

Video Quality Measurements in the Crowd

A Case Study from Germany

Werner Robitza^{1,2}, Alexander Dethof¹, Steve Göring², Alexander Raake², André Beyer³, Tim Polzehl³

¹ AVEQ GmbH, Vienna, Austria

² Ilmenau University of Technology, Ilmenau, Germany

³ Crowdee GmbH, Berlin, Germany



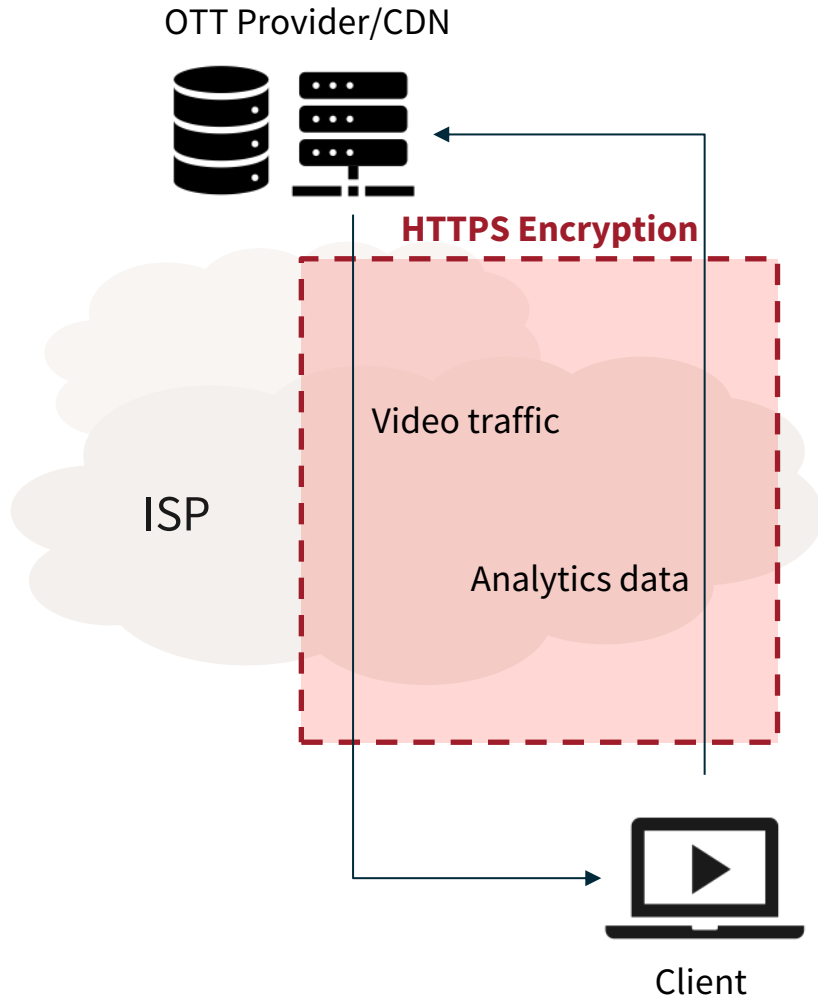
Overview

- We conducted a **crowdsourcing study** on Over the Top (OTT) video streaming (YouTube, Netflix, Amazon Prime Video)
- Data was collected throughout 2019, overall: ~400,000 video playbacks from ~2,000 users across 5 large, national ISPs
- Source: Desktop PCs and laptops, dedicated web browser extension (Chrome & Firefox)
- Goals:
 - Collection of usage behavior
 - Independent comparison of ISP and OTT quality



Population density in Germany (2017),
demografie-portal.de

Motivation – Why Crowdsourcing?



- OTTs have a large amount of customer-specific **user behavior and quality-related data**, but data is inaccessible to operators and regulators
- Challenge: How to obtain representative and valid data on the performance/quality of video streaming over different ISP networks?
- Laboratory and stationary probes cannot give the full picture
- **Crowdsourcing is a viable solution for obtaining such information** that regulators/ISPs cannot get otherwise

What did we measure?

Context Data

- Internet Service Provider via IP address and manual surveys
- Geolocation via IP address and browser API
- Device information via browser API



Bandwidth Data

- Speed test results from speedtest.net
- Upload, download speeds
- Latency (ping times)



Video Data

- Active, user-initiated measurements
- Passive, user-initiated measurements (all watched videos are tracked)



Key Performance Indicators & Quality of Experience according to ITU-T Rec. P.1203

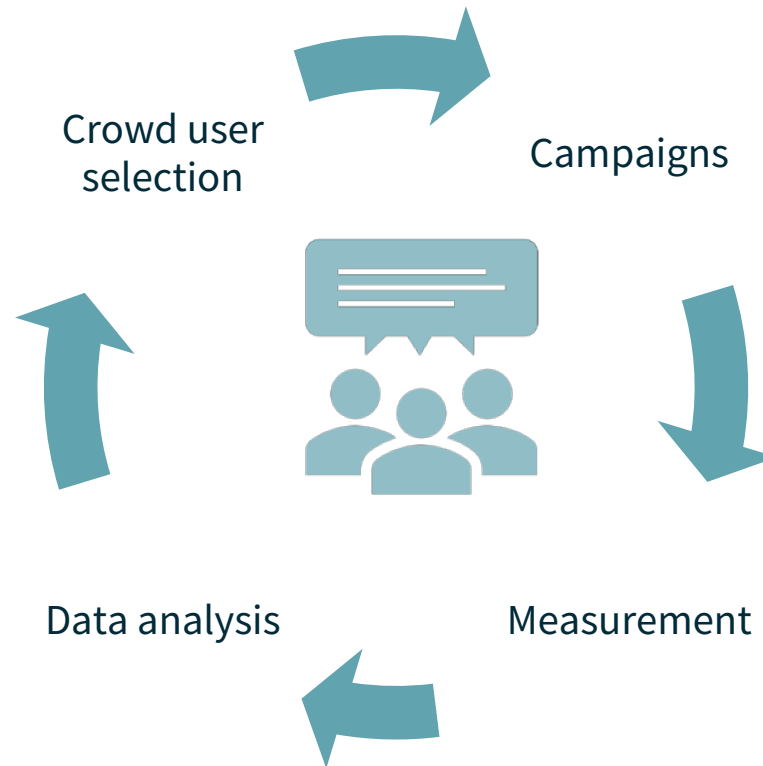
Measurement Campaigns

1 Selection of crowd users from a nationwide pool of 135,000

Based on ISP and previous knowledge about user profile

4 Filtering, enrichment, and data analysis using context data and KPIs/QoE

Verification of correct crowd tasks



2 Users can participate in crowd jobs to voluntarily install measurement software and run tests

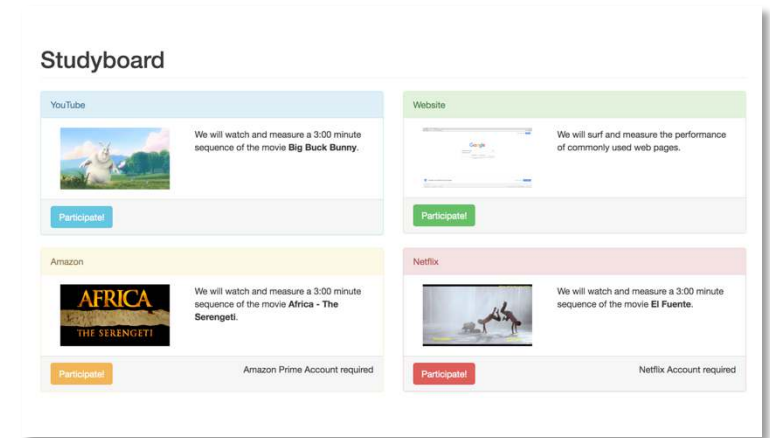
Target: 35 users per ISP per week, weekly tasks

3 The web browser extension performs active and passive video quality measurements

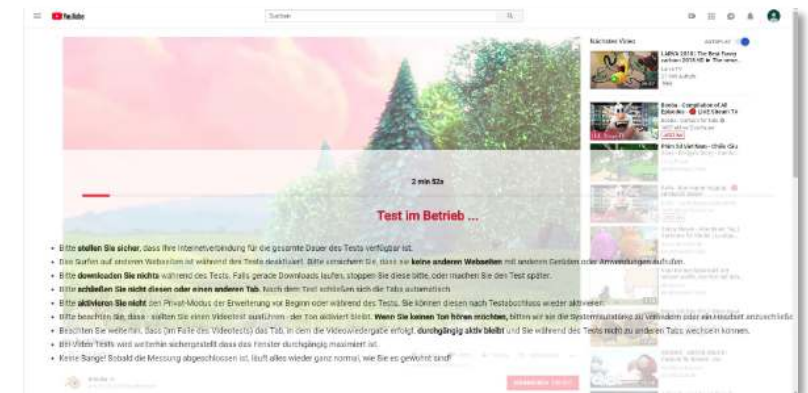
Users get remunerated

Measurement Software

- Measurements are run with dedicated web browser extension, developed by AVEQ
- One measurement run contains:
 - Conduction of speed test on speedtest.net
 - YouTube video (1.5 min)
 - Amazon Prime video (1.5 min, trailer)
 - Netflix video (1.5 min, trailer)
- Completely automatic active measurement without user involvement
- Blocking of disturbing factors during the measurement (e.g., parallel downloads, background traffic)

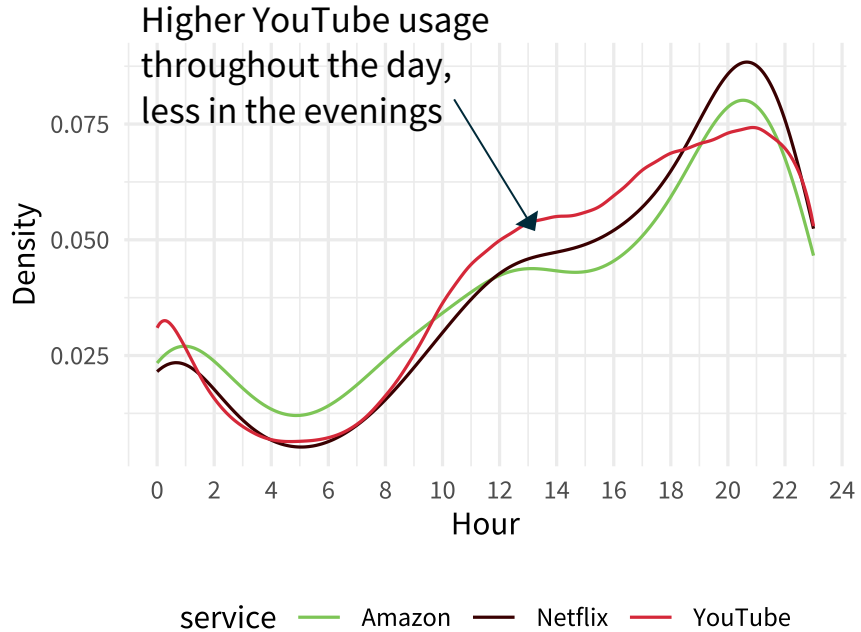


Example of active measurements that can be selected by the user.

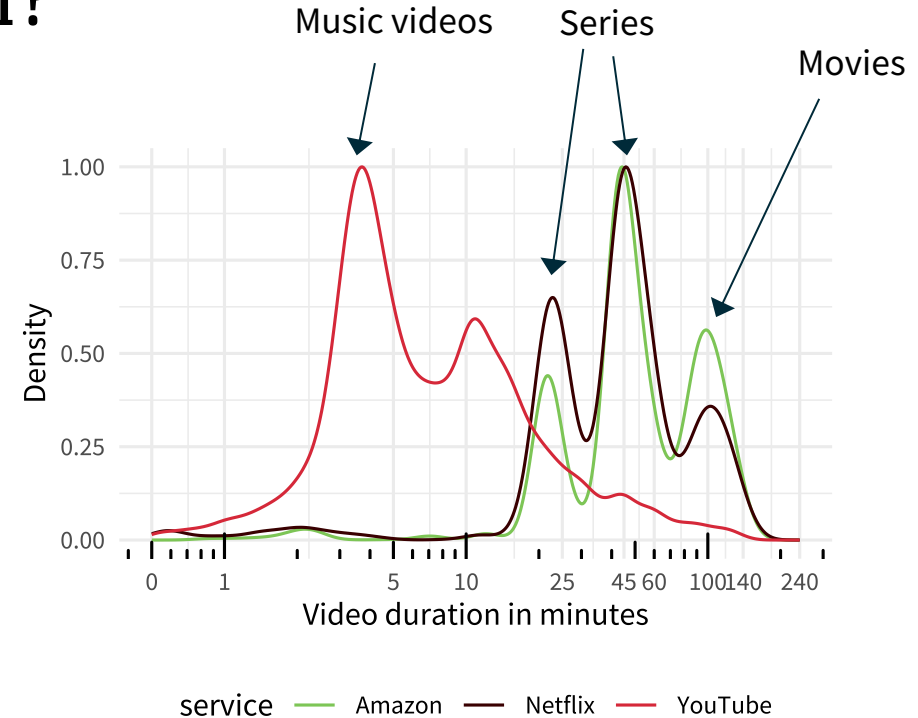


Active video test in progress, with instructions.

When and what do users watch?

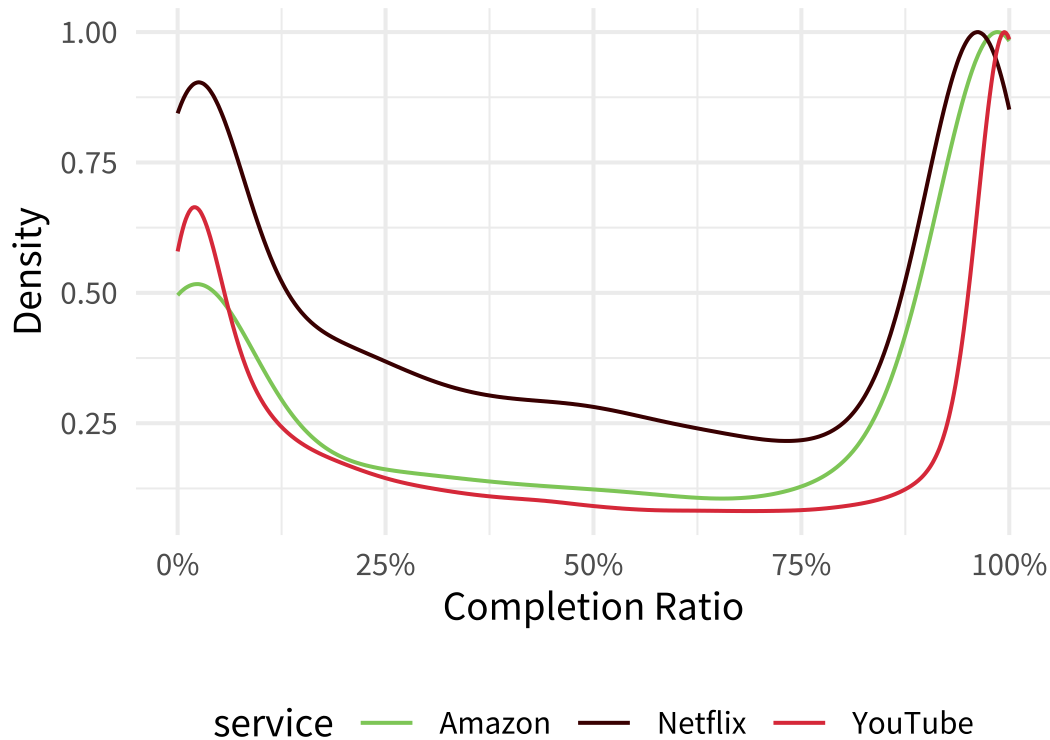


- Users watch mostly in the evenings
- General peaks between 18–22h



- Strong difference in video duration between YouTube and NFLX/AIV
- Different classes of content: Short videos (incl. music, 5–10 min), Series (25, 45 min), Movies (100 min)

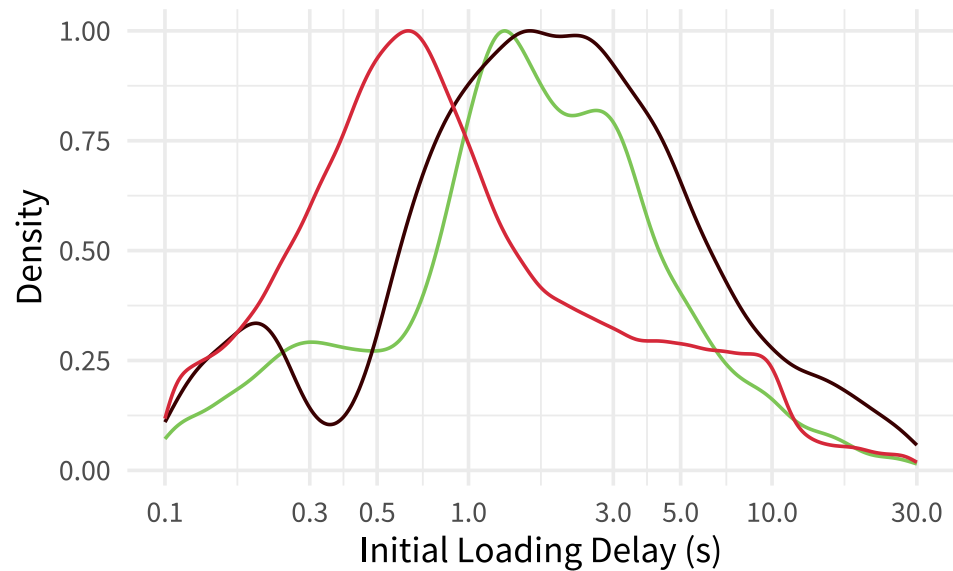
User Engagement and Video Completion



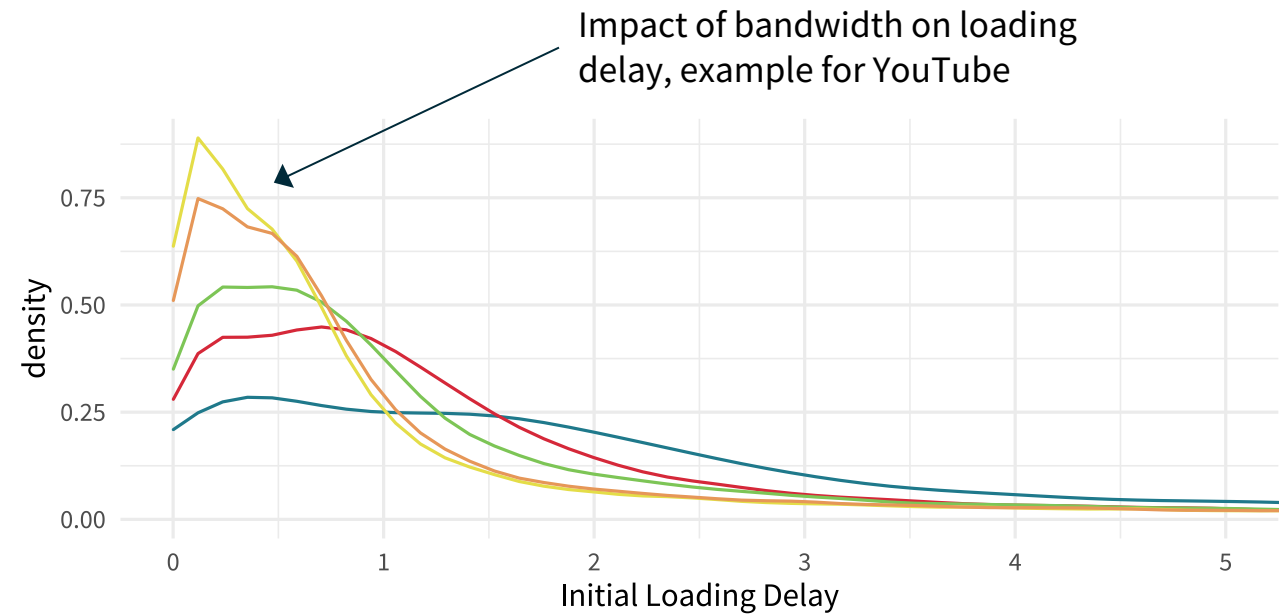
- Completion Ratio: How many % of a video are completed?
- Videos are either:
 - quit in the beginning
 - watched (almost) fully
- Evidence that completion depends on the length of the content
- Future research: reasons for low completion ratio. First indications: Stalling is a determining factor.

Streaming performance

- Initial loading differs between YouTube and Amazon/Netflix
- Bandwidth has a strong influence; ≤ 8 Mbit/s causes noticeable delays (> 1 s)



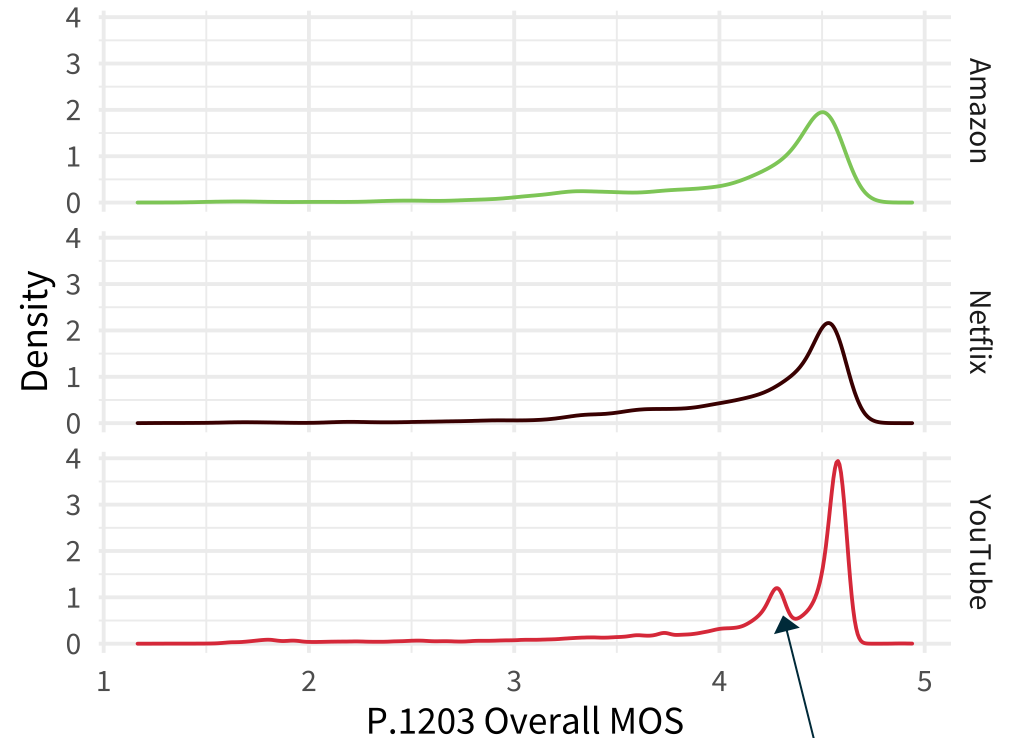
Service — Amazon — Netflix — YouTube



speed_group — 0-4 — 4-8 — 8-16 — 16-30 — 30-50

Quality of Experience

- Collection of input data according to ITU-T Rec. P.1203, allows calculating Mean Opinion Score (MOS)
- Caveats:
 - Only first 5 minutes can be calculated for each video
 - Direct comparison between OTTs is not possible due to characteristics of used video model (P.1203.1 Mode 0)
- General observations:
 - Streaming quality is very good with few issues, stalling only occurs in a few percent of sessions
 - Bandwidths above 8–16 Mbit/s provide already sufficient quality for streaming (HD)



Advantages and Challenges of the Approach

Advantages:

- It is possible to get an in-depth and unbiased look at user behavior, network/streaming performance and experienced quality
- Unique combination of active, foreground/user-initiated measurements and passive usage data
- Highly useful data for:
 - Independent ISP benchmarking
 - Identification of service outages, finding weak spots in bandwidth provisioning
 - Knowing “who the customer is”; transparency reports to customers

Challenges:

- Strong incentivization is needed, e.g. remuneration, vouchers, data about own network performance, gamification
- Users need to be motivated to participate and provide passive measurement data
- Heterogeneous field: measurements may be skewed by a low number of heavy users (→ statistical filtering/aggregation required)

A large crowd of people is gathered at a festival, with a thick cloud of colorful powder (yellow, blue, pink, green) filling the air. Many people are holding up their phones to take photos or videos. The sky is bright blue with scattered white clouds. The overall atmosphere is festive and celebratory.

Thank you!