Orange
Innovation

Characterization and troubleshooting of low-latency applications on cellular networks

First step: anomaly detection of cloud-gaming sessions

DataCamp 2022 – Research session 20/09/22

Joël Ky - INNOV/NET/WNI/IPN/ITEQ

Table of contents

1. PhD objectives

- **PhD Title :** Characterization, classification and troubleshooting of cloud gaming applications.
  - University of Lorraine
  - **Start date :** 17/11/2021

- **Motivation:**
- Emergence of many low-latency applications (cloud-gaming, tactile internet, metaverse, …) with the development of Internet.
- These applications come with stringent network requirements that hinder the quality of user sessions, especially on cellular networks.
  - Crucial for ISP to detect and mitigate the possible anomalies.

- **Objectives:**
  - Identify, collect and analyze relevant metrics in network traffic, base stations, UEs… to characterize those applications in time-varying network conditions.
  - Use machine/deep learning approaches to identify anomalies in the collected metrics and go back to the root causes
- **Keywords :** low-latency, cloud-gaming, AI, troubleshooting

2. Generation of realistic cellular network conditions

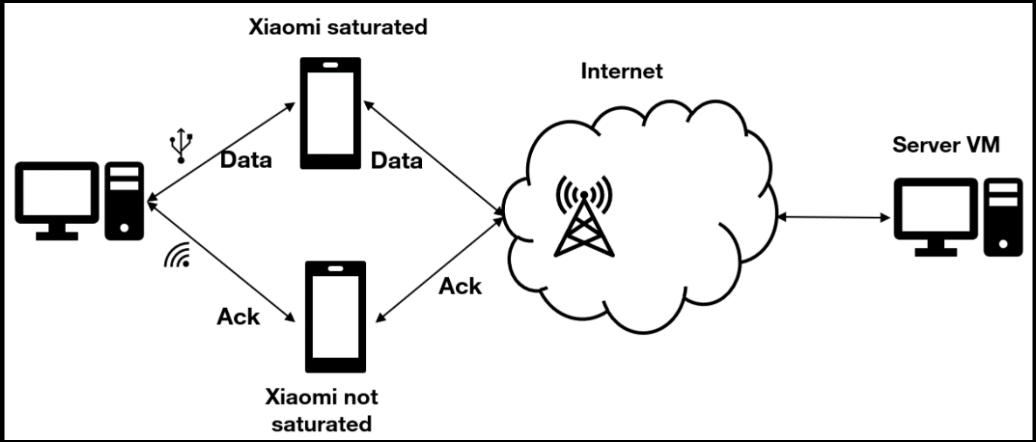
2-1. Motivation :

- How can we conduct controlled experiments on realistic network conditions ?
  - The framework Mahimahi developed by MIT researchers.

- Transmission opportunities (txops) files, used by Mahimahi to emulate time-varying capacity network, are old and not representative of current cellular network capacities (Verizon LTE - TMobile 2016).
  - Current downlink throughput according to [ARCEP] are about 71Mbps while those on the txops are about 5-10Mbps.
  - We want more recent txops file to perform better evaluations.

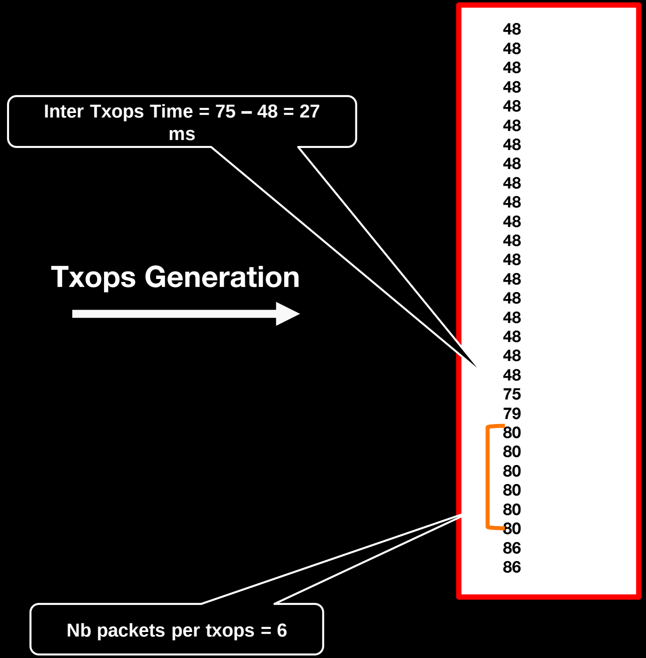
- How to generate txops files that can emulate current and realistic cellular network conditions ?
  - Use Saturatr tool to make measurements from 4G/5G base station.

2-2. Protocol for experiments on time-varying capacity networks



$$if\ (rtt < RTT_{lower}\ \&\&\ congestion\ window < window_{upper})\ then\ congestion\ window++$$

$$if\ (rtt > RTT_{upper}\ \&\&\ congestion\ window > window_{lower})\ then\ congestion\ window\ -= 20$$

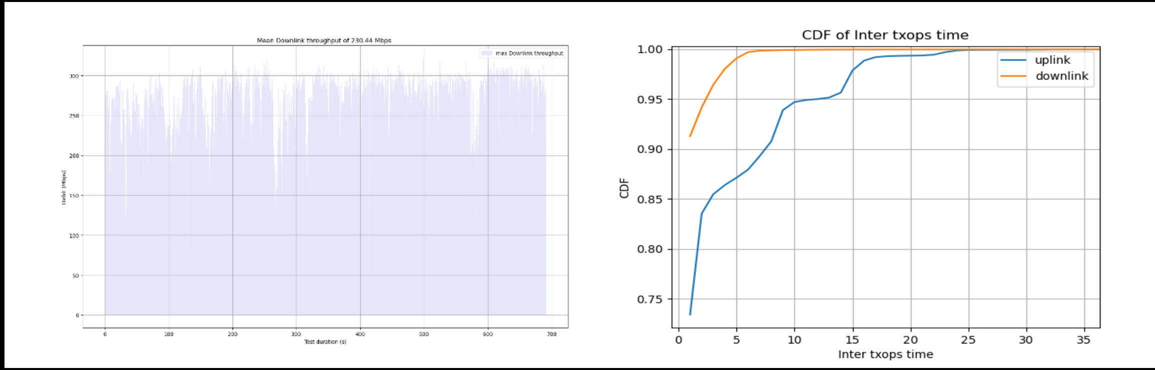


Saturator tool [Saturatr] to generate transmission opportunities (txops) by saturating link radio.

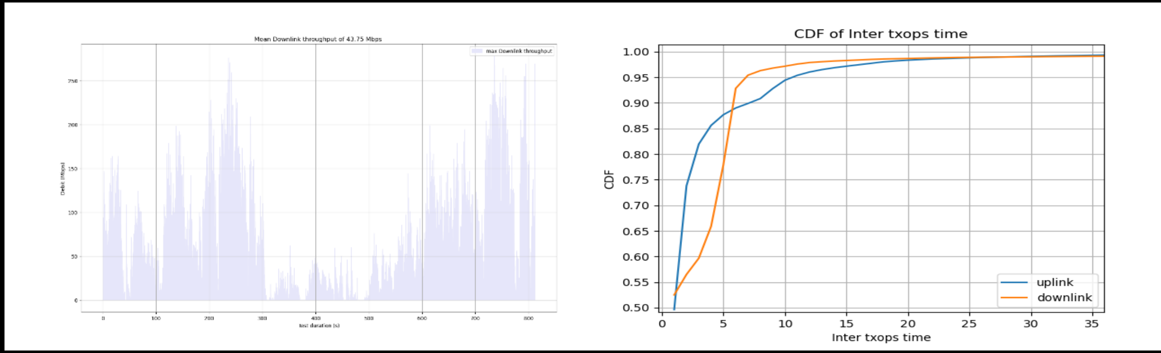
2-3. Characteristics of the measured cellular networks condition

Conditions	Throughput (Mbps)	Location
File 1	220	Orange
File 2	160	Orange
File 3	120	Brélévenez
File 4	80	Brélévenez
File 5	40	Plemeur-Bodou
File 6 (Highway)	45	Guingamp - Lannion

Measurements conditions



File 1

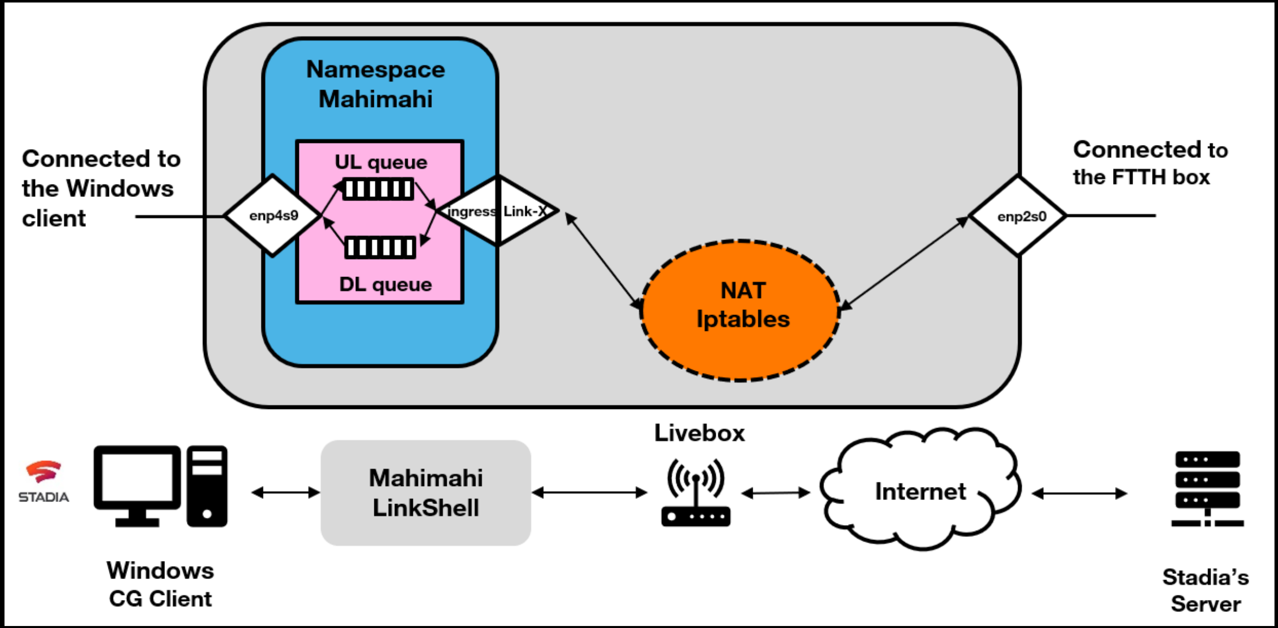


File 6

3. Data collection of cloud gaming sessions on emulated 4G network conditions

- Collect QoS/QoE metrics to characterize cloud gaming applications over cellular network conditions:
    - Use of Mahimahi tool [Mahimahi] for network emulation
    - Use of WebRTC API that provides a lot of client-side metrics through Chrome desktop client.

- Game sessions with Google Stadia on different networks emulated conditions.



4. Evaluation of Unsupervised ML models for Anomaly Detection in Cloud-Gaming sessions

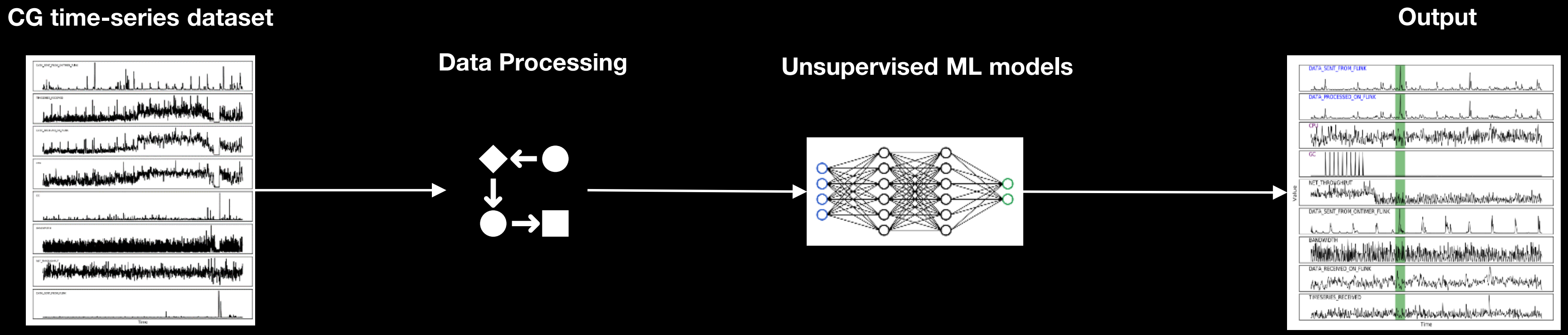
- Detecting abnormal network behaviors requires experts knowledge.
  - Impractical due to the increasing network complexity.

- Use of Unsupervised ML approaches to bypass the need of labeled data.

- Based on the aforementioned metrics collected, we evaluate the performance of 5 different unsupervised ML for the detection of cloud gaming performance degradation.

- Unsupervised ML model assume that the training data is free from anomalies but this is not the case in real-life scenarios.
  - We then assess the robustness of unsupervised ML models to data contamination and the impact of data splitting strategies on the performance.

4-1. Methodology



- We consider in this work, **point anomalies** only.
- To objectively compare the models with well-known ML performance metrics, ground truths are required.
- An observation is as a ground-truth anomaly if **FPS < 60 or Resolution < 1080p or a freeze occurs.**

4-2. Data processing

- **Mixed-dataset splitting**

X_{220} X_{160} X_{120}

X_{80} X_{40} $X_{highway}$

Train — 50% of normal

Contamination set — 40% of abnormal observations

Test — 50% of normal + 60% of abnormal

TABLE I
COLLECTED DATASETS

Data split Strategy	# Train instances	Train Anomalies ratio (%)	# Test instances	Test Anomalies ratio (%)
Mixed-dataset	138021	48.8	171704	58.85
High-bitrate dataset	168408	34.56	141318	77.98

- **High-bitrate splitting:**

X_{220} X_{160} X_{120} → Train, Contamination set

Normal obs. for train
Abnormal for contamination

X_{80} X_{40} $X_{highway}$ → Test

4-3. Unsupervised ML models

- **OC-SVM:** Support Vector Machines based approach to separate the normal data from anomaly data with an hyper-sphere.

- **Isolation Forest:** Performs splits based on features to isolate anomalies from normal instances.

- **PCA:** Reconstruction of the data with principal components.

- **Auto-Encoder (AE):** Constitued of an encoder, that learns from inputs a low-dimensional representation of data, and a decoder that reconstruct original data from latent variable.

- **LSTM-VAE:** Combination of LSTM and a VAE (AE with bayesian inference).

Reconstruction-based approach that output anomaly score for the detection.

4-4. Evaluations & Results

Performance evaluation metrics :

- Precision
- Recall
- F1-Score
- AUC

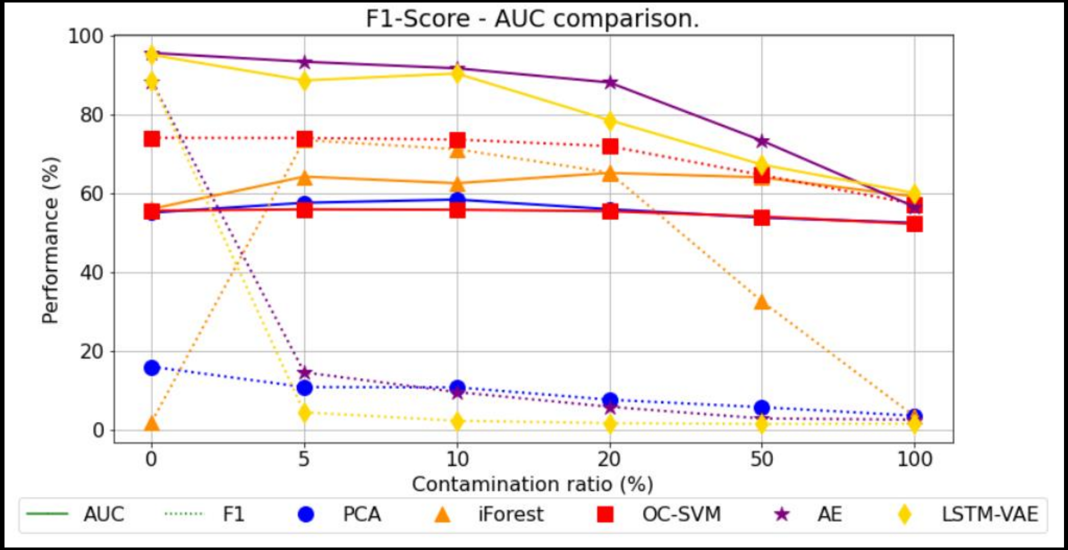
The best models without data contamination are the AE and LSTM-VAE.
For real-life situations, OC-SVM or iForest should be preferred since they are more robust to data contamination.

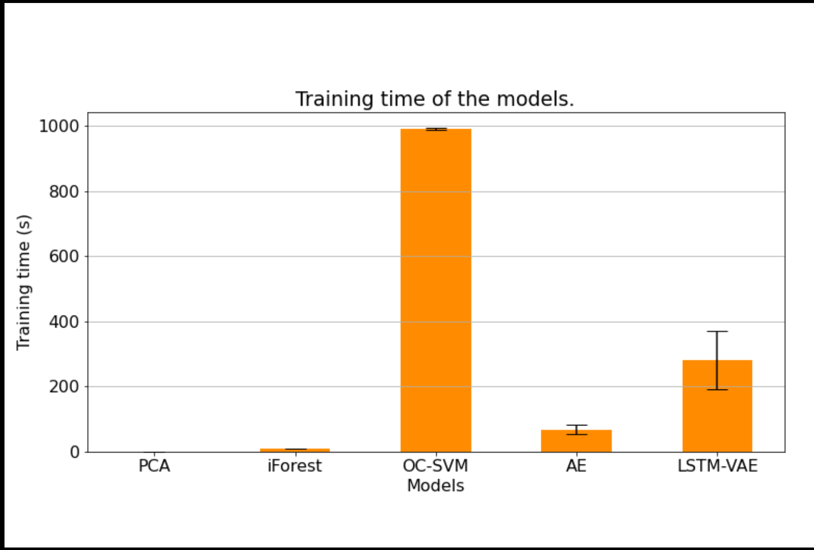
The same conclusions for the both data splitting strategies.

TABLE II
OVERALL PERFORMANCE (MEAN AND STANDARD DEVIATIONS OVER THE 5 RUNS) ON THE MIXED-DATASETS STRATEGY. δ IS THE ANOMALY CONTAMINATION RATIO IN THE TRAINING DATASET.

		Mixed-datasets		
	δ (%)	Precision	Recall	F1-score
PCA	0	82.01±0.14	8.83±0.1	15.94±0.16
	5	88.76±4.7	5.77±1.56	10.8±2.79
	10	83.92±2.75	5.74±0.9	10.73±1.6
	20	73.44±1.07	4.00±0.23	7.59±0.42
	50	65.36±0.89	2.97±0.12	5.69±0.22
	100	53.72±0.45	1.82±0.02	3.53±0.03
iForest	0	68.18±2.8	1±0.41	1.97±0.8
	5	62.18±0.58	89.77±1.96	73.47±0.96
	10	63.19±0.35	81.5±3.34	71.15±1.34
	20	68.44±1.05	62.53±5.47	65.23±3.32
	50	77.21±2.03	20.85±4.29	32.58±5.35
	100	74.61±3.25	1.76±0.74	3.43±1.42
OC-SVM	0	59.29±0.01	98.59±0.07	74.01±0.02
	5	59.5±0.02	97.82±0.14	74±0.03
	10	59.86±0.02	95.52±0.23	73.6±0.07
	20	60.51±0.04	88.65±0.25	71.93±0.07
	50	60.98±0.05	68.65±0.54	64.59±0.24
	100	60.28±0.05	54.47±0.31	57.23±0.19
AE	0	99.02±0.05	79.65±0.06	88.28±0.03
	5	95.55±0.43	7.86±0.07	14.52±0.12
	10	94.09±0.79	5.02±0.08	9.53±0.14
	20	91.45±1.28	3.00±0.07	5.80±0.14
	50	80.61±1.86	1.44±0.04	2.83±0.07
	100	77.68±1.12	1.28±0.03	2.52±0.06
LSTM-VAE	0	98.44±0.76	80.27±1.3	88.42±0.83
	5	98.58±0.7	2.25±1.02	4.38±1.95
	10	98.43±4.01	1.15±0.8	2.25±1.56
	20	83.95±25.31	0.83±0.43	1.65±0.83
	50	88.83±14.22	0.72±0.37	1.43±0.74
	100	95.37±2.21	0.75±0.26	1.48±0.51

4-4. Evaluations & Results





The comparison between F1-score and AUC show how misleading the AUC score can be when the test set is imbalanced.

The training time for OC-SVM is very high compared to those of the iForest that has the same performance with data contamination.

4-5. Conclusion

- Our comparative analysis show that data contamination have a high impact on unsupervised ML models.
  - AE and LSTM-VAE are the best without data contamination.
  - OC-SVM and iForest are the most robust to data contamination even if OC-SVM has a longer training time.

- Our evaluation however has some limitations:
  - The reconstruction-based approach are evaluated with the 3-sigma rule for threshold selection.
  - The point-wise anomaly detection is not well-suited for the detection of CG quality degradation.

- In future work we will
  - Perform additional evaluations with state-of-the-art approaches for anomaly detection
  - Use sequences of observations instead of point observations to better model an anomaly for cloud-gaming sessions.
  - Study the impact of the threshold for the performance of reconstruction-based models.
  - Explain the anomalies for root-cause analysis.

5. Conclusion

- A paper about this work on the detection of Cloud Gaming is accepted and will be presented at the IEEE CNSM Workshop, HiPNet :

  J. Ky, B. Mathieu, A. Lahmadi, R. Boutaba, "Assessing Unsupervised Machine Learning solutions for Anomaly Detection in Cloud Gaming Sessions", 4th International Workshop on High-Precision, Predictable, and Low-Latency Networking (HiPNet), 18th International Conference on Network and Service Management (CNSM), Thessaloniki, Greece, 31 October - 4 November 2022.

- The code for the experiments is available at : https://github.com/joelromanky/cg-ano-detect-eval

- The data for emulated networks conditions is available as open data at: https://cloud-gaming-traces.lhs.loria.fr/cellular.html

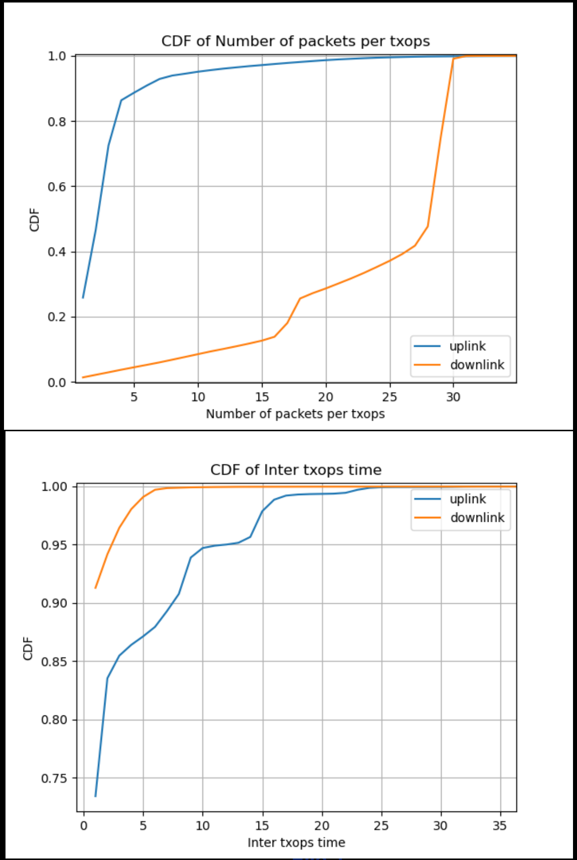
Orange
Innovation

Thank you

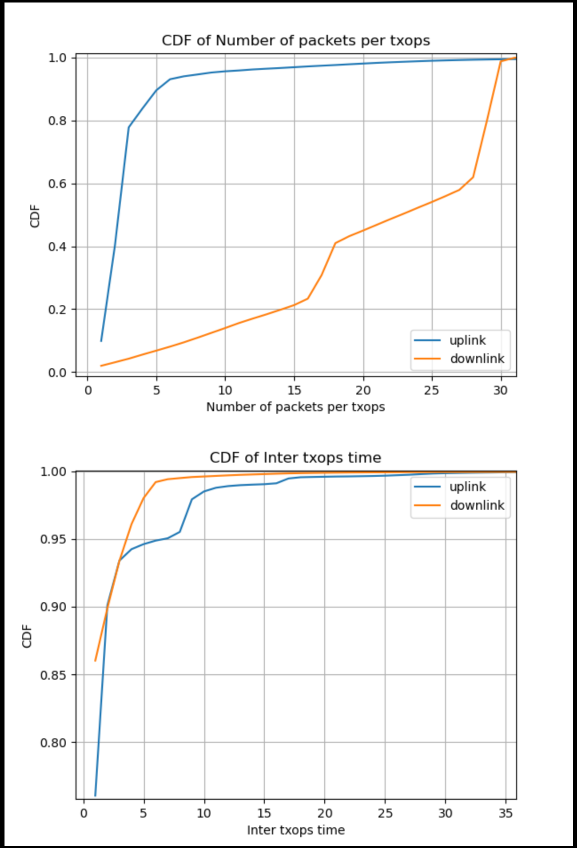
6. References

- [Mahimahi] : http://mahimahi.mit.edu/
- [Saturatr] : https://github.com/keithw/multisend/blob/master/sender/saturatr.cc
- [ARCEP] : Qualité des services mobiles | Arcep

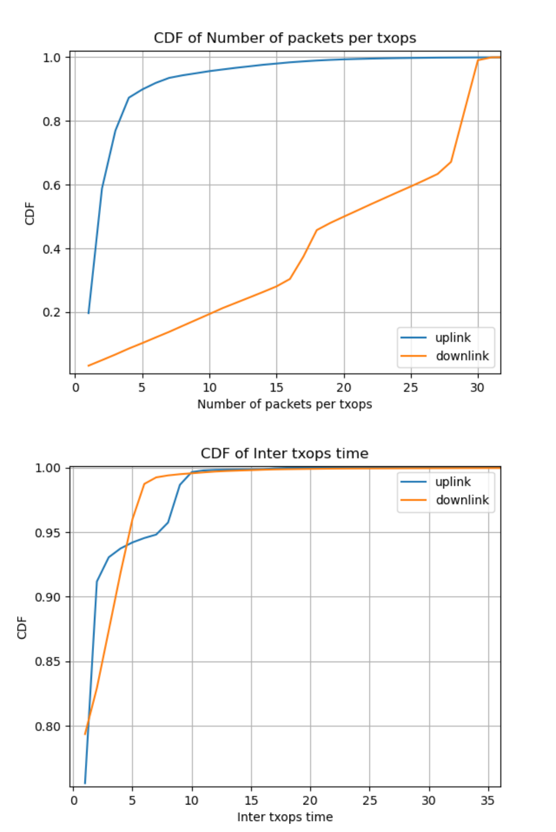
A-1. Characterization of 4G txops measured



File 1

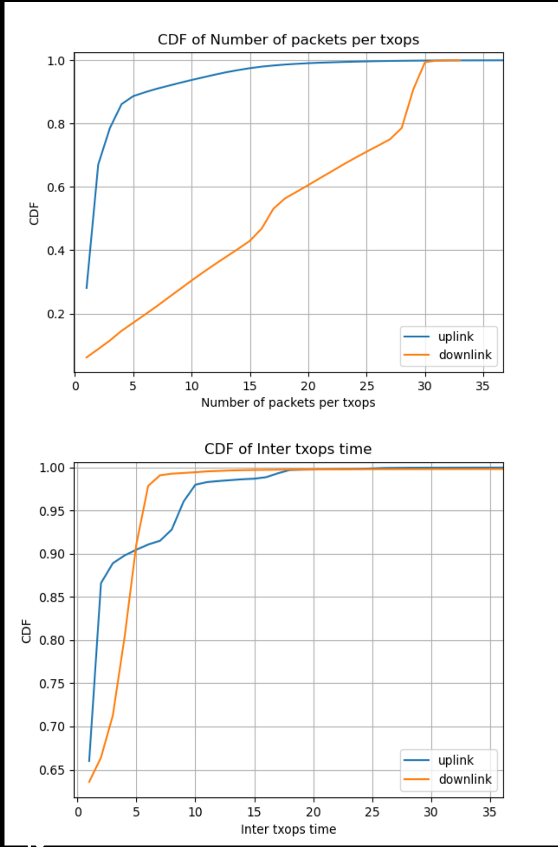


File 2

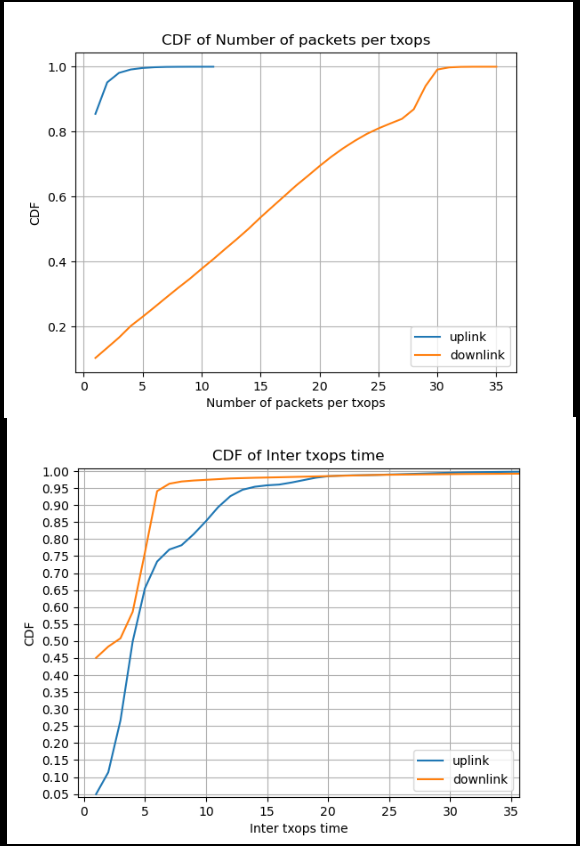


File 3

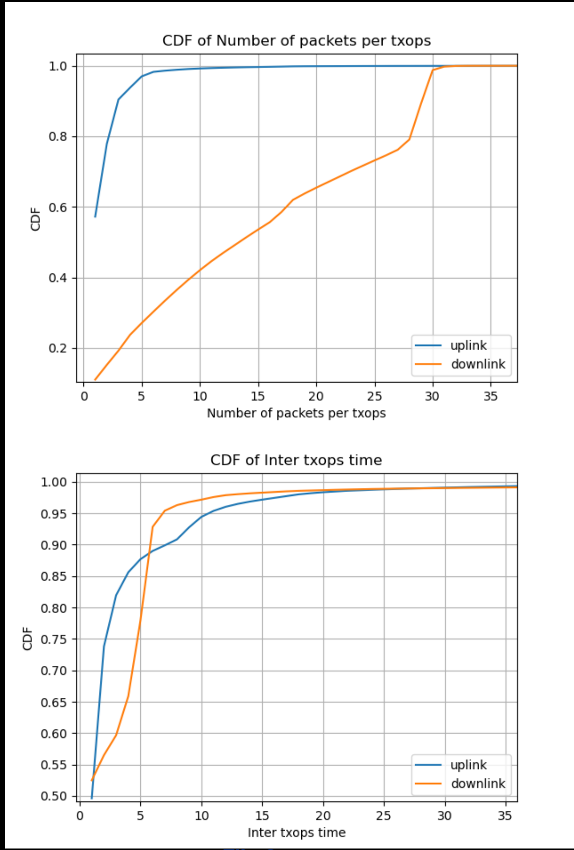
A-1. Characterization of 4G txops measured



File 4

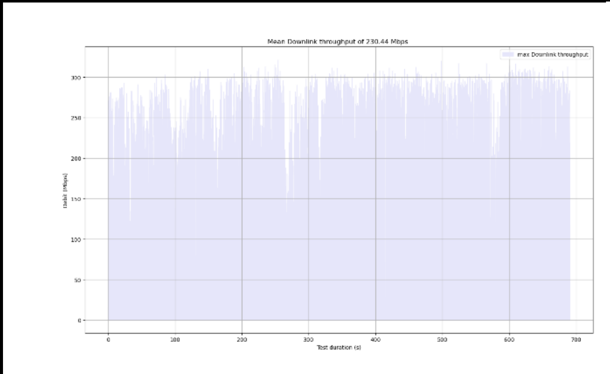


File 5

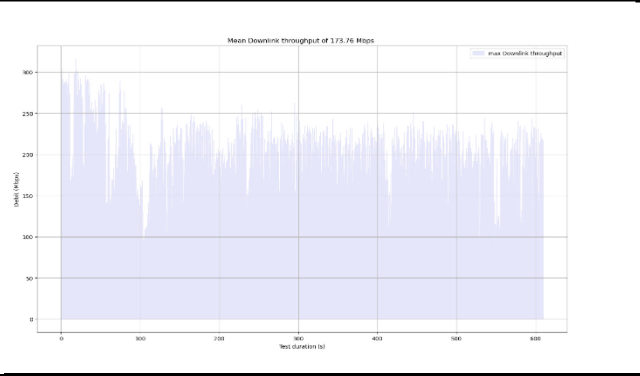


File 6

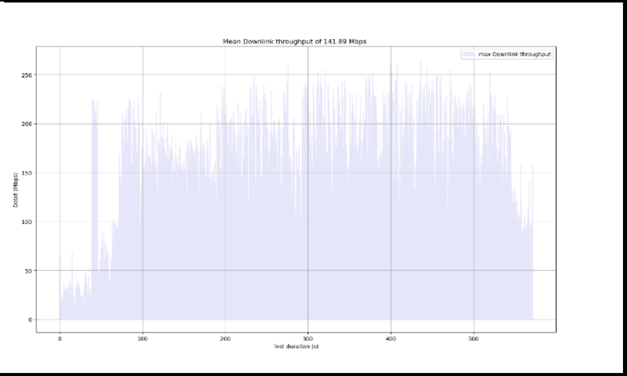
A-2. Max downlink throughput on the txops files



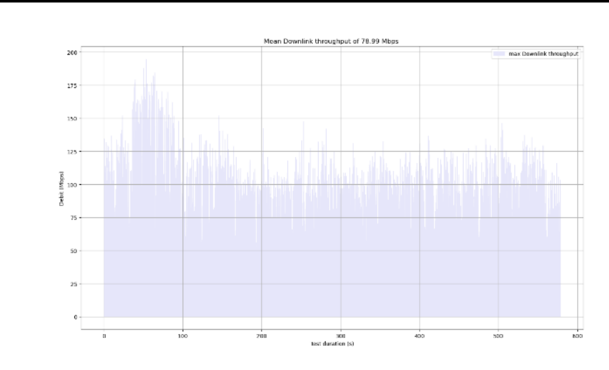
File 1



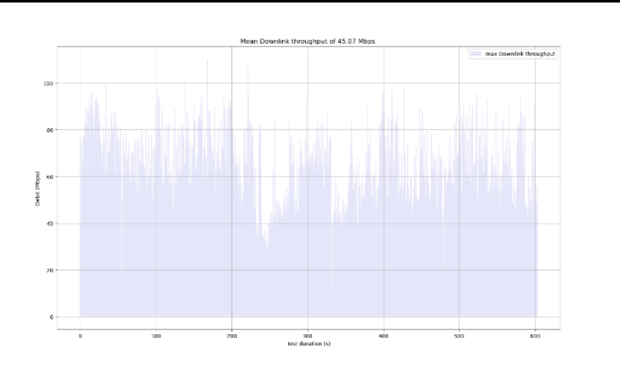
File 2



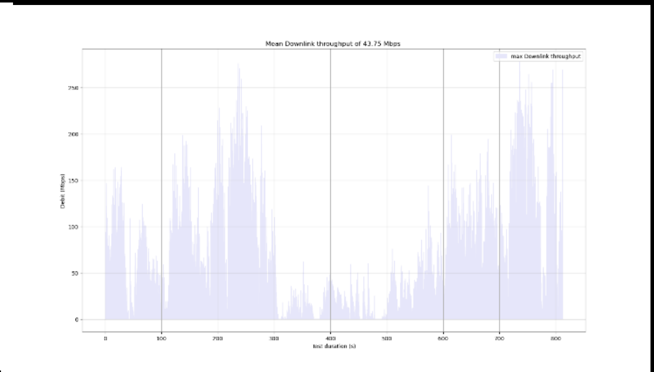
File 3



File 4



File 5



File 6

B. Results with the high-bitrate splitting strategy

TABLE III

OVERALL PERFORMANCE (MEAN AND STANDARD DEVIATIONS OVER THE 5 RUNS) ON THE HIGH-BITRATE TRAINING SET STRATEGY. δ IS THE ANOMALY CONTAMINATION RATIO IN THE TRAINING DATASET.

		High-bitrate training datasets		
	δ (%)	**Precision**	**Recall**	**F1-score**
PCA	0	98.01±0	10.89±0	19.6±0
	5	95.36±3.64	7.38±0.24	13.7±0.42
	10	97.84±1.25	8.1±0.65	14.96±1.11
	20	90.90±2.87	7.89±0.21	14.51±0.36
	50	89.88±0.34	5.85±0.2	10.99±0.36
	100	92.21±0	5.93±0	11.14±0
iForest	0	88.92±1.33	38.21±4.47	53.33±4.56
	5	98.18±0.42	88.94±0.56	93.33±0.26
	10	98.63±0.48	88.21±0.4	93.13±0.16
	20	98.56±0.15	89.16±0.51	93.62±0.26
	50	99.37±0.13	87.5±0.29	93.06±0.12
	100	99.16±0.48	78.3±4.26	87.43±2.57
AE	0	99.68±0.11	86.77±0.22	92.77±0.08
	5	99.43±0.12	8.48±1.87	15.57±3.26
	10	99.28±0.29	3.95±0.52	7.58±0.97
	20	99.23±0.08	1.82±0.27	3.58±0.53
	50	99.03±0.28	0.77±0.02	1.53±0.04
	100	99.17±0.01	0.76±0.01	1.51±0.01
LSTM-VAE	0	99.79±0.1	86.59±0.72	92.72±0.41
	5	97.5±3.47	7.59±11.53	12.03±16.95
	10	89.7±19.85	1±0.54	1.98±1.06
	20	99.49±0.16	1.12±0.54	2.21±1.06
	50	94.32±9.62	0.93±0.3	1.83±0.59
	100	99.49±0.27	1.55±1.43	3.02±2.72