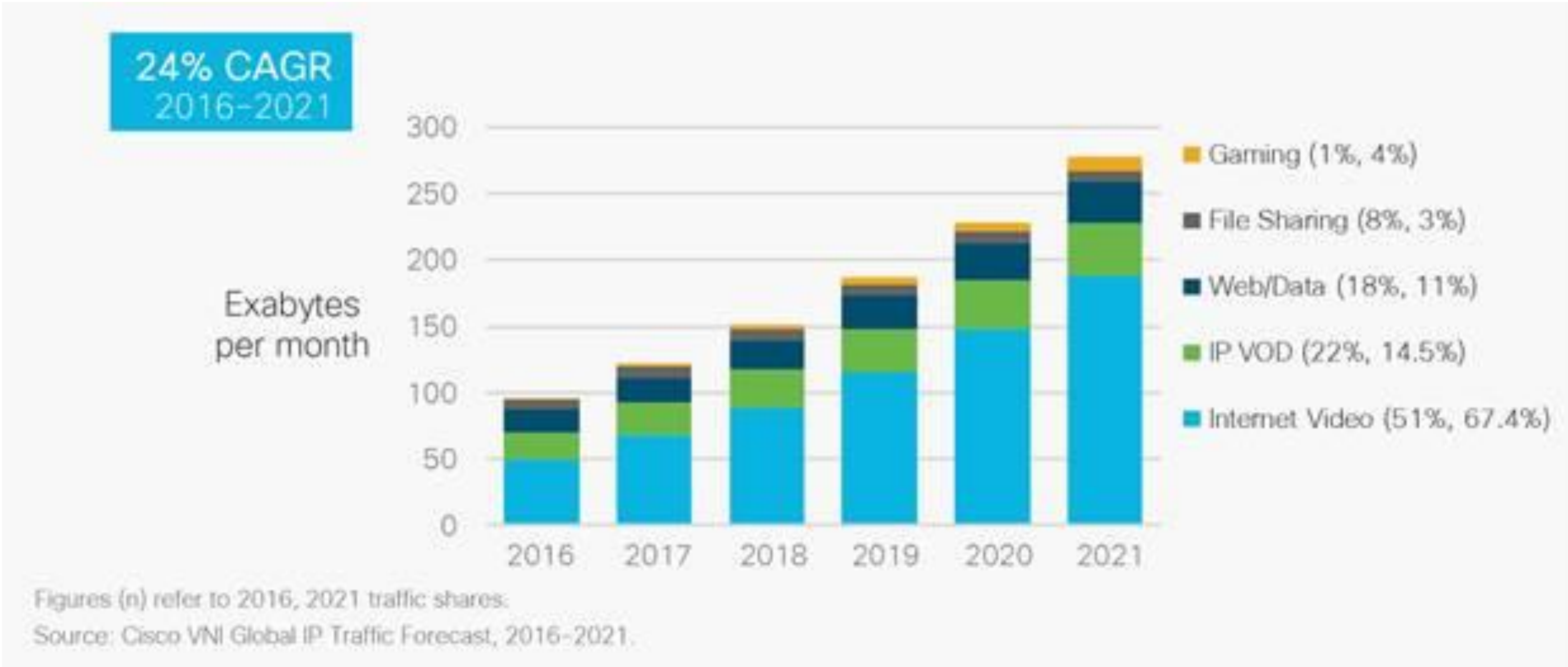
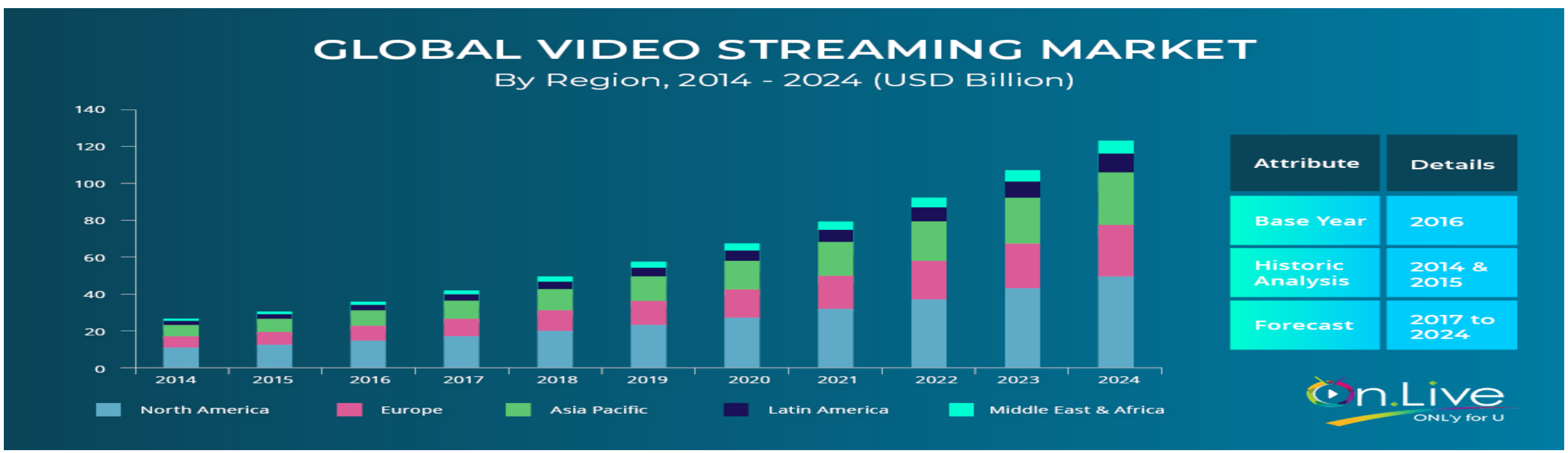
Bridging the gap between network measurements and quality of experience: the video streaming case

Chadi.Barakat@inria.fr
http://team.inria.fr/diana/chadi/

Workshop on Systems (WOS), Rennes
October 12th, 2021

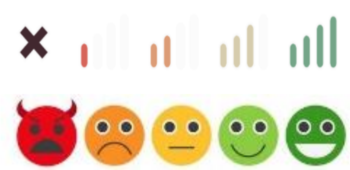
Video streaming: The Internet service





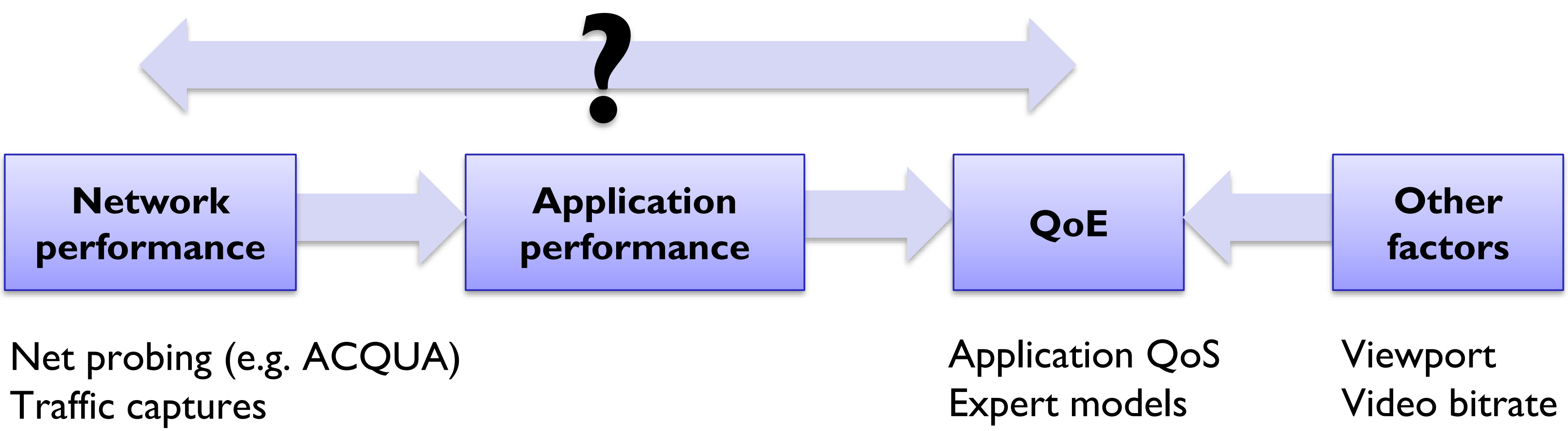
How well it performs?

- This depends on many factors
  - Network performance – bandwidth, delay, packet loss
  - Protocols – DASH and its variants, TCP vs QUIC, HTTP1/2/3, caching
  - Content – encoding, video category, chunking
  - Context – mobile, landline, outdoor, indoor
- Quality of end-user experience
  - Subjective measurement towards end users – Mean Opinion Score
  - Objective measurement – bitrate, join time, stalls, resolution switches
  - Expert models
    - QoE = function (application_level_QoS)
    - Calibrated with MOS measurements
    - Example: ITU-T P.1203 Recommendation (score from 1 to 5)

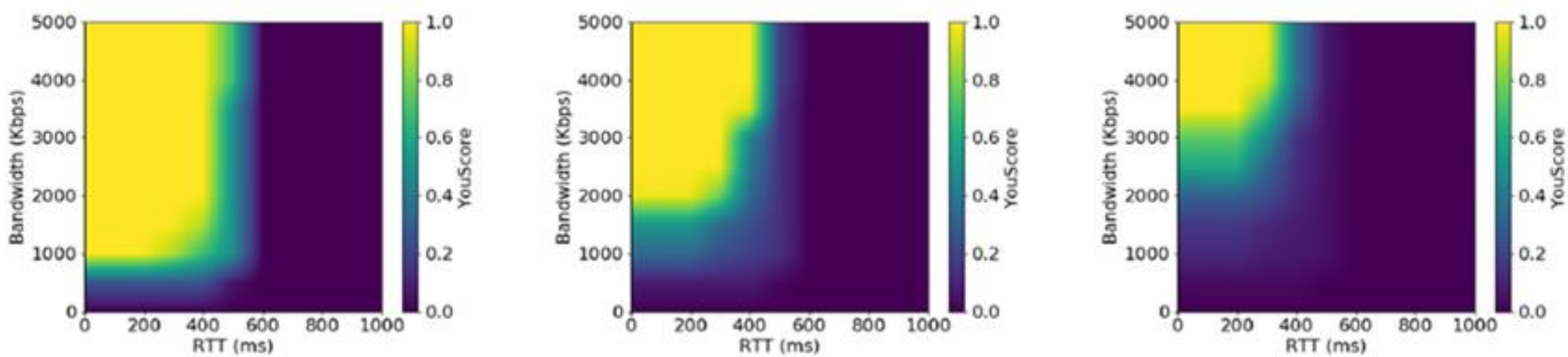


Objectives

- ❑ Data-driven models relating user-level experience to network performance – mainly video streaming, but also audio and web
  - For ISPs and CPs: better view on user experience, and better network management (traffic engineering, troubleshooting, provisioning)
  - For users: improved transparency, diagnosis, forecasting



To give an idea
YouScore: likelihood of video interruption



(a) 480p (b) 720p (c) 1080p

	4G	3G	2G	LAN	WLAN
$YouScore_{240p}$	0.97	0.82	0.05	0.93	0.92
$YouScore_{360p}$	0.95	0.76	0.04	0.89	0.88
$YouScore_{480p}$	0.92	0.65	0.02	0.83	0.83
$YouScore_{720p}$	0.82	0.43	0.01	0.68	0.67
$YouScore_{1080p}$	0.68	0.22	0.01	0.51	0.49

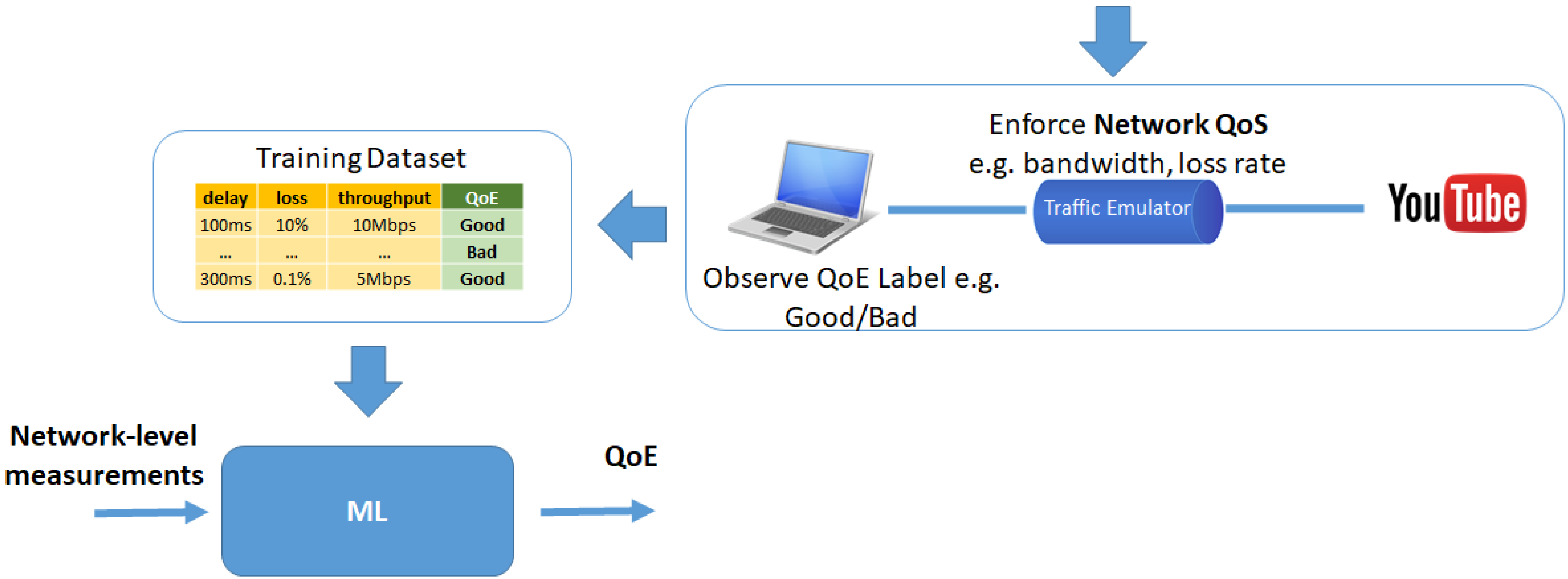
Network performance data from RTR-NetTest

M.J. Khokhar, T. Spetebroot, C. Barakat, "A Methodology for Performance Benchmarking of Mobile Networks for Internet Video Streaming", in proceedings of ACM MSWIM, 2018.

Data-driven approach
Controlled experimentation and machine learning

A pool of network conditions to emulate

delay	loss	throughput	QoE
100ms	10%	10Mbps	?
...	...	...	?
300ms	0.1%	5Mbps	?

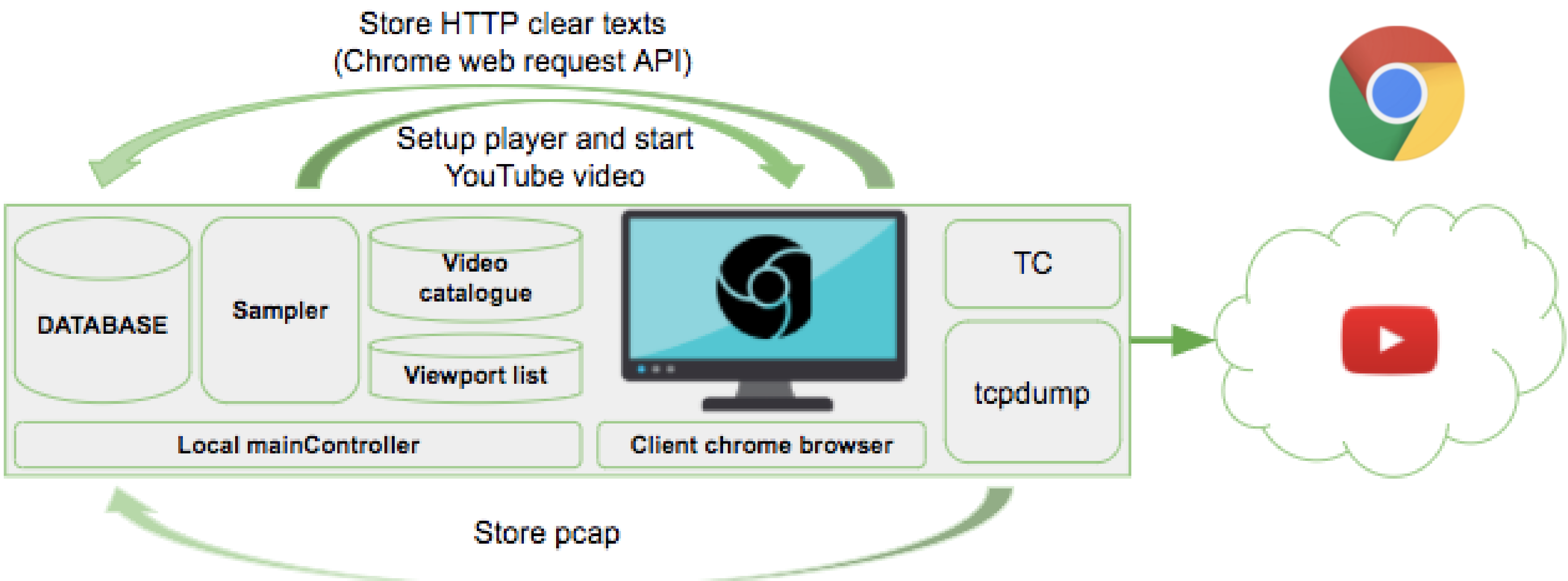


Training Dataset

delay	loss	throughput	QoE
100ms	10%	10Mbps	Good
...	...	...	Bad
300ms	0.1%	5Mbps	Good

Experimentation framework

- ❑ Chrome-based automatic Youtube playout + network emulation
- ❑ Dumping of streaming events (Chrome API) + video traffic pcap



- ❑ Video catalog: one million trending Youtube videos (> 720p)

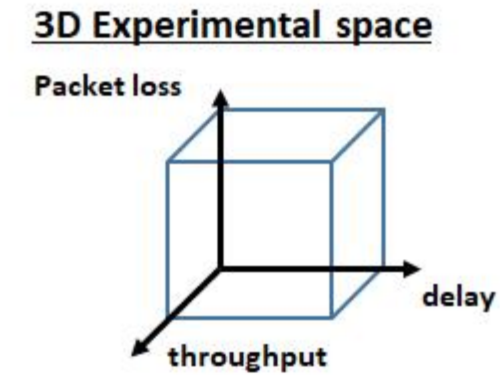
M. J. Khokhar, T. Ehlinger, C. Barakat, "From Network Traffic Measurements to QoE for Internet Video", in proceedings of IFIP Networking Conference, 2019.

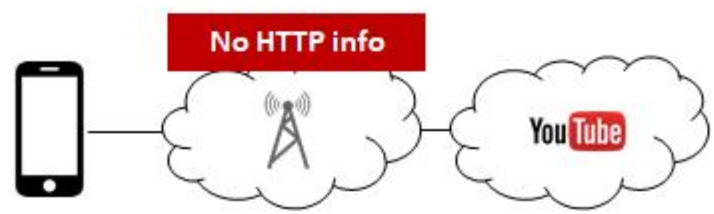
Case studies

- Predictive models for QoE
  - From network to application, anticipate the experience
  - Out-of-band network measurements as input
- Estimation models of QoE
  - From network to application, estimate the quality of experience
  - Encrypted in-band application traffic as input
- Estimation models of network performance
  - From application behavior to network performance, avoid probing

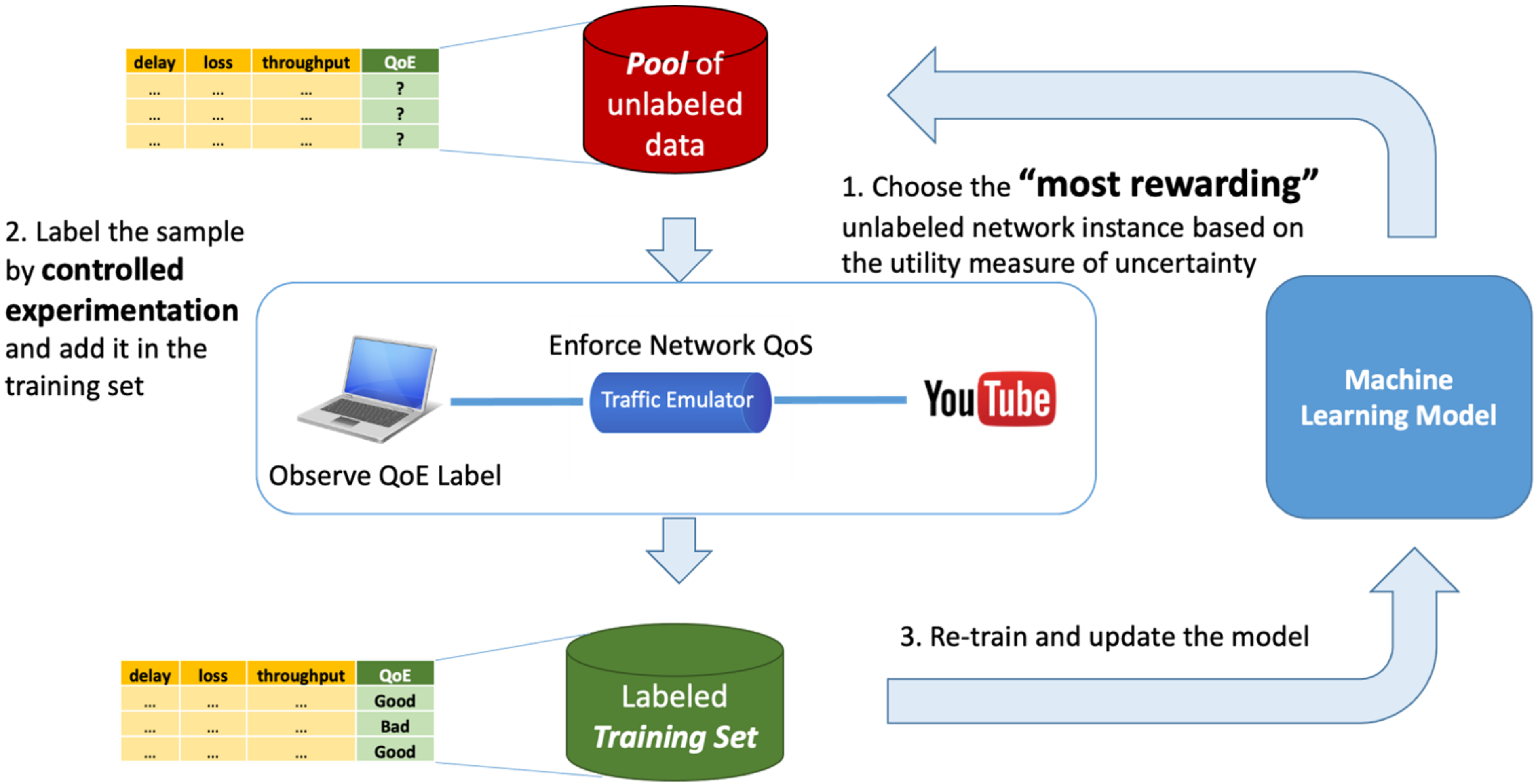
Among the challenges

- ❑ The **large** experimental space to cover in **controlled experimentation**
- ❑ Data **acquisition**
- ❑ Traffic **encryption**
- ❑ Content **diversity**
- ❑ **Complexity** of intermediate protocols (DASH, HTTP, TCP, etc)





Intelligent experimentation framework based on active learning



Utility measures for choosing the most rewarding sample from the Pool

Model's classification probability for each output label/class

Pool of Unlabeled Data

	$P(\hat{y}^{(1)})$	$P(\hat{y}^{(2)})$	$P(\hat{y}^{(3)})$	$P(\hat{y}^{(4)})$	$P(\hat{y}^{(5)})$
$x^{(1)}$	0.1	0.35	0.1	0.2	0.25
$x^{(2)}$					
...					
$x^{(p)}$					

$$\sum_i P(\hat{y}^{(i)}) = 1$$

Least Confident: $\text{argmin}_x P(\hat{y}_{max})$

Minimal Margin: $\text{argmin}_x [P(\hat{y}_{max1}) - P(\hat{y}_{max2})]$

Maximum Entropy: $\text{argmax}_x - \sum_y P(y) \log P(y)$

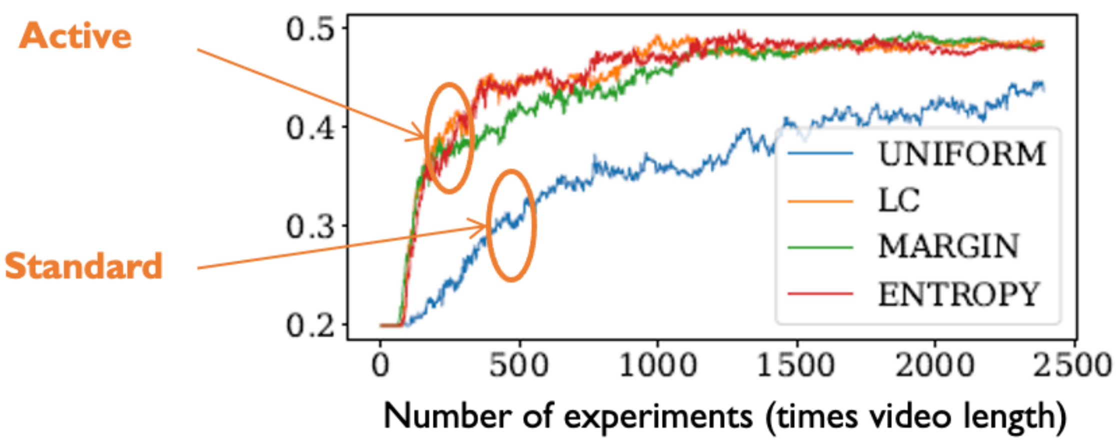
Experiment less, model faster

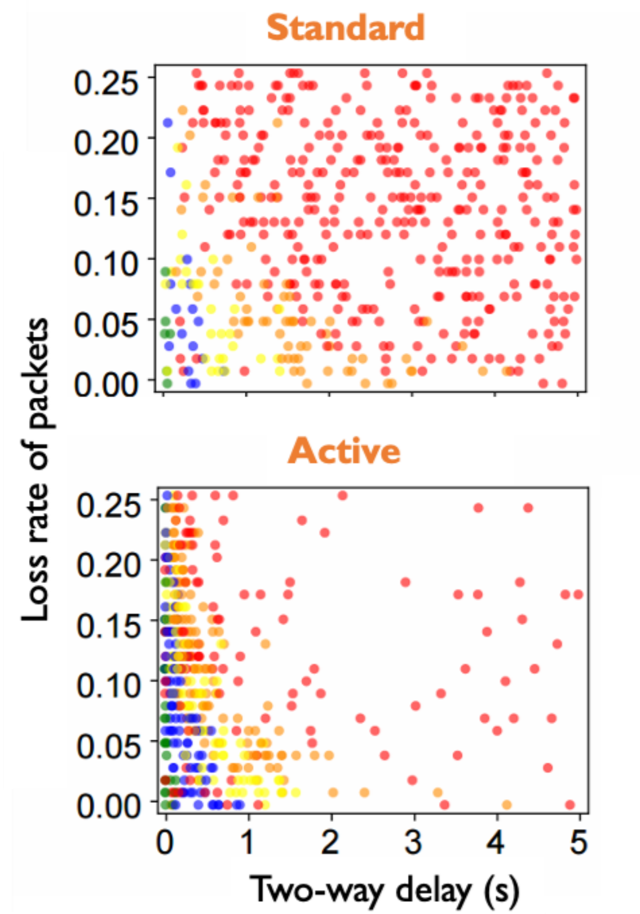
Fixed resolution 720p

Expert QoE model ~ $e^{-\textit{stall\_duration}}$

1 - Poor 2 - Bad 3 - Fair 4 - Good 5 - Excellent

Classification accuracy – Decision trees





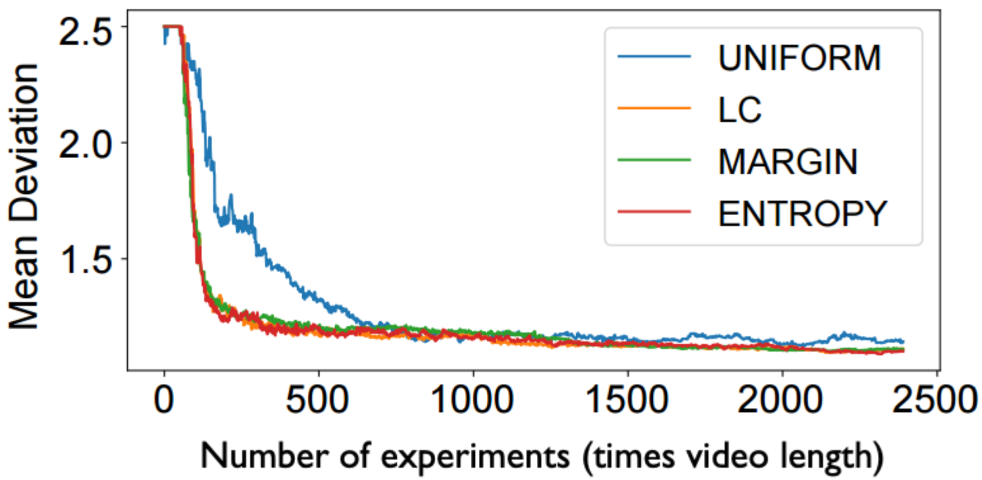
M. J. Khokhar, N.A. Saber, T. Spetebroot, C. Barakat, "An Intelligent Sampling Framework for Controlled Experimentation and QoE Modeling", in Computer Networks, vol. 147, pp. 246-261, December 2018.

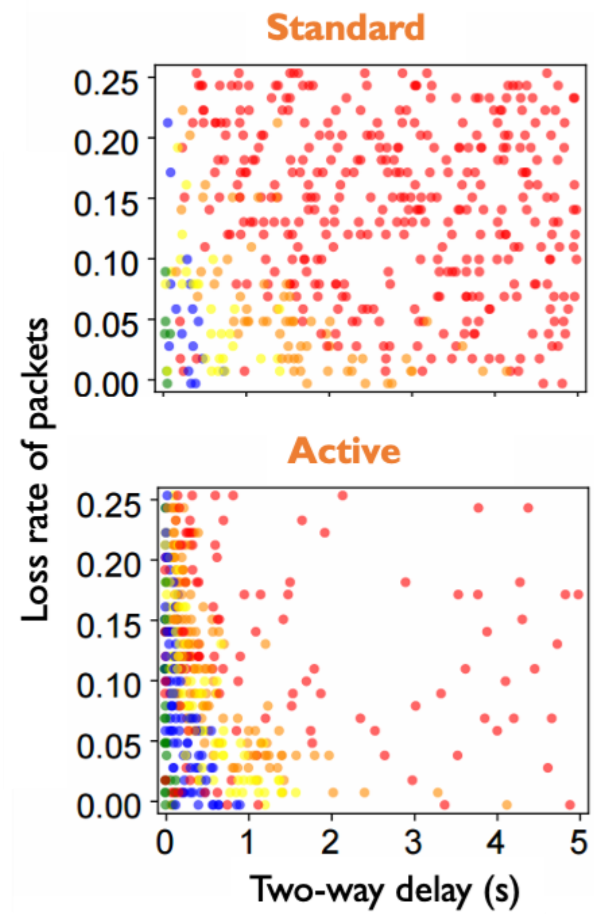
Experiment less, model faster

Fixed resolution 720p

Expert QoE model ~ $e^{-\,stall\_duration}$

1 - Poor 2 - Bad 3 - Fair 4 - Good 5 - Excellent



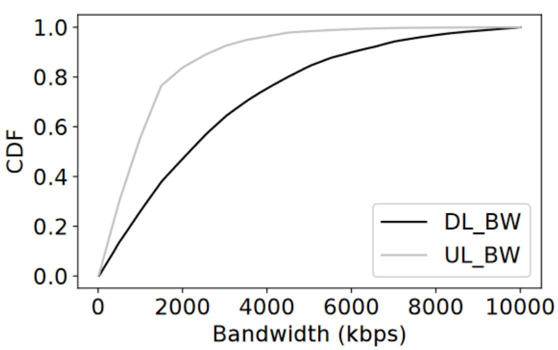


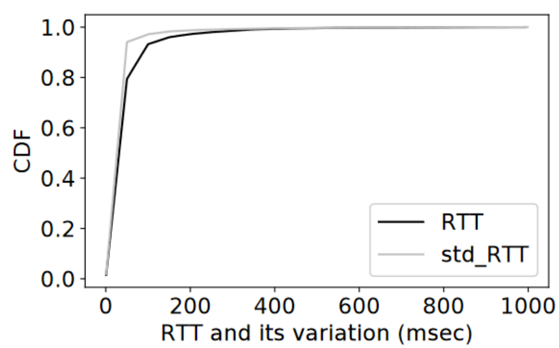
M. J. Khokhar, N. A. Saber, T. Spetebroot, C. Barakat, "An Intelligent Sampling Framework for Controlled Experimentation and QoE Modeling", in Computer Networks, vol. 147, pp. 246-261, December 2018.

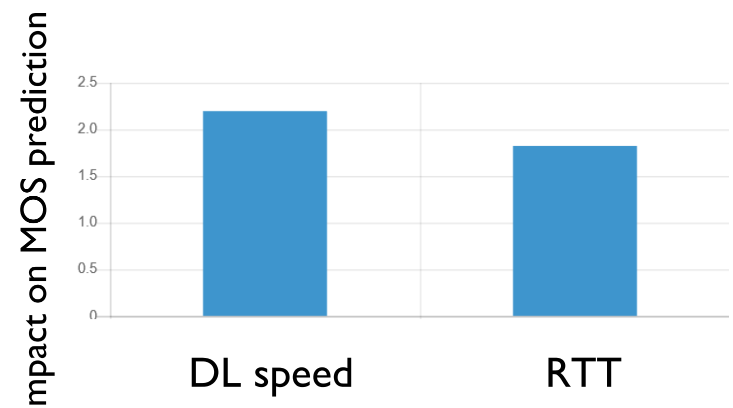
Network perf metrics: the out-of-band case

- A total of seven network metrics, enforced with ' tc '
  - Input features for QoE prediction
- Sampled from empirical datasets
  - Active measurement
  - ACQUA (see next)
  - RTR-NetTest and MobiPerf
- ~ 100K streaming experiments
  - Different network instances
  - Random videos

RTT
Download loss rate
Upload loss rate
Download jitter
Upload jitter
Download throughput
Upload throughtput

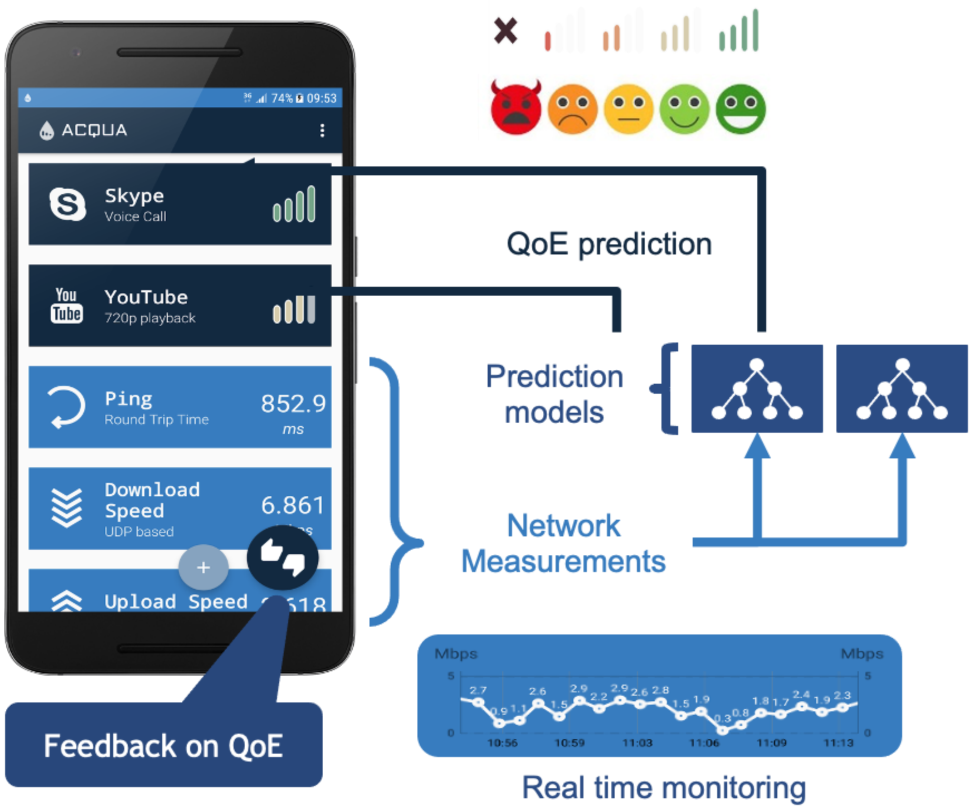


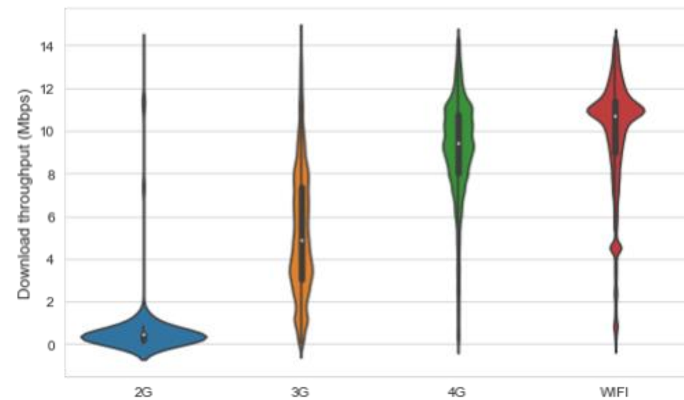




The ACQUA mobile app

http://project.inria.fr/acqua/





~ 3 million network snapshots

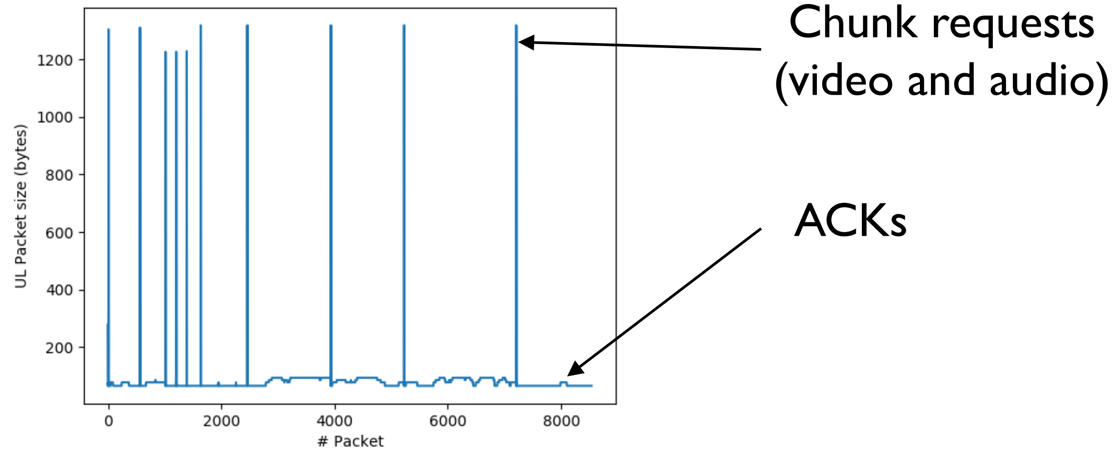
O. Belmoukadam, T. Spetebroot, C. Barakat, "ACQUA: A user friendly platform for lightweight network monitoring and QoE forecasting", in proceedings of QoE Management Workshop, 2019.

Network perf metrics: the inband case

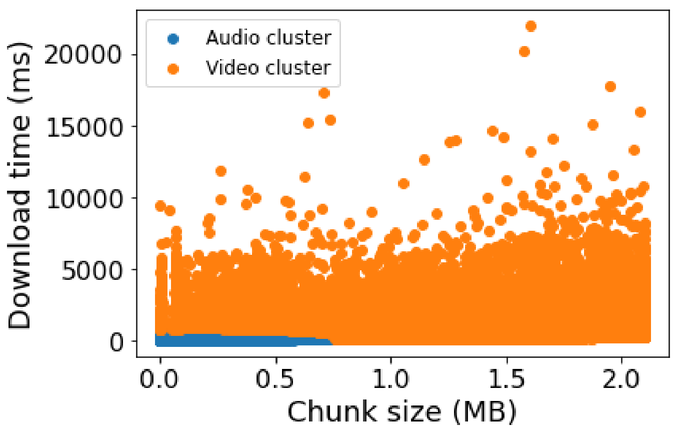
- Features extracted from encrypted video traffic traces
- Ground-truth on QoE from within the browser
- Packet-level metrics
  - **DL throughput** [avg, max, standard deviation, percentiles (10^{th} to 90^{th} in steps of 10)]
  - **DL interarrival times** [avg, max, standard deviation, percentiles (10^{th} to 90^{th} in steps of 10)]
  - **UL interarrival times** [avg, max, standard deviation, percentiles (10^{th} to 90^{th} in steps of 10)]
  - **DL packet sizes** [avg, max, standard deviation, percentiles (10^{th} to 90^{th} in steps of 10)]
- Chunk-level metrics
  - **Chunk sizes** (avg, max, standard deviation, minimum, 25^{th}, 50^{th} ,75^{th} percentiles)

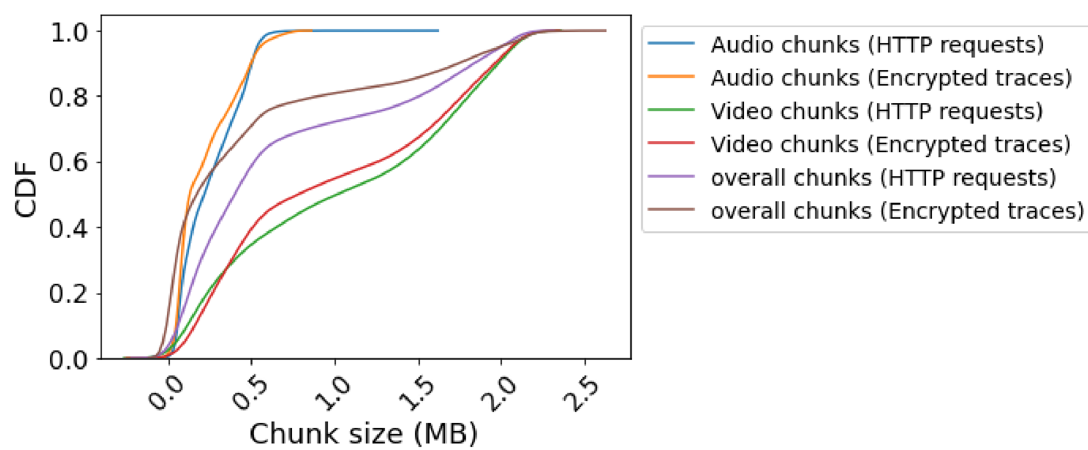
Identifying video chunks from encrypted traffic

- K-means to separate requests from ACKs
- Chunk = DL data between two consecutive requests



- GMM (Gaussian Mixture Model) to separate audio and video chunks

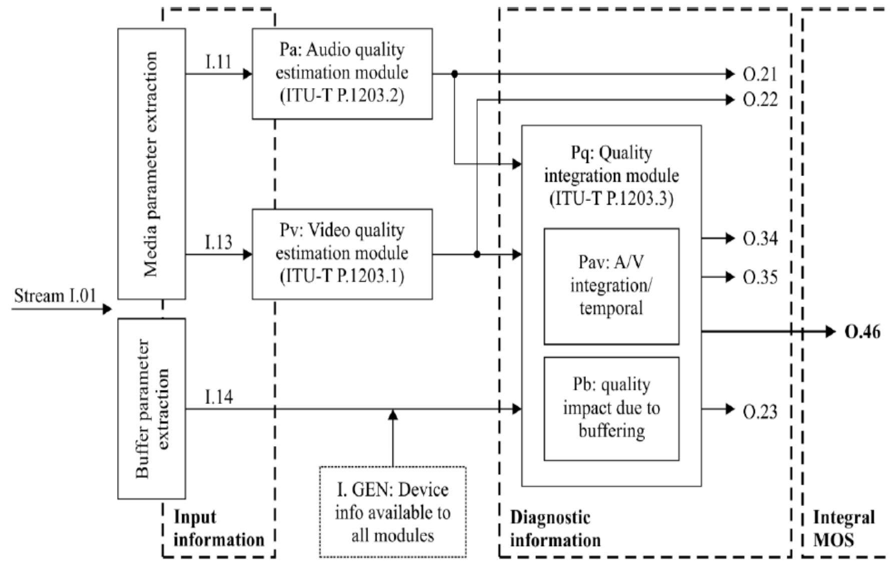


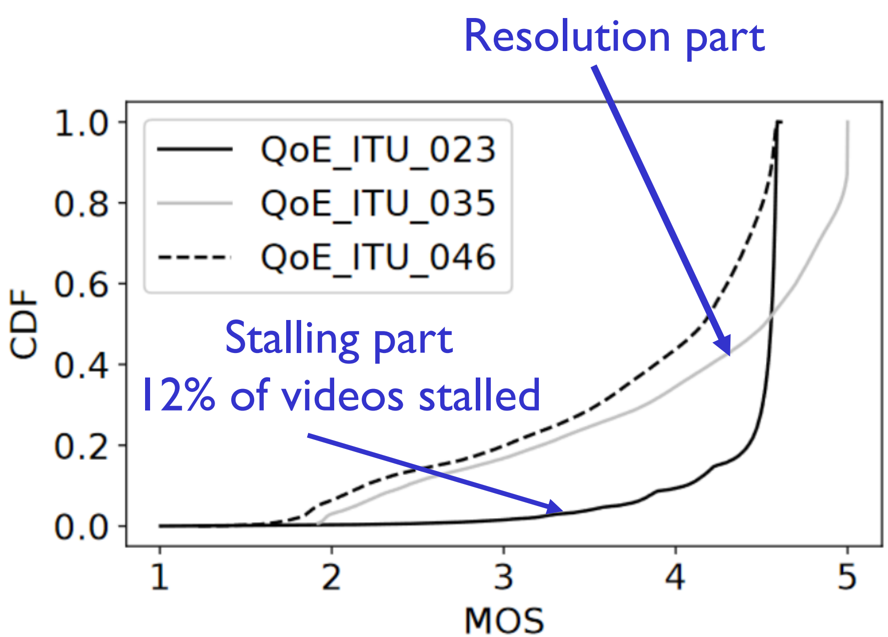


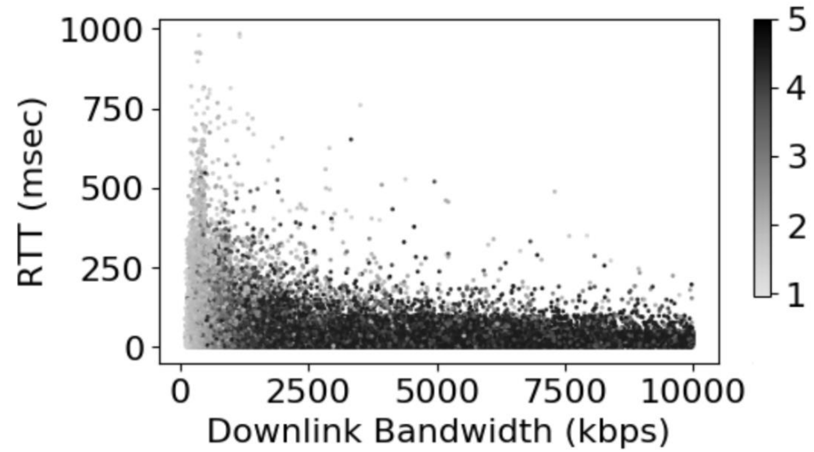
ITU P.1203 model for Video QoE

- Meta Data for each chunk required to estimate the final MOS (O.46):
  - Bitrate, Codec, Duration, Frame rate, Resolution (O.35)
  - Buffering/stall timestamps and their durations (O.23)

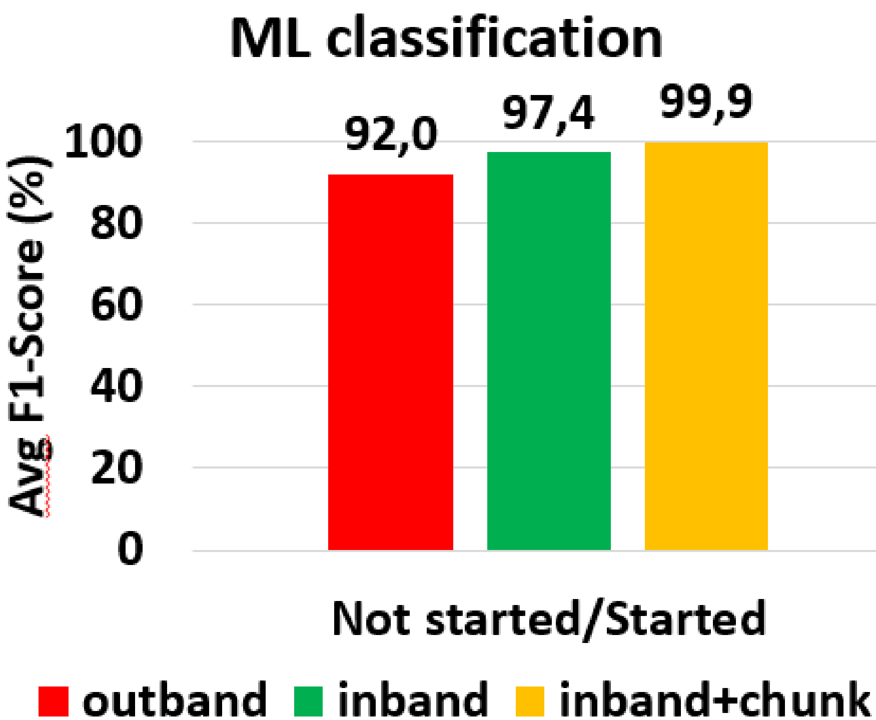
[Building blocks of the ITU-T P.1203 model]()

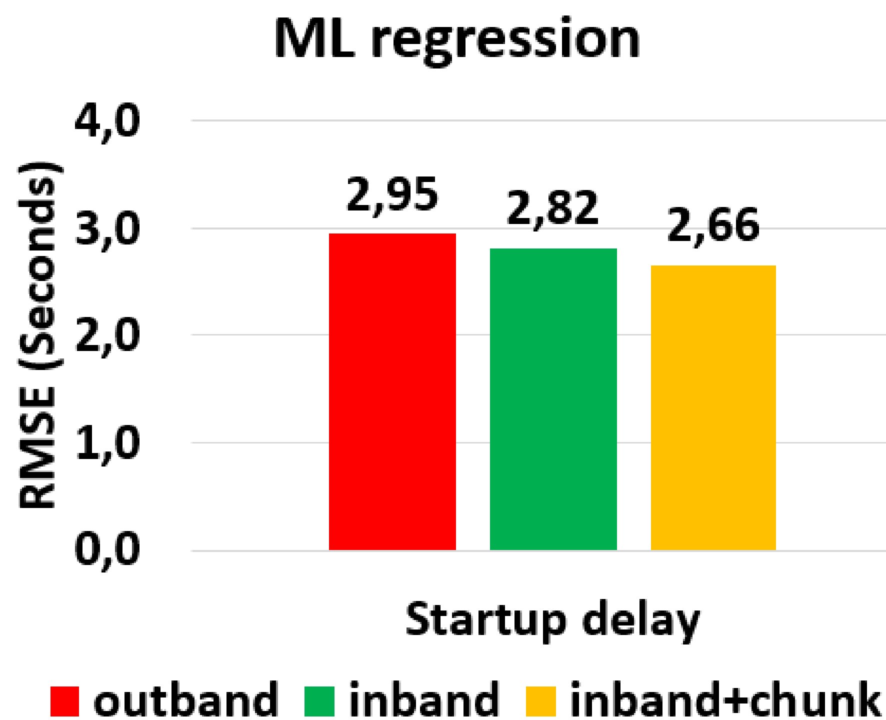






Predicting start-up delay

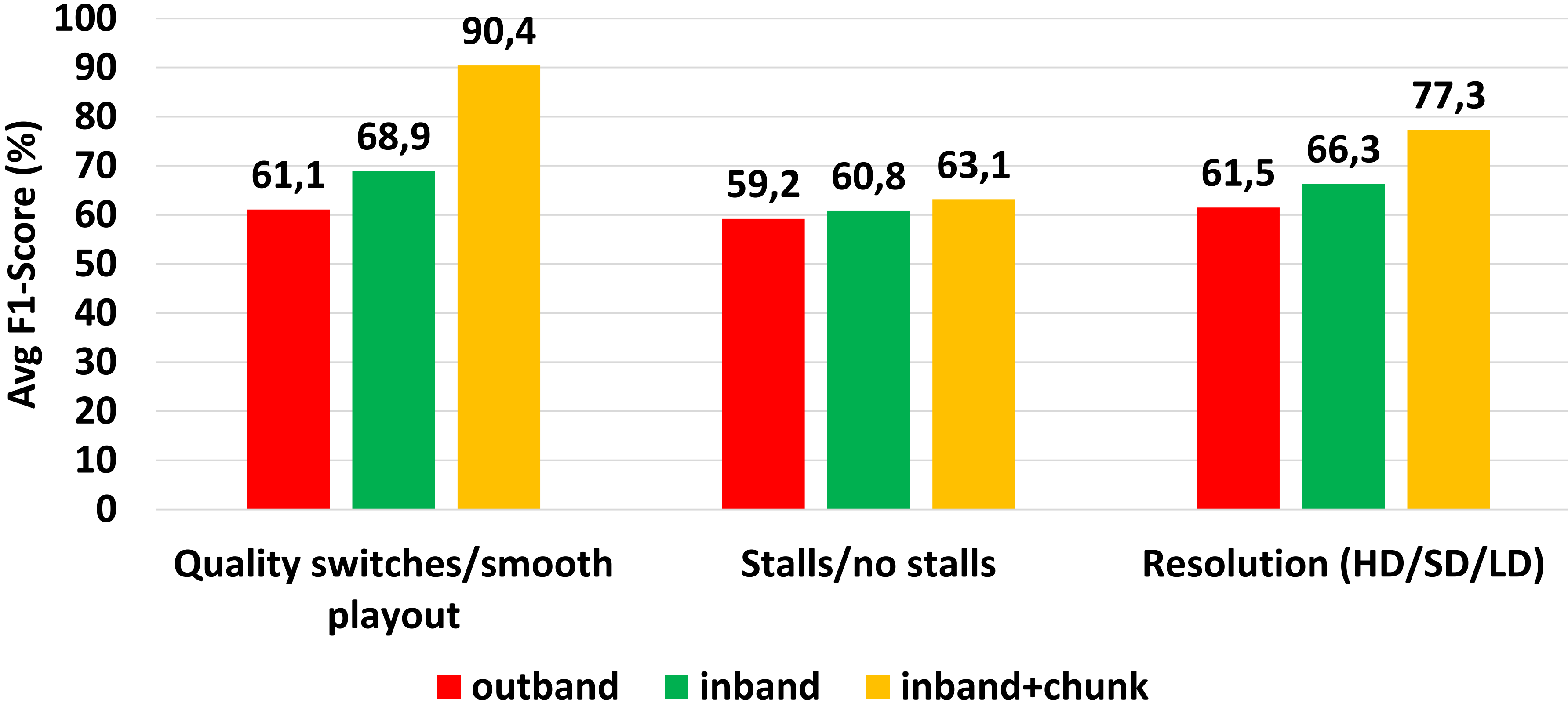




* Random Forests with default configuration

M. J. Khokhar, T. Ehlinger, C. Barakat, "From Network Traffic Measurements to QoE for Internet Video", in proceedings of IFIP Networking Conference, 2019.

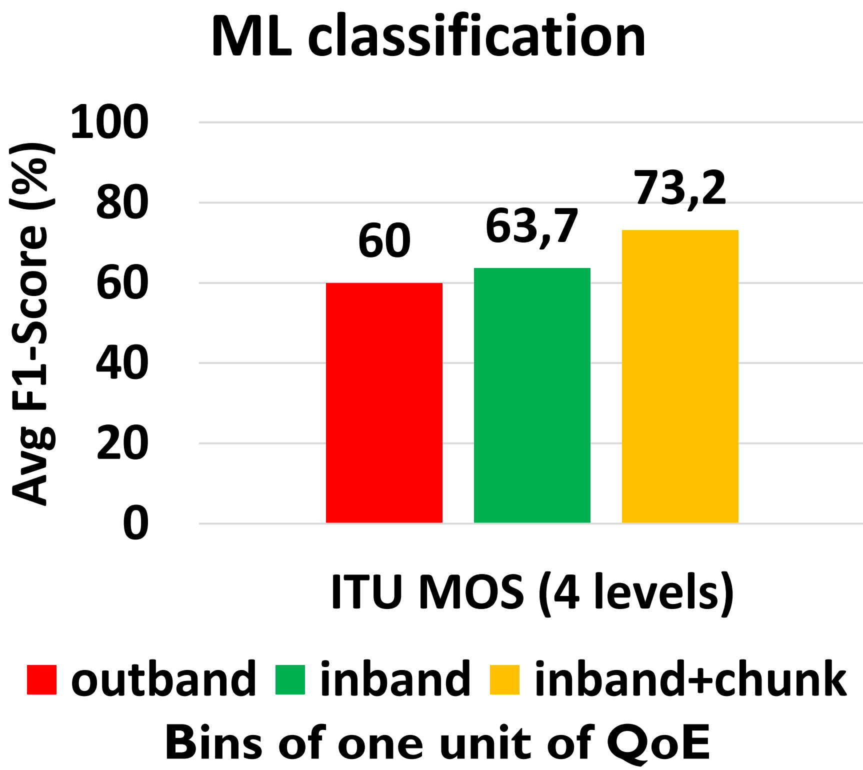
Predicting switches, stalls and resolution

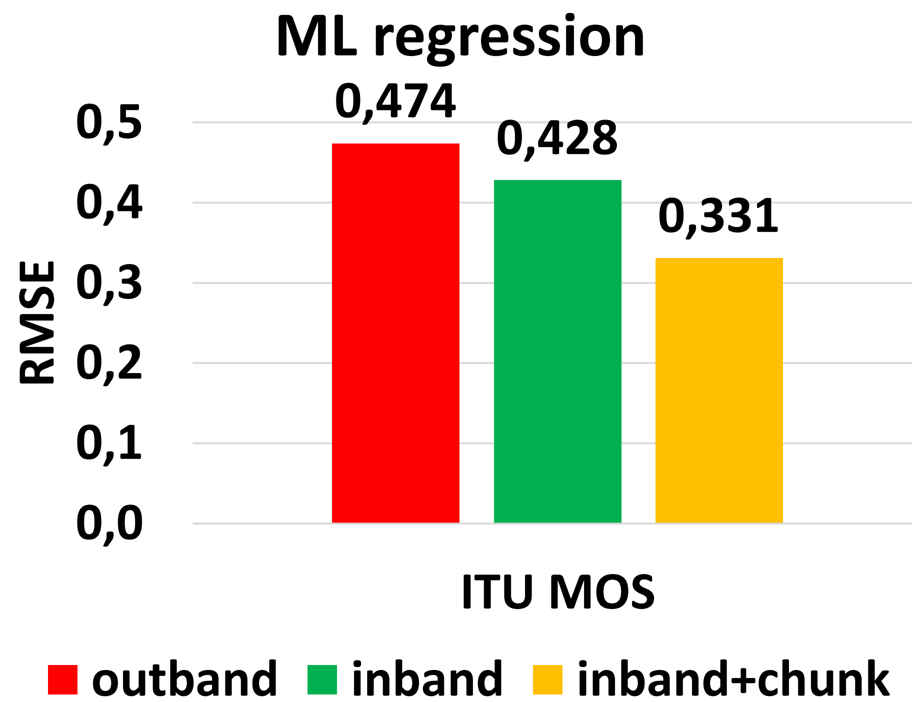


* Random Forests with default configuration

M. J. Khokhar, T. Ehlinger, C. Barakat, "From Network Traffic Measurements to QoE for Internet Video", in proceedings of IFIP Networking Conference, 2019.

Predicting ITU MOS

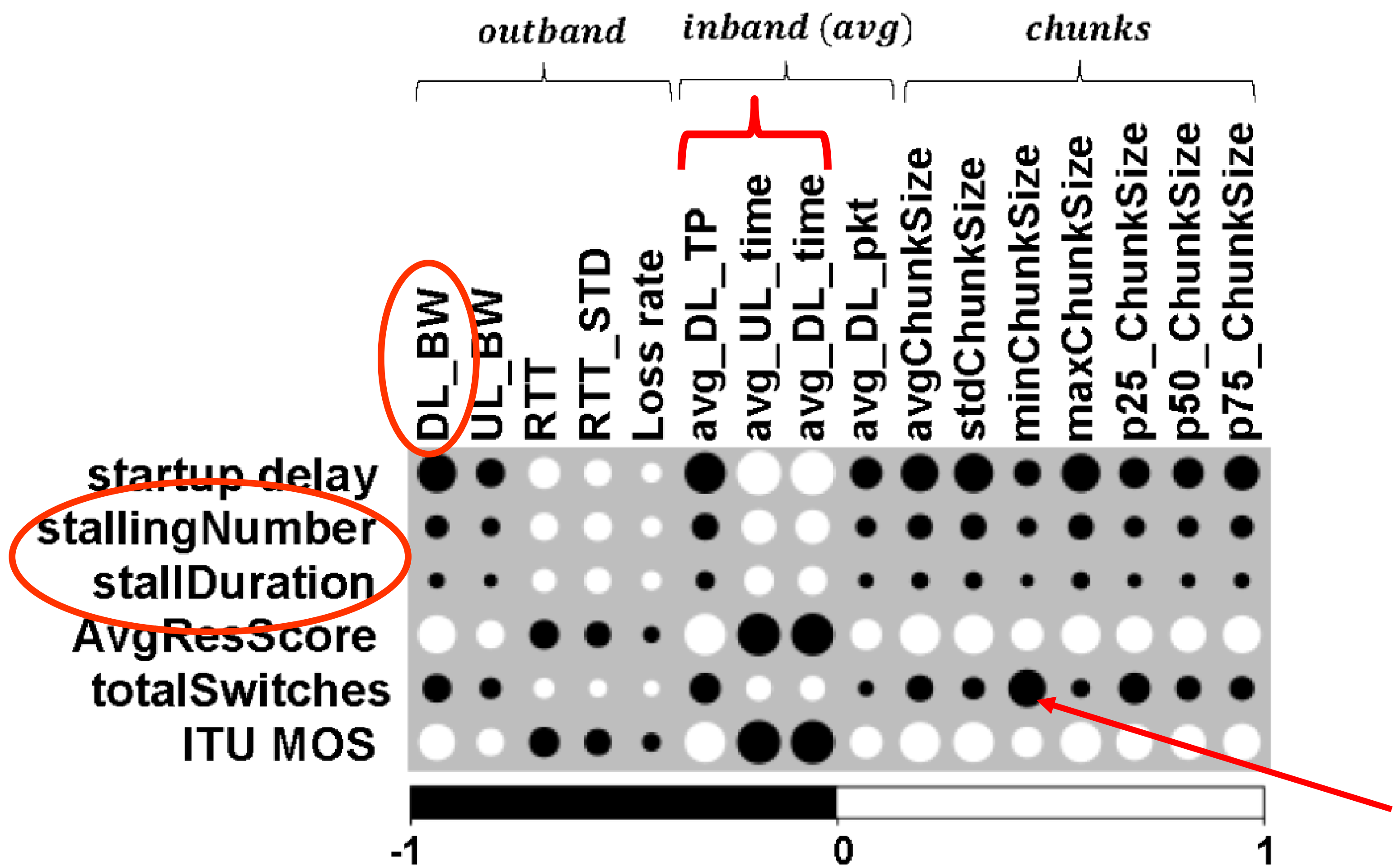




* Random Forests with default configuration

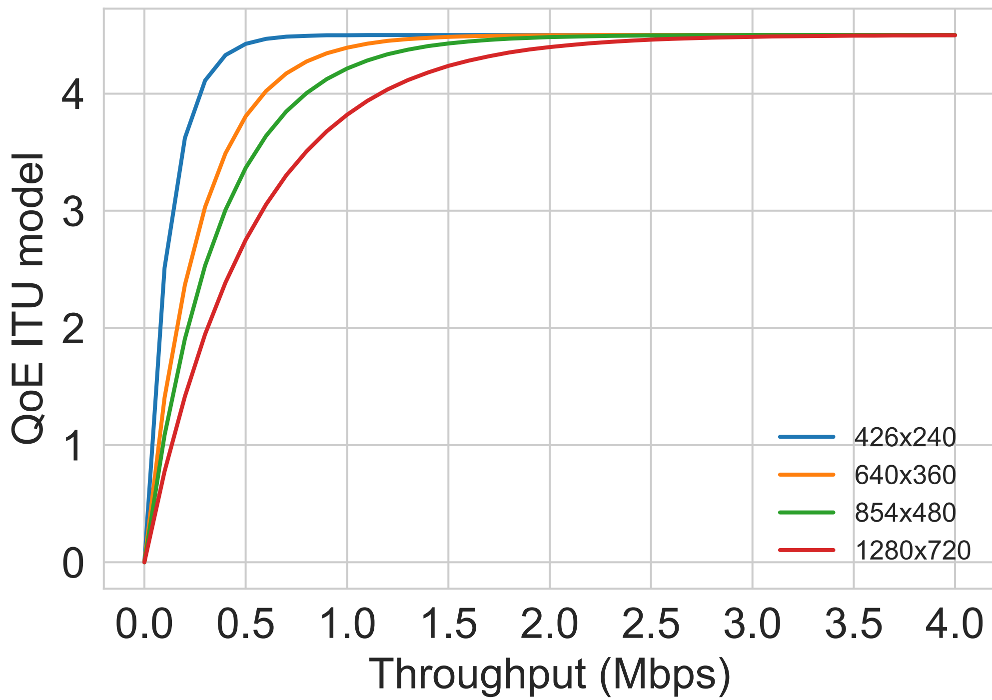
M. J. Khokhar, T. Ehlinger, C. Barakat, "From Network Traffic Measurements to QoE for Internet Video", in proceedings of IFIP Networking Conference, 2019.

Correlogram for network QoS, app QoS & MOS



QoE is also a matter of viewport

Part of the screen where the video is played out

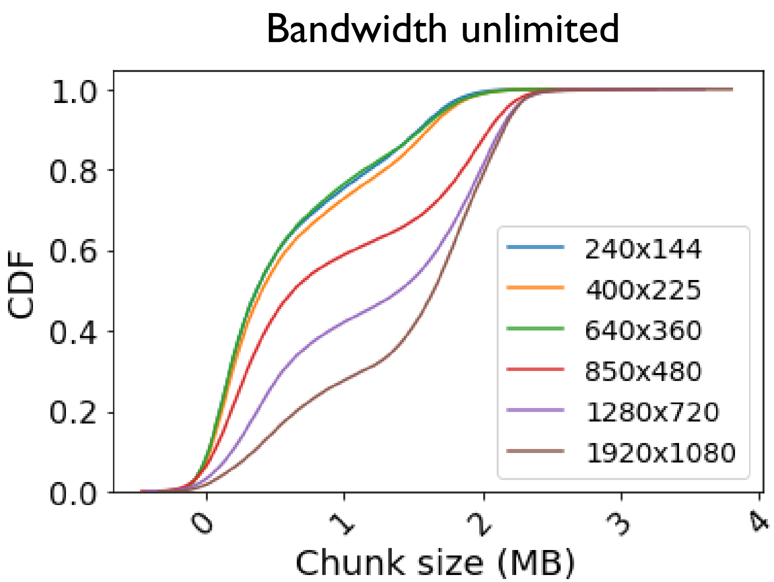


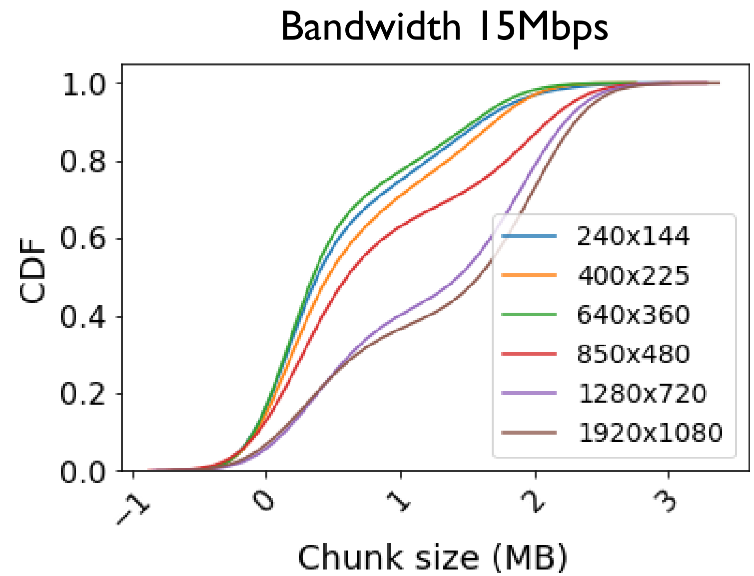
$$QoE_s \underset{exp}{=} QoE_{max}(1 - e^{-\beta_s x})$$

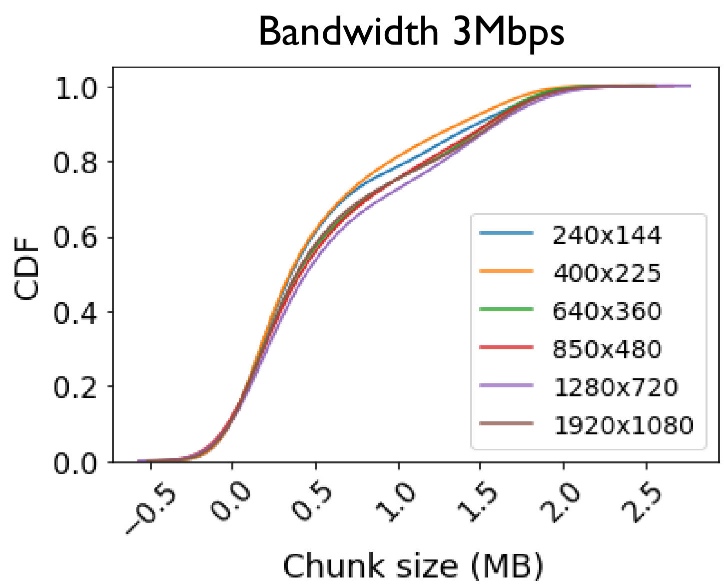
Screen resolution s	Estimated parameter β
426x240	8.17
640x360	3.73
850x480	2.75
1280x720	1.89

O. Belmoukadam, M. J. Khokhar, C. Barakat, "On Accounting for Screen Resolution in Adaptive Video Streaming: A QoE-Driven Bandwidth Sharing Framework", in proceedings of CNSM, 2019.

Viewport and network traffic







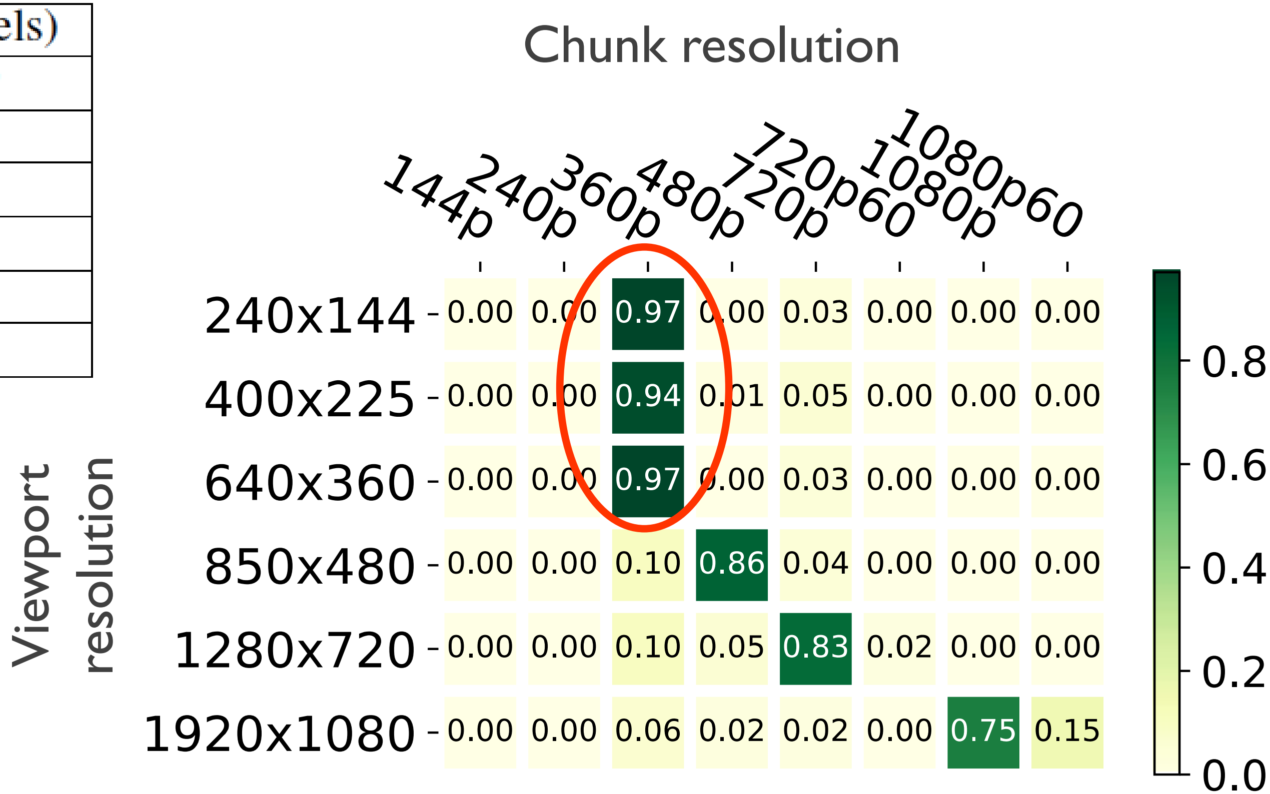
Bandwidth unlimited

→ →

Is the viewport well respected?

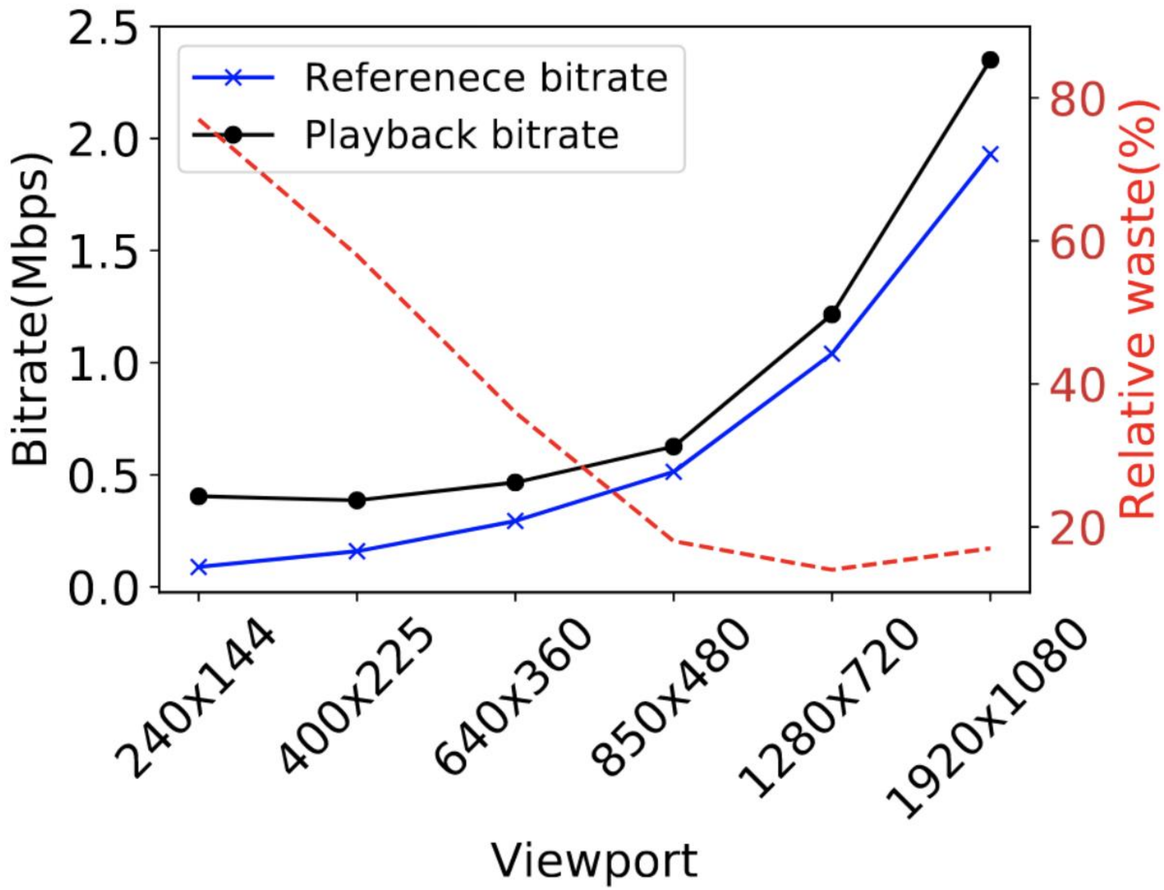
Bandwidth waste = downloaded resolutions finer than the viewport

Video resolution	Viewport (pixels)
1080p	1920x1080
720p	1280x720
480p	850x480
360p	640x360
240p	426x240
144p	240x144

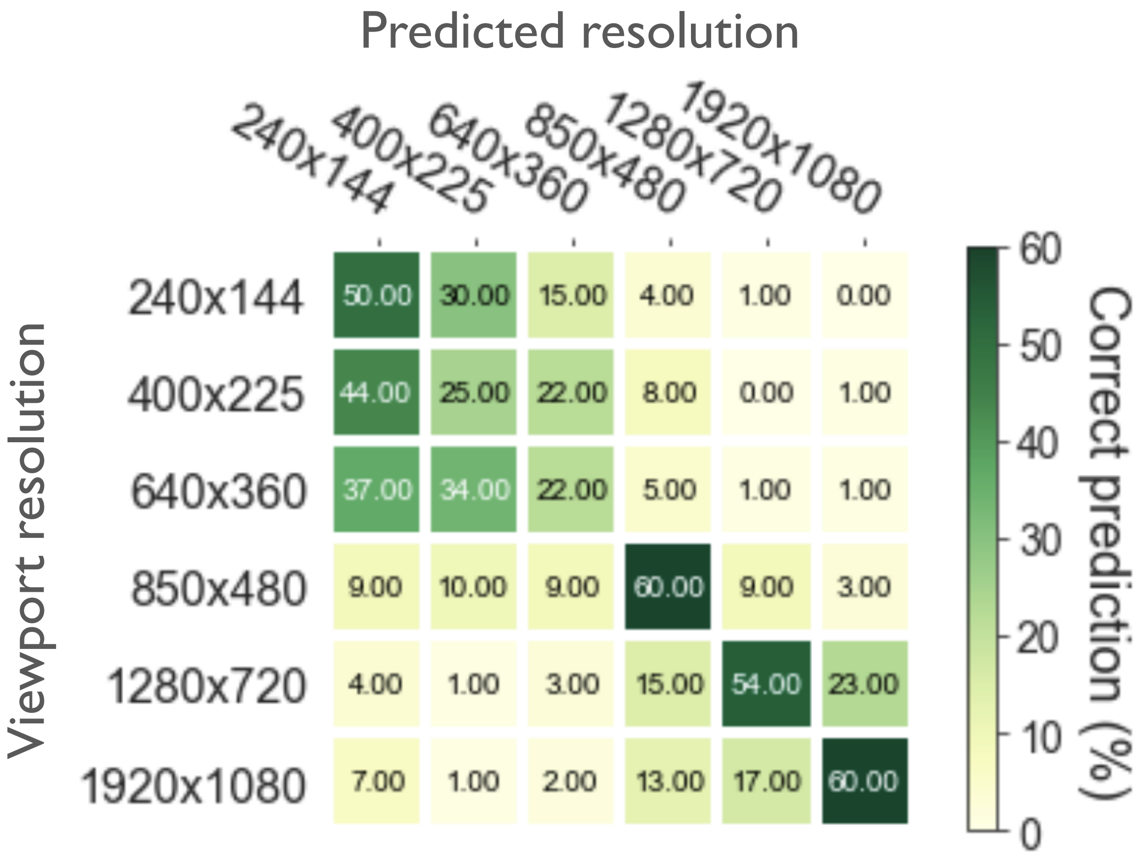


O. Belmoukadam, M. J. Khokhar, C. Barakat, "On excess bandwidth usage of video streaming: when video resolution mismatches browser viewport", in proceedings of the NoF Conference, 2020.

Bandwidth waste – Chrome and YouTube



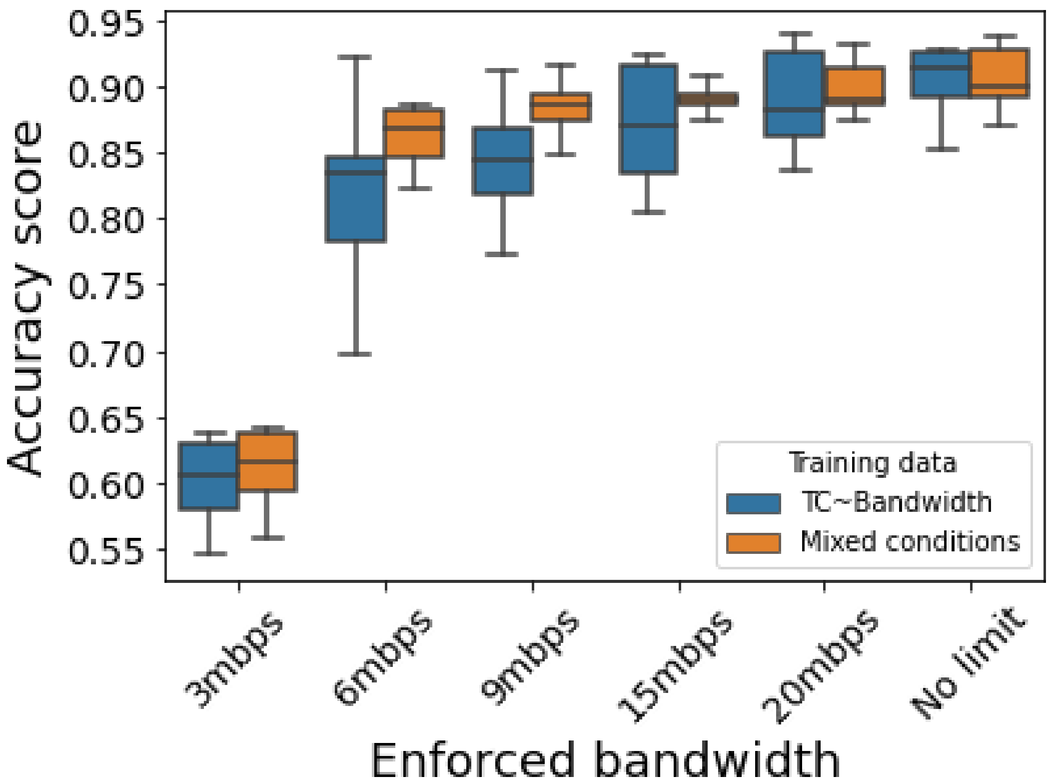
Viewport inference from traffic captures



* Random Forests with default configuration, unlimited bandwidth scenario

O. Belmoukadam, C. Barakat, "Unveiling the end-user viewport resolution from encrypted video traces", to appear in IEEE Transactions on Network and Service Management.

Viewport class inference



Viewport resolution (pixels)	Viewport class
400x225	SD
640x360	SD
850x480	SD
1280x720	HD
1920x1080	HD

Customized model per bandwidth value
General model for any bandwidth

* Random Forests with default configuration

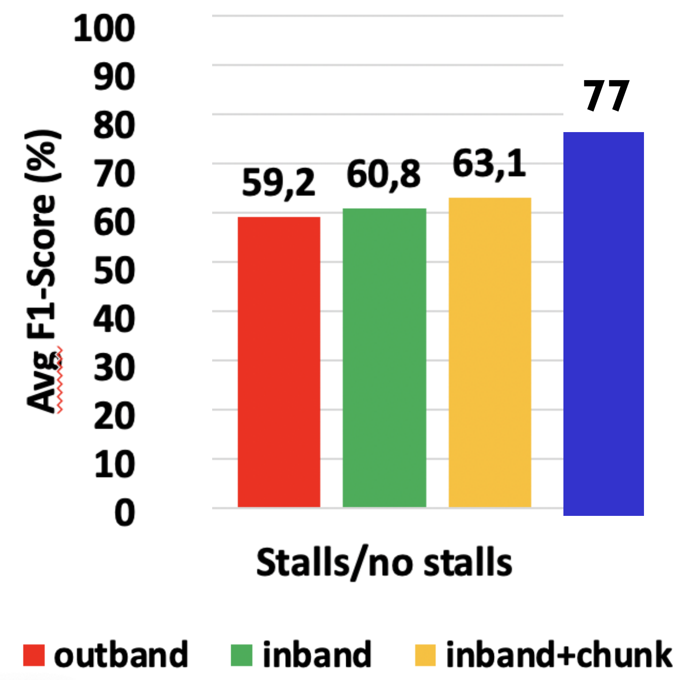
O. Belmoukadam, C. Barakat, "From Encrypted Video Traces to Viewport Classification", in proceedings of CNSM, 2020.

Wrap-up

- A new intelligent experimentation framework based on active learning
- A set of new models to predit and estimate QoE for video streaming
- Study of viewport impact, and of viewport inference
- Two case studies: out-of-band and inband
  - Techniques to isolate encrpyed video chunks
  - Predicting QoE from out-of-band measurements behaves slightly worse than estimating QoE from passive traffic captures

Ongoing and future work

- ❑ Enhancing the out-of-band method with content related information
- ❑ Consideration of other streaming platforms, protocols, and access technologies (mobile)
    - QoE perspective benchmarking
- ❑ Extension to other services than video streaming
- ❑ Closing the loop – Model-driven QoE-aware network management



Adding the video bitrate to QoE prediction (stalls as QoE proxy)