatment (Figure 3b) works in a similar way, but subjects initially are not exposed to a random quality level, but to the best video quality level available. After the minimum stimulus exposure time, the video is randomly downgraded to a lower video quality level to simulate a loss experience. As for the Control Treatment, after the minimum stimulus exposure time, the subject is asked whether he or she wants to share his or her upload capacity in exchange for the best possible quality.

After each video, the subjects are asked to fill in a survey measuring the motivation for giving or denying consent. In particular, two questions are asked, one for the role the increase in QoE plays for the decision to accept or deny, the other for the role of bandwidth sharing for the decision to accept or deny. The combination of both allows to evaluate the motivation of subjects with respect to both factors.

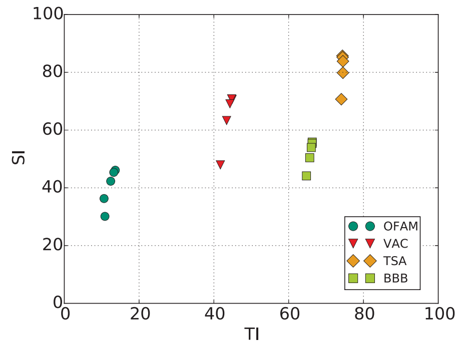
3.5 User Interface

The video experiments are embedded in a simple User Interface (UI) (see Figure 4) to avoid distraction of the subjects. For the same reason, the video is played without audio. The UI shows information on the current state of playback, i.e., the current visual quality level and timers on the current playback position and the total duration of the video. Controls for the video player are not visible to the subjects. In a separate box, the state of upload capacity utilization and the two buttons for choosing to accept or deny the offer are displayed after the minimum stimulus exposure time. The buttons are randomly switched for each video experiment to be able to tell apart altruistic subjects always choosing to accept from unreliable subjects trying to finish the experiment as fast as possible by always clicking the left or the right button.

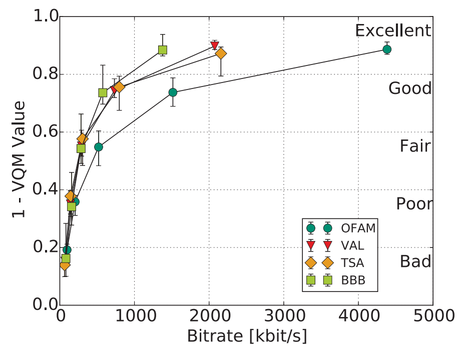
3.6 Test Video Selection and Preparation

As a test video database, Lederer et al.'s DASH test set [19] is used. Videos in the test set are chosen from the genres animation, sports, and movies. In the following, we refer to the test sequences as Valkaama (VAL), Big Buck Bunny (BBB), The Swiss Account (TSA), and Of Forest and Men (OFAM).

Following the recommendations in ITU-BT.500 [2] the Temporal Information (TI) and Spatial Information (SI) metrics of the test videos as defined in [1] are calculated before the experiment in order to ensure that the whole TI/SI



(a) SI and TI measures of video sequences. The figure contains all encoded quality versions, where a quality version closer to TI= 0 and SI= 0 denotes a lower quality due to loss of temporal and spatial information.



(b) VQM score of scenes (95% confidence intervals) in test material and MOS mapping versus encoding bit rate of all test video sequences. As the highest quality setting is prohibitively large for complete preloading in a subject's web browser, the "Excellent" level was not taken into consideration for the experiments.

Figure 5: Overview on TI/SI spectrum, quality, and encoding of test video sequences.

space is covered, as both metrics influence the perceptibility of quality differences. The results depicted in Figure 5a show that the set of chosen videos contains sequences with a high TI/SI as well as a low TI/SI value.

The videos are encoded using H.264 (libx264) with a constant rate factor encoding, i.e., the visual quality degradation is triggered by enforcing a certain quantization parameter for the encoding. In order to quantify the quality loss between the different quality versions, we measure the difference between the reference video from the test database and the encoded video. For this purpose, the video is cut into different shots using shotdetect[8]. Subsequently, the VQM metric [22, 23] is applied on a per scene basis. The constant rate encoding factor is chosen in a way such that the median VQM score over all scenes in the test sequence for each quality version falls within one of the intervals ($0 \leq 0.2 \leq 0.4 \leq 0.6 \leq 0.8 \leq 1.0$), as these intervals can be mapped linearly to the MOS scale [36]. The mapping for all four video sequence is depicted in Figure 5.

The examination of the bit rate/VQM relation reveals that the highest quality level ("Excellent") is prohibitively large to be preloaded entirely in the subject's browser. Consequently, the four levels "Bad" to "Good" are used throughout the video experiments.

3.7 Upload Traffic Generation

Once a subject accepts to share upload capacity, a P2P chunk exchange like traffic pattern is generated on the user's uplink. More precisely, as many 200 kilobyte chunks as possible are uploaded to a measurement server as long as the video is running after the subject has taken the decision to accept. With the test videos described in Section 3.6, the workload is roughly equivalent to uploading 2 second DASH segments in the highest quality used for the experiments. This depicts the worst case of a fully satisfied uplink from the subject's perspective. We visualize the utilization of the uplink in the UI as depicted in Figure 2 and Figure 4.

3.8 Closing Survey

The closing survey is presented to the subject after the last video experiment is finished. In this survey, the subject is asked general questions on the experiment from two categories. The first category contains video quality related questions. In particular, the subject is asked whether a difference in video quality could be noticed during the experiment and whether the upgrade in quality after clicking accept satisfied the user.

The second category comprises questions related to upload capacity. The subject is asked to rate the value of upload capacity and the acceptability of a utilization of upload capacity without asking for consent. This particular question is intended to answer research question (d). Moreover, subjects are asked for noticeable side effects of upload capacity sharing (e.g., slower network connection, etc.).

Additionally, one question on the subject's age is added to the closing survey. The question was already asked in the initial survey and is used for reliability filtering. However, in the closing survey, the scale of age intervals to select from is flipped.

4. EVALUATION

This section describes the reliability filters applied to the raw data and the demographic properties of the sample before and after filtering. Moreover, the results with respect to differences between the Control Treatment and the Endowment Treatment as well as differences in conversion rates with respect to the impact of QoE as an incentive are discussed. Finally, the insights gained on user motivation based on the survey data are described.

4.1 Reliability Filtering

Reliability filtering is a crucial aspect of crowd working studies, as crowd working platforms attract unreliable subjects. The authors of [13] define four filtering methods to detect inconsistent behavior of subjects, all of which are applied in this work.

Consistency Tests (CT): This test requires subjects to answer slightly varied questions multiple times. As an example, subjects are asked for their age twice, once in the initial survey and once in the closing survey, but in the latter case a flipped age scale is used. Consequently, subjects providing inconsistent answers can be filtered.

The probability to pass all consistency checks in the user study by chance is below 3%. As depicted in Figure 6, the CT filtering group filters 4% of the subjects (see Figure 6).

[8]http://johmathe.name/shotdetect.html, last visited 11/01/2016

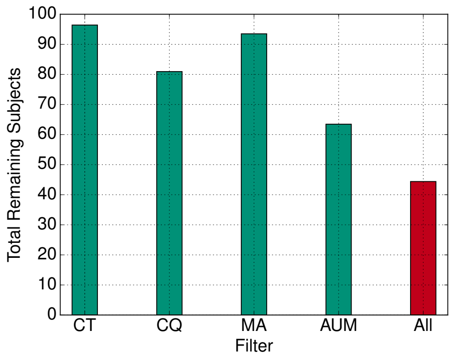


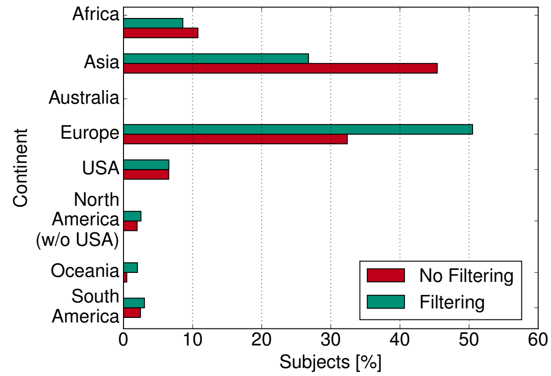
Figure 6: Fraction of remaining subjects after applying each single filtering step (CT, CQ, MA, AUM) and all (All) filtering steps at once.

Content Questions (CQ): For this test, users are asked to answer questions relating to the content of a text or a video seen during the study. The study design contains two different types of content questions. First, the upload capacity training step is based on answering questions regarding a text. The text explains the implications of sharing upload capacity on the application level. Subjects needing multiple attempts to answer all questions correctly are filtered. This is done to ensure that subjects understand the implications of upload capacity sharing and can take the implications into account during the user study. The probability to pass this step by chance is below 4%. Second, after each video a user watches during the study, the subjects are asked to select a description of the content of the video from a list of descriptions (e.g., *Person climbing rocks*). Workers not passing all of the content questions are filtered. The probability of passing this filter by chance for all videos is below 1%. The CQ filter filters 19% of the subjects (see Figure 6).

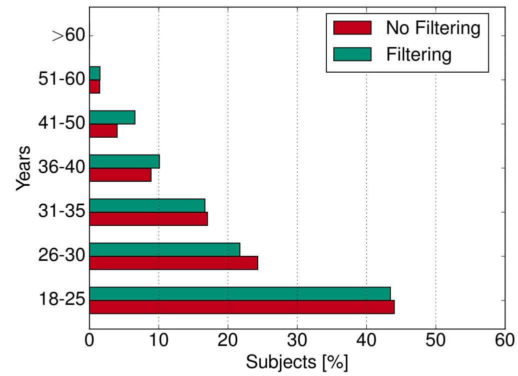
Mixed Answerss (MA): This type of test aims to identify users following fixed click patterns (e.g., always the left button). For the video experiments described in Section 3, we flip the position of the accept and deny buttons between the experiments. Otherwise, subjects always clicking on the left or right button would deliver very consistent answers and are not distinguishable from subjects always accepting or subjects always denying. Consequently, we filter all workers providing an accept/deny/accept/deny or deny/accept/deny/accept pattern for the four video experiments. The MA filter filters 6% of the subjects (see Figure 6).

Application Usage Monitoring (AUM): This type of filter aims at identifying unreliable subjects based on their application usage. In particular, workers completing the survey in an unreasonable short time frame and workers not having the browser tab on top of their desktop all the time during the study are filtered. Moreover, mobile users, and users having a screen size smaller than 800x600 pixels are filtered as video QoE changes are perceived differently on mobile devices and small screens [21]. The AUM filter has the highest impact of all filters, excluding 36% of the subjects.

All filtering steps applied at the same time disqualify 56% of the subjects as depicted in Figure 6. Notably, this share is not unusual. As an example, the authors of [13] filter 75% of the subjects.



(a) Regional distribution of subjects before and after filtering.



(b) Distribution of subjects' age before and after filtering.

Figure 7: Summary of sample characteristics.

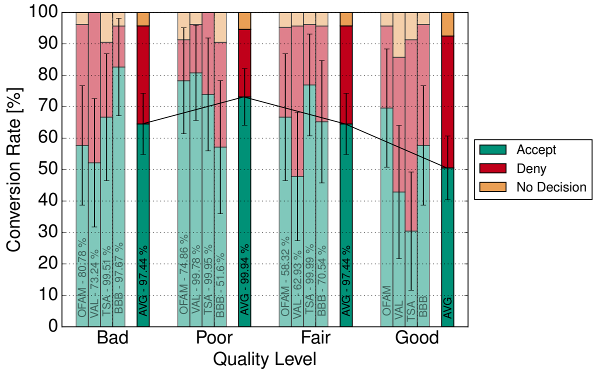
4.2 Sample Characteristics

In the following, the sample is characterized according to demographic factors. We aim at a young, international sample for two reasons: first, users of online video services are predominantly young [4]. Second, streaming platforms like YouTube have an international audience. Figure 7 summarizes the most important characteristics before and after reliability filtering.

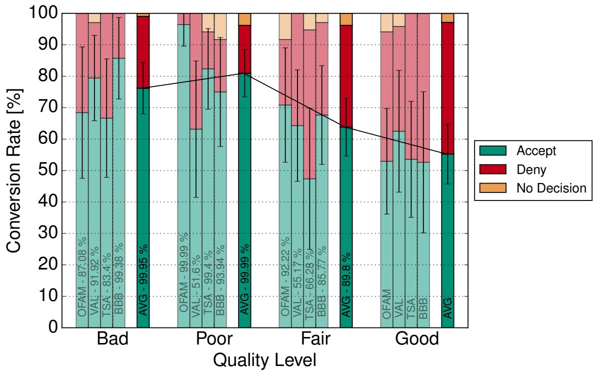
With respect to the regional distribution of the subjects after filtering, Europeans represent 33% and Asians represent 30% of the sample (see Figure 7a). Thus, 63% of the sample are from one of these two regions, while the remaining sample is represented by subjects from Africa, the Americas, Oceania and Australia. Notably, the reliability filtering process removes 20% of Asian subjects, which may be caused by click farming[9]. In total, 363 subjects participate of which 192 remain after filtering.

The median age of the sample is between 26 and 30 (see Figure 7b), i.e., the sample is dominated by young subjects. Filtering does not change this characteristic considerably. Probably correlated to the young age of the sample is the affinity to Internet technology. More than 50% of the subjects agree or strongly agree to be affine to Internet technology. The median consumption of online video streams is as high as 6 to 10 consumed videos per day.

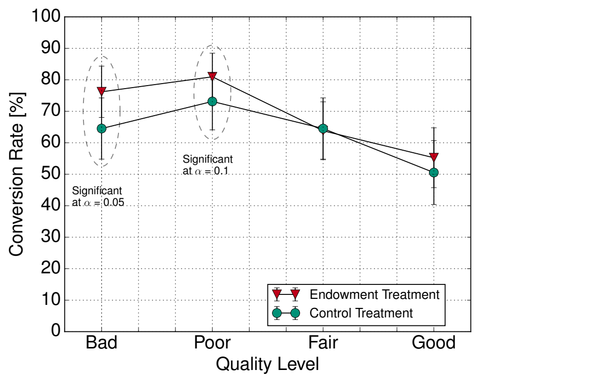
[9]http://www.theguardian.com/technology/2013/aug/02/click-farms-appearance-online-popularity, 11/01/2016



(a) Control Treatment group conversion rate results.



(b) Endowment Treatment group conversion rate results



(c) Control vs. Endowment Treatment. Subjects have an up to 10% higher probability to convert for the Endowment Treatment compared to the Control Treatment.

Figure 8: Conversion rate for all four test video sequence for the Control and Endowment Treatment group. The numbers represent the confidence that the measured result is different from the rightmost reference case according to a t-Test. The error bars represent the 95% confidence interval. The AVG bar represents the average rate over all videos.

4.3 Endowment vs. Control Treatment

Figure 8 shows a summary of the conversion rates, i.e., the fraction of users accepting upload capacity sharing under the two treatments. Figure 8a and Figure 8b show the results for the two treatments separately and for each test video. The numbers on each of the bars represent the confidence that the measured result is different from the rightmost case

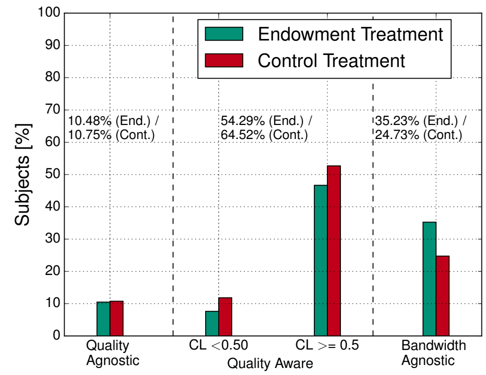


Figure 9: Fraction of Quality Agnostic, Quality Aware and Bandwidth Agnostic subjects. The group of Quality Aware subjects is subdivided into subjects with a weakly consistent set of answers (CL<0.5) and a highly consistent set of answers (CL$\geq$ 0.5) according to the consistency criterion defined in the appendix.

where no increase in quality was offered. If the confidence is larger than 95%, the result is significantly different at a significance level $\alpha = 0.05$. If the confidence is lower than 95%, there is no significant difference or the difference was too small to be significant for the sample size. The statistical evaluation is based on a two-sided t-Test.

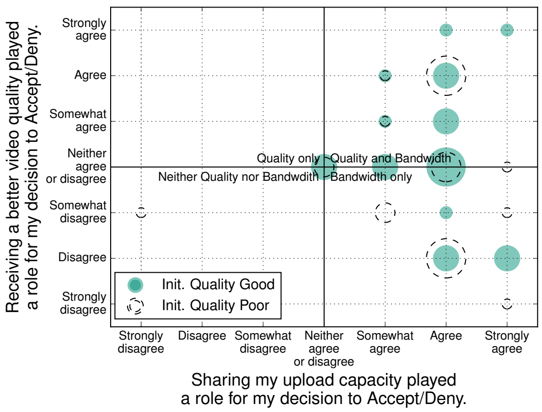
Notably, for both treatments (Figures 8a and 8b), a comparably large share of roughly 50% of the subjects acts altruistically and chooses to share upstream capacity even if no increase in quality is offered. The motivation for this behavior is discussed in Section 4.4. Moreover, both treatments show that subject's willingness to accept peaks at the "Poor" quality level, which is however not significantly different from the "Bad" quality level. As expected, the willingness to accept the offer decreases with increasing video quality. From the "Poor" to the "Good" quality level, around 25% of acceptance rate is lost for both treatments.

Figure 8c compares both treatments. For lower quality levels, the Endowment Treatment shows a significantly higher conversion rate. For the "Bad" quality setting, subjects have a 12% higher probability to accept sharing upload capacity compared to the Control Treatment, while for the "Poor" quality level a significant difference of 9% could be found at $\alpha = 0.1$. For all other quality versions, no significant difference could be found from the Control Treatment. Consequently, showing the subject the achievable increase in quality first and simulating a loss experience by downgrading to a lower quality can be utilized to optimize the conversion rate.

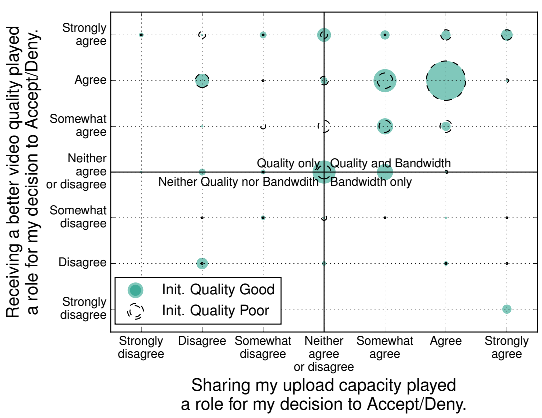
Summing up, there is a share of around 50% of the subjects sharing their upload capacity in an altruistic manner. An additional share of 25% can be incentivized by offering an increase in video quality. Moreover, utilizing the Endowment Effect for the lower quality layers is feasible and increases conversion rates by up to 12%. All confidence values for all videos and the comparison of the Endowment and the Control Treatment can be found in the appendix.

4.4 User Motivation Evaluation

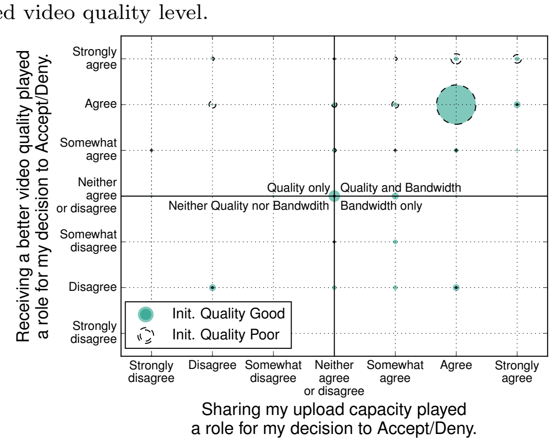
Based on the results presented beforehand three categories of subjects are identified: *Quality Agnostic*, *Quality Aware*,



(a) Motivation of Quality Agnostic subjects to deny. The role of quality is answered highly diverse regardless of the offered quality level. However, upload capacity plays a large role.



(b) Motivation of Quality Aware subjects to accept. Quality and upload capacity play a role in the decision depending on the offered video quality level.



(c) Motivation of Bandwidth Agnostic subjects to accept.

Figure 10: Motivation charts for Altruistic, Quality Agnostic and Altruistic subjects with respect to the role of increased video quality and upload capacity sharing on the decision.

and *Bandwidth Agnostic* subjects. Quality Agnostic subjects always deny to share upload capacity, Quality Aware subjects accept or deny depending on the offered quality, Bandwidth Agnostic subjects always accept to share upload capacity regardless of the quality enhancement offered. Figure 9 depicts the share of all three groups as indicated by the numbers between the dashed lines.

Additionally, Figure 9 subdivides the Quality Aware group of users into two classes according to the consistency of their decisions. For that purpose, a consistency criterion is used, which is formally defined in the appendix. More precisely, the consistency criterion assumes subjects to act rationally by picking a certain quality level satisfying their quality needs and denying upload capacity sharing for all quality levels above this quality level. The Consistency Level (CL) of the decision measures the deviation from this behavior, where a value close to 1 denotes consistent behavior and a value close to 0 denotes inconsistent behavior. For Quality Aware subjects, the majority of subjects takes a decision with a consistency level larger or equal to 0.5, thus the majority of the subjects in this group take a rationally justifiable decision on upload capacity sharing.

The group of Quality Agnostic subjects represents 10/11% of the sample, the group of Quality Aware subjects is as large as 54/65% and the Bandwidth Agnostic group is found to be as high as 11/14% of the sample.

For each of the decisions to be taken throughout the user study, subjects are asked for their motivation to accept or deny an offer using two statements:

(a) Receiving a higher video quality played a role for my decision to accept/deny.
(b) Sharing my upload capacity played a role for my decision to accept/deny.

Subjects indicate their agreement with each of the statements on a Likert scale ranging from “strongly agree” to “strongly disagree”. For all groups, the motivation chart is depicted in Figure 10. The size of the circles indicates the number of answers received for a certain combination of Likert values for both statements.

Figure 10a compares the motivation for Quality Agnostic subjects to deny for two given initial quality versions (“Poor” and “Good”). Notably, this group is highly decisive on the role of upload capacity, but highly indecisive regarding the role of quality increase regardless of the quality level offered initially. This behavior indicates that the drawbacks of sharing upload capacity is a dominant factor for this group.

For Quality Aware subjects both, upload capacity and video quality play a role for the decision (see Figure 10b). Interestingly, the motivation of Bandwidth Agnostic subjects is even more decisive for quality and bandwidth playing a role in the decision than Quality Aware subjects. This result allows for two conclusions: either Bandwidth Agnostic subjects are always dissatisfied with the quality offered or they have no understanding or no valuation of upload capacity. As the majority of subjects in both categories is satisfied with the video quality after clicking accept, the latter explanation seems to be more likely.

In the closing survey, subjects are asked for their agreement with the statement: *My upload capacity can be shared without my consent.* The resulting answers for the three groups are depicted in Figure 11. As expected, for the majority of Bandwidth Agnostic users a sharing of upload capacity without consent is slightly acceptable. As opposed to that, for Quality Aware and Quality Agnostic users, upload capacity sharing without consent is unacceptable. However, the latter two groups make up for more than two thirds of the subjects in total. Thus, asking users for their consent to

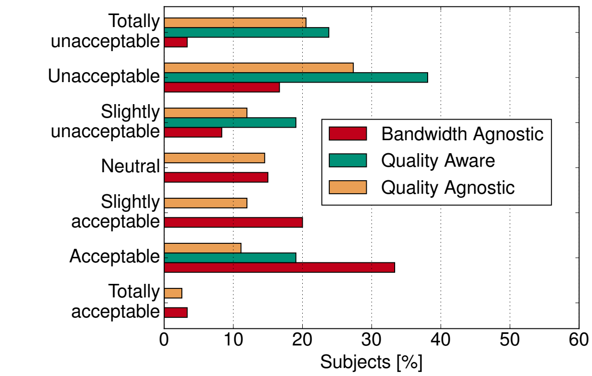


Figure 11: Distribution of agreement with the statement: *My upload capacity can be shared without my consent.*

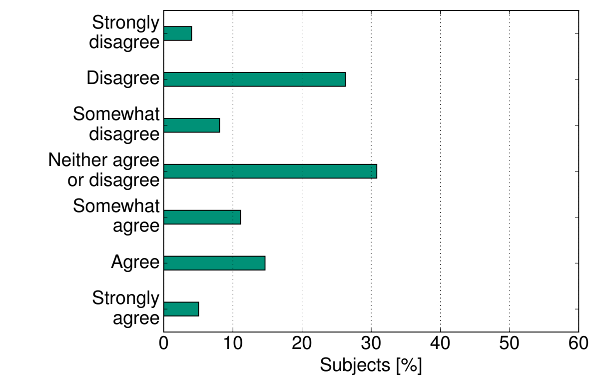


Figure 12: Distribution of agreement with the statement: *I noticed side effects of sharing my upload capacity (e.g., a slower Internet connection).*

utilize upstream capacity is advisable in any case to prevent an unsatisfied user base.

Moreover, we were interested whether subjects perceive any drawbacks of upload capacity sharing. Therefore subjects were asked for their agreement with the statement *I noticed side effects of sharing my upload capacity (e.g., a slower Internet connection)*. In particular, the answers to this question showed to be diverse with a median in the neutral statement. This is an indicator that the cost for upload capacity sharing perceived by subjects is mostly a psychological factor.

5. CONCLUSIONS AND OUTLOOK

This work is motivated by the steep increase of video traffic volume and the spreading of inter browser communication frameworks like WebRTC. The latter offers possibilities for CDNs and content providers to utilize a CDN/P2P hybrid architecture for content distribution based on pure, website-embedded JavaScript. As a result, a large share of the backend capacity can be offloaded to end users [35].

However, utilizing user's upload capacity raises the question of user's consent to do so. From a purely rational, economic perspective users should demand compensation as they are buying upstream bandwidth from their Internet Service Provider (ISP). In particular, four research questions are defined and answered by this work.

(a) How high is the fraction of altruistic users giving consent without further benefits?
(b) How high is the fraction of non-altruistic users that can be convinced to give their consent in exchange for a better QoE?
(c) Can incentive mechanisms based on findings from behavioral economics be utilized to increase consent?
(d) How sensitive are users to a utilization of upload capacity without consent?

For answering these questions, a video streaming study was designed. During the study, subjects are exposed to different quality versions of video streams. While streaming, the subjects are offered an increase in visual quality in exchange for giving consent to providing their upload capacity. The study was conducted utilizing a crowd sourcing platform. 363 subjects participated; after reliability filtering, 192 subjects remained for a deeper analysis.

Regarding question (a), three types of behavior were identified: a share of around 30% of the subjects always accepts to share upload capacity regardless of the type of proposed quality increase and even if no increase of quality is offered at all. This group is clearly bandwidth agnostic. Moreover, a smaller share of 10% always denies to share upload capacity. The majority of subjects decides on upload capacity sharing based on the offered quality increase.

A closer investigation of the motivation of the subjects reveals that bandwidth agnostic subjects do not seem to have a good understanding of the technical implications of sharing upload capacity or no valuation of upstream capacity, even though this was part of a training step conducted before showing the video stream.

As an answer to question (b), setting an incentive by offering a higher video quality was found to significantly increase the rate of consent by up to 25% if an optimal initial quality setting is chosen before upgrading.

Regarding question (c), an Endowment effect was found: if the upgraded version of the stream is shown for a short period of time before downgrading quality and placing the offer to increase quality again, a significantly higher share of up to 12% of subjects is willing to accept the offer. However, this statement only holds if the subject receives a low quality version of the stream on denial. Content providers can utilize this effect, but should be aware that a very aggressive form of incentive may dissatisfy users.

With respect to question (d), the study captured subject's opinion on the acceptability of sharing upload capacity without an explicit consent of the user. For the majority of users, utilizing the upload capacity without consent is slightly to totally unacceptable.

The data gathered in this study allows for recommendations how incentive mechanisms for WebRTC based distributed streaming websites should be designed. First, the content distributor should seek for user's explicit consent before utilizing the upstream. Otherwise, a majority of users will remain unsatisfied and may spread bad word of mouth. Second, video quality can be used as an incentive to incentivize users to give consent. The content distributor should learn, which users belong to which of the aforementioned user groups and the available upstream capacity of users. This can be done by creating user profiles based on cookies or other tracking methods. Based on this information, an offer to increase video quality in exchange for upload capacity sharing should only be made to quality aware subjects with a high upstream capacity. All other users should be asked for consent without offering an increase in quality. Third, behavioral patterns like the Endowment effect can be utilized to increase consent rates. However, this only works for very low initial quality on denial and does not make a difference for higher quality levels.

While the trends identified in this study are stable for the sample, they are a snapshot raising the question for long-term stability. In particular, users might get accustomed to the incentive scheme, if offers are made frequently on streaming websites. As a future research direction, the long-term stability can be investigated by frequently repeating the study with the same subjects.

6. ACKNOWLEDGMENT

This work has been supported in parts by the European Union (FP7/#317846, SmartenIT and FP7/#318398, eCousin) and the German Research Foundation (DFG) within the Collaborative Research Center (CRC) 1053 MAKI.

7. REFERENCES

[1] ITU-T Recommendation P.910: Subjective Video Quality Assessment Methods for Multimedia Applications. Tech. rep., International Telecommunication Union, 2008.

[2] ITU-R Recommendation BT.500: Methodology for the Subjective Assessment of the Quality of Television Pictures. Tech. rep., International Telecommunication Union, 2009.

[3] Global Internet Phenomena Report. Tech. rep., Sandvine Inc., 2014.

[4] 9th Digital Democracy Survey. Tech. rep., Deloitte, 2015.

[5] WebRTC 1.0: Real-time Communication Between Browsers (Draft 06). Tech. rep., World Wide Web Consortium, 2015.

[6] Chan, E. Y. Endowment Effect for Hedonic but not Utilitarian Goods. *International Journal of Research in Marketing 32*, 2 (2015), 439–441.

[7] Chen, K.-T., Chang, S.-J., Wu, C.-C., Chang, Y.-C., and Lei, C.-L. Quadrant of Euphoria: A Crowdsourcing Platform for QoE Assessment. *IEEE Network 4*, 24 (2010), 28–35.

[8] Chikkerur, S., Sundaram, V., Reisslein, M., and Karam, L. J. Objective Video Quality Assessment Methods: A Classification, Review, and Performance Comparison. *IEEE Transaction on Broadcasting 57*, 2 (2011), 165–182.

[9] Chu, Y.-h., Chuang, J., and Zhang, H. A Case for Taxation in Peer-to-Peer Streaming Broadcast. In *ACM SIGCOMM Workshop on Practice and Theory of Incentives in Networked Systems (PINS)* (2004).

[10] Dobrian, F., Sekar, V., Awan, A., Stoica, I., Joseph, D. A., Ganjam, A., Zhan, J., and Zhang, H. Understanding the Impact of Video Quality on User Engagement. *SIGCOMM Computer Communication Review 41*, 4 (2011), 362–373.

[11] Hardin, G. The Tragedy of the Commons. *Science 162*, 3859 (1968), 1243–1248.

[12] Hossfeld, T., Keimel, C., Hirth, M., Gardlo, B., Habigt, J., Diepold, K., and Tran-Gia, P. Best Practices for QoE Crowdtesting: QoE Assessment With Crowdsourcing. *IEEE Transactions on Multimedia 16*, 2 (2014), 541–558.

[13] Hossfeld, T., Seufert, M., Hirth, M., Zinner, T., Tran-Gia, P., and Schatz, R. Quantification of YouTube QoE via Crowdsourcing. In *IEEE International Symposium on Multimedia (ISM)* (2011).

[14] Hu, H., Guo, Y., and Liu, Y. Peer-to-Peer Streaming of Layered Video: Efficiency, Fairness and Incentive. *IEEE Transactions on Circuits and Systems for Video Technology 21* (2011), 1013–1026.

[15] Huang, C., Wang, A., Li, J., and Ross, K. W. Understanding Hybrid CDN-P2P: Why Limelight Needs its Own Red Swoosh. In *ACM International Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV)* (2008).

[16] Kahneman, D., and Tversky, A. Prospect Theory: An Analysis of Decision under Risk. *Econometrica: Journal of the Econometric Society* (1979), 263–291.

[17] Keimel, C., Habigt, J., and Diepold, K. Challenges in Crowd-Based Video Quality Assessment. In *IEEE International Workshop on Quality of Multimedia Experience (QoMEX)* (2012).

[18] Keimel, C., Habigt, J., Horch, C., and Diepold, K. QualityCrowd - A Framework for Crowd-Based Quality Evaluation. In *IEEE Picture Coding Symposium (PCS)* (2012).

[19] Lederer, S., Müller, C., and Timmerer, C. Dynamic Adaptive Streaming over HTTP Dataset. In *ACM Multimedia Systems Conference (MMSys)* (2012).

[20] Liu, Z., Shen, Y., Ross, K. W., Panwar, S. S., and Wang, Y. LayerP2P: Using Layered Video Chunks in P2P Live Streaming. *IEEE Transactions on Multimedia 11*, 7 (2009), 1340–1352.

[21] Moorthy, A., Choi, L. K., Bovik, A., and de Veciana, G. Video Quality Assessment on Mobile Devices: Subjective, Behavioral and Objective Studies. *IEEE Journal of Selected Topics in Signal Processing 6*, 6 (2012), 652–671.

[22] Pinson, M. H., and Wolf, S. A New Standardized Method for Objectively Measuring Video Quality. *IEEE Transaction on Broadcasting 50*, 3 (2004), 312–322.

[23] Pinson, M. H., and Wolf, S. Application of the NTIA General Video Quality Metric (VQM) to HDTV Quality Monitoring. In *Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM)* (2007).

[24] Rhinow, F., Veloso, P. P., Puyelo, C., Barrett, S., and Nuallain, E. O. P2P Live Video Streaming in WebRTC. In *IEEE World Congress on Computer Applications and Information Systems (WCCAIS)* (2014).

[25] Roverso, R., Reale, R., El-Ansary, S., and Haridi, S. SmoothCache 2.0: CDN-Quality Adaptive HTTP Live Streaming on Peer-to-Peer Overlays. In *ACM Multimedia Systems Conference (MMSys)* (2015).

[26] Sackl, A., Zwickl, P., and Reichl, P. The Trouble with Choice: An Empirical Study to Investigate the Influence of Charging Strategies and Content Selection on QoE. *Internet Charging and QoS Technologies (ICQT)* (2013).

[27] Seufert, M., Egger, S., Slanina, M., Zinner, T., Hossfeld, T., and Tran-Gia, P. A Survey on Quality of Experience of HTTP Adaptive Streaming. *IEEE Communications Surveys & Tutorials 17* (2015), 469–492.

[28] Thaler, R. Toward a Positive Theory of Consumer Choice. *Journal of Economic Behavior & Organization 1*, 1 (1980), 39–60.

[29] Tian, G., Xu, Y., Liu, Y., and Ross, K. Mechanism Design for Dynamic P2P Streaming. In *IEEE Conference on Peer-to-Peer Computing (P2P)* (2013).

[30] Tversky, A., and Kahnemann, D. Loss Aversion in Riskless Choice: A Reference-Dependent Model. *Quarterly Journal of Economics 106*, 4 (1991), 1039–1061.

[31] Vogt, C., Werner, M. J., and Schmidt, T. C. Leveraging WebRTC for P2P Content Distribution in Web Browsers. In *IEEE International Conference on Network Protocols (ICNP)* (2013).

[32] Wichtlhuber, M., Wicklein, G., Wilk, S., Effelsberg, W., and Hausheer, D. RT-VQM: Real-Time Video Quality Assessment for Adaptive Video Streaming Using GPUs. *ACM Multimedia Systems Conference (MMSys)* (2016).

[33] Yin, H., Liu, X., Zhan, T., Sekar, V., Qiu, F., Lin, C., Zhang, H., and Li, B. Design and Deployment of a Hybrid CDN-P2P System For Live Video Streaming: Experiences with LiveSky. In *ACM International Conference on Multimedia (MM)* (2009).

[34] Zhang, X., and Hassanein, H. A Survey of Peer-to-Peer Live Video Streaming Schemes: An Algorithmic Perspective. *Elsevier Computer Networks 56*, 15 (2012), 3548 – 3579.

[35] Zhao, M., Aditya, P., Chen, A., Lin, Y., Haeberlen, A., Druschel, P., Maggs, B., Wishon, B., and Ponec, M. Peer-Assisted Content Distribution in Akamai Netsession. In *ACM Internet Measurement Conference (IMC)* (2013).

[36] Zinner, T., Hohlfeld, O., Abboud, O., and Hossfeld, T. Impact of Frame Rate and Resolution on Objective QoE Metrics. *IEEE International Workshop on Quality of Multimedia Experience (QoMEX)* (2010).

APPENDIX

A Confidence Tables

Tables 1-3 list the results of confidence calculations for all experiments performed during the user study.

B Consistency Criterion Definition

We define the consistency level of a subject's decisions as the deviation from a predefined assumption of rational behavior. More precisely, let $V = (B, P, F, G)$ be the subject's decision, where B (Bad) $<$ P (Poor) $<$ F (Fair) $<$ G (Good) define the quality levels for which the offer on sharing upload capacity was proposed. Moreover, let B, P, F, G be 1 if the subject denied to share upload capacity for the given quality level and 0 otherwise.

An entirely rational tuple V of decisions is given, if the subject has selected a single quality level above which all offers were denied, because the quality was good enough to satisfy the subject compared to the cost of sharing upload capacity.

More formally, an entirely consistent decision is given if

$$CL(V) = \max\left(\frac{B+P+F+G}{4}, \frac{P+F+G}{3}, \frac{F+G}{2}, G\right) = 1. \quad (1)$$

The closer $CL(V)$ is to 1, the more consistent is the decision made by the subject. Notably, the decision to always accept yields a consistency level of 0, while the decision to always deny yields a consistency level of 1. However, these two cases are treated separately in this work and are thus excluded from the definition, i.e., $B + P + F + G \neq 4$ and $B + P + F + G \neq 0$.

Table 1: Confidence of difference to the "Good" to "Good" case calculated using a two-sided t-test for the Control Treatment group. Bold numbers satisfy $\alpha = 0.05$.

Control Treatment			
Video	**Bad→Good**	**Poor→Good**	**Fair→Good**
OFAM	80.78%	74.86%	58.32%
VAL	73.24%	**99.78%**	62.93%
TSA	**99.51%**	**99.95%**	**99.99%**
BBB	**97.67%**	51.60%	70.54%
AVG	**97.44%**	**99.94%**	**97.44%**

Table 2: Confidence of difference to the "Good" to "Good" case calculated using a two-sided t-test for the Endowment Treatment group. Bold numbers satisfy $\alpha = 0.05$.

Endowment Treatment			
Video	**Bad→Good**	**Poor→Good**	**Fair→Good**
OFAM	87.08%	**99.99%**	92.22%
VAL	91.92%	51.60%	55.17%
TSA	83.40%	**99.40%**	66.28%
BBB	**99.38%**	93.94%	85.77%
AVG	**99.95%**	**99.99%**	89.80%

Table 3: Confidence for difference between Endowment and Control Treatment calculated using a two-sided t-test. Bold numbers satisfy $\alpha = 0.05$.

Endowment vs. Control Treatment				
Video	**Bad**	**Poor**	**Fair**	**Good**
OFAM	77.04%	**97.50%**	61.79%	90.15%
VAL	**98.54%**	90.32%	88.30%	90.99%
TSA	50.00%	77.34%	**98.17%**	**95.73%**
BBB	61.79%	89.97%	57.53%	63.31%
AVG	**96.41%**	90.49%	53.98%	74.54%