Dynamic Predictive Streaming of 360 Degree Video

Yong Liu

joint work

Yao Wang, Liyang Sun, Fanyi Duanmu, Chenge Li, Lifan Mei

Electrical & Computer Engineering

Tandon School of Engineering

New York University

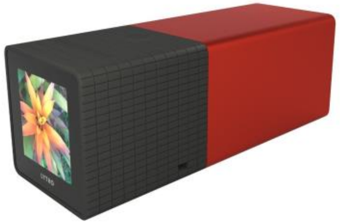
Outline

- Motivation and Challenges
- Two-Tier 360V Streaming
- Predictive & Prioritized Buffer Control
- System Settings and Evaluations
- Conclusion

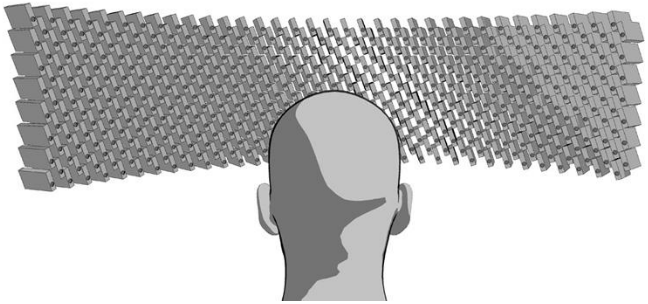
VR/AR next generation “killer-apps”

- ❑ Virtual Reality: immersive three-dimensional environment which can be explored and interacted with by a person

- ❑ 360° View Scene Generation
  - Real environment captured by 360° Camera: recorded, live broadcast



  - computer-generated environment, e.g., multi-player game
- ❑ View Scene Display



VR/AR Networking Challenges

- high bandwidth/low delay/reliability/mobility…
- example: 360 degree video streaming

	Entry-Level	Advanced	Ultimate
Resolution	8K 2D (7680*3840)	12K 2D (11520*5760)	24K 3D (23040*11520)
Equivalent TV resolution	480P	2K	4K
Frame rate	30	60	120
Compression ratio	165:1	215:1	350:1 (3D)
Typical video bit rate	64 Mbit/s	279 Mbit/s	3.29 Gbit/s
Network bandwidth requirement	100 Mbit/s	418 Mbit/s	1 Gbit/s (smooth play)(FoV) 2.35 Gbit/s (interactive)(FoV)
Latency requirement	30 ms	20 ms	10 ms

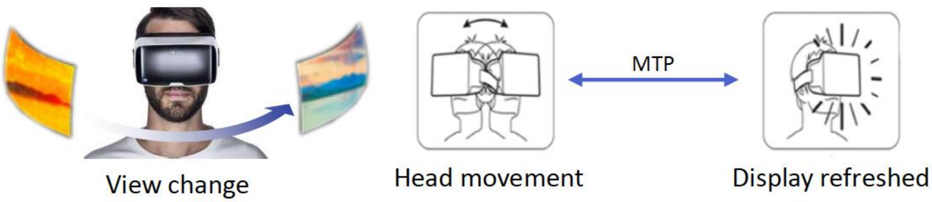
Reference

- http://www-file.huawei.com/~/media/CORPORATE/PDF/white%20paper/whitepaper-on-the-vr-oriented-bearer-network-requirement-en.pdf

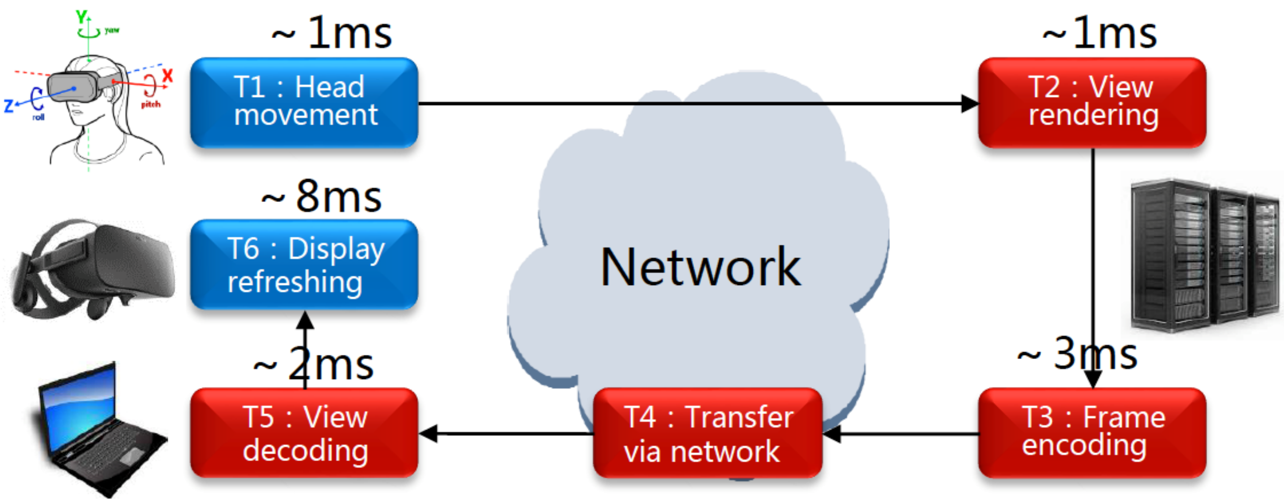
VR/AR Networking Challenges

- high bandwidth/low delay/reliability/mobility…
- example: 360 degree video streaming

Motion-to-Photon (MTP) delay ≤ 20ms



Budget for network delay ≤ 10ms!

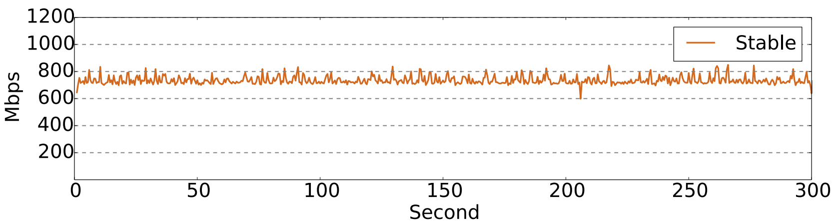


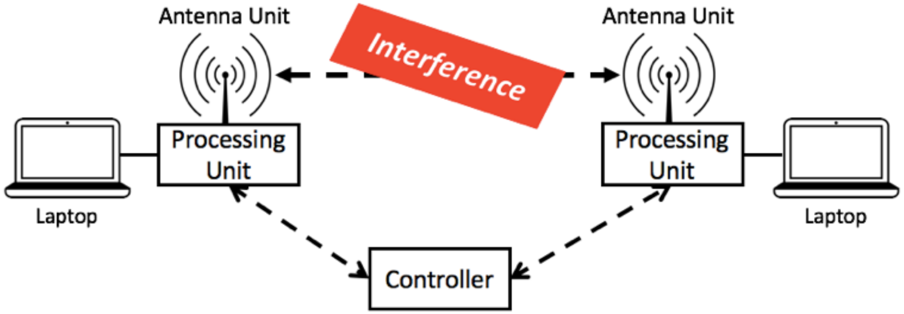
How to Make It Happen?

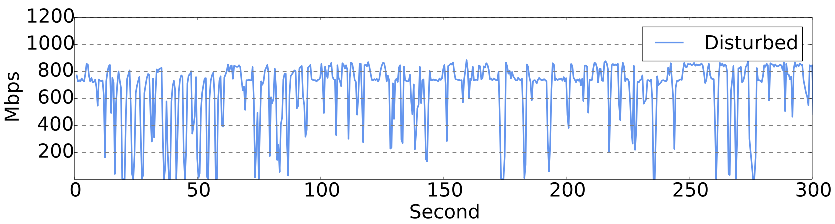
- Innovations at all layers of protocol stack?
  - Massive MIMO, millimeter wave, SDN, NFV, QUIC, HTTP/2, HTTP/3 etc.
- Novel Network Architecture?
  - Cloud, Fog, MEC, ……
- Creative Service/Business Models?
  - ISP&CSP, CDN, App, ……
- Interdisciplinary and User-centric Designs?

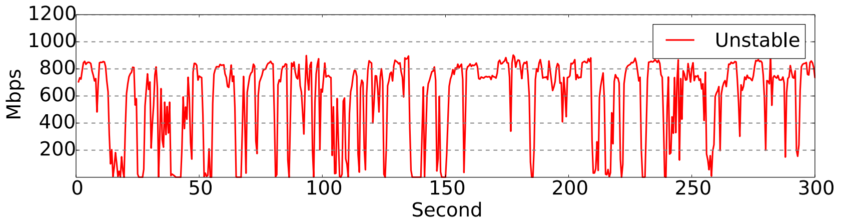
5G Promises & Challenges

- Special Properties
  - High speed, up to 10 Gbps
  - Low latency, down to 1ms
  - High volatility: 28GHz, 60GHz => path loss quadratic increase, vulnerable to blockage, "on-off"
- Testbed
  - WiGig (802.11ad, wireless communication, 60Ghz)
  - Trace Collection







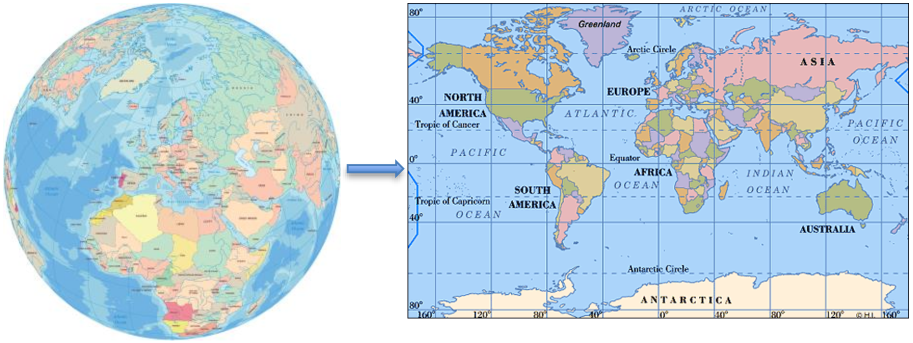


Outline

- Motivation and Challenges
- Two-Tier 360V Streaming
- Predictive & Prioritized Buffer Control
- System Settings and Evaluations
- Conclusion

360° Video Streaming

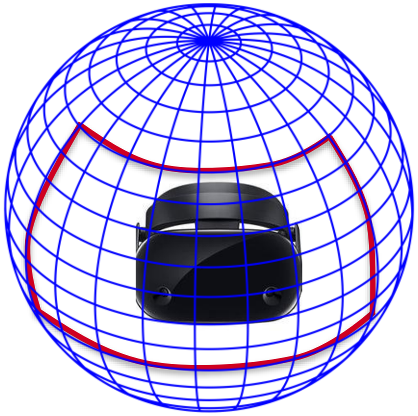
- ❑ 360° Video Scene Capturing
- ❑ Projection to 2D Planar Video
  - e.g. equirectangular mapping



  - pyramid, cube-map, ……
- ❑ Coding and Streaming Projected 2D video
- ❑ Rendering based on user Field-of-Vision (FoV)

Bandwidth Inefficiency

- A user only watches a small fraction of video within her current Field-of-View (FoV)



Fraction of area within FoV: 120°/360° x 90°/180°=1/6

FoV Streaming

- Only stream video within current FoV
  - b.w. challenge: five-fold saving! ☺
  - delay challenge: respond to user FoV dynamics?
    - download video in updated FoV within 20 ms (MTP) ☹?
- FoV Pre-fetching in VoD:
  - predict user FoV in near future, e.g., seconds.
  - pre-fetch video in predicted FoV
  - render with zero latency ☺ if pre-fecthed FoV matches actual FoV
  - download actual FoV if prediction is wrong ➔ long latency ☹

Two-tier 360° Video Streaming

- Two-tier video encoding:
  - Base-tier (BT) chunks cover entire 360° view scene in low quality
  - Enhancement-tier (ET) chunks cover view-ports in different directions at multiple quality levels
- Two-tier video streaming
  - download BT chunks with long prefetching buffer
  - download ET chunks in predicted FoV with short prefetching buffer, quality level dynamically selected
    - pre-fetched ET chunks useless if not falling into user FoV
- Two-tier video rendering
  - if buffered ET chunks match user actual FoV, render high quality video in FoV
  - otherwise, render low quality video in FoV based on buffered BT chunks.

Two-tier 360° Video Streaming

- Challenges: network b.w. & user FoV dynamics
- Realtime streaming decisions
  - given current bandwidth and FoV
    - download BT or ET chunks? (instant quality vs. long-term robustness)
    - which ET chunks?
      - quality level? (DAS problem)
      - FoV? (FoV prediction)
  - trade-offs
    - rendered video quality, streaming continuity, and responsiveness to network & user dynamics

Outline

- ❑ Motivation and Challenges
- ❑ Two-Tier 360V Streaming
- ❑ Predictive & Prioritized Buffer Control
- ❑ System Settings and Evaluations
- ❑ Conclusion

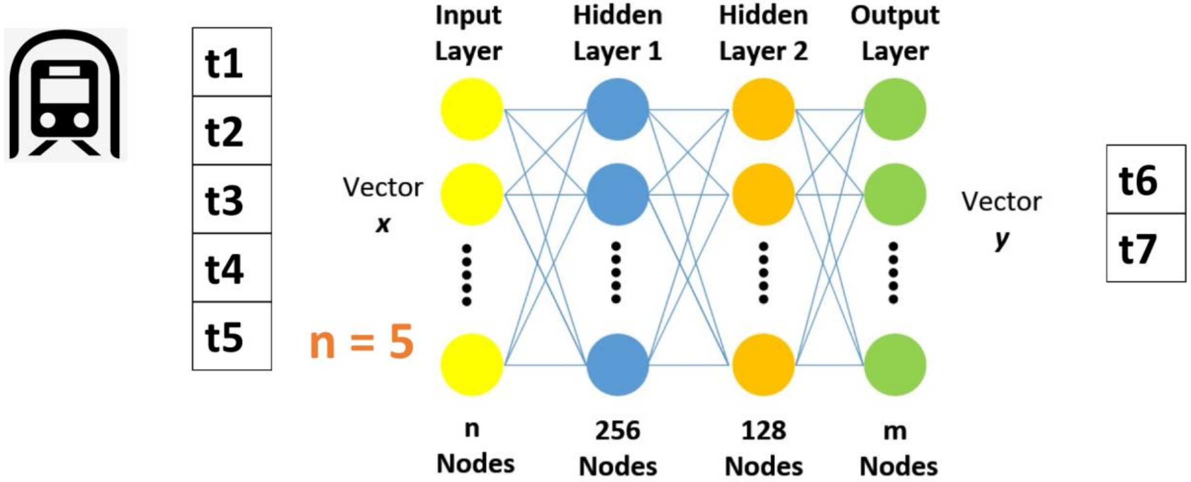
Predictive & Prioritized Buffer Control

- Prioritized buffer-based strategy
  1. prioritize base-tier (BT) downloading to ensure video playback continuity.
     - if buffered BT video time $q^b(t)<q^b_{ref}$, sequentially download BT chunks
  2. enhancement-tier (ET) downloading to utilize residual bandwidth
  3. ET quality-level determined by b.w. prediction and accumulated ET buffer length
  4. ET FoV determined by FoV prediction algorithm

Realtime Mobile Bandwidth Prediction

- Predict available bandwidth for mobile user in next five seconds
  - collect mobile 4G b.w. traces from different transport methods/routes
    - subway, bus, train, ferry, ….
  - offline train LSTM RNNs, one for each transport method/route
  - online prediction based on b.w. measurement in previous five seconds

LSTM Model Training



Realtime Mobile Bandwidth Prediction

- Predict available bandwidth for mobile user in next five seconds
  - collect mobile 4G b.w. traces from different transport methods/routes
    - subway, bus, train, ferry, ...
  - offline train LSTM RNNs, one for each transport method/route
  - online prediction based on b.w. measurement in previous five seconds

Table 1: Evaluation Results on NYC MTA Traces

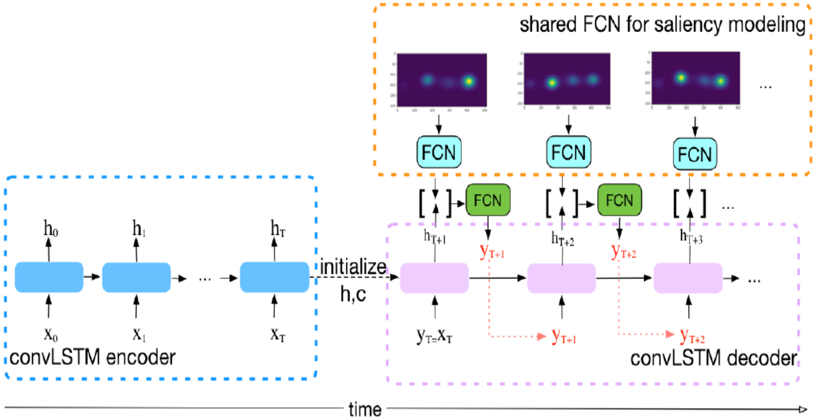
	7A Train	7B Train	Bus 57	Bus 62	N Train
Testset Average	6.41	4.55	9.43	2.62	8.94
RLS RMSE	2.83	2.12	2.71	0.85	3.22
Harmonic RMSE	2.99	2.56	2.86	0.98	3.35
LSTM RMSE	**2.29**	**2.05**	**2.50**	**0.75**	**2.95**
RLS Error Ratio	44.15%	46.59%	28.74%	32.44%	36.02%
HAR Error Ratio	46.65%	56.26%	30.33%	37.40%	37.47%
LSTM Error Ratio	35.73%	45.05%	26.51%	28.63%	33.00%

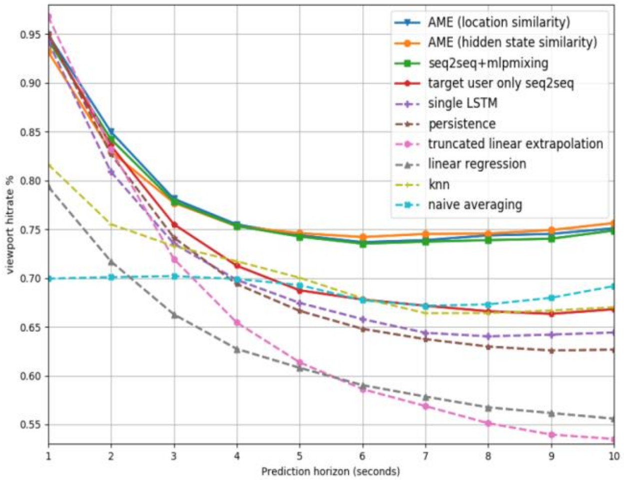
NYU-METS Dataset: https://github.com/NYU-METS/Main

"Realtime Mobile Bandwidth Prediction using LSTM Neural Networks", Lifan Mei, Runchen Hu, Houwei Cao, Yong Liu, Zifa Han, Feng Li, and Jin Li, Passive and Active Measurement (PAM), 2019

Very Long-Term FoV Prediction

- Predict FoV of target user in next ten seconds
  - own past FoV trajectory
  - other users' past and future trajectories
  - video saliency map
  - ……





"Very Long Term Field of View Prediction for 360-degree Video Streaming", Chenge Li, Weixi Zhang, Yong Liu, and Yao Wang, 2019 IEEE Conference on Multimedia Information Processing and Retrieval

Outline

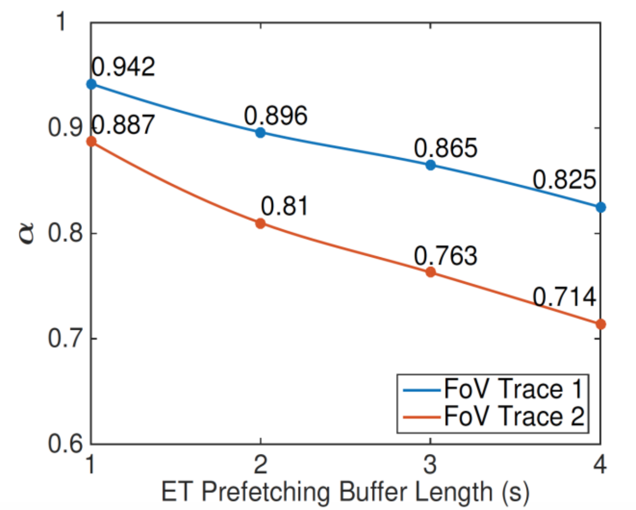
- Motivation and Challenges
- Two-Tier 360V Streaming
- Predictive & Prioritized Buffer Control
- System Settings and Evaluations
- Conclusion

Setting Target ET Buffer Length

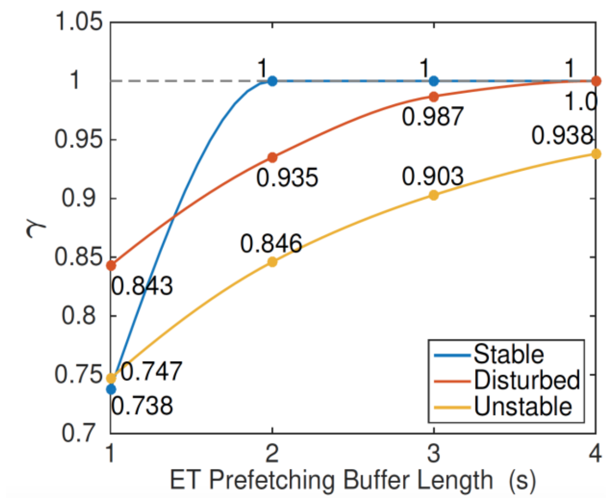
- Setting of Target ET buffer length q^e_{ref}
  - long ET buffer
    - increase robustness against b.w. variations
    - vulnerable to FoV dynamics
      - hard to predict FoV long into future
      - pre-fetched FoV wasted if prediction wrong
  - short ET buffer
    - more accurate FoV prediction
    - vulnerable to b.w. variations

Setting Target ET Buffer Length

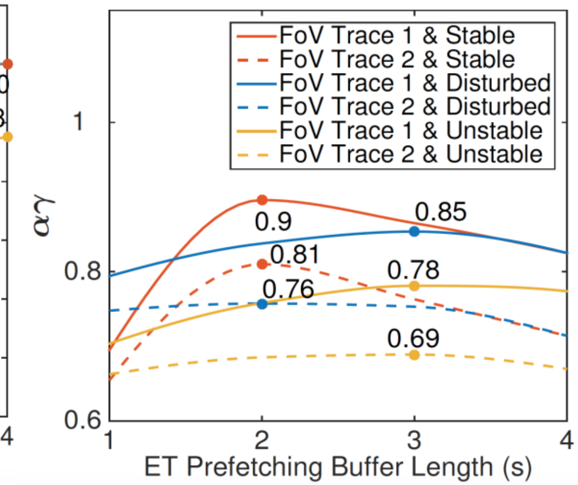
- maximizing average rendered video quality
  - $$Q(R_b; \alpha, \gamma, R_t) = \alpha\gamma Q_e(\tilde{R}_e) + (1 - \alpha\gamma)Q_b(\tilde{R}_b)$$
    $$= \alpha\gamma Q_e\left(\frac{R_b}{A_b} + \frac{R_t - R_b}{A_e}\right) + (1 - \alpha\gamma)Q_b\left(\frac{R_b}{A_b}\right)$$



FoV prediction accuracy



ET chunk delivery ratio



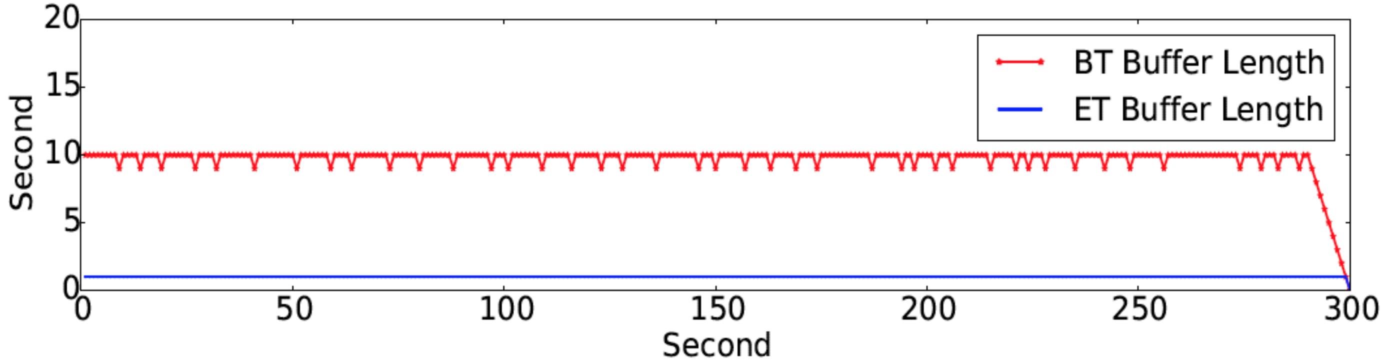
Effectiveness of ET chunk

Optimal Rate Allocation for WiGig

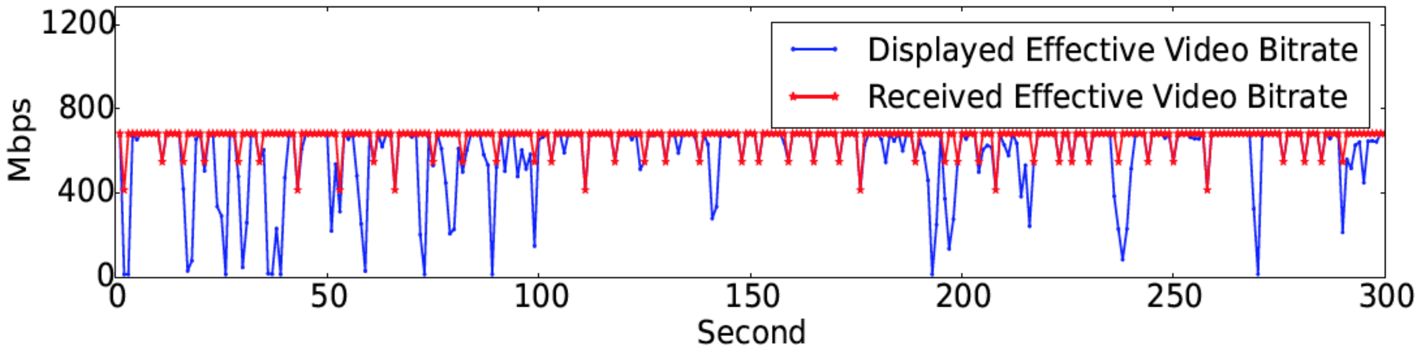
Table 3: Optimal Rate Allocation

5G Trace	FoV Trace	$B_e{}^*$(s)	$\alpha\gamma$	Rate Allocation (Mbps)			
				R_b	R_{e_1}	R_{e_2}	R_{e_3}
Stable	FoV 1	2	0.90	45.7	433.9	578.5	723.1
	FoV 2	2	0.81	89.8	400.8	534.4	668.0
Disturbed	FoV 1	3	0.85	62.7	373.5	498.0	622.5
	FoV 2	2	0.76	103.4	343.0	457.4	571.7
Unstable	FoV 1	3	0.78	83.3	310.7	414.3	517.8
	FoV 2	3	0.69	120.9	282.5	376.6	470.8

WiGig Stable Network Trace

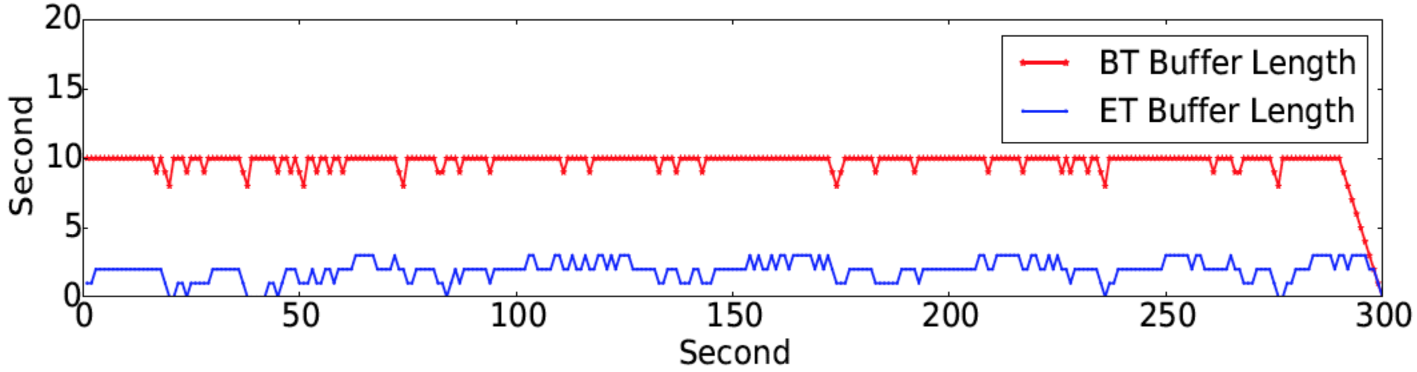


(a) Buffer Length

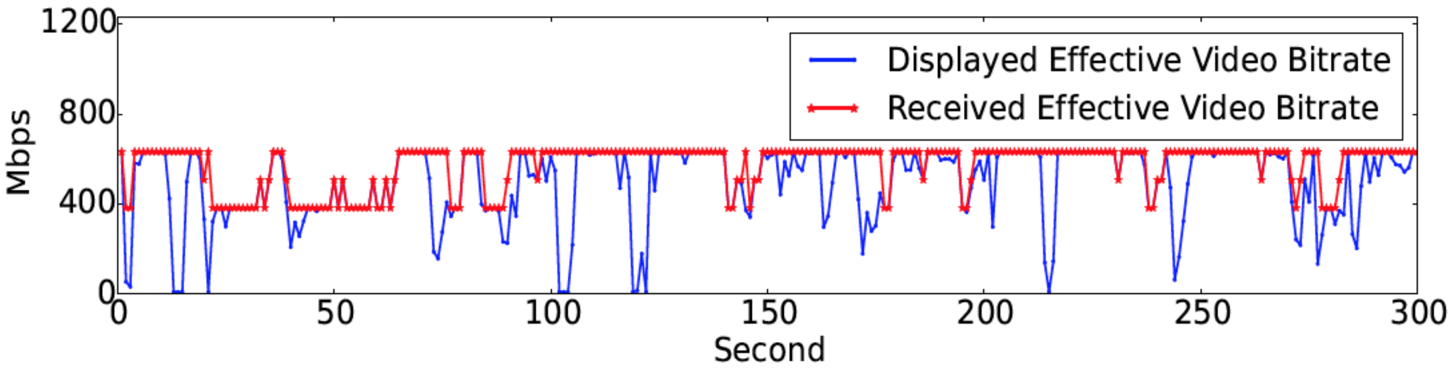


(b) Video Rate

WiGig Disturbed Network Trace



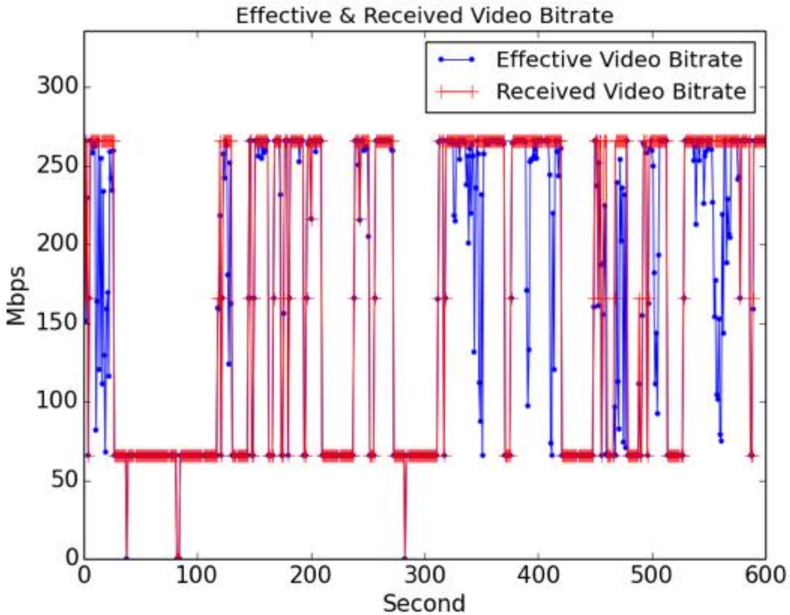
(a) Buffer Length

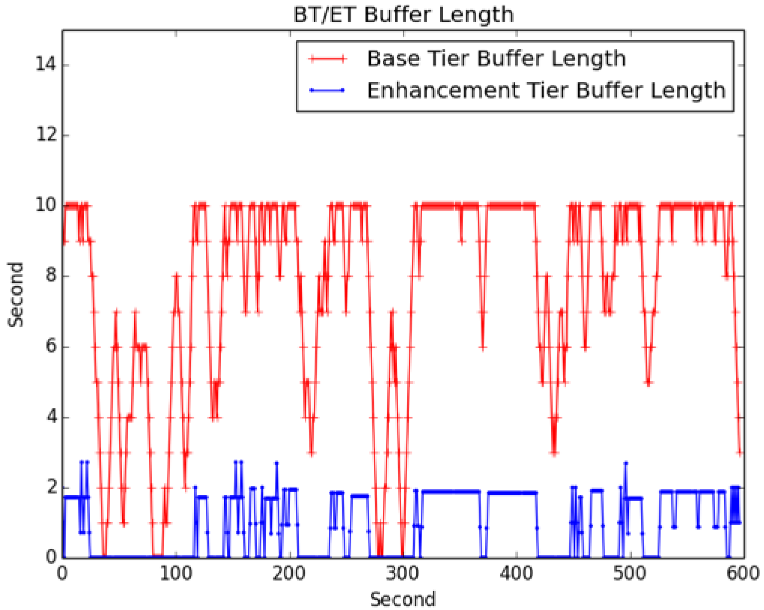


(b) Video Rate

WiGig Unstable Case 1: thick base tier

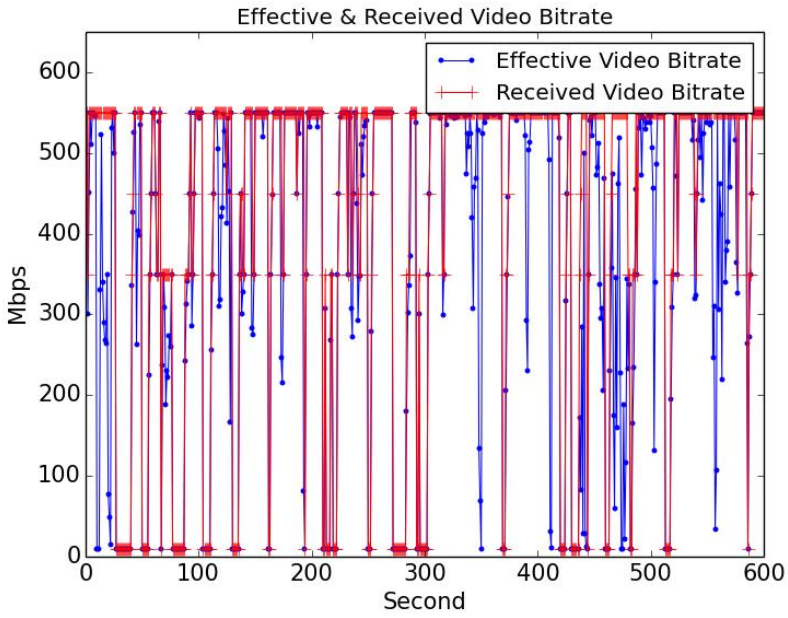
- 5G channel properties: Average bandwidth: 590Mbps, Peak bandwidth: 902Mbps
- Base Tier Bitrate: 400Mbps (66.67Mbps)
- Enhancement tier: 100Mbps, 150Mbps and 200Mmbps (166.7, 216.7 and 266.7 Mbps)

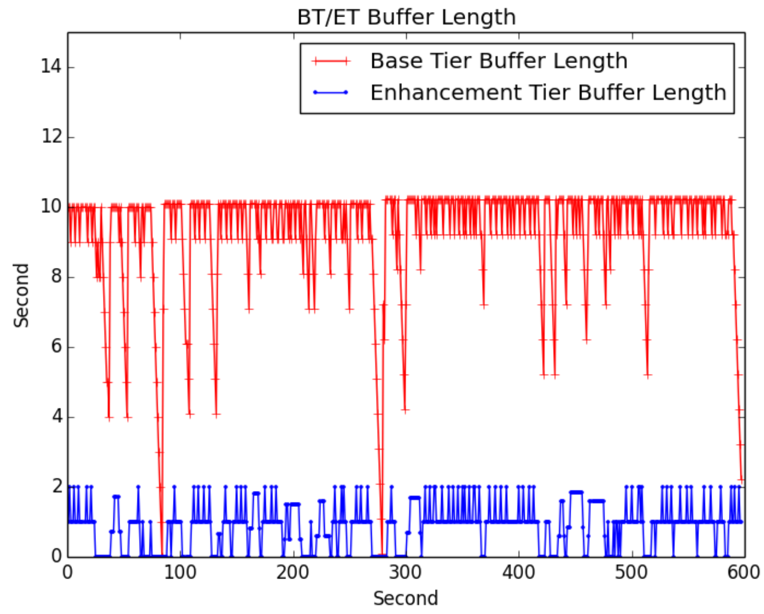




WiGig Unstable Case 2: thin base tier

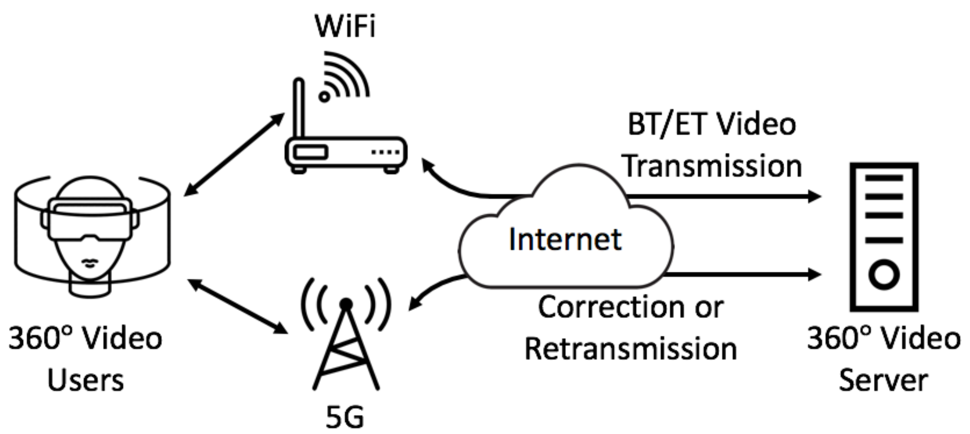
- 5G channel properties: Average bandwidth: 590Mbps, Peak bandwidth: 902Mbps
- Base Tier Bitrate: 60Mbps (10Mbps)
- Enhancement tier: 340Mbps, 440Mbps and 540Mmbps (350, 450 and 550 Mbps)





Multi-path Multi-tier Extension

- Multi-path streaming
  - Multiple available connections for users, e.g. Wi-Fi, 4G/LTE, 5G in future.
    - Bandwidth aggregation
    - Robustness against volatility/failure on individual paths
    - QoS multipath routing



Increase throughput and robustness

More Details ……

"A Two-Tier System for On-Demand Streaming of 360 Degree Video Over Dynamic Networks",
Liyang Sun, Fanyi Duanmu, Yong Liu, Yao Wang, Hang Shi, Yinghua Ye, and David Dai,
IEEE Journal on Emerging and Selected Topics in Circuits and Systems (March 2019)

“Multi-path Multi-tier 360-degree Video Streaming in 5G Networks”,
Liyang Sun, Fanyi Duanmu, Yong Liu, Yao Wang, Hang Shi, Yinghua Ye, and David Dai,
in the Proceedings of ACM Multimedia Systems 2018 Conference (MMSys 2018),

“Prioritized Buffer Control in Two-tier 360 Video Streaming”,
Fanyi Duanmu, Eymen Kurdoglu, S. Amir Hosseini, Yong Liu and Yao Wang,
in the Proceedings of ACM SIGCOMM Workshop on Virtual Reality and Augmented Reality Network, August 2017;

Conclusion & Open Questions

- Multi-tier predicative 360V streaming simultaneously handles both network b.w. variations and user viewing direction dynamics.
- On-going Projects
  - VoD (live) streaming vs. interactive/tele-presence
    - multicast, caching, MEC, P2P, NDN, SDN/NFV, …..
  - Common Design Principles:
    1. FoV-based Quality Differentiation
       - coding bits, streaming b.w., cpu, battery
    2. Predictive Decision-making
       - network QoS and user FoV prediction
       - rate selection, pre-fetching, caching, ……
    3. Joint Coding-delivery Adaptation: maximizing user QoE
    4. Robustness-first Design
       - network and user dynamics
       - prediction errors

Thanks!

Q&A