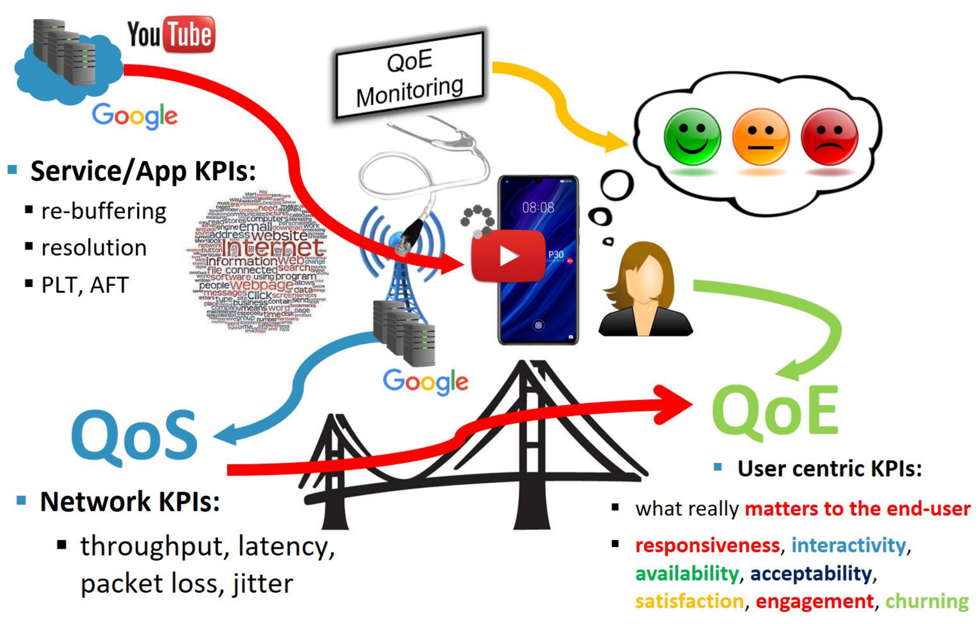
ViCrypt: Real-time, Fine-grained Prediction of Video Quality from Encrypted Streaming Traffic

Sarah Wassermann, Michael Seufert, Pedro Casas, Li Gang, Kuang Li

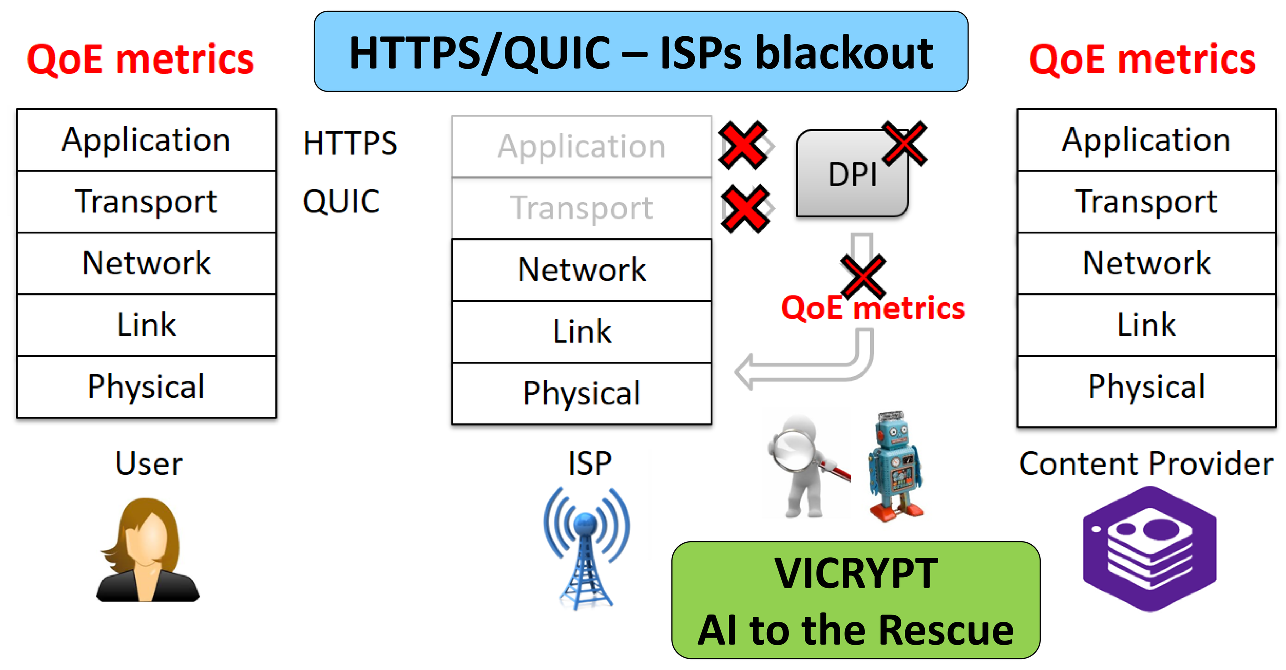
VICRYPT: Real-time, Fine-grained Prediction of Video Quality from Encrypted Streaming Traffic

S. Wassermann (1), M. Seufert (2), P. Casas (1), L. Gang (3), K. Li (3)
(1) AIT Austrian Institute of Technology, Vienna, Austria
(2) University of Würzburg, Würzburg, Germany
(3) Huawei Technologies, Shenzhen, China

QoE–BASED NETMON (QoE–MON)



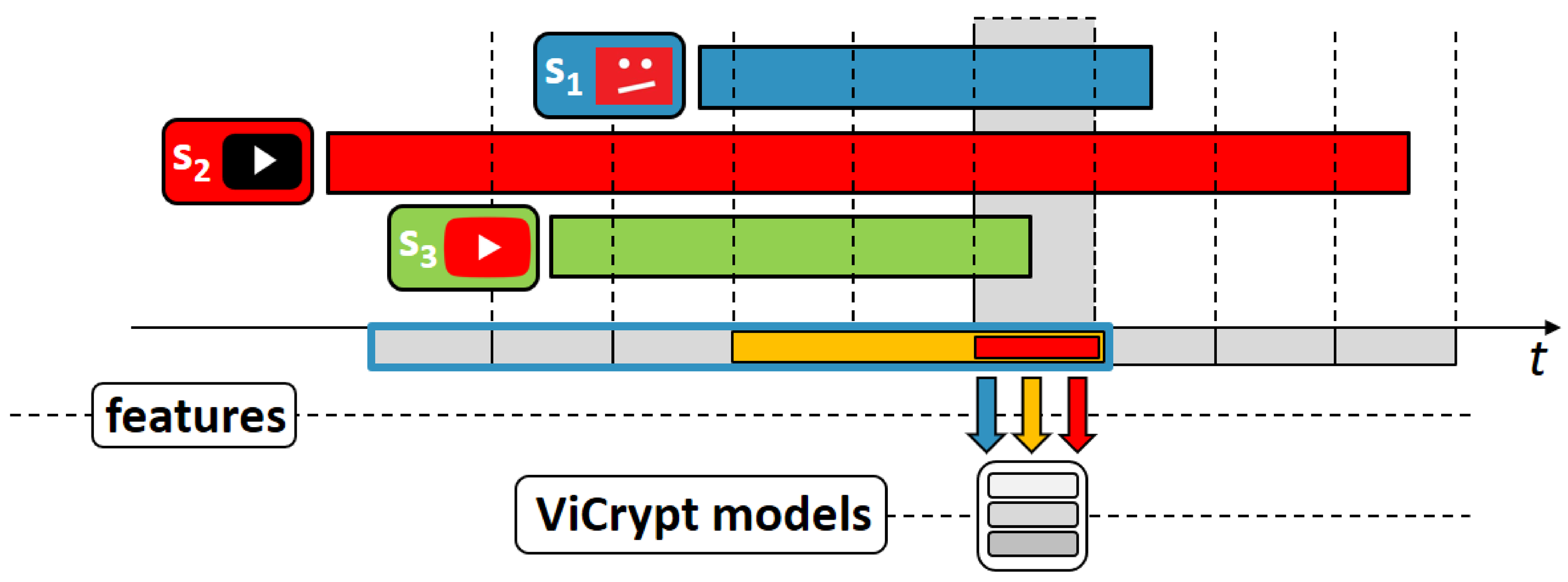
MOTIVATION & CHALLENGE



VICRYPT: ML–BASED QoE–MON

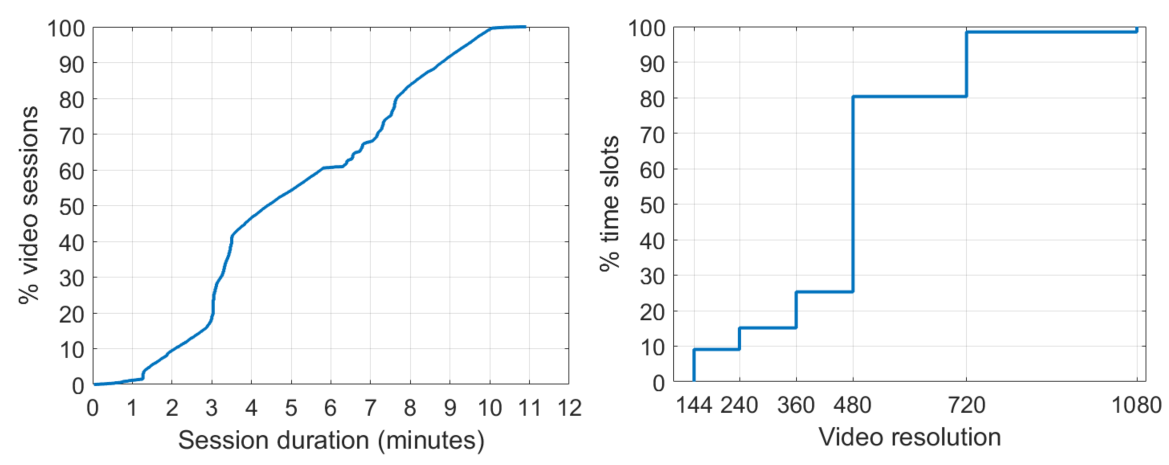
- **Real-time** (1-sec resolution) **prediction of KQIs** for video streaming
- **Video chunk detection NOT NEEDED** features are packet size/time based
- ML models for prediction of **instant, per-sec**:
  - re-buffering events
  - video resolution
  - video bitrate

STREAM-BASED PREDICTION OF VIDEO KQIs

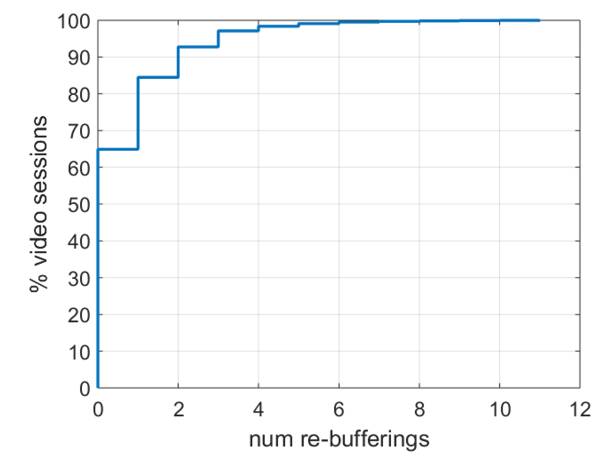


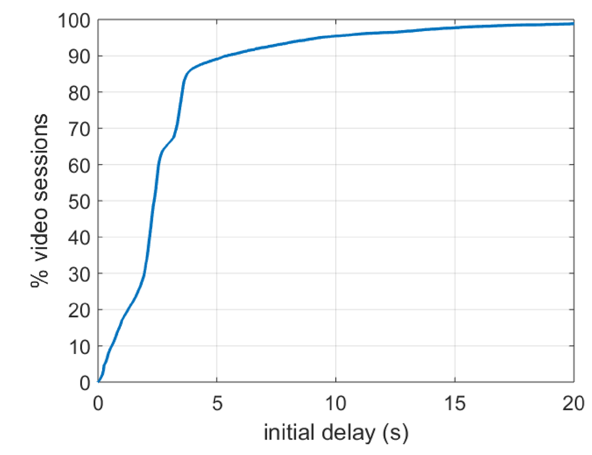
- **Video stream-based analysis,** using multiple sliding windows, capturing different temporal phenomena (**current time**, **short-term trend**, **session-aggregated**)
- **Analysis is done in real time:** for every video session and **for every new time slot of 1 second**, we consider the following set of **207 features**:
  - Features extracted from **current time slot (C) – 69 features**
  - Short-memory (trend) based features, extracted from **last T (3) slots (CT) – 69 features**
  - Cumulative based features, extracted from **all past traffic for this video session (CS) – 69 features**
- **Feature computation** is done **continually**, in **constant-memory** boundaries, ***using sketches***

YouTube DATASET FOR TRAINING & VALIDATION



- Data generation through **semi-controlled testbeds**
- **15.000+ YouTube video sessions** streamed and recorded in late 2018/early 2019
- **Different ISPs**, different **geographic locations** (Austria, Italy, Germany, China)
- Home/corporate **WiFi** networks, **LTE** networks
- **QUIC and TCP sessions**
- **Bandwidth limitations**: 20Mbps, 5Mbps, 3Mbps, 1Mbps, 300kbps + fluctuations
- JavaScript-based monitoring script to **measure ground truth at the player**



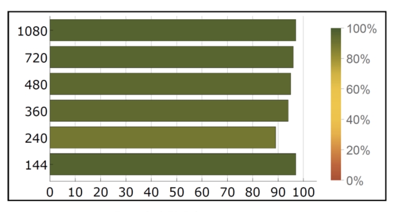


ONLINE PREDICTION OF VIDEO RESOLUTION

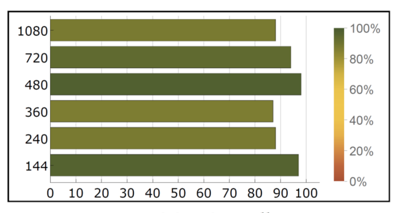
- Training multiple ML models over more then 4.6M individual, 1 sec. slots (***5-fold cross validation***) – here using all 207 inputs
- **Classification task:** per second video resolution, ***6-classes:*** 144p, 240p, 360p, 480p, 720p, 1080p

	Training time (min)	Accuracy (%)
DT	43	92
RF10	2	92
ADA	125	68
ERT10	1	90
BAGGING	37	95
BAYES	1	42
KNN	9	73
NN	507	58
SVM	194	54

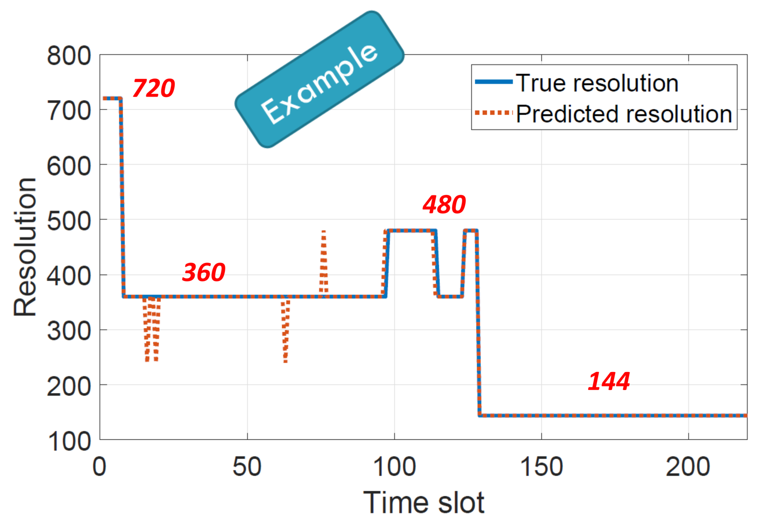
Benchmarking of different ML models



BAGGING precision



BAGGING recall

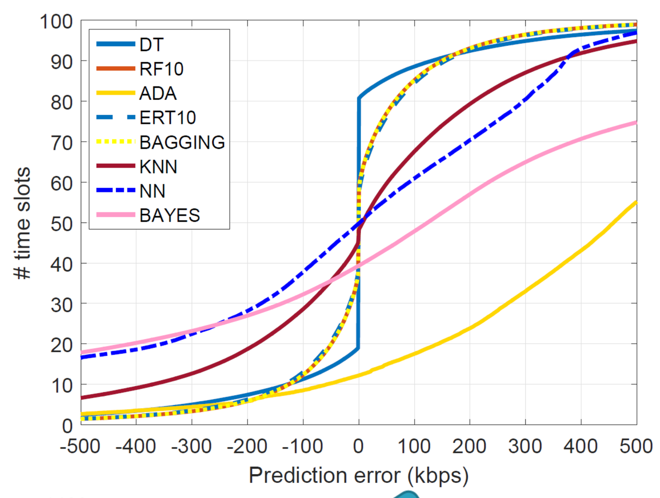


Video-resolution prediction.

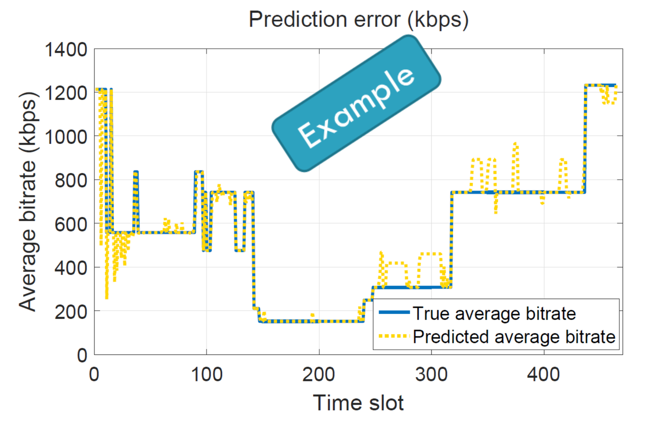
ONLINE PREDICTION OF VIDEO BITRATE

- **Regression task:** estimation of per second average video bitrate

	5-CV time (minutes)	MAE (kbps)	RMSE (kbps)	MRE (%)	PLCC
DT	31	94	246	18	0.88
RF10	36	89	179	18	0.93
ADA	126	492	573	130	0.59
ERT10	7	93	182	19	0.93
BAGGING	22	89	179	17	0.93
BAYES	3	2,540	6,530	545	-0.14
KNN	6	229	353	42	0.70
NN	305	333	489	70	0.20
SVM	143	10^{23}	$2 \cdot 10^{23}$	$2 \cdot 10^{23}$	0.12



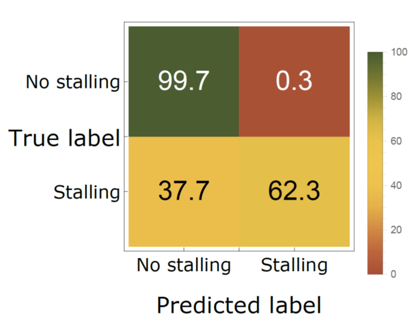
- ERT10 & BAGGING realize **MAE below 100kbps**, and **RMSE below 190kbps** (penalizes larger errors)
- **80% of the slots** are estimated with **errors below 100kbps**
- Predictions are ***highly correlated*** with the target (***PLCC = 0.93***)

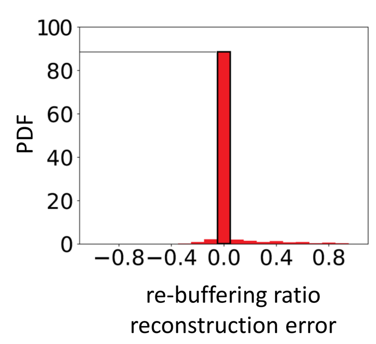


ONLINE PREDICTION OF STALLING

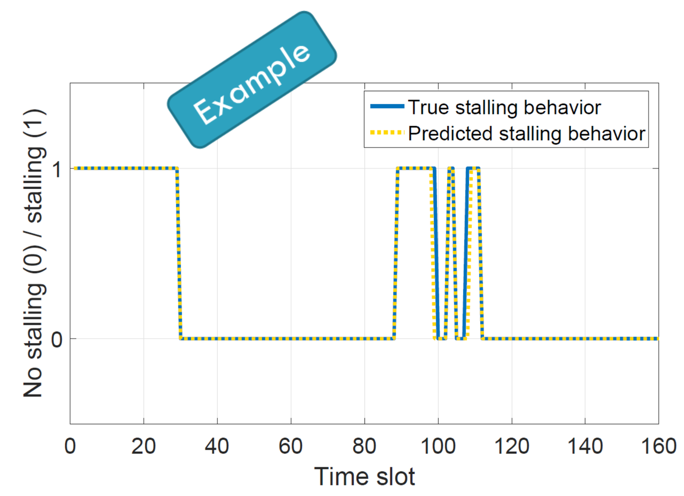
- **Binary classification task:** playback stalled/not-stalled at every new slot

	Accuracy (%)	Recall (%)	Precision (%)	5-CV time (minutes)
DT	96	64	68	57
RF10	97	55	88	3
ADA	95	29	61	154
ERT10	97	54	88	1
BAGGING	97	65	87	63
BAYES	50	86	9	1
KNN	96	48	71	10
NN	94	0	0	600
SVM	84	62	21	36
ISO10	86	13	8	4
LOF	86	11	6	46



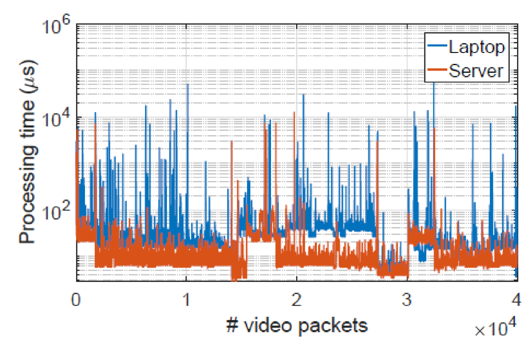


- **per-slot re-buffering** estimation errors are high, **stalling slots under-estimated…**
- …but **estimation of re-buffering ratio** is ***perfect for almost 90% of the videos***

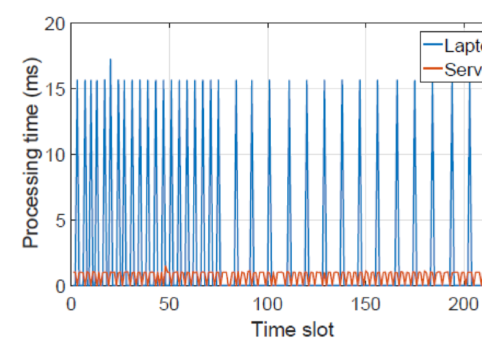


COMPUTATIONAL TIME & IMPACT OF FEATURE SELECTION

- Evaluation of ***full feature set update time*** (done at every new incoming packet) and ***prediction time*** (done for every 1s slot), using an upper bound **with all 207 features**
- ***laptop*** (i5 CPU, 8GB RAM) vs. ***server*** (Xeon Silver, 48 cores, 128GB RAM)
- server: avg. **duration of full feature set update is 13 µs, prediction time below 1.4ms**
- Laptop: **avg. feature update takes 37 µs, prediction time below 16ms**



Features update time at each new packet.



Prediction time.

Features	Accuracy (%)
All	92
F_C	70
F_T	73
F_S	96
F_{DOWN}	90
F_{UP}	90
F_{TOP20}	95

Features	MAE [kbps]	RMSE [kbps]	MRE [%]	PLCC
All	93	182	19	0.93
F_C	275	407	55	0.58
F_T	253	377	51	0.64
F_S	68	157	14	0.95
F_{DOWN}	105	195	21	0.92
F_{UP}	106	198	21	0.92
F_{TOP20}	81	175	16	0.93

Features	Accuracy (%)	Recall (%)	Precision (%)
All	97	54	88
F_C	96	10	30
F_T	97	17	51
F_S	99	72	91
F_{DOWN}	98	41	87
F_{UP}	98	47	74
F_{TOP20}	97	56	86

- Automatic Feature Selection – **CS features (F_S)** alone provide the best results (69 features), improving overall performance. **Top 20 features (F_{TOP20})** provide ***similar improvement with much less features***