FORTH Institute of Computer Science

UNIVERSITY OF CRETE

Quality of Experience for Internet Services

Main focus on Video Streaming

Maria Papadopouli
Professor
University of Crete
Foundation for Research & Technology – Hellas (FORTH)
http://www.ics.forth.gr/mobile

Growth opportunities for services & new markets

Trends that cause this growth:

- mobile device capacity growth
- advances in networks
- **cloud services**
- user-generated content
- mobile services
- **virtualization**

Larger, more complex & heterogeneous markets are formed

- Service oriented technologies
- **Complex service systems** modeling & simulation
- **Service quality of experience**
- Business aspects of **service composition**
- People in services
- Service innovation management

Dramatic growth of mobile data, streaming services, telepresence

Networks experience periods of severe impairments (PSIs) due to various reasons, such as **contention, handovers, channel impairment, outages, congestion**

- The network degradations affect user engagement & satisfaction
- For **churn prevention**, providers need to understand QoE
- To better adapt & monetize a service, QoE prediction is needed

Quality of Experience (QoE)

Definition: The **degree of delight or annoyance** of a person whose experiencing involves an application, service, or system. It results from the evaluation of the user **fulfilment of the expectations and needs** with respect to the utility and/or enjoyment in the light of the user **context, personality and current state**.

Brunnström,et al. “Qualinet White Paper on Definitions of Quality of Experience” 2013.

User-centric & Contextual Aspects

Affected by techno-socio-economic-cultural-psychological aspects, e.g.:

- Preferences QoE & Price
- Willingness-to-pay, intrinsic indicators towards a service provider (e.g., band name, perceived value, reliability)
- Content (e.g., richness, diversity, searching mechanisms)
- Integration with popular services (e.g., social networking apps)

Human	Context	System	QoE
User's mood	Physical context	Navigation	Effectiveness
Motivation	Social context	Content reliability	Trust
Needs	*Distraction*	Page loading time	Aesthetics
Gender	Mobility	CPU resources	Usability
Knowledge	Crowd	***Quality of Information***	Quality of information
Expectations	*Noise*	Loading strategy	Loading time
Prior Experiences	Task	*Battery lifetime*	Pleasure
Visual and auditory Acuity	Security	*Screen size*	Acceptability
Attention Level	Temporal Context	Content type	Satisfaction
Educational background	Environment	Jitter	Efficiency
Emotions	Memory effect	Color	
Personality traits	Price	Easy of use	
Age		Element order	
Culture		***Bandwidth***	
		Reliability	
		Screen brightness	
		Screen Resolution	

Important aspects in the user Quality of Experience analysis

User engagement

Can be characterized by their usage pattern:

- time spent using their service (total or session duration),
- “revisits” to that service (e.g., session inter-arrival),
- use of features (application/service dependent)
- number of downloads, clicks, pauses, FF/RW
- type of abandonments,
- number of active processes running in parallel
- user attention type of metrics (e.g., eye-trackers)

1. C. Moldovan and F. Metzger, "Bridging the Gap between QoE and User Engagement in HTTP Video Streaming"28th International Teletraffic Congress (ITC 28), 2016.
2. M. Seufert et al., "Unsupervised QoE field study for mobile YouTube video streaming with YoMoApp" IEEE 9th International Conference on Quality of Multimedia Experience (QoMEX), 2017.
3. S. Wassermann et al., "QoE in Cellular Networks through in-Smartphone Measurements"12th IFIP Wireless and Mobile Networking Conference (WMNC), 2019.
4. S. Krishnan and R. Sitaraman, "Video stream quality impacts viewer behavior: inferring causality using quasi-experimental designs" IEEE/ACM Transactions on Networking, 2013.

Technical

- Smarter systems
- Improved network performance
- Rapidly roll-out new services
- Reduce maintenance
- Improved user satisfaction
- Real-time QoE inference
- Early-warning
- Real-time statistics & trends
- …

Business

- Assess agreements
- Pricing
- Advertising & marketing
- Booking
- Optimize value-generating mobility & wireless services
- Monetize mobility & wireless service sessions
- …

Vision

- **Smarter Commerce**
- **Analytics & Big Data**
- **Smarter Network Management**
- **Support of augmented reality & virtual reality services**

Convenience → **Time** → **Customer Satisfaction** → **Money**

2.2.2. Consumer empowerment: boosting consumer choice and protection

All stakeholders in general support a broader focus on more empowerment of end users. Some of them believe that well informed consumers with a choice of suppliers will enable a more dynamic and responsive market to the benefit of consumers and industry. In that aspect all contributors pay great attention to the transparency especially in the context of bundled offers and net neutrality measures. In that respect several of the stakeholders think that BEREC should further work on avoiding the unjustified traffic management practices, stressing that measures taken out of commercial motivations might lead to discriminatory practices with a direct negative effect towards the consumers.

Despite the general support for strengthening the consumer protection, some of the stakeholders point out that consumer protection measures should complement and not supersede the legal framework for competition. One of the stakeholders is of the opinion that BEREC should not adopt decisions in the field of privacy and data protection in order not to cause confusion and legal uncertainty.

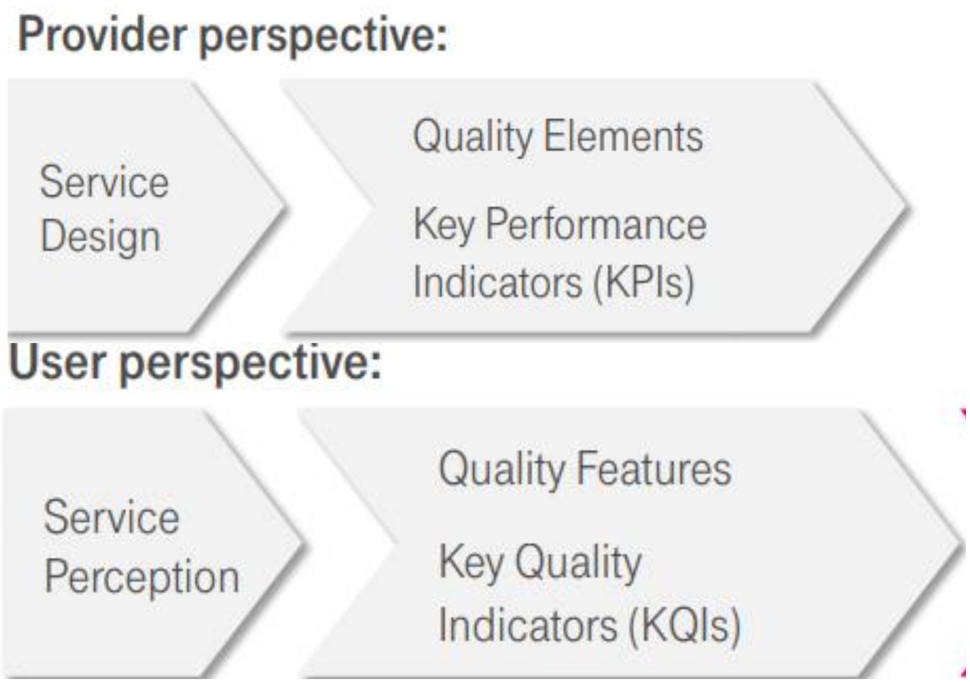
2.2.3. Service related developments

In its draft Strategy BEREC has envisaged undertaking additional work in the field of international roaming, net neutrality, special rate and/or cross-border services, mainly through developing common concepts, gathering and analysing data and will focus on the elaboration of better methodologies to ensure comparability of data with a view to ensuring better and monitoring.

In addition the stakeholders propose BEREC to undertake additional measures, as follows:

- In the field of international roaming - to work more in order to guarantee development and growing of competitive alternatives to mobile international roaming;
- In the fields of net neutrality and transparency, including quality of service - to envisage Pan-European transparency measures related to network performance (including disclose traffic management information and the quality of Internet access);
- In the field of cross-border services – dedicating more efforts to facilitating their provision, including through dissemination of the best practices existing in that field.

QoS vs. QoE



QoS

QoE

QoE Metrics

- **Network level**
  - packet loss
  - buffer state
  - throughput
  - jitter
  - retransmission
  - SNR

- **Application level**
  - startup delay
  - rebufferings
  - resolution changes
  - advertisements
  - skips
  - downloaded %
  - screen mode
  - location
  - context

- **Chunk level**

(for encrypted traffic due to HTTPs)
  - start time
  - time to first byte
  - download time
  - slack time
  - chunk duration
  - chunk size

How do **users perceive the network**?

Network QoS

Throughput, delay, packet loss, number of resource units,…

Application QoS

Startup delay, rebuffering events, bitrate changes, video resolution, termination status, …

User engagement

Fast forward, rewind, pause, abandonment, watching dur. %, revisit …

Keyboard activity, applications running, …

Subjectivity & **techno-socio-economic-psychological** aspects

Willingness-to-pay, preference on QoS vs. price, intrinsic indicators towards a provider (e.g., popularity among peers, reliability)

Categories of Measurement Studies

- **"Out in the wild" participatory studies**
  - **Diverse conditions**
  - Relatively small costs, larger scales
  - Limited contextual knowledge
- **Controlled field studies**
  - **Homogenous fixed** conditions
  - **Large cost** & overhead, relatively small scale
  - Often not reaching representative conditions

- **Subjective** feedback vs. **objective** measurements
  Subjectivity vs. Objectivity vs. Reliability
  Intrusiveness vs. Privacy
- **Explicit feedback vs. Inferring** QoE/user engagement from measurements

Latency Requirements (2/2)

≤ 1ms

- **Remote control / telepresence** with real-time, synchronous **haptic** feedback
- **Industrial moving robots**
- **Industrial closed loop** control systems (e.g. 1ms cycles of polling data from sensors + actuators)
- **Negotiated automatic cooperative-driving** manoeuvres
- **Smart grid**: synchronous co-phasing of power suppliers (< 1ms)

≤ 10ms

- **Shared Haptic Virtual Environments**: several users perform tasks that require fine-motor skills
- **Tele-medical** applications (e.g. tele-diagnosis, tele-rehabilitation)
- **Augmented reality**
- **Education: Haptic overlay** trainer / learner for fine motor skills (e.g. for medical)
- **Smart grid** (3ms)
- **Process automation** (5ms)

≤ 50ms

- **Serious gaming** (20ms)
- **Cognitive assistance** (20-40ms)
- **Virtual reality**
- **Cooperative driving** (20ms)
- **UAV control** (10 - 50ms)
- **Remote robot control** with haptic feedback (25ms)

≤ 100ms

- **Vehicle safety apps** (mutual awareness of vehicles for warning/alerting)
- **Assisted driving** (cars make cooperative decisions, but driver stays in control)

From: Simone Mangiante, Through the Fog Workshop, Feb. 2017

Classification based on the Time of QoE estimation

- **Real-time estimation (while the session takes place)**
  - LSTMs (to model system dynamics)
  - Chunk based approach (per second activity from the IP header)
  - Deep learning approach (per second aggregated activity)

- **Post-data collection estimation (after the completion of the sessions)**
  - Analysis of the session characteristics (e.g., rebufferings, startup delay, bitrate changes)
  - Use of statistical tests to find interactions of impairments
  - Use of ML techniques to predict the QoE or the user engagement of session characteristics

1. N. Eswara *et al.*, "Streaming Video QoE Modeling and Prediction: A Long Short-Term Memory Approach", IEEE Transactions on Circuits and Systems for Video Technology, 2019.
2. C. Gutterman *et al.*, "Requet: real-time QoE detection for encrypted YouTube traffic"10th ACM Multimedia Systems Conference (MMSys '19).
3. S. Wassermann *et al.*, "QoE in Cellular Networks through in-Smartphone Measurements", 12th IFIP Wireless & Mobile Networking Conference (WMNC), 2019.

- **Explicit QoE**: through opinion scores
  - After or while watching a video, the participants are asked to rate the video
  - Absolute Category Ratings (ACR)

- **Implicit QoE**: inferred by the QoS information
  - No exact knowledge about what the user liked or not
  - Use the available network- & application-level information & models of QoS → QoE

1. C. Moldovan and F. Metzger, "Bridging the Gap between QoE and User Engagement in HTTP Video Streaming", In 28th International Teletraffic Congress (ITC 28), 2016.
2. V. Menkovski *et al*., "Predicting quality of experience in multimedia streaming.", In Proceedings of the 7th International Conference on Advances in Mobile Computing and Multimedia. ACM, 2009.

Absolute Category Rating (ACR)

- A test method used in quality tests
- The levels of the scale are, sorted by quality:
  - Excellent (5)
  - Good (4)
  - Fair (3)
  - Poor (2)
  - Bad (1)
- A single test condition (generally an image or a video sequence) is presented to the viewers ***once only***. They should then give a quality rating on an ACR scale.

  Test conditions should be presented in a **random order per test person**.
- Mean opinion score: the average numeric score over all experiment participants, for each test condition that was shown
- Used for telephony voice quality to give a mean opinion score.

Web Browsing

1. Impact of network performance on **waiting times** (resulting from page/element loading times)

   - Partial download ratio
   - Abandonment
   - Session length

2. HCI & user experience studies addressed user preferences & experiences to perceived usability & aesthetics of Web-content

3. **Usefulness of content and adequacy of information**
   - Value (relevancy & clearness)
   - Reliability (accuracy, dependability, & consistency)
   - Currency (timeliness and continuous update)
   - Accuracy (degree to which the system information is free of error)
   - Extent of completeness of information, since Web sites need to provide information to facilitate user's content understanding

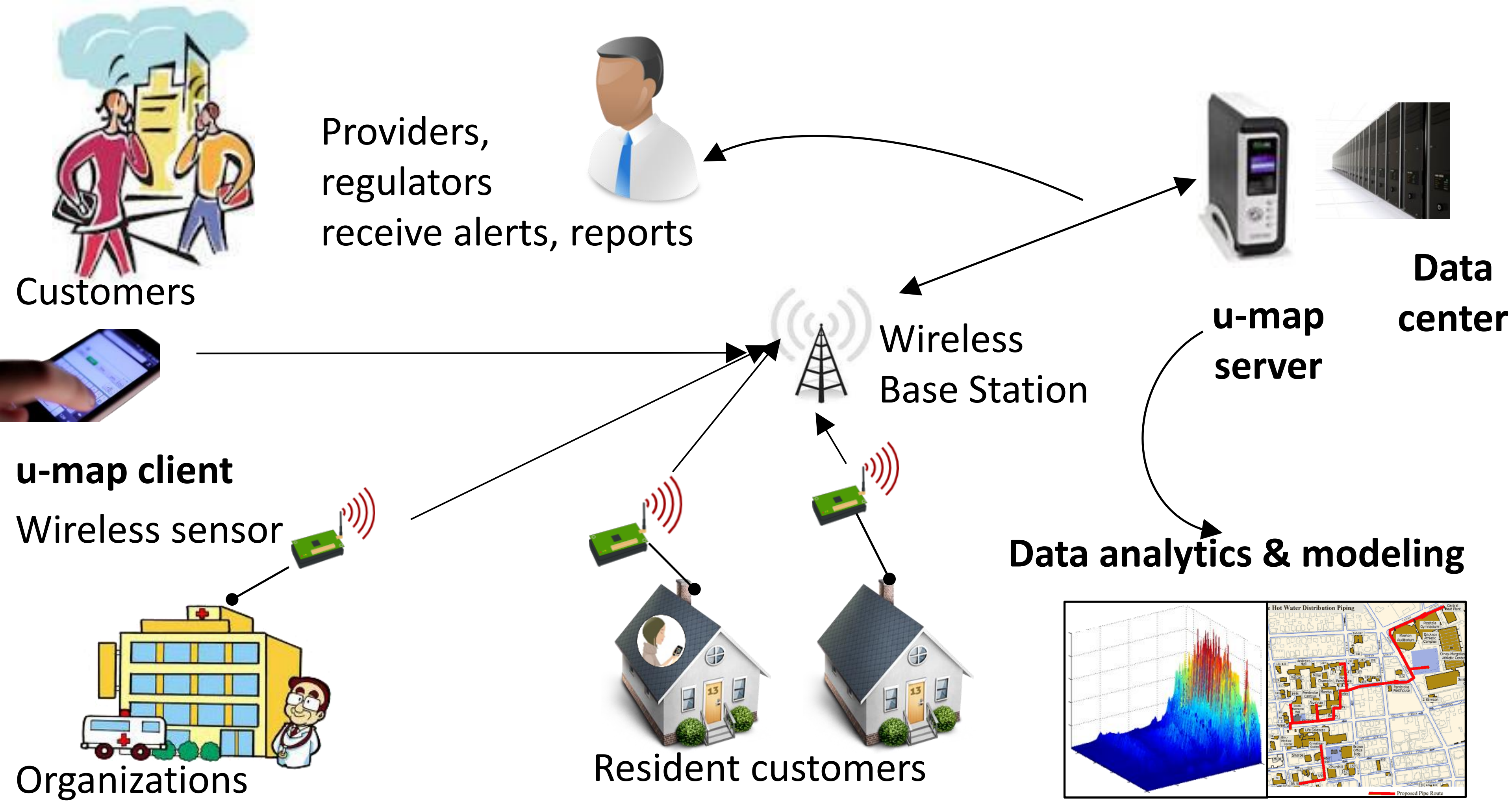
2D & 3D Video Streaming

- 2D video: fixed viewing direction
- 3D video brings immersive experience & stereoscopic vision
  - Users can change the position of the viewport, bringing an interactive experience
  - The high resolution & extremely large bit rate requirements have prevented their wide spread

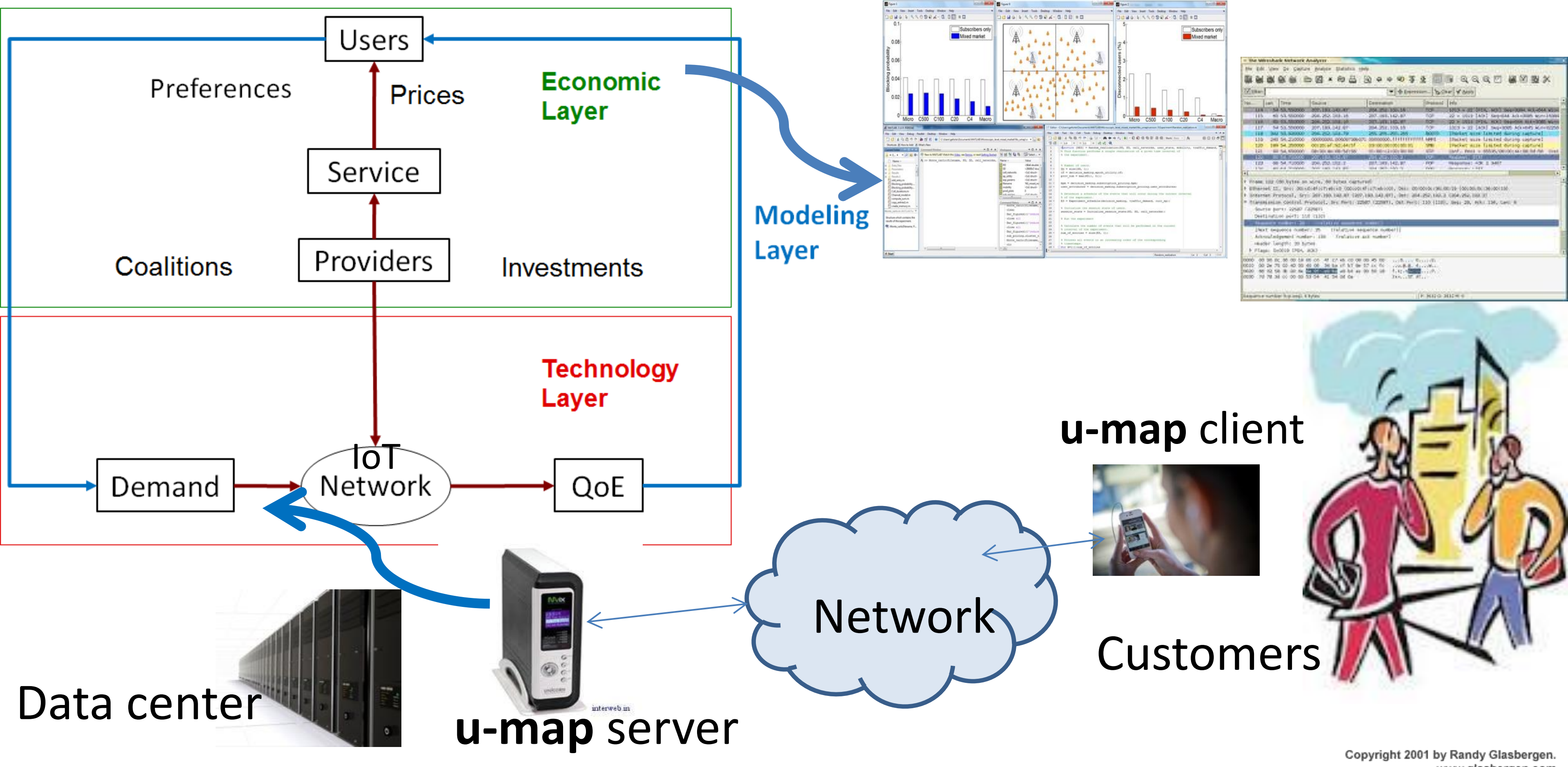
QoE-aware 3D Video Streaming via Deep Reinforcement Learning in Software Defined Networking Enabled Mobile Edge Computing. Zhou et al. IEEE Trans. on Network Science & Engineering, November 2020

Different Models for QoE-driven Measurement Analysis

- **Signal-based models** (or media-based models) work on the levels of pixels and samples only
  - They assume full access to data and decoding capabilities
- **Hybrid models** combine signal information & bitstream-level information (e.g., packet headers)
- **Parametric models** operate on transmitted packet-level or bitstream-level information
- **Models with access to extra information** (e.g., decoded video frames instead of just packet headers) provide a more accurate estimation of the quality, but in practice, the amount of information accessible is influenced by several **extrinsic factors beyond control of the ISP.**

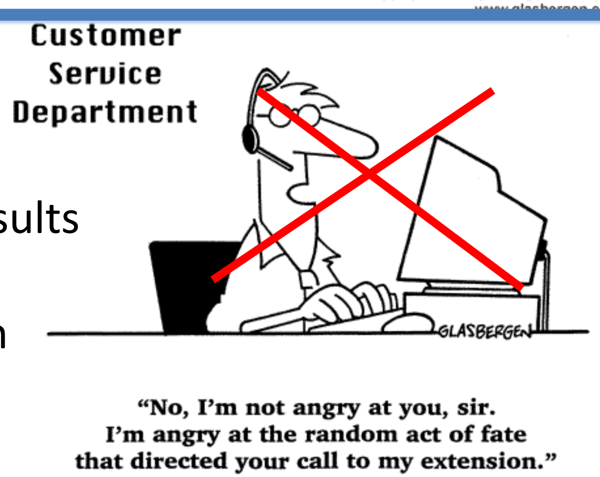


1. Monitor infrastructure & upload **objective measurements** on server
2. Customers upload their feedback using u-map clients on smartphones
3. Analysis of **objective measurements & subjective feedback**
4. **Recommendations**, alerts, and reports are sent to all stakeholders



No extra cost for monitoring

Reduces overhead, administrative support & results in **faster responses, better resource management,** with **lower cost**



u-map: user-centric QoE geo-database for recommendations & feedback

Client-to-Server architecture

- **u-map clients** on smart-phones
  - Collect network **measurements, opinion scores, customer feedback** and store them locally
  - Upload traces to u-map server
  - Query u-map server
- **u-map server**
  - Collects traces & stores them in spatio-temporal geo-DB
  - Responds to queries sent by users/customers & service/content providers

☞ Can be designed according to different **business** models:

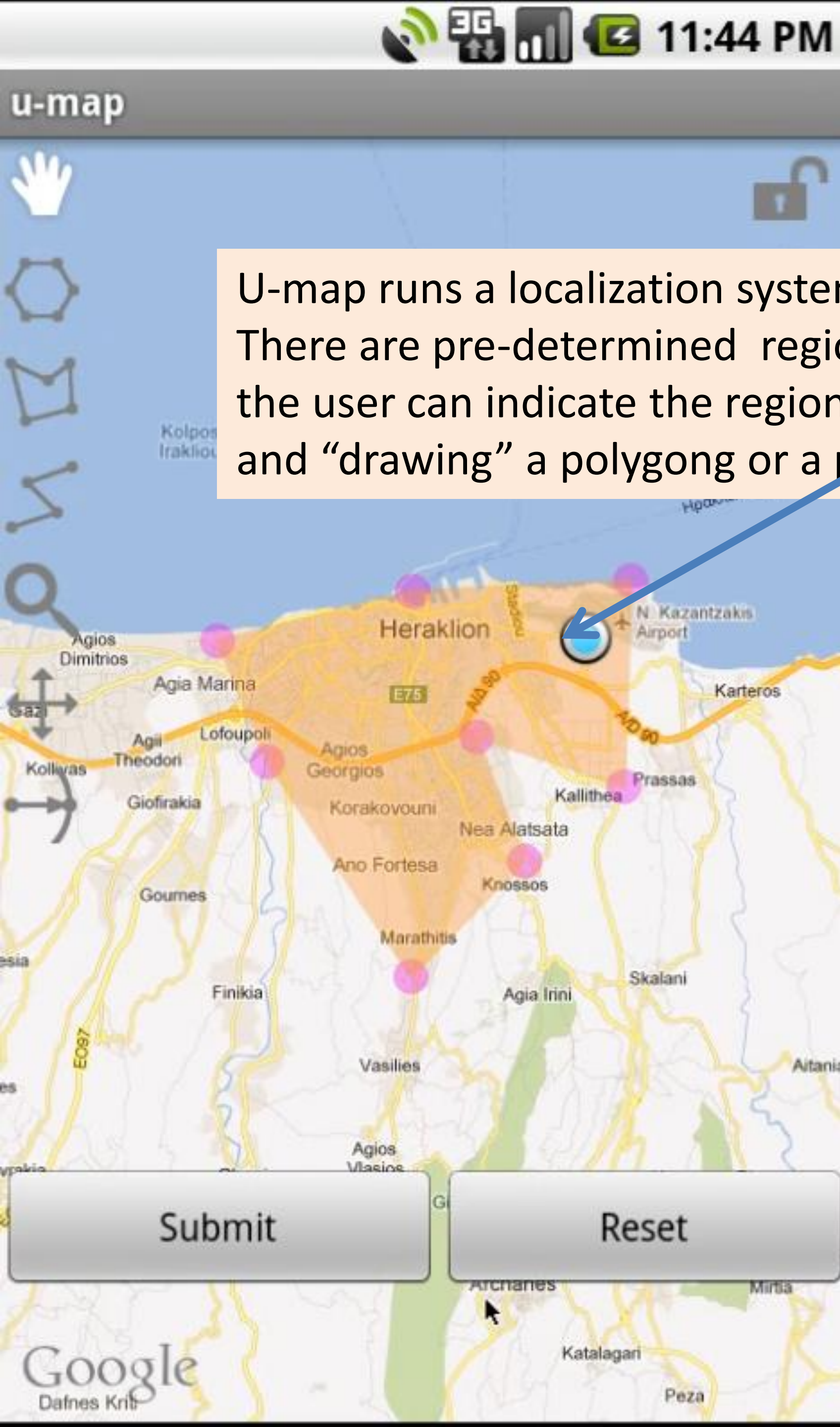
e.g., provider-driven or a third-party

- Strong **access control** that applies privacy rules
- Provision of **incentives**

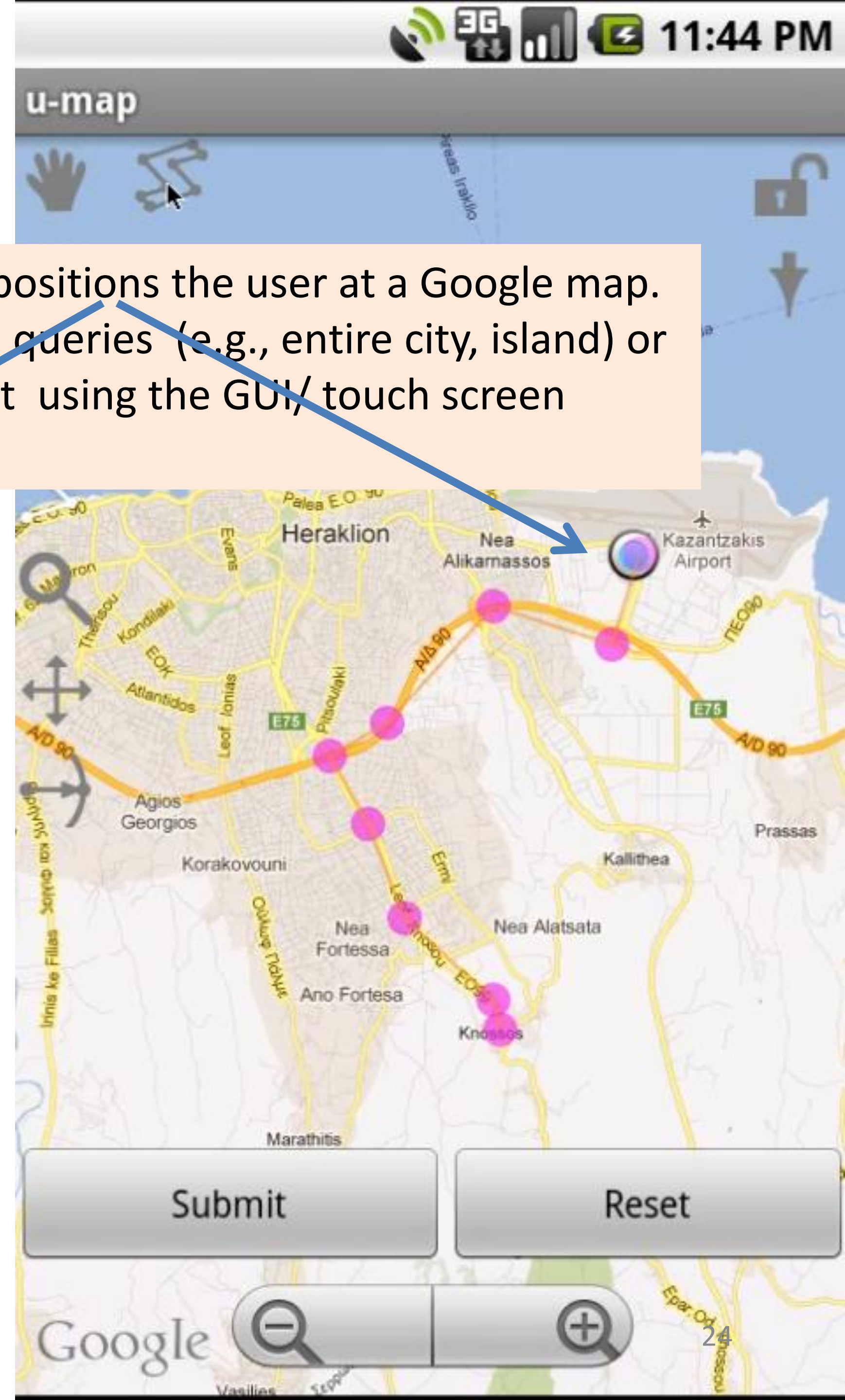
  improved QoE in services

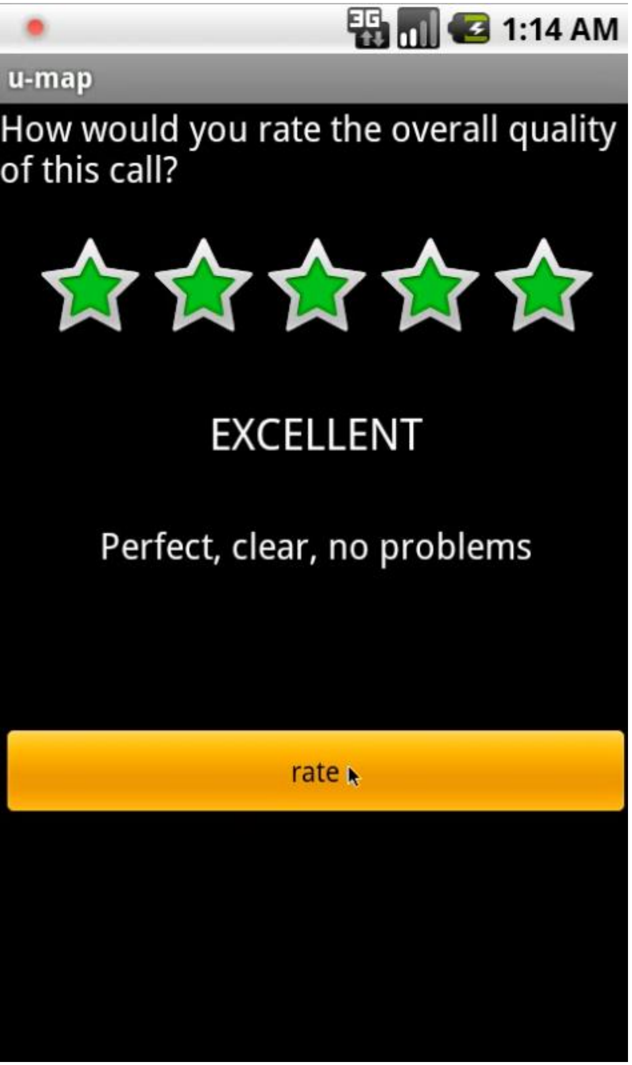
  reputation, altruism,

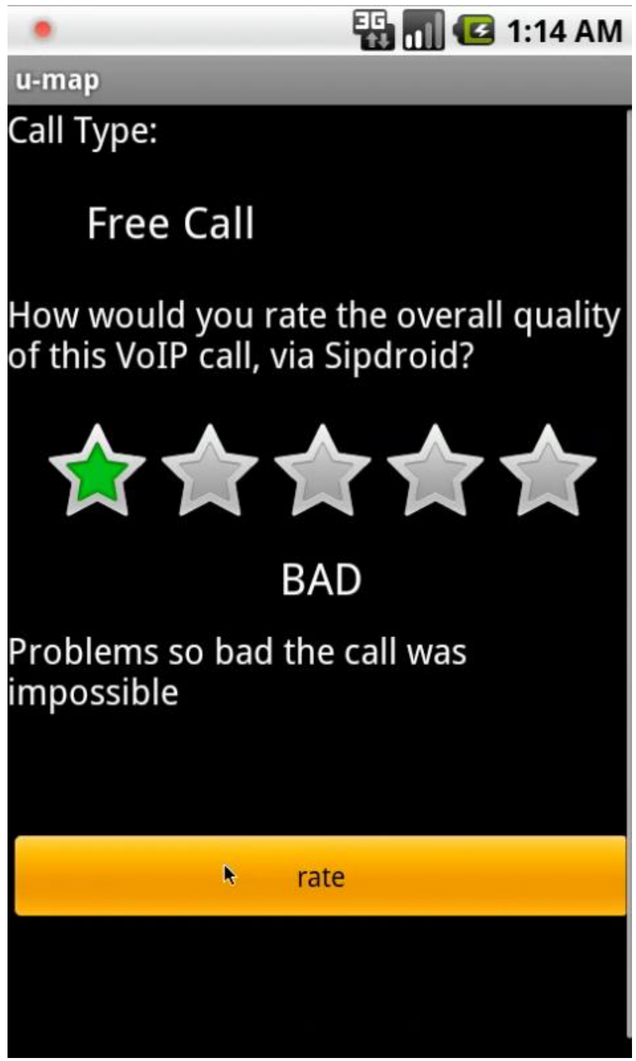
  payment (e.g., free SMS, calls)

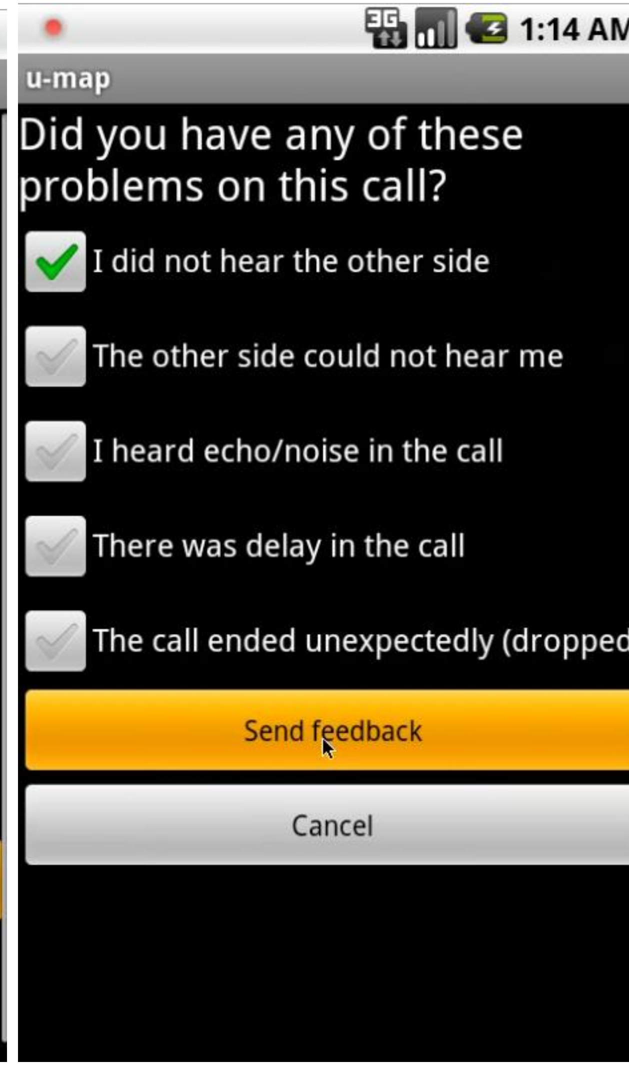


U-map runs a localization system, which positions the user at a Google map. There are pre-determined regions for the queries (e.g., entire city, island) or the user can indicate the region of interest using the GUI/ touch screen and “drawing” a polygong or a path.

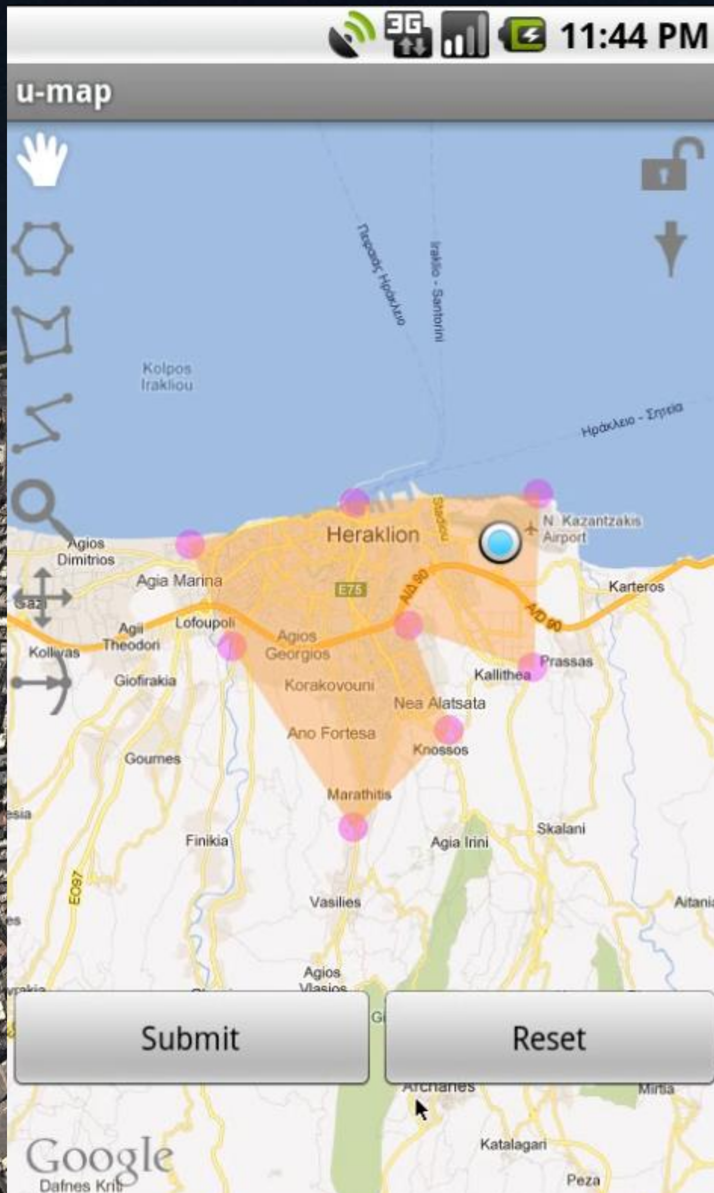


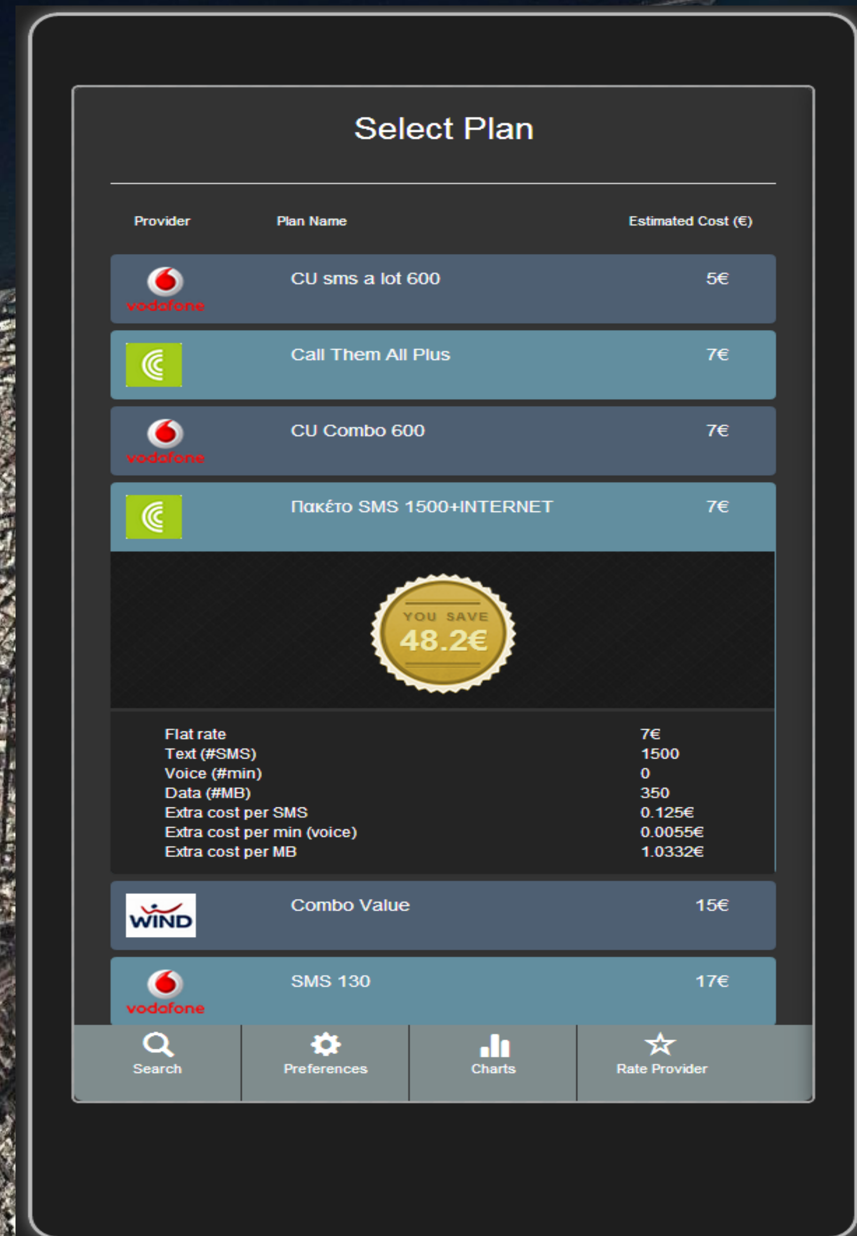

1:14 AM
u-map
How would you rate the overall quality of this call?
EXCELLENT
Perfect, clear, no problems
rate


1:14 AM
u-map
Call Type:
Free Call
How would you rate the overall quality of this VoIP call, via Sipdroid?
BAD
Problems so bad the call was impossible
rate


1:14 AM
u-map
Did you have any of these problems on this call?
I did not hear the other side
The other side could not hear me
I heard echo/noise in the call
There was delay in the call
The call ended unexpectedly (dropped)
Send feedback
Cancel

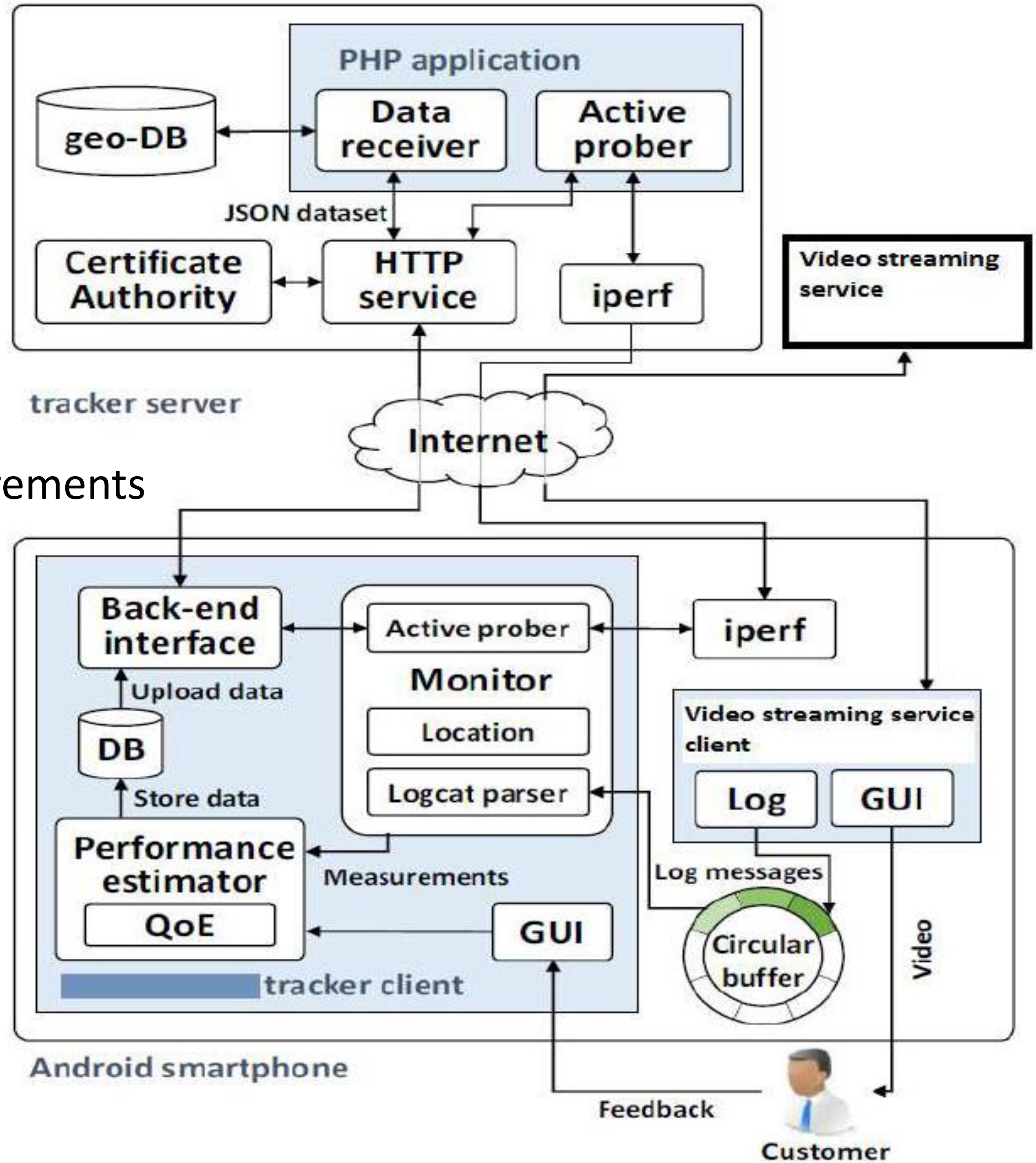
u-map finds the best plans/tariffs based on demand & user preference



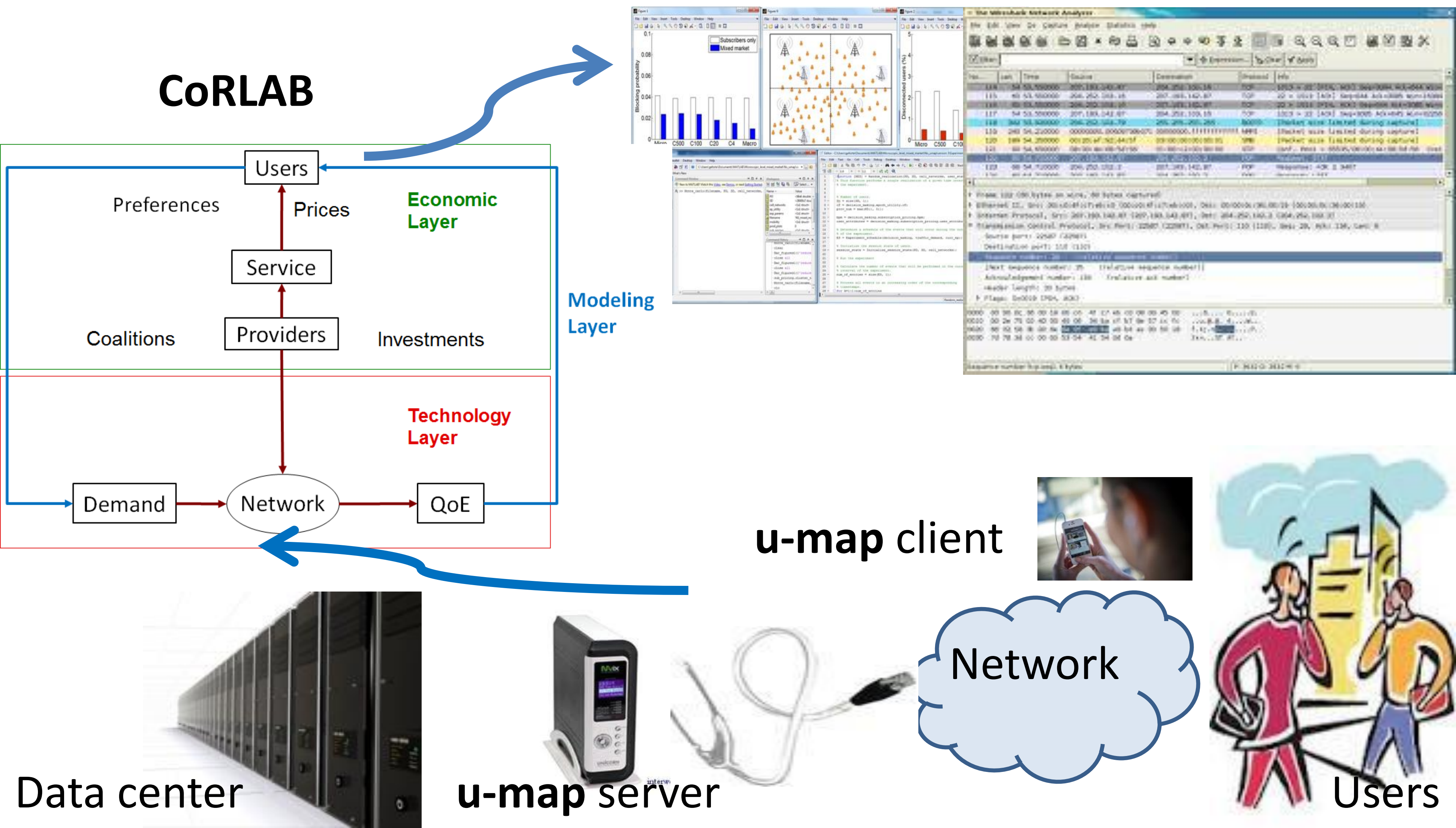


QoE tracker

Network-QOS Measurements
Active prober (iperf)



u-map feeds the analysis platform with **real customer data,** so the business-driven assessment becomes more **accurate, relevant**, **faster.**



Case 1: Field study – completely uncontrolled

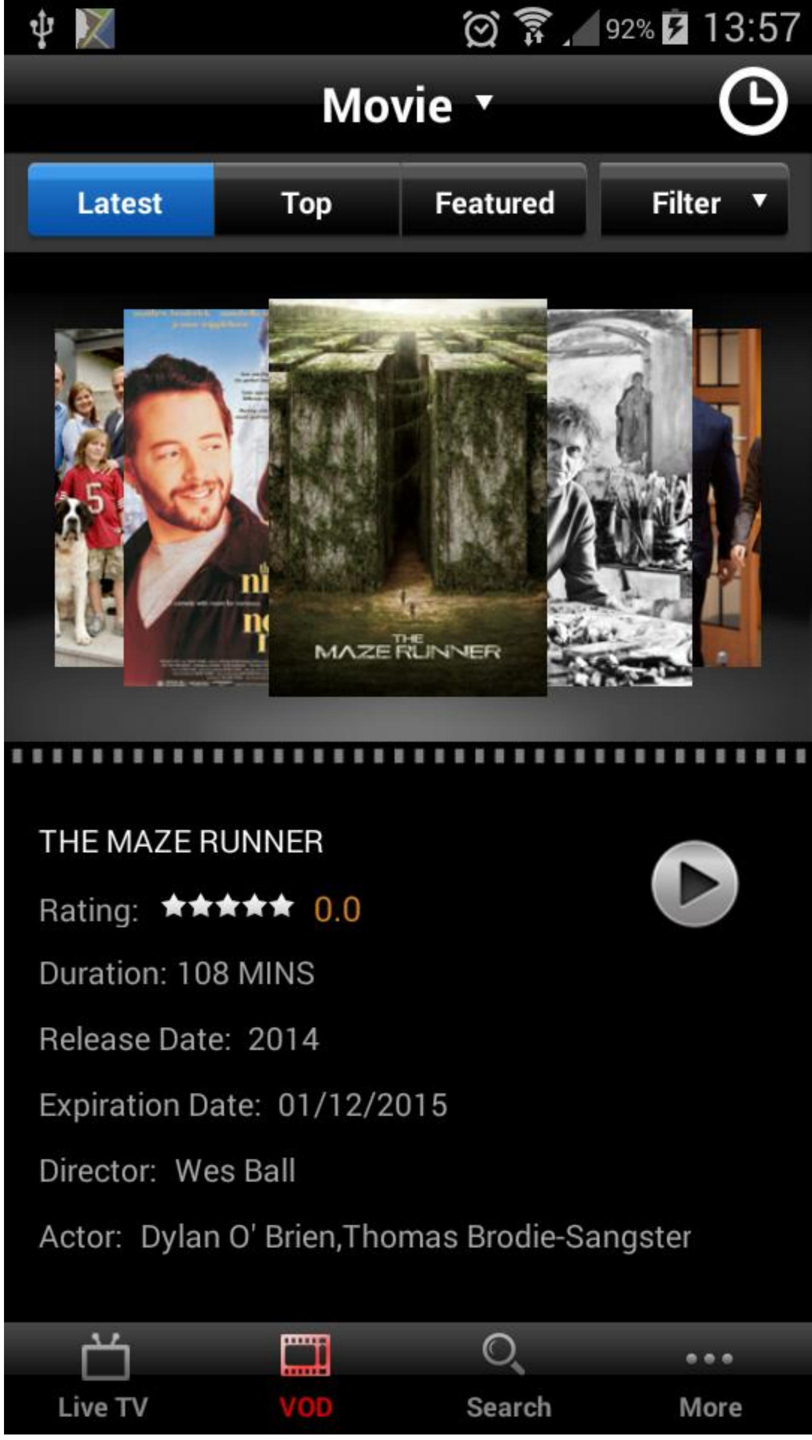
Duration: 56 days

Volunteers: 20 customers

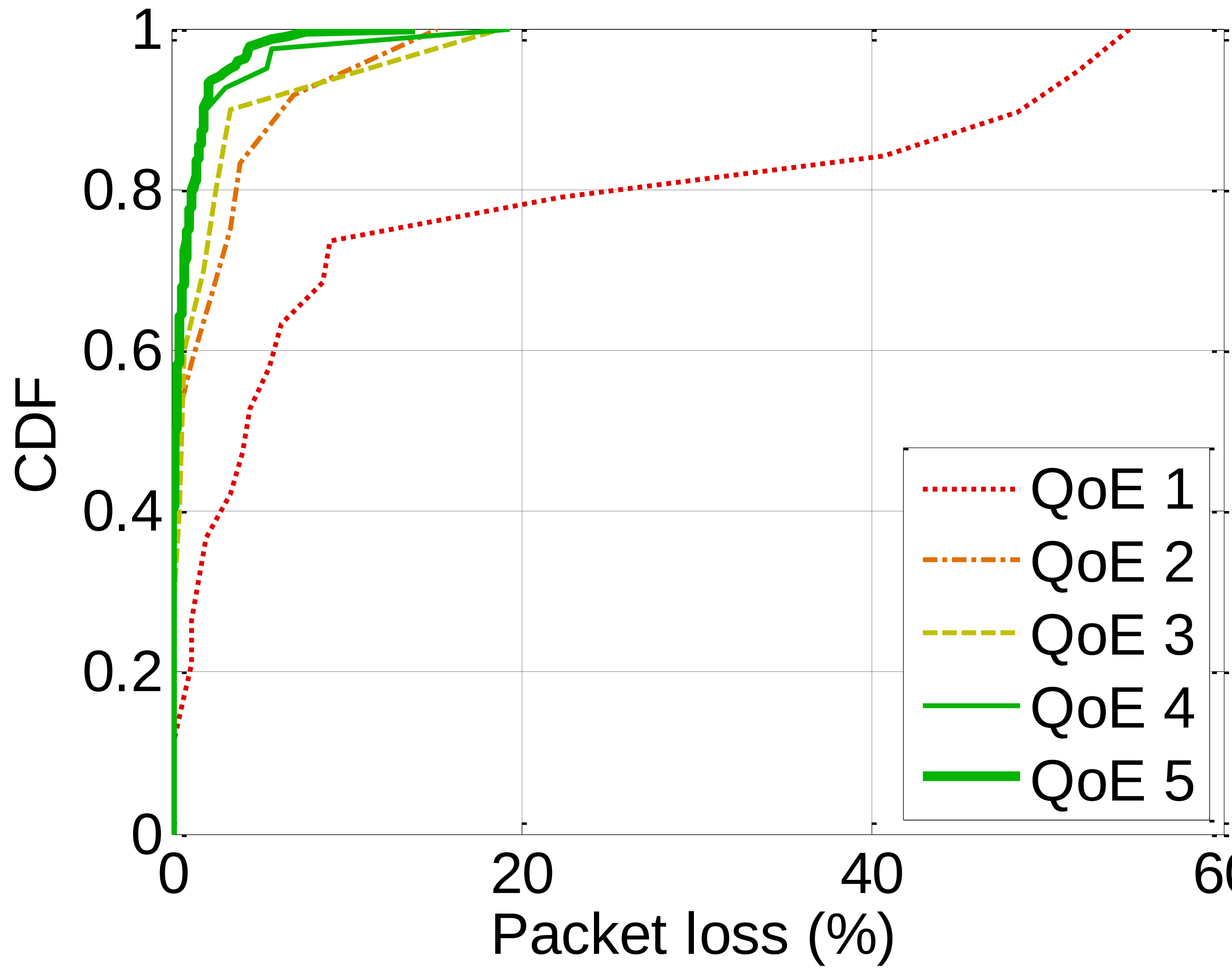
- 20 with ≥ 1 labeled video session
- 13 with > 5 labeled video sessions
- Their devices vary in terms of manufacturer, model, display size & Android version

Volunteers free to

- select any videos of their video streaming service
- move & connect via wireless APs
- view the video



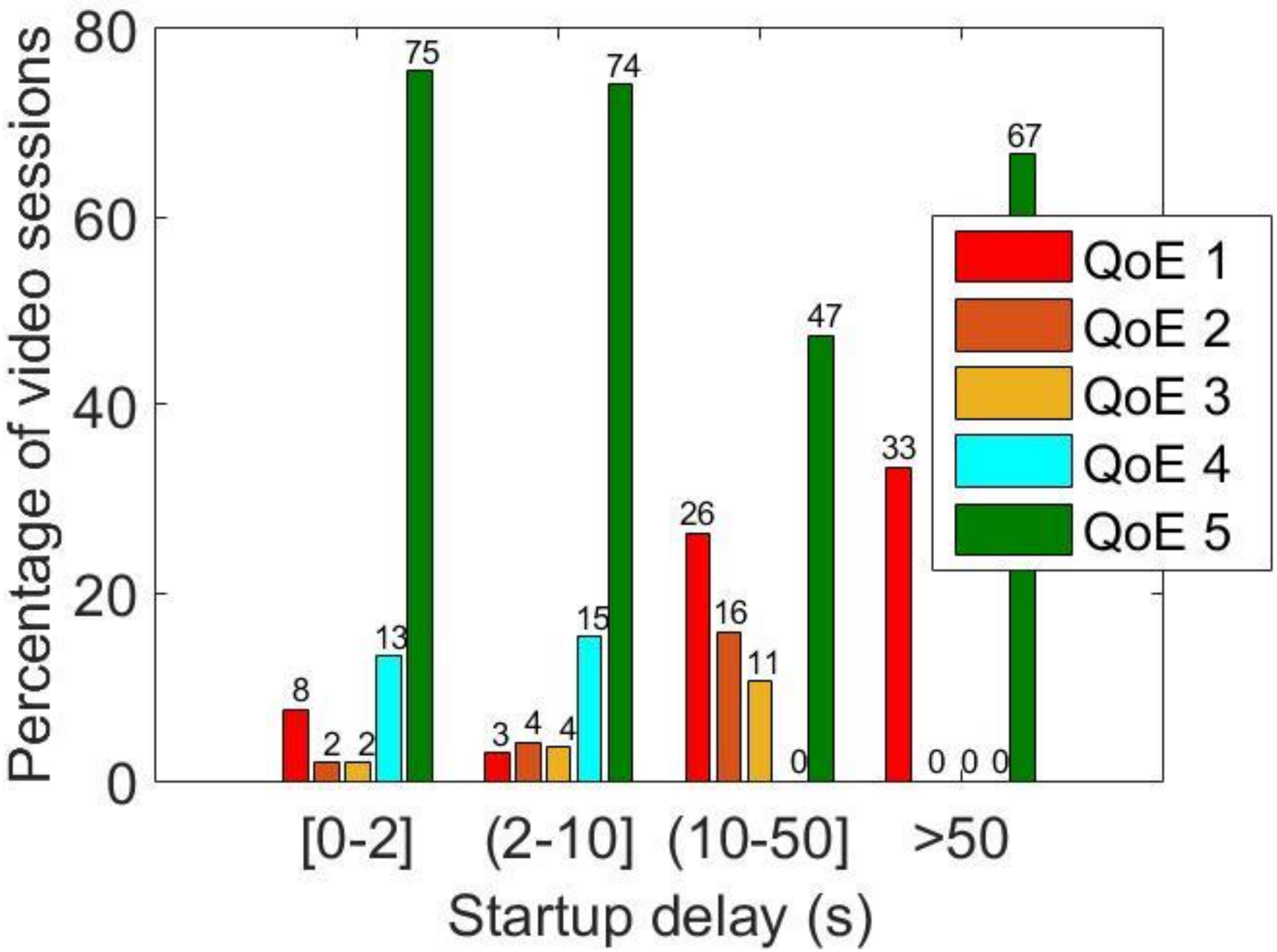
Sessions with higher startup delay, buffering ratio & lower network performance have lower QoE



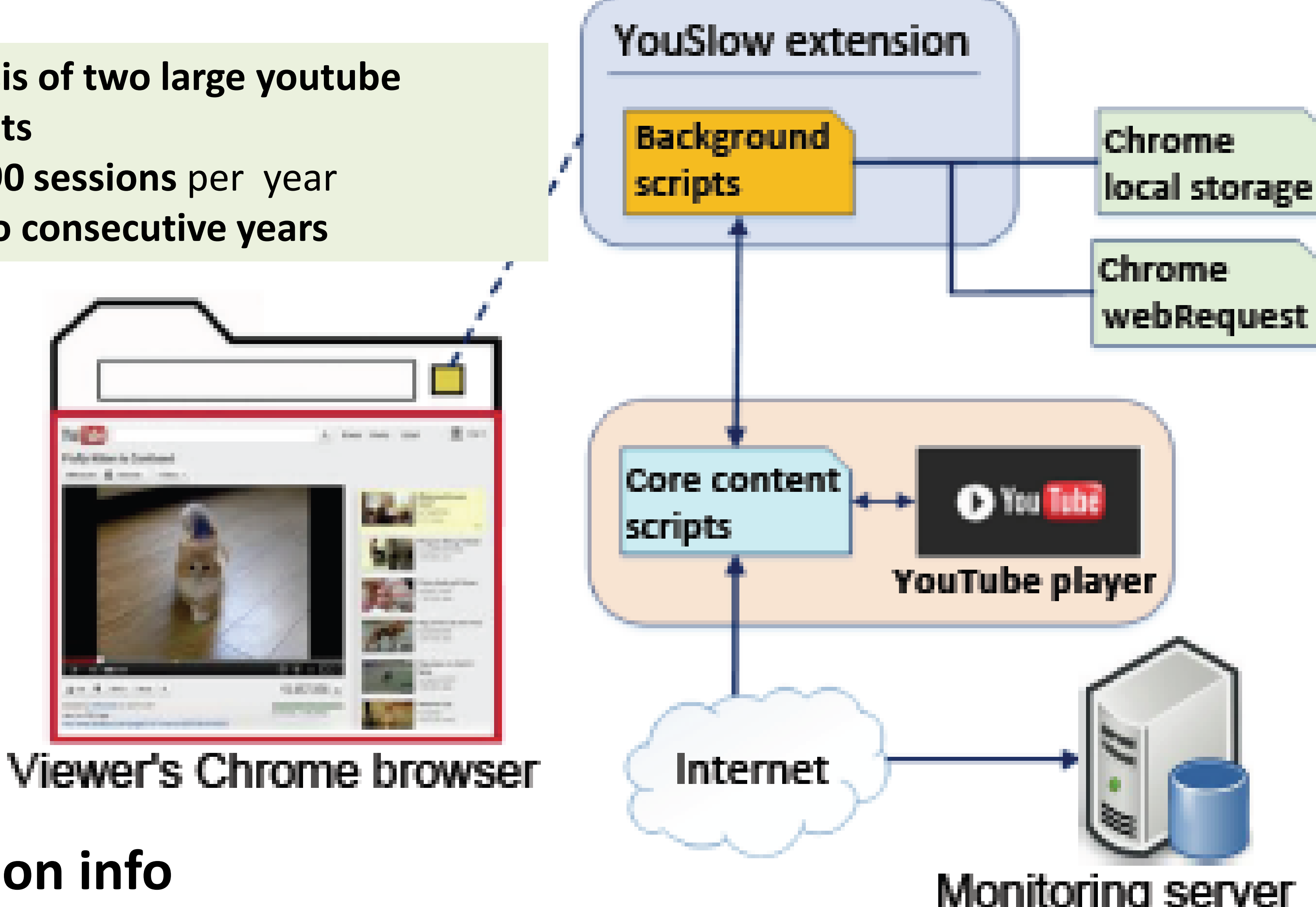
Users perceive the degradation for startup delay ≥ 10 sec

Startup delay above 2 sec causes viewers to abandon the video [Krishnan 13]

Our speculation: smartphone users are more tolerant



Analysis of two large youtube datasets
470,090 sessions per year
for t**wo consecutive years**



Session info

- ISP, date & time, city & region, country , latitude & longitude
- Type of abandonment, video **resolutions,** total viewing duration, **initial Buffering**, video duration, # of FWD & BWD skips, r**ebuffering** events

Case 2: Controlled study

Scenarios of **different types of impairment**

- 4 reference **videos** of high quality
- Each video consists of 4 chunks and lasts 5sec

50 playback video parameterized based on

- **Startup delay**
- **Buffering events**
- **Video resolution**

- 20 participants viewed all videos

using the same device in a controlled environment

Three types of prominent impairments
- large startup delay
- number of buffering events
- low resolution

Depending on the type of impairment:
- some users are consistently more tolerant/strict than others
- some users are more tolerant to **some types** of impairment & more strict to others
- **statistically significant difference** of the scores of users for the various types of impairment

Modeling QoE

- **Mathematical models** using QoS parameters
  - **Weber Fehner Law (WFL), IQX** hypothesis
  - **ITU-T.P1202** based on **log-logistic models** (based on initial delays and stalling)
  - **ITU-T.P1203**

    https://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-P.1203-201710-I!!PDF-E&type=items

- **Signal processing** techniques (e.g., using PESQ, PSNR)
  - Compare two video signals & estimate the **perceptual difference between them**
  - Need for reference files

- **Data-mining** algorithms, such as:

  Support Vector Regression, Gaussian Naïve Bayes, Random Forests, Deep learning algorithms

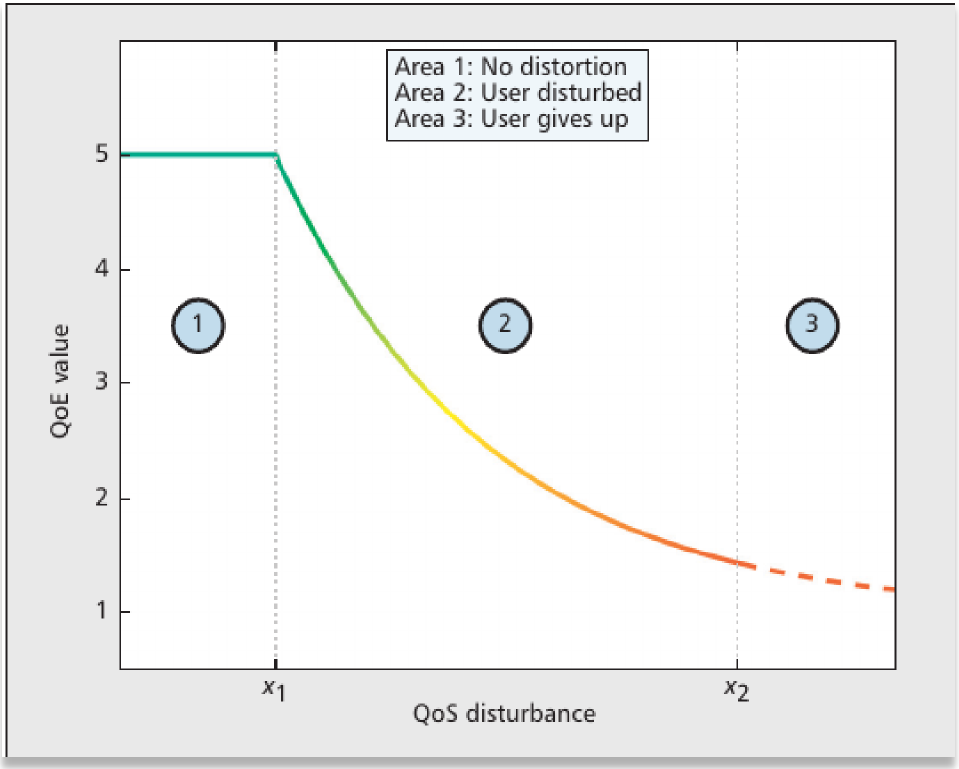
Depending on the Availability of Source Signal

- No-Reference (NR), Reduced-Reference (RR) &Full-Reference (FR) models
- FR models compare the **source signal with the received one**

e.g., an automated telephone call that was recorded both at the source & receiving end

  - Data transfer of the signals is often required
  - Impractical to be used at remote ends, especially when large video traffic has to be inspected
- To reduce the amount of information needed, RR & NR models typically operate on the client side only
  - RR models receive an auxiliary channel of information
  - NR models only inspect received signals

- Signal-based models work on the level of pixels & samples only
  - They assume full access to data and decoding capabilities
- Hybrid models combine signal information with bit-stream-level information, such as packet headers
- Parametric models only operate on transmitted packet-level or bitstream-level information
- A model with access to more information (e.g., decoded video frames instead of just packet headers) should provide a more accurate estimation of the quality, but in practice, the amount of information accessible is often influenced by several extrinsic factors beyond control of the ISP



Weber-Fechner Law: as an actual stimulus increases linearly, the intensity of our perception increases only **logarithmically**.

$$\frac{\partial QoE}{\partial QoS} \sim -\frac{1}{QoS}$$

Elements of psychophysics

- Perceived loudness/brightness is proportional to logarithm of the actual intensity measured with an accurate nonhuman instrument.
- The eye senses brightness approximately logarithmically over a moderate range (but more like a power law over a wider range), and stellar magnitude is measured on a logarithmic scale.
  - This magnitude scale was invented by the ancient Greek astronomer Hipparchus in about 150 B.C. He ranked the stars he could see in terms of their brightness, with 1 representing the brightest down to 6 representing the faintest, though now the scale has been extended beyond these limits; an increase in 5 magnitudes corresponds to a decrease in brightness by a factor of 100.
- There is a new branch of the literature on public finance hypothesizing that the Weber–Fechner law can explain the increasing levels of public expenditures in mature democracies.

Mathematical models of QoE

Weber-Fechner Law

$$\frac{\partial QoE}{\partial QoS} \sim -\frac{1}{QoS}$$

$$QoE = \log(aQoS + b)$$

and integrating

IQX

$$\frac{\partial QoE}{\partial QoS} \sim -(QoE - \gamma)$$

$$QoE = ae^{-bQoS} + \gamma$$

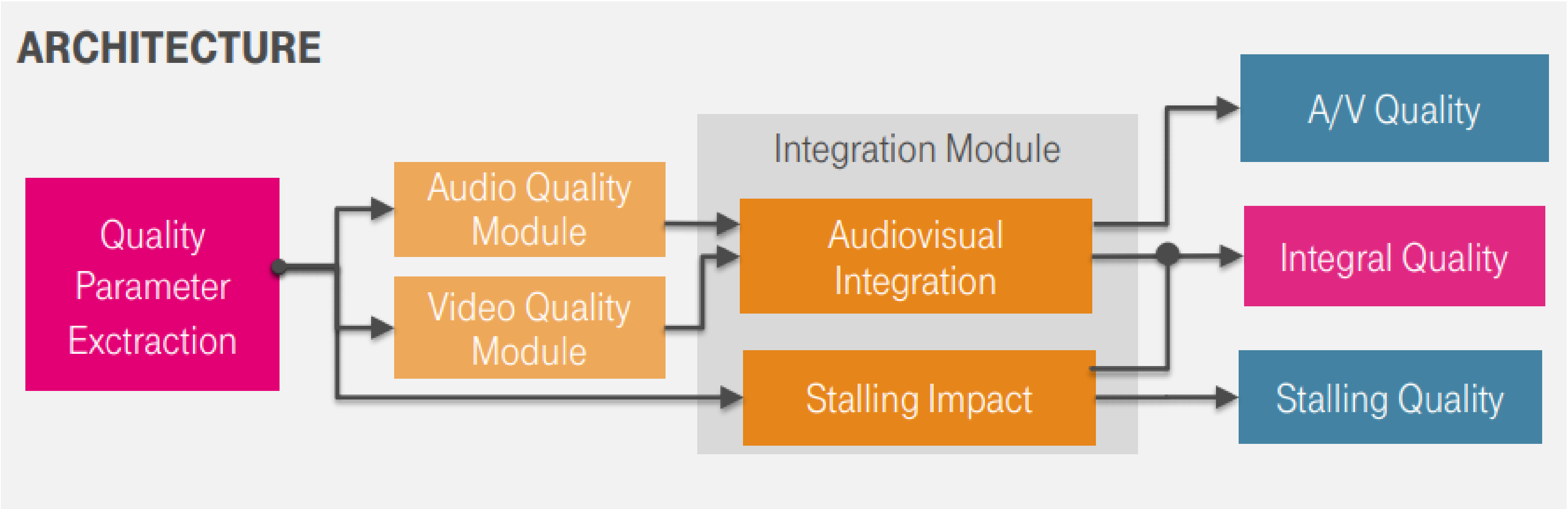
The QoE of video streaming has a **logarithmic dependence on bitrate**
A logarithmic dependence also exists between QoE & **startup delay** for video streaming
The logarithmic dependence in the Weber-Fechner law: the perception is proportional to the relative change of the stimulus [Reichl2010]

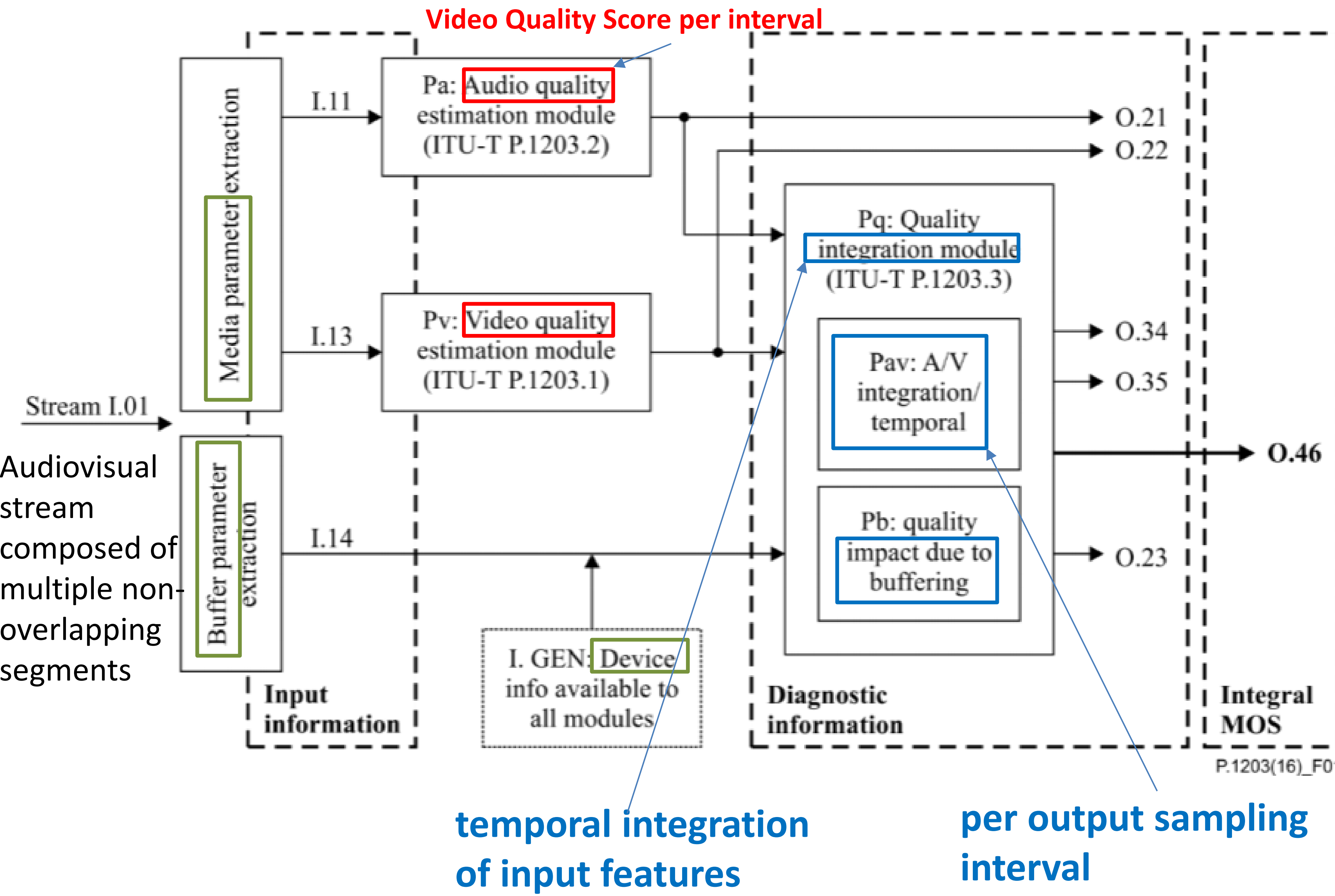
ITU-T.P1203

- For **adaptive-** & **progressive-download**-type media streaming
- Can be deployed both in **end-point** locations & at **mid-network monitoring** points
- **Cannot** provide a comprehensive evaluation of **user perceived transmission quality** because its scores reflect the **perceived impairments due to coded audiovisual media data** being transmitted with certain performance & **does *not* include *specific terminal devices* or user info**
- The scores reflect **average perceptual impairments**
- Benchmarking of different service implementations. However, it ***cannot*** be **used for direct benchmarking of different encoder implementations**.
- Effects such as those due to ***audio levels, signal noise*** and effects due to source generation (e.g., video shake, certain colour properties) and other impairments related to the payload are not reflected in the scores.

ITU-T RECOMMENDATION P.1203 (2017)

First standardized QoE model for HAS
Predicts Mean Opinion Scores for sequences up to 5 min length



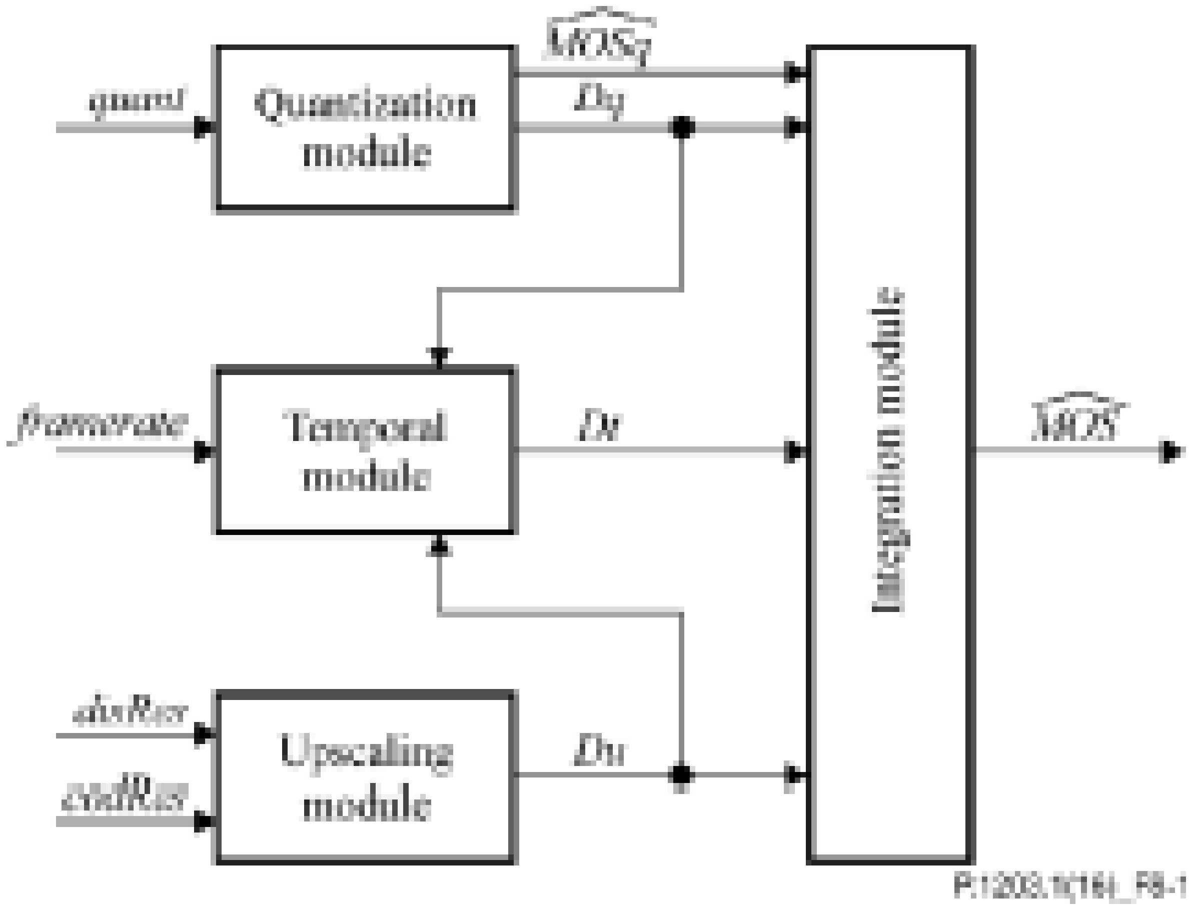


Sliding measurement window for the input data acquisition & output score calculation

Test factors for which the model has been validated
Video compression degradations: ITU-T H.264/AVC High profile, 75 kbit/s – 12.5 Mbit/s For details regarding codec parameters see the Pv module recommendation [ITU-T P.1203.1]
Audio compression degradations tested during standard development: AAC-LC, 32-196 kbit/s For details regarding codec parameters see the audio module Pa [ITU-T P.1203.2] NOTE – The audio quality module Pa is assumed to be valid also for other codecs, since it is identical to the audio coding component in [ITU-T P.1201.2] and [ITU-T P.1201], which has been tested for a larger number of audio codecs. Further audio codecs validated as part of the development of [ITU-T P.1201] are, with the bitrate range from 24-196 kbit/s: AAC-LC, HE-AACv2, AC3, MPEG-LII. See [ITU-T P.1203.2] for details.
Video content: Video contents of different spatio-temporal complexity For details regarding tested video content see the Pv module [ITU-T P.1203.1]
Initial loading delay and stalling degradations: For details regarding specifics of initial loading delay and stalling see the Pq module [ITU-T P.1203.3]
Display Resolutions: Full HD (1920x1080)
Display device: PC/TV monitors, mobile phone (Samsung Galaxy S5)
Media adaptation: Video quality variations caused by switching between different quality levels. For details regarding quality layer properties see [ITU-T P.1203.1]
Frame Rates: 8-30 frames per second

Table 3 − Application areas, test factors, and coding technologies for which ITU-T P.1203 is not intended to be used

Applications for which the model is not intended
In-service monitoring of video UDP-based streaming, where packet loss introduces visible quality degradations
Direct comparison/benchmarking of encoder implementations, and thus of services that employ different encoder implementations
Evaluation of visual quality including display/device properties
Test factors for which the model should not be applied
Audio/video sync distortions
Packet loss distortions
Video codecs for which the model is not validated (MPEG2, ITU-T H.265, VP9, etc.)
Transcoding solutions
The effects of noise, delay, colour correctness



$$SI = e^{-numStalls/s_1} \cdot e^{-\frac{totalBuffLen}{T \cdot s_2}} \cdot e^{-\frac{avgBuffInterval}{T \cdot s_3}}$$

Probes

- Devices that extract and process information sent over a network (e.g., counting and forwarding packets)
- May implement quality models but can also be exclusively used for simple network traffic monitoring
- Active and passive
  - Active probes initiate data transfer
  - Passive probes inspect traffic that passes through them without interfering & not generating additional media traffic

Platforms with Dedicated Probes

- Hardware probes are able to gather round-the-clock measurements, while software measurements are more susceptible to resource contention from other applications and are harder to calibrate have **lower distribution costs**
- SamKnows, BISmark, RIPE Atlas are platforms that deploy dedicated hardware-based probes
- Dasu, Netradar, Portolan, perfSONAR rely on software installations for some hardware systems

SamKnows (Company)

internet performance, beyond speed.

Knowing the speed to download a file from a server, that's already close to you, won't help you fix your home Wi-Fi problems, optimise your connection to Fortnite, or prevent poor Netflix performance. At SamKnows, we build measurements ...r today's internet. We do measure

Monitor your internet performance over time to know how it changes throughout the day. Check you're getting what you paid for from your service provider.

Measure internet performance on a national scale. Check ISPs deliver on their promises and that customers are treated fairly. Create a booming digital economy with effective and informed policy.

ISPs

Understand your customers' internet experience. Gain perfect sight of your entire network and beyond. Isolate faults and deploy targeted fixes. Embed SamKnows into your existing infrastructure.

Manufacturers

Enhance your CPE with the SamKnows internet performance measurement suite. Anticipate your customers' requirements. Create a brand new revenue stream with the Manufacturer Partner Program.

SamKnows

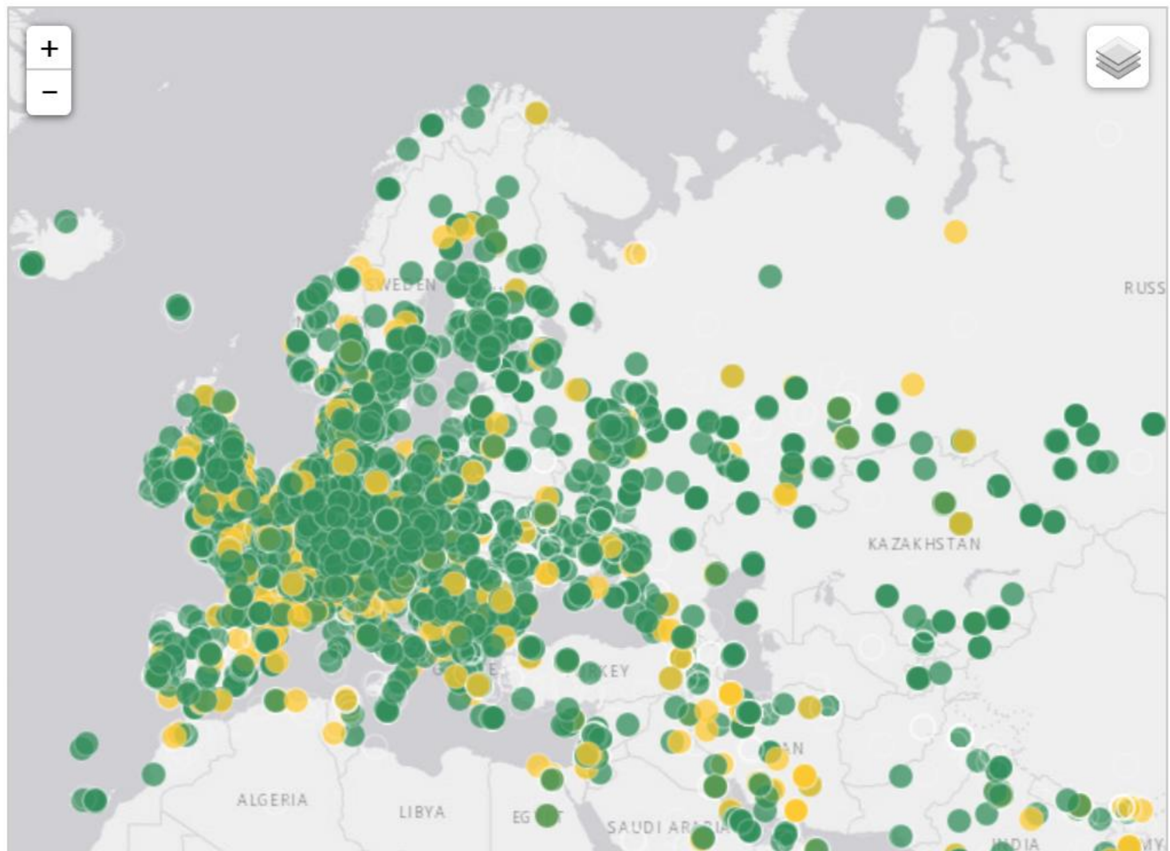
- For ISPs to use accurate, third-party data to help them see what was going on, both inside and outside of their networks.
- Includes our full range of measurement agents for fixed and cellular internet
- Global test infrastructure and cloud-based platform
- Securely stores and visualises performance data in real-time

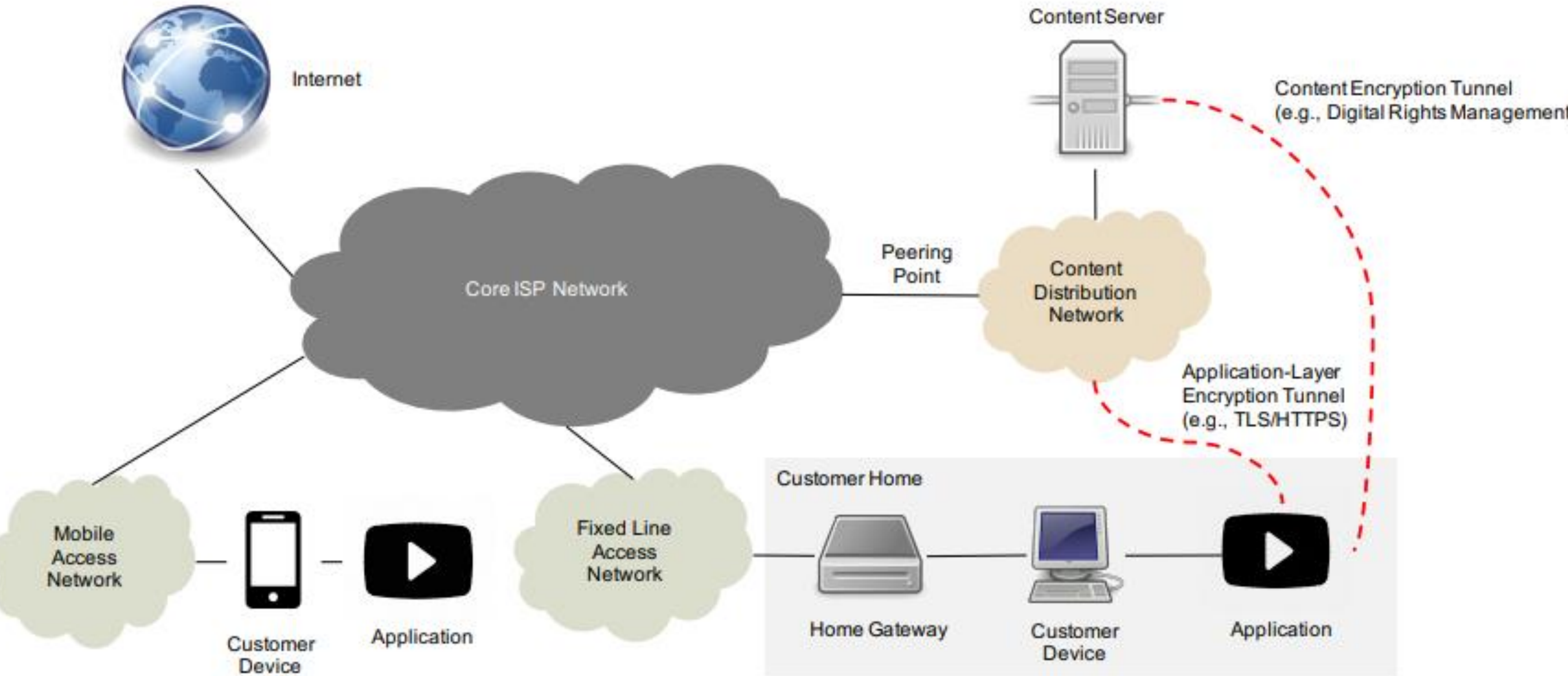
RIPE ALTAS

- Not-for-profit membership association,supporting the Internet through technical coordination.
- Aims to build the largest Internet measurement network ever made
- Employs a global network of probes that measure Internet connectivity & reachability in real time
- Probes are small, USB-powered hardware devices that connect to an Ethernet port on their router. They perform active measurements, e.g., ping, traceroute, DNS, SSL/TLS, NTP
- The aggregated measurements are made publicly available

Measurements currently running

	Built-in	User-defined		
		Total UDM	Anchoring	DNSMON
Ping	41	7268	2386	0
Traceroute	45	6461	2389	882
DNS	158	6102	1	3528
SSL/TLS Certificate	4	386	0	0
NTP	0	100	0	0
HTTP	4	2447	2387	0
WiFi	0	14	0	0





OTT providers have switched to application-level encryption to offer better privacy to their users, such as with the use of SSL/TLS for HTTP or RTP. E.g., YouTube force-redirects most of its users to a HTTPS version of their portal. Their mobile transmissions are mostly encrypted, too.

Sprint communication: Epitiro 4150 probe

- measure QoE for Ethernet, Wi-Fi & LTE
  - with a single probe for fixed & mobile users
- help enterprises & service providers capable to measure parameters which affect the QoE, the performance of commercial services & applications e.g., Gmail, Facebook, YouTube
- network parameters, e.g., latency and speed in LTE, Wi-Fi, Ethernet

MULTI-SERVICE QUALITY MONITORING SYSTEM (MSQMS)

- Distributed quality monitoring system
- Fully automated probes across Germany
- Measure video streams and website calls, directly from the browser, **like a customer** would do
- Different KPIs and KQIs measured:
  - DNS times, ping
  - Upload and download speed tests
  - Video loading times
  - **Video quality according to P.1203**

Toolbox

- Apply **machine learning and data mining algorithms for prediction** e.g., Decision Trees, Support Vector Regression, Artificial Neural Networks, Gaussian Naïve Bayes, Random Forests

- **Feature discovery**: determine the parameters with the most predictive power using Bayesian networks, regression

- **Train the models** based on **empirical measurements** collected from real-world studies

Our approach for **predicting the QoE:**

Develop **user-centric, service-oriented** models based on **network metrics**

- Apply **machine learning and data mining algorithms**, such as:

  Decision Trees, Support Vector Regression, Artificial Neural Networks, Gaussian Naïve Bayes
- Find the set of predictors that minimizes the mean absolute error of a model (**feature selection**)
- **Train the models** based on **empirical measurements** collected from field studies

We have demonstrated this methodology for **VoIP, audio & video streaming**

Main aspects of MLQoE

- User-centric
- Training the models based on collected **network measurements & user feedback**
- **Automatic** selection of the best algorithm & parameter tuning for predicting QoE
- Robustness on the number of **dominant factors** for predicting QoE
- Conservative bound on its future performance (using nested cross-validation)
- Performs dimensionality reduction using feature selection algorithms
- Addresses the over-fitting

Youtube Measurement Studies

youslow plugin over Chrome browser developed by Nam and Schulzrinne

<u>Data collection</u>

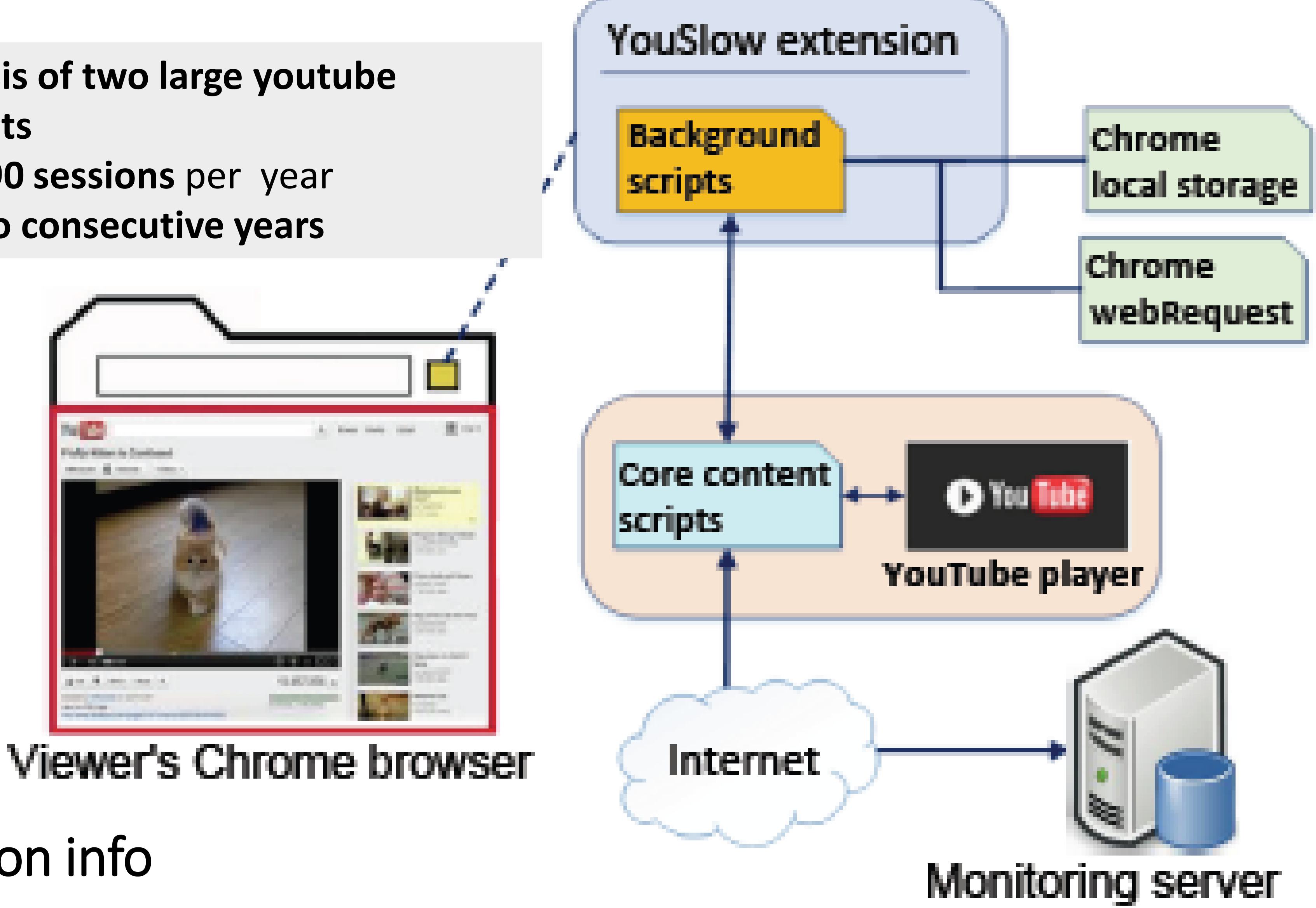
Over 1,400,000 youtube views from more than 1,000 viewers located in more than 110 countries

Two studies:

February 2015 and July 2016 and

January 2017 and July 2017

Analysis of two large youtube datasets
470,090 sessions per year
for t**wo consecutive years**



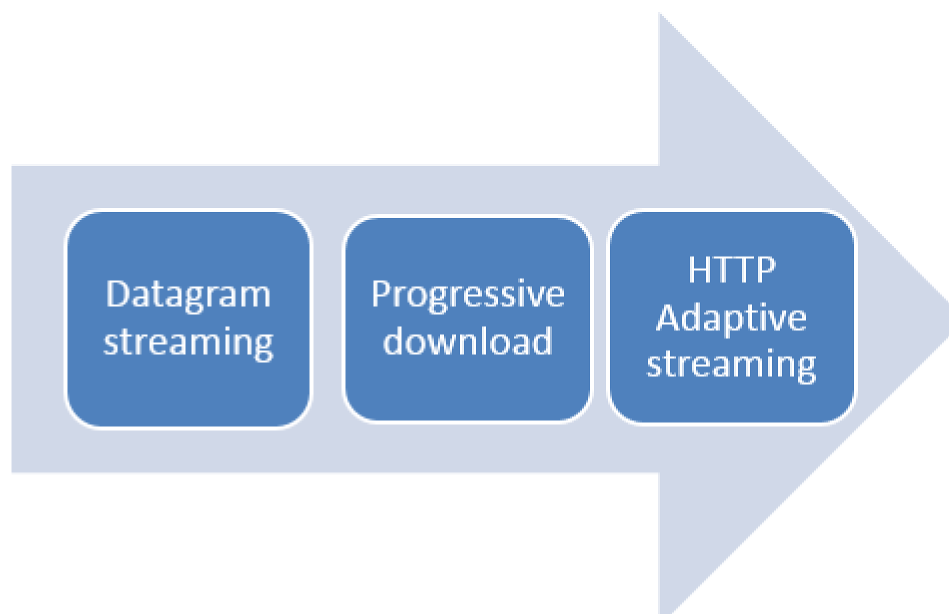
Session info

- ISP, date & time, city & region, country , latitude & longitude
- Type of abandonment, video **resolutions,** total viewing duration, **initial Buffering**, video duration, # of FWD & BWD skips, r**ebuffering** events

Popular Streaming Technologies

- Real Time Messaging Protocol (RTMP) / Real Time Streaming Protocol (RTSP)
- Progressive download
- ABR streaming

Video streaming evolution



RTMP/RTSP chunk based delivery

- A series of video chunks and a **Flash player** consumes the content **instantly without any local caching**
- The streaming server using dynamic RTMP contains multiple bitrates for a single video file & allows the player to **automatically change the bitrates** during playback based on the network conditions
- RTMP/ RTSP streaming requires a **special Flash-based media server**.

Progressive Download

- A video is delivered by a **regular web** server using **HTTP (**no streaming server)
- **No quality adjustment**
- Easy to setup and cost-effective
- Upon a video request, an HTTP server pushes the video content as quickly as it can
- The playback starts as soon as enough content has been downloaded
- FF or skipping ahead is only possible for the downloaded content
- There is a security concern since the player **caches the video content on the viewer's device**

Adaptive Bitrate Streaming (ABR)

- Popular video streaming services, e.g., YouTube, Netflix, HBO GO, BBC, use it
- **Automatic quality switching** and ease of delivery over **HTTP**
- Popular ABR technologies:

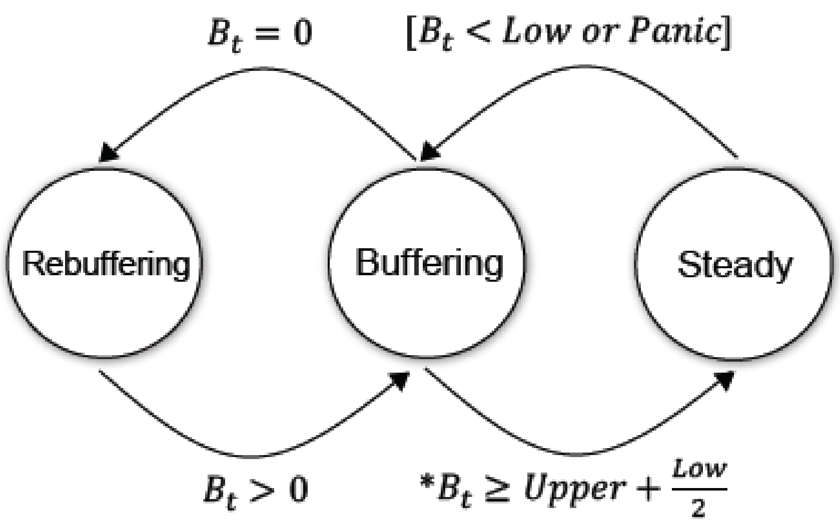
  Apple's HTTP Live Streaming (HLS), **Microsoft's Smooth Streaming (SS)**

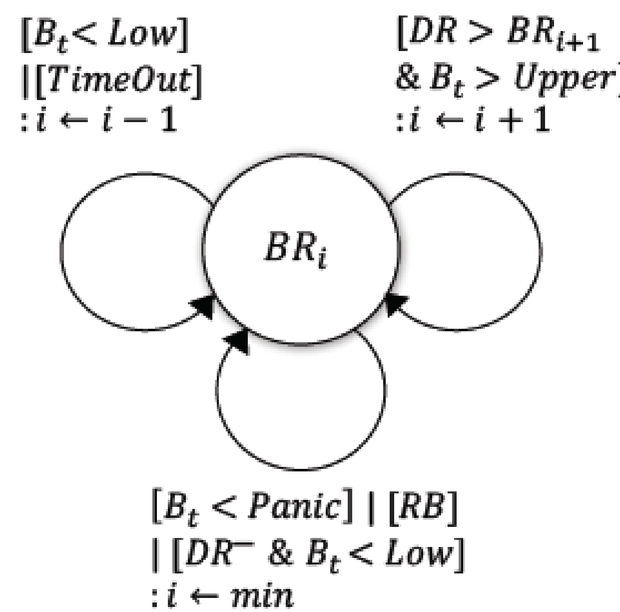
  Adobe's HTTP Dynamic Streaming (HDS), 3GPP/MPEG DASH
- A video server contains **multiple bitrates** encoded for a single video object
- Each bitrate file is split into small **segments**
- A segment size is measured ***in seconds*** (**not bytes**), typically **2-10 sec**
- A video player **dynamically adjusts** bitrates based on ***estimated network condition***s, ***buffer occupancy, hardware specifications*** of viewers' devices (e.g., distinguishing smartphones from desktops)

Finite state machine (FSM) of state change & bitrate switching behavior of Microsoft's SS players

While the video is being played, the state of player can be **Buffering, Steady** or **Rebufferring**

BRi : the video bitrate (in kb/s) selected by a player during playback





Bt: how much video content is currently left in the **playout buffer** (in sec).

* It depends on ABR configuration

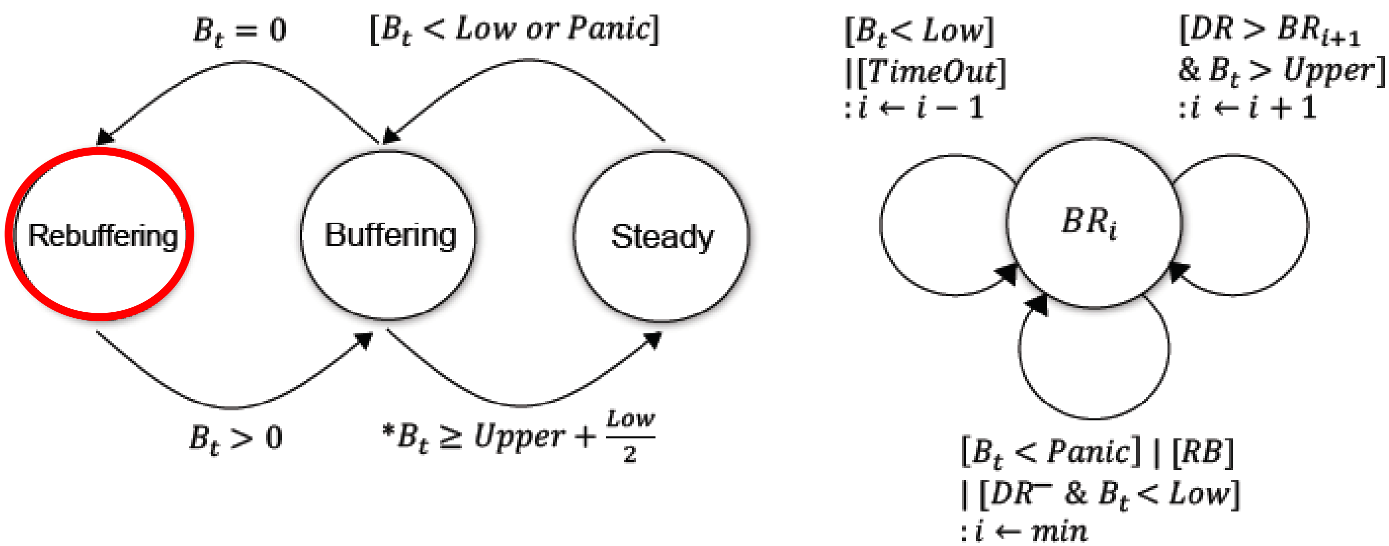
Buffering state: the player **aggressively downloads** video content into its playout buffe
The player requests the next segment right after it completely downloads the current segment (**back-to-back HTTP requests**) so that the buffer can be filled as quickly as possible.

Steady state: the player tries to **keep the buffer full**, instead of increasing the playout buffer level by downloading the segments back-to-back. Request one segment for ever segment duration.

When the playout buffer is running low, the state will switch to Buffering again.

Rebuffering: **no video content** available in the playout buffer during playback.

Rebuffering: no video content

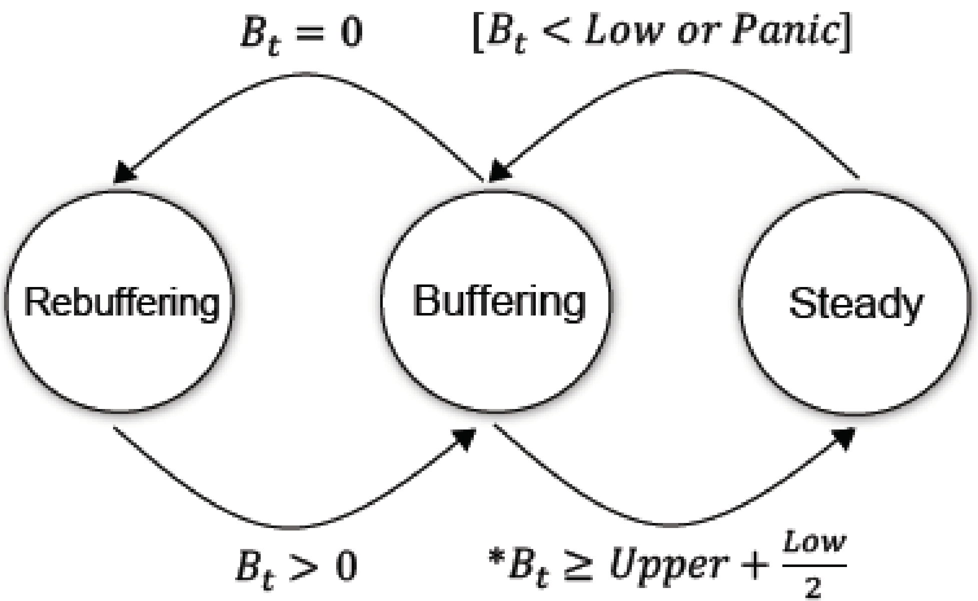


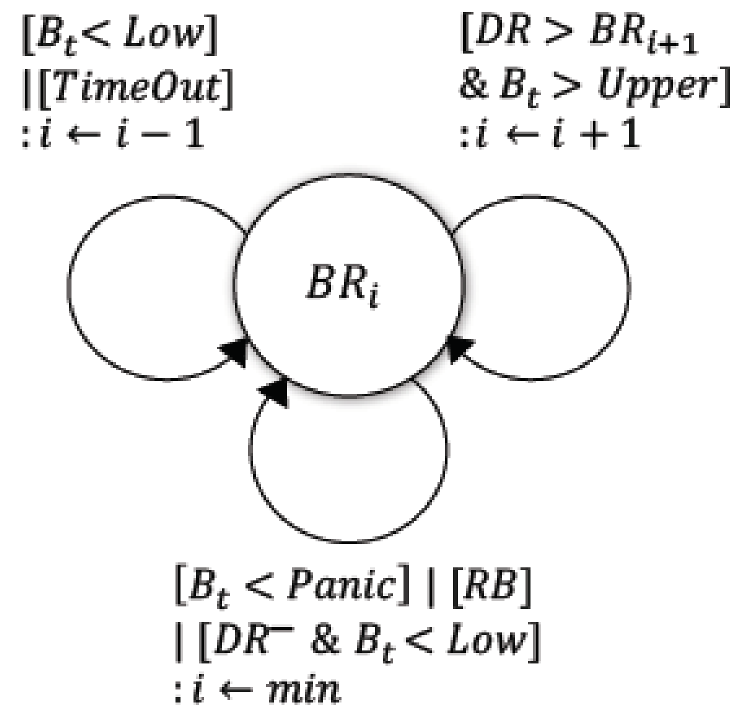
* It depends on ABR configuration

DR: current **downloading data rate** measured by a bandwidth estimator in an ABR player
DR-: available bandwidth in the network is decreasing

BRi : the video bitrate (in kb/s) selected by a player during playback

Bt: how much video content is currently left in the **playout buffer** (in sec).





* It depends on ABR configuration

TimeOut: The timer is set to estimate network conditions. It activates when the elapsed time for downloading a requested segment is longer than expected. In such case, the bitrate is decreased for the next request.

Selecting the Best Available Bitrate during Playback

- Real-time available network bandwidth
- Amount of video remaining in the playout buffer during playback
- Screen resolution & video rendering capabilities of viewers' devices
- Frame rate & viewers' interactive actions (e.g., resizing the browser window) during playback

A player may experience frequent frame drops when a system is running multitasking that requires significant RAM & CPU

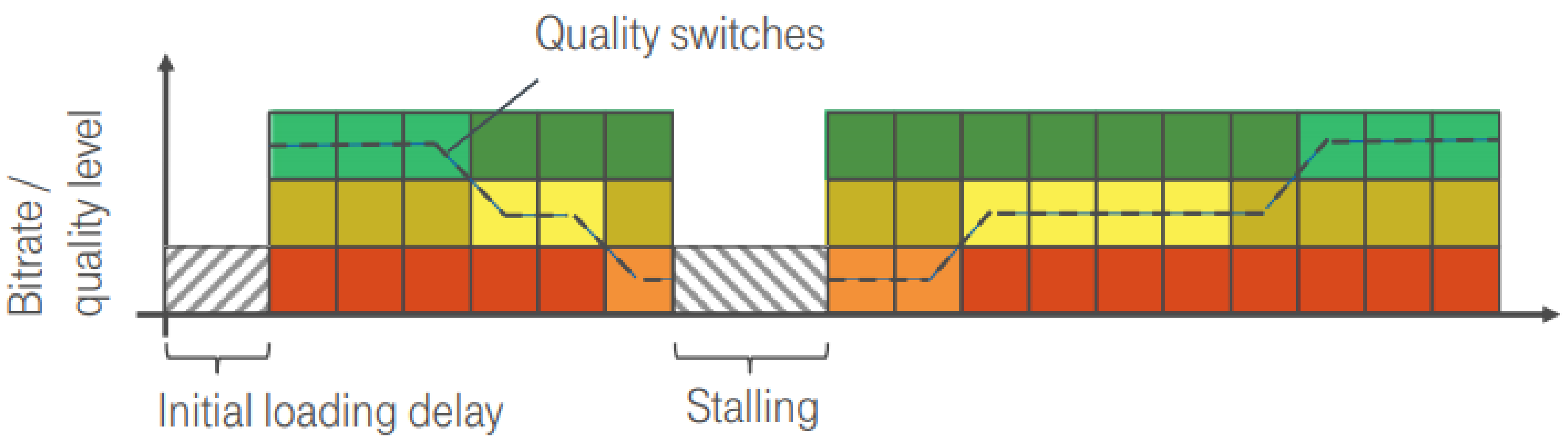
- Under a large number of frame dropped:

1. the player flushes its buffer
2. re-downloads the discarded segments at **lower playback** rates to provide a good video quality

HTTP Adaptive Streaming (HAS) as the leading technology for video streaming today.

QoE may be degraded due to:

- Inaccessible services
- Initial loading delays, stalling events
- Reduced video resolution / quality switching

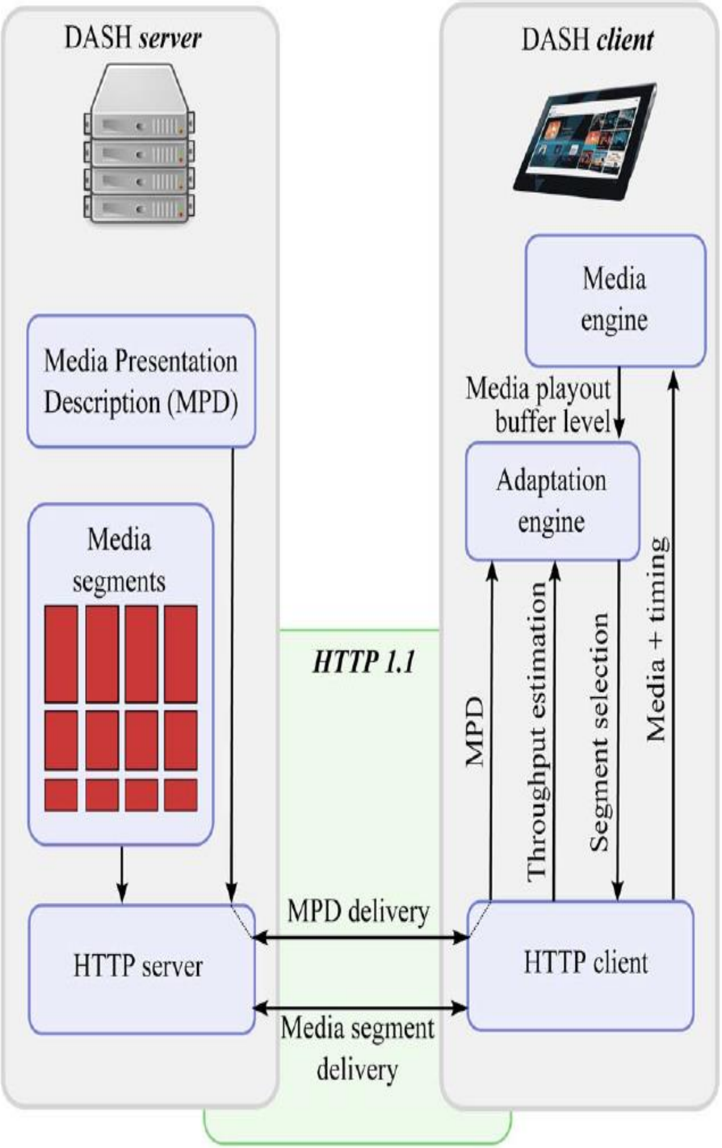


HTTP Adaptive streaming

- Video chunks in different bitrate at the server
- Client selects bitrate based on throughput & buffer availability

Impairments:

- Start-up delay
- Rebuffering events
- Positive & negative BR changes
- Low video resolutions



[Seufert *et al*., "A Survey on Quality of Experience of HTTP Adaptive Streaming", IEEE Communication Surveys & Tutorials, 2014]

Youtube Measurement Studies

- RB: rebuffering
- BR: bitrate change
- Pos. BR: increase of the bitrate
- Neg. BR: decrease of the bitrate

Number of youtube sessions that have the corresponding types of impairments

Scenario	Description of sessions	Number of Sessions (%)
1	No RB, No BR	18278 (86.53%)
2	**RBs**, No BR	1340 (6.34%)
3	No RB, only **Pos. BRs**	1033 (4.89%)
4	No RB, only **Neg. BR**s	226 (1.07%)
5	**RBs**, only **Pos. BR**s	175 (0.83%)
6	**RBs**, only **Neg. BRs**	72 (0.34%)

Key Findings of QoE for Video Streaming (Related Work)

- Rebufferings (RB) are critical:
  - Duration & Number of the events
- Startup delay > 2 sec has negative impact
- Bitrate changes (BR) are important:
  - Negative BR -> poor QoE specially after long periods of high quality
  - Positive BR -> can also dissatisfy users
  - Same average BR is perceived differently, depending on the level fluctuation patterns
  - The annoyance of staying at low quality grows exponentially with the duration it is maintained

1. T. Hoßfeld *et al.* , "Pippi Longstocking calculus for temporal stimuli pattern on YouTube QoE: $1 + 1 = 3$ and $1 * 4 \neq 4 * 1$." 5th Workshop on Mobile Video. ACM, 2013.
2. S. Krishnan and R. Sitaraman, "Video stream quality impacts viewer behavior: inferring causality using quasi-experimental designs", IEEE/ACM Transactions on Networking, 2013.
3. A. Moorthy *et al.*, "Video Quality Assessment on Mobile Devices: Subjective, Behavioral and Objective Studies", IEEE Journal of Selected Topics in Signal Processing, 2012.
4. Y. Liu *et al.*, "Deriving and Validating User Experience Model for DASH Video Streaming", IEEE Transactions on Broadcasting, 2015.
5. B. Lewcio *et al.,* "Video quality in next generation mobile networks - Perception of time-varying transmission", 2011 IEEE International Workshop Technical Committee on Communications Quality and Reliability (CQR), 2011.

QoE on Different Devices

Mobile phones/tablets

- Higher impact of stallings than image quality on MOS
- BR- after a long period of high quality results to poor QoE
- Users of mobile devices more tolerant to startup delay compared to desktops users
- The environmental context is important for entertainment & information ratings

Desktop computers/laptops

- Users are more likely to abandon videos with multiple rebufferings compared to a single re-buffering event although the rebuffering ratio is the same.
- Users are more likely to abandon a video when startup delay larger than 2 sec

1. A. Moorthy *et al*., "Video Quality Assessment on Mobile Devices: Subjective, Behavioral and Objective Studies", IEEE Journal of Selected Topics in Signal Processing, 2012.
2. H. Nam *et al.,* "QoE matters more than QoS: Why people stop watching cat videos", In IEEE Infocom, 2016.
3. P. Casas *et al*., "Next to You: Monitoring Quality of Experience in Cellular Networks From the End-Devices", IEEE Transactions on Network and Service Management, 2016.
4. S. Wassermann *et al.,* "QoE in Cellular Networks through in-Smartphone Measurements", 12th IFIP Wireless and Mobile Networking Conference (WMNC), 2019.
5. S. Jumisko-Pyykkö and M. Hannuksela, "Does context matter in quality evaluation of mobile television?", In Proceedings of the 10th international conference on Human computer interaction with mobile devices and services, 2008.

Our own analysis of the user-engagement in the context of youtube video streaming

Should I stay or should I go: Analysis of the Impact of Application QoS on User Engagement in YouTube

Maria Plakia [2], Evripides Tzamousis [2], Thomais Asvestopoulou [1,2], Giorgos Pantermakis [2], Nick Filippakis [2], Henning Schulzrinne [3], Yanna Kane-Esrig and Maria Papadopouli [1,2]

[1]Department of Computer Science, University of Crete, Heraklion, Greece

[2]Institute of Computer Science, Foundation for Research and Technology-Hellas, Heraklion, Greece

[3] Columbia University

Under Review in ACM Transactions on Modeling & Performance Evaluation of Computing Systems
https://arxiv.org/abs/1901.01603

Methodological Contributions

User engagement metrics

- video watching duration ratio & abandonment ratio
- time elapsed from the occurrence of an impairment to the end of the session
- % of sessions that get abandoned within a certain time (e.g., 60 sec) after the occurrence of an impairment

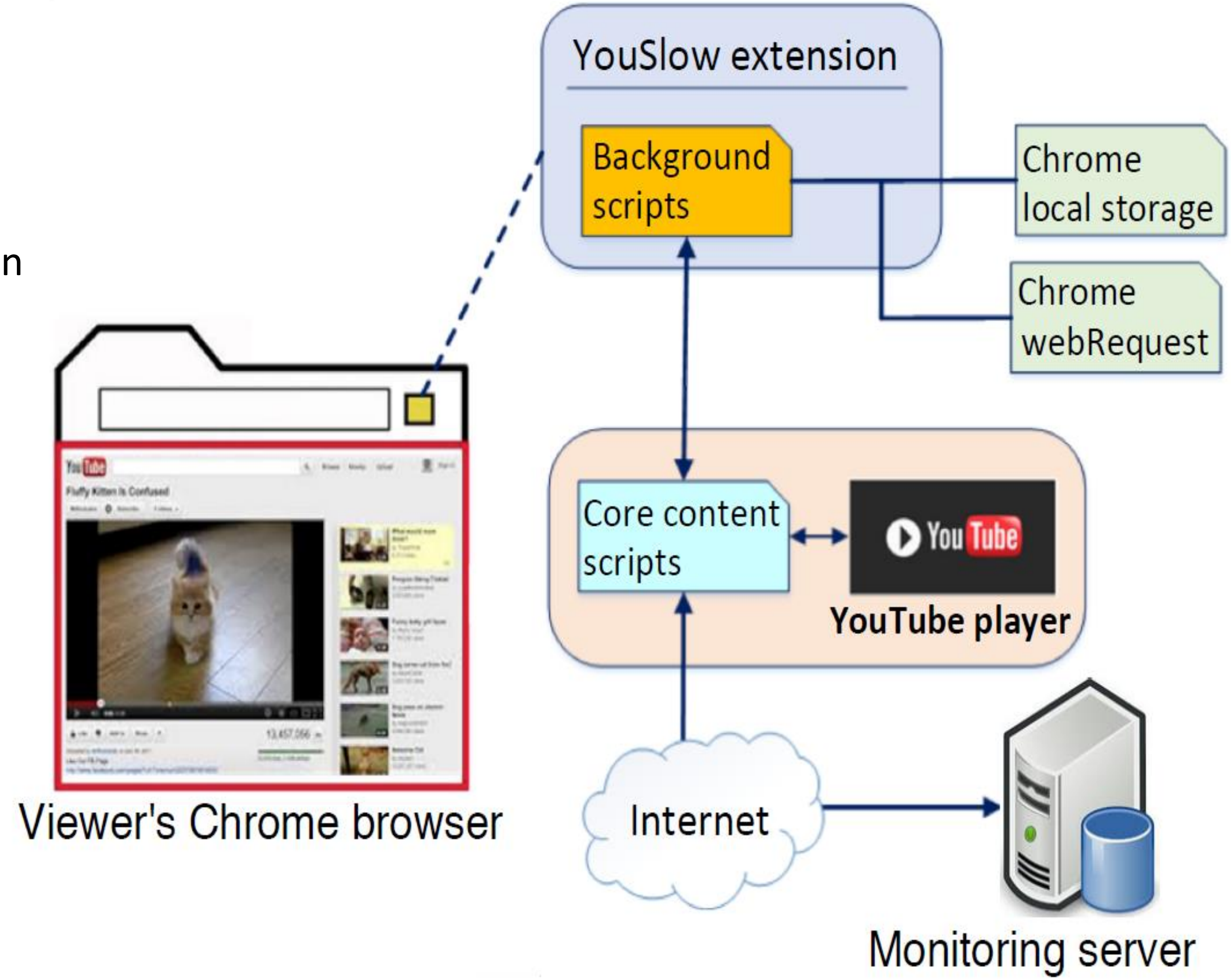
Identified scenarios with specific types of impairments

Performance Analysis

Relationship among different types of impairments & user engagement metrics, considering covariates of sessions (e.g., video duration, mean data rate)

YouSlow

Google Chrome plug-in



H. Nam, K. Kim and H. Schulzrinne, "QoE matters more than QoS: Why people stop watching cat videos", IEEE INFOCOM 2016

Datasets

- Desktop computers and laptops (no mobility)
- Two datasets
  - collected between February 2015 and July 2016
  - collected between January 2017 and July 2017
- More than 1,400,000 sessions

Each video session includes:

- Date and location
- Internet Service Provider
- Played bitrates during the session
- Video duration
- Session Duration
- Start-up delay: presence and duration
- Rebuffering : time and duration
- Abandonment

We performed dataset pre-processing and sanitization

Key Findings of Our Study

1. BR- has a severe negative impact on video watching percentage & abandonment ratio
2. BR+ in sessions with low initial resolution is not well-received
3. High RB ratio has even more prominent impact than BR-
4. Compared to startup delay, RBs have larger impact on the video watching percentage
5. Features with predictive power for the video watching percentage include the number of RBs, number of BR changes, number of negative BR changes, mean weighted bit rate
6. An impairment prior to a BR- increases the likelihood of abandonment

Scenarios Examined:

1. No RB, no BR (*baseline*)
2. RBs, no BR
3. No RB, BR+
4. No RB, BR-
5. RBs, BR+
6. RBs, BR-

Thresholds considered:

For video watching percentage:

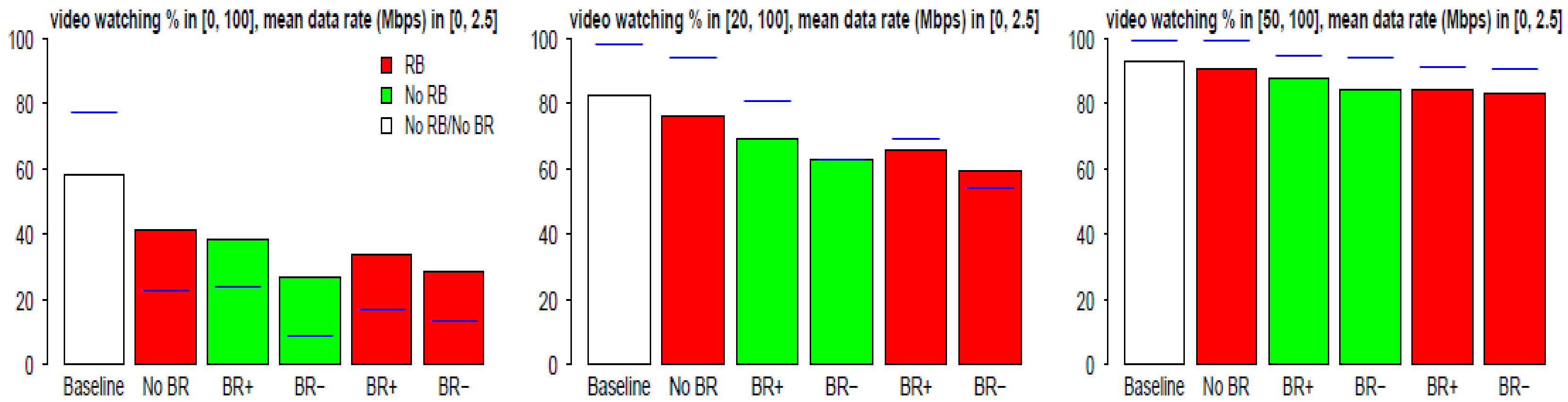
- [0, 100]
- [20, 100]
- [50, 100]

For mean weighted bitrate:

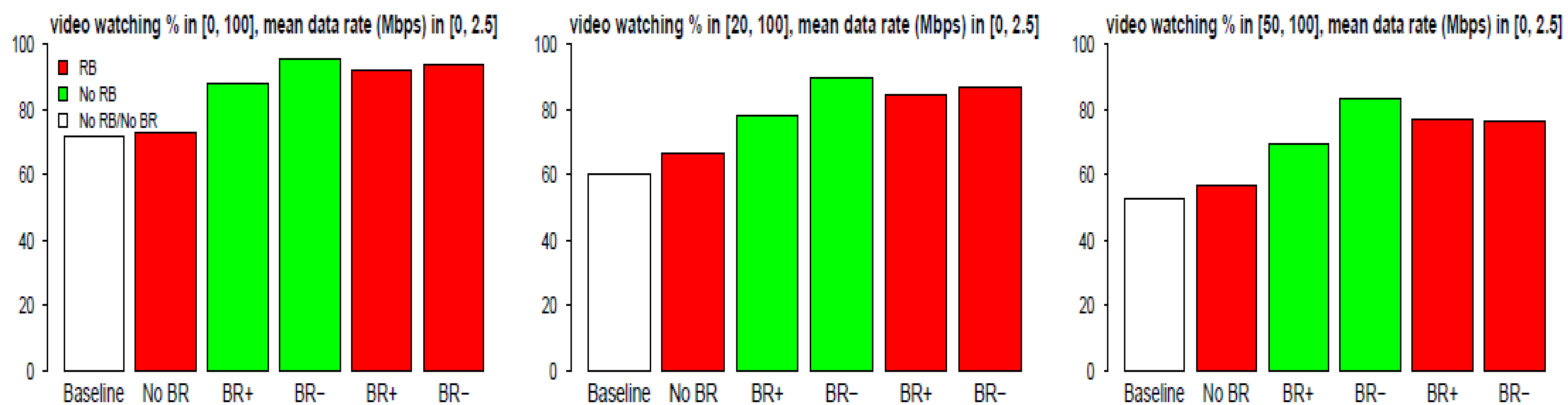
- [0, 2.5] Mbps
- [1, 8] Mbps
- [2.5, 8] Mbps
- [6, 8] Mbps

BR- has a severe negative impact on video watching percentage & abandonment ratio
BR+ in sessions with low initial resolution is not well-received

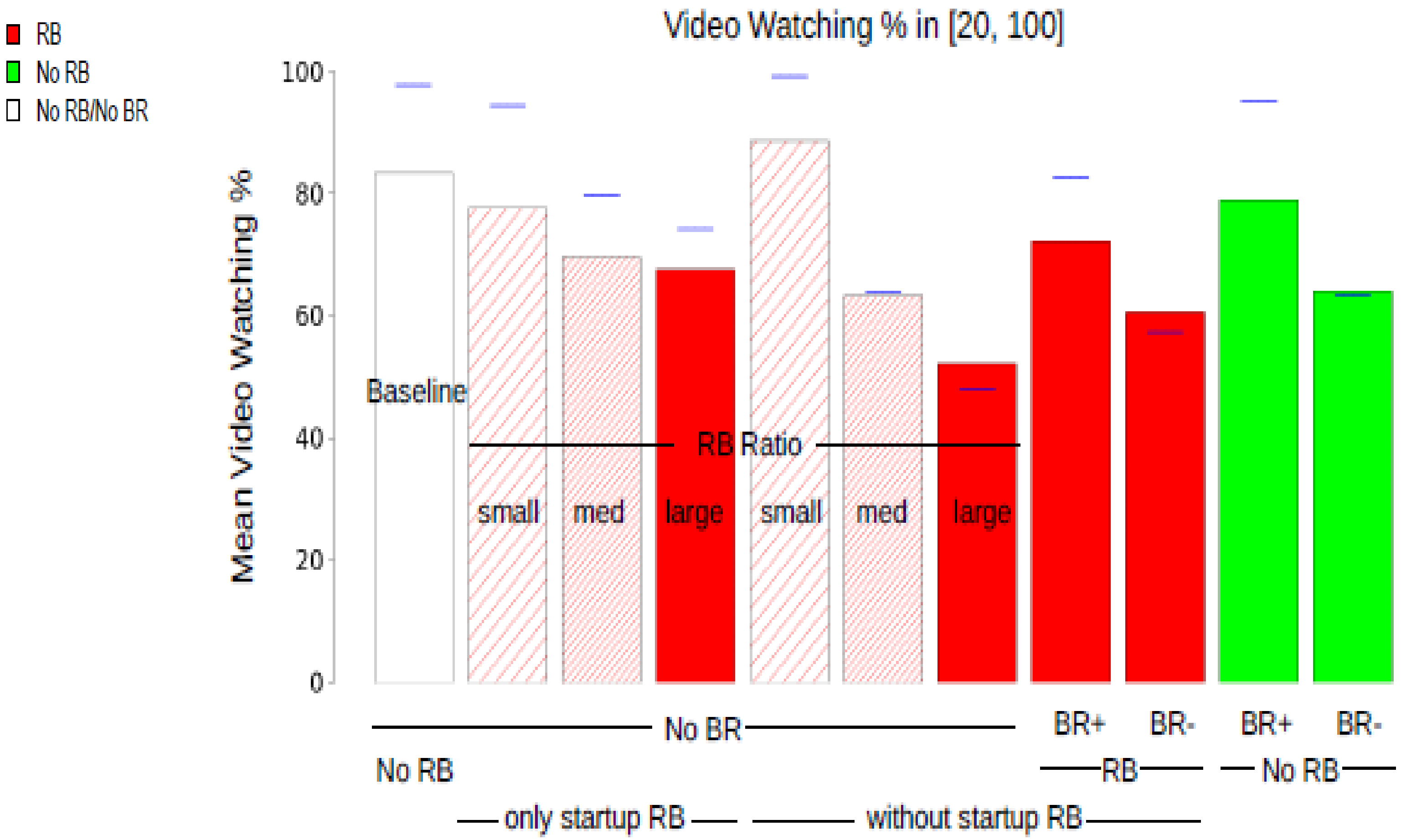
Mean Video Watching Percentage



Abandonment ratio



Compared to startup delay, RBs have larger impact on the video watching percentage

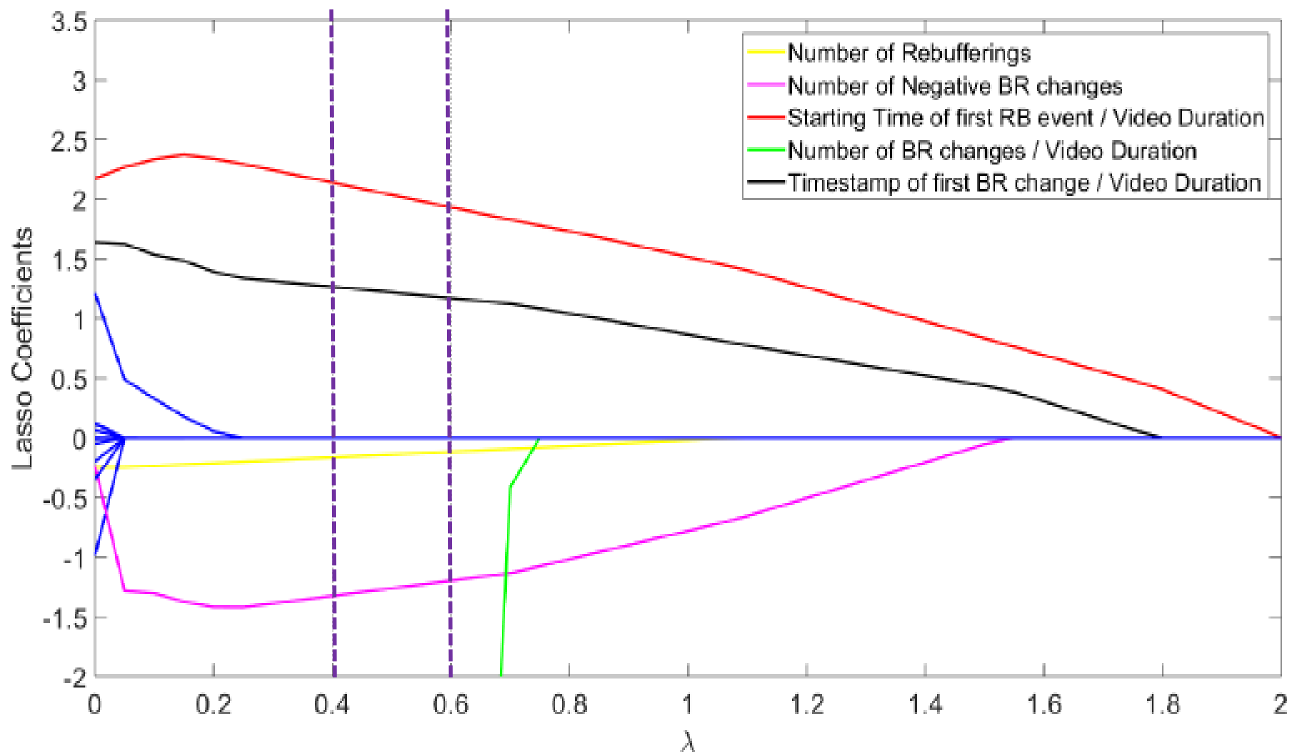


Features with predictive power for the video watching percentage include the number of RBs, number of BR changes, number of negative BR changes, mean weighted bit rate

- Used LASSO regression to find the dominant features

$$\beta^{LASSO} = \arg\min_{\beta}\left\{\frac{1}{2N}\sum_{i=1}^{N}\left(y_i - \beta_0 - \sum_{j=1}^{p} x_{i,j}\beta_j\right)^2 + \lambda\sum_{j=1}^{p}|\beta_j|\right\}$$

As λ gets higher, a smaller number of features are taken into account



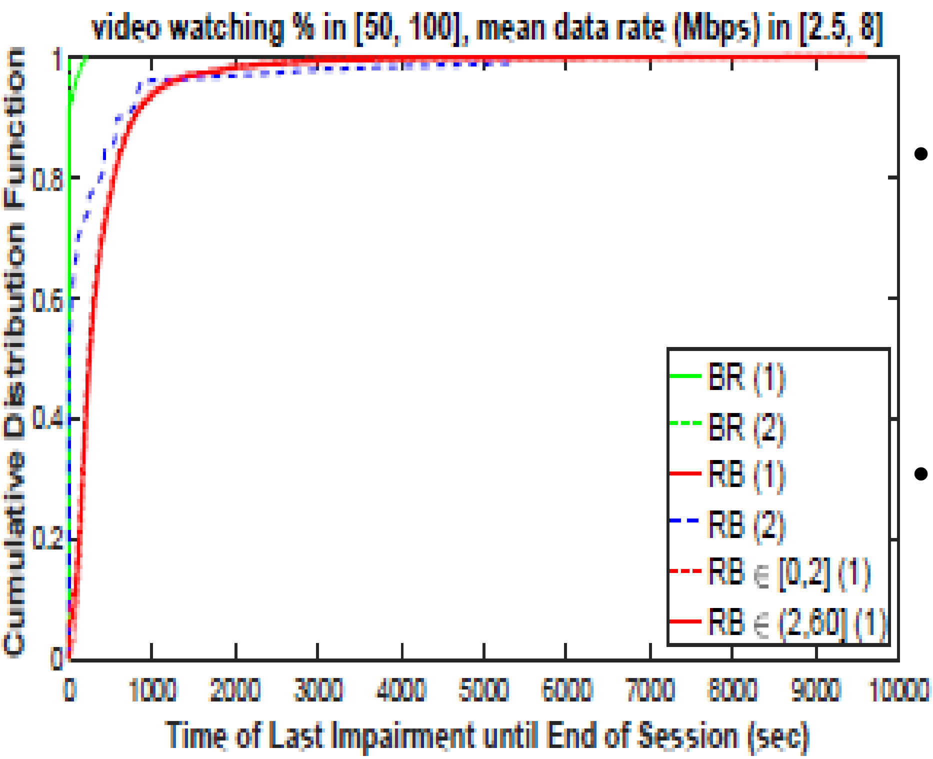
The dominant parameters are the ones with non-zero LASSO coefficients for λ in [0.4, 0.6].

An impairment prior to a BR- increases the likelihood of abandonment

Scenarios examined:

- Exactly one impairment being BR-
- Exactly two impairments the second being BR- , the first being RB or BR

Time margin between them ≥ 30 sec → they are perceived as different impairments



- User that experiences a BR- as second impairment is more prone to abandon the session than when experiences a BR- for the first time
- BR- more direct impact than RB

Conclusions

The "worst offenders" for the design of adaptation algorithm are:

- the combination of BR- and RB
- the large RB ratio
- the combination of BR+ and low resolution
- and the BR- change of two or more levels

Reporting average statistics, such as average bitrate per session, is not enough

We need to monitor RB and BR changes at the user level to quantify and improve the user engagement

per-user statistics about the revisit and viewing duration per video, info about the user device, context (e.g., time-of-the-day, position), content type to improve not only the adaptation but also caching

→ user-centric ABR player

Future Work

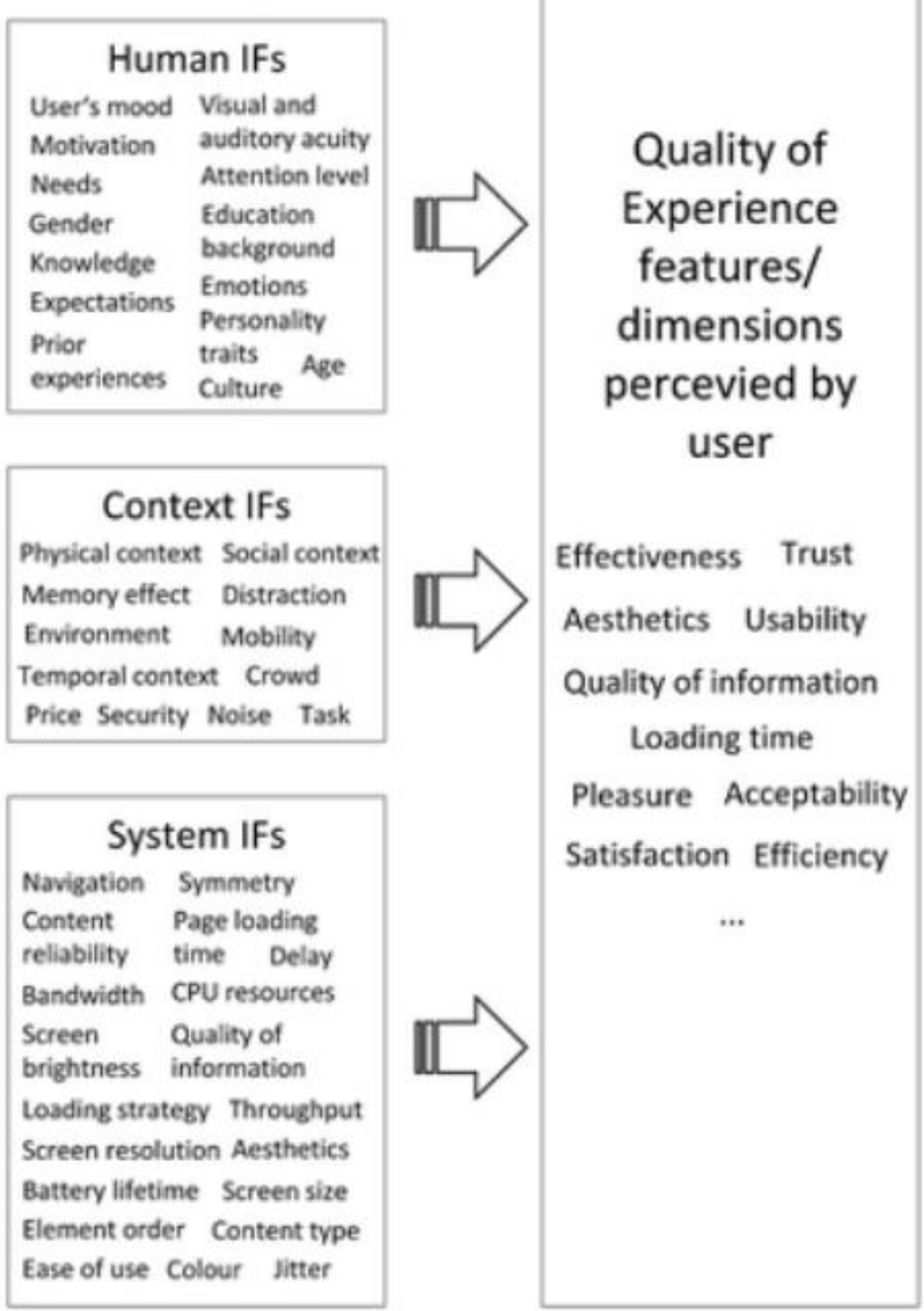
Collect and integrate information about:

- the content type
- the interest of the user in the video content
- whether or not the user manually changed the bitrate during a playback

Also, specific cases with relatively small number of sessions or the presence of confounding variables , such as

- early and late BR- sessions
- sessions with RB larger than 20sec
- sessions with one BR- in [1, 2]Mbps

do not allow us to draw definite conclusions


Human IFs
User's mood
Motivation
Needs
Gender
Knowledge
Expectations
Prior experiences
Visual and auditory acuity
Attention level
Education background
Emotions
Personality traits
Age
Culture
Context IFs
Physical context
Social context
Memory effect
Distraction
Environment
Mobility
Temporal context
Crowd
Price
Security
Noise
Task
System IFs
Navigation
Symmetry
Content reliability
Page loading time
Delay
Bandwidth
CPU resources
Screen brightness
Quality of information
Loading strategy
Throughput
Screen resolution
Aesthetics
Battery lifetime
Screen size
Element order
Content type
Ease of use
Colour
Jitter
Quality of Experience features/ dimensions percevied by user
Effectiveness
Trust
Aesthetics
Usability
Quality of information
Loading time
Pleasure
Acceptability
Satisfaction
Efficiency
...

QoE for AR/VR Services

- Physiological data, e.g., eye measurements, electrodermal activity (EDA), EEG, heart-rate

- Characteristics of the perceptual and cognitive processes, such as effort required, response time, errors, and interaction
- No agreed methods or benchmarks for accessing the QoE in AR (G.QoE-AR, ITU-T work program)
- Borrowed aspects from ITU-T standards, integrating physiological measurements

Mobile Cloud Gaming Framework

- Shift mobile user load to cloud server due to the inherent hardware constraint of mobile devices (memory and graphics processing)
- Objective factors that impact on the QoE :

  cloud server, source video, wireless network & client
- subjective QoE:

Game Mean Opinion Score (GMOS) for measurement of end user's QoE

Applications	QoE Parameters	Future considerations
Multimedia services	AQoS (e.g. codec,frame rate) NQoS (e.g. bitrate)	monitoring of client device, design of algorithms for analysis of accurate QoE and network policy change
Network services	NQoS (Packet loss & reorder)	SLAs, automatic network monitoring, dynamic policy
VoIP	NQoS (delay, etc) & AQoS (audio codec, etc)	QoE performance parameters per service type
Web development	NQoS (loading time)	New protocols such as Multi-Path Transmission Control Protocol (MPTCP)
Games	NQoS, AQoS, PSNR & VGA	speculation-based technology
Cloud	NQoS & AQoS (data retrieval)	SLA, sentiment analysis

Additional References

- https://www.researchgate.net/profile/Arslan_Ahmad3/publication/331013590_QoE--aware_Multimedia_Service_Management_and_Monitoring_through_OTT_and_ISP_Collaboration/links/5c61455292851c48a9c98a46/QoE--aware-Multimedia-Service-Management-and-Monitoring-through-OTT-and-ISP-Collaboration.pdf
- https://link.springer.com/content/pdf/10.1007%2Fs11213-018-9471-x.pdf
- https://link.springer.com/article/10.1007/s41233-017-0009-2