Assessing Unsupervised Machine Learning solutions for Anomaly Detection in Cloud Gaming Sessions

Joël Roman Ky, Bertrand Mathieu – Orange Innovation Lannion
Abdelkader Lahmadi – LORIA
Raouf Boutaba – University of Waterloo

HiPNet '22
Thessaloniki, 04/11/22

Table of contents

1. Context & Motivation

- Stringent network requirements of low-latency applications (CG) :
  - Network issues for end-users.

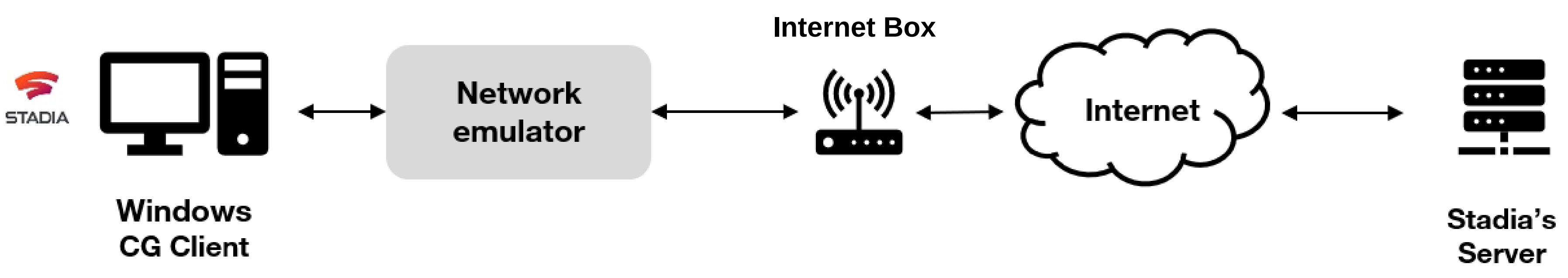
- Need to detect user quality degradation.

- Machine-learning approaches (ML) efficient in anomaly detection but supervised learning require labeled data.
  - Impractical due to the increasing network complexity.
  - **=>** Use of unsupervised ML models.

- Evaluation of 5 unsupervised ML models with datasets collected on Google Stadia CG server under 6 different 4G emulated network conditions.

2. Testbed

- Public Google Stadia platform with the traffic routed through the Internet.
- WebRTC API to provide client-side QoS/QoE metrics.
- Played on 4G network conditions emulated with the Mahimahi tool [Mahimahi], based on real 4G conditions from the commercial french ISP, Orange.



3. Datasets collected

- 5 static scenarios
- 1 mobility scenario on highway

220 Mbps	**160 Mbps**
140 Mbps	**120 Mbps**
80 Mbps	**Highway**

X_{220}	X_{160}	X_{120}
X_{80}	X_{40}	$X_{highway}$

14 QoS/QoE features with a time-step of 5ms :

- Bitrate, RTT, client-processing delay, frame-rate, resolution, freeze occurrences, frames dropped, video reendering jitter

Chromium WebRTC API

- Downlink throughput reachable on the 4G emulated network condition.

4. Evaluation of Unsupervised ML models for Anomaly Detection

- 4-1. Unsupervised ML models
- 4-2. Data processing
- 4-3. Performance assessment
- 4-4. Evaluations & Results

4-1. Unsupervised ML models

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

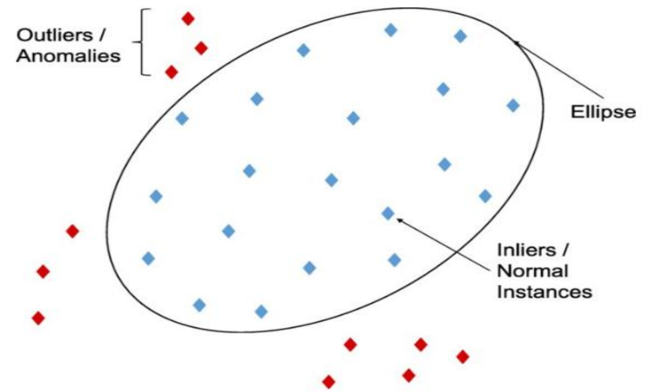


Fig : OC-SVM

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

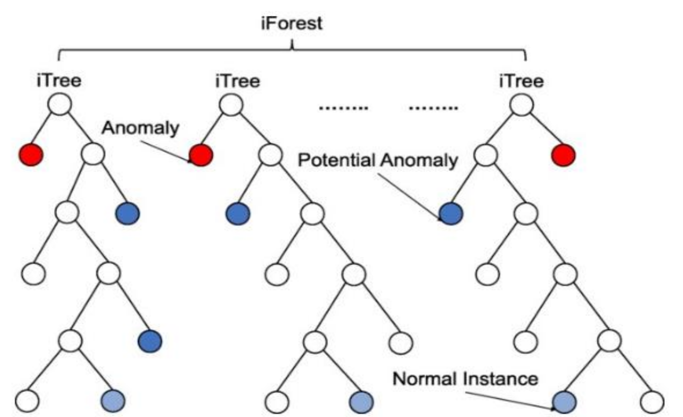
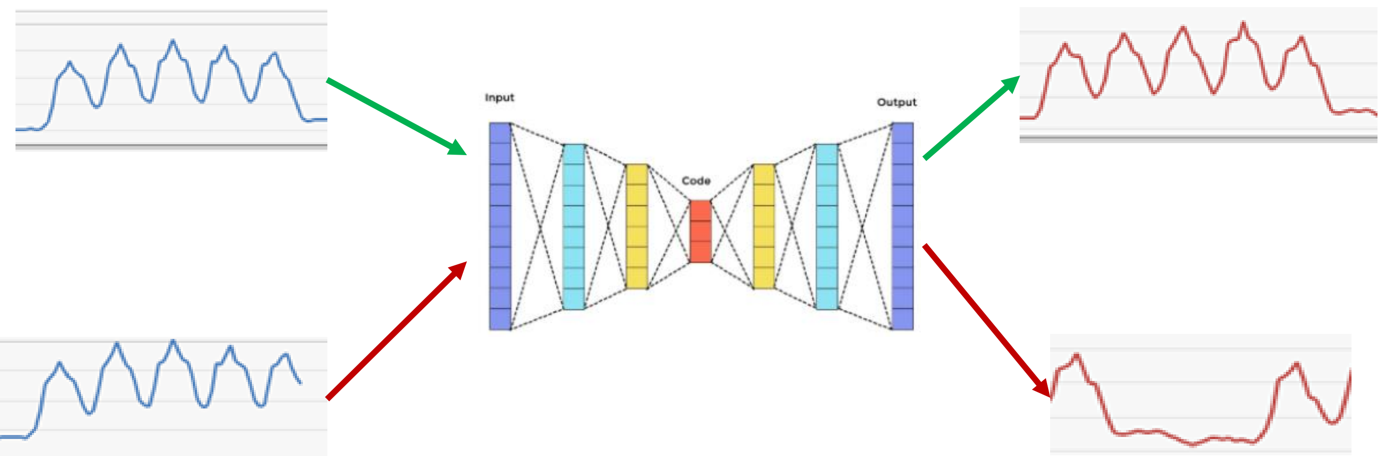


Fig : Isolation Forest

4-1. Unsupervised ML models : Reconstruction approaches



Good reconstruction

Bad reconstruction

Input

Reconstruction

$$anomaly_{score} = (X_{input} - X_{reconst})^2$$

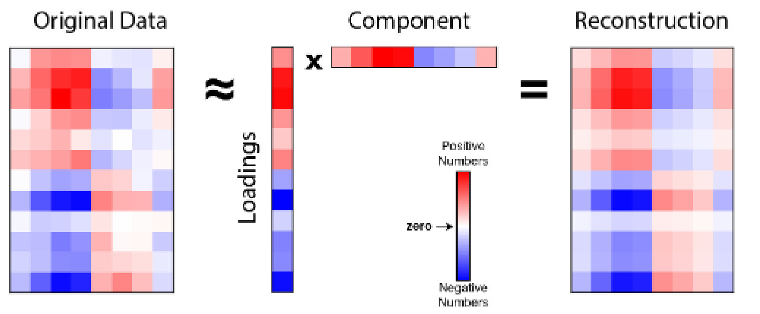
3-sigma thresholding rule

If >= threshold
anomalous obs.

If < threshold
normal obs.

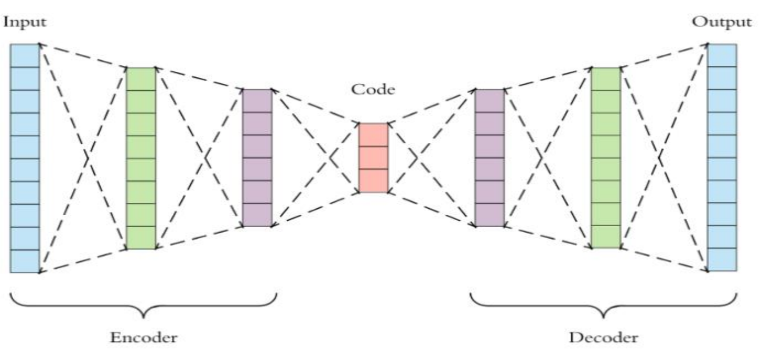
4-1. Unsupervised ML models : Reconstruction approaches

- PCA: Reconstruction of the data with principal components.



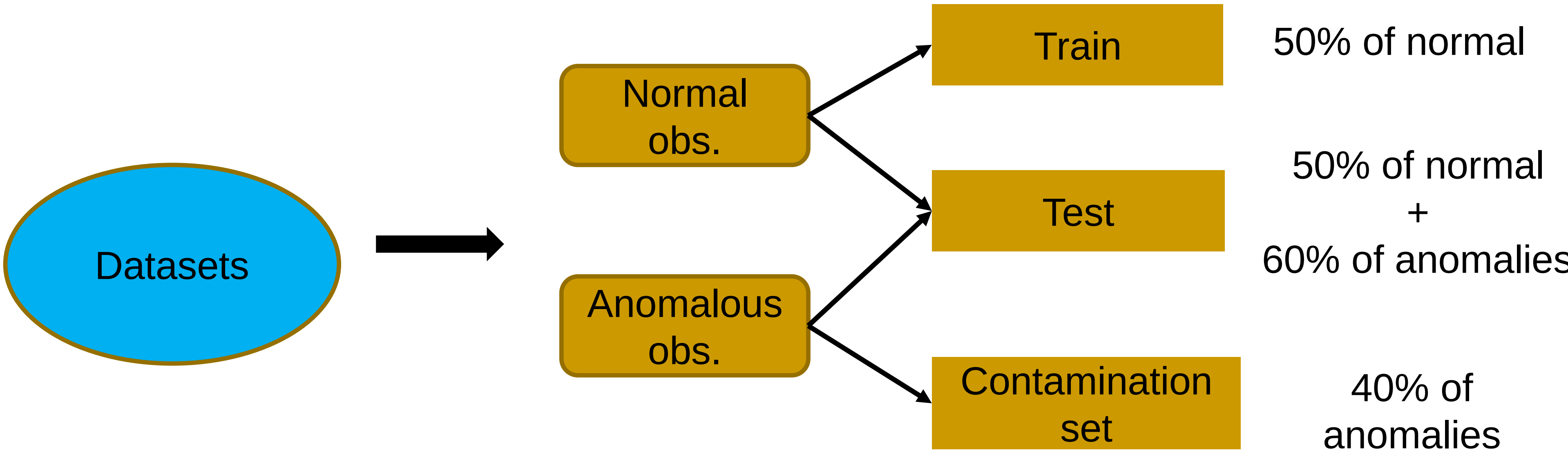
- Auto-Encoder (AE): Constitued of an encoder, that learns from inputs a low-dimensional representation of data, and a decoder that reconstruct original data from latent variable.

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

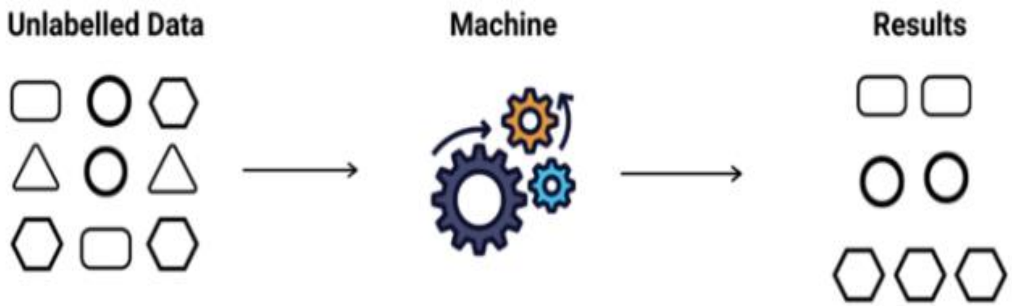


4-2. Data processing

- Train/Test datasets splitting.
- Contamination set to assess the robustness of unsupervised ML models to data contamination.
  - Real-life datasets not free of anomalies.



4-3. Performance assessment



After training, how can we know if the model correctly predicted ?

Ground truths required to objectively assess the performance.

Ground truths created for the performance assessment

Observation defined as anomalies if satisfies one of the following criteria

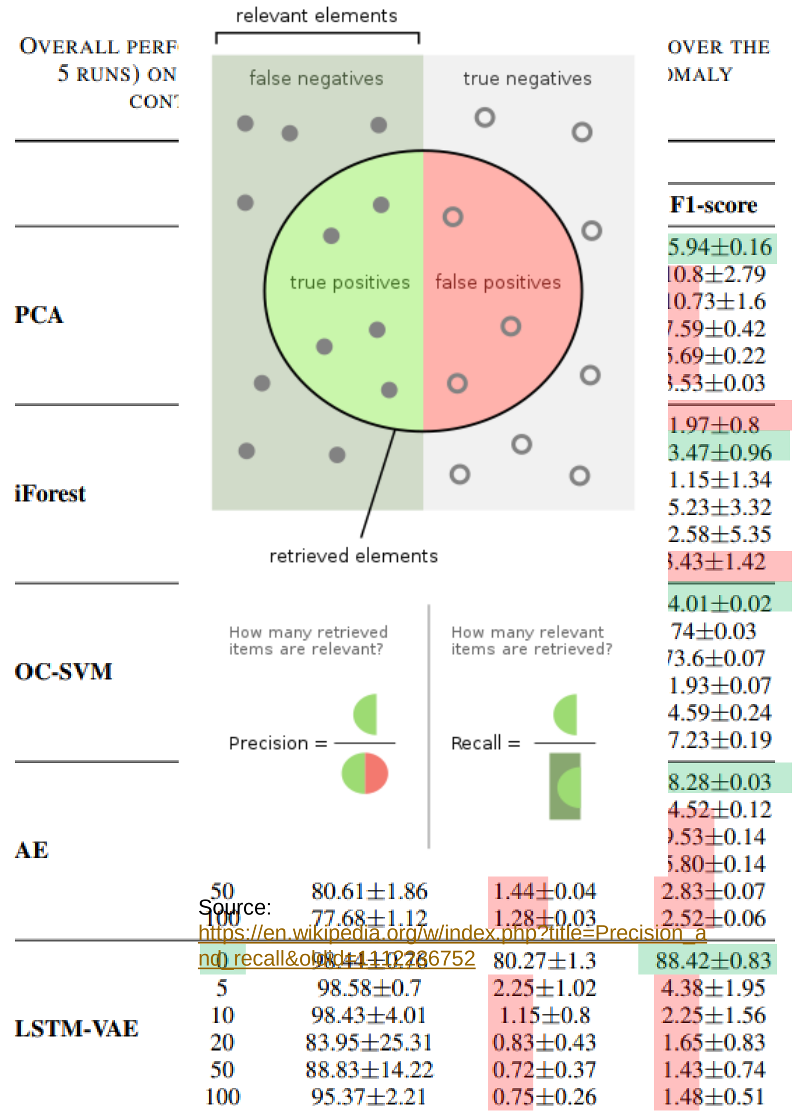
- Frame rate < 60 FPS
- Resolution < 1080p
- Freeze occurence

4-4. Evaluations & Results

Performance evaluation metrics :

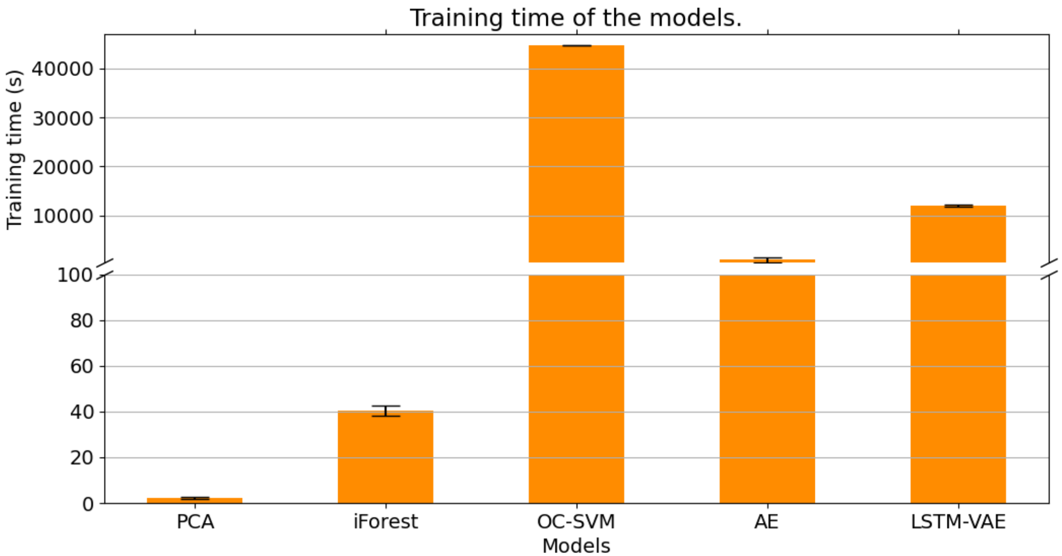
- Precision: $P = \frac{TP}{TP+FP}$
- Recall: $R = \frac{TP}{TP+FN}$
- F1-Score: $F1 = 2\frac{P \, . \, R}{P+R}$

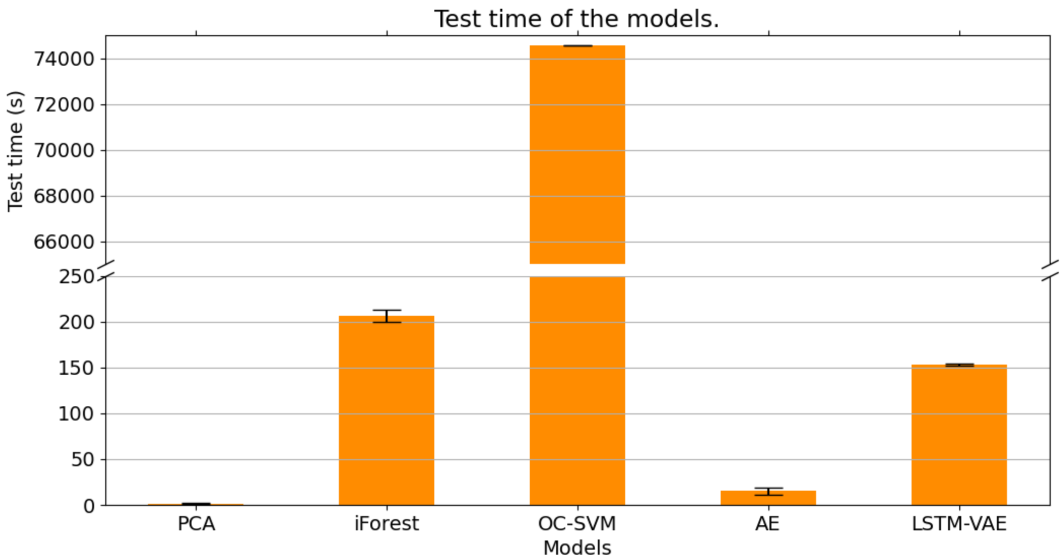
- Best models without data contamination: AE and LSTM-VAE.
- OC-SVM or iForest more robust to data contamination.



Source: https://en.wikipedia.org/w/index.php?title=Precision_and_recall&oldid=1112286752

4-4. Evaluations & Results





The training/test time for OC-SVM very high compared to iForest for the same performance with data contamination.

5. Conclusion

- High impact of data contamination on unsupervised ML models.
  - AE and LSTM-VAE better without data contamination.
  - OC-SVM and iForest more robust to data contamination but OC-SVM has a longer training/test time.

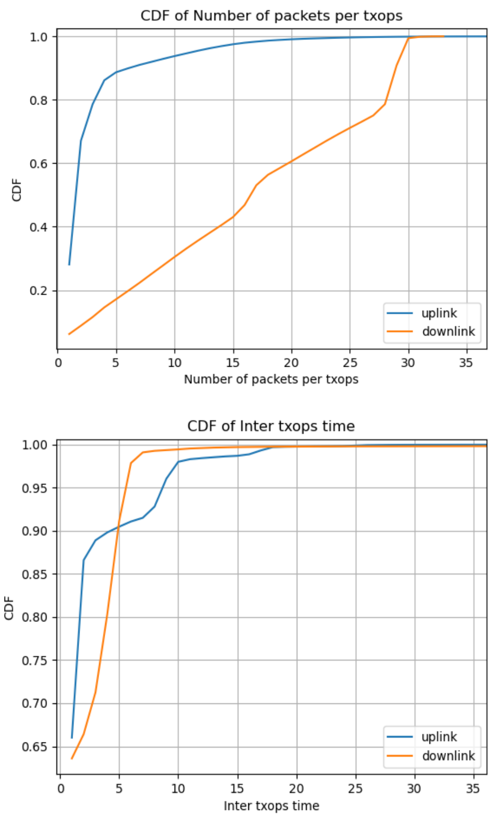
- Some current limitations:
  - Reconstruction-based approach evaluated with the 3-sigma rule for threshold selection.
  - Point-wise anomaly detection not well-suited for the detection of CG quality degradation.

- Future work:
  - Additional evaluations with state-of-the-art approaches
  - 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.

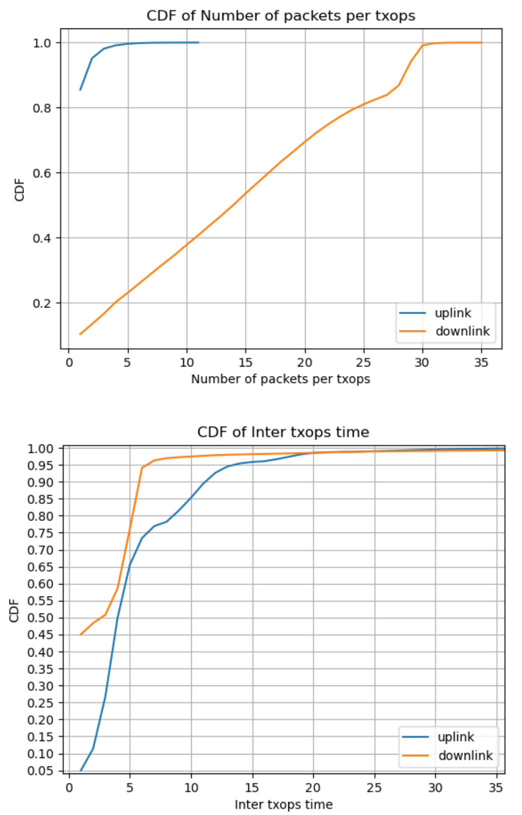
Questions

Q & A

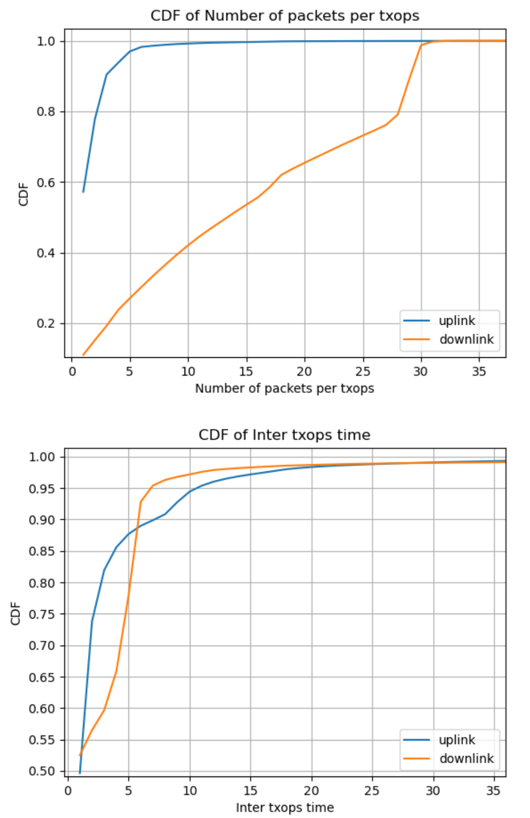
A-1. Characterization of 4G txops measured



File 4

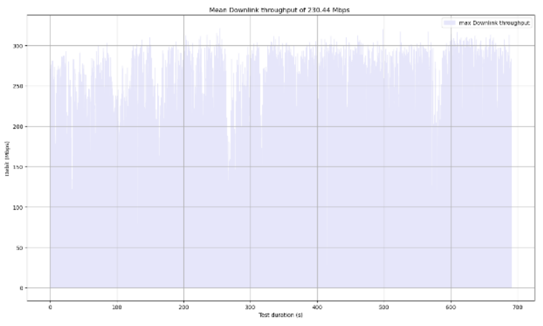


File 5

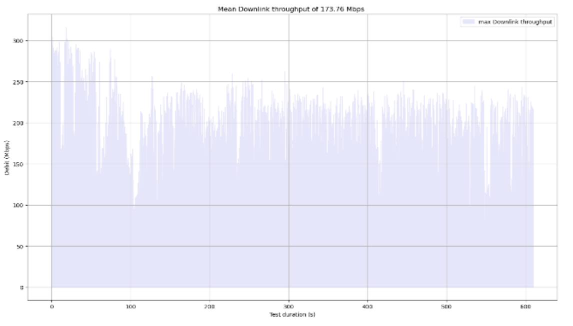


File 6

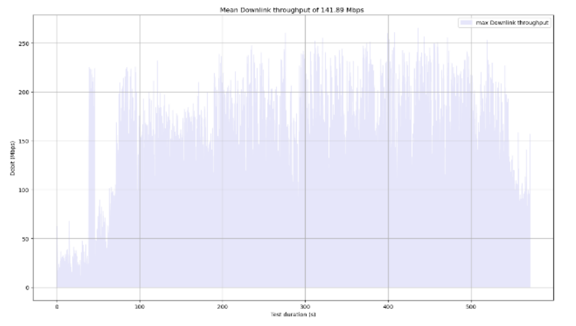
A-2. Max downlink throughput on the txops files



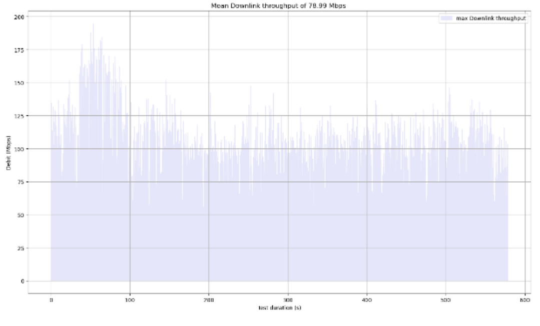
File 1



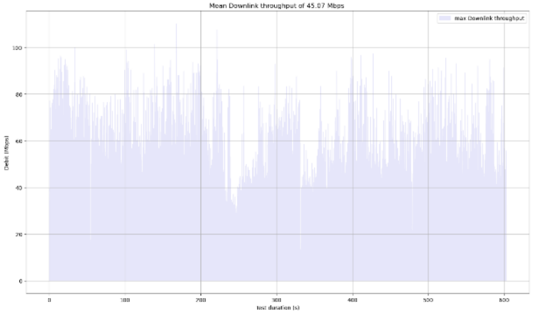
File 2



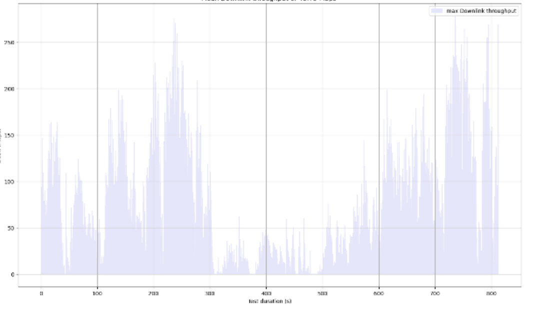
File 3



File 4



File 5



File 6

B. Generation of realistic cellular network conditions

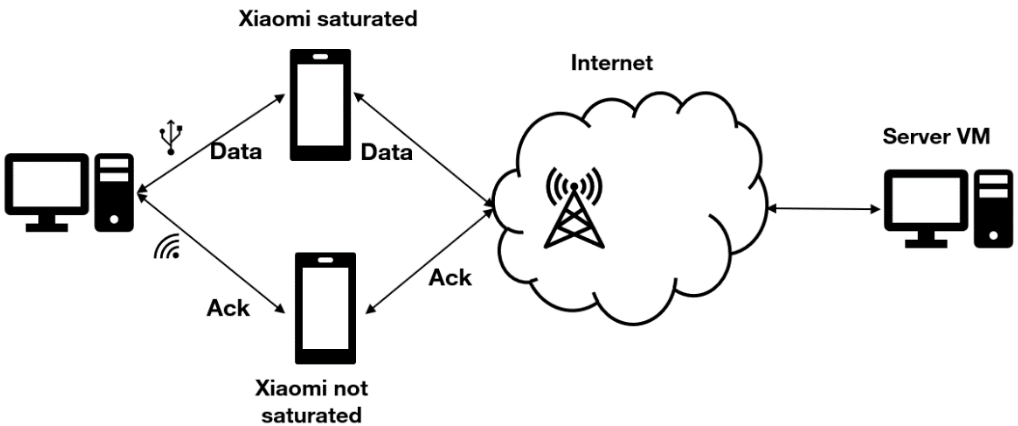
2-1. Motivation :

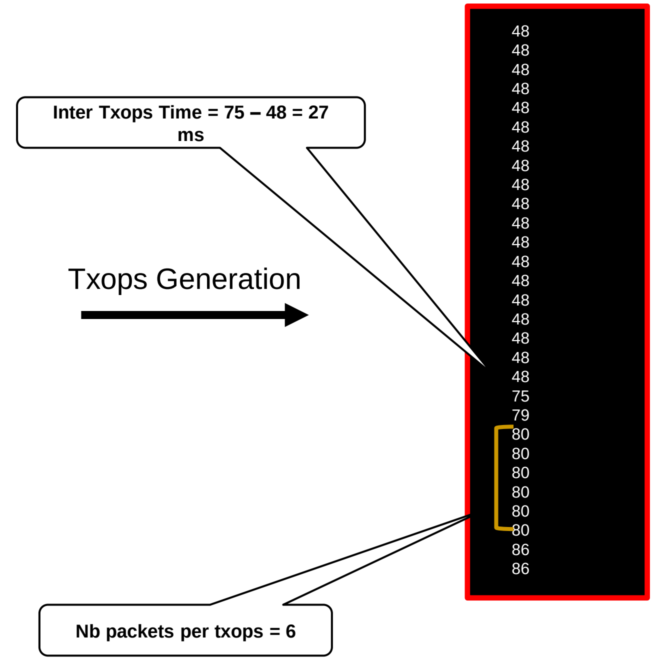
- How can we conduct controlled experiments on realistic network conditions ?
  - The framework Mahimahi developed by MIT researchers.

- 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.

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





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

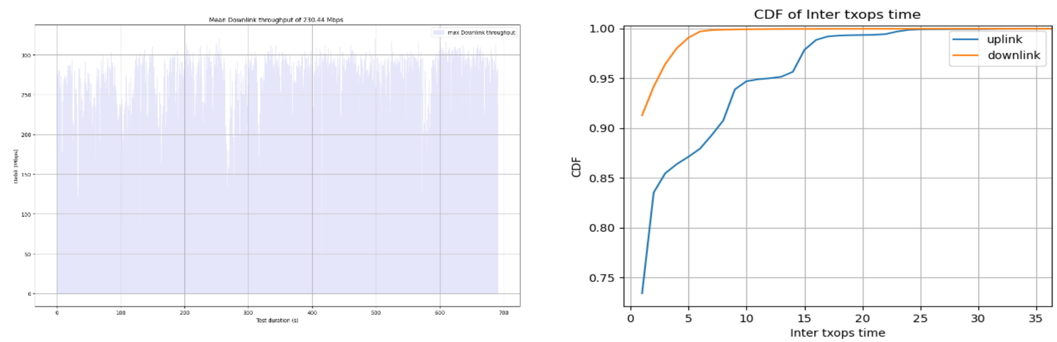
$$if\ (rtt > RTT_{upper}\ \&\&\ congestion\ window > window_{lower})\ then\ congestion\ window -= 20$$

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

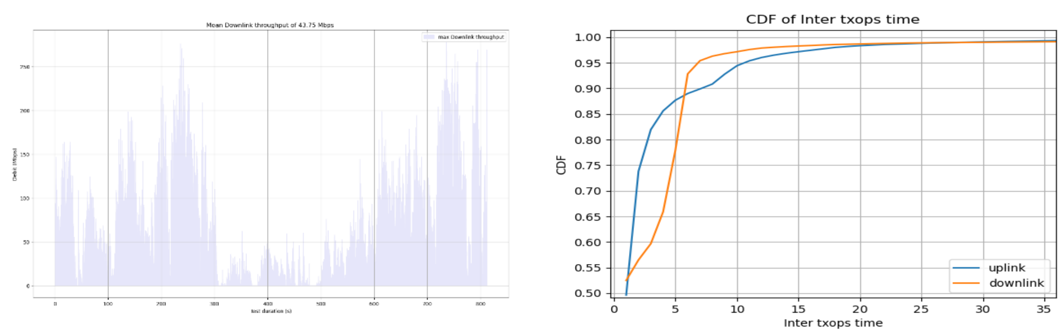
B-3. Characteristics of the measured cellular networks condition

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

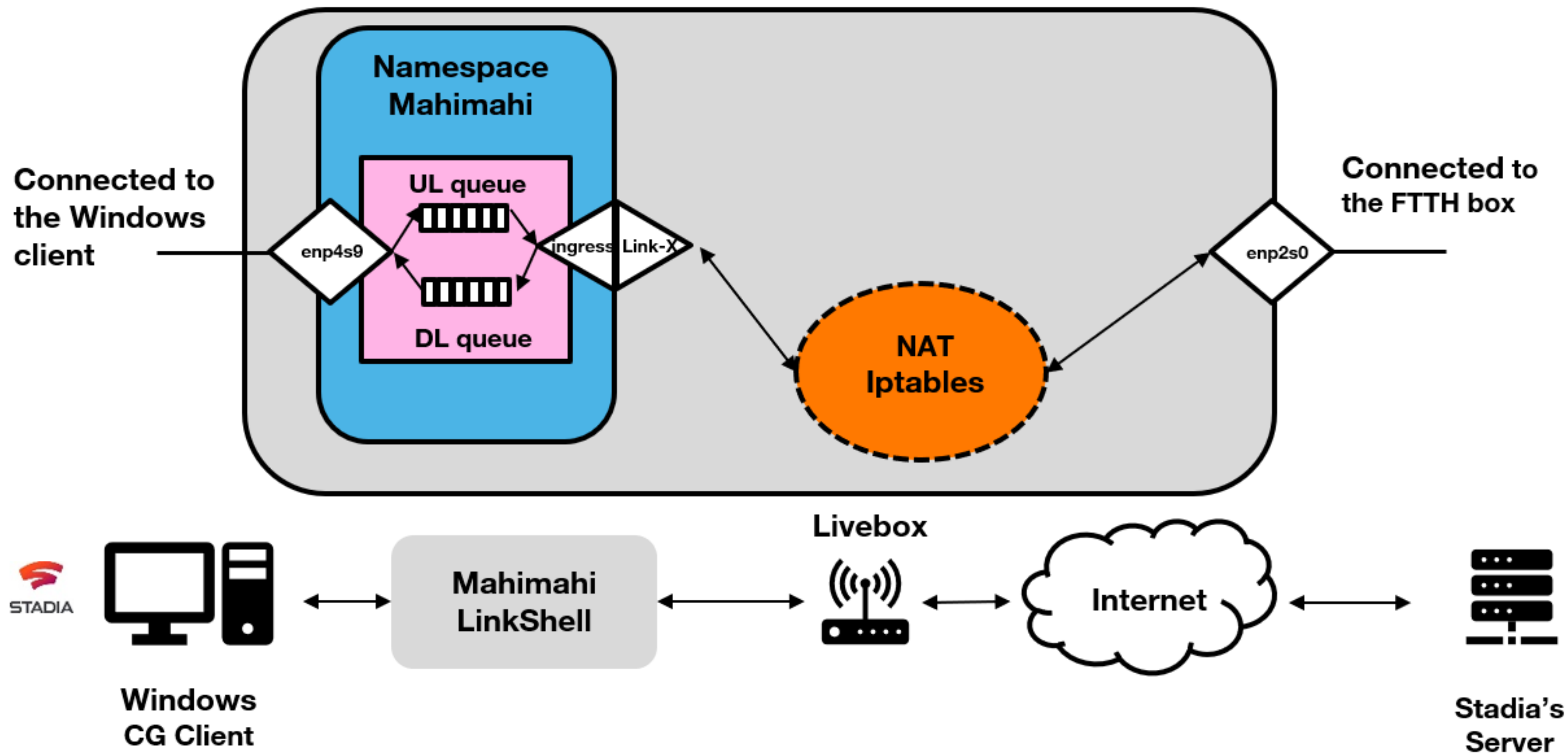


File 1



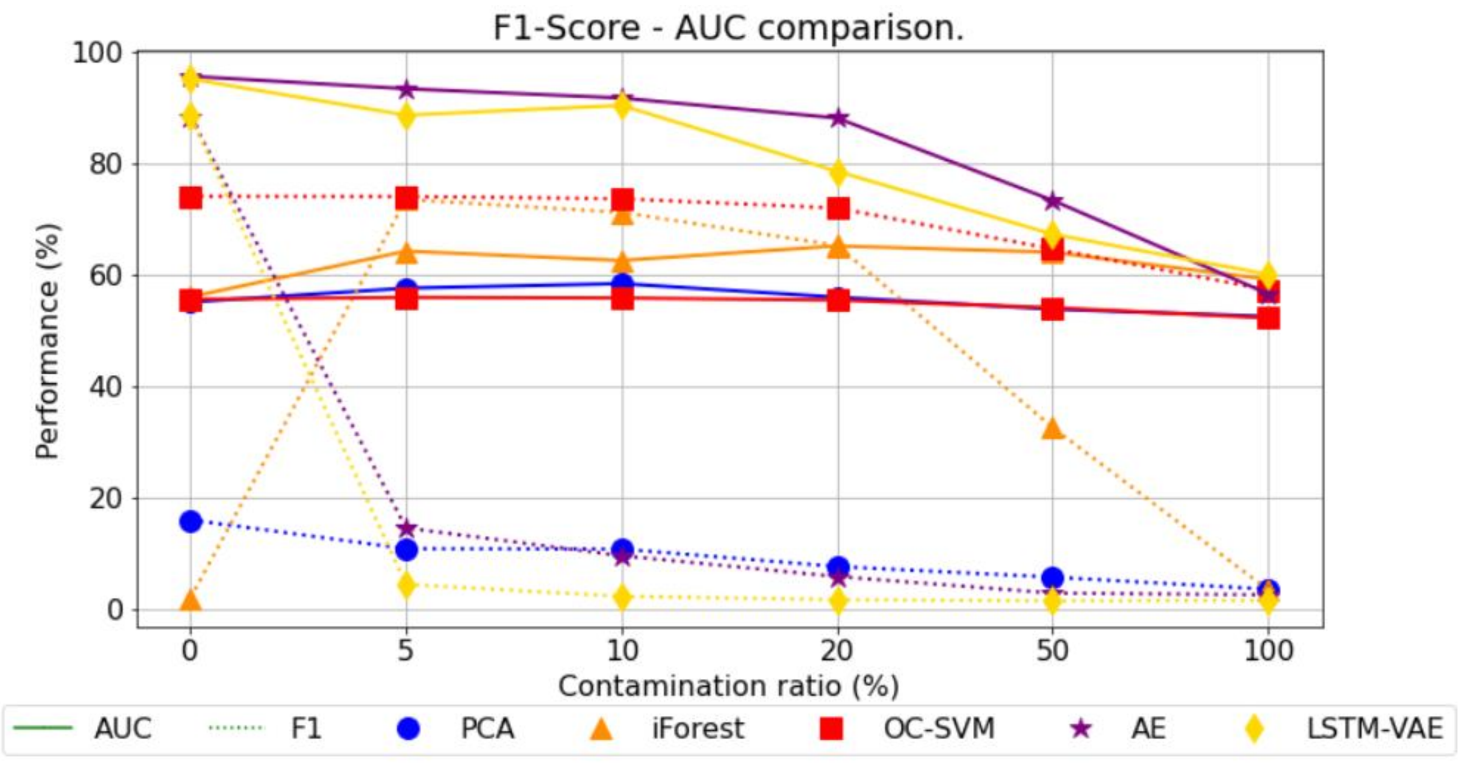
File 6

B-4. Testbed



Namespace Mahimahi
UL queue
DL queue
enp4s9
ingress
Link-X
NAT Iptables
enp2s0
Connected to the Windows client
Connected to the FTTH box
STADIA
Windows CG Client
Mahimahi LinkShell
Livebox
Internet
Stadia's Server

C. Evaluations & Results



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

D. Datasets

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

E. Evaluations & Results

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.

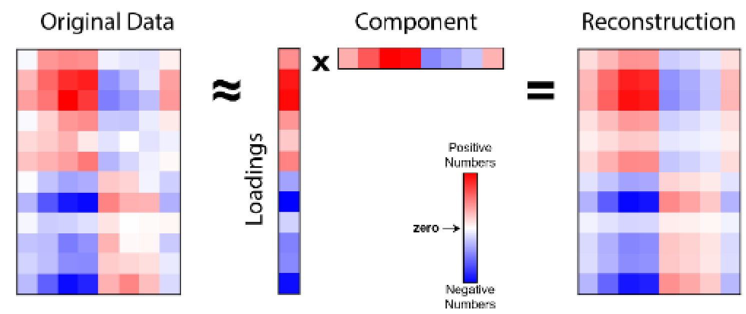
		High-bitrate training datasets		
	δ (%)	**Precision**	**Recall**	**F1-score**
PCA	0	98.01±0	10.89±0	19.6±0
	5	95.36±3.64	7.38±0.24	13.7±0.42
	10	97.84±1.25	8.1±0.65	14.96±1.11
	20	90.90±2.87	7.89±0.21	14.51±0.36
	50	89.88±0.34	5.85±0.2	10.99±0.36
	100	92.21±0	5.93±0	11.14±0
iForest	0	88.92±1.33	38.21±4.47	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

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

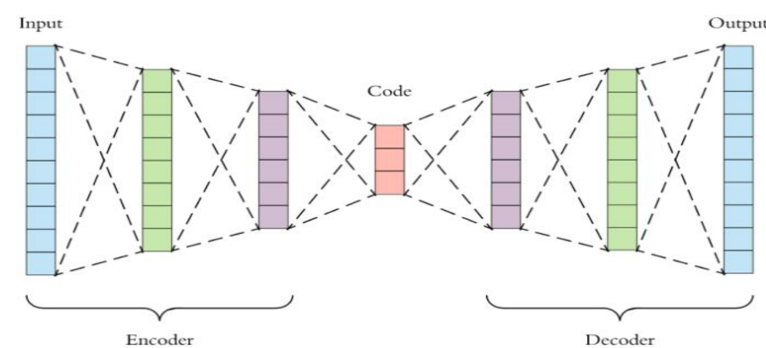
4-1. Unsupervised ML models : Reconstruction approaches

- **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.

$$anomaly\_score = (X_{true} - X_{recons})^2$$

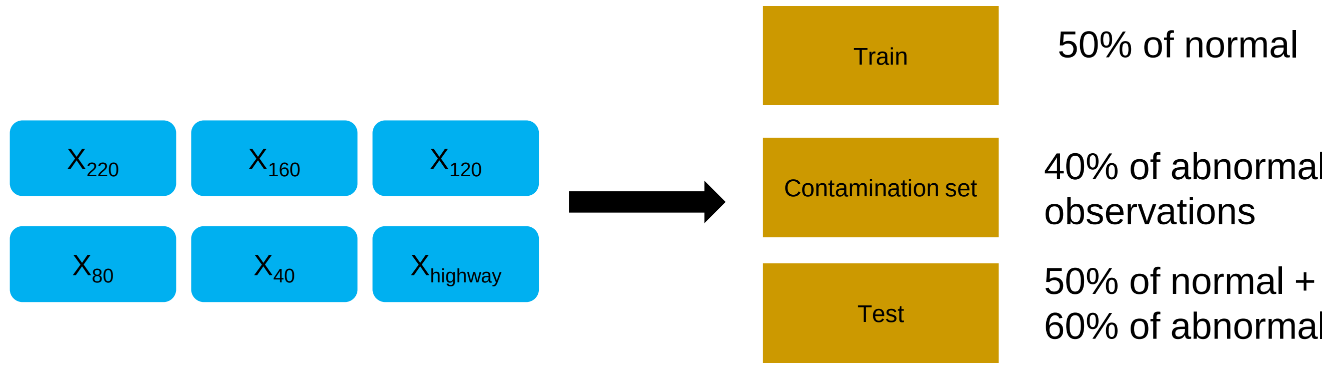


- **LSTM-VAE:** Combination of LSTM and a VAE (AE with bayesian inference).

4-2. Data processing

➢ Assessing the robustness of unsupervised ML models to data contamination

- **Mixed-dataset splitting**



- **High-bitrate splitting:**