Orange Innovation

Détection d'Anomalies et Diagnostic des Causes Racines des Applications à Faible-Latence sur les Réseaux à Capacité Variable

Soutenance de thèse pour l'obtention du titre de Docteur en Informatique de l'Université de Lorraine

Par Joël Roman Ky

29 avril 2025

Jury:

Isabelle Chrisment (Directrice de thèse)

Raouf Boutaba (Co-directeur)

Abdelkader Lahmadi (Encadrant)

Bertrand Mathieu (Encadrant)

Sandrine Vaton (Rapporteuse)

Nadjib Aitsaadi (Rapporteur)

Hind Castel (Examinatrice)

Yacine Ghamri (Examinateur)

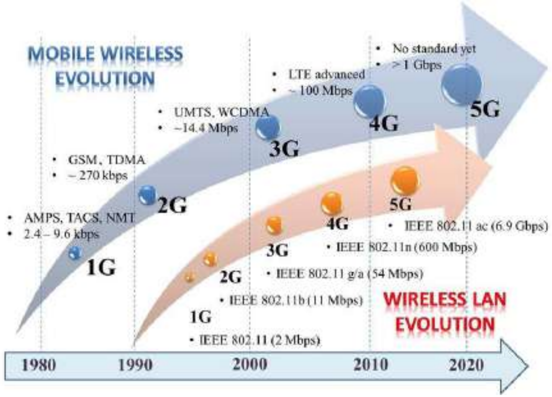
Claudia Ignat (Invitée)

Table of contents

1. Context & Motivation

1. Motivation

- Recents network evolutions (Wi-Fi and mobile networks).



- Low latency (LL) applications (Cloud Gaming, Cloud VR) which is gaining importance.

- Current time varying capacity networks (4G/5G, WiFi) focus on bandwidth, neglecting latency
  - LL applications need more than speed: stability is required for better Quality of Experience (QoE).

luna

GEFORCE NOW

CLOUD GAMING

STADIA

1. Motivation

- LL applications suffer in real-world time varying capacity networks :
  - Wi-Fi: higher RTT and jitter may occurs due to attenuation.
  - Cellular networks: delay spikes due to signal drops or handovers.

- We need smart detection and diagnostic solutions to improve QoE for LL applications.

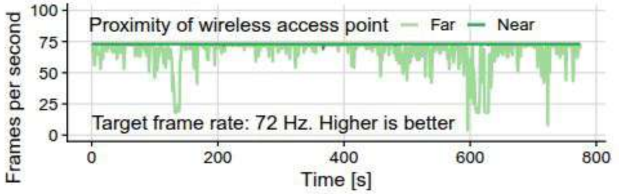
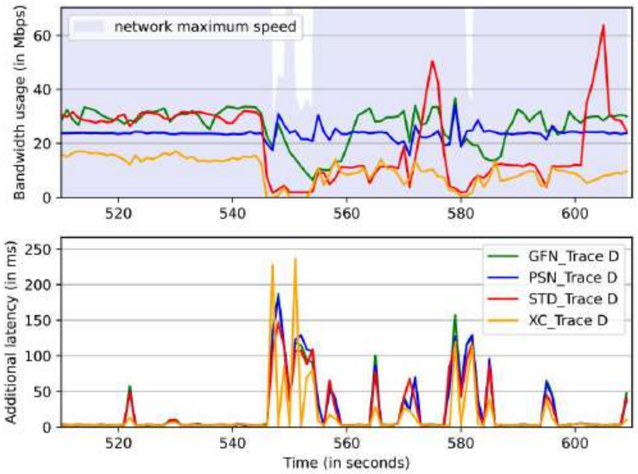


Figure 1: An example result from our performance characterization, showing the effect of wireless access point proximity on VR performance.

[Jansen et al. 2023]



[Marchal et al. 2023]

2. Problem Statement & Objectives

2. Problem Statement & Objectives

Goal of this thesis: Propose efficient and robust methods to detect and diagnose the causes of performance degradation in LL applications based on KPIs collected.

- Improve QoE on LL applications for individual and enterprises clients.
- Improve or design networks infrastructures (5G and WiFi) to support LL applications requirements (latency and jitter especially).

Data collection

Anomaly Detection

Root Cause Diagnosis

3. Background

3-1. Background: Data Collection

- No existing datasets capture LL applications behavior under realistic network conditions.

- Network issues faced by LL applications that prone to QoE degradation:
    - Cellular (congestion, coverage, interference, handovers, …)
    - WiFi (interference, congestion, signal attenuation, hidden terminals, …)

3-2. Background: Anomaly Detection

- An anomaly is an observation that deviates considerably from some concept of normality [Chandola et al.].
- Expert-defined rule-based techniques are fast but no longer scale.
- Use of ML/DL solutions to circumvent these limitations.
  - Given a multivariate time-series dataset $X = \{x_1, x_2, \ldots, x_T\}$ with $x_t \in \mathbb{R}^m$ we train f_θ that for each new observation outputs an **anomaly score** $s(\tilde{x}_t)$

$$\tilde{y}_t = \begin{cases} 1, & \text{if } s(\tilde{x}_t) > \delta \\ 0, & \text{otherwise.} \end{cases}$$

  - Performance metrics:

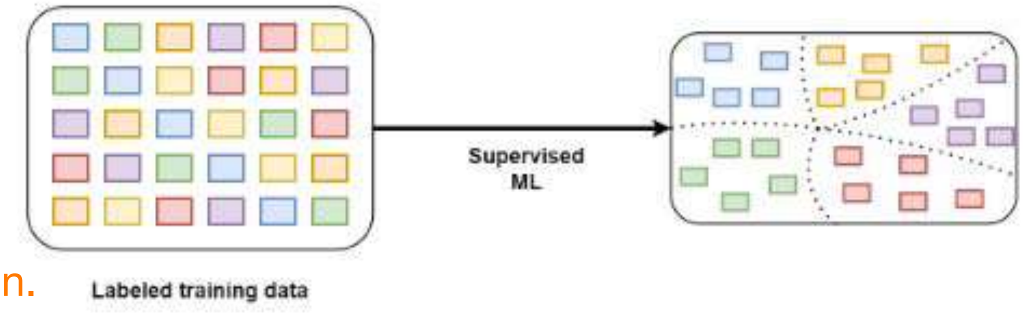
$$P = \frac{TP}{TP + FP} \qquad F_1 = 2.\frac{P.R}{P + R}$$

$$R = \frac{TP}{TP + FN} \qquad MCC = \frac{TP.TN - FP.FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

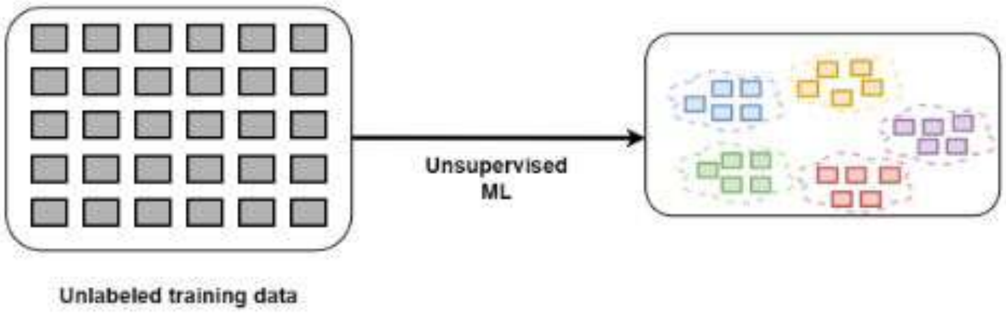
3-2. Background: Anomaly Detection

Two time series AD approaches:

- Supervised AD: learns from labeled anomalies.
  - ✔ Very efficient
  - ✘ Need extensive labels often unavailable or expensive to obtain.



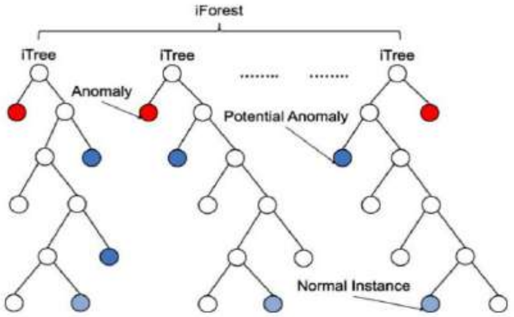
- Unsupervised AD: learns from unlabeled data.
  - ✔ No labels required
  - ✘ Assume that training data are free of anomalies => Risk of data contamination.



3-2. Background: Anomaly Detection

Classical unsupervised ML methods:

- ✔ Lightweight and efficient for small datasets.
- ✘ Struggle with high-dimensional data or complex datasets.
- Methods:
  - PCA [Paffenroth et al.], Isolation Forest [Liu et al.] OC-SVM [Schölkopf et al.]
  - Distance-based: LOF [Breunig et al.], DBSCAN [Ester et al.]
  - Statistical models: ARIMA [Yaacob et al., Cao et al.]



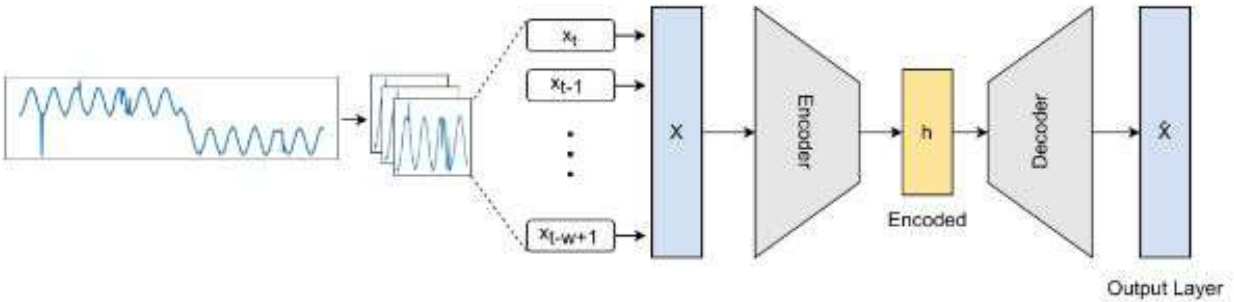
3-2. Background: Anomaly Detection

Unsupervised **Deep Learning** (DL) for AD:

- Handle high-dimensional data with neural networks.

- Require computational resources and lacks interpretability.
- Vulnerable to data contamination.

- Methods
  - Autoencoder-based: USAD [Audibert et al.], LSTM-VAE [Park et al.]
  - One-class: Deep-SVDD [Ruff et al.]



3-3. Background: Root Cause Diagnosis

Detecting an anomaly do not give the reason why it happened

- Root Cause Diagnosis (RCD)

Traditional (expert-based) approaches:

✔ • Rely on rules and domain knowledge.

✘ • Struggle with evolving network conditions
- Require frequent manual updates
- May interfere with the causes (e.g., active probe-based)

- Many techniques:
    - Cellular: [Watanabe et al., Kan et al.]
    - Wi-Fi: [Rayanchu et al., Kanuparthy et al.]

3-3. Background: Root Cause Diagnosis

ML-based techniques

✔ • Automatically learn from data

✘ • Mostly supervised => require labeled data
- Generalizability problem
- May require computational resources

- Many techniques:
  - Cellular: Neural networks to detect faults from KPI data. [Shi et al., 2022; Hasan et al., 2024]
  - Wi-Fi: ML models for impairments detection [Salinas et al., 2018; Syrigos et al., 2019; Salik et al. 2023]

3-3. Background: Conclusion & Research Questions (RQ)

- **Data Collection**
  - RQ: How to collect KPI datasets for LL applications under realistic network conditions ?
- **Anomaly Detection**
  - Unsupervised AD models performance model-dependent.
  - Training data often contaminated.
  - Industrial deployments needs: fast, robust detection.
  - RQ: Can we propose an AD model that outperforms existing solutions and remain efficient under data contamination ?
- **Root Cause Diagnosis**
  - Rule-based methods no longer scale.
  - ML-based RCD relies heavily on labeled anomalies.
  - RQ: Can we design a RCD method efficient with minimal labeled data, well-suited for real-world deployments ?

4. Key Contributions

4. Key Contributions

A. Realistic data collection for Cloud Gaming (CG) and Cloud VR under cellular/WiFi networks

B. CATS: Contrastive learning for Anomaly detection on Time Series

C. RAID: Root-cause Anomaly Identification and Detection

A. Data collection

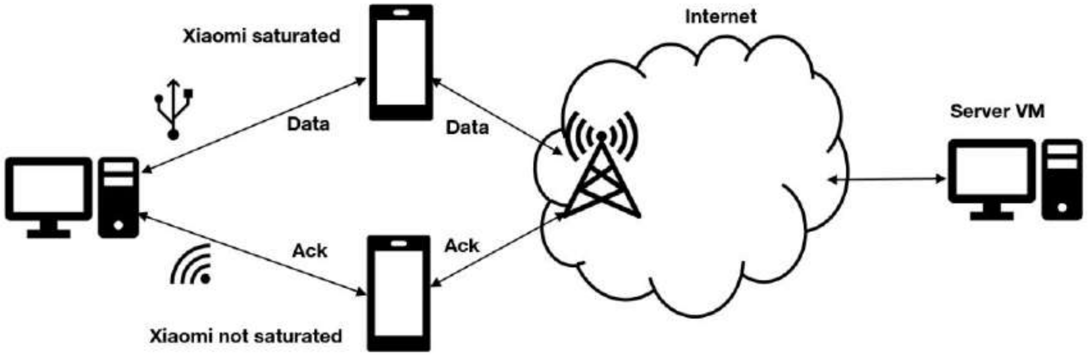
An analysis of Cloud Gaming Platforms Behaviour under Synthetic Network Constraints and Real Cellular Networks Conditions. Xavier Marchal, Philippe Graff, Joël Roman Ky, Thibault Cholez, Stéphane Tuffin, Bertrand Mathieu and Olivier Festor.
Journal of Network and Systems Management, 2023.
OpenData: https://cloud-gaming-traces.lhs.inria.fr/data.html

A-1. 4G network conditions on Orange commercial network

- Use Mahimahi framework to conduct controlled experiments on time-varying network conditions.

- Collect recent transmission opportunities (txops) files representing current 4G network capacities.

- Use Saturatr tool to record 4G/5G base station behavior.



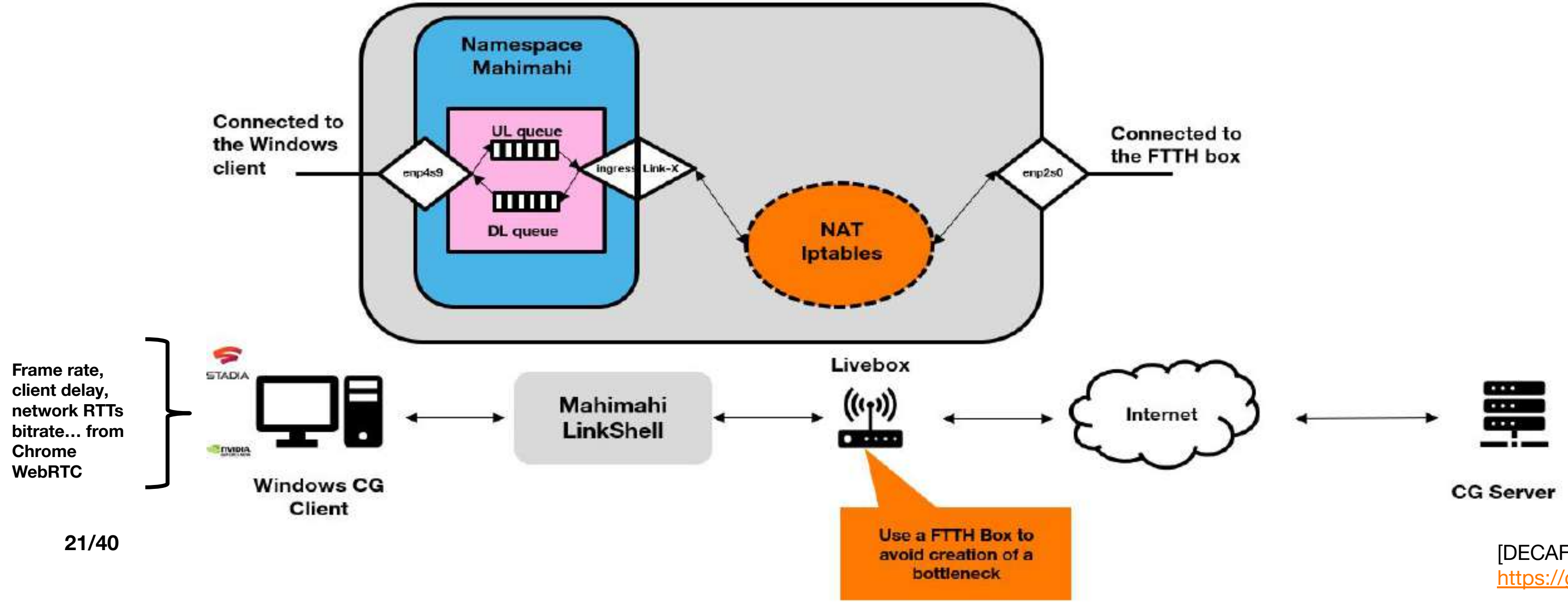
Conditions	Throughput (Mbps)	Location
File 1	220	Orange
File 2	160	Orange
File 3	120	Brélévenez
File 4	80	Brélévenez
File 5	40	Plemeur-Bodou
File 6 (Highway)	45	Guingamp - Lannion

[Mahimahi]: http://mahimahi.mit.edu/

A-2. Cloud Gaming KPIs collection on 4G networks

- Leveraging the previous txops files and commercial CG platforms to collect QoS/QoE KPIs via WebRTC API.

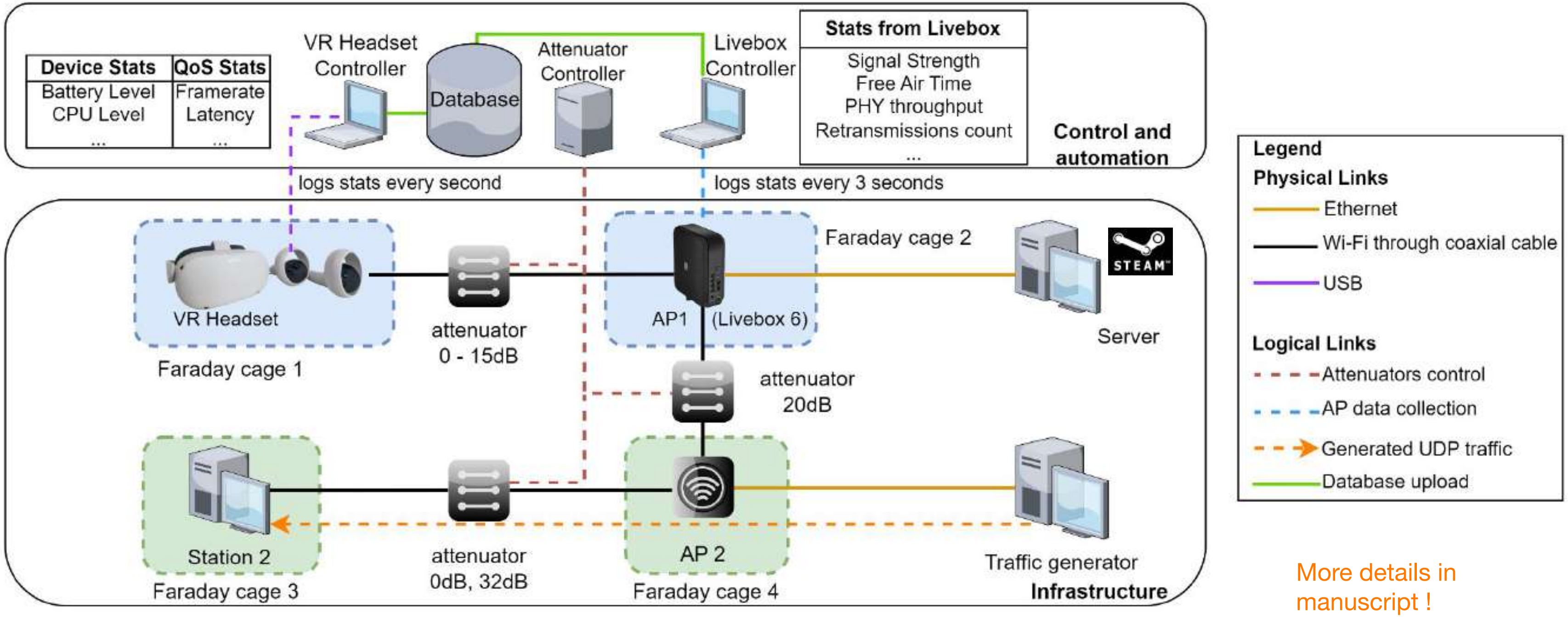
- Use Mahimahi-LinkShell and DECAF tool.



[DECAF]: https://github.com/decafCG/decaf

A-3. Data collection of Cloud VR data over Wi-Fi networks

- CG datasets collected make it challenging to isolate the root causes
- Cloud VR applications are more valuable for ISP like Orange
- Use CloudXR + Oculus tool + Livebox 6



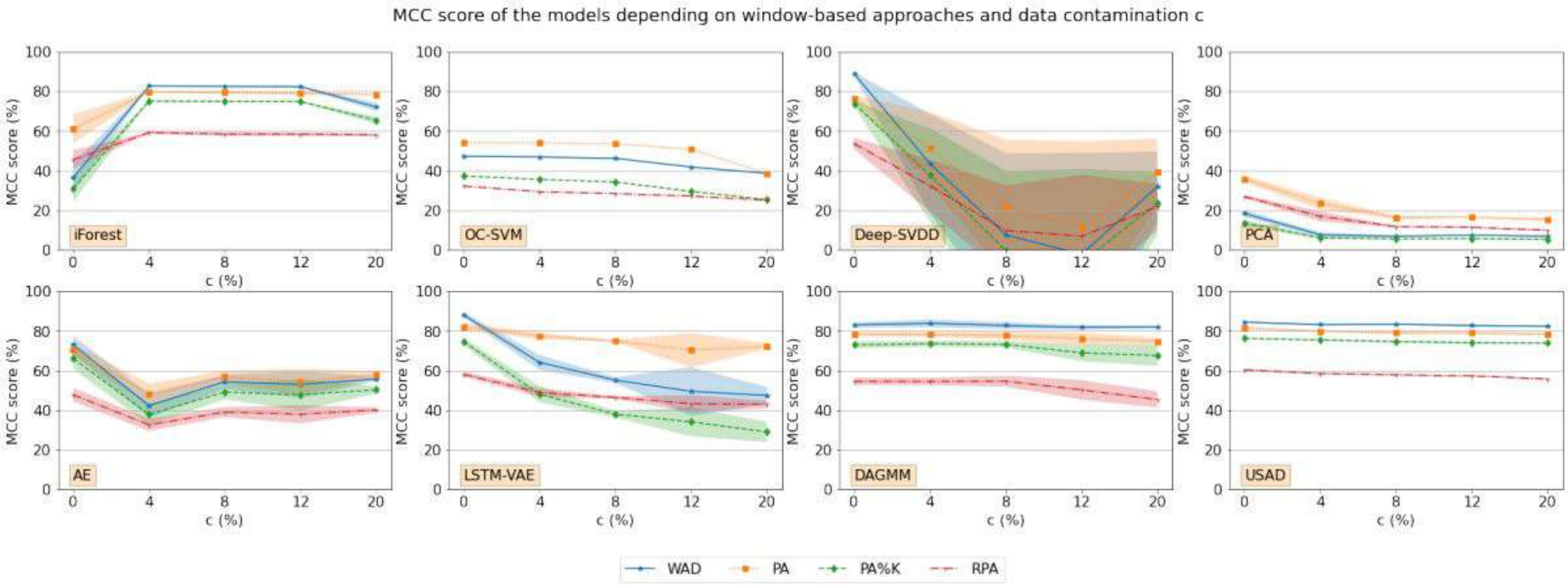
More details in manuscript !

B. CATS: Contrastive learning for Anomaly detection in Time Series

CATS: Contrastive learning for Anomaly detection on Time Series. Joël Roman Ky, Bertrand Mathieu, Abdelkader Lahmadi and Raouf Boutaba.
2024 IEEE International Conference on Big Data (BigData 2024), Washington DC, USA, December 15 - December 18 2024.
Code: https://github.com/joelromanky/cats

B-1. CATS: Motivation

- Existing unsupervised AD models may suffer to discriminate anomalies close to normal samples while being impacted by data contamination.



B-1. CATS: Motivation

- Contrastive learning (CL) gained popularity in many domains and is now applied to time series AD.

- Existing CL-based AD methods can be improved:
    - Do not exploit the temporal aspect of multivariate time series
    - Not robust to data contamination.

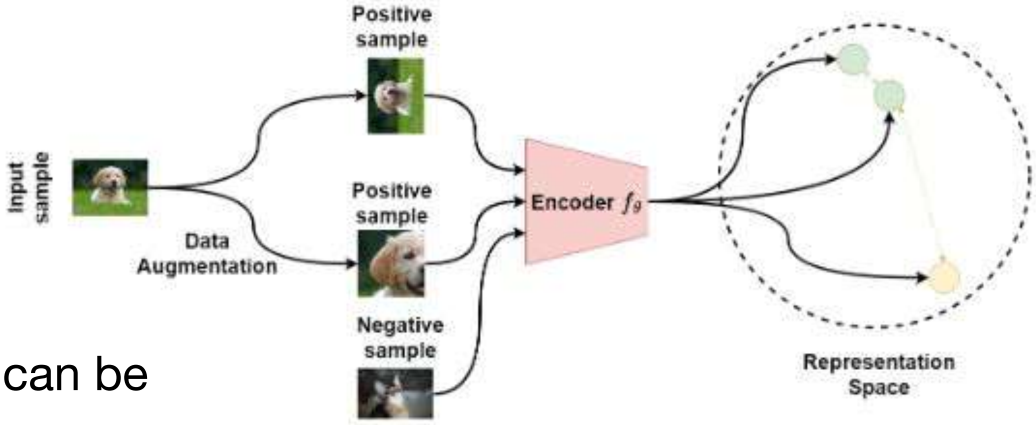
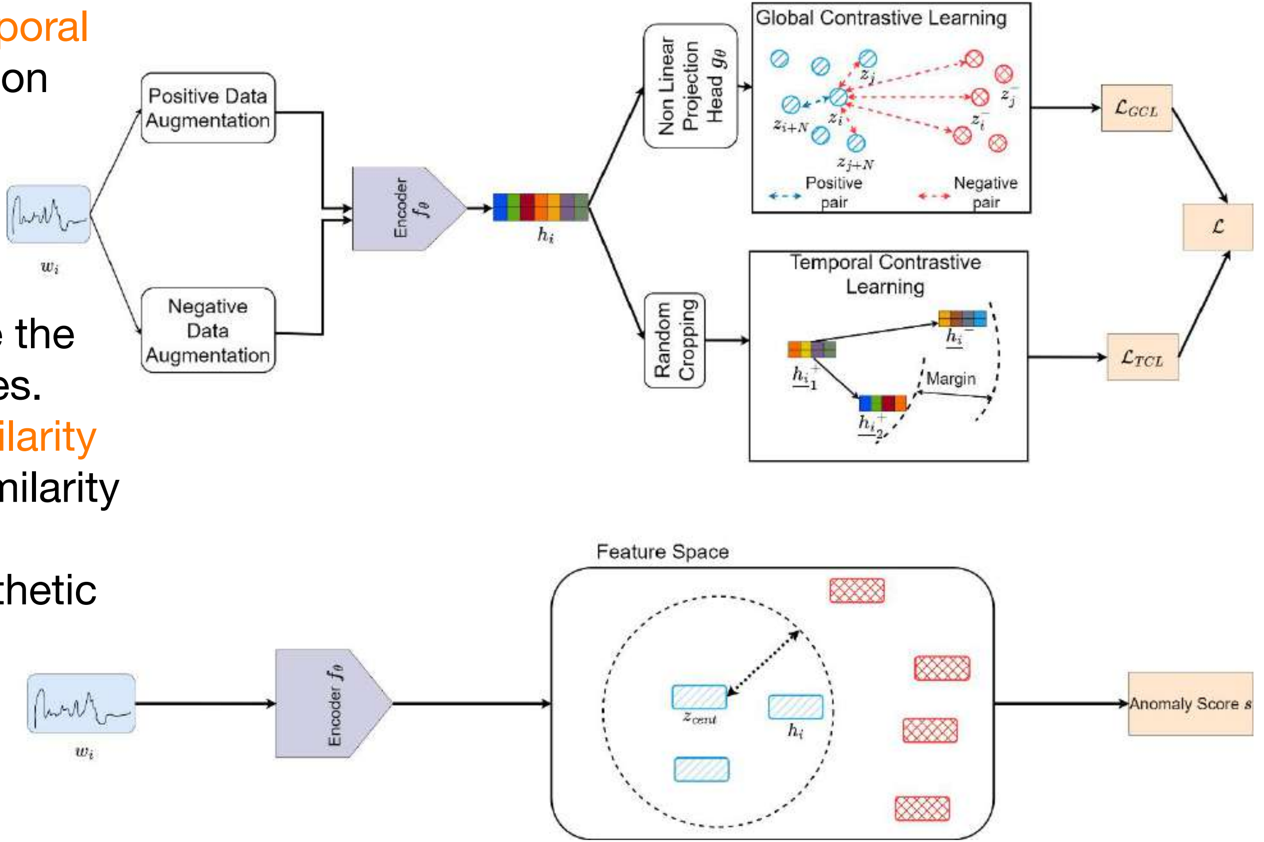


Figure 1.5: Contrastive learning

B-2. CATS: How it works ?

- Idea: Leverage CL with temporal similarity awareness for AD on time series.
- Core techniques:
  - Generate synthetic anomalies to introduce the knowledge of anomalies.
  - Use a DTW-based similarity to enforce temporal similarity (TCL)
  - Improve GCL with synthetic anomalies.



TCL: Temporal Contrastive Learning
GCL: Global Contrastive Learning
DTW: Dynamic Time Warping

B-3. Experiment Results: Performance

- Validated on real-world CG KPIs data previously presented (cf A-2)

Table 4.2: Performance comparison on the datasets. Mean and standard deviation computed over all entities for benchmark datasets and over five runs for case-study datasets. Bold values indicate best results and underlined values the second best results.

	Models	IForest	Deep-SVDD	AE	USAD	SimCLR	SimSiam	TS2Vec	CATS
STD	AUPR	75.16 (±1.28)	95.24 (±0.59)	97.57 (±0.16)	97.55 (±0.04)	97.46 (±0.21)	79.79 (±8.14)	<u>97.65 (±0.99)</u>	**98.72 (±0.09)**
	AUC	74.57 (±1.63)	91.19 (±1.08)	96.04 (±0.27)	<u>96.09 (±0.08)</u>	95.78 (±0.39)	75.65 (±11.3)	95.63 (±1.94)	**97.93 (±0.13)**
	F1	75.79 (±1.42)	87.18 (±1.24)	90.35 (±0.51)	90.02 (±0.24)	90.15 (±0.52)	74.21 (±9.22)	<u>92.83 (±1.92)</u>	**94.06 (±0.45)**
	MCC	39.56 (±3.66)	71.83 (±2.77)	78.93 (±1.14)	77.89 (±0.36)	78.48 (±1.17)	39.31 (±19.8)	<u>84.33 (±4.12)</u>	**86.72 (±0.88)**
GFN	AUPR	76.97 (±0.58)	87.81 (±1.51)	88.60 (±0.40)	88.65 (±0.14)	<u>90.19 (±0.65)</u>	84.49 (±3.33)	89.63 (±1.99)	**93.60 (±0.52)**
	AUC	61.97 (±0.87)	71.78 (±3.41)	74.05 (±0.84)	74.84 (±0.42)	<u>78.50 (±1.95)</u>	67.07 (±3.25)	74.91 (±4.32)	**84.35 (±1.23)**
	F1	74.12 (±0.71)	75.51 (±2.11)	74.05 (±0.84)	77.80 (±0.38)	<u>81.20 (±2.61)</u>	74.25 (±2.93)	76.76 (±2.71)	**82.88 (±0.96)**
	MCC	17.07 (±1.27)	24.26 (±6.56)	28.08 (±0.14)	31.40 (±1.22)	<u>37.46 (±3.87)</u>	17.86 (±6.27)	28.19 (±8.39)	**47.27 (±1.49)**
XC	AUPR	67.19 (±1.61)	61.99 (±7.61)	84.21 (±3.24)	83.34 (±0.31)	80.55 (±2.93)	76.44 (±13.4)	**95.01 (±1.75)**	<u>93.65 (±0.41)</u>
	AUC	78.71 (±1.13)	67.32 (±6.52)	89.18 (±2.31)	89.97 (±0.26)	85.81 (±3.17)	83.35 (±10.6)	**96.96 (±1.36)**	<u>96.10 (±0.41)</u>
	F1	63.33 (±1.18)	50.83 (±7.69)	75.94 (±3.30)	77.59 (±0.58)	70.58 (±3.45)	69.09 (±13.4)	**89.60 (±2.03)**	<u>86.69 (±0.83)</u>
	MCC	43.42 (±2.43)	27.40 (±11.4)	63.95 (±4.63)	65.35 (±0.72)	56.59 (±4.89)	52.30 (±21.3)	**84.07 (±2.94)**	<u>79.67 (±0.11)</u>

- CATS on average outperforms AD models thanks to temporal similarity and anomaly class knowledge.

B-3. Experiment Results: Ablation study

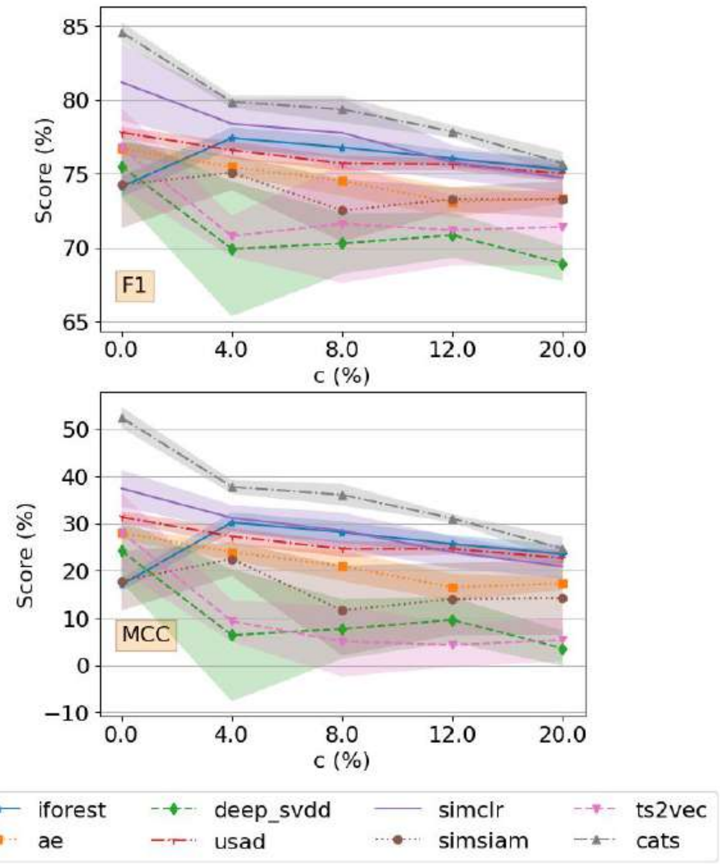
Table 4.3: Ablation study on loss components.

	GFN		XC	
Loss	F1	MCC	F1	MCC
$\mathcal{L}_{NTXent}$	$81.20_{(\pm 2.61)}$	$37.46_{(\pm 3.87)}$	$70.58_{(\pm 3.45)}$	$56.59_{(\pm 4.89)}$
$\mathcal{L}_{GCL}$	$82.52_{(\pm 1.77)}$	$40.73_{(\pm 2.32)}$	$85.68_{(\pm 2.52)}$	$78.31_{(\pm 3.67)}$
$\mathcal{L}_{TCL}$	$79.93_{(\pm 2.69)}$	$38.12_{(\pm 8.32)}$	$76.57_{(\pm 5.68)}$	$65.71_{(\pm 8.34)}$
$\mathcal{L}_{GCL} + \mathcal{L}_{TCL}$	$\mathbf{82.88}_{(\pm 0.96)}$	$\mathbf{47.27}_{(\pm 1.49)}$	$\mathbf{86.69}_{(\pm 0.83)}$	$\mathbf{79.67}_{(\pm 0.11)}$

- Combination of GCL and TCL enhance the performance of CATS.

B-3. Experiment Results: Data contamination robustness

- Even with contaminated training sets, CATS outperforms other models.
- Robustness limited when contamination rate is high.

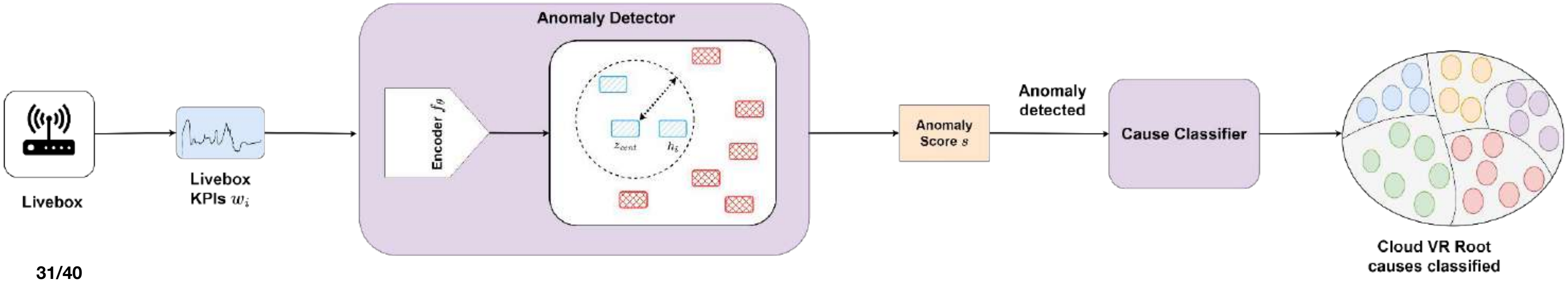


C. RAID: Root cause Anomaly Identification

RAID: Root cause Anomaly Identification. Joël Roman Ky, Bertrand Mathieu, Abdelkader Lahmadi, Minqi Wang, Nicolas Marrot and Raouf Boutaba.
Under review at ECML PKDD 2024.
Code: https://github.com/joelromanky/raid

C-1. RAID: Motivation & Strategy

- ML models for RCD need labeled datasets.
- Contrastive learning improve multivariate time series classification.
- RAID: a two-stage RCD pipeline combining self-supervised and supervised steps:
  - Anomaly detection based on CATS
  - Cause classification using a shallow classifier (SVM).



C-2. Experimental Results: Performance

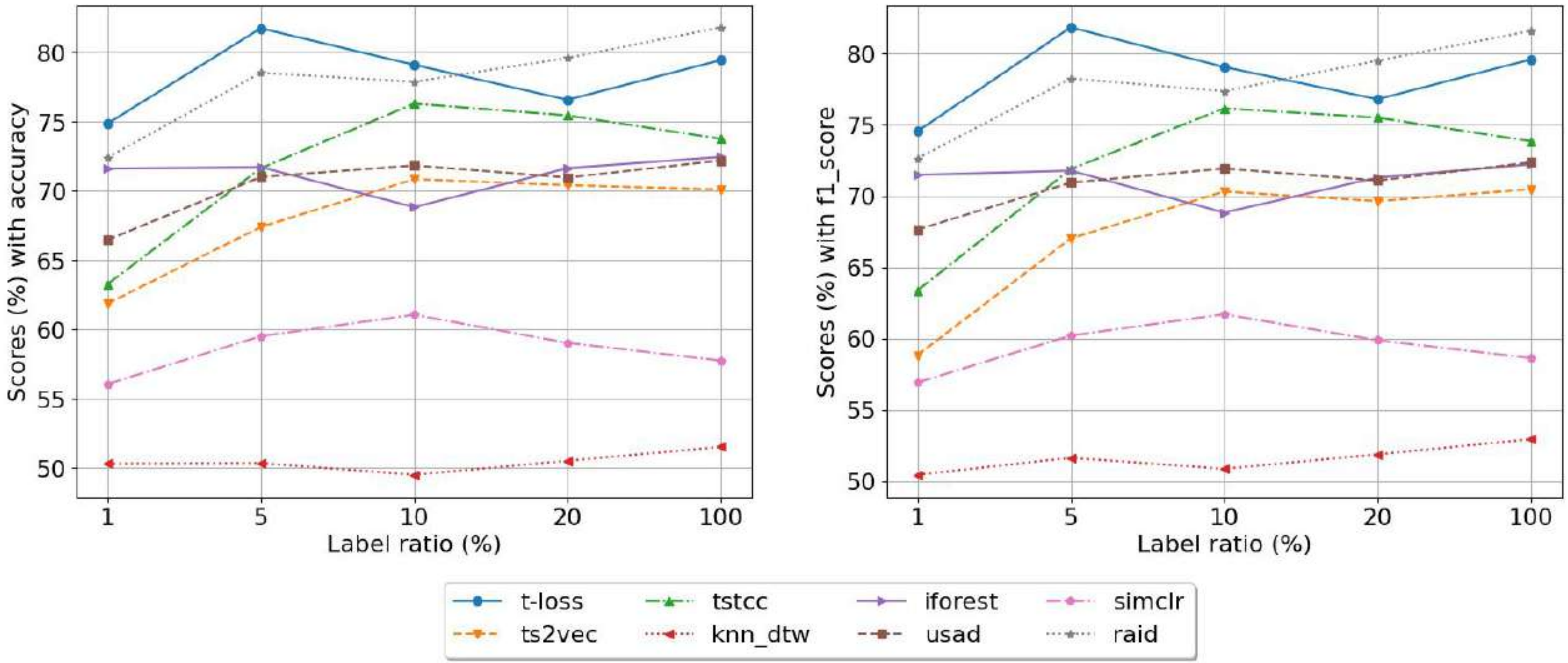
- Using Wi-Fi datasets (cf A-3) RAID is evaluated against:
  - One-stage models (including self-supervised models)
  - Two-stage models.
- RAID achieves the highest performance compared to other approaches.

Table 2. Performance comparison on the datasets. Mean and standard deviation computed over five runs for Cloud VR datasets. Bold values indicate best results and underlined values the second best.

Models	Metrics	Accuracy	N-Accuracy	Precison	Recall	F1-score
One-stage	1-NN-DTW	$51.54_{(\pm 0.11)}$	$26.36_{(\pm 0.17)}$	$56.74_{(\pm 0.09)}$	$51.54_{(\pm 0.11)}$	$52.96_{(\pm 0.10)}$
	T-Loss	$\underline{79.47_{(\pm 4.39)}}$	$\mathbf{75.22_{(\pm 5.58)}}$	$\mathbf{83.98_{(\pm 4.74)}}$	$\underline{79.47_{(\pm 4.39)}}$	$\underline{79.60_{(\pm 4.53)}}$
	TS2Vec	$70.12_{(\pm 5.28)}$	$55.71_{(\pm 6.53)}$	$75.42_{(\pm 3.22)}$	$70.12_{(\pm 5.28)}$	$70.49_{(\pm 4.88)}$
	TS-TCC	$73.78_{(\pm 6.38)}$	$66.17_{(\pm 8.12)}$	$79.29_{(\pm 6.07)}$	$73.78_{(\pm 6.38)}$	$73.86_{(\pm 6.68)}$
Two-stage	iForest	$72.48_{(\pm 2.69)}$	$62.22_{(\pm 4.69)}$	$72.26_{(\pm 3.14)}$	$72.48_{(\pm 2.69)}$	$72.24_{(\pm 2.99)}$
	USAD	$72.22_{(\pm 0.80)}$	$63.39_{(\pm 1.36)}$	$72.72_{(\pm 0.97)}$	$72.22_{(\pm 0.80)}$	$72.38_{(\pm 0.84)}$
	SimCLR	$57.76_{(\pm 3.25)}$	$37.59_{(\pm 4.84)}$	$61.01_{(\pm 2.60)}$	$57.76_{(\pm 3.25)}$	$58.65_{(\pm 3.06)}$
	RAID	$\mathbf{81.83_{(\pm 2.96)}}$	$\underline{74.80_{(\pm 4.19)}}$	$\underline{81.85_{(\pm 3.02)}}$	$\mathbf{81.83_{(\pm 2.96)}}$	$\mathbf{81.60_{(\pm 3.05)}}$

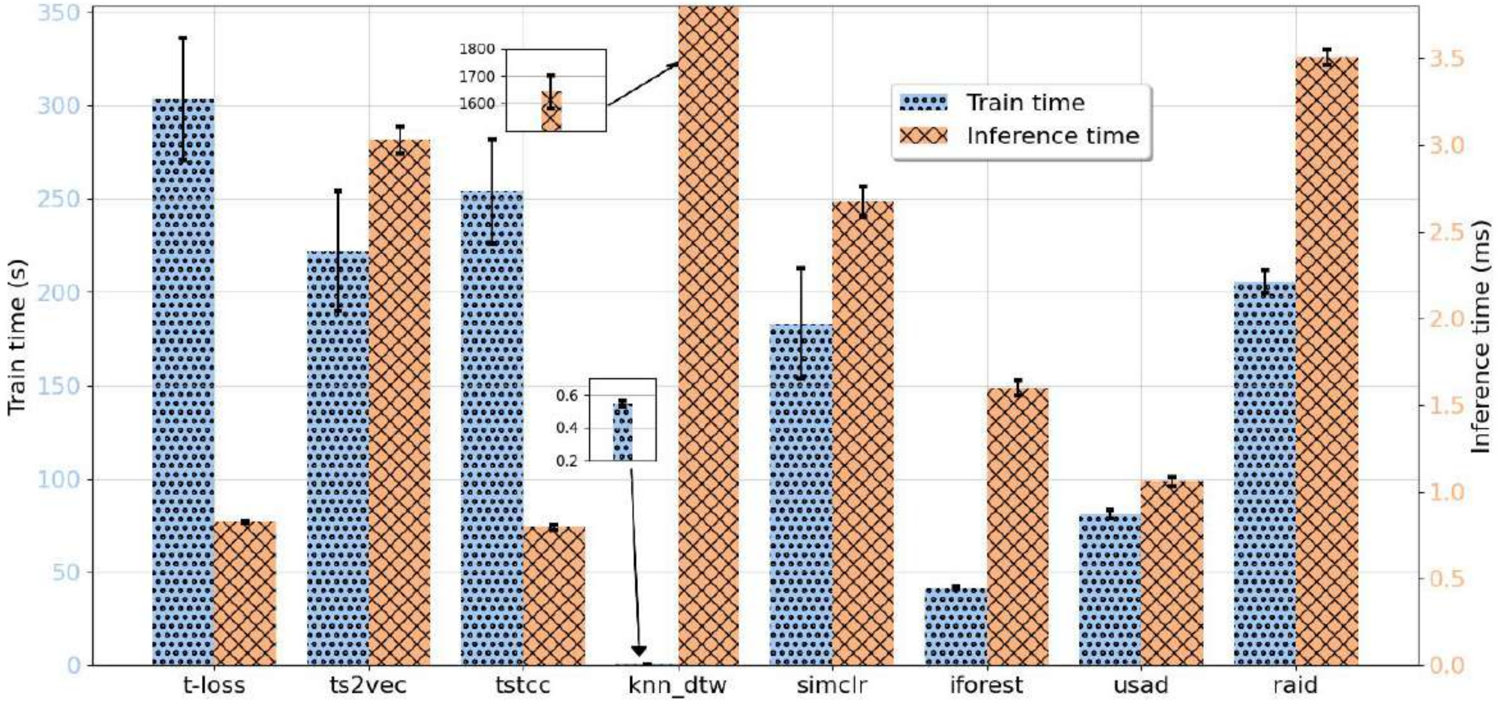
D-2. Experimental Results: Label efficiency

- RAID is performant with limited labeled data but benefit from more labels.



D-2. Experimental Results: Time complexity

- RAID achieves a practical balance between training efficiency and inference speed.
  - Training time around **200s** and inference time of **3.5ms**
  - Only **the classifier needs to be retrained** if new classes of anomalies.



D-3. Comparison RAID vs TLoss

- Better recall for normal scenarios.
- Struggles in discriminating coverage and normal scenarios
- Efficient in detecting interference scenarios (more impact on QoE during experiment).
  - More practical as it avoids more useless countermeasures deployments.

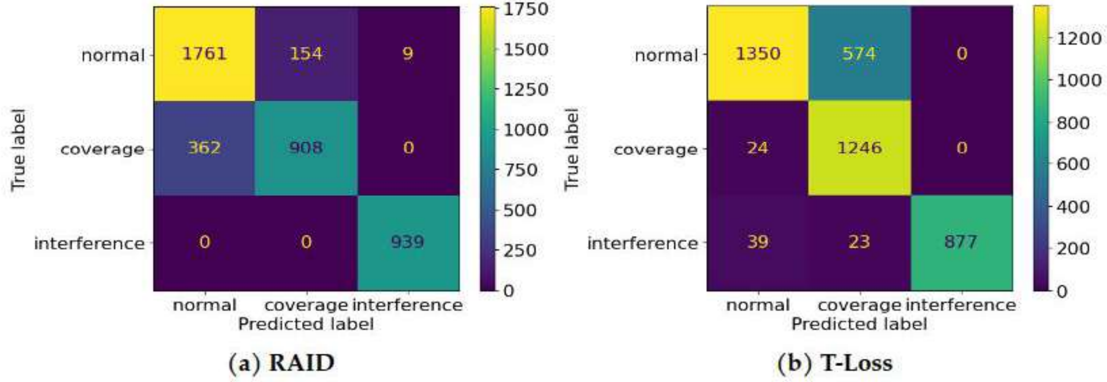


Figure 5.4: Confusion matrix

5. Conclusion & Perspectives

5-1. Conclusion

In this thesis, we make the following contributions:

- **Data Collection**
  - Real-world datasets for QoE of CG applications and labeled Wi-Fi datasets collected over Cloud VR
- **Anomaly Detection on CG (CATS)**
  - Contrastive learning model that exploits temporal structure
  - Robust to data contamination and generalizes well to different LL datasets.
- **Root-Cause Diagnosis for Cloud VR (RAID)**
  - 2-stage pipeline for AD with CATS and cause classification
  - More efficient than two-stage and SSL classification techniques even with few labeled data
  - Reasonable training & inference time => practical for real deployments

5-2. Perspectives

- Improve CATS for AD in time series
  - Increased training time due to the DTW-based loss time complexity $O(N^2)$.
  - Temporal modeling efficiency hindered due to the use of 1 negative in TCL triplet loss.
  - Improve robustness (uncertainty estimation).

- Further data collection for Cloud VR experiments
  - RAID has been tested on a controlled Cloud VR testbed with only two types of impairments.

- Leverage multiple sources of data for RCD
  - Only Wi-Fi metrics used for RCD while our testbed provides much more data sources.

5-2. Perspectives

- Few-shot learning for efficient labeling
- Novel class Discovery
- Causal Discovery

Orange
Innovation

Merci

Bibliography

- [Jansen et al., 2023]: Can my wifi handle the metaverse? a performance evaluation of meta's flagship virtual reality hardware. In Companion of the 2023 ACM/SPEC International Conference on Performance Engineering (pp. 297-303).
- [Marchal et al., 2023]: An analysis of cloud gaming platforms behaviour under synthetic network constraints and real cellular networks conditions. Journal of Network and Systems Management, 31(2), 39.
- [Yaacob et al. 2010]: Arima based network anomaly detection. In 2010 Second International Conference on Communication Software and Networks (pp. 205-209). IEEE.
- [Cao et al., 2016]: One-class classification for anomaly detection with kernel density estimation and genetic programming. In Genetic Programming: 19th European Conference, EuroGP 2016, Porto, Portugal, March 30-April 1, 2016, Proceedings 19 (pp. 3-18). Springer International Publishing.
- [Breunig et al., 2000]: LOF: identifying density-based local outliers. In Proceedings of the 2000 ACM SIGMOD international conference on Management of data (pp. 93-104).
- [Ester et al., 1996]: A density-based algorithm for discovering clusters in large spatial databases with noise. In kdd (Vol. 96, No. 34, pp. 226-231).
- [Paffenroth et al., 2018]: Robust pca for anomaly detection in cyber networks. arXiv preprint arXiv:1801.01571.
- [Liu et al., 2008]: Isolation forest. In 2008 eighth ieee international conference on data mining (pp. 413-422). IEEE.
- [Schölkopf et al., 1999]: Support vector method for novelty detection. Advances in neural information processing systems, 12.
- [Audibert et al., 2020]: Usad: Unsupervised anomaly detection on multivariate time series. In Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining (pp. 3395-3404).
- [Park et al., 2018]: A multimodal anomaly detector for robot-assisted feeding using an lstm-based variational autoencoder. IEEE Robotics and Automation Letters, 3(3), 1544-1551.
- [Wu et al., 2020]: Developing an unsupervised real-time anomaly detection scheme for time series with multi-seasonality. IEEE Transactions on Knowledge and Data Engineering, 34(9), 4147-4160.
- [Ruff et al., 2018]: Deep one-class classification. In International conference on machine learning (pp. 4393-4402). PMLR.
- [Kim et al., 2023]: Contrastive time-series anomaly detection. IEEE Transactions on Knowledge and Data Engineering, 36(10), 5053-5065.
- [Darban et al., 2025]: CARLA: Self-supervised contrastive representation learning for time series anomaly detection. Pattern Recognition, 157, 110874.

Bibliography

- [Watanabe et al., 2008]: UTRAN O&M support system with statistical fault identification and customizable rule sets. In NOMS 2008-2008 IEEE Network Operations and Management Symposium (pp. 560-573). IEEE.
- [Kane et al., 2015]: U.S. Patent No. 9,172,593. Washington, DC: U.S. Patent and Trademark Office.
- [Rayanchu et al., 2008]: Diagnosing wireless packet losses in 802.11: Separating collision from weak signal. In IEEE INFOCOM 2008-The 27th Conference on Computer Communications (pp. 735-743). IEEE.
- [Kanuparthy et al., 2012]: Can user-level probing detect and diagnose common home-wlan pathologies. ACM SIGCOMM Computer Communication Review, 42(1), 7-15.
- [Liu et al., 2019]: KQis-driven QoE anomaly detection and root cause analysis in cellular networks. In 2019 IEEE Globecom Workshops (GC Wkshps) (pp. 1-6). IEEE.
- [Shi et al., 2022]: Towards automatic troubleshooting for user-level performance degradation in cellular services. In Proceedings of the 28th Annual International Conference on Mobile Computing And Networking (pp. 716-728).
- [Hasan et al., 2024]: Root Cause Analysis of Anomalies in 5G RAN Using Graph Neural Network and Transformer. arXiv preprint arXiv:2406.15638.
- [Salinas, 2018]: Home Wi-Fi Impairments (Doctoral dissertation, UPMC (Paris 6)-Sorbonne Université).
- [Syrigos et al., 2019]:On the employment of machine learning techniques for troubleshooting WiFi networks. In 2019 16th IEEE Annual Consumer Communications & Networking Conference (CCNC) (pp. 1-6). IEEE.
- [Salik et al., 2023]: Non-WiFi interference detection and throughput estimation at the WiFi edge for 2.4 and 5 GHz bands with machine learning. In 2023 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom) (pp. 371-378). IEEE.
- [Bauskar, 2024]: Enhancing System Observability with Machine Learning Techniques for Anomaly Detection. 14. 64-70.

Publications

- **Root cause diagnosis in Cloud VR applications over Wi-Fi networks,** J. Ky, B. Mathieu, A. Lahmadi, M. Wang, N. Marrot, and R. Boutaba. *Under review at ECML-PKDD 2025.*
  - **Code: https://github.com/joelromanky/raid**

- **CATS: Contrastive learning for Anomaly detection in Time Series,** J. Ky, B. Mathieu, A. Lahmadi, and R. Boutaba. *IEEE International Conference on Big Data (Big Data), IEEE.,* Washington DC, USA, December 15-18, 2024
  - **Code: https://github.com/joelromanky/cats**

- **Segment Routing for Chaining Micro-Services at Different Programmable Network Levels,** B. Mathieu, O. Dugeon, J. R. Ky, P. Graff, and T. Cholez. *27th Conference on Innovation in Clouds, Internet and Networks (ICIN), IEEE,* Paris, France, March 11-14, 2024.

- **ML models for detecting QoE degradation in low-latency applications: a cloud-gaming case study,** J. Ky, B. Mathieu, A. Lahmadi, and R. Boutaba. *IEEE Transactions on Network and Service Management,* 2023.
  - **Code: https://github.com/joelromanky/unsupervised-ml-ad-qoe-deg**

Publications

- **A hybrid P4/NFV architecture for cloud gaming traffic detection with unsupervised ML,** J. R. Ky, P. Graff, B. Mathieu, and T. Cholez. *IEEE Symposium on Computers and Communications (ISCC), IEEE,* Gammarth, Tunisia, July 9-12, 2023.

- **An analysis of cloud gaming platforms behaviour under synthetic network constraints and real cellular networks conditions,** X. Marchal, P. Graff, J. R. Ky, T. Cholez, S. Tuffin, B. Mathieu, & O. Festor. *Journal of Network and Systems Management,* 2023.
    - **OpenData: https://cloud-gaming-traces.lhs.inria.fr/data.html**

- **Assessing unsupervised machine learning solutions for anomaly detection in cloud gaming sessions,** J. Ky, B. Mathieu, A. Lahmadi, and R. Boutaba. *Workshop on High-Precision, Predictable, and Low-Latency Networking (HiPNet '22), colocated with 18th International Conference on Network and Service Management (CNSM), IEEE,* Thessaloniki, Greece, October 31 - November 4, 2022.
    - **Code: https://github.com/joelromanky/cg-ano-detect-eval**

- **Characterization and troubleshooting of CG applications on mobile network,** J. R. Ky, B. Mathieu, A. Lahmadi, R. Boutaba. *Poster presentation at Network Trafic Measurement and Analysis Conference (TMA 2022),* Jun 2022, Enschede, Netherlands.

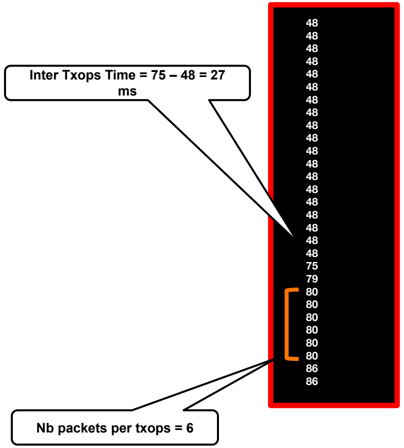
Appendixs

Collection of 4G network conditions on Orange commercial network

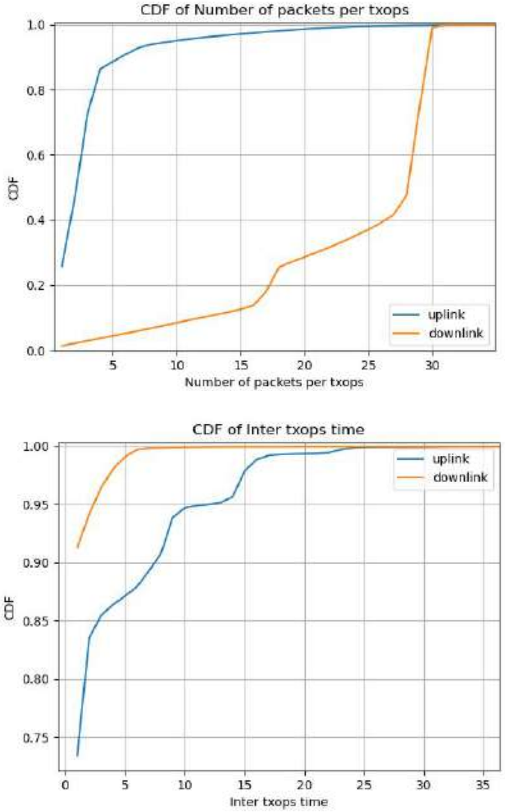
Txops Generation

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

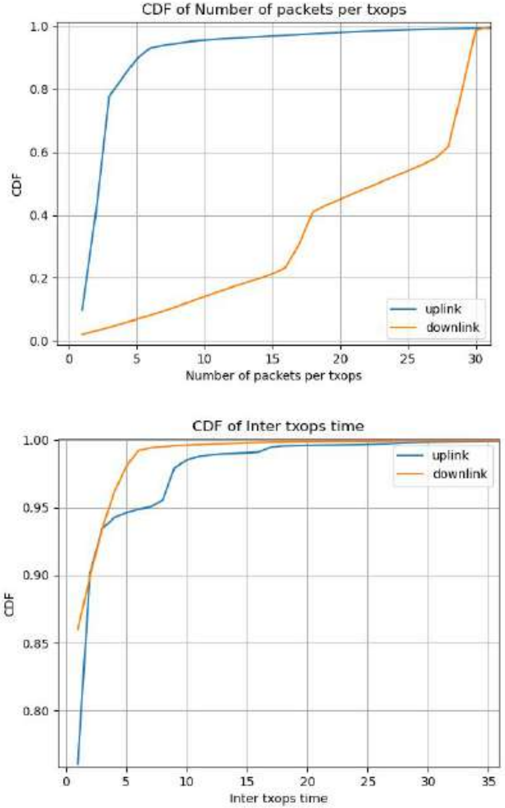
Measurements conditions



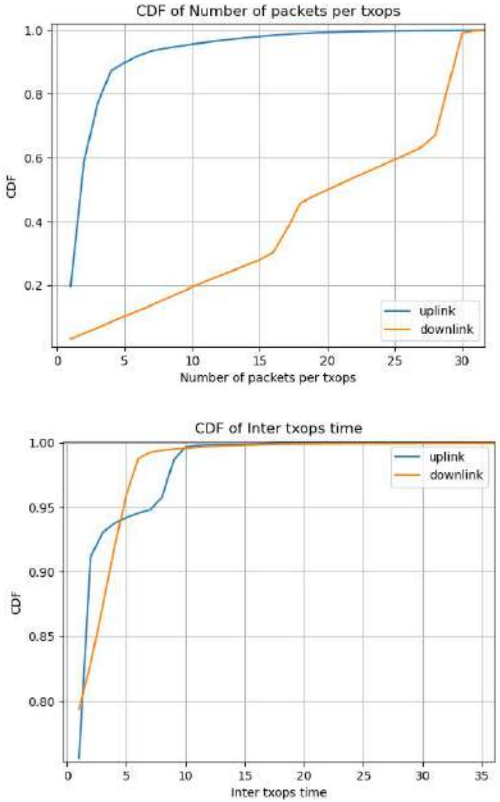
Characterization of 4G txops measured



File 1

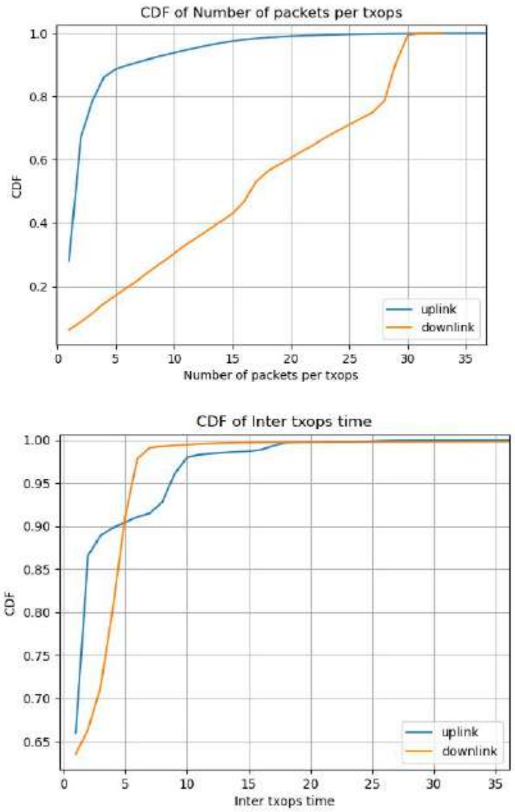


File 2

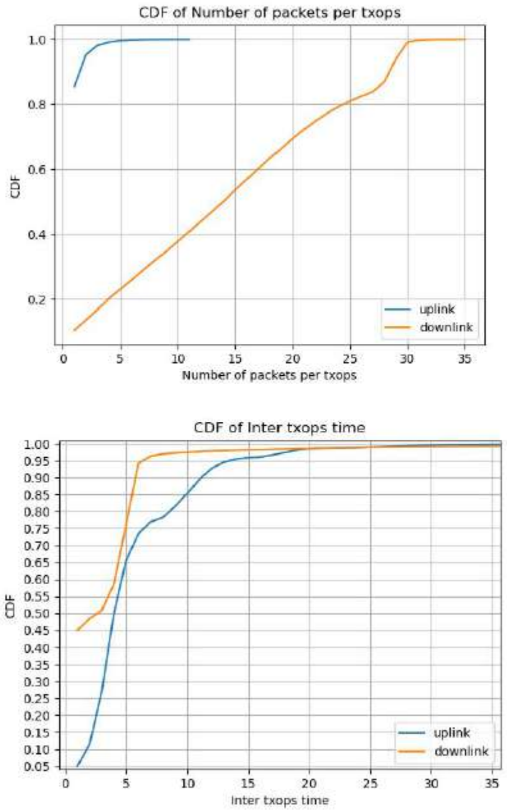


File 3

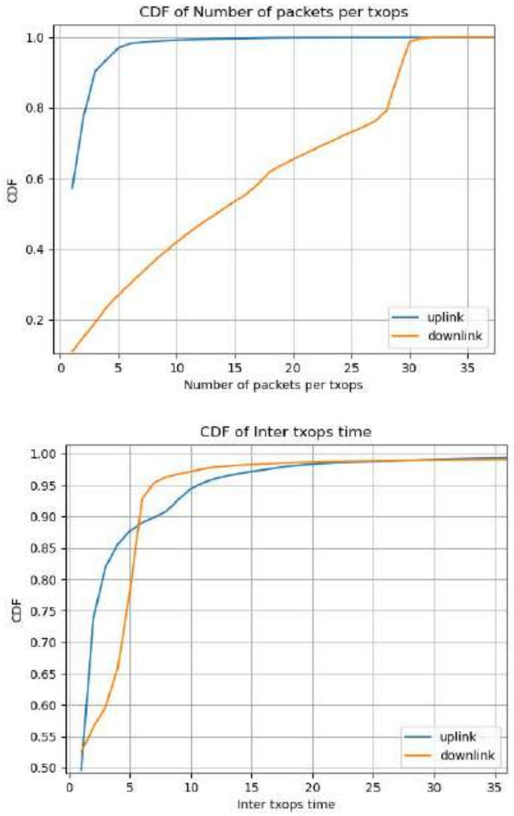
Characterization of 4G txops measured



File 4

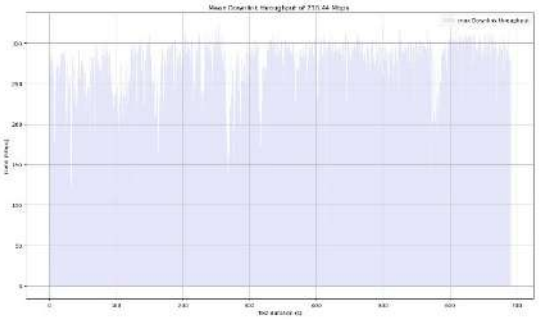


File 5

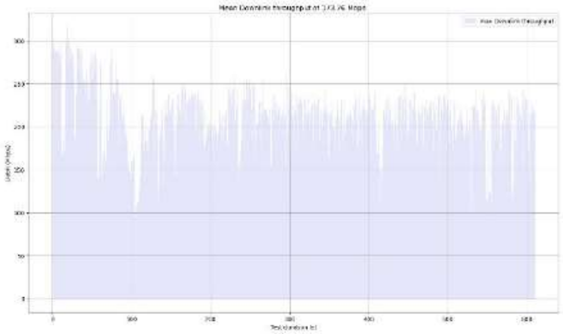


File 6

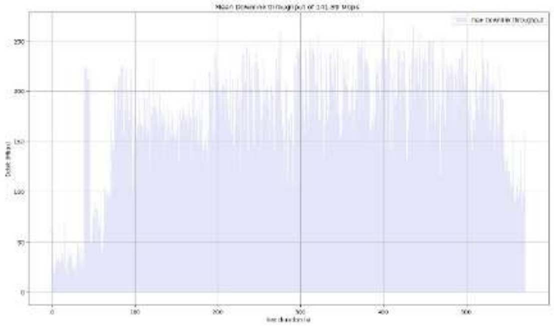
Max downlink throughput on the txops files



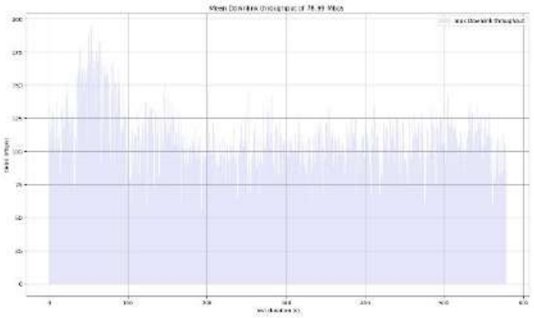
File 1



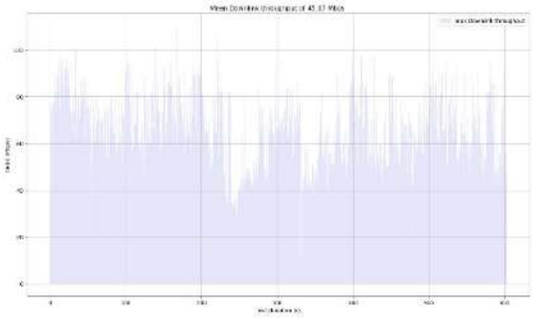
File 2



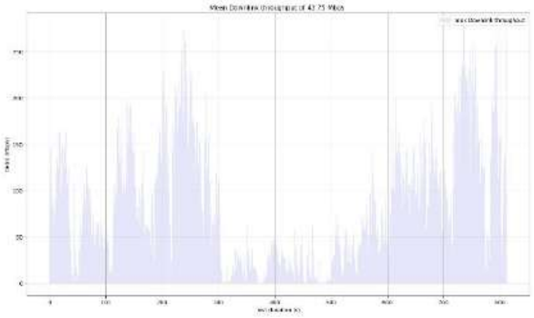
File 3



File 4



File 5



File 6

Cloud Gaming KPIs collection on 4G networks

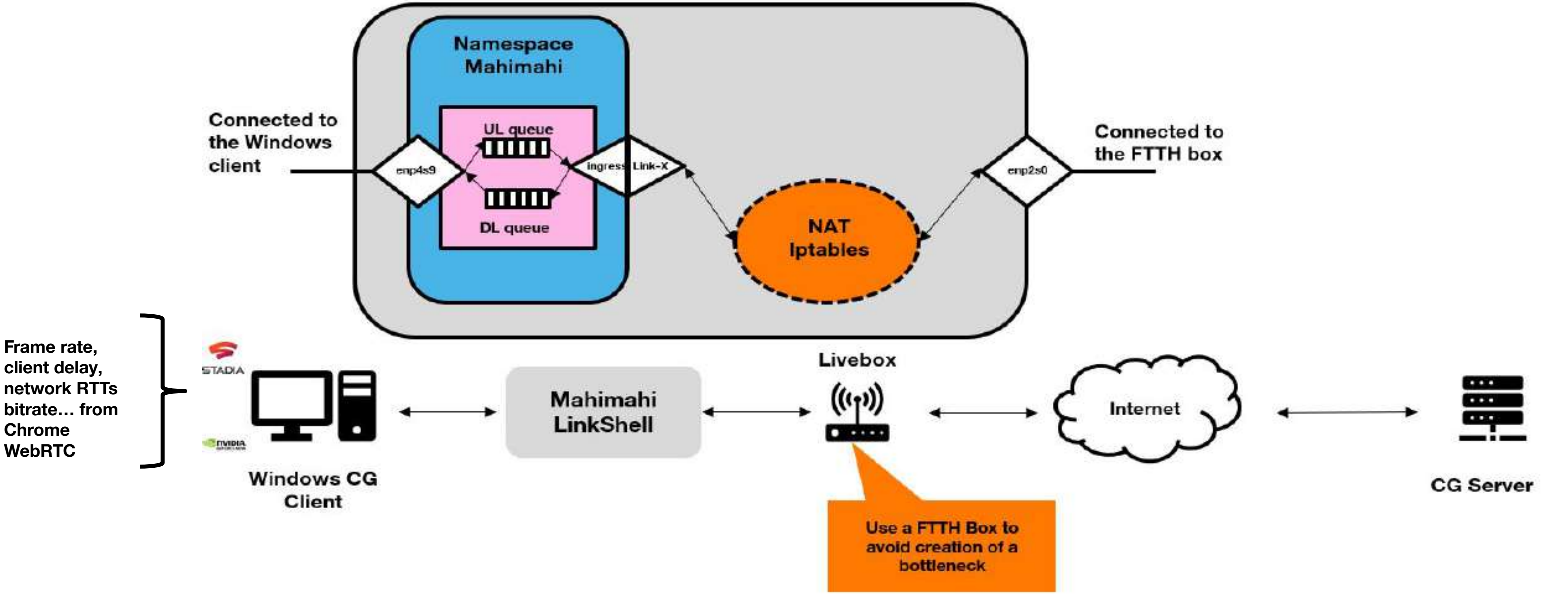


TABLE I: Datasets summary.

	Dataset	Train	Test	Dimensions	Anomalies (%)
Cloud Gaming	STD	80486	169706	14	52.57
	GFN	27415	22667	14	55.36
	XC	83611	17918	14	24.32

Data collection of Cloud VR data over Wi-Fi networks

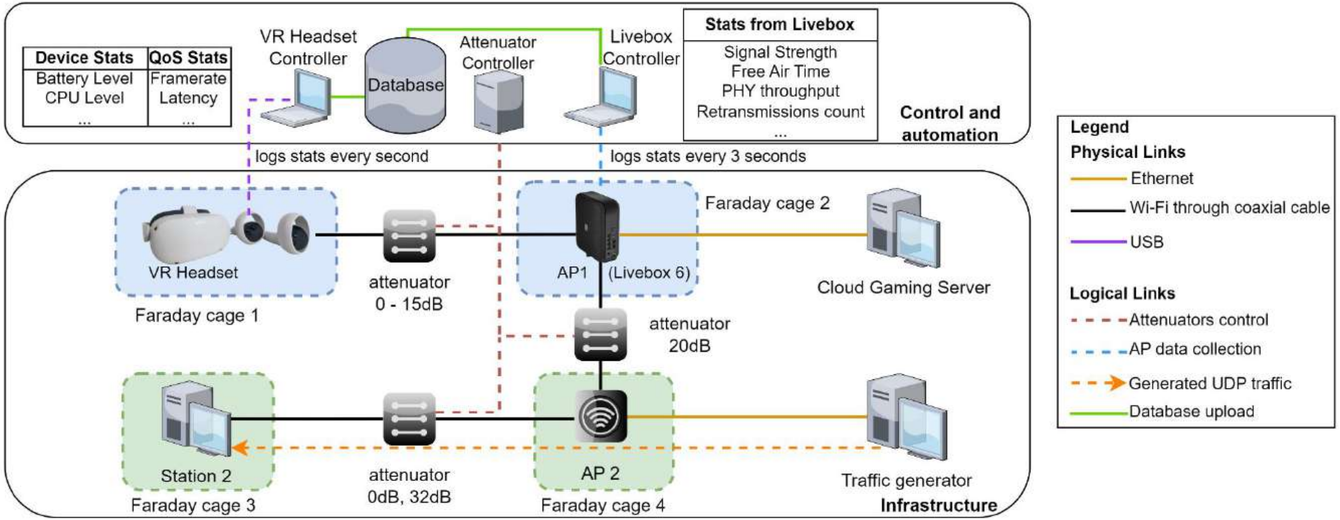


Table 5.1: Dataset Summary

Class	Train Size	Test Size	Number of Features	Number of Time Steps
Normal	4718	1924	112	10
Coverage	2984	1270	112	10
Interference	1822	939	112	10
Overall	9524	4133	112	10

3-1. Motivation

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

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

- Anomalies occur in the form of *windows* and metrics to better evaluate performance of AD models are inacurrate.

3-2. Contributions

- **Demonstrate on synthetic models that existing window metrics wrongly estimate AD models performance and propose WAD (Window Anomaly Decision) approach.**

- **Exhaustive evaluation of 8 unsupervised ML models with real-world datasets collected on 3 commercial CG platforms servers under 6 different 4G emulated network conditions.**
  - **Study data contamination and window size impact on AD models.**

- **Recommendation to network management experts on best models regarding different industrial requirements.**

Unsupervised ML models for anomaly detection

We compare several unsupervised ML models on their performance, robustness and time complexity on AD on the CG time series datasets.

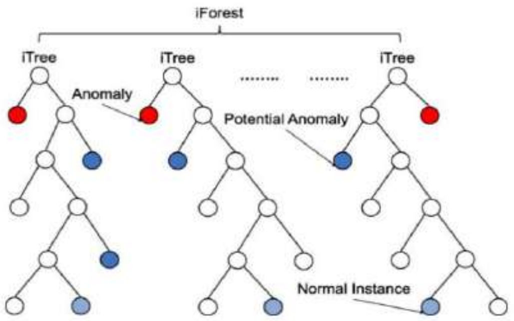
- **Isolation based models**
  - **iForest**
- **One classification based models**
  - **OC-SVM**
  - **Deep-SVDD**
- **Reconstruction-based models**
  - **PCA**
  - **Auto Encoder**
  - **LSTM-VAE**
  - **DAGMM**
  - **USAD**

Performance evaluation metrics :

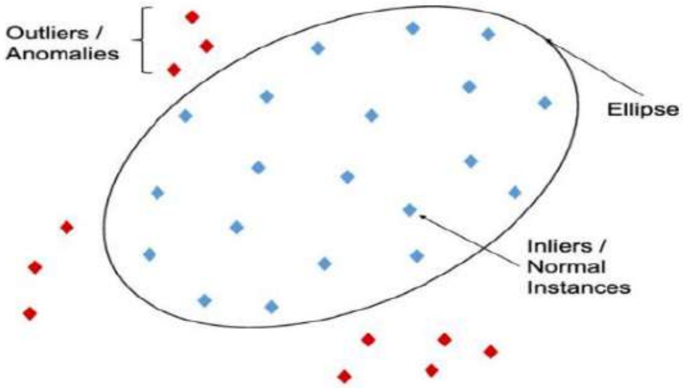
- **Precision:** $P = \frac{TP}{TP+FP}$
- **Recall:** $R = \frac{TP}{TP+FN}$
- **F1-Score:** $F1 = 2\frac{P\,.\,R}{P+R}$
- **MCC:** $\mathrm{MCC} = \frac{TP.TN - FP.FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$

Unsupervised ML models: Isolation-based models

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



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



- **Deep-SVDD:** Deep-learning implementation of OC-SVM that benefits from DL efficiency on high-dimensional data.

A) Unsupervised ML models : Reconstruction approaches

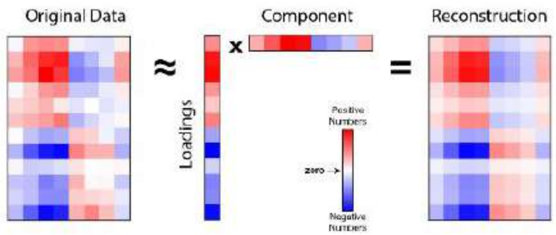
- **PCA:** Reconstruction of the data with principal components.

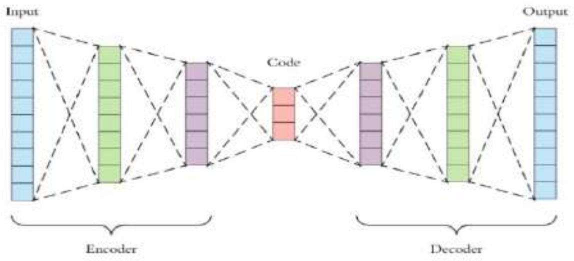
- **AutoEncoder (AE):** Constitued of an encoder, that learns from inputs a low-dimensional representation of data, and a decoder that reconstruct original data from latent variable.

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

- **DAGMM**: Combination of AE and a gaussian mixture model.

- **USAD:** Two AE adversely trained and sharing the same encoder to reconstruct and discriminate for better representation learning.





CG 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} $X_{highway}$

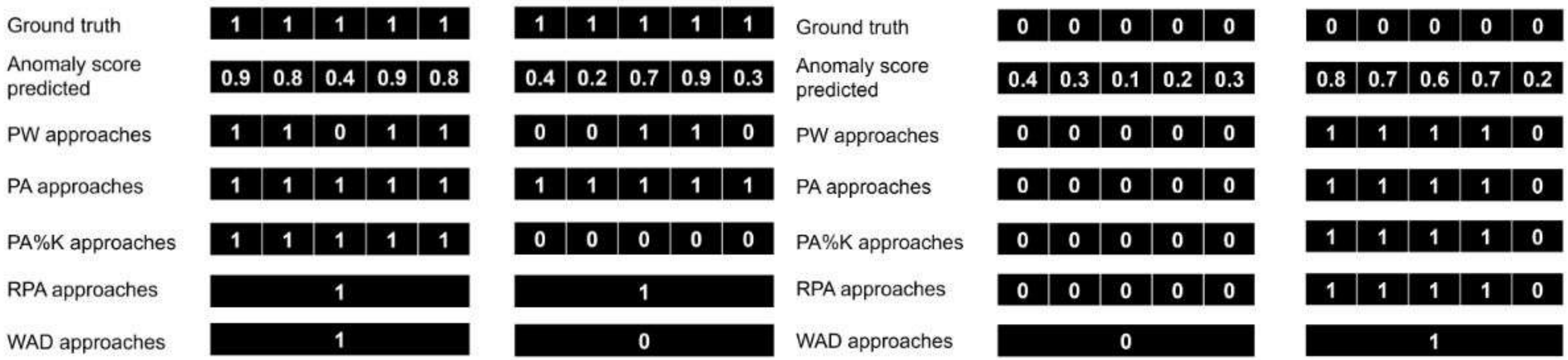
14 QoS/QoE features with a time-step of 5ms :

- 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.

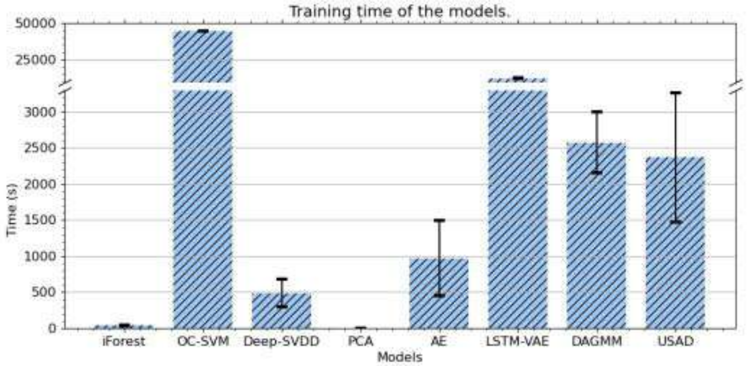
Window approaches



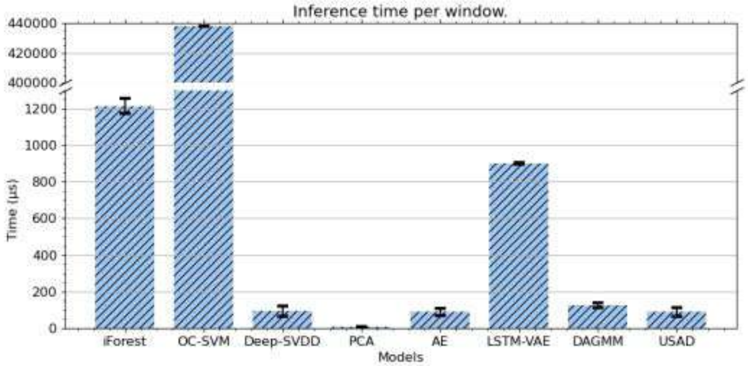
(a) Prediction on anomaly windows (b) Prediction on normal windows

Figure 3.1: Illustration of PW, PA, RPA and WAD approaches. 0 is normal and is 1 anomalous. The anomaly score threshold to decide if an observation is anomalous or not is $\delta = 0.5$. Window size $p = 5$, $\alpha = 0.8$.

Unsupervised ML models for anomaly detection



(a) Training time



(b) Inference time per window

Fig. 5. Train and inference time.

Unsupervised ML models for anomaly detection

TABLE VI
ML MODELS RECOMMENDATION

Model	Performance	Robustness	Deployment	Explainability
IForest	+	++	-	-
OC-SVM	-	+	- -	- -
Deep-SVDD	++	- -	++	- -
PCA	- -	++	++	++
AE	+	-	++	- -
LSTM-VAE	++	- -	-	- -
DAGMM	++	++	+	- -
USAD	++	++	+	- -

++: good; +: somewhat good; -: somewhat bad; - -: bad.

3-8. Conclusion

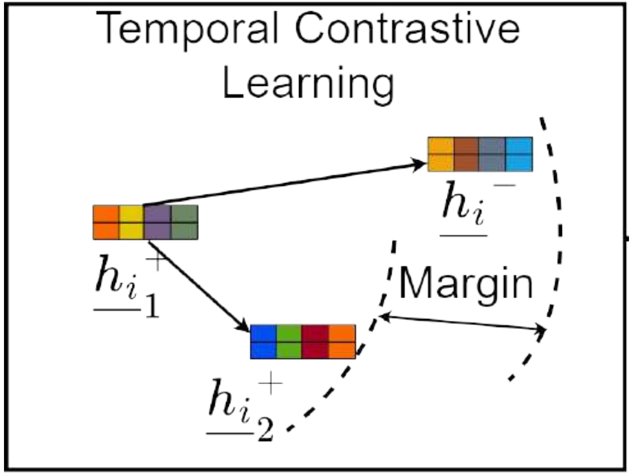
- F1-score has some limitations and should be coupled with MCC metric to avoid erroneous conclusions on model performance.

- Data contamination has different impact on unsupervised ML models.
    - Isolation-based benefit from it until a certain level
    - One class and reconstruction-based see their performance degrade (except DAGMM and iForest)

- ML models usually do not necessarily meet industrial considerations such as robustness, performance, explainability, energy consumption…
    - Future work will consist in ML models for low-latency applications anomaly detection from PCAP files.

Temporal Contrastive Learning (TCL)

- **Dynamic Time Warping (DTW): a similarity measure between time series that seeks for the temporal alignment that minimizes Euclidean distance between aligned series.**
  - **However, DTW is not differentiable.**
  - **Soft-DTW was introduced using the soft-min operator to make DTW differentiable.**



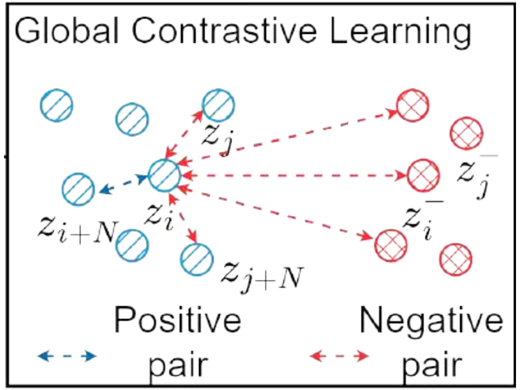
- **TCL learns a temporal representation using a triplet loss with Soft-DTW and is defined as follows:**

$$L_{TCL} = \sum_{i=1}^{N} \max(d(h_i, h_i^+) - d(h_i, h_i^-) + m, 0)$$

$$d(h_i, h_j) = softDTW(h_i, h_j) - \frac{1}{2}(softDTW(h_i, h_i) + softDTW(h_j, h_j))$$

Global Contrastive Loss (GCL)

- GCL learn representations at the instance level using the NT-Xent loss while considering more negative pairs.
  - NT-Xent loss consider two views of same instance as positive and view of different instances as negative.
  - GCL also include the views generated through negative data augmentation.
  - Consequently, instead of contrasting one positive pair and N-1 negative pairs in NT-Xent =, GCL contrasts one pair and 2N-1 negative pairs.



$$L_{GCL} = -\frac{1}{2N} \sum_{i \in B_a \cup B^+} \log \frac{\exp(sim(z_i, z_i^+)/\tau)}{\sum_{j \in B \text{ and } j \neq i} \exp(sim(z_i, z_j)/\tau)}$$

$$sim(z_i, z_j) = \frac{z_i^T z_j}{\|z_j\| \|z_i\|}$$

$$B = \{B_a, B^+, B^-\}$$

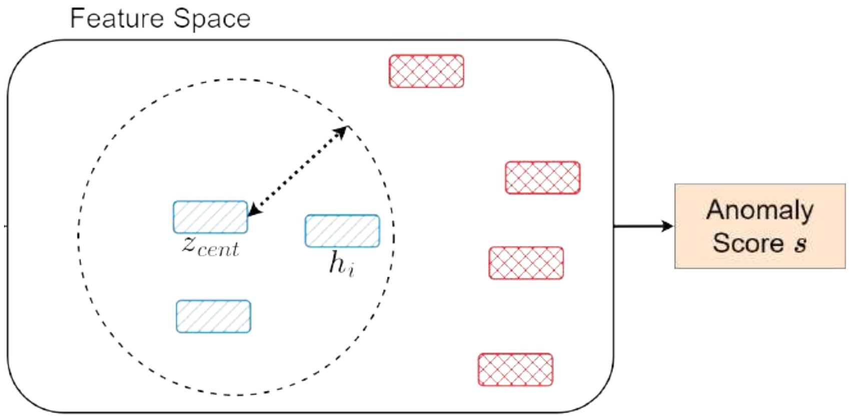
Anomaly score

- After training, we assume that the encoder has learn sufficient information to be efficient for our downstream task (AD).

- Anomaly can be identified using a simple anomaly score computed as follows:

$$s(w_t) = D(f_\theta(w_t), z_{cent})$$

$$z_{cent} = \frac{1}{N_{train}} \sum h_i$$



CATS: Data augmentation impact

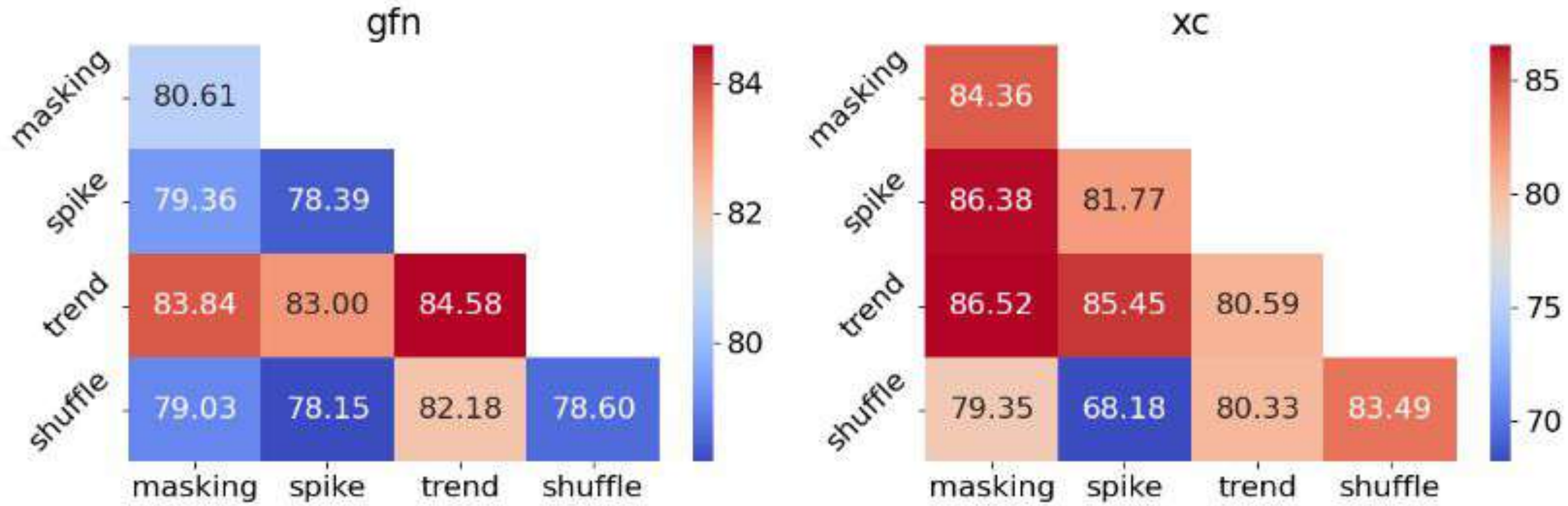
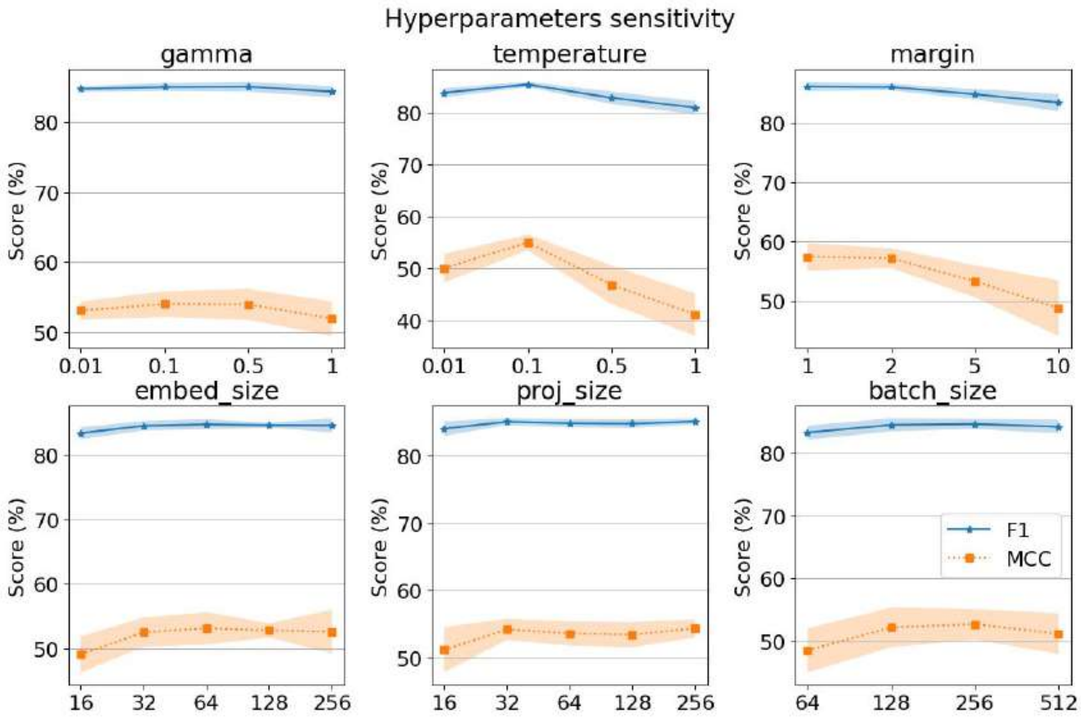
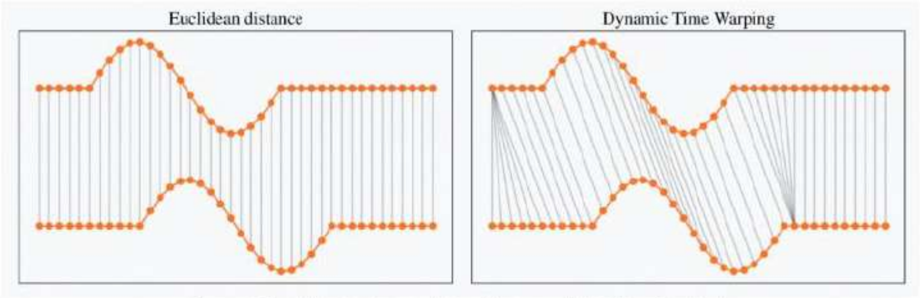


Figure 4.2: Ablation of negative data augmentations using F1-score for GFN and XC datasets.

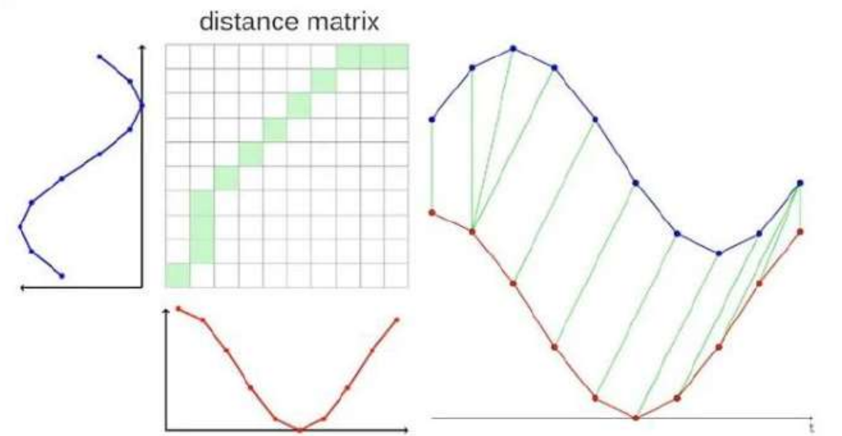
CATS: Hyper-parameters sensitivity



Dynamic Time Warping



Dynamic Time Warping (source : https://rtavenar.github.io/blog/dtw.html)



$$DTW_q(x, x') = \min_{\pi \in \mathcal{A}(x,x')} \left(\sum_{(i,j)\in\pi} d(x_i, x'_j)^q \right)^{\frac{1}{q}}$$

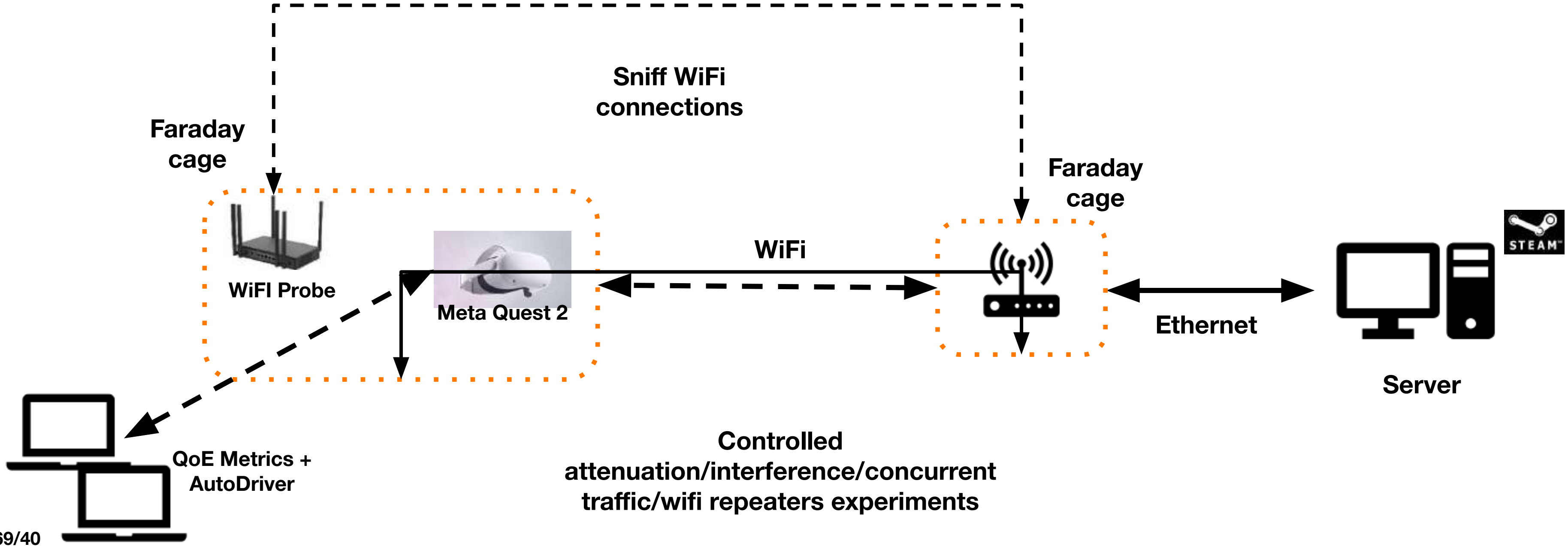
$$\text{soft-}DTW^{\gamma}(x, x') = \min_{\pi \in \mathcal{A}(x,x')}{}^{\gamma} \sum_{(i,j)\in\pi} d(x_i, x'_j)^2$$

$$\min{}^{\gamma}(a_1, \ldots, a_n) = -\gamma \log \sum_i e^{-a_i/\gamma}$$

4. Conclusion

- CATS addresses the limitations of traditional CL with temporal similarity and negative data augmentation.
- Empirical evaluations demonstrate performance in AD tasks on different datasets while being robust to data contamination.
- Some limitations remain:
  - Increased training time due to the SoftDTW time complexity $O(N^2)$
  - Triplet loss in TCL hinders the efficiency of temporal modeling due to the use of 1 negative.

Root-cause analysis on Cloud VR



6. Root-cause analysis on Cloud VR

- Generate different WiFi degradation scenarios:
  - Signal attenuation
    - 2.4 GHz: between -45dB and -70dB
    - 5 GHz: between -65dB and -90dB
  - Interference: using a neighboring LAN, we generate a traffic to reduce the txops on the main LAN
    - Leave 9-15% of txops on the main AP

RAID: Anomaly detector

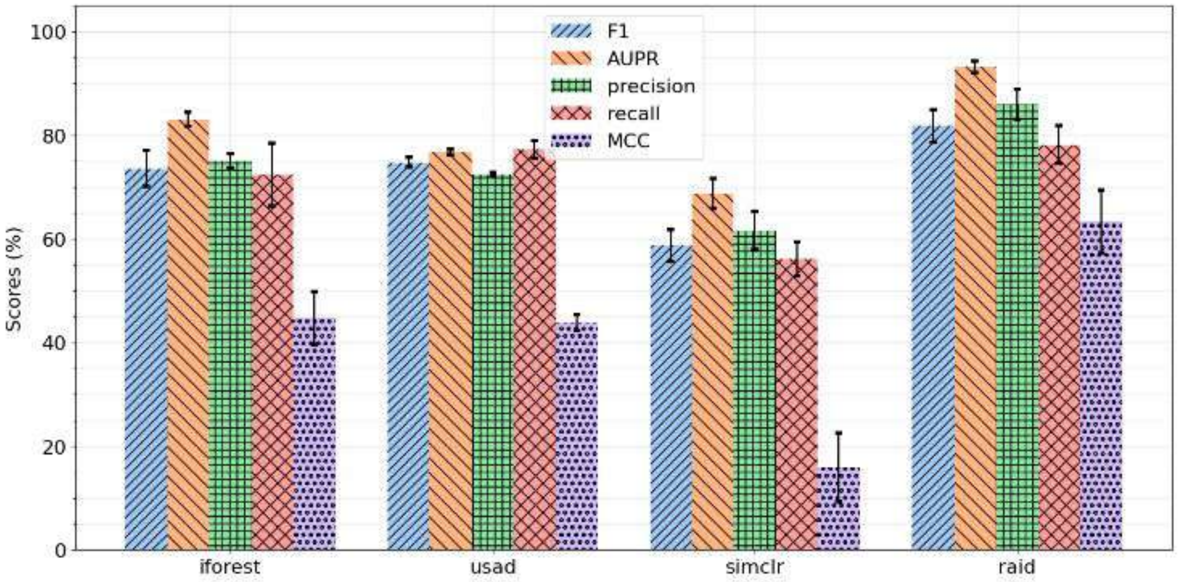
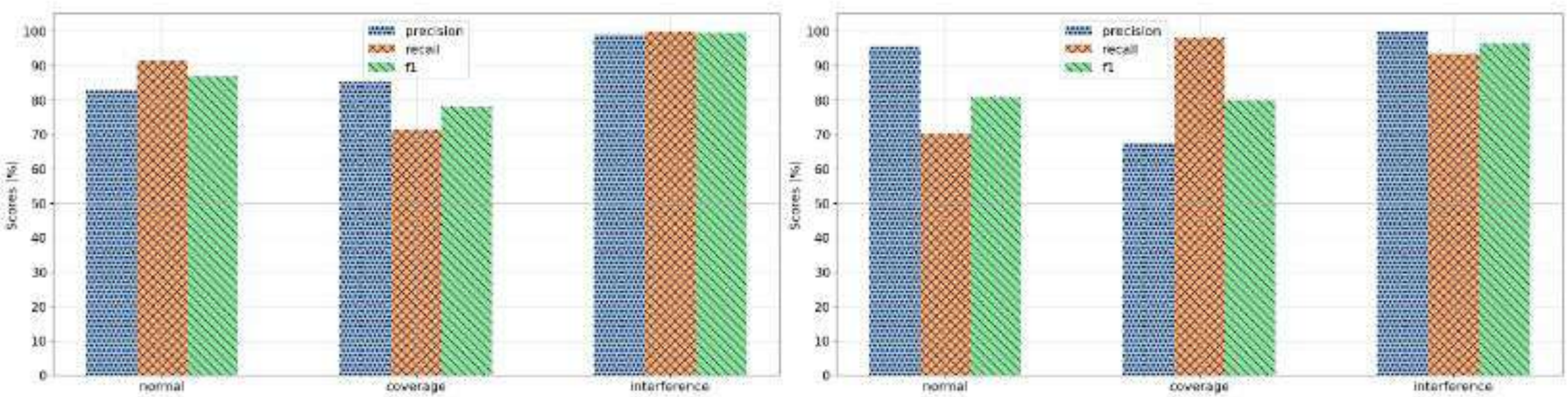


Figure 5.3: Results of anomaly detectors of two-stage models.

RAID: Per-class accuracy



(a) RAID (b) T-Loss

Figure 5.5: Per-class precision, recall and F1-score.

8. Highlights

- Public Prize at Orange challenge « Ma thèse en 3 minutes » at Orange Open Tech Days, November 2023, Châtillon, France
- Demo « Low-latency made easy » as part of ANR MOSAICO project at Orange Open Tech Days, November 2023, Châtillon, France.
- Awards at NeurIPS’22 competition track « Cross-Domain MetaDL »
- Best poster presentation award at 10th TMA PhD school colocated with TMA conference.