MDM 2026 · 27TH IEEE INTERNATIONAL CONFERENCE ON MOBILE DATA MANAGEMENT

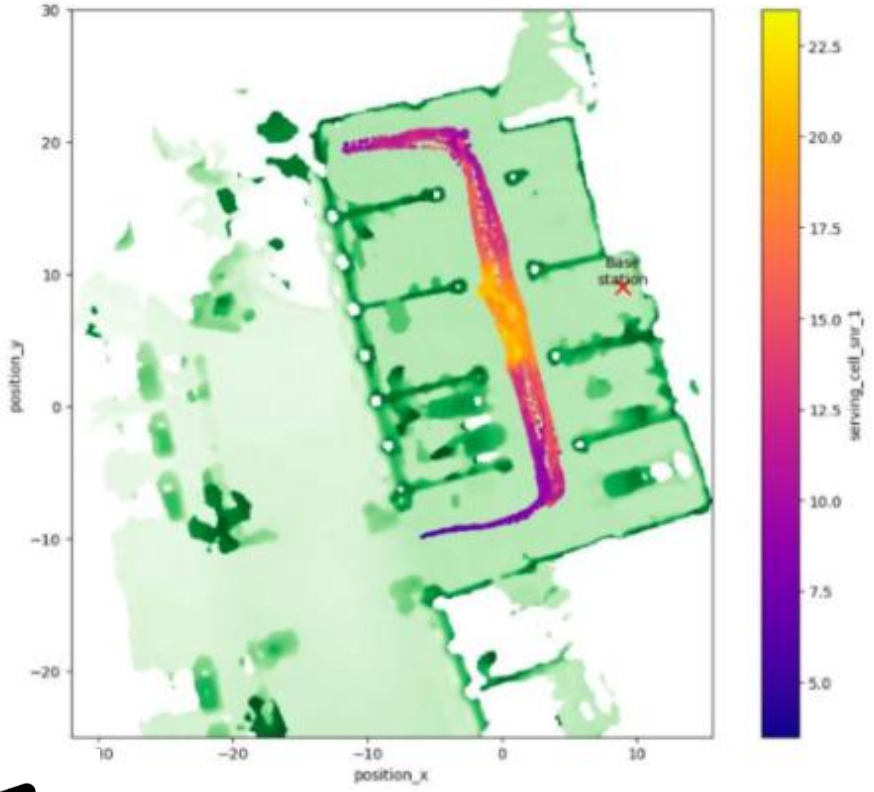
Memory-Aware Federated Learning for Anomaly Detection over IoT Data Streams

Raafat Osman[1] Zahraa El Attar[1] Nikolaos Papadakis[1] Georgios Bouloukakis[1,2]

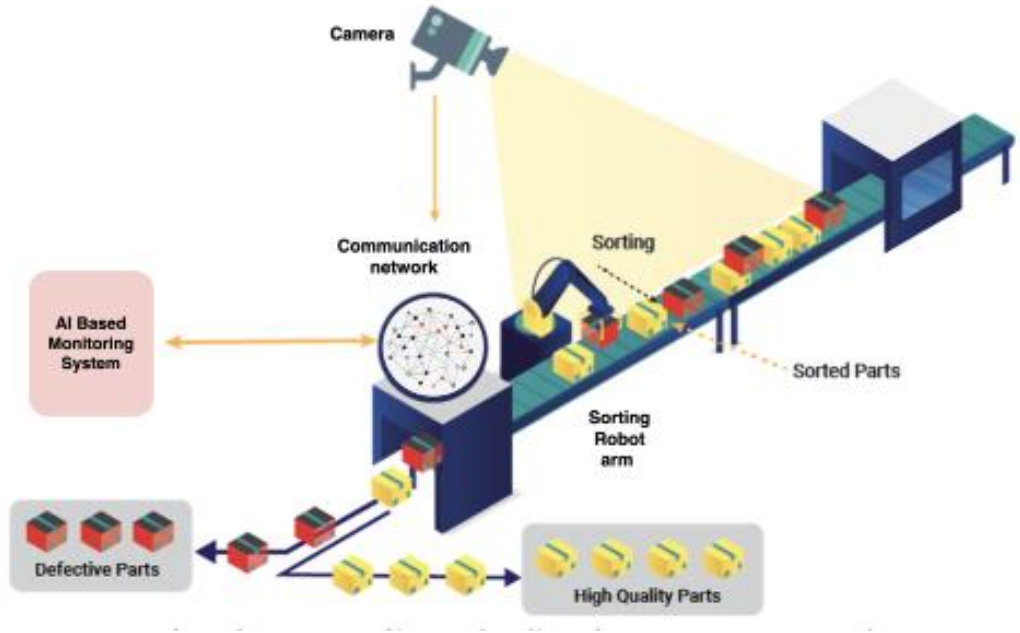
[1] Télécom SudParis · Institut Polytechnique de Paris [2] University of Patras

Partners use-cases

ROBOT TRAJECTORY (POSITION DATA) - ERICSON



SORTING MACHINE (CAMERA AND NETWORK DATA) - SIEMENS



ERICSSON

SIEMENS

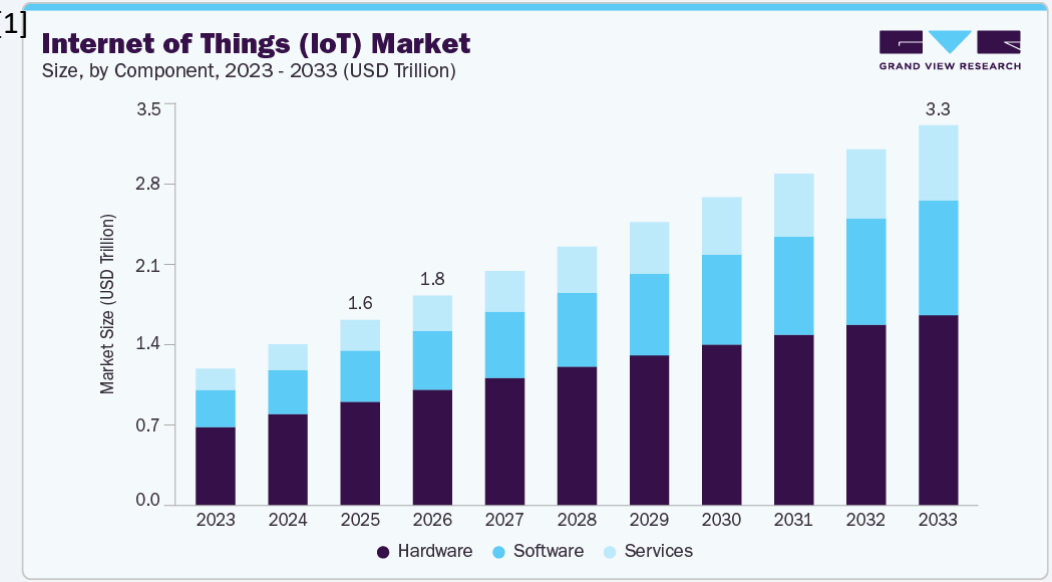
The new age of AIoT

The growth of IoT:

- The investment in IoT is highly increasing
- The increase of IoT devices implies the increase of data

➢ **The need to handle big data**

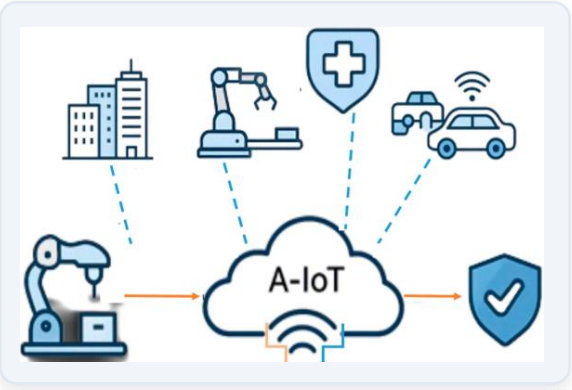
[1]



The integration of AI:

- Modern IoT devices integrates more and more AI
- AI and ML models are implemented on the devices

➢ **The need for quick predictions**



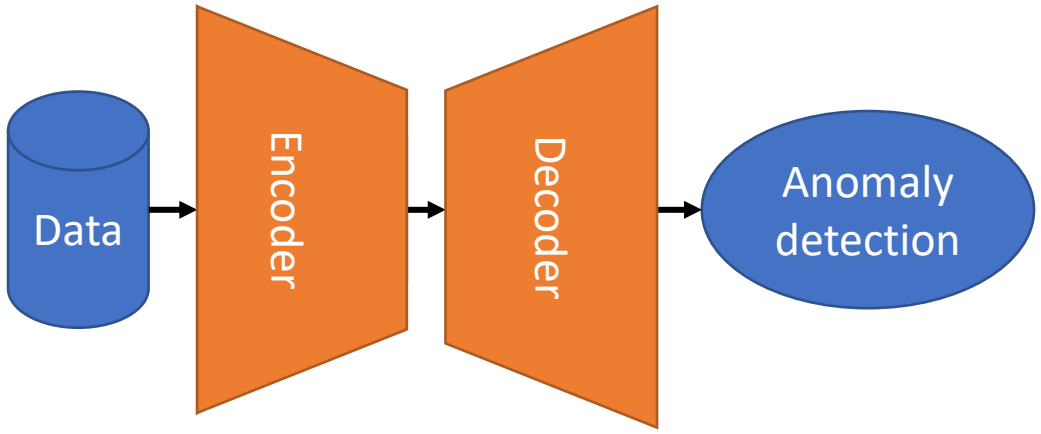
- Intrusion detection
- Autonomous response
- Anomaly detection

[1]: https://www.grandviewresearch.com/industry-analysis/iot-market

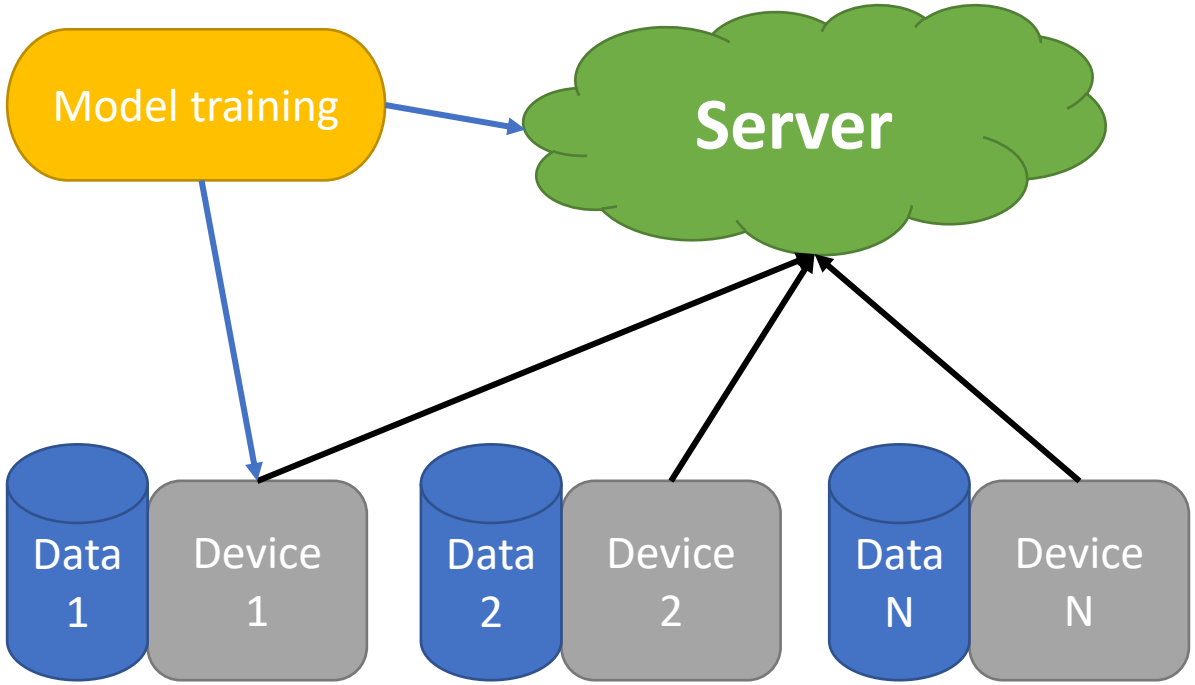
Model training in IoT systems for anomaly detection

ANOMALY DETECTION

AUTO-ENCODER



MODEL TRAINING

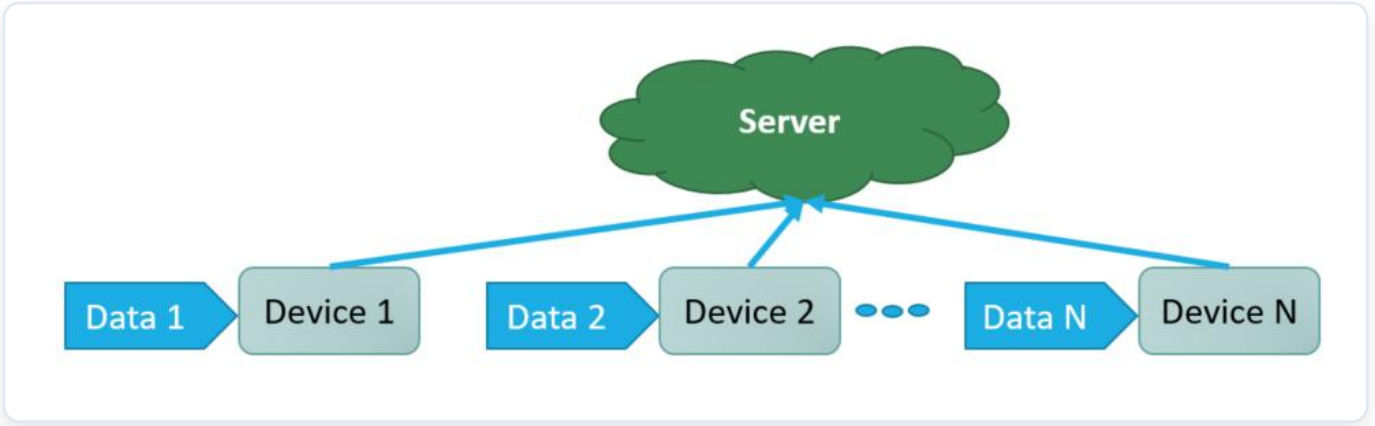


THE PROBLEM

Limitations & Challenges

- All devices connect to a central server
- Devices must make quick predictions

- **Data is streamed continuously**
- **Data is different across devices**



1

Limited Computation

Longer training time or lower accuracy

2

Limited Data Capacity

Data dropping → information loss

3

Data Streaming

Massive, heterogeneous datasets

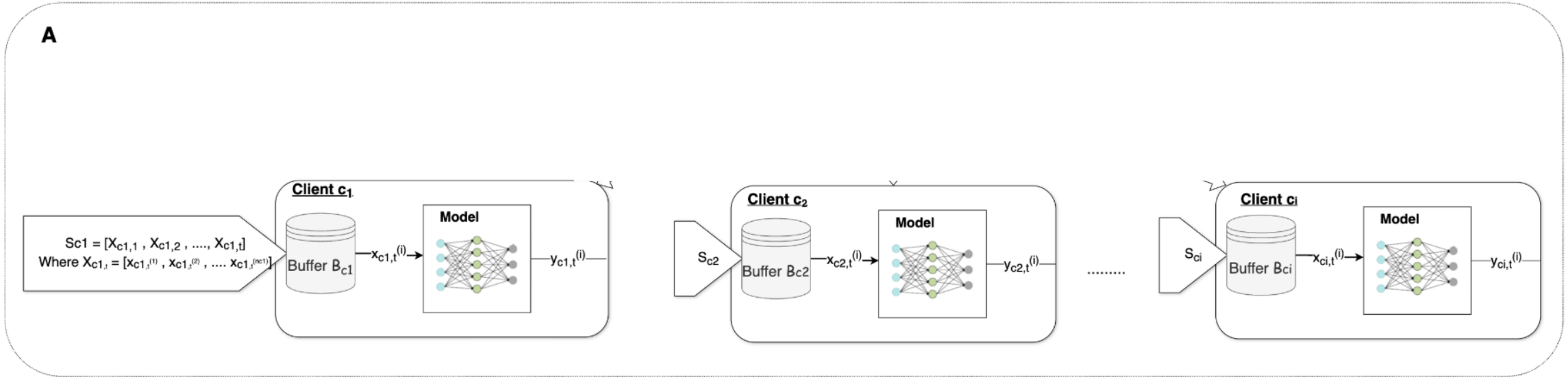
1

CHALLENGE 01

Limited Computation power

SOLUTION

Federated Learning based framework



Distributes the training

Aggregates the weights

Sends back the updated weights

➢ **Doesn’t solve the limited data capacity on the devices**

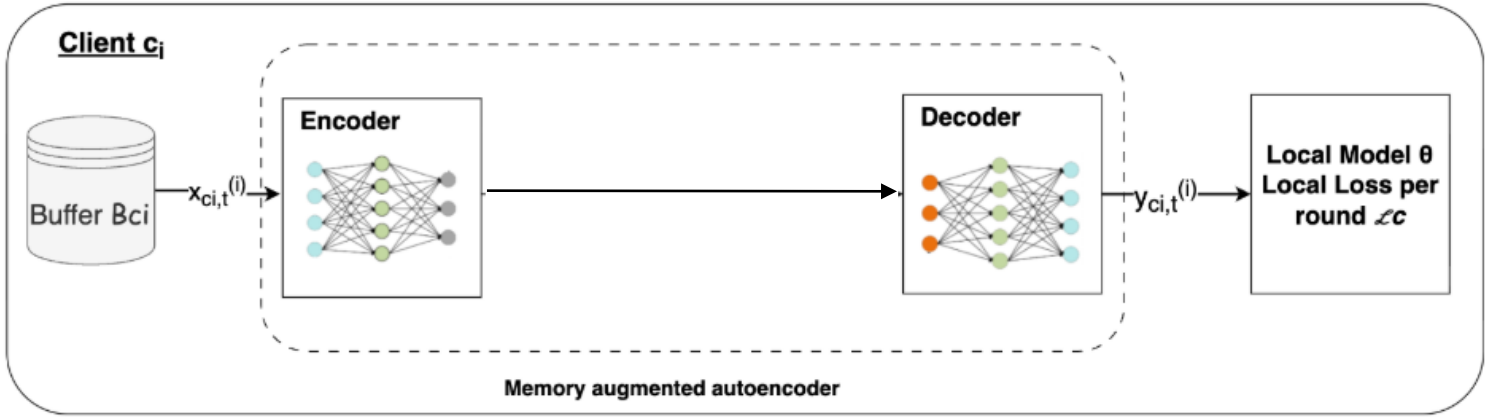
2

CHALLENGE 02

Limited Data Capacity

SOLUTION

Memory-Augmented Auto-Encoder

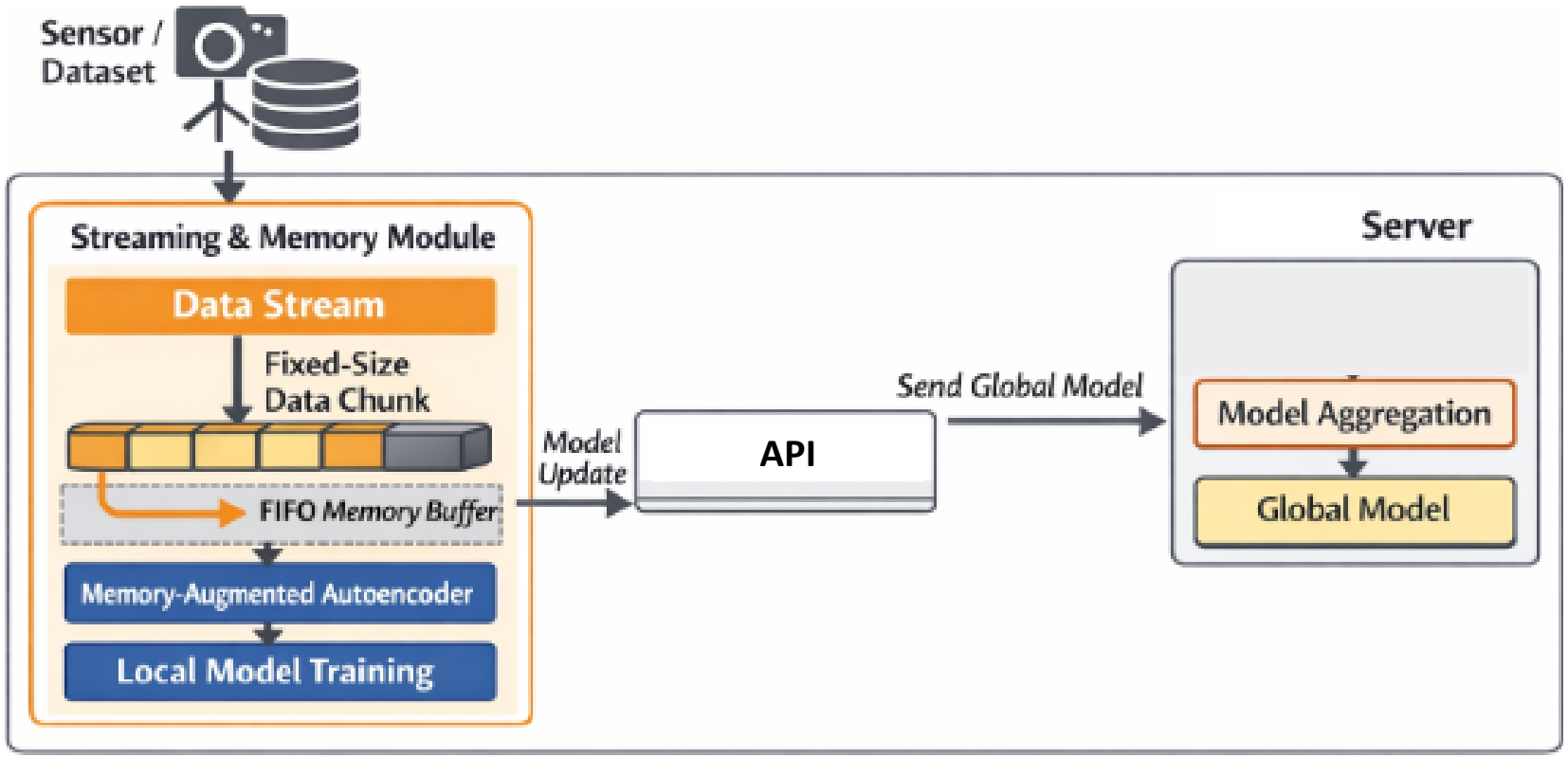


Limited data capacity → Data dropping → **Information loss** → Decrease in accuracy

To avoid Information loss, we integrated:
Memory-augmented Auto-encoder (MemAE)

→ Memory layer to preserve the most relevant information

Platform overview



- **Doesn’t solve the heterogeneity of data**

3

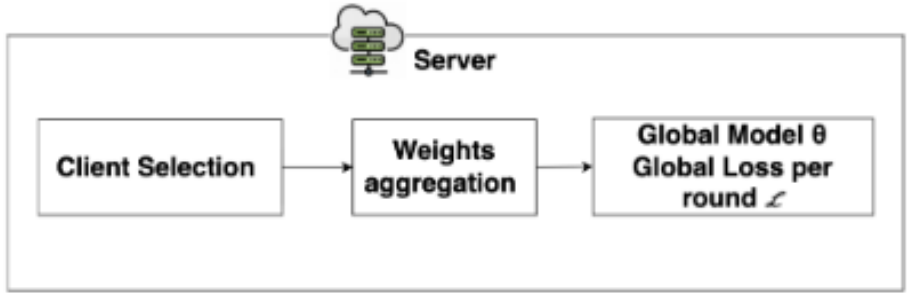
CHALLENGE 03

Data Streaming

SOLUTION

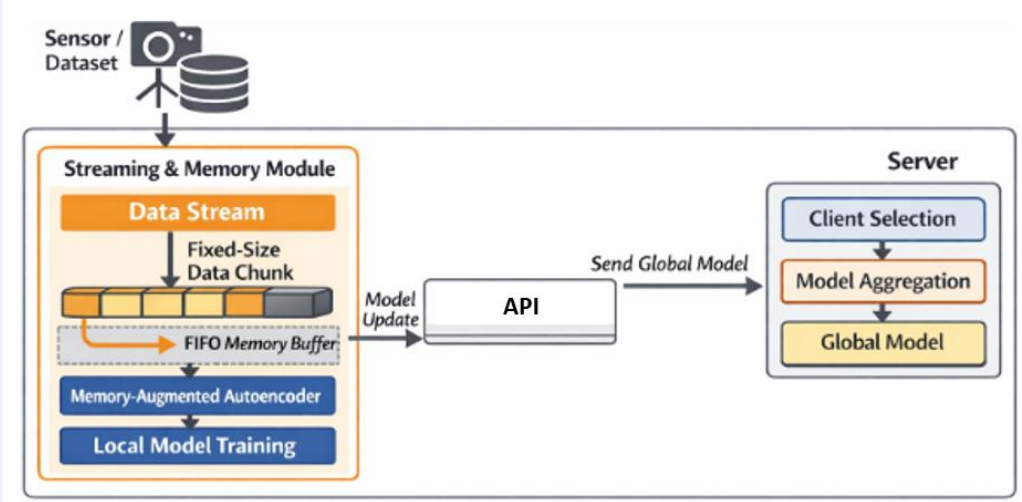
Client Selection Strategies

Every client receives a different data



Different data → Different training → Different performance

- **Client selection strategies ensure the prioritization of the most informant clients**
- **Different strategies depending on the related goals**

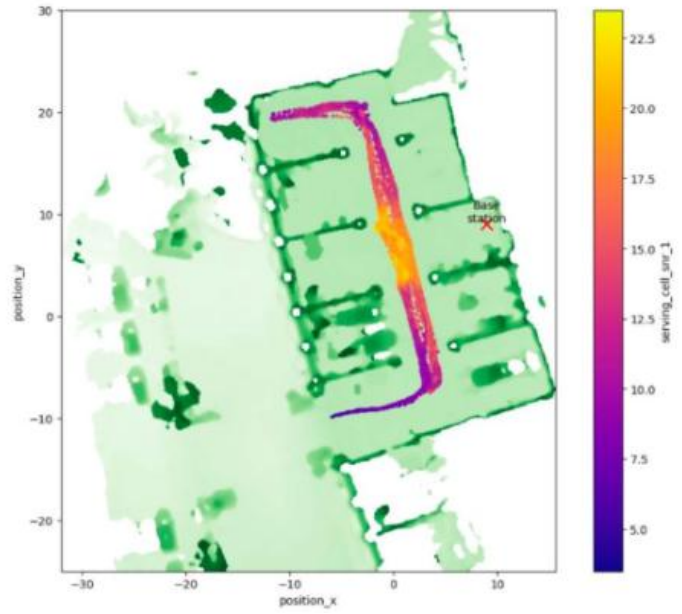


EVALUATION · USE CASE 1

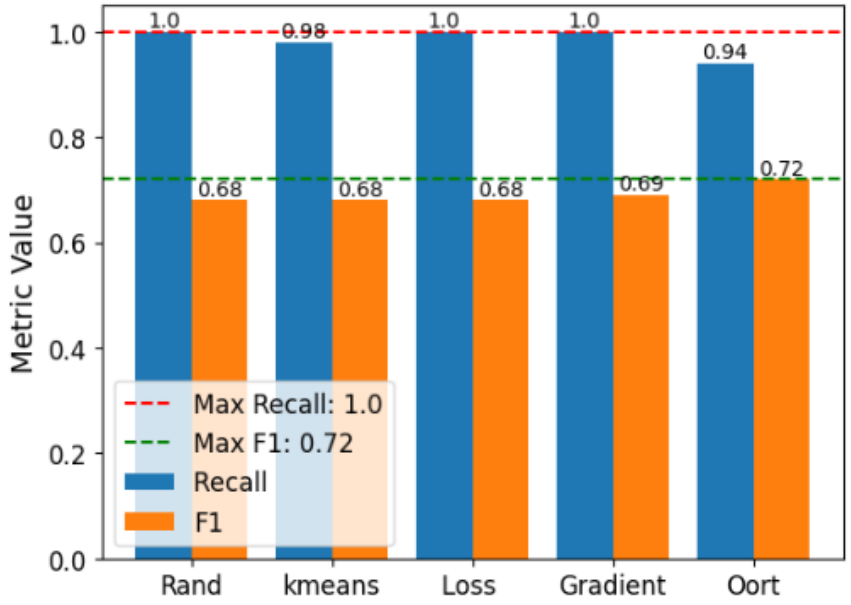
Roaming Robot

We validate the system on three real-world use cases. First: a roaming robot navigating with position and network data.

ROBOT TRAJECTORY (POSITION DATA)



CLIENT-SELECTION PERFORMANCE



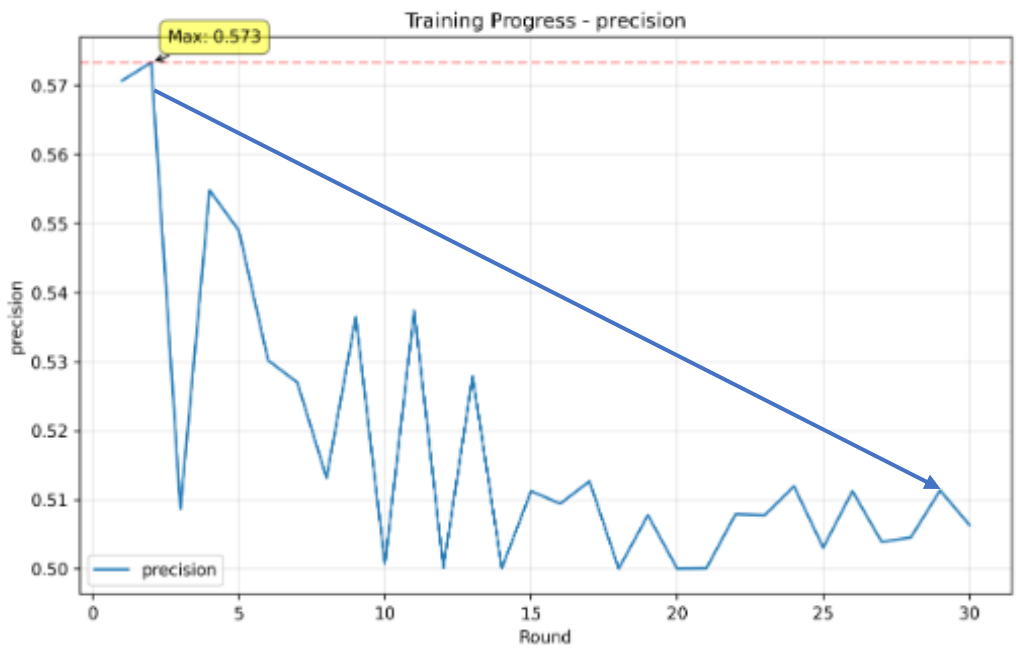
KEY TAKEAWAY **Client-selection strategies yield different performance depending on the use case.**

EVALUATION · USE CASE 1

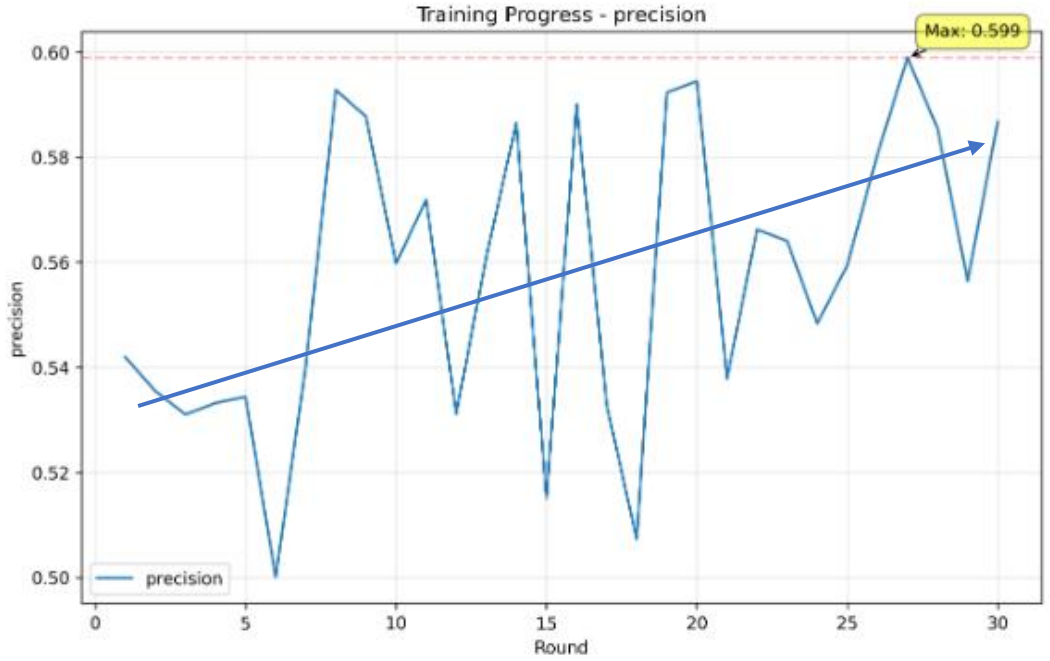
Roaming Robot

PRECISION VARIATION ON TRAINING ROUNDS

Auto-encoder



Memory-augmented Auto-encoder



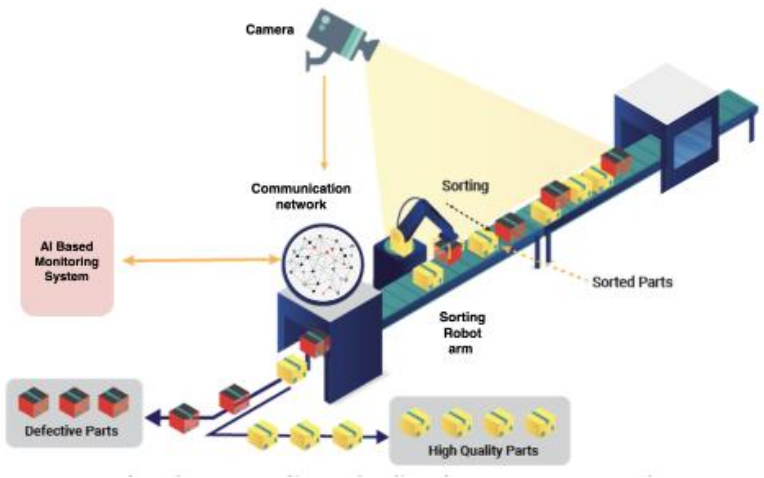
KEY TAKEAWAY **The memory-augmented auto-encoder ensured performance increase through training rounds.**

EVALUATION · USE CASE 2

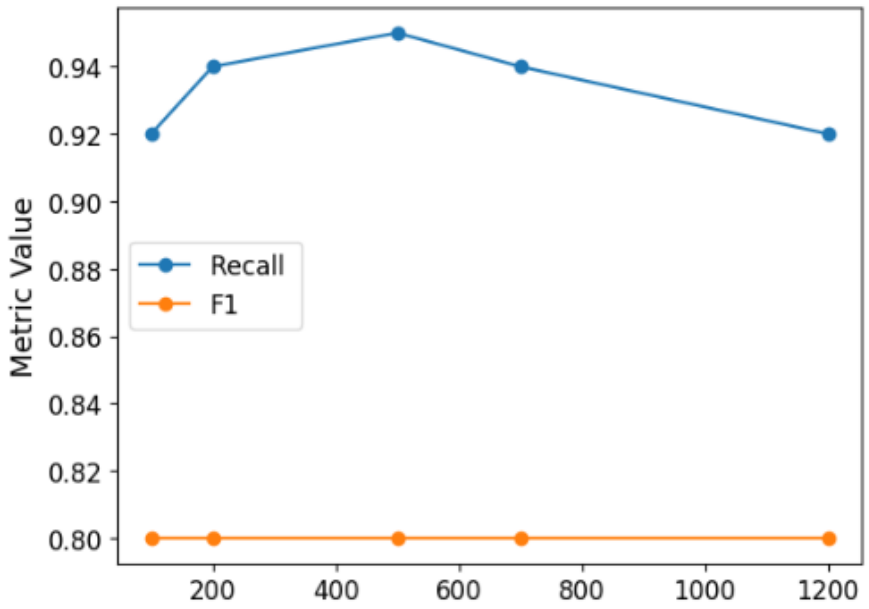
Sorting Arm

We validate the system on three real-world use cases. Second: a sorting machine with camera, sensors and network data.

SORTING MACHINE (CAMERA AND NETWORK DATA)



RECALL AND F1 SCORES WITH CHUNK SIZE



KEY TAKEAWAY **The system ensured a steady performance even with higher amounts of data entering the system.**

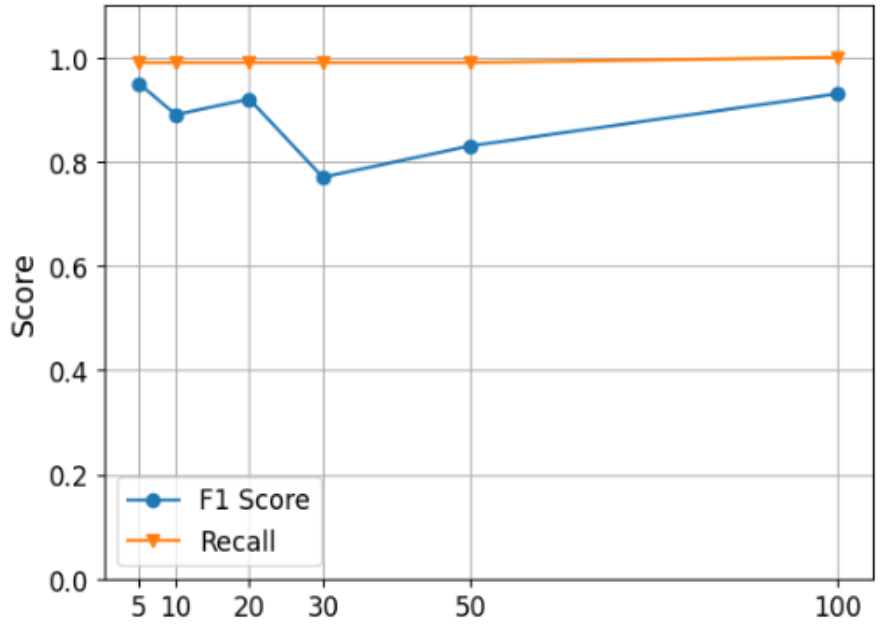
EVALUATION · USE CASE 3

KDD on 100 Clients

We validate the system on three real-world use cases. Third: KDD [1] dataset for cyber-attacks.

CYBER-ATTACKS (NETWORK DATA)

RECALL AND F1 SCORES THROUGH THE NUMBER OF CLIENTS



KEY TAKEAWAY **The system ensured a steady performance even with large numbers of clients training on it.**

[1]: https://www.kaggle.com/datasets/hassan06/nslkdd

WRAP-UP

Conclusion & Future Work

Solution:

Creating a system including:

- Memory augmented auto-encoder
- Federated learning
- Client selection strategies

Future work:

- Creating learning based client selection algorithms
- Going beyond anomaly detection to classification and time series forecasting

THANK YOU

Questions & Discussion

Memory-Aware Federated Learning for Anomaly Detection over IoT Data Streams
https://inria.hal.science/hal-05641969/

Raafat Osman · Zahraa El Attar · Nikolaos Papadakis · Georgios Bouloukakis

Télécom SudParis · Institut Polytechnique de Paris University of Patras

https://pandora-heu.eu/

ERICSSON SIEMENS IP PARIS TELECOM SudParis ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ UNIVERSITY OF PATRAS PANDORA MDM 2026 27th Conference on Mobile Data Management