# Machine Learning Assisted DASH Video QoE Inference Through Network QoS Features in 4G and 5G scenarios

**PhD candidate:** Raza Ul Mustafa
**Supervisor:** Prof. Dr. Christian Esteve Rothenberg (FEEC/UNICAMP)

**Committee members:**

1. Prof. Dr. Aldebaro Klautau (UFPA)
2. Dr. Fuad Mousse Abinader Junior (CPQD)
3. Prof. Dr. Paulo Cardieri (FEEC/UNICAMP)
4. Prof. Dr. Rafael Pasquini (UFU)

Department of Computer Engineering and Industrial Automation (DCA)
School of Electrical and Computer Engineering (FEEC)
University of Campinas (UNICAMP)

1 Introduction

2 Frameworks

3 Machine Learning

4 YouTube QoE & Datasets

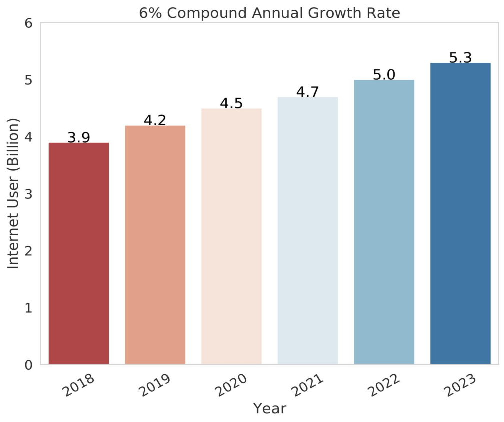
5 Contributions & Publications

6 Conclusions & Future Work

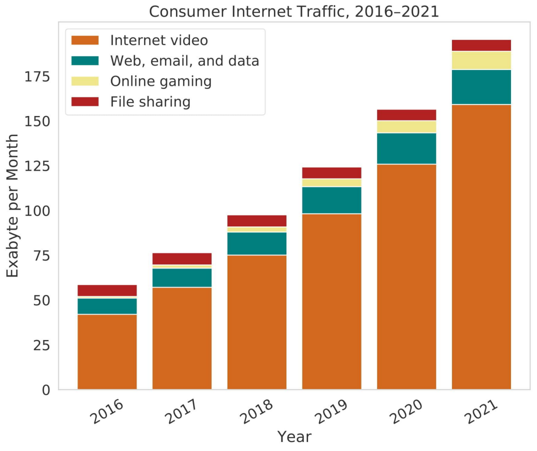
# 1 Introduction, Background, Motivation & Research Questions

# Internet consumers

- **Internet User:** ⅔ will have internet access by the end of 2023 in terms of global population.*
- **Dominant Consumer Traffic:** Internet video (incl. short-form, long-form, live internet video, etc.).**




*Cisco Annual Internet Report (2018–2023)



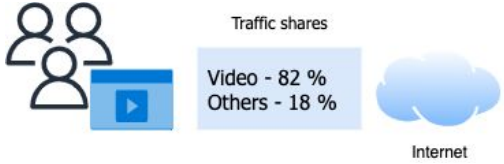

**Cisco VNI: Forecast and methodology (2016-2021)

# Today's internet traffic

Till 2020:

- 65% downstream video traffic accounted overall on the internet.*
  - Streaming services (incl. Youtube, Netflix, etc.)
  - Social network (incl. Facebook video)

By 2022: Internet video will represent 82% of all internet traffic. Mobile video traffic is supposed to increase by 73 % in 2023 **.




*The stats indicate the predominance of internet video streaming traffic in the foreseeable future*

* Sandvine, The mobile internet phenomena report, Technical Report, 2020
*** Cisco, Cisco visual networking index: Forecast and trends, 2017-2022

# 4G & 5G

- From the first-generation, voice-only, mobile cellular communication systems to the current 4G - Long Term Evolution - LTE, development has progressed at a steady pace.
- But, applications utilizing social media, gaming, and recent advances in Augmented/Virtual Reality, has accelerated the demands for 5G.
    - High data rates 10x increase than 4G
    - 10x lower latency
    - Supports tens of thousands of devices for future IoT
    - *The largest prediction being that the number of connected Internet of Things(IoT) devices is expected to reach 75.44 billion* *
- At the heart of this growth in throughput demand is video traffic.
    - Video on Demand (VoD), live video streaming



*Tanweer Alam. 2018. A Reliable Communication Framework and Its Use in Internet of Things (IoT).

# Video streaming services

→ Video-on-Demand (VOD) services.

- A wide variety of high-quality online content libraries
- Always accessible
- Paid subscription basis or free

→ Over-the-Top (OTT) platform.

- Video content is delivered directly to viewers via the internet
- Viewers can access content from a wide range of devices
- Youtube, Netflix, Amazon Prime, etc

- **HTTP Adaptive Streaming (HAS) standard.**
  - De-facto standard to carry video traffic for VoD services
  - **Adaptive Bitrate Streaming (ABS)** algorithms aim to provide interrupt free video streaming service based on the network status

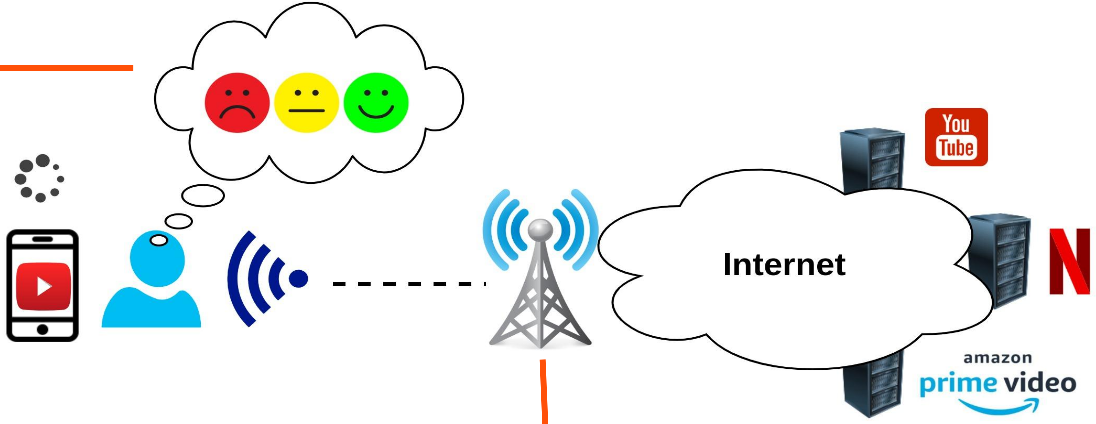
# User satisfaction and stakeholders

Client / Server Side Log

- Content Providers

Quality of Experience (QoE)

**User Satisfaction**



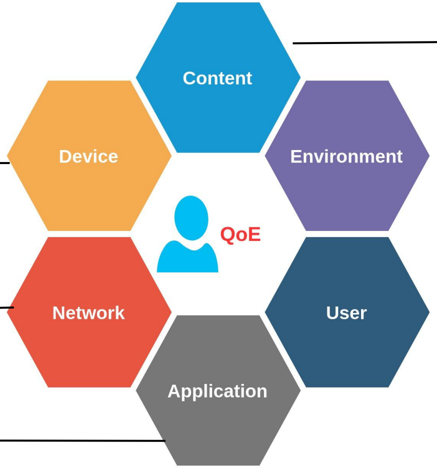

**Network Operators (e.g., MNO)**

Need to assess the key factors that impact on user experience

(QoE monitoring and management)

# Quality of Experience (QoE)

*"**The degree of delight or annoyance of the user of an application or service**. It results from the fulfillment of his or her expectations with respect to the utility and/or enjoyment of the application or service in the light of the user's personality and current state"***




**Type, quality, popularity**

**PC/mobile, screen size, browser/app**

**Viewing environment .....**

**Quality of Service (QoS) offered by transmission network**

**Expectation, mode, demography.**

**Multimedia, web, game**

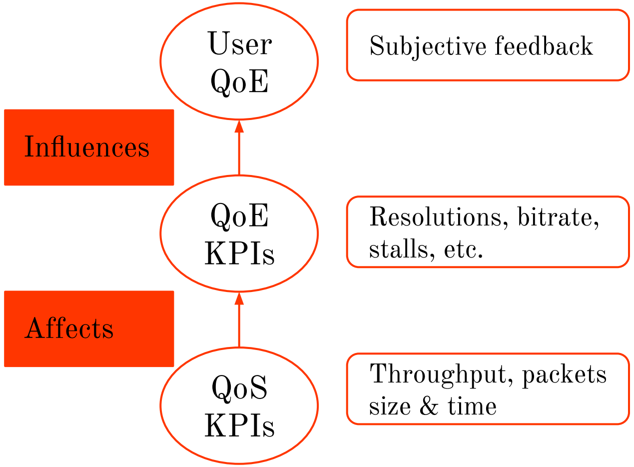
* S. M. Patrick Le Callet and e. Andrew Perkis, "Qualinet white paper on definitions of quality of experience(2012)," European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003), Lausanne, Switzerland, Version 1.2, March 2013

# Quality of Service (QoS) to QoE

- **For video streaming,** in most cases, application-level KPIs such as video bitrate, resolution, and buffering/stall has a strong influence on QoE.*
- **QoS parameters** as **network performance** indicators (KPIs) have a direct impact on application-level KPIs.**

**Network element's ability to satisfy stated and implied needs of the user of the service.**

- **MNOs can track QoS KPIs.**



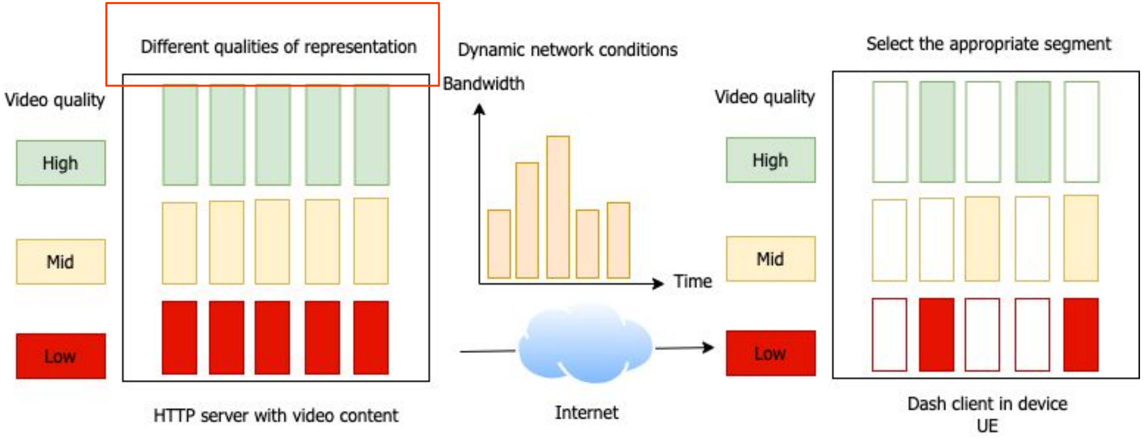

* F. Dobrian, V. Sekar, A. Awan, I. Stoica, D. Joseph, A. Ganjam, J. Zhan, and H. Zhang,“Understanding the impact of video quality on user engagement,” in ACM SIGCOMM Computer Communication Review, vol. 41, no. 4. ACM, 2011, pp. 362–373.
** M. Fiedler, T. Hossfeld, and P. Tran-Gia, “A generic quantitative relationship between quality of experience and quality of service,”IEEE Network, vol. 24, no. 2, pp. 36–41, 2010.

# HTTP Adaptive Streaming (HAS) standard for VoD services

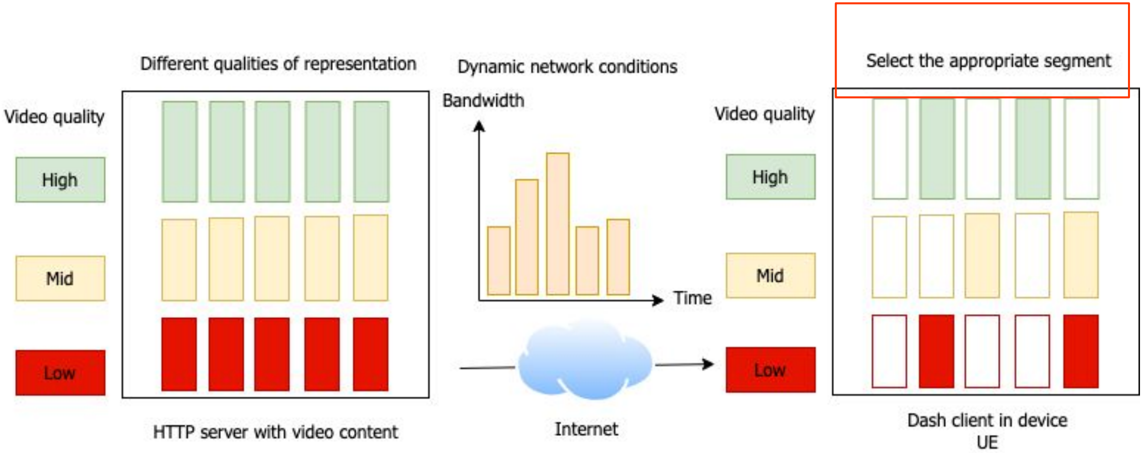
HAS works by split video file into many small segments of same duration (2-10 sec). Each segment encoded with different bitrates and resolutions.

**Dynamic Adaptive Streaming over HTTP (DASH)** is the most dominating format for implementing HAS.

In DASH, each segments' structure describes in **Media Presentation Description (MPD)** file.



# The role of Adaptive Bitrate Streaming (ABS) algorithm

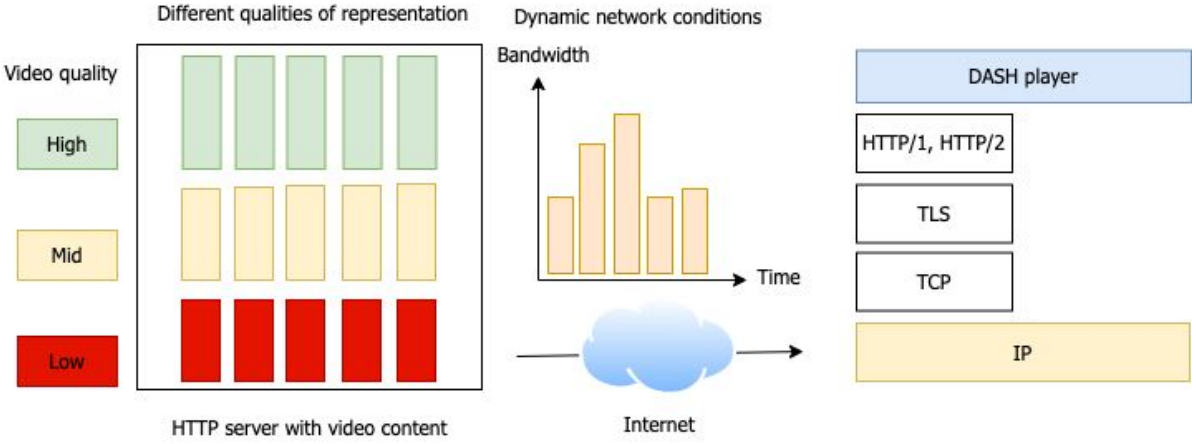



ABS algorithms are responsible for dynamically selecting the appropriate segments based on network conditions.

The purpose of this dynamic segment selection to adapt to changes in network conditions and provide an interrupt-free (e.g., stall) service.

**Rate-based:** Bandwidth estimation
**Buffer-based:** State of the playback buffer
**Hybrid:** Mixed

# Transport options: TCP

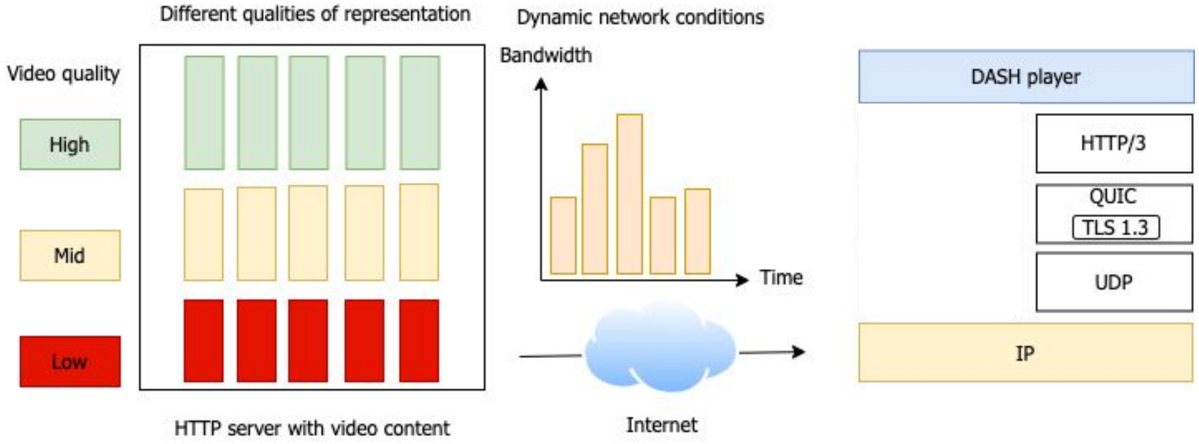



Reliability and inorder delivery. TCP required a 3-way handshake to establish a connection.

TLS over TCP requires an additional handshake for secure connection.

TCP can be an expensive network tool as it includes absent or corrupted packets and protects data delivery with controls like acknowledgments, connection startup, and flow control.

# Transport options: QUIC




Reliable compared to UDP
Automatics retransmission, congestion control.

QUIC uses UDP as its base, it involves loss recovery. This is because QUIC behaves like TCP and checks each stream separately re-transmits data when it gets lost.

Improves performance during network switching events Wifi to mobile.

# Machine Learning (ML) and QoS to QoE mapping

- ML/AI achieved maturity over the period of time.
  - Therefore, ML techniques are widely used to increase end users satisfaction
- Supervised ML based model for objective assessment, i.e,
  - Decision Tree
  - Random Forest
  - K-nearest neighbors
  - Artificial Neural Network (ANN)
- Mapping between QoS and QoE KPIs for certain actions, i.e., resources optimization, SLA, SDN decisions.

**Objective QoE**

- Stalls
- Bitrate
- Shifts

**QoE = Given the input features, predict QoE values**

**Models accuracy = Highly correlated features**

**Models – Regressions (continuous) and classifications (categorical)**

# Motivation (1/2)

Now, the technology is evolving towards its fifth-generation (5G)
- High throughput
- Low latency, realtime information processing & low network management complexity

Moreover, 5G and Beyond (5GB) networks are expected to equip with the Edge Computing
- Computation near to edge of the network
- Increase user satisfaction by reducing latency, providing realtime response

In fact, the share of video streams in the overall Internet traffic is continuously growing
- YouTube, NetFlix, and others, which use the Dynamic Adaptive Streaming over HTTP (DASH) technology
- Mobile video traffic is supposed to increase by 73 % in 2023 *

*Ragimova, Kamila, Vyacheslav Loginov, and Evgeny Khorov. "Analysis of youtube dash traffic." *2019 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom)*. IEEE, 2019.

# Motivation (2/2)

MNOs responsibility

- Provide satisfactory experience
- Maximize Mean Opinion Score (MOS)

However, Deep Packet Inspection (DPI) not available because of TLS encryption

- Limited QoS KPIs
- Packet time and size
- Chunk level statistics are expensive and vary overtime by various video service providers

Moreover, limited availability to conduct large scale 4G & 5G experiments

- Realtime 4G and 5G traces with Channel Level Metrics (CLM)
- Realtime QoE KPIs
- Realtime objective QoE prediction

1

# Research questions (1)

**How to conduct large-scale experiments using real 4G and 5G use cases, which requires low system requirements and at the same time, provide necessary data for analysis of the streaming sessions?**

- ❑ A light weight DASH QoS and QoE evaluation framework with open datasets
- ❑ Emulation based 4G and 5G open dataset with QoS and QoE in the form Jupyter Books
- ❑ 4G and 5G commercial dataset to run different use cases, i) Pedestrian, ii) Mobility, iii) Indoor, iv) Outdoor
- ❑ Equipped with all the dependencies to run 4G and 5G experiments
- ❑ Both protocols TCP & QUIC, with many other scalability options

# Research questions (2)

**What is the performance footprint of the current state-of-the-art ABS algorithms under 5G, and how does it compare with 4G? Which ABS algorithm is more similar to the most popular streaming platform YouTube?**

- Analysis of ABS algorithms under diverse 4G and 5G network conditions
  - Shifts, Stalls, Bitrate, QoE Model ITU P.1203, etc
- **Conventional** – Provides more similarity with YouTube
- **Buffered** - Maximum QoE in 5G compared to 4G
- **Elastic** - Similarities in QoE in both the technologies

# Research questions (3)

**Which QoS features can be effectively used from the network level of encrypted and unencrypted DASH traffic to estimate the QoE of adaptive video streaming?**

- ❑ Per-segment QoS features RTT, Packets, Throughput relation to QoE
- ❑ Inter Packet Gap (IPGs) provide new possibilities to estimate QoE
- ❑ Shifts metrics EMA and CUSUM are highly correlated with QoE derived from IPGs
- ❑ QoS features extraction using a real time based method, i.e., (0.5-5) seconds

# Research questions (4)

**How can ML techniques effectively predict QoE using network-level QoS features?**

- Linear model for the prediction of QoE based on resolutions, bitrate and stall
  - Regression with multiple variables, Decision tree, Random forest
- QoE classification using QoS features derived from IPGs
  - MOS scale (1-5)
  - Artificial Neural Network
  - Decision Tree
  - Random Forest
  - K nearest neighbors
- QoS to QoE mapping performance under different time windows

# Research questions (5)

**How can we correlate 4G and 5G Channel Level Metrics (CLM) to the QoE KPIs of real YouTube Traffic?**

- 4G and 5G performance footprint using YouTube as a baseline
- Prediction of objective QoE KPI stall using only CLM metrics,
  - i.e., CQI, RSRQ, RSRP
- 4G and 5G (NSA & SA) dataset over 06+ months in different regions.
  - France
  - USA
  - Brazil
- Reproducibility with open datasets, include CLM, context and YouTube QoE with 1 second granularity

## 2 Frameworks, Encrypted (EFFECTOR) & Unencrypted

2

# DASH QoS to QoE evaluation frameworks

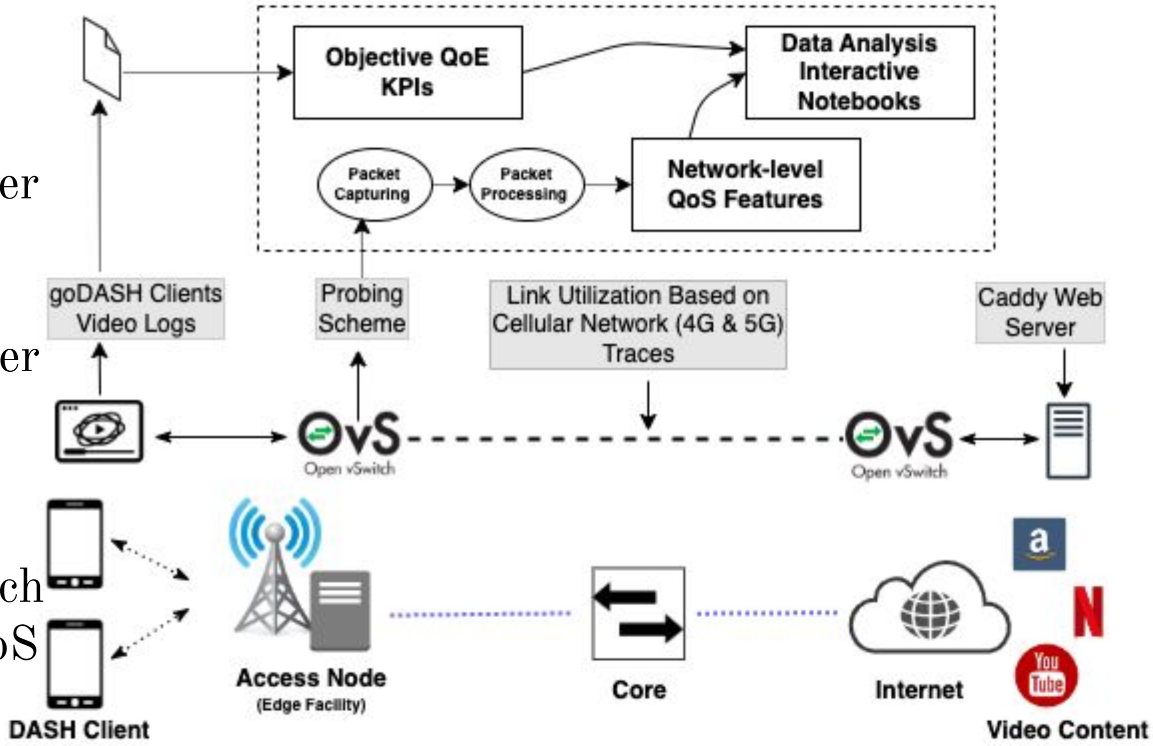
We present a data-driven framework for DASH QoE Performance Evaluation over real 4G & 5G cellular network traces collected in the wild.

Reproducible framework  with a series of pre-installed DASH tools to analyze state-of-the-art ABS algorithms by varying  QoS KPI in i) Indoor, ii) Outdoor, iii) Pedestrian, iv) Mobility use cases.

Interactive Jupyter notebook and Binder service providing an executable live analytical environment to processes the output dataset of the framework and compares the relationship between QoE and QoS KPIs.

# Framework architecture

- **Mininet-WiFi**- a network emulation tool
- **goDASH**- a DASH video player
- **Caddy**- a web server hosting DASH video content*
- **Linux TC**- a traffic controller in the Linux kernel
- **Tcpdump**- a passive network traffic sniffer
- **Python Scripts**- Scripts to fetch encrypted and unencrypted QoS KPIs

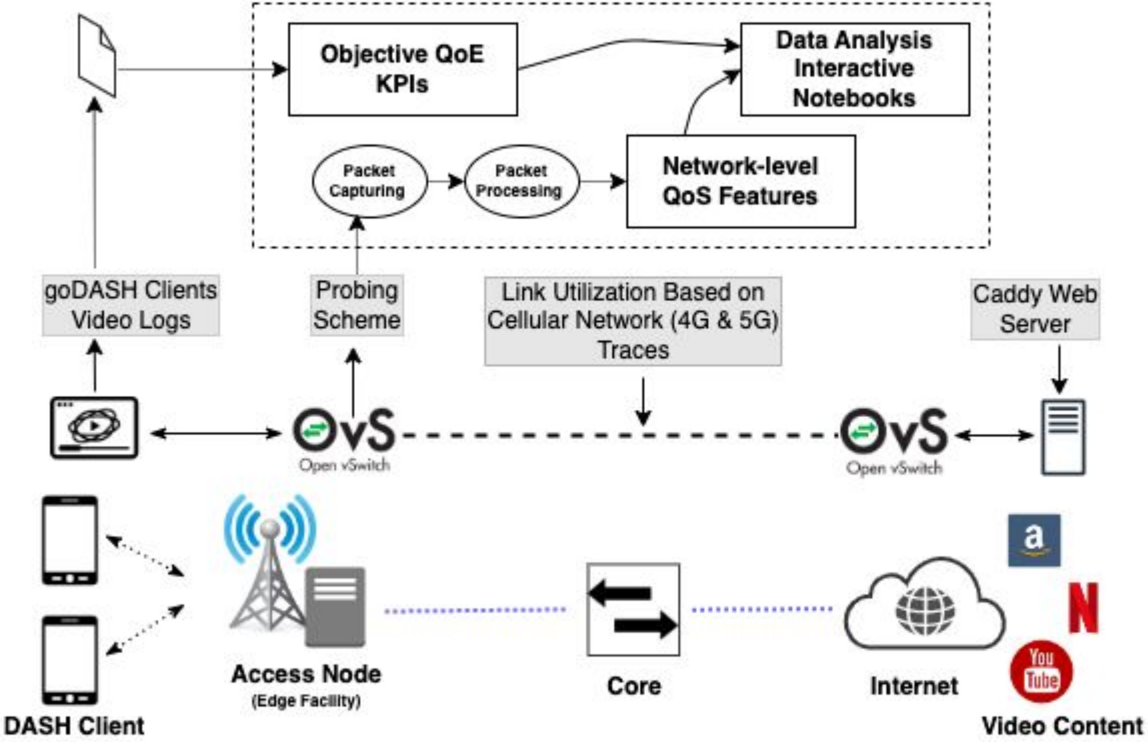


# Experimental parameters

| Technology | Use cases | Protocols |
|---|---|---|
| 4G | Indoor<br>Outdoor<br>Pedestrian<br>Mobility | TCP<br>QUIC |
| 5G | | |

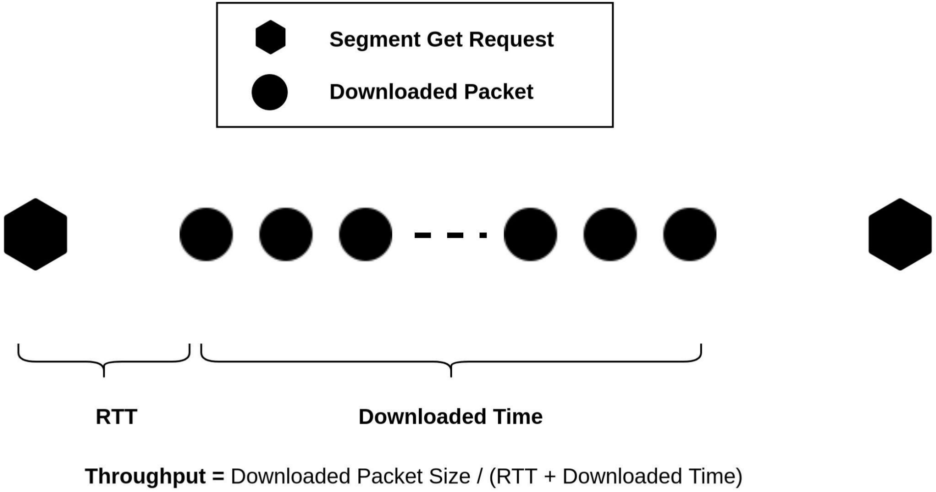
4G and 5G dataset with channel and context

https://github.com/razaulmustafa852/youtubegoes5g/



# QoS features extraction - unencrypted

1. Per-segment based QoS features.
   a. RTT
   b. Throughput
   c. Packets
2. Adaptation algorithms correlation with different network use cases.
   a. Buffered – BBA
   b. Hybrid – Elastic
   c. Throughput – Conventional




**Limitation:** *E2E encryption limits segment level KPIs*

# EFFECTOR: DASH QoE and QoS Evaluation Framework For EnCrypTed videO tRaffic

# Introduction

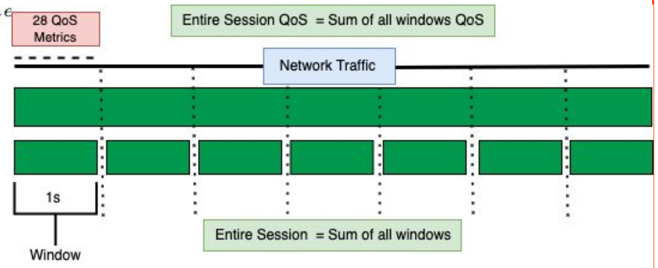
- E2E encryption poses many challenges for Mobile Network Operators (MNOs).
- MNOs must be aware of the end user's QoE by exploring network level QoS.
- To conduct large-scale experiments using real 4G and 5G use cases, we provide EFFECTOR.
  - A framework to showcase lightweight QoS features measurement technique at edge nodes from encrypted DASH video traffic
  - EFFECTOR uses an emulated environment with real 4G and 5G drive test traces to generate video traffic
  - It requires low system requirements and provide necessary data for analysis of the streaming sessions

2

# QoS features extraction approach - EFFECTOR (1/2)

**Algorithm 1** QoS features extraction approach for a time window of a video streaming session

```
Session ← n_minutes                                   ▷ Streaming session
Ctime ← 0                                             ▷ Current time – Begin from 0
Step ← window                                         ▷ Step (1,2,3,4,5)-seconds
Tpackets ← 0                                          ▷ Total packets
Tpackets_GT100 ← 0                                    ▷ Total packets size > 100
while Ctime ≤ Session do
    Array_Size () ← Packet_Size
    Array_Time () ← Packet_Time
    BITS ← BITS + len(p[IP]) * 8)
    if Packet_Size > 100 then
        Array_GT100_Size () ← Packet_Size
        Array_GT100_Time () ← Packet_Time
    end if
    Ctime ← Ctime + Step
end while
while P = 10 to 90 do
    P percentile of Array_Size
end while
while P = 10 to 90 do
    P percentile of throughput in a time window
end while
Throughput ←BITS/window
Tpackets ← Count (Array_Size)
Tpackets_GT100 ← Count (Array_GT100_Size)
IAT_w ← Last − First                ▷ Last Packet Time - First Packet Time in a window
IPG_avg ← Average (IPG (Array_Time))
IPG_avg_gt100 ← Average (IPG (Array_GT100_Time))
Std ← STD (Array_Size)
Std_gt100 ← STD (Array_GT100_Size)
Std_IPG ← STD (IPG (Array_Time))
Std_IPG_gt100 ← STD (IPG (Array_GT100_Time))
```




**28 QoS features – of a single window**

1. Packet size distribution 10-90 percentile – 9 features
2. Throughput distribution 10-90 percentile – 9 features
3. Throughput
4. Total packets [w/ or gt 100B)
5. IPG average of a window [w/ or gt 100B)
6. Standard deviation of packet size [w/ or gt 100B)
7. IPG standard deviation of a window [w/ or gt 100B)
8. Inter Arrival Time (IAT) of a window

2

# QoS features extraction approach - EFFECTOR (2/2)

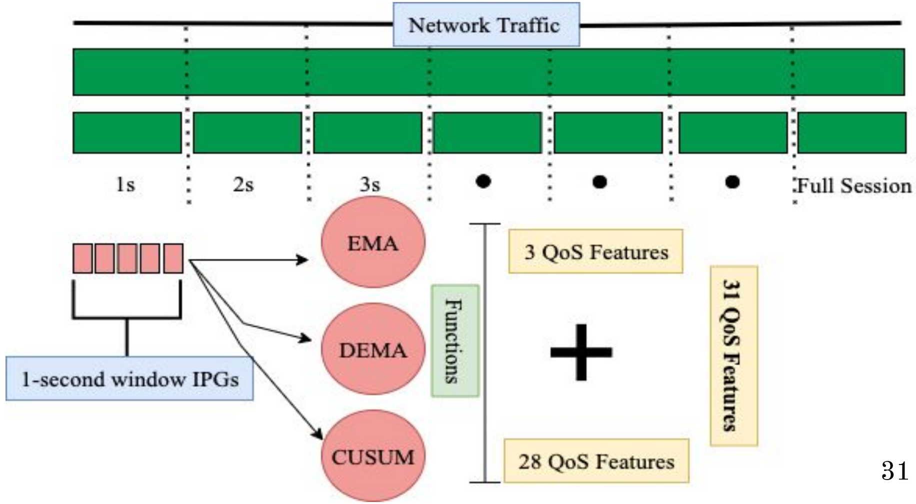
**Algorithm 2** EMA calculation, function takes IPG values of a time window as an array

```
FUNCTION ← START
α ← 0.99
EMA_array () ← IPG[0]
I ← 0
while I < count (IPG) do
    EMA_array () ← IPG[I] * (1 − α) + IPG[I − 1] * α
    I + +
end while
return array_sum(EMA_array)
FUNCTION ← END
```

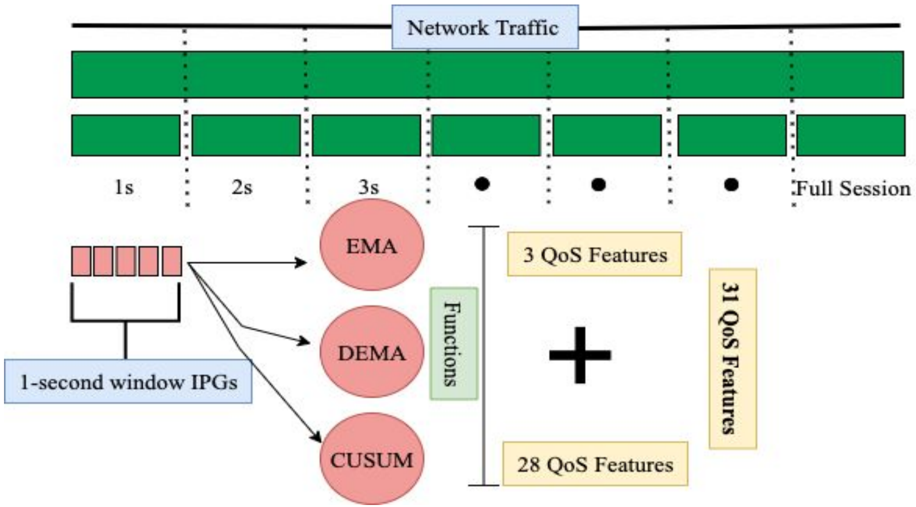
**Algorithm 4** CUSUM calculation, function takes IPG values of a time window as an array

```
FUNCTION ← START
I ← 0
CUSUM ← 0
while I < count(IPG) do
    CUSUM+ = I
    I + +
end while
return CUSUM
FUNCTION ← END
```

- Given the IPGs of a time-window, function will return three QoS features which provides maximum IG for video QoE estimation.
- DEMA is a technical indicator devised to reduce the lag in the results produced by a traditional moving average.



# QoS features extraction approach - EFFECTOR (2/2)




Given the IPGs of a time-window, function will return three QoS features which provides maximum IG for video QoE estimation.

The moving average is designed as such that older observations are given lower weights. The weights fall exponentially as the data point gets older.

# Network level temporal QoS features

2

Encryption

| Real Time (Window) | Comments |
|---|---|
| Packets count (total) [ w/ gt 100B] | *Ignoring ack packets of size 100B* |
| Packet size distribution [w, (10-90)p] | *10-90 percentile packet size distribution in a window* |
| Throughput [w, distribution (10-90)p] | *10-90 percentile throughput distribution in a window* |
| Packet Time [IPGs, Inter Arrival Time] | *Inter Packet Gap (IPGs) of a window* |
| IPGs features [EMA, DEMA, CUSUM] | *See the continuity of packets* |
| IPGs [Avg, Std, w/ gt100B] | *Average, Standard deviation of window* |

# Application level QoE KPIs

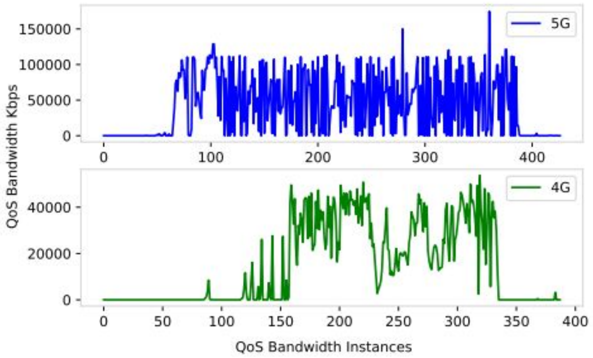
- Application-level QoE KPIs as ground truth
  - Per segment KPIs by godash : Taken into account for entire video session.

| Seg | Algo | Seg_Dur | Width | Heigh | Arr_Time | Del_Time | Stall | Rep_Rate | Act_Rate | Buffer | P.1203 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Conv | 2000 | 320 | 180 | 16273 | 104 | 0 | 237 | 496 | 3541 | 1.871 |
| 2 | Conv | 2000 | 736 | 414 | 18313 | 866 | 0 | 1786 | 3489 | 3501 | 2.228 |
| 3 | Conv | 2000 | 1280 | 720 | 24016 | 4249 | 2201 | 3031 | 5098 | 2000 | 2.336 |
| 4 | Conv | 2000 | 1280 | 720 | 29421 | 3636 | 0 | 3031 | 2311 | 2000 | 2.253 |
| 5 | Conv | 2000 | 1280 | 720 | 30593 | 43 | 0 | 2375 | 1810 | 2828 | 2.343 |

| | | |
|---|---|---|
| Arrival time ms | Time spent for delivery of this segment in ms | Stall in ms |
| Representation rate of downloaded segment in Kbit/s (taken from MPD file); | Delivery rate of the network in Kbit/s (segment size divided by time for delivery); | The actual bitrate of this segment (segment size divided by the segment duration) in Kbit/s; |

# QoS features correlation with QoE

- QoS features correlation with the objective **QoE stall** and **quality shifts.**
    - Shifts and stalls are the main QoE indicators *
- Commercial 4G and 5G datasets collected in the wild with Channel Level Metrics (CLM) and emulate them in EFFECTOR.




YouTube as a baseline QoE metrics:

- Shifts
- Resolutions
- Bytes downloaded
- Loaded percentage
- Stall

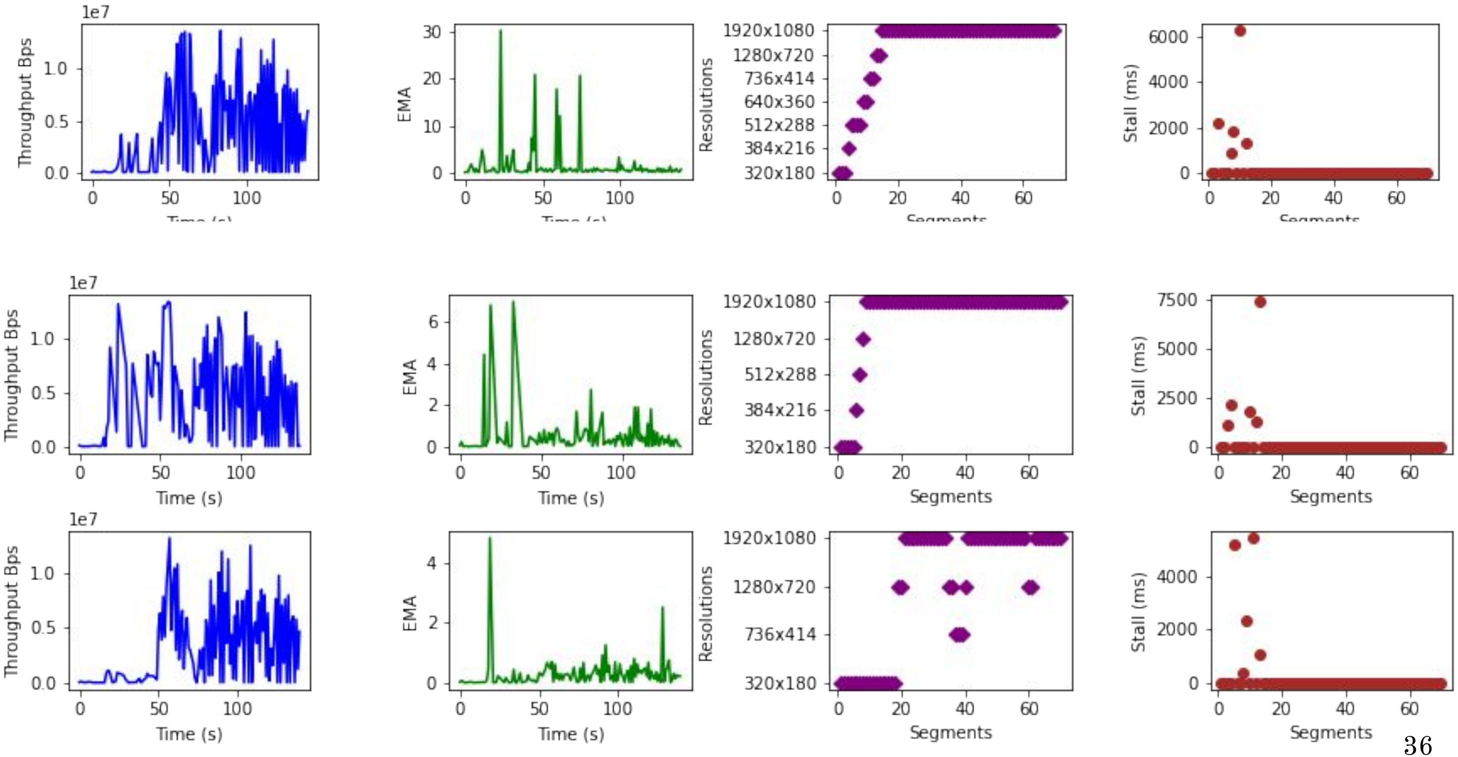
* Fan Zhang, Long Xu, and Qian Zhang. 2013. Maximum-likelihood visual quality based on additive log-logistic model. In 2013 IEEE 15th International Workshop on Multimedia Signal Processing (MMSP). IEEE, 470–475.

# 4G: ABS QoS and QoE correlation

BBA: Buffered

Conv: Throughput

Elastic: Hybrid



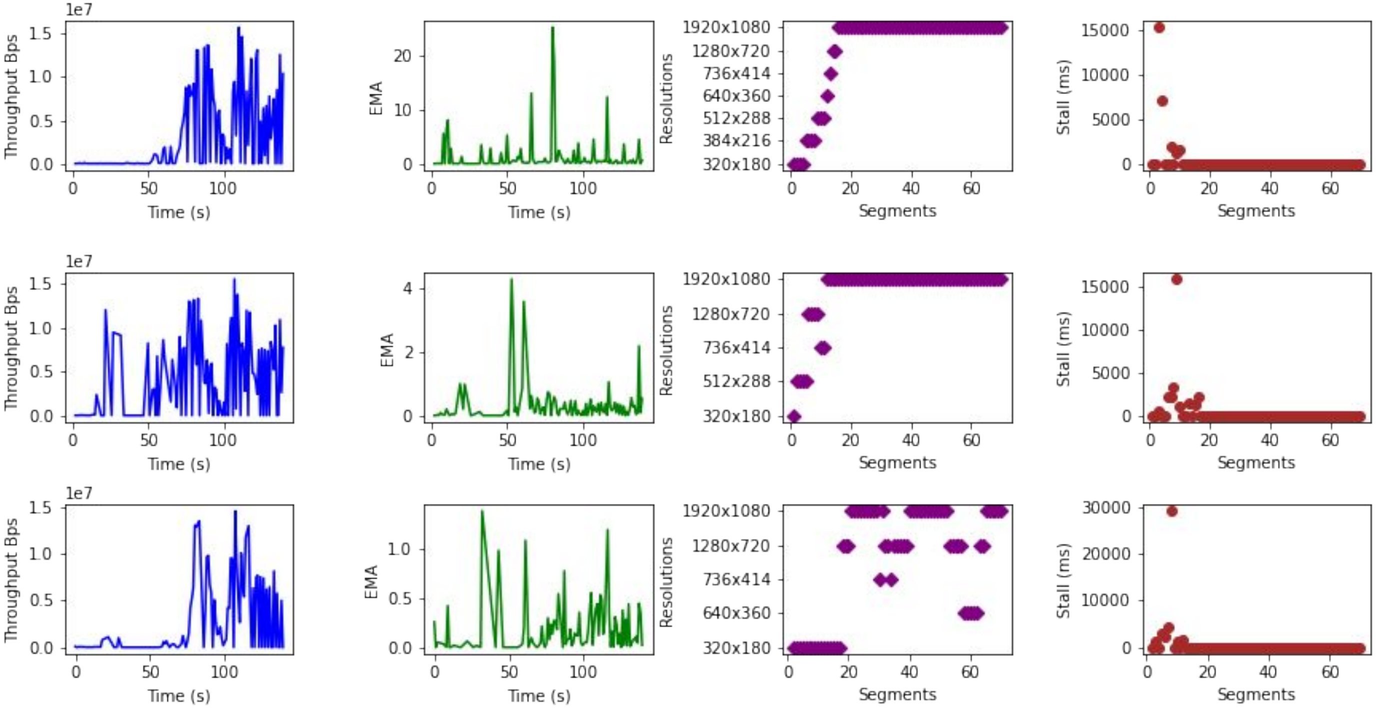
2

# 5G: ABS QoS and QoE correlation

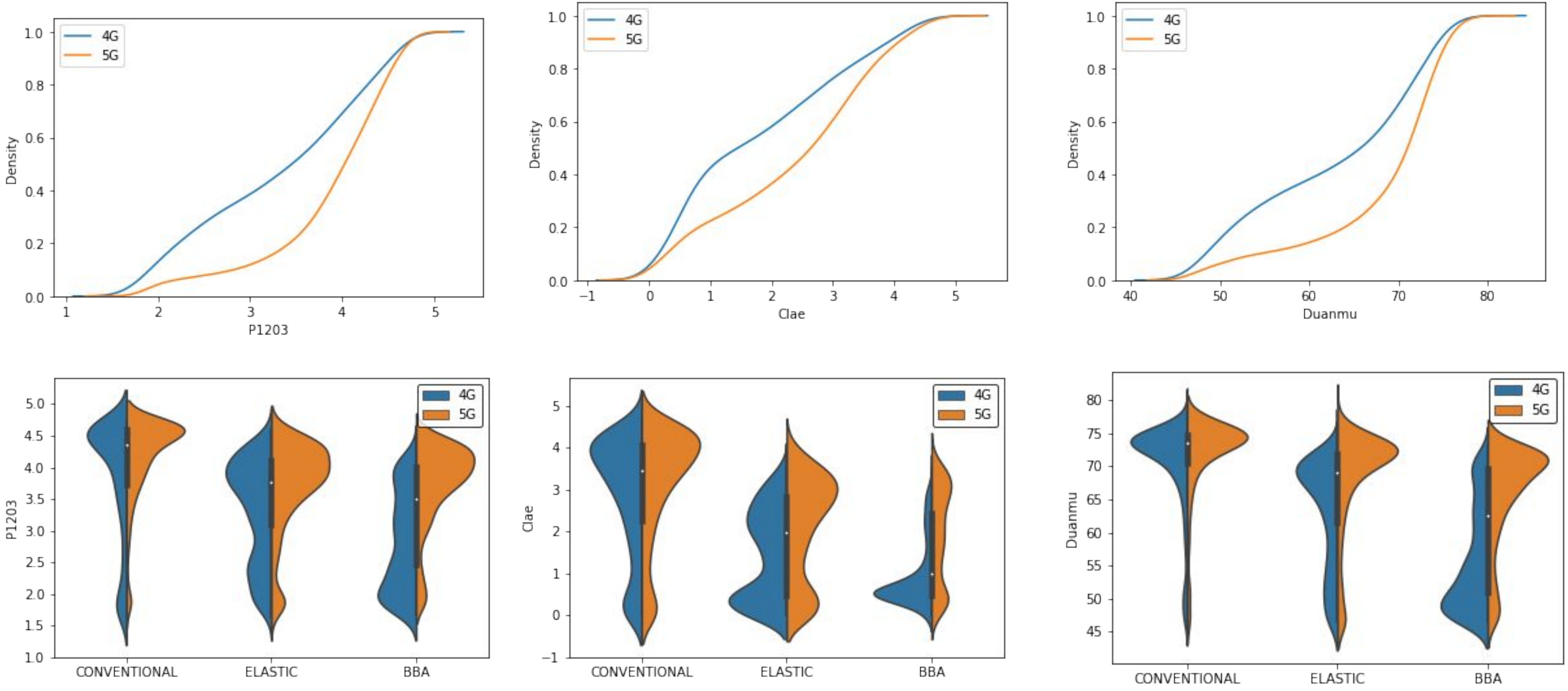
BBA: Buffered

Conv: Throughput

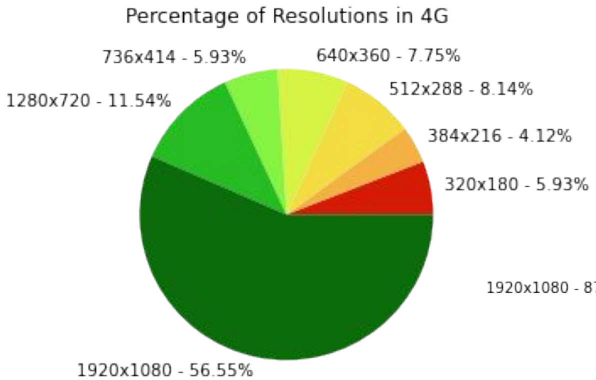
Elastic: Hybrid

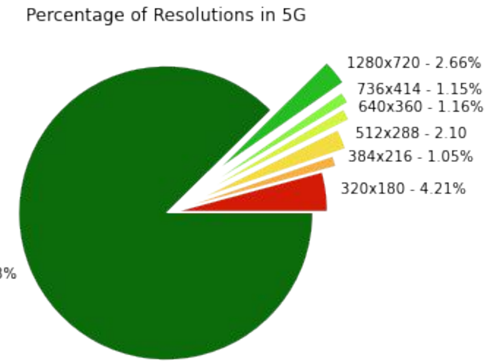


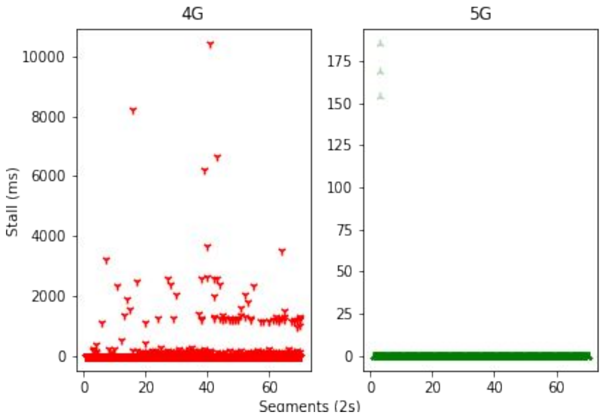
# Network capacity: 4G vs. 5G performance footprint - QoE models



# Network capacity: 4G vs. 5G shifts and stalls




4G vs. 5G Shifts:

4G experience more quality shifts as compared to 5G. We see 56 % the higher resolution – 1920x1080 for 4G whereas 87 % for 5G.

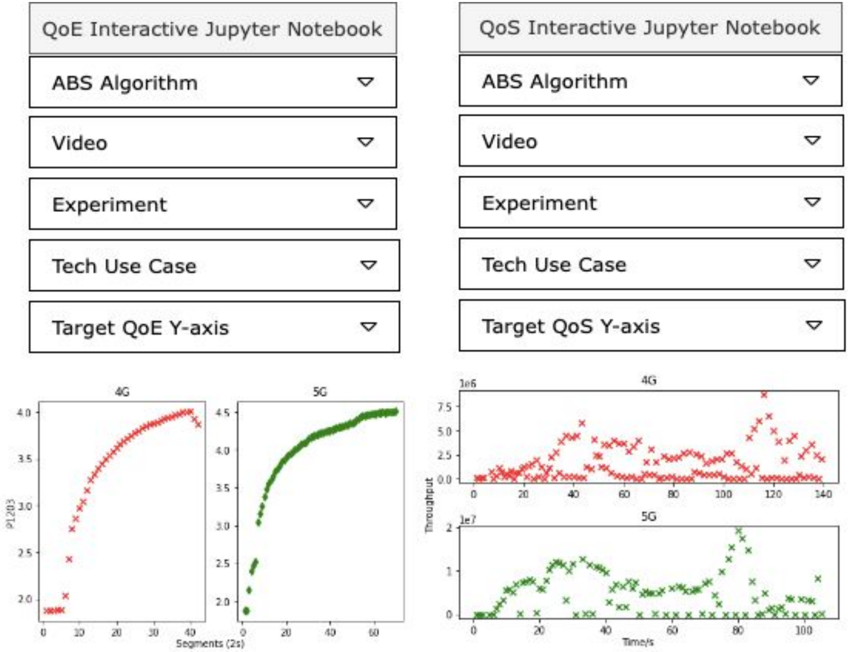
In 5G we see few quality shifts, but their percentage compared to 4G is very less.

4G vs. 5G Stalls:

4G experience more stalling events as compared to 5G.

# Interactive Jupyter Notebooks

- Users can visualize and control changes in the data.
- The Jupyter Notebook is a web-based interactive computing platform.
- We provide Interactive Jupyter Notebook for the dataset generated during experimental phase.
    - QoE: Objective QoE KPIs – Stall, Resolutions, Bitrate, 5 QoE model, Delivery rate of network
        - Context: ABS Algorithms, Video, Use case, Experimental Iteration
    - QoS: Target variable  on Y-axis – 30+ features
        - Context: ABS algorithms, Video, Use case, Experiment Iteration



# Frameworks

Per-segment

https://github.com/razaulmustafa852/5G

EFFECTOR

https://github.com/razaulmustafa852/EFFECTOR

# 3 QoE prediction using QoS KPIs in encrypted and unencrypted video traffic

# QoE prediction using QoS KPIs - Regression

- Per-segment unencrypted QoE prediction
  - Used per-segment QoS KPIs RTT, Throughput and Packets
  - Multilinear Regression, Random Forests, Decision Tree
    - Model input: RTT, Throughput, Packets – QoE score
    - Train/Test split 70% / 30%

| ABS | RTT/s | Throughput/bps | Packets | **QoE score** |
|---|---|---|---|---|
| BBA | 0.054 | 828169 | 2 | **1.87** |
| BBA | 0.46 | 479897 | 49 | **1.89** |
| BBA | 0.115 | 527584 | 64 | **1.903** |
| BBA | 0.10 | 528615 | 62 | **1.91** |
| BBA | 0.106 | 459642 | 38 | **1.90** |

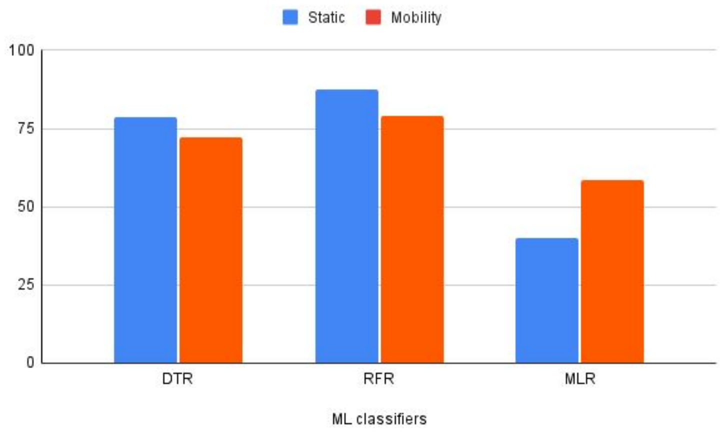
# MAE: Static & Mobility

$$\text{MAE} = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n}$$

| Algorithm | Classifiers | MAE [%] [Static] | MAE [%] [Mobility] |
|---|---|---|---|
| Arbiter +, Elastic | DTR | 0.20 | 0.31 |
| | **RFR** | **0.17** | **0.31** |
| | MLR | 0.55 | 0.55 |
| BBA, Logistic | DTR | 0.12 | 0.01 |
| | **RFR** | **0.07** | **0.01** |
| | MLR | 0.12 | 0.19 |
| Conventional, Exponential | DTR | 0.23 | 0.13 |
| | **RFR** | **0.10** | **0.07** |
| | MLR | 1.03 | 0.70 |

# Results: static & mobility

| Case | Classifiers | Accuracy |
|---|---|---|
| Static | DTR | 78.68 % |
| | **RFR** | **87.63 %** |
| | MLR | 40.01 % |
| Mobility | DTR | 72.37 % |
| | **RFR** | **79.00 %** |
| | MLR | 58.67 % |



# QoE prediction using QoS KPIs - Classification

3

- The process of identifying and and grouping objects or ideas into predetermined categories.
  - We extracted QoS features from packet time and size
  - Input QoS features to ML classifiers to estimate QoE class into three categories
    - Poor
    - Good
    - Excellent
  - Used Artificial Neural Network (ANN), Decision Tree, Random Forest, k-Nearest Neighbors (KNN)
  - 5-fold cross validation
  - Train/Test 70% / 30 %
  - Highest accuracy ANN
  - Accuracy on different time windows (1-5) seconds

3

# Classifier results on different time windows in %

TCP

| Window | ANN | KNN | DT | RF |
|---|---|---|---|---|
| 1 | 73 | 71 | 72 | 72 |
| 2 | 77 | 75 | 76 | 77 |
| 3 | 79 | 79 | 73 | 79 |
| 4 | 79 | 77 | 71 | 76 |
| 5 | 78 | 77 | 76 | 76 |

QUIC

| Window | ANN | KNN | DT | RF |
|---|---|---|---|---|
| 1 | 73 | 77 | 78 | 76 |
| 2 | 78 | 78 | 77 | 80 |
| 3 | 77 | 78 | 76 | 80 |
| 4 | 79 | 77 | 71 | 76 |
| 5 | 77 | 82 | 77 | 81 |

# Classifiers results (%) on different ABS with time windows - TCP

Conventional

| Window | ANN | KNN | DT | RF |
|---|---|---|---|---|
| 1 | 60 | 60 | 59 | 60 |
| 2 | 67 | 67 | 64 | 66 |
| 3 | 68 | 67 | 62 | 70 |
| 4 | 70 | 65 | 55 | 64 |
| 5 | 67 | 69 | 61 | 66 |

BBA

| Window | ANN | KNN | DT | RF |
|---|---|---|---|---|
| 1 | 73 | 72 | 72 | 71 |
| 2 | 84 | 81 | 80 | 84 |
| 3 | 84 | 83 | 83 | 83 |
| 4 | 85 | 88 | 82 | 83 |
| 5 | 84 | 83 | 83 | 83 |

Elastic

| Window | ANN | KNN | DT | RF |
|---|---|---|---|---|
| 1 | 74 | 74 | 73 | 71 |
| 2 | 75 | 75 | 71 | 74 |
| 3 | 75 | 79 | 76 | 78 |
| 4 | 80 | 80 | 74 | 80 |
| 5 | 75 | 79 | 74 | 79 |

# Classifiers results (%) on different ABS with time windows - QUIC

Conventional

| Window | ANN | KNN | DT | RF |
|---|---|---|---|---|
| 1 | 70 | 70 | 65 | 66 |
| 2 | 68 | 67 | 68 | 63 |
| 3 | 64 | 63 | 65 | 70 |
| 4 | 72 | 70 | 68 | 70 |
| 5 | 67 | 62 | 68 | 67 |

BBA

| Window | ANN | KNN | DT | RF |
|---|---|---|---|---|
| 1 | 89 | 88 | 90 | 86 |
| 2 | 85 | 81 | 85 | 82 |
| 3 | 83 | 82 | 77 | 81 |
| 4 | 82 | 83 | 77 | 84 |
| 5 | 86 | 81 | 87 | 86 |

Elastic

| Window | ANN | KNN | DT | RF |
|---|---|---|---|---|
| 1 | 81 | 82 | 73 | 80 |
| 2 | 86 | 85 | 81 | 83 |
| 3 | 88 | 87 | 82 | 85 |
| 4 | 85 | 85 | 80 | 84 |
| 5 | 82 | 78 | 73 | 79 |

# 4 YouTube goes 5G:QoE Benchmarking and ML-based Prediction

# Introduction

- 5G technology New Radio (NR) is developed  to address high bandwidth, low latency, and massive connectivity requirements of enhanced Mobile Broadband (eMBB) compared to 4G LTE.
- In order to provide a 5G network while addressing compatibility with previous cellular systems, there are two 5G deployment options, Non-Standalone (NSA) and Standalone (SA).
- In NSA, 5G control plane relies on a pre-existing 4G core network, while SA on a dedicated 5G core network.

# Motivation

- The QoE of the YouTube video streaming from MNOs perspective is ideal and challenging.
  - To ensure better QoE, understanding and monitoring KPIs that impact users' perceived QoE has become a trending topic.
  - *Moreover, to support EFFECTOR with more use cases, i.e., Pedestrian (low mobility), Mobility (high mobility), Indoor, Outdoor.*

- Therefore, we carry out a massive 4G and 5G dataset collection campaign using a commercial 4G and 5G network, where we consider YouTube as baseline for video streaming to collect Channel Metrics and YouTube QoE logs with 1-second granularity.

# Setup & architecture

Collection of dataset in the wild – both channel level and YouTube level performance metrics:

- Use of YouTube IFRAME data API to extract player information i.e., stalls and quality shifts.
  - https://knwl.website/
- An android application to collect CLM e.g., CQI, RSRQ, RSRP, SNR, application download bitrate among other 100 + features.

- We open-source our dataset & **framework** to allow for further analysis of network coverage and end-user performance aspects.
- We carry out an extensive benchmarking of 4G and 5G using YouTube as a baseline for objective QoE i.e., stalls, quality shifts among other features.

# A short video illustrating dataset collection campaign

# Detail of solution (1/4)

4G and 5G dataset collection using the most popular video streaming website YouTube with 1-second granularity.

**YOUTUBE FEATURES:**

- Stalls, Resolution, Video Bytes Downloaded that can further provide per-session Objective QoE (i) Total Stalling Event, ii) Stalling Ratio, iii) Stalling Time, iv) Quality Shifts or Percentage of Time in a single Resolution, v) Dominant Resolution, etc.

# Detail of solution (2/4)

## CHANNEL METRICS

1-second granularity, few features below

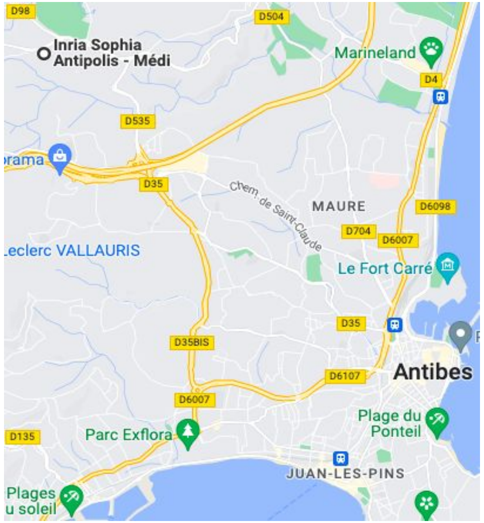
- Timestamp, Longitude, Latitude, Velocity, Operator Name, Cellid, Network Mode, Download bitrate, Upload bitrate, RSRQ, RSRP, SNR, RSSI, CQI, RSRQ and RSRP values for the neighbouring cell, among 100 + other channel metrics *.

* https://gyokovsolutions.com/manual-g-nettrack/

# Detail of solution (3/4)

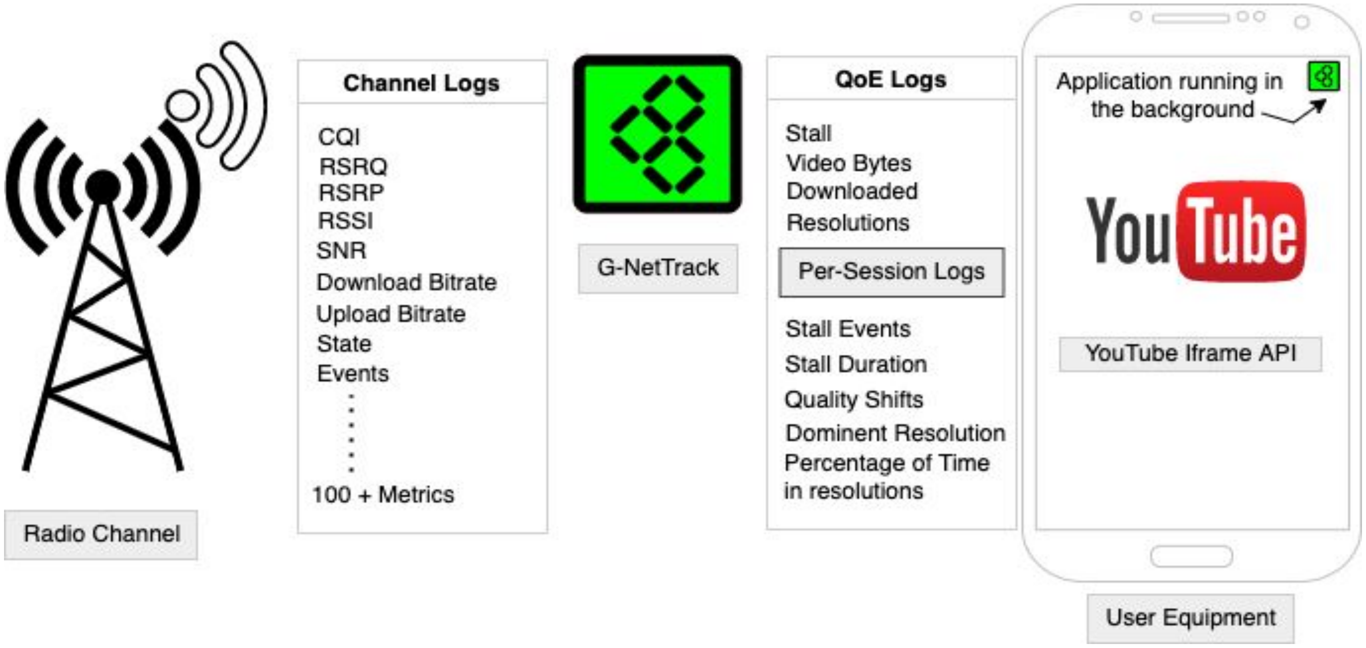
Data Collection Use Cases

1) Pedestrian – Low mobility
2) Driving – High mobility
3) Static – Bus and railway terminals
4) Static Outdoor – High crowd




# Detail of solution (4/4)

- YouTube IFRAME API for YouTube QoE Logs
- G-NetTrack Pro - Wireless network monitor and drive test tool



# Dataset statistics

| | |
|---|---|
| **Mobility – Total kilometers** | 300 + |
| **Pedestrian – Total kilometers** | 100 + |
| **Number of videos** | 13 |
| **Total video sessions** | 300 +, 1500 + Minutes streaming |
| **4G and 5G data consumed** | 300 + GB |
| **5G smartphone** | Samsung Galaxy S21 5G |
| **4G smartphone** | Samsung Galaxy S8 |

# A look at QoE & CLM KPIs

| Tech | RSRP | RSRQ | SNR | CQI | Dl_bitrate | Altitude | Height | State | Events |
|---|---|---|---|---|---|---|---|---|---|
| 5G | -102 | -3 | 19 | 13 | 538 | 146 | 146 | D | Periodic |
| 5G | -106 | -3 | 19 | 13 | 5022 | 146 | 146 | D | Periodic |
| 5G | -108 | -3 | 15 | 13 | 5022 | 145 | 145 | D | Periodic |
| 5G | -108 | -3 | 15 | 10 | 32800 | 145 | 145 | D | Periodic |
| 5G | -101 | -3 | 19 | 10 | 56346 | 145 | 145 | D | Periodic |

Events:
1. PERIODIC
2. HANDOVER_DATA_5G5G
3. IRAT_HANDOVER_DATA_5G4G
4. HANDOVER_DATA_4G4G
5. IRAT_HANDOVER_DATA_4G5G

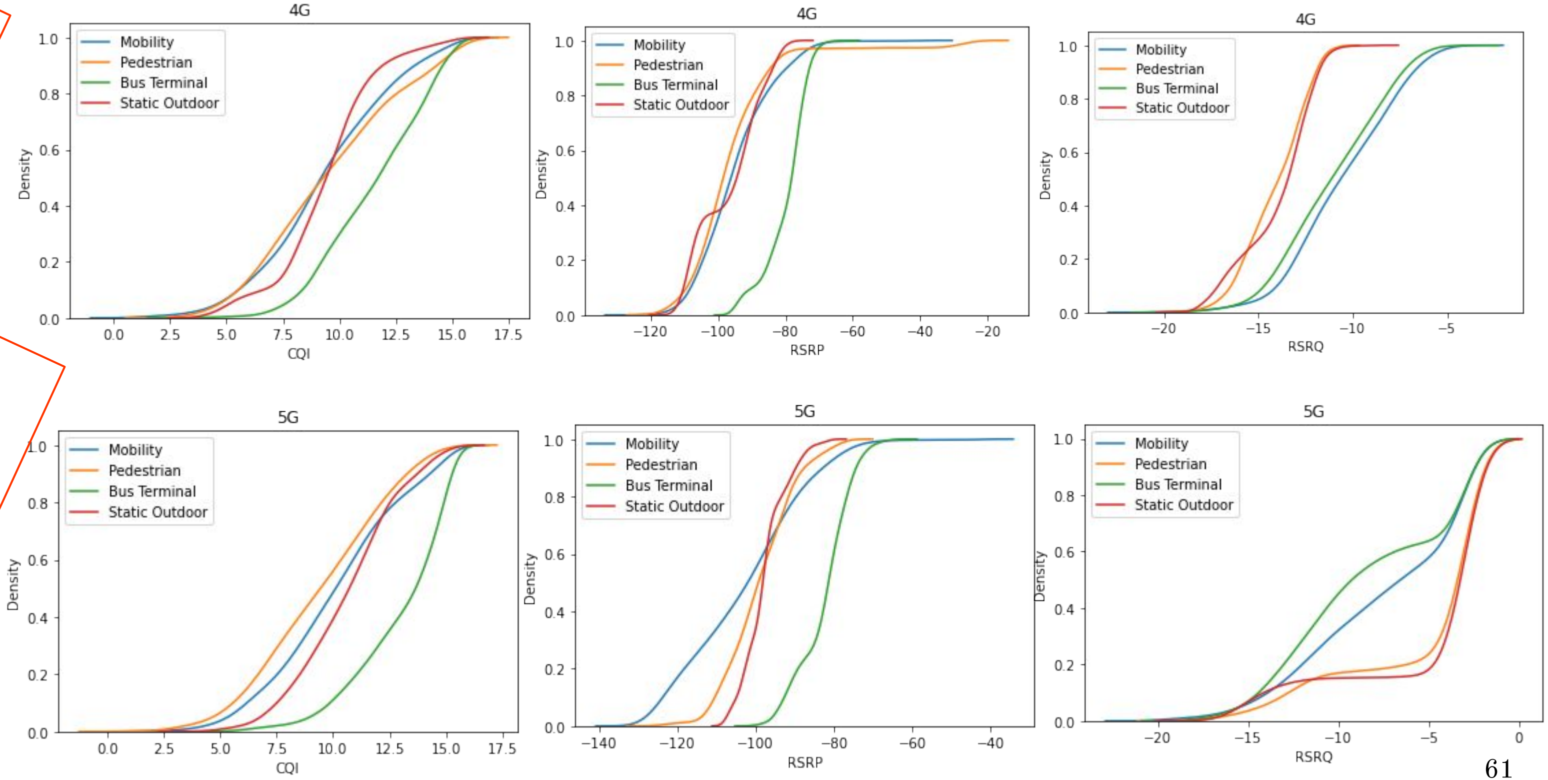
States:
1. D- Downloading
2. I - Idle

4G and 5G dataset with Channel and Context
https://github.com/razaulmustafa852/youtubegoes5g/

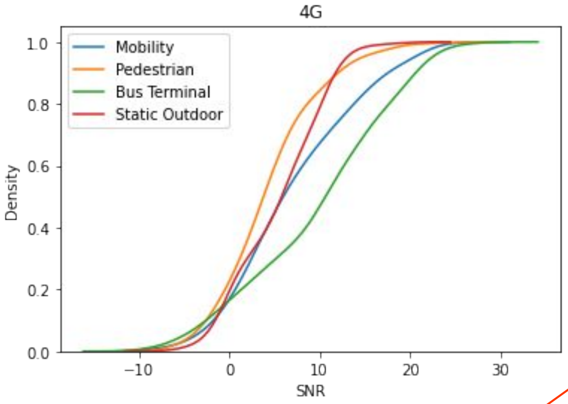
# CLM CQI, RSRP, RSRQ: 4G vs. 5G on different use cases (1/3)

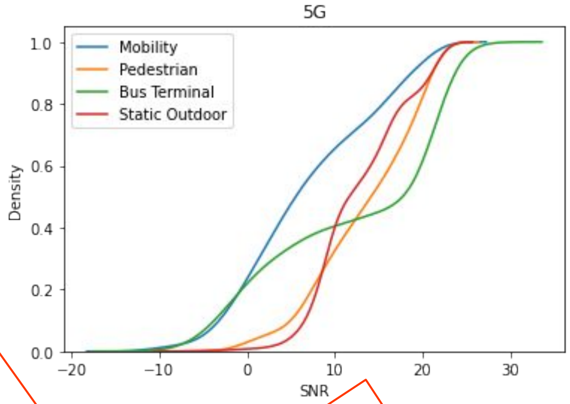
CQI is a feedback provided by UE to eNodeB

RSRP, RSRQ cell signal strength or coverage



# CLM SNR: 4G vs. 5G on different use cases (2/3)



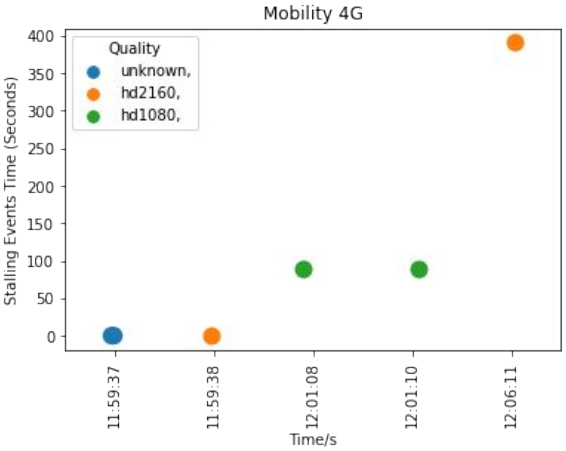

Better CQI, RSRP at bus and railway terminals

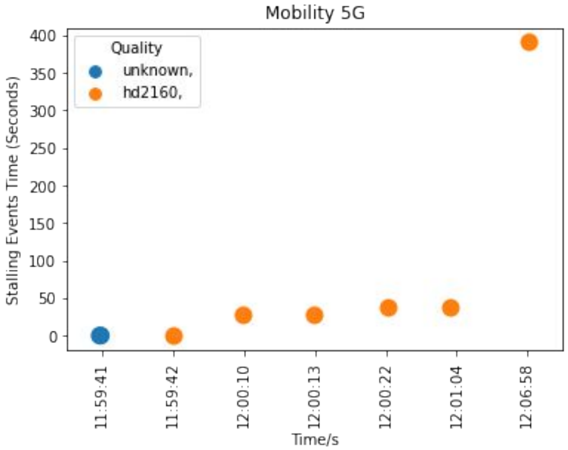
4G shows different RSRQ in mobility and pedestrian than 5G

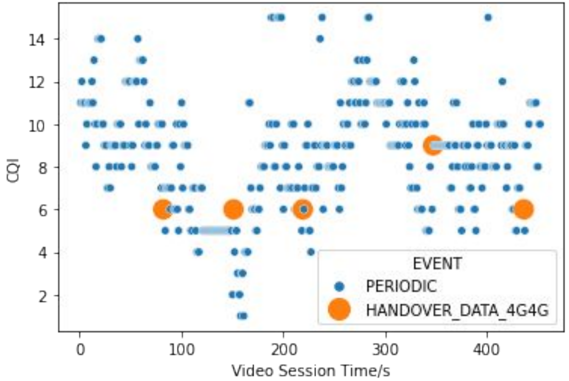
5G provides better CQI compared to 4G

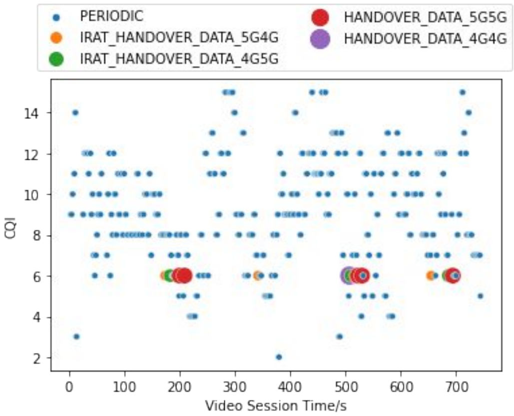
At terminals both 4G and 5G experiences better SNR

# Results (3/3) stalling and handoff event




5G in mobility suffers more stalling events as compared to 4G.

5G experiences more handover events.

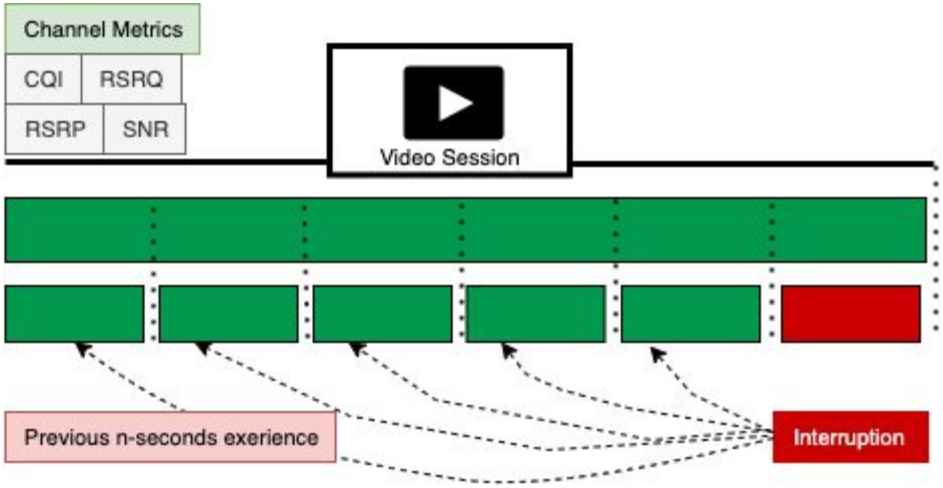
5G continues streaming at higher resolutions, even there are stalling events as compared to 4G, where we see multiple quality shifts.

Most handover occurs at CQI value 6.

# Stalling events prediction using CLM

- Look at the application QoE, and when interruption came, see the previous n-second channel metrics.
- CLM → RSRQ, RSRP, SNR, CQI
- Windows – (1-7) seconds

- CQI, RSRP, RSRQ, SNR – previous n-time (window)
- 25 %, 50 %, 75 % of a window
- Majority of a window
- Standard deviation of a window



# CQI values of 7-second window for Target Class – stall, Yes/No.

| Res | CQI-1 | CQI-2 | CQI-3 | CQI-4 | CQI-5 | CQI-6 | CQI-7 | Stall |
|---|---|---|---|---|---|---|---|---|
| **tiny** | **7** | **7** | **4** | **4** | **4** | **4** | **8** | **Yes** |
| **hd2160** | **7** | **7** | **5** | **5** | **4** | **5** | **5** | **Yes** |
| **hd2160** | **5** | **5** | **5** | **4** | **8** | **8** | **5** | **Yes** |
| **hd2160** | **8** | **5** | **5** | **5** | **4** | **4** | **4** | **Yes** |
| **hd2160** | **4** | **6** | **5** | **5** | **5** | **4** | **4** | **Yes** |
| **hd2160** | **15** | **15** | **15** | **14** | **14** | **13** | **9** | **No** |
| **hd2160** | **15** | **14** | **14** | **13** | **13** | **9** | **9** | **No** |
| **hd2160** | **14** | **14** | **13** | **13** | **9** | **9** | **13** | **No** |
| **hd2160** | **14** | **13** | **13** | **9** | **9** | **13** | **13** | **No** |
| **hd2160** | **13** | **13** | **9** | **9** | **13** | **13** | **14** | **No** |

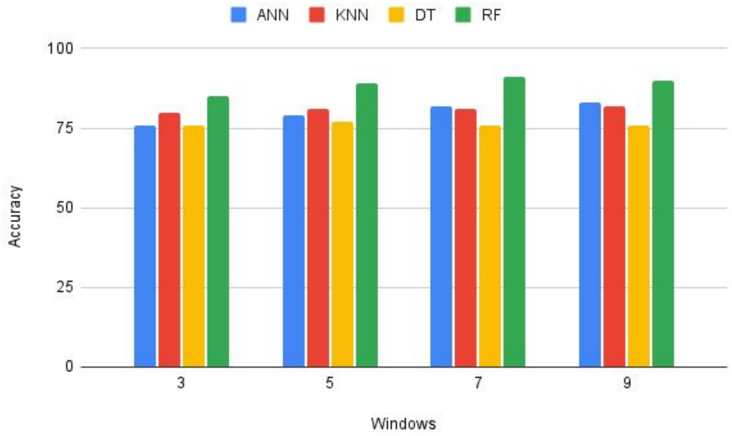
Used player events information to look at CLM metrics.

CLM metrics when there are no stalling events.

CQI values along with other metrics, i.e., RSRP, RSRQ, SNR predict stalls.

# Accuracy: Classification results / Stall vs. No Stall

| Windows | ANN | KNN | DT | RF |
|---|---|---|---|---|
| 3 | 76 | 80 | 76 | 85 |
| 5 | 79 | 81 | 77 | 89 |
| **7** | **82** | **81** | **76** | **91** |
| 9 | 83 | 82 | 76 | 90 |



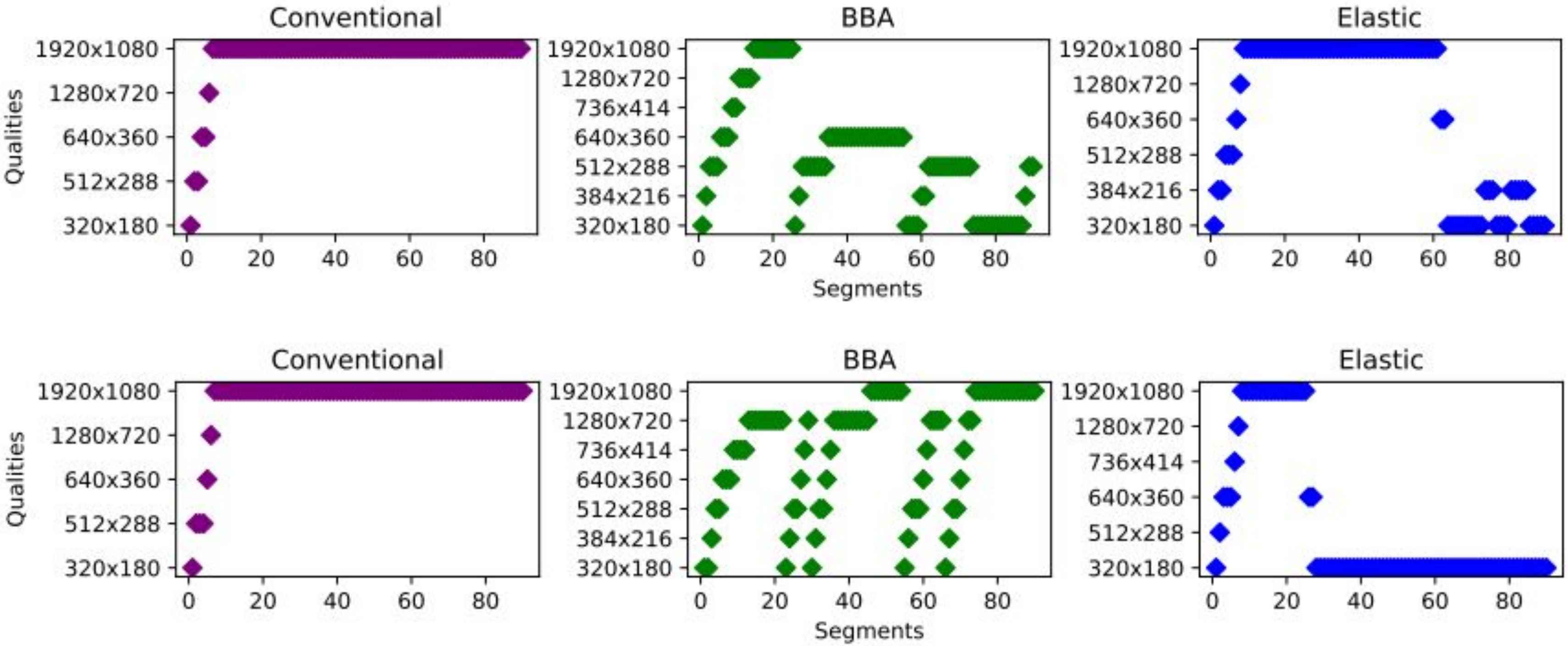
# YouTube similarity with adaptive bitrate streaming algorithms

# Real YouTube vs. emulation based QoE experiments - ABS

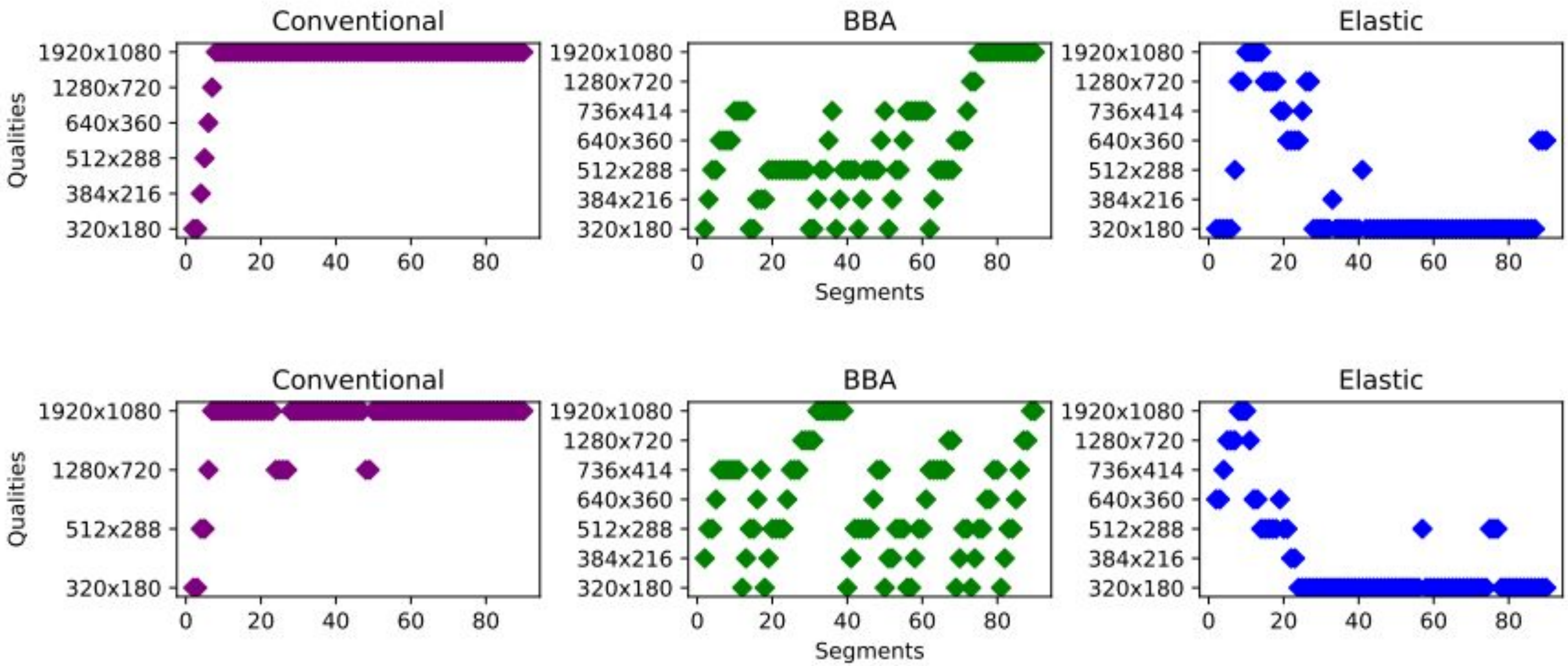
- We show the comparison of different ABS algorithms, i.e., i) Buffered – BBA, ii) Conventional – Throughput, iii) Elastic – Hybrid with YouTube player.
- Our findings show that Conventional shows more similarity with a YouTube player in terms of quality shift and dominant resolution throughout the video streaming session.
- Therefore, instead of doing experiments in the wild to draw a complex relationship between QoE and QoS, research can be done using emulation based experiments *.

* Christian Esteve Rothenberg, Danny Alex Lachos Perez,, Nathan F. Saraiva de Sousa, Raphael Rosa, **Raza Ul Mustafa**, Md Tariqul Islam, Pedro Henrique Gomes. Intent-based Control Loop for DASH Video Service Assurance using ML-based Edge QoE Estimation. In 6th IEEE International Conference on Network Softwarization (NetSoft'20) - Demo Session, Ghent, Belgium. Jun 2020.

# TCP: Quality shifts in all ABS with 4G & 5G mobility use cases



# QUIC: Quality shifts in all ABS with 4G & 5G mobility use cases



# Conventional: goDash player vs. YouTube QoE KPIs

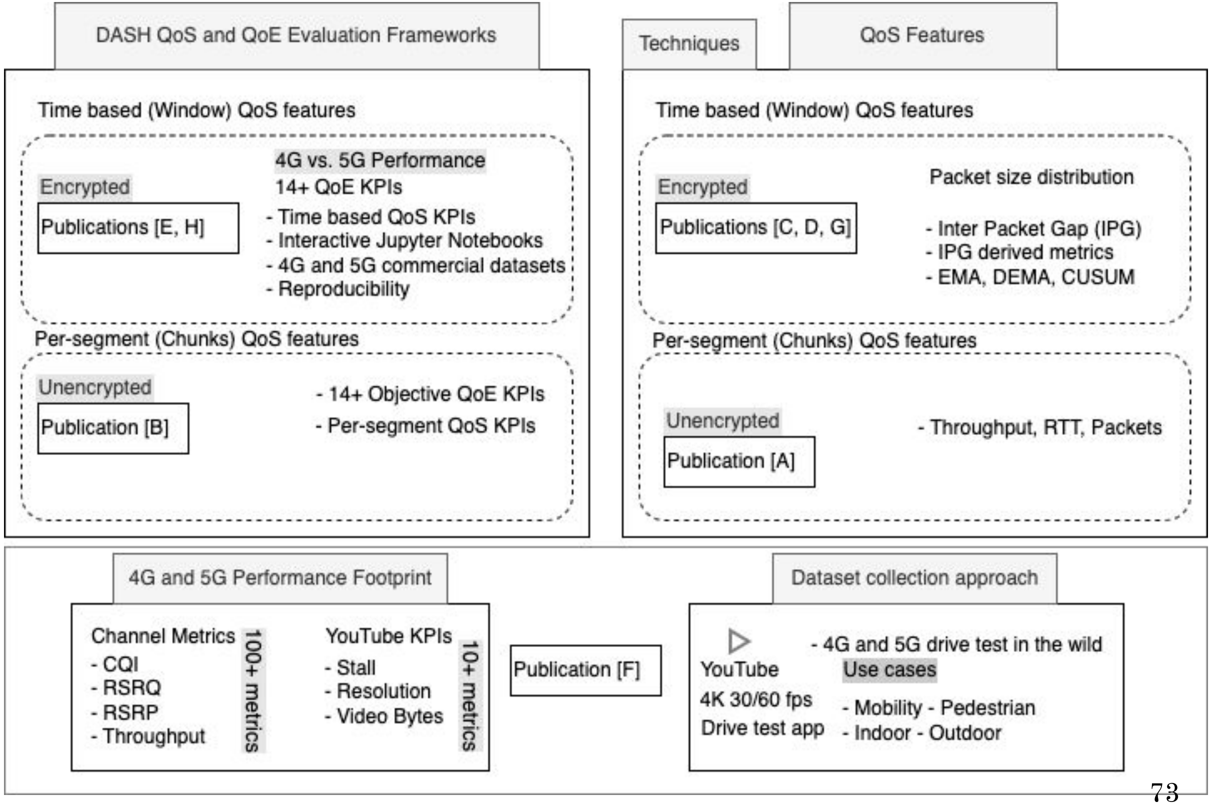
| Parameters | Emulation | YouTube |
|---|---|---|
| QoE degradation time | 20s | 20s |
| Max resolutions | 88.4% | 90.4% |
| Low resolutions | 7.1% | 8.6% |
| Stalls | Yes | Yes |
| Stalls ratio | Max resolutions – 78 % | Max resolutions – 100% |

- Time spent with QoE degradation
- Percentage of maximum time of streaming in higher resolutions
- Percentage of maximum time in lower resolutions
- Stalling events
- Stall in maximum resolution during the streaming session

# 5 Contributions & Publications

# Contributions vs. Publications

- DASH Frameworks
  - Per-segment
  - Realtime- EFFECTOR
- 4G vs. 5G performance footprint and QoS to map QoE using ML.
- Jupyter books to analyze various 4G and 5G use cases.
- A framework for the collection of real 4G and 5G dataset
  - A massive dataset with CLM and YouTube KPIs
  - YouTube stall prediction with CLM



# Publications

A. **Raza Ul Mustafa**, Simone Ferlin, Christian Esteve Rothenberg, Darijo Raca, Jason J. Quinlan, ``**A Supervised Machine Learning Approach for DASH Video QoE Prediction in 5G Networks**'' Proceedings of the 16th ACM Symposium on QoS and Security for Wireless and Mobile Networks, ACM, 2020.

B. **Raza Ul Mustafa**, Md Tariqul Islam, Christian Rothenberg, Darijo Raca, Jason J. Quinlan, ``**DASH QoE performance evaluation framework with 5G datasets**'' 2020 16th International Conference on Network and Service Management (CNSM) AnServApp Workshop. IEEE, 2020.

C. **Raza Ul Mustafa**, David Moura, Christian Esteve Rothenberg ``**Machine Learning Approach to Estimate Video QoE of Encrypted DASH Traffic in 5G Networks**''. In IEEE Signal Processing (SSP2021) Workshop, 2021.

D. **Raza Ul Mustafa**, Christian Esteve Rothenberg ``**Machine Learning Assisted Real-time DASH Video QoE Estimation Technique for Encrypted Traffic**'' In ACM MHV, 2022.

E. **Raza Ul Mustafa**, MD. Tariqul Islam, Christian Esteve Rothenberg, Pedro Henrique Gomes ``**A Framework For QoS and QoE Assessment of Encrypted Video Traffic With 4G and 5G Open Datasets**'' In IEEE Globecom Demo Session, 2022.

F. **Raza Ul Mustafa**, Chadi Barakat, Christian Esteve Rothenberg, "**YouTube Goes 5G: Benchmarking YouTube in 4G vs 5G Through Open Datasets**", In IEEE Globecom Demo Session, 2022.

G. **Raza Ul Mustafa**, Christian Esteve Rothenberg, ``**Machine Learning Assisted Real-time DASH Video QoE Estimation Technique for Encrypted Traffic**'', In Submission Journal of Network and Systems Management (JNSM), 2023.

H. **Raza Ul Mustafa**, Md Tariqul Islam, Christian Rothenberg, Pedro Henrique Gomes. ``**EFFECTOR: DASH QoE and QoS Evaluation Framework For EnCrypTed videO tRaffic**'' In Submission IEEE/IFIP Network Operations and Management Symposium (NOMS), 2023.

# Further results & collaborative activities

- Framework Repository Github, Dash Quality of Experience Open Source Evaluation Framework,
  - https://github.com/razaulmustafa852/dashframework-1
- Framework Repository Github, EFFECTOR
  - https://github.com/razaulmustafa852/EFFECTOR
- YouTube and Channel Metrics Repository Github
  - https://github.com/razaulmustafa852/youtubegoes5g
- IEEE Dataport Submission
  - DOI: 10.21227/h00h-ew92
- Christian Esteve Rothenberg, Danny Alex Lachos Perez,, Nathan F. Saraiva de Sousa, Raphael Rosa, **Raza Ul Mustafa**, Md Tariqul Islam, Pedro Henrique Gomes. Intent-based Control Loop for DASH Video Service Assurance using ML-based Edge QoE Estimation. In 6th IEEE International Conference on Network Softwarization (NetSoft'20) - Demo Session, Ghent, Belgium. Jun 2020.
- Nathan F. Saraiva de Sousa, Md Tariqul Islam, **Raza Ul Mustafa**, Danny Alex Lachos Perez, Christian Esteve Rothenberg, Pedro Henrique Gomes, ``Machine Learning-Assisted Closed-Control Loops for Beyond 5G Multi-Domain Zero-Touch Networks'', JNSM, 2022.

# 6 Limitations, Conclusions & Future Work

# Limitations – frameworks & ML work

- DASH QoS to QoE evaluation frameworks are equipped with all the dependencies to run 4G and 5G use cases with commercial 4G and 5G datasets collected in the wild.
    - More videos and topologies can be used to generalize the QoS features extraction approach
    - We use a headless goDASH player, however; in reality, users are streaming video content from different OTT Platforms, e.g., YouTube, Amazon, and Netflix
    - Devices also impact QoE, i.e., Mobile, PC, and Tablets. QoS is also impacted by various other factors, which include streaming using official apps, or by using Browsers – Chrome, Mozilla, etc
    - Increase the number of DASH client to see the impact of QoS - QoE
    - The proposed technique requires practical deployment for the evaluation. We are unaware of the computational complexity, such as CPU, Memory, and Storage, for large deployments

# Limitations - CLM & YouTube

- The dataset collected in this work is collected using a web-based application, which uses YouTube IFRAME API.
- We used the browser to open the application.
  - Therefore there might be a chance of a different QoE than YouTube Android / iOS application
- Moreover, the dataset collection is done using two android devices, one for 4G and one for 5G.
  - However, we do not consider multi-user streaming of the same content simultaneously
- Moreover, during the dataset collection campaign, we consider the full width of YouTube player, which automatically adjusts to the viewport of the device.
  - However, different screen sizes may influence QoE

# Challenges throughout the research work

- Upto 30+ minutes to do a single experiment in DASH environment
  - Experiments combinations
    - 3+ ABS
    - Technology - 4G , 5G
      - Different use cases
      - Repetitions
    - Videos - BBB, Sintel, Tears
- 40+ minutes required to do a single dataset collection campaign
  - Use cases
    - Pedestrian - 15KM walk on daily basis
    - Mobility - Driving 30+ KM on daily basis
    - Indoor
    - Outdoor

# Conclusion (1/5)

- Throughout this thesis, the entire line of reasoning fits according to finding highly correlated Quality of Service (QoS) Metrics to map QoE followed by Machine Learning (ML) techniques.
- We proposed a window based QoS features extraction approach of DASH videos using only Inter Packet Gap (IPG) as baseline.
  - We derive more features from IPG, such as EMA, CUSUM to observe the continuity of datapacket overtime
  - We find that IPGs along with other basic QoS metrics are highly correlated to objective QoE KPIs
- Moreover, to conduct large scale 4G and 5G experiments, we provide DASH QoS to QoE evaluation frameworks, which provides highly correlated QoS features to investigate QoS and QoE.
- To run large scale experimentation we provide a large number of commercial 4G and 5G use cases to emulate them in frameworks.
  - Mobility, pedestrian, indoor, outdoor, railway and bus terminals
  - We also made the setup open-source to create more realistic use cases and then to emulate them, i.e., dataset with extreme mobility, weather conditions, multiple devices etc

# Conclusion (2/5)

- We compare the performance footprint of 4G and 5G using different state-of-the-art ABS algorithms, i) throughput, ii) buffered, iii) hybrid.
  - We find that throughput based adaptive bitrate streaming algorithms are quite similar with YouTube streaming environment
    - Thus, YouTube QoS and QoE can be emulated in the frameworks to draw more complex relationship between QoS and QoE, i.e., shifts, stalling events, bitrate
  - We also compared the performance footprint of TCP vs. QUIC using mobility use cases, and found that throughput and hybrid based ABS algorithms are showing similarities. However, differences in technology
  - Buffered based experience more shifts

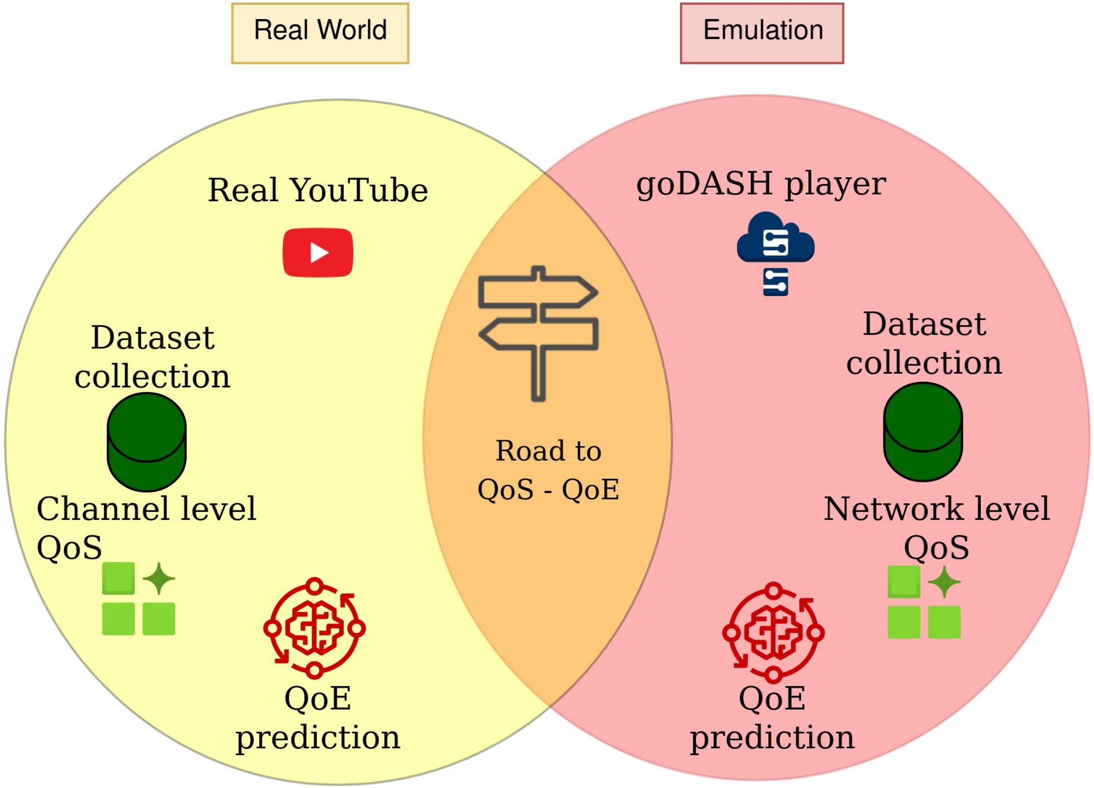
# Conclusion (3/5)

- We build datasets over TCP and QUIC for DASH video content and use various Machine Learning techniques. The dataset is composed of various time windows in seconds, i.e., (0.5, 1, 2, 3, 4, 5).
- We conclude that Random Forests and Artificial Neural Network (ANN) provides the best results, when dealing with encrypted QoS features to estimate QoE KPIs.
- Apart from finding highly correlated QoS features to estimate QoE, we provide real 4G and 5G datasets collected in the wild in different regions.
- We concluded that 5G outperforms 4G in video streaming, however, this is not the true case most of the time.
  - 5G requires a stable connection to provide maximum perceived video quality
  - However, it suffers from stalling events in the case of Mobility due to frequent Intra-RAT HO and Inter-RAT HO
- Therefore, we conclude to provide better QoE in 5G; YouTube players must be aware that 5G does not always provide the best video QoE experience; thus, it requires following the 4G ABR algorithm.

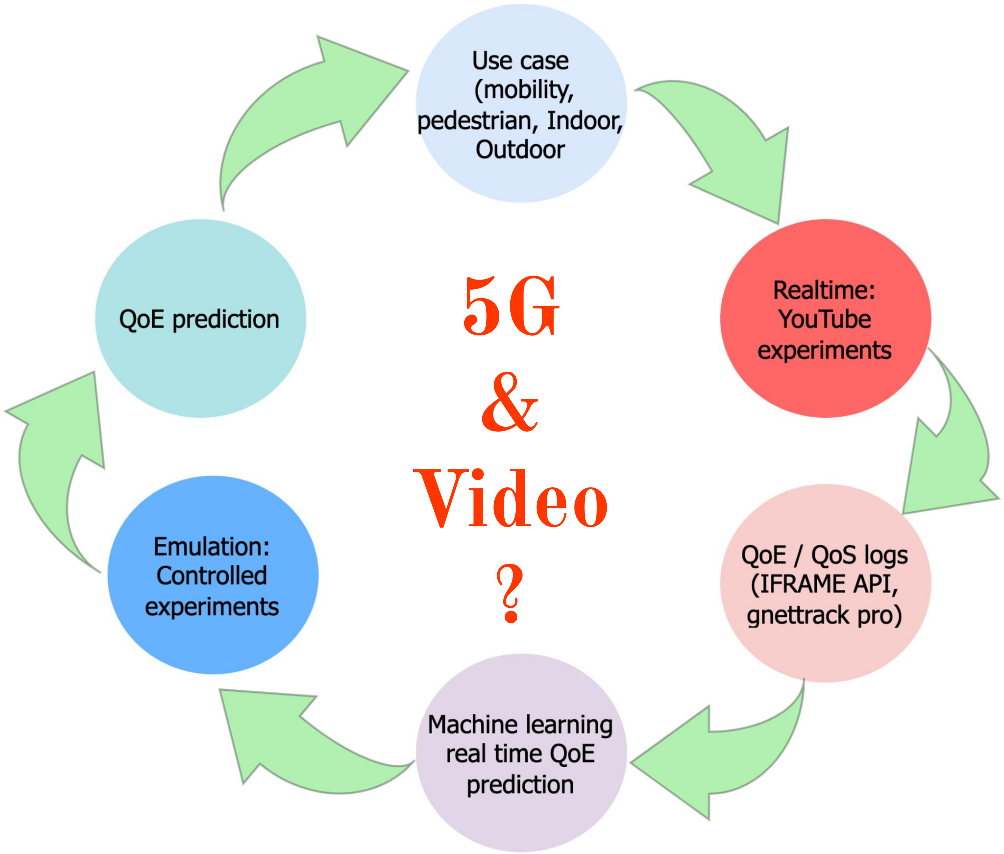
# Conclusion (4/5)

- Moreover, we also conclude that Channel Level Metrics (CLM), i.e., CQI, SNR, RSRQ, and RSRP, are correlated to stalling events in real YouTube traffic.
- We propose a window based stalling event prediction technique to predict the binary classification of Stall vs. No Stall.
- We evaluated the window up to 9 seconds and found the best slot, which is highly correlated with stalling events along with other features.

# Conclusion (5/5)



# Closing the loop



# Future work

6

- Extension of EFFECTOR with more complex network scenario and scalability, Mininet-WiFi access node to replicate 4G and 5G trace's channel condition (e.g., SINR, RSRP/Q, CQI).
- Find more lightweight QoS features such as Progressive Mean and EMA-CUSUM (mix) using IPGs as a baseline.
- In real-time YouTube QoE estimations, future work can be done in many directions. For instance, Quality-shifts has well known QoE metrics that influence the MOS.
  - Investigate the CLM factors affecting the shifts – i) Up, ii) Down
  - For 5G-aware streaming a recommender system
  - Application continuously monitors the location and mobility patterns to avoid stalls, thus, maximum QoE
- A new ABR algorithm for a famous video streaming platform – YouTube, which takes decisions based on current QoE experience rather than based on technology, i.e., (5G).
- Address identified limitations.

# Collaborators

Jason J. Quinlan
University College Cork, Ireland

Simone Ferlin
Ericsson AB, Sweden

Darijo Raca
University of Sarajevo, Sarajevo, BiH, Bosnia

Chadi Barakat
INRIA Cote d'azur France

# Thanks

Questions ?

razaulmustafa@outlook.com

INFORMATION & NETWORKING TECHNOLOGIES RESEARCH & INNOVATION GROUP