Designing a Forensic-ready Wi-Fi Access Point for the Internet of Things

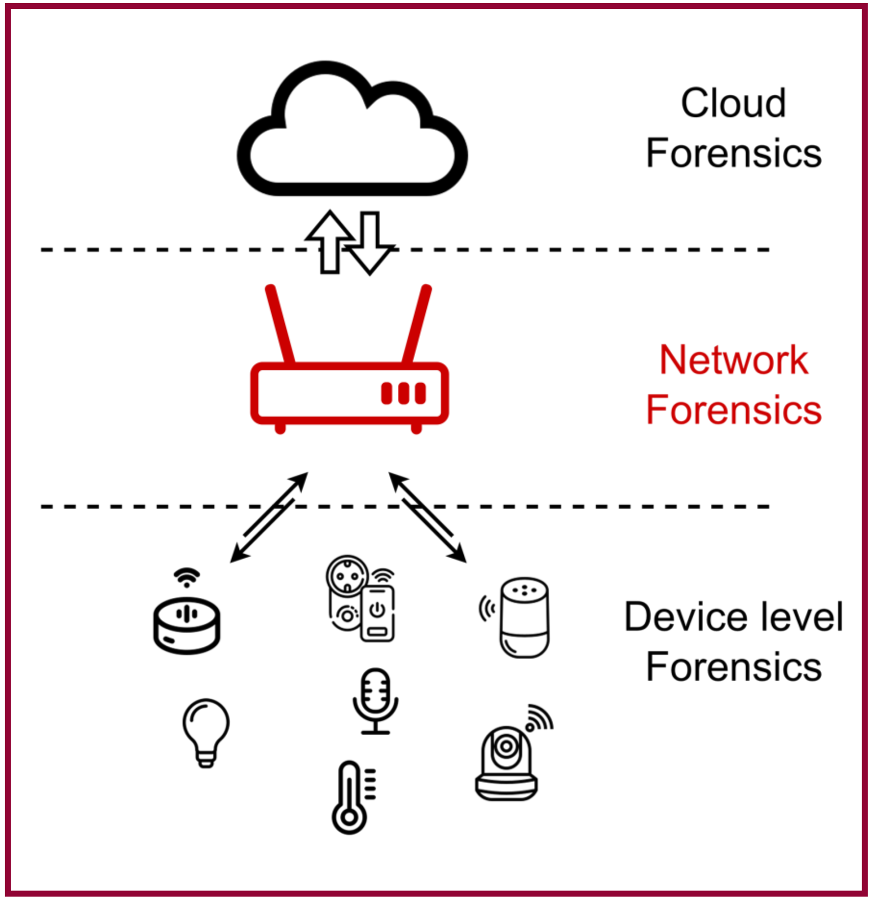
Fabio Palmese
Ph.D. Student
Advisor: Prof. Alessandro E. C. Redondi
ANTLab, DEIB
Politecnico di Milano
fabio.palmese@polimi.it

6 months visiting in **UCL** from September
Advisor: Anna Mandalari

Introduction: IoT Forensics

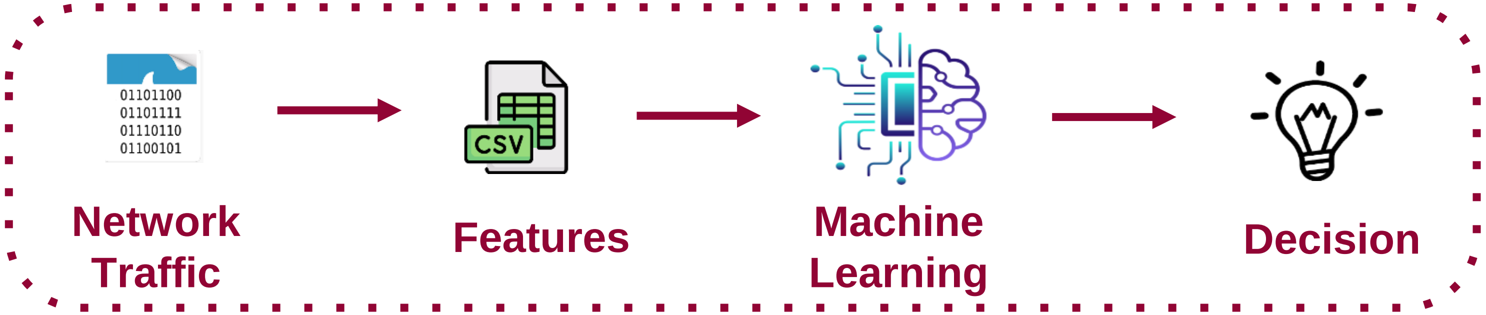
IoT Forensics: Branch of Digital Forensics with the goal of identifying and extracting information from IoT devices, to be used as source of evidence

The IoT device as **witness** of user daily activities



IoT Forensics layers

State-of-the-art



Limitations:

- Huge *space* needed to store all the network traffic packets
- Considerable *time* for PCAP processing for feature extraction
- Need setup to collect the traffic as close as possible to where it is produced

Our solution: Capture and compute features on-the-fly as the traffic flows through the Access Point with an easily configurable tool

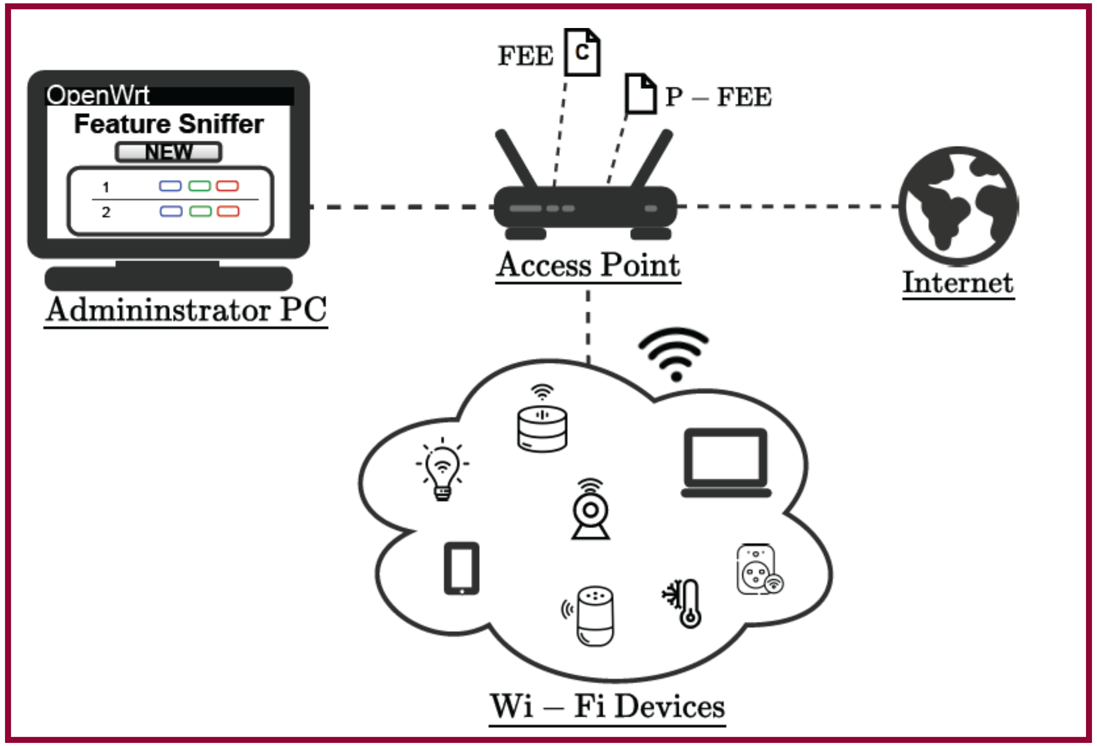
ANTLAB

Feature-Sniffer: Project Overview

Idea: Directly in the Access-Point **aggregate packets in time windows** and compute statistical features per device **(on-the-fly)**

Three different components:

- Easy to use web interface
- Feature Extraction Engine (FEE) for Network/Transport layer features (C program)
- Physical layer FEE for RSSI and CSI-based features



Architecture of *Feature-Sniffer*

Can we afford running it in Access Points?

ANTLAB

Performance Evaluation

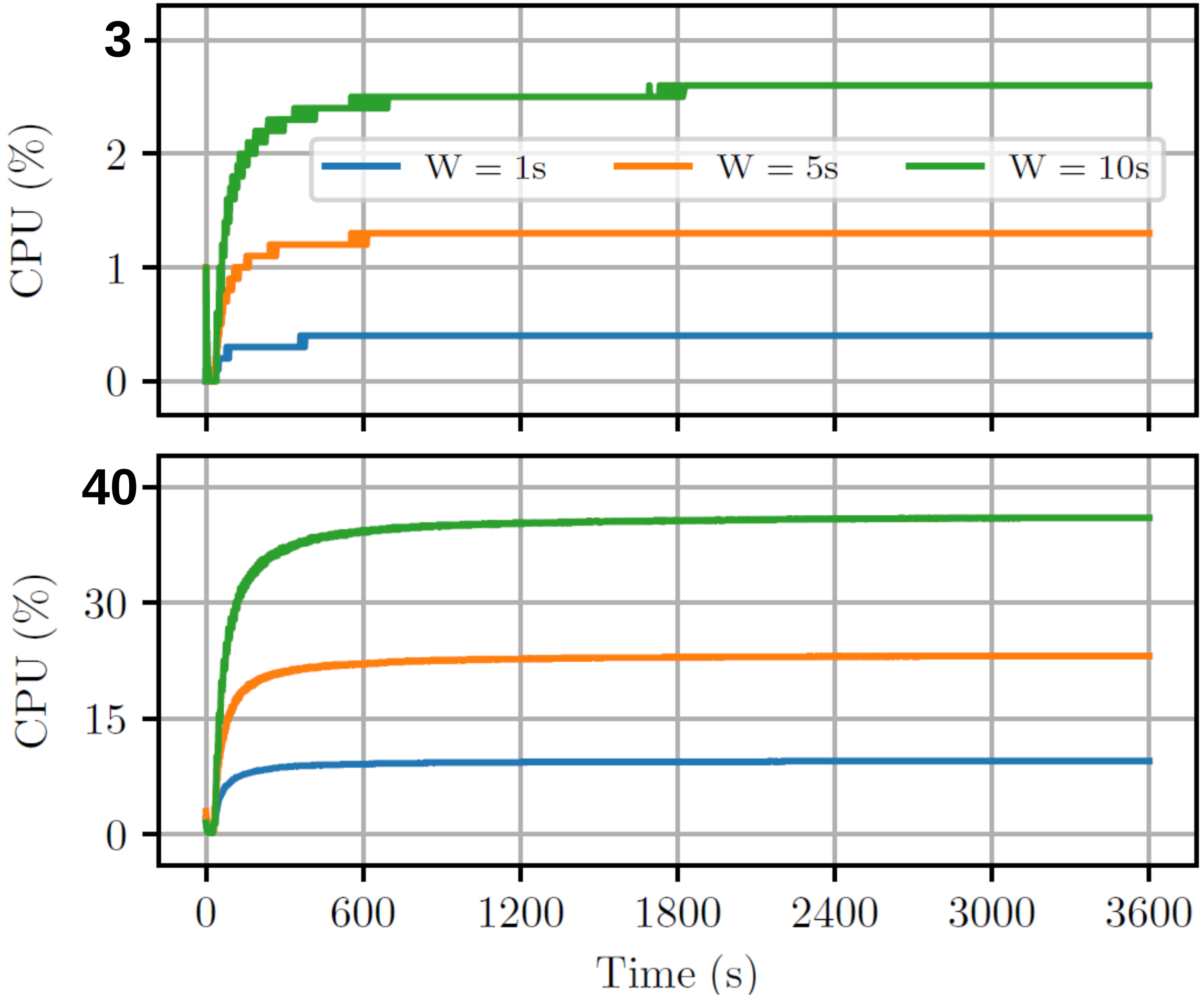
We test the tool performance (CPU) into two different Access Points in a network with **30 IoT Devices**, enabling **all features** with **different window lengths**

Linksys WRT3200ACM

- 512 MB RAM
- 1.8 GHz CPU (4 cores)

Netgear R6120

- 64 MB RAM
- 560 MHz CPU (2 cores)

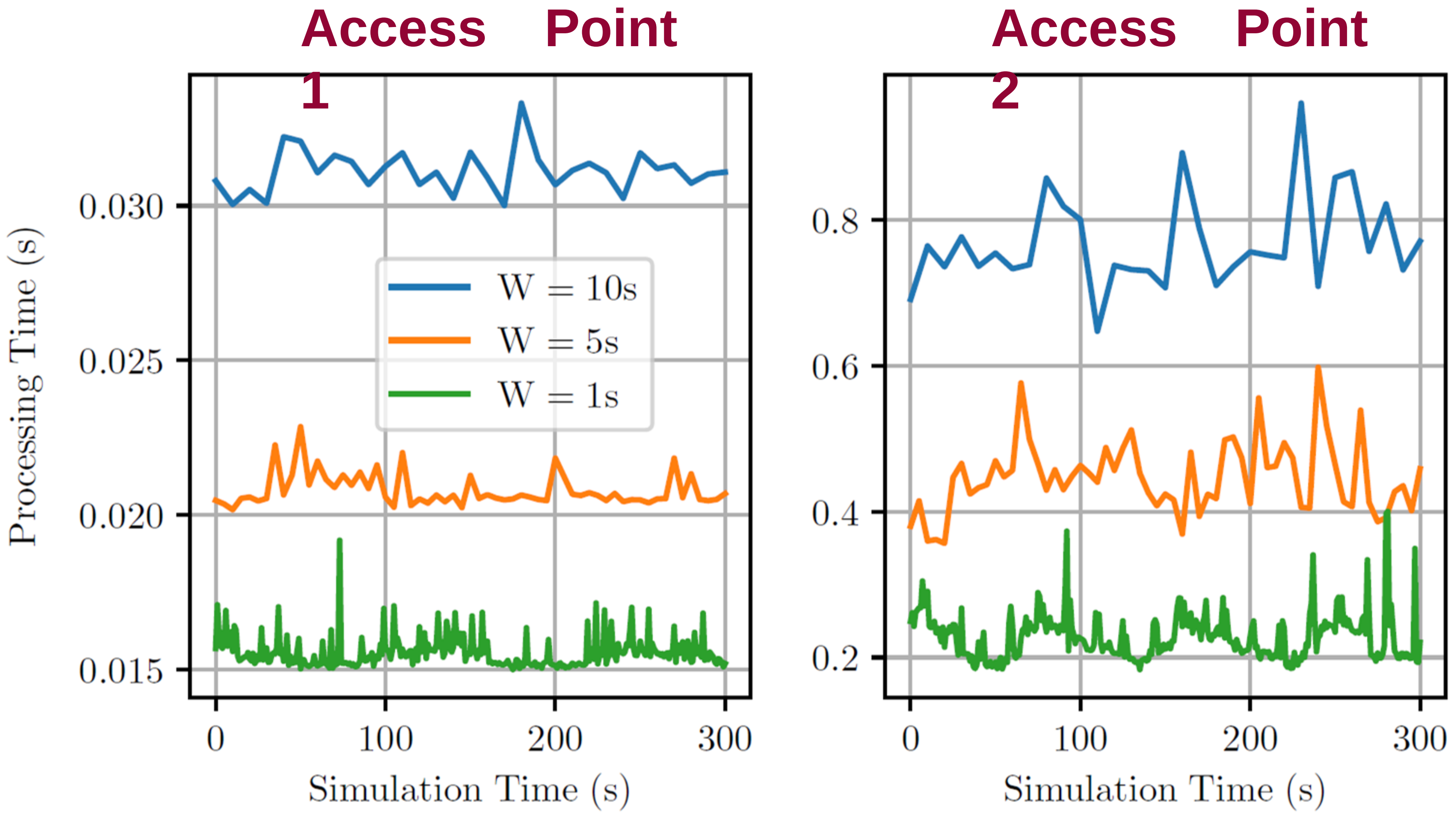


Can we produce the features real-time?

ANTLAB

Performance Evaluation: Real-time?

We report the processing time of each window for all the connected devices



Application Cases

ANTLAB

Application Cases

We use the tool output for performing different tasks:

1. **IoT Device Identification (F1 94%):**

   Goal: Identify the device producing the traffic

2. **IoT Cameras Human Activity Recognition (F1 85%):**

   Goal: Identify different activities of the user in front of smart cameras

3. **Amazon Echo Analysis:**
   a. Interaction Detection (**F1 99%)**
   b. English vs. Italian: language recognition (F1 84%)
   c. Real vs. Synthetic: voice recognition (F1 73%)

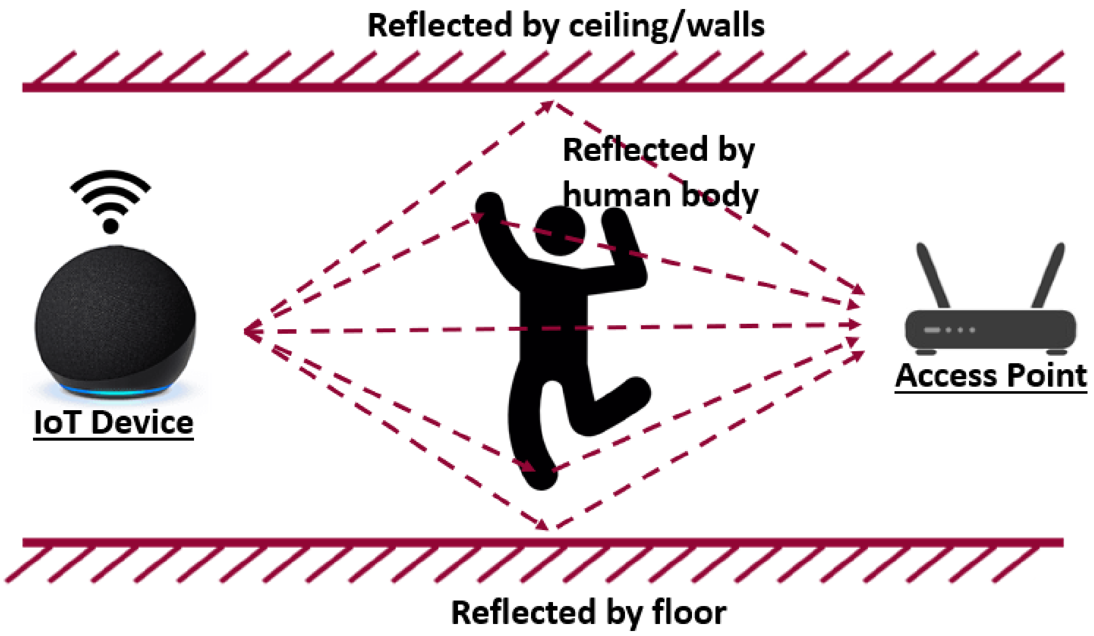
1. **Human Passage Detection with CSI:**

   Goal: Detecting a human passing through the room door using CSI extracted from an IoT device

ANTLAB

Wi-Fi Channel State Information (CSI)

- Describes the propagation of the signal from the sender to the receiver
- Discriminates multipath characteristics: suitable for **human activities sensing**



CSI returns a complex value **for each subcarrier** for each packet

$$H_i = |H_i| e^{j \sin(\angle H_i)} \quad i \in [1, N]$$

Amplitude: $|H_i|$

Phase: $\angle H_i$

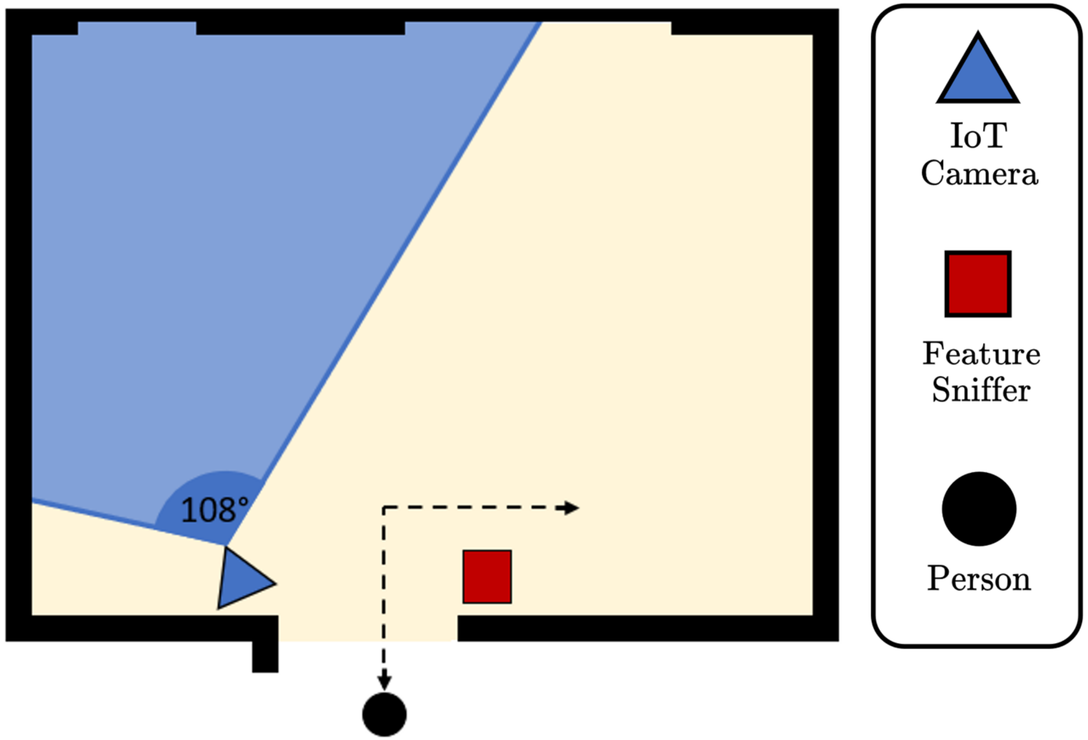
ANTLAB

Task 4: Human Passage Detection

Goal: Detect human **presence** in the room using CSI data from a generic IoT device

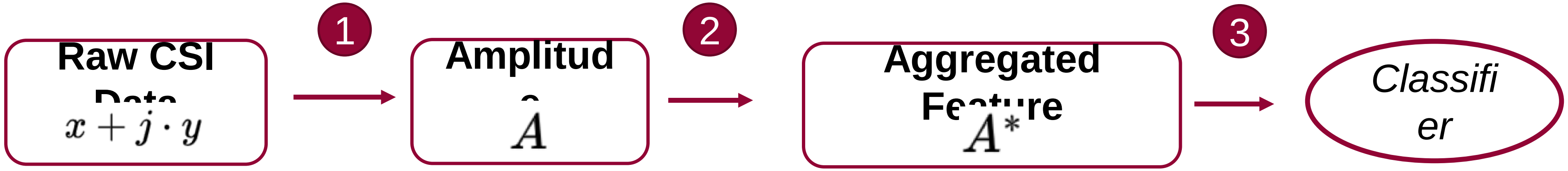
We use an IoT camera to generate traffic and collect data for **50 total passsages** throgh the door

IMPORTANT: the person is not in the camera FOV while moving



Frame N.	Timestamp	Device	CSI i=-32	CSI i=-31	….	CSI i=30	CSI i=31	Label
1	1686396012.00	AA:AA:AA:AA:AA:AA	242+168	73-282j	….	71-217j	69-244j	1
2	1686396012.31	AA:AA:AA:AA:AA:AA	236+152j	76-252j	….	76-261j	66-231j	1
3	1686396015.31	AA:AA:AA:AA:AA:AA	266+164j	83-268j	….	74-245j	61-263j	0

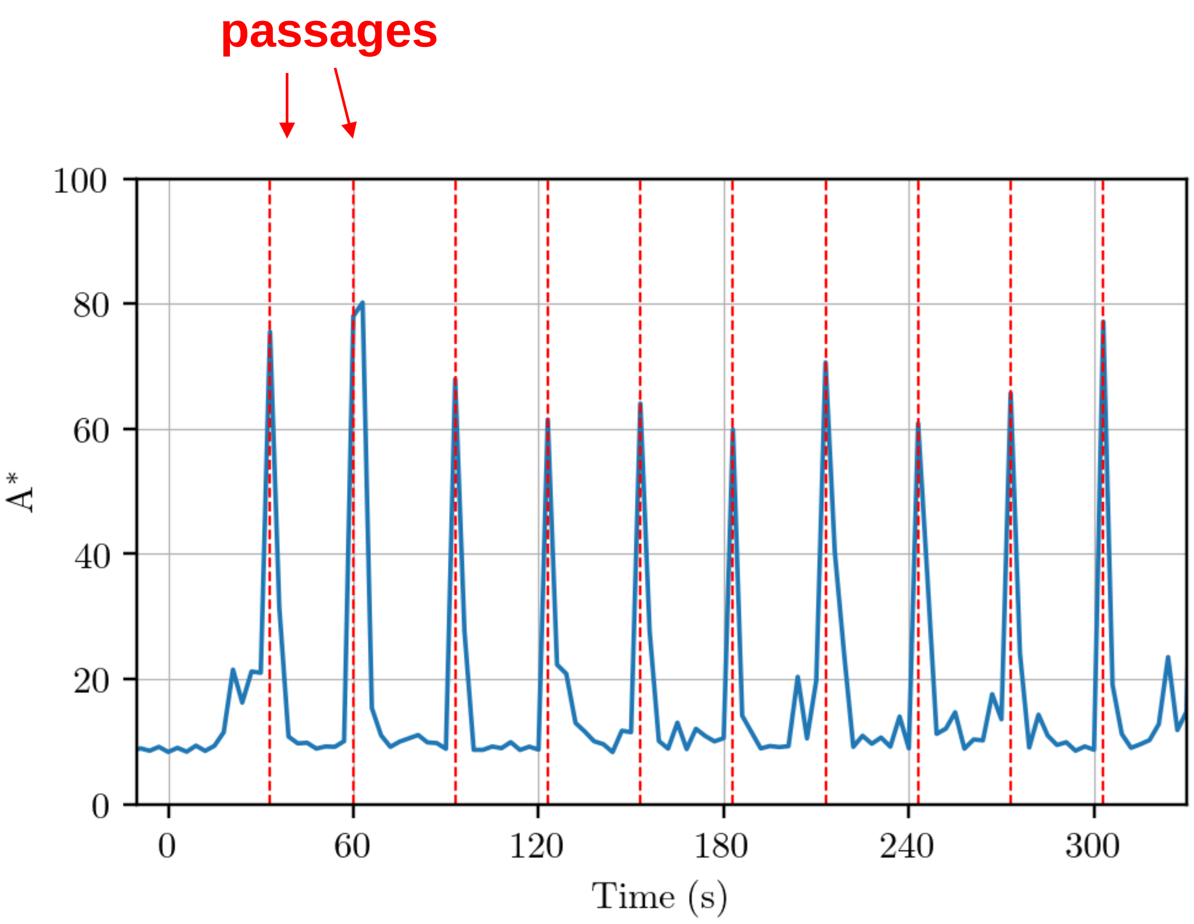
Analysis Pipeline



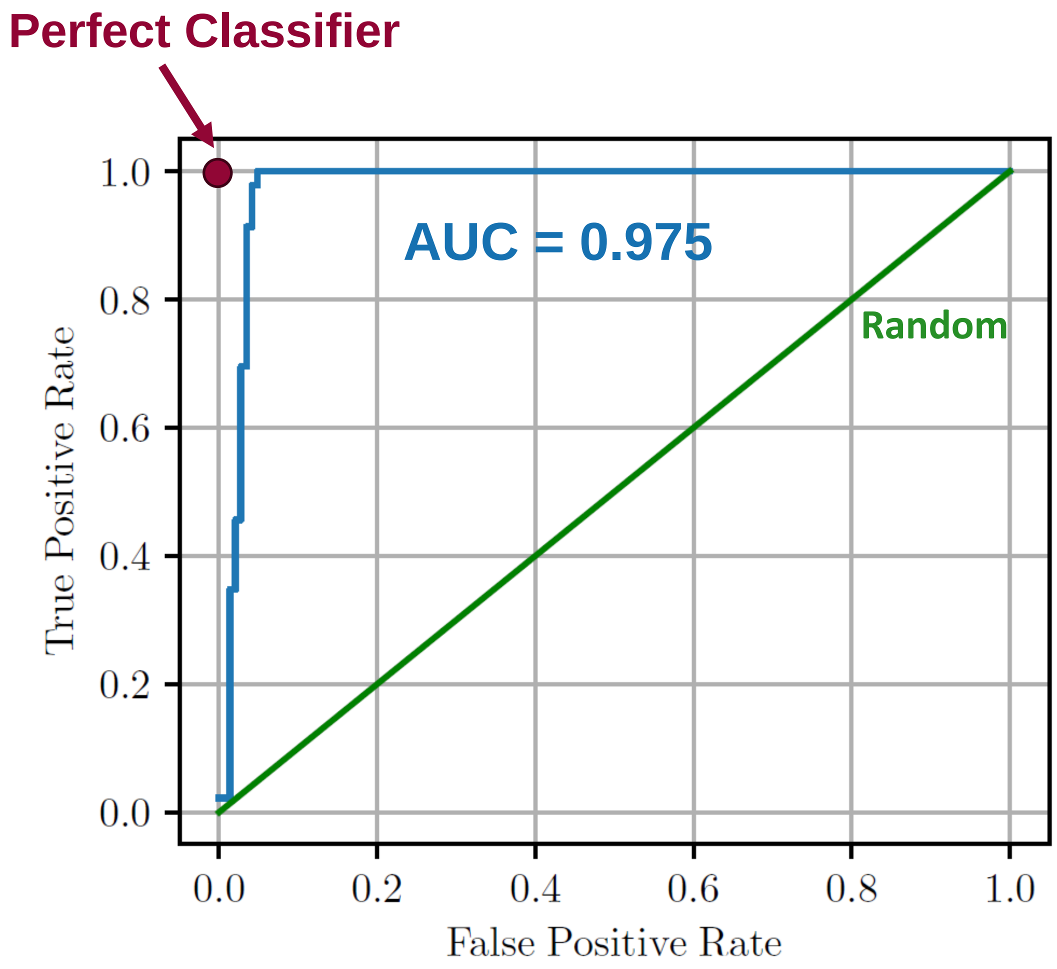
1. Extract CSI Amplitude and Phases from raw CSI data
2. Frames are grouped into time windows of 3 seconds and we extract ***A****:
   1. For each window compute the st. dev. of the amplitude in the different frames for each subcarrier: $|\sigma(A_0), \sigma(A_1) \ldots \sigma(A_N)|$ (vertical st. dev)
   2. For each window compute the mean over all N subcarriers to have a single value for each time window: $A^*_t = \mu(|\sigma_t(A_0), \sigma_t(A_1) \ldots \sigma_t(A_N)|)$ (horizontal mean)
3. Values of A* are passed to a binary threshold classifier and compared with the ground truth to extract resulting performance: ROC and AUC.

[2] S. M. Hernandez and E. Bulut, "**Adversarial Occupancy Monitoring using One-Sided Through-Wall WiFi Sensing**," in 2021 IEEE International Conference on Communications (ICC): IoT and Sensor Networks Symposium (IEEE ICC'21 - IoTSN Symposium), Montreal, Canada, Jun. 2021.

Task 3: Results



A* over time



ROC Curve

Can we optimize the storage?

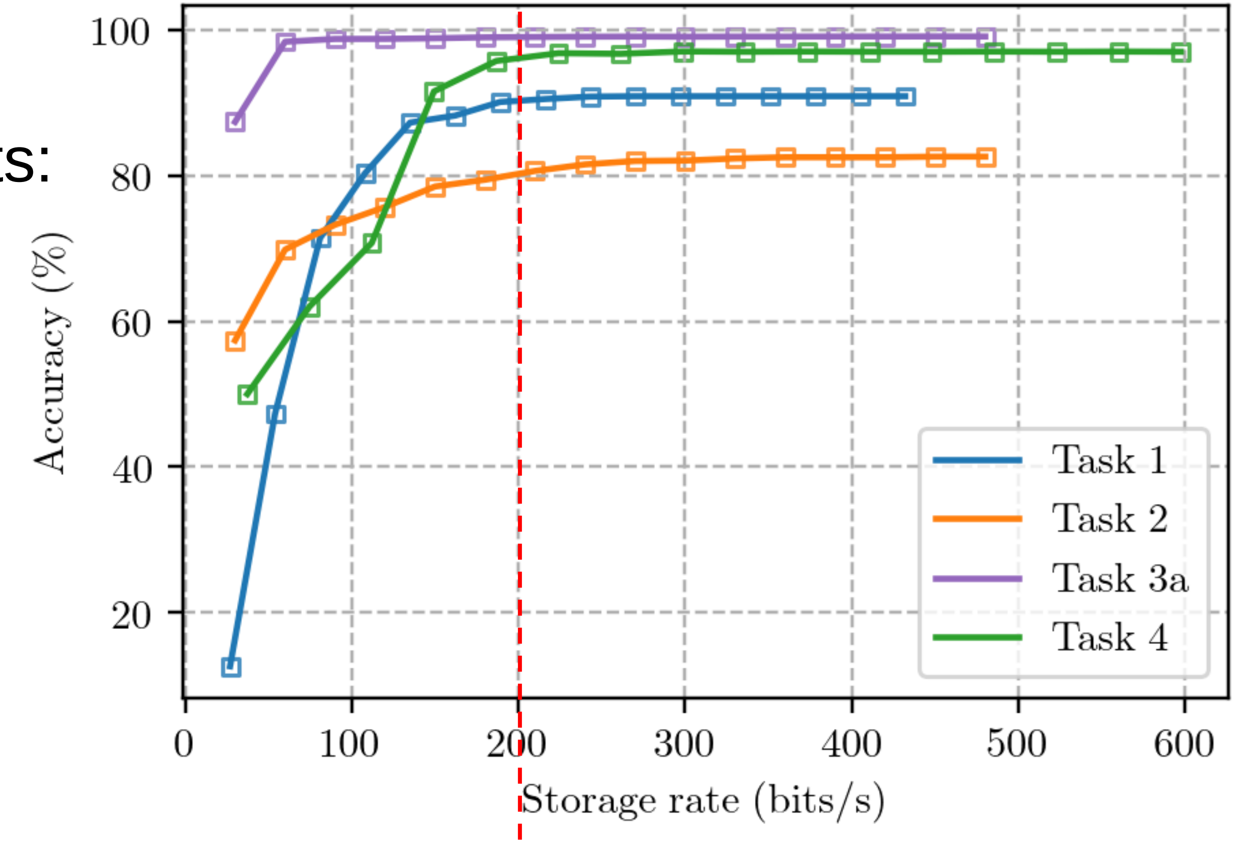
Storage-Accuracy with Lossy Compression

We apply Scalar Quantization for each value in the dataset of each task

Each value is represented with B bits:

$$v_i = \lfloor \frac{v_i - v_{\min}}{v_{\max} - v_{\min}} \cdot (2^B - 1) \rceil$$

We use different values of B ranging in [1,16], and extract the corresponding accuracy



200 bits/s are enough

Future Directions

Towards my visiting period in UCL

Investigating on Privacy and Security for IoT devices in the smart home: Integration in Wi-Fi access points

Thank you for your attention!

ANTLAB | ADVANCED NETWORK TECHNOLOGIES LAB

antlab.polimi.it

antlabpolimi

Fabio Palmese
fabio.palmese@polimi.it

ANTLAB