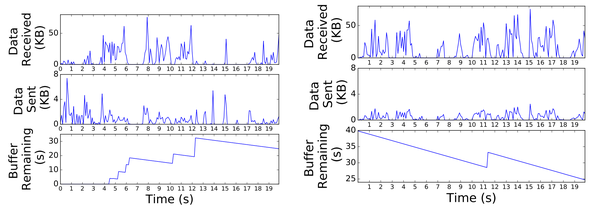
Inferring YouTube Streaming QoE from Encrypted Traffic

Trey Gilliland*, Craig Gutterman*, Sarthak Arora*, Katherine Guo^{o}, Les Wu^{o}, Xiaoyang Wang^{o}, Gil Zussman*
*Electrical Engineering, Columbia University; ^{o}Bell Labs

Objective

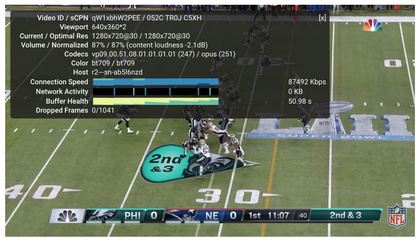
- Content providers can measure Quality of Experience (QoE) metrics
- End-to-end encryption does not allow network providers to examine video session information using deep packet inspection
- Design basic features based on IP header information to predict events that cause QoE impairments ahead of time so network operators can proactively use adaptive resource provisioning

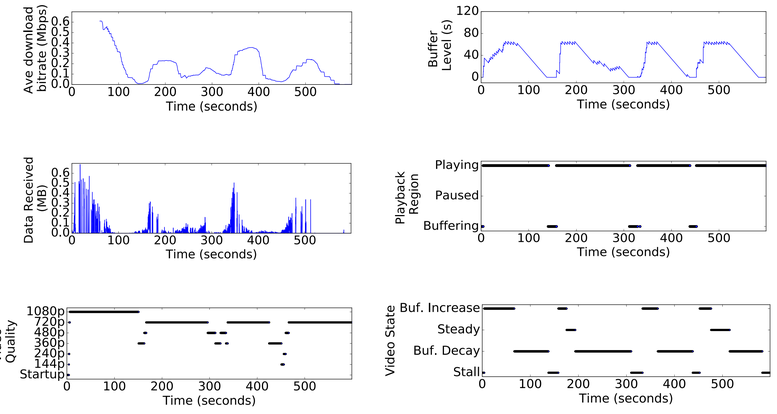


Network throughput does NOT provide accurate QoE prediction

Quality of Experience (QoE)

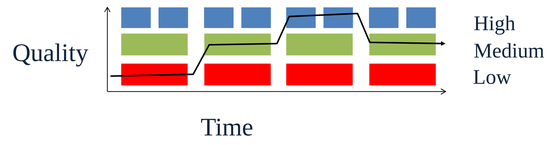
- Stall events have the largest negative impact on end user engagement
- Higher average video playback bitrate improves user engagement





Video Streaming

- Adaptive BitRate (ABR) streaming: Each clip is encoded in multiple resolutions. A clip with a given resolution is then divided into a number of chunks of variable length in playback time

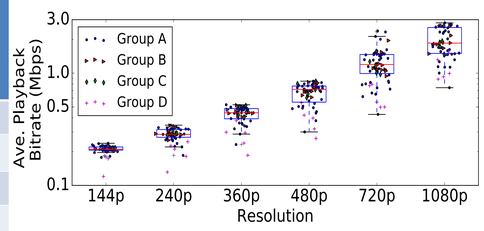


Experimental Setup

- YouTube IFrame API/automated ADB program collects video statistics information from “Stats for Nerds”
- Tshark collects network level data

Data Sets

	Clip Length (min)	Session Length (min)	No. of Unique Clips
A	8-12	10	40
B	8-12	10	10
C	3-5	5	5
D	25-120	30	5



Machine Learning Models

- Use 4-fold cross validation to determine accuracy of best ML algorithm and features
- Basic Network Features (e.g., protocol, Number of uplink/downlink packets and bytes)
- Chunk Based Features (e.g., chunk size)
- Random Forest

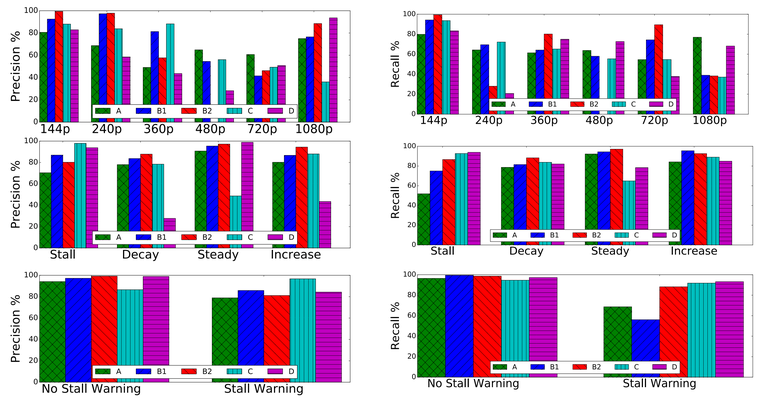
Evaluation

- Build model from sessions in lab environment
  - Use ~400 sessions from 40 video IDs in lab environment
  - Use 4-fold cross validation (10 non-overlapping video IDS in each fold)
    - Training: Sessions from 30 video IDs
    - Testing: Sessions from 10 video IDs
- Further testing
  - Use chunk model trained in lab environment to test on unseen video IDs in different environments
  - Dataset B1 in US residential, B2 in India residential, C and D in lab environment

	Basic		Chunk	
	Precision	Recall	Precision	Recall
Stall	31.1%	7.6%	70.4%	51.9%
Buffer Decay	32%	16.3%	78.0%	78.7%
Buffer Steady	57.6%	80.2%	90.7%	92.2%
Buffer Increase	64.1%	57.6%	80.2%	84.2%

	Basic		Chunk	
	Precision	Recall	Precision	Recall
No Stall Warning	86%	98.1%	94.1%	96.5%
Stall Warning	51%	11.1%	79.0%	68.7%

	Basic		Chunk	
	Precision	Recall	Precision	Recall
144p	13%	7.6%	80.6%	79.9%
240p	14.6%	10.1%	68.7%	64.3%
360p	14.1%	9.9%	49.2%	64.4%
480p	24.7%	33.3%	64.9%	63.8%
720p	24.5%	30.3%	60.6%	54.5%
1080p	22.2%	20.1%	75.0%	76.9%



Supported in part by NSF grant CNS-1650685