Special Topics: Machine Learning (ML) for Networking

COL867

Holi, 2025

Application Performance Monitoring

Tarun Mangla

Recap

- **What** is application performance monitoring?

① long-term capacity planning ② Traffic Engineering

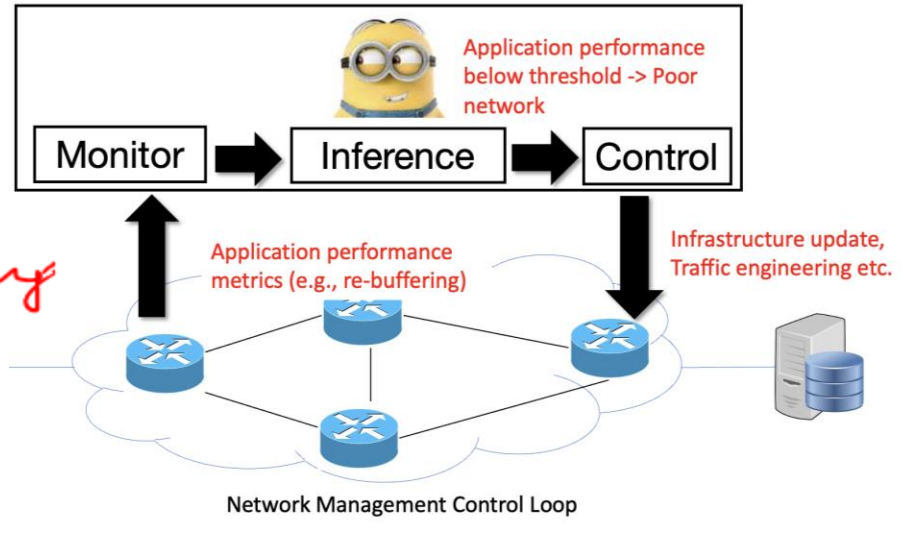
- **Why:** Enable application-aware network management

(Speedtest) Active measurement vs Passive

- **How:**
  - Passive traffic monitoring → application identification → performance inference [SNI or DNS or ML] (application-specific)
  - Modeling for video streaming in case of unencrypted traffic

↳ DASH : Dynamic Adaptive Streaming over HTTP

↳ Rebuffering : modeling buff occupancy

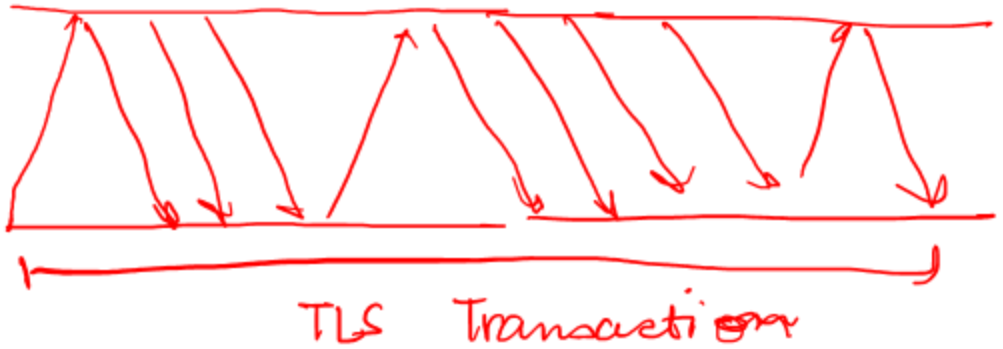


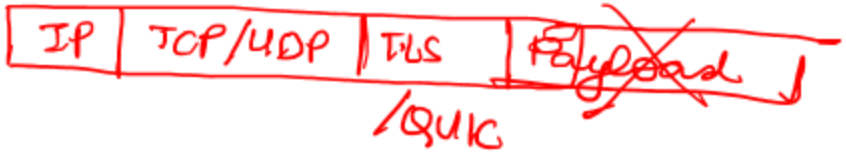
/videoapp/path/**V0987654321**/**track03**/**segment101**.ts?p=324&token=**32543563654645**

Content ID | Chunk quality | Chunk ID | Session ID

When Traffic is Encrypted

No longer have access to HTTP transactions from network measurements





When Traffic is Encrypted

No longer have access to HTTP transactions from network measurements

Solution: learn patterns from video streaming network traffic

What kind of ML solutions?
NetMicroscope, ViCrypt

Problem Statement

- Given packet-level network traffic for a video session, infer application-level performance metrics

f(Network data) → → Stalls duration
→ → Resolution (Avg) → Bitrate
→ Startup delay → Resolution switches

- How would you formulate it as a machine learning problem?

→ Supervised : labeled data to train
→ classification (high or low) (slow / poor, med, high)
→ Regression → Value of start up delay

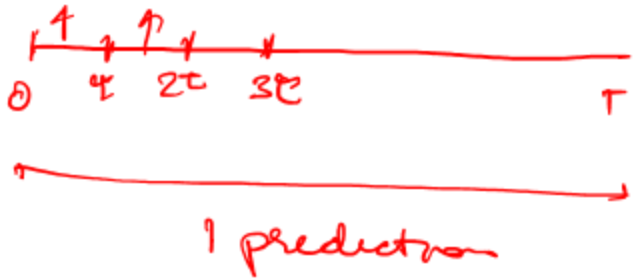
Prediction Model of NetMicroscope

- What is the granularity of prediction?

Features
① Avg / STD Packet Size → Uplink and downlink
② IAT
③ # PKTs, Tput

- Startup delay:
  - One-time prediction
- Video resolution:
  - Time series prediction

Time-series vs entire session

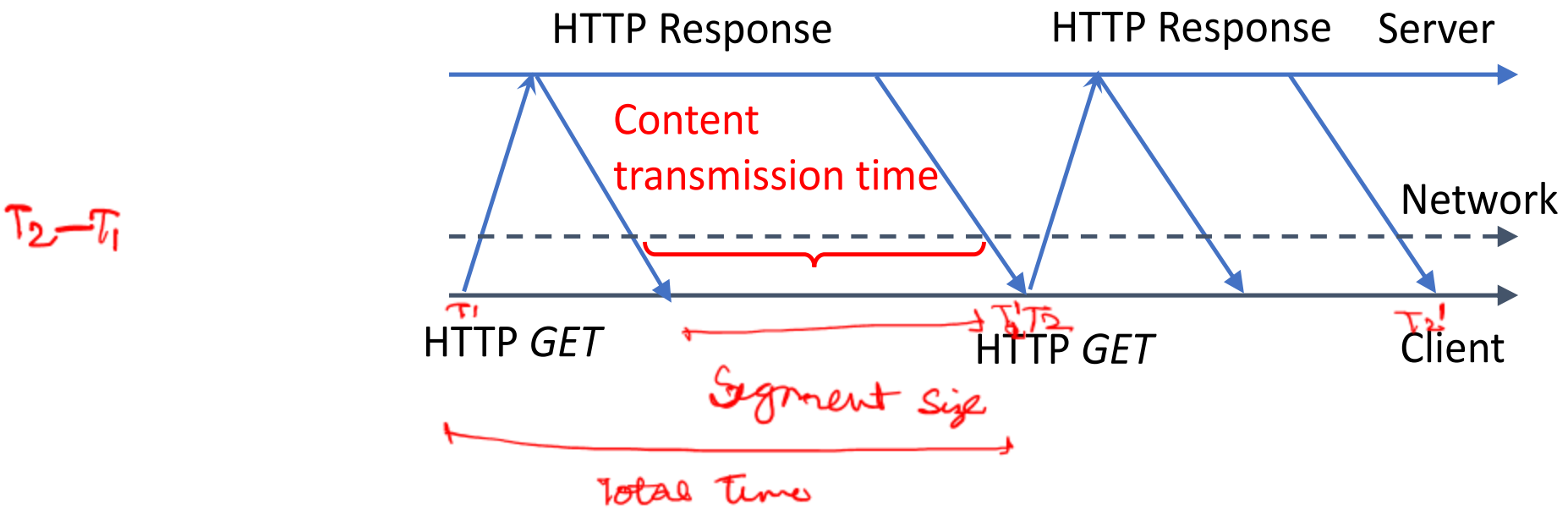


Features used by NetMicroscope

Network Layer	Transport Layer	Application Layer
throughput up/down (total, video, non-video)	# flags up/down (ack / syn / rst /push / urgent)	segment sizes (all previous, last-10, cumulative)
throughput down difference	receive window size up/down	segment requests inter arrivals
packet count up/down	idle time up/down	segment completions inter arrivals
byte count up/down	goodput up/down	# of pending request
packet inter arrivals up/down	bytes per packet up/down	# of downloaded segments
# of parallel flows	round trip time	# of requested segments
	bytes in flight up/down	
	# retransmissions up/down	
	# packets out of order up/down	

Can we reconstruct HTTP segments for encrypted traffic?

Insight: *Use directionality of data flow in an HTTP transaction*



How to train and validate the model?

- Need labeled ground truth data: Build automation framework

Video Service

- Browser-based automated streaming
- Collect ground truth QoE metrics

- Collect packet traces

- Collect data both in-lab and from home networks
- Consider 4 services: YouTube, Netflix, Amazon, Twitch
- Random Forest, 10-fold cross validation

Inference Accuracy: Startup Delay

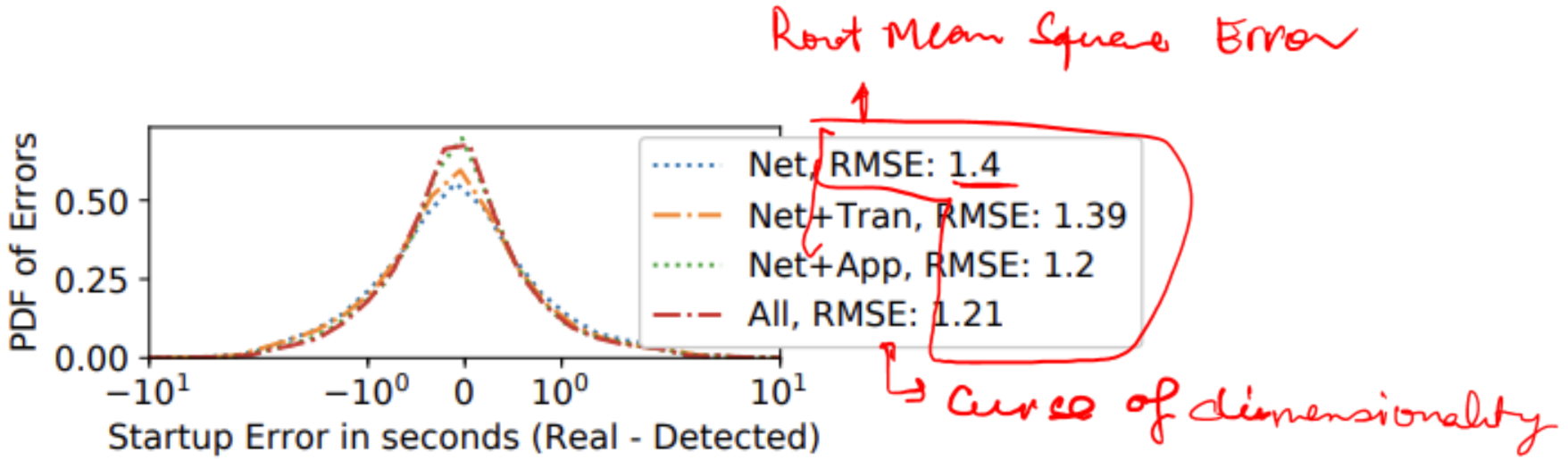
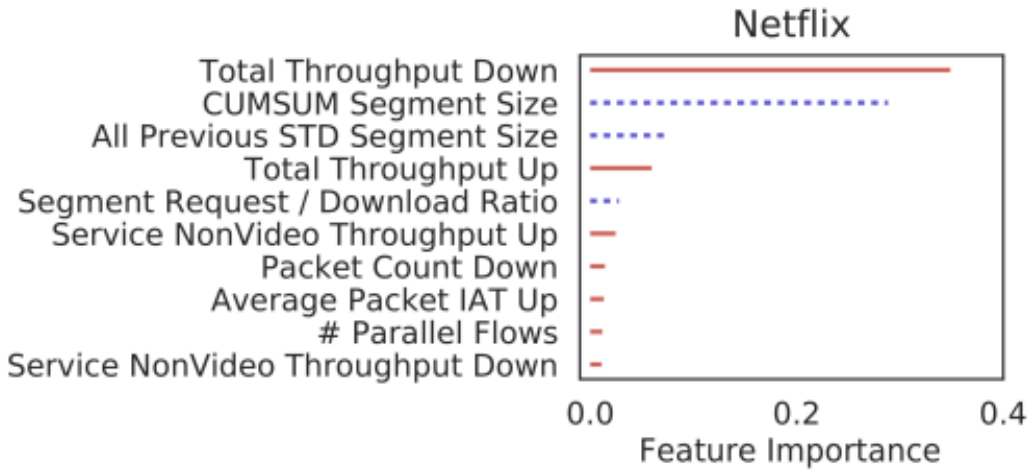
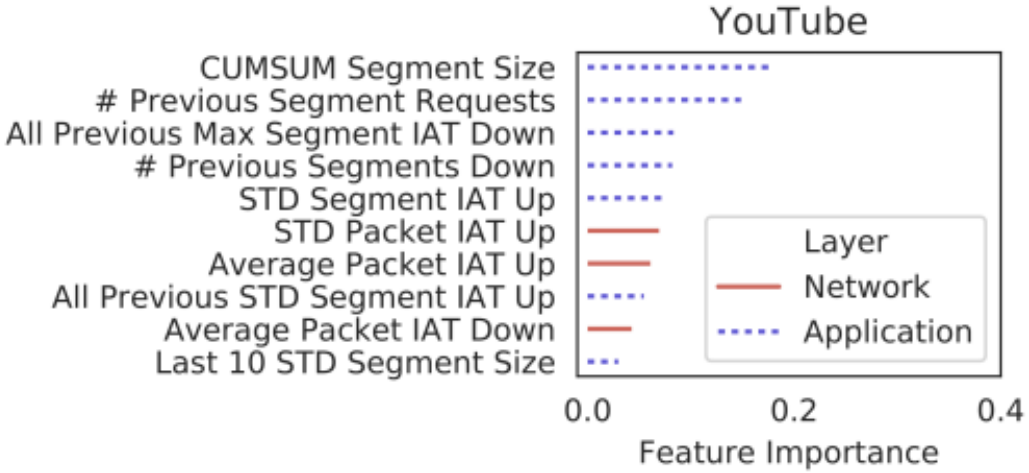


Fig. 2. *Startup delay inference error across different feature sets.*

Which features are the most important? [Explainability]



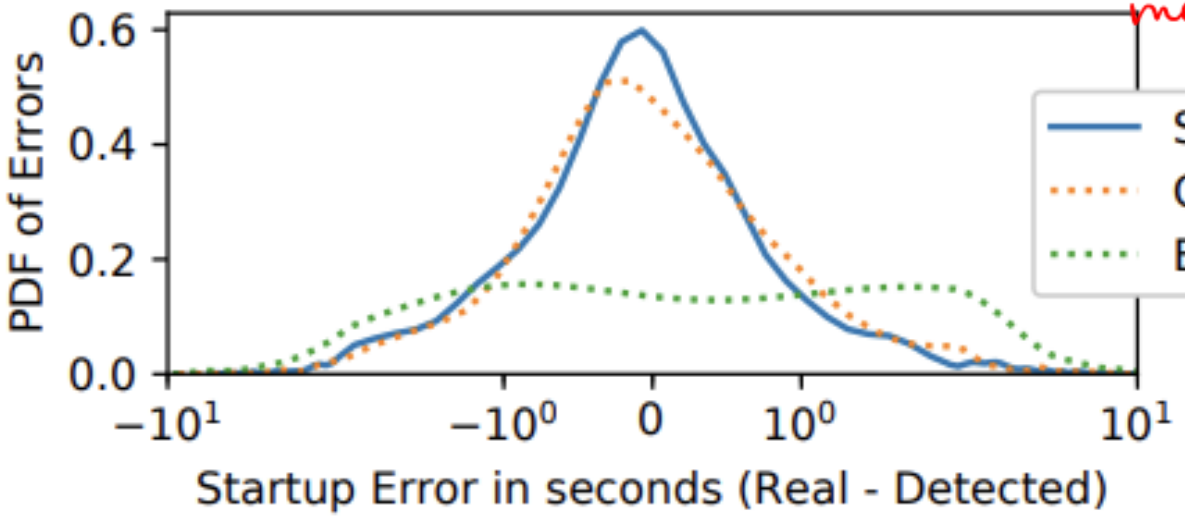


Does the model generalize? [Generalizability]

Startup latency is app-specific

4 services

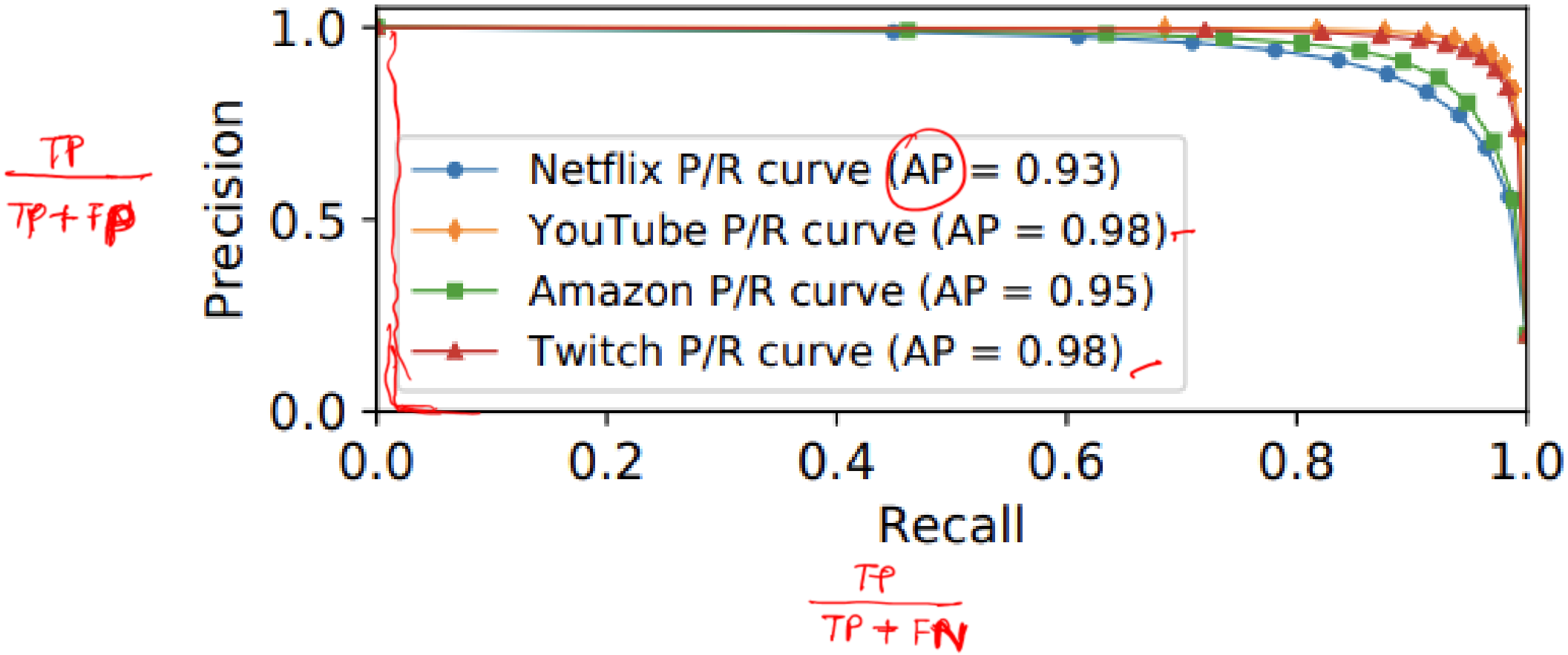
Train a single model for each service vs a common model for all



(2) A model trained can predict on a new service

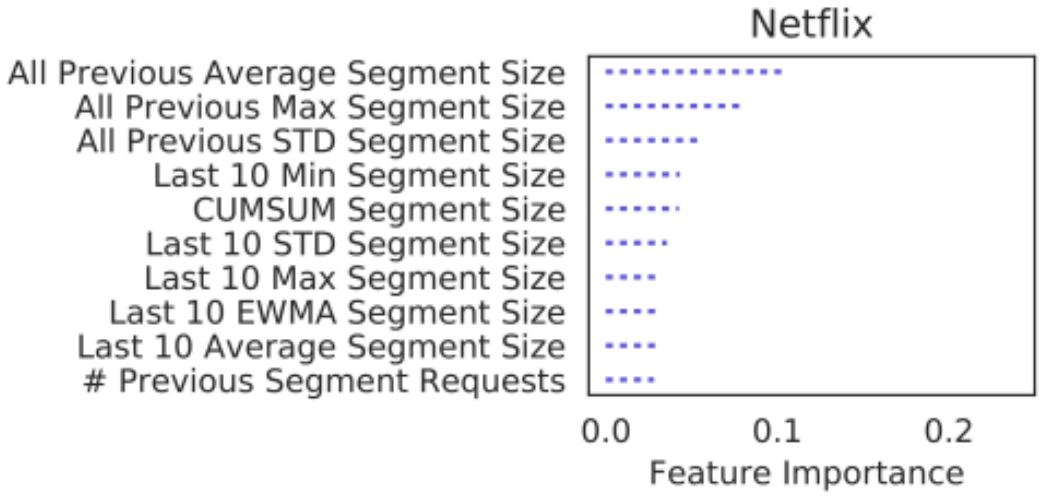
→ Transfer learning

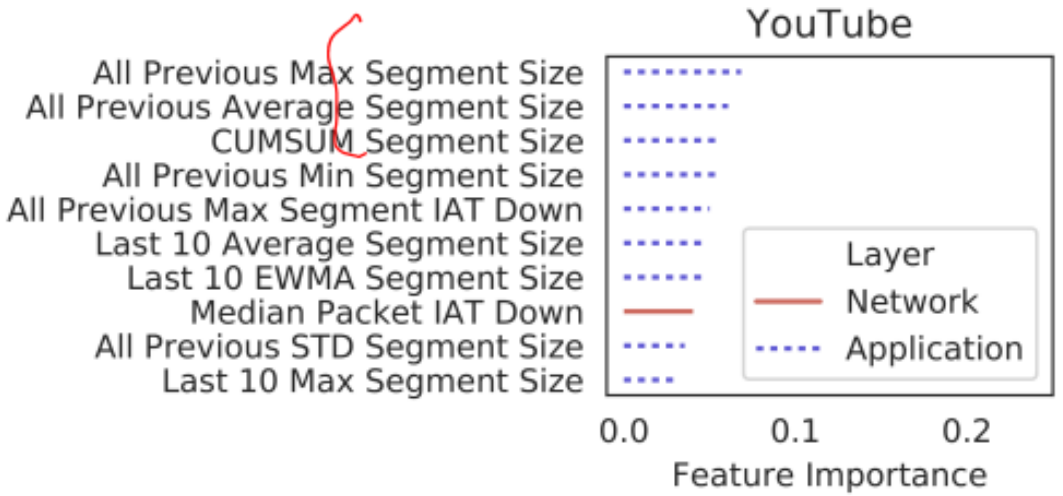
Inference Accuracy: Video Resolution



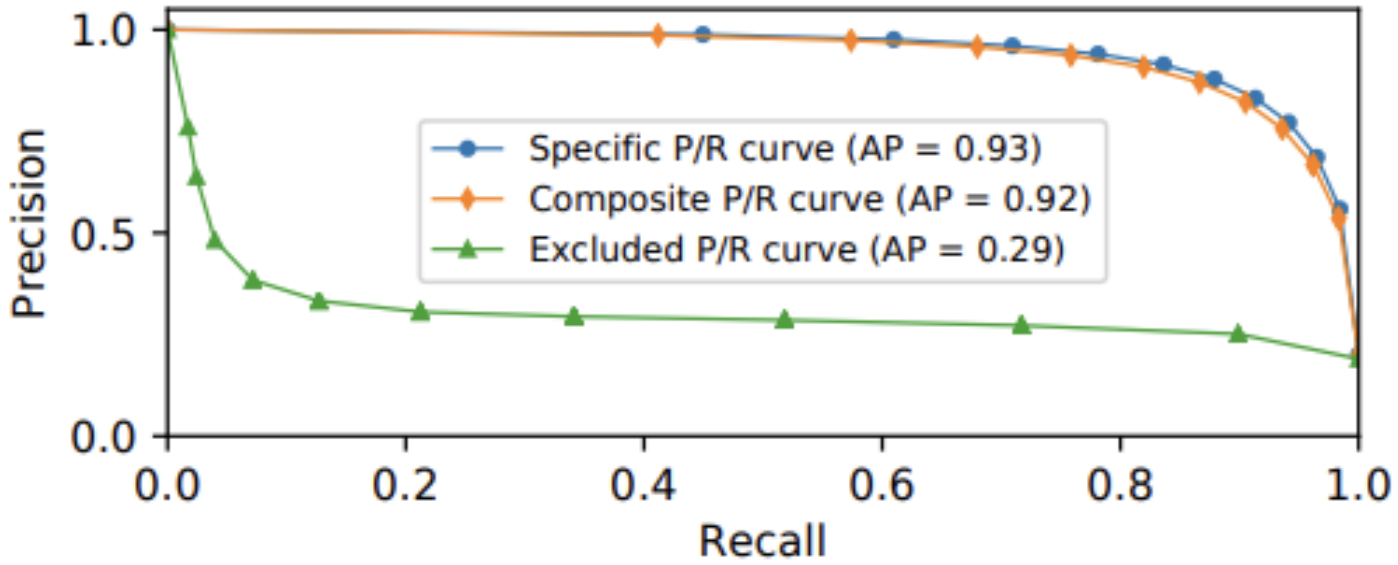
Explainability: Video Resolution

- What do you think are the most important features?

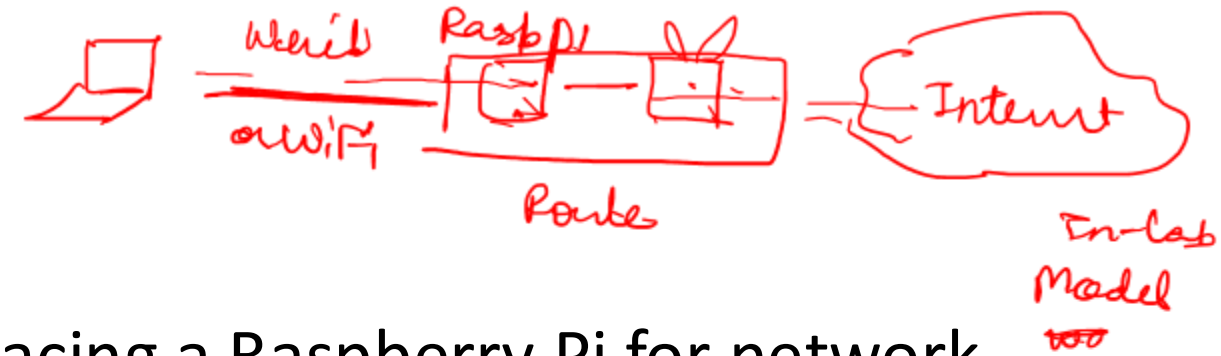




Generalizability: Video Resolution



Practical Deployment

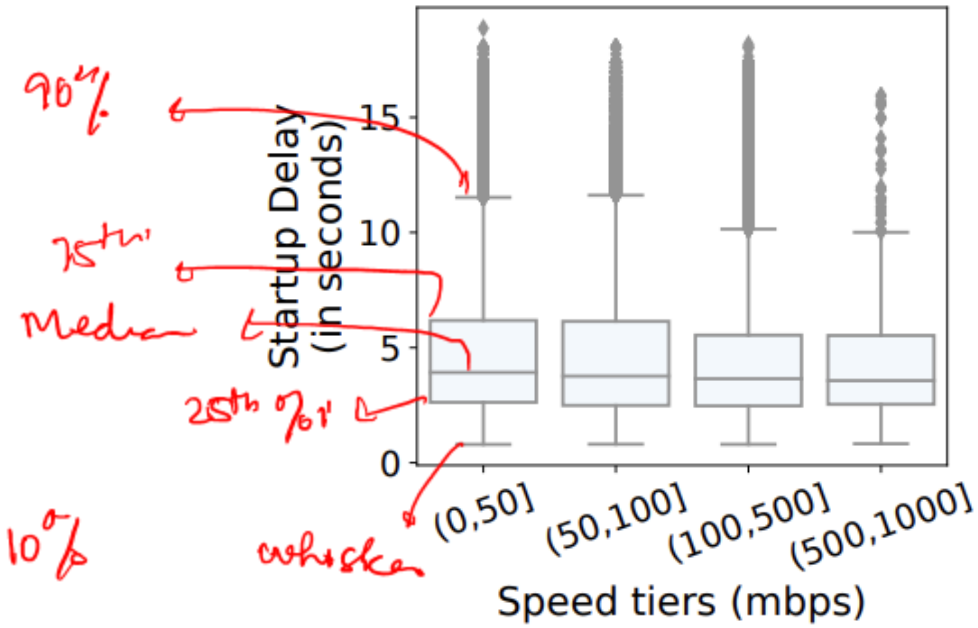


- Collected data from 66 homes by placing a Raspberry Pi for network data collection

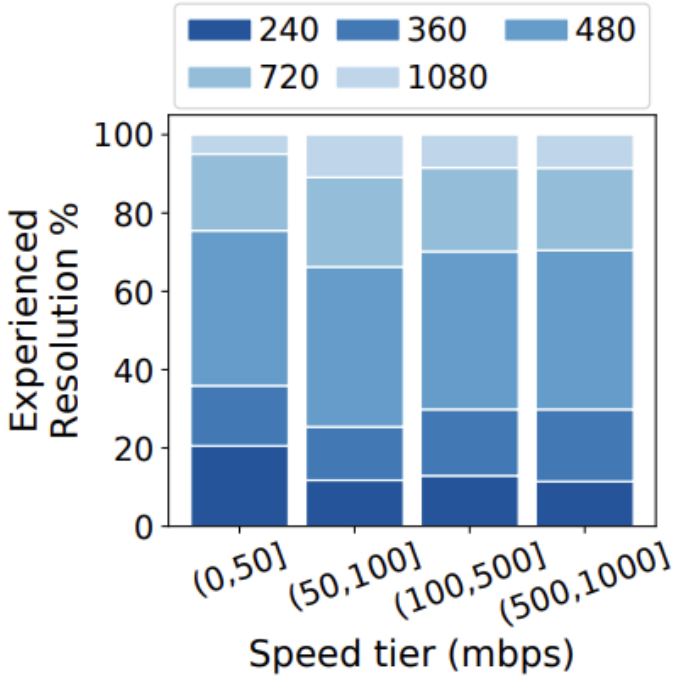
- **Challenge**: Data available at 5-second granularity. Did not have precise estimate of the session start time
  - Solution: Used domain adaptation

- Asked interesting policy questions

Does subscribing to high-speed tier improve performance?



(b) *YouTube.*



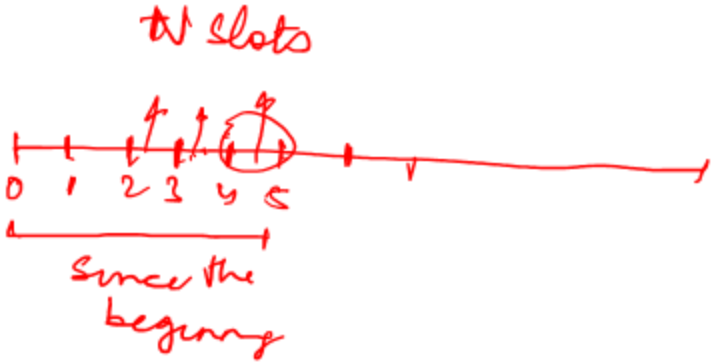
(b) *YouTube.*

ViCrypt

		ViCrypt
KQI estimation targets	Initial delay (# classes)	✓ (continuous)
	Stalling (# classes)	✓ (binary detection/continuous estimation)
	Resolution (# classes)	✓ (6 levels, 144p–1080p)
	Bitrate (# classes)	✓ (continuous)
Input features	Chunk detection required?	✗
	# features	208
	Feature selection	✓ (down to 20 features)
Network monitoring	Real time	✓
	Temporal resolution	1 second
	Feature computational efficiency	✓
Training/Evaluation data	Streaming service	YouTube
	# video sessions	15,000+
	Access network	WiFi & Cellular
	# ISPs | geo-location	4 ISPs | 4 EU countries
	Devices	laptop & smartphone (native app)
	Time span	9 months in 2018/2019

Features

- Divide time into slots
- Short-term memory vs trend



up/down/total	Volume	Throughput	Distribution	Protocol Shares
packets size	✓	✓	✓	✓
packet count	✓			✓
packet IAT		✓	✓	

Features : Streaming manner

up/down/total	Volume	Throughput	Distribution	Protocol Shares
packets size	✓	✓	✓	✓
packet count	✓			✓
packet IAT		✓	✓	

Procedure 2 Online Update of Distribution Metrics, Used for Computation of Distribution Features. The Procedure is Executed When New Values for the Corresponding Statistics are Observed

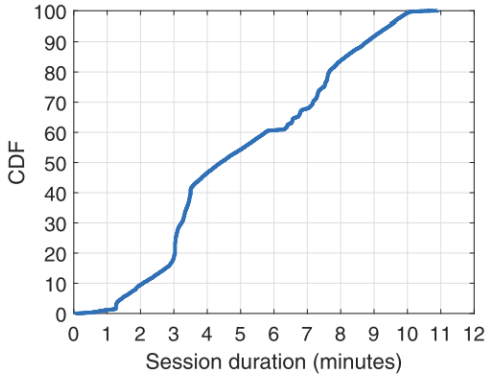
procedure UPDATEDISTRIBUTIONS(x)
- $n \leftarrow n + 1$
- $d_x \leftarrow x - mean$
- $d_n \leftarrow \frac{d_x}{n}$
- $mean \leftarrow mean + d_n$
- $sdm_4 \leftarrow sdm_4 + [d_x\, d_n\, (n-1)\, d_n^2 (n^2 - 3n + 3)] + (6\, d_n^2\, sdm_2) - (4\, d_n\, sdm_3)$
- $sdm_3 \leftarrow sdm_3 + [d_x\, d_n\, (n-1)\, d_n\, (n-2)] - (3\, d_n\, sdm_2)$
- $sdm_2 \leftarrow sdm_2 + [d_x\, d_n\, (n-1)]$
- **if** $x < min$ **then**
  - $min \leftarrow x$
- **if** $x > max$ **then**
  - $max \leftarrow x$

Procedure 3 Computation of Distribution Features

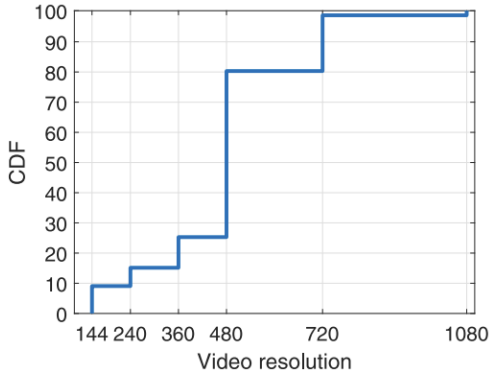
procedure COMPUTEDISTRIBUTIONFEATURES
- $var \leftarrow \frac{sdm_2}{n-1}$
- $std \leftarrow \sqrt{var}$
- $cvar \leftarrow \frac{std}{mean}$
- $skew \leftarrow \sqrt{\frac{n}{sdm_2^3}} \cdot sdm_3$
- $kurt \leftarrow n \cdot \frac{sdm_4}{sdm_2^2} - 3$

Dataset

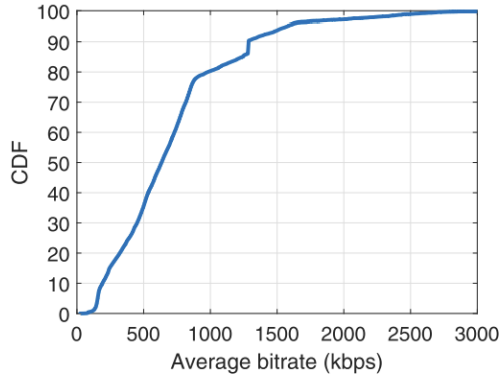
- One service: YouTube
- Different network conditions



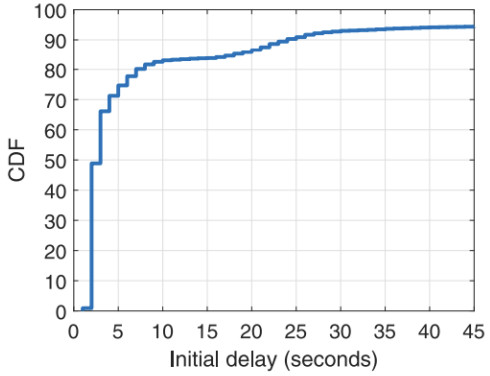
(a) Video session duration.



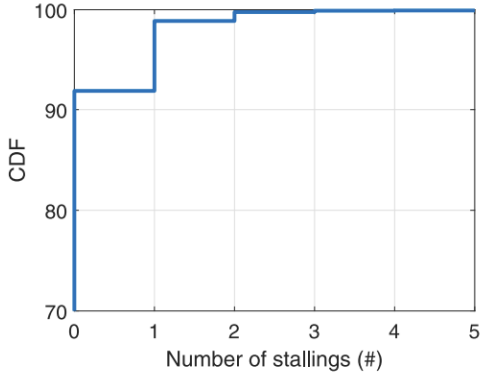
(b) Video resolution.



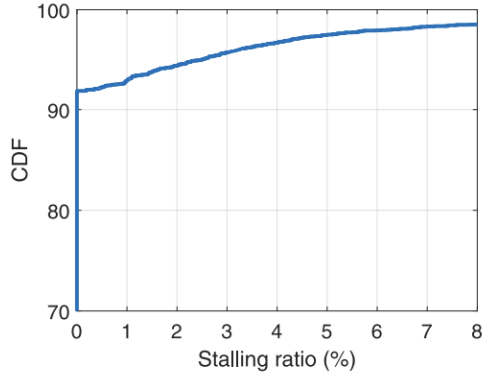
(c) Average video bitrate.



(d) Initial delay.



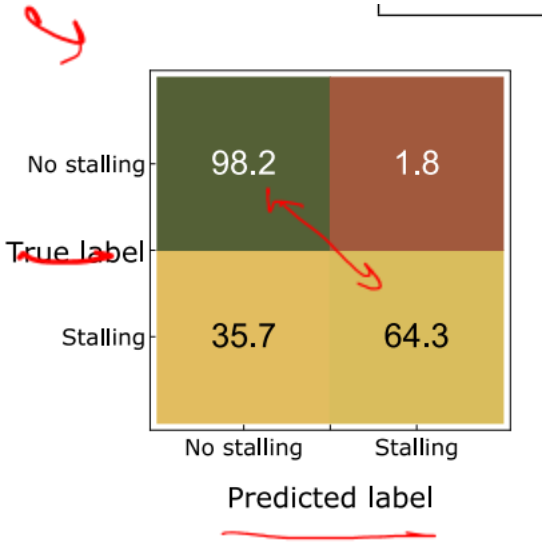
(e) Number of stallings.



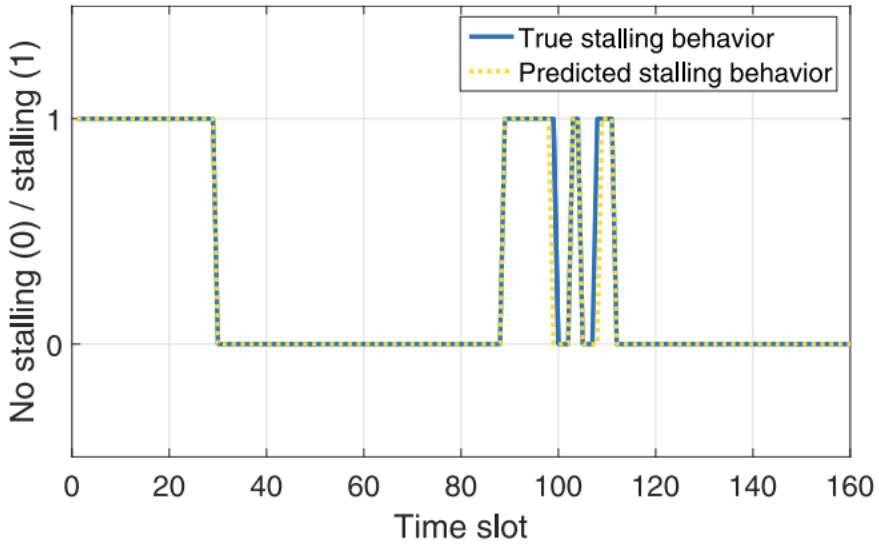
(f) Stalling ratio.

unbalanced

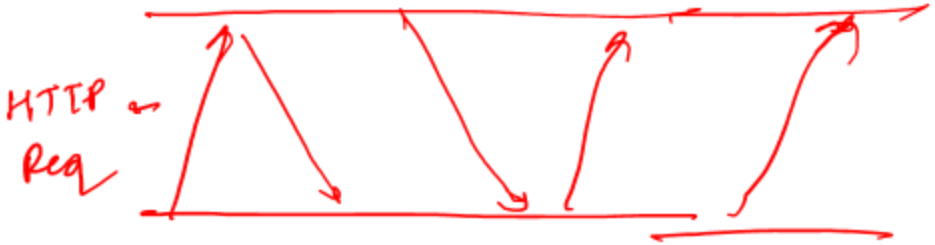
Performance: Stall Prediction



(a) DT confusion matrix.

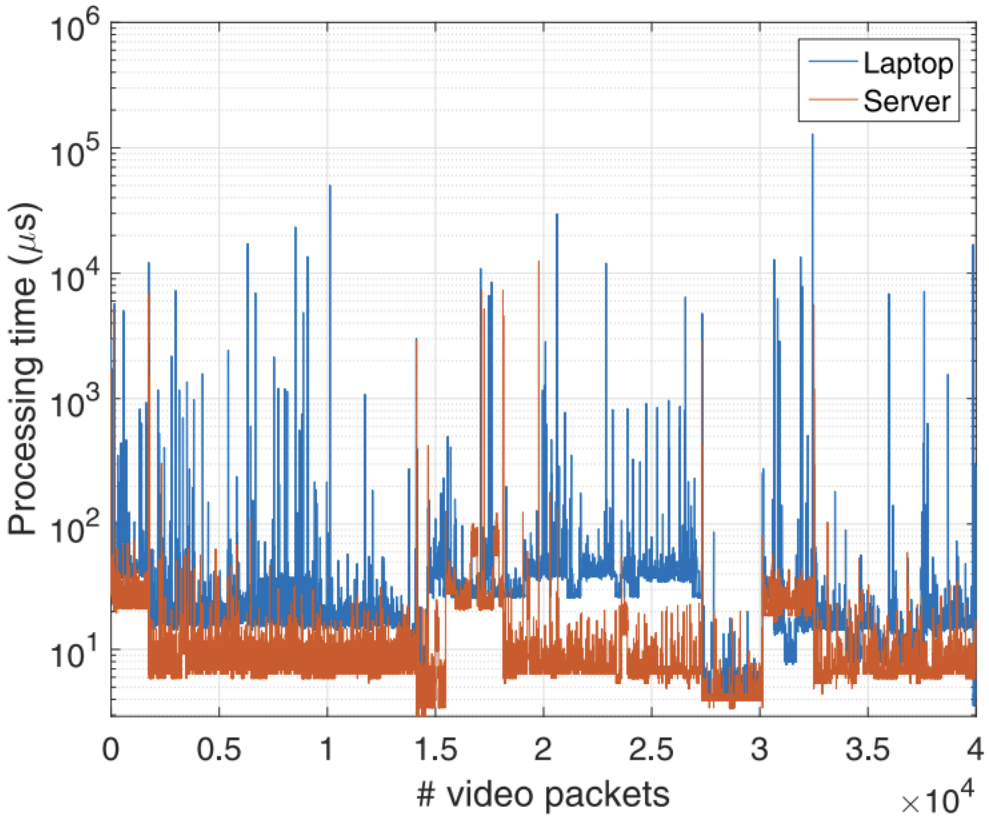


Feature Importance

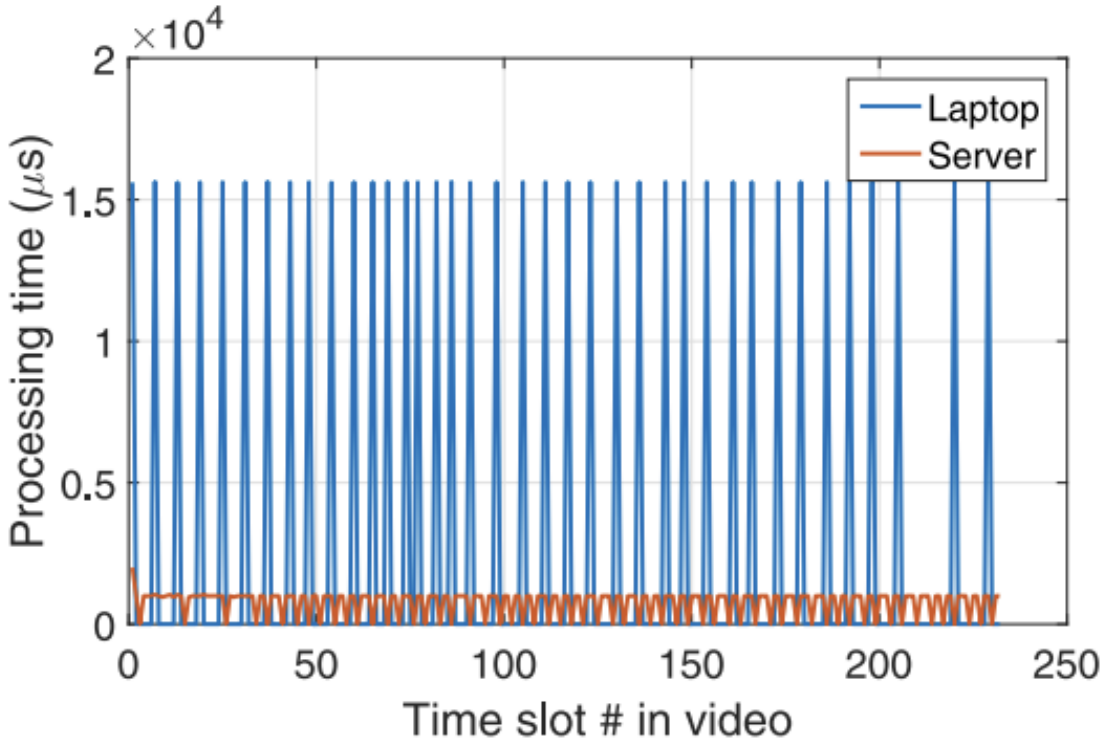


	Stalling	Video resolution	Average bitrate
#1 feature	maximum upload packet size (trend) [0.03]	throughput (session) [0.04]	throughput (session) [0.07]
#2 feature	standard deviation of upload packet size (session) [0.02]	burst throughput (session) [0.03]	burst throughput (session) [0.05]
#3 feature	upload volume (session) [0.02]	mean IAT of download packets (session) [0.02]	skewness of upload packet-size distribution (session) [0.04]
#4 feature	standard deviation of download packet size (session) [0.02]	burst throughput of download traffic (session) [0.02]	mean IAT of download packets (session) [0.04]
#5 feature	skewness of upload packet-size distribution (session) [0.01]	coefficient of variation of IAT of download packets (session) [0.02]	download burst throughput (session) [0.04]

System cost



Update Features



Prediction by the model