The Trace You Leave Behind — and What We Can Do About It





Pablo Serrano
pablo@it.uc3m.es

IMDEA Networks annual workshop 2026

uc3m | Universidad Carlos III de Madrid

Talk Overview

- Campus WLAN as a Human Sensor
- Privacy Threats
- Practical Defences

DATA PROTECTION BOARD SUPERVISED

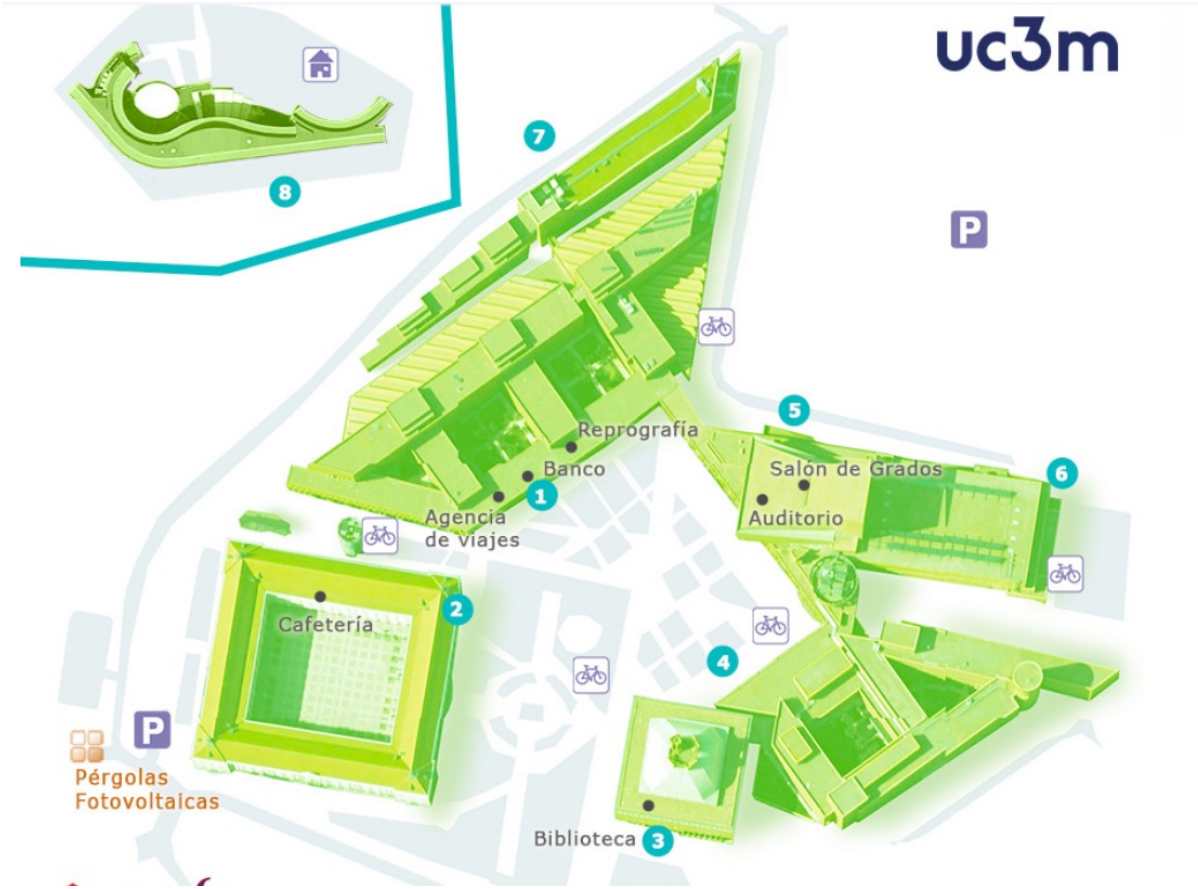
(Most of) The activities described in this presentation have been conducted under the strict supervision of the Data Protection Board at UC3M

No actual users were re-identified in the making of this research (actually: a few)

PART 1: WLAN AS A HUMAN SENSOR

Leganés Campus – eduroam WiFi

Building	# APs
1	78
2	72
3	24
4	50
5	17
6	8
7	36
Total	285



The Dataset

- UC3M eduroam network
  - 24k students (16k undergrad), 2.2k faculty, 1k admin
  - Leganes campus: 1/3rd total population (10k users)

- RADIUS logs <timestamp · user ID · device MAC · AP ID>
  - One record per AP change or 15 minutes
  - Identifiers hashed with MD5

Validation: April 28 Blackout

- 28 April 2025: Iberian Peninsula's power failure
  - Mobile networks collapsed
  - UC3M campus WiFi remained (mostly) operational
    - Eduroam or Public-Wifi (based on SMS)



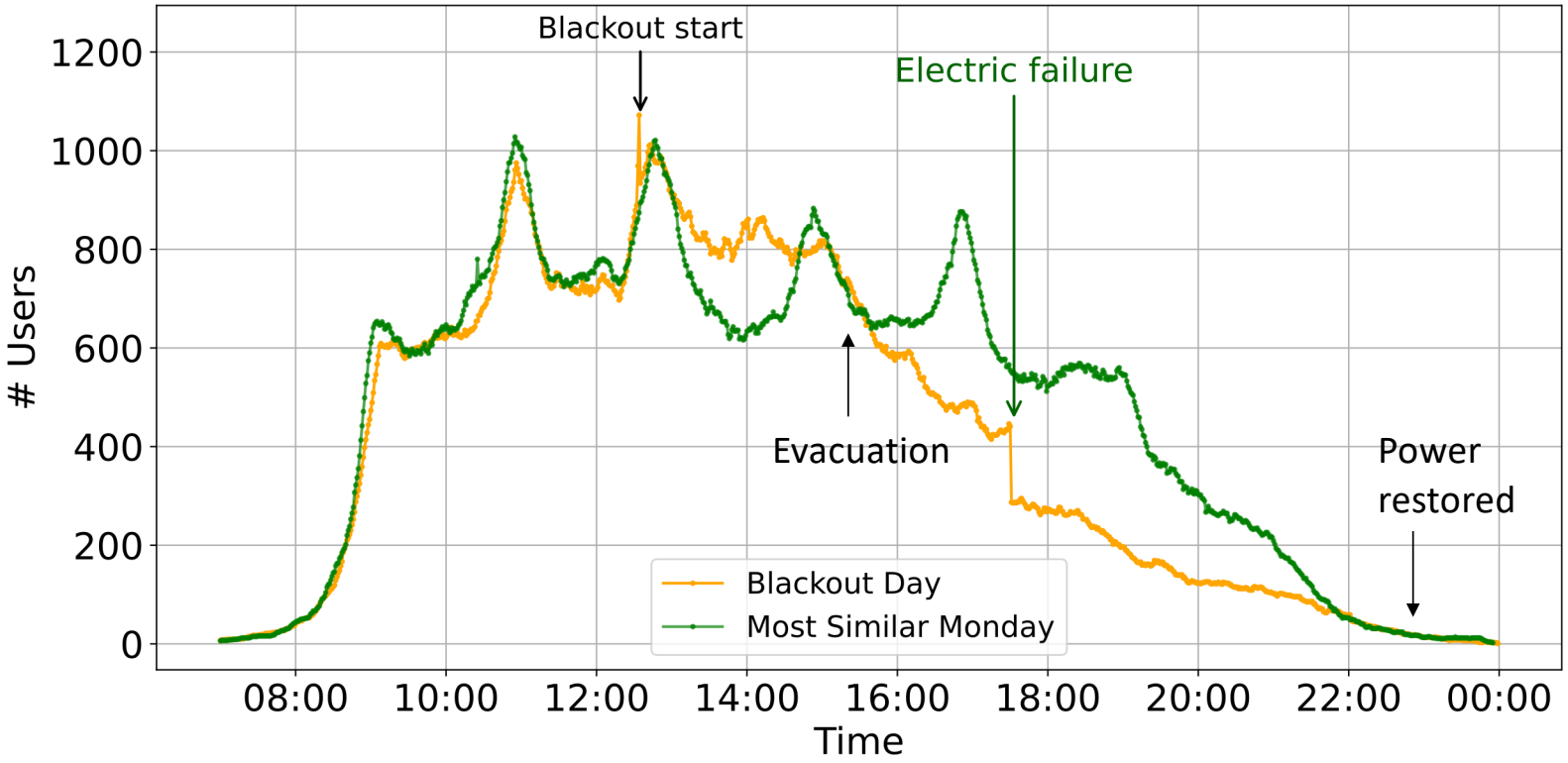
Connected devices and traffic

- Few new devices, almost no new users

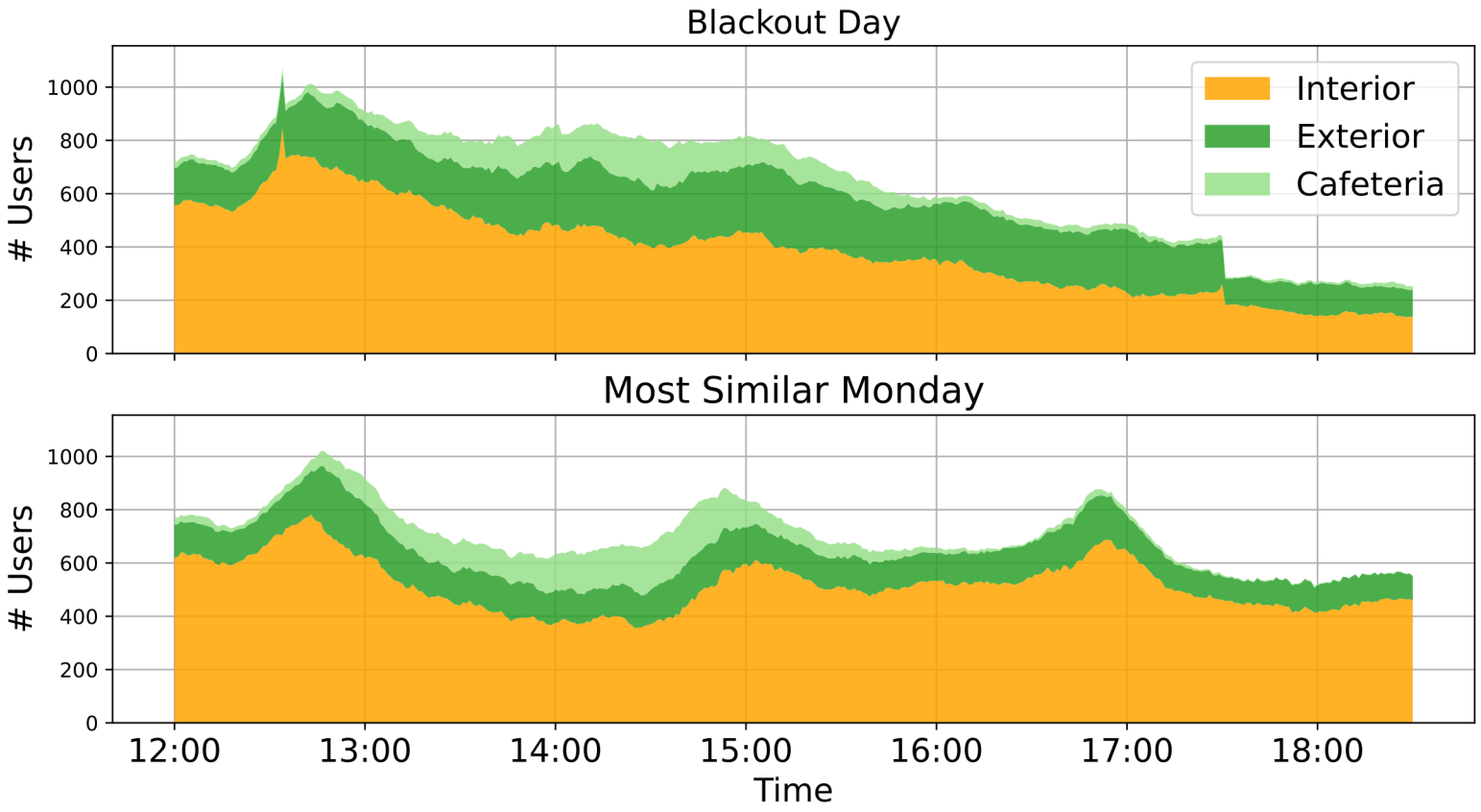
Users: 2 819 (new: 0.6%)	Devices: 4 890 (new: 4%)
2 802 existing users	4 692 existing devices
-> 156 users added a new device	-> 179 new devices of existing users
17 new users	19 new devices of new users

- More traffic was generated

Campus occupation (one mobile/user)



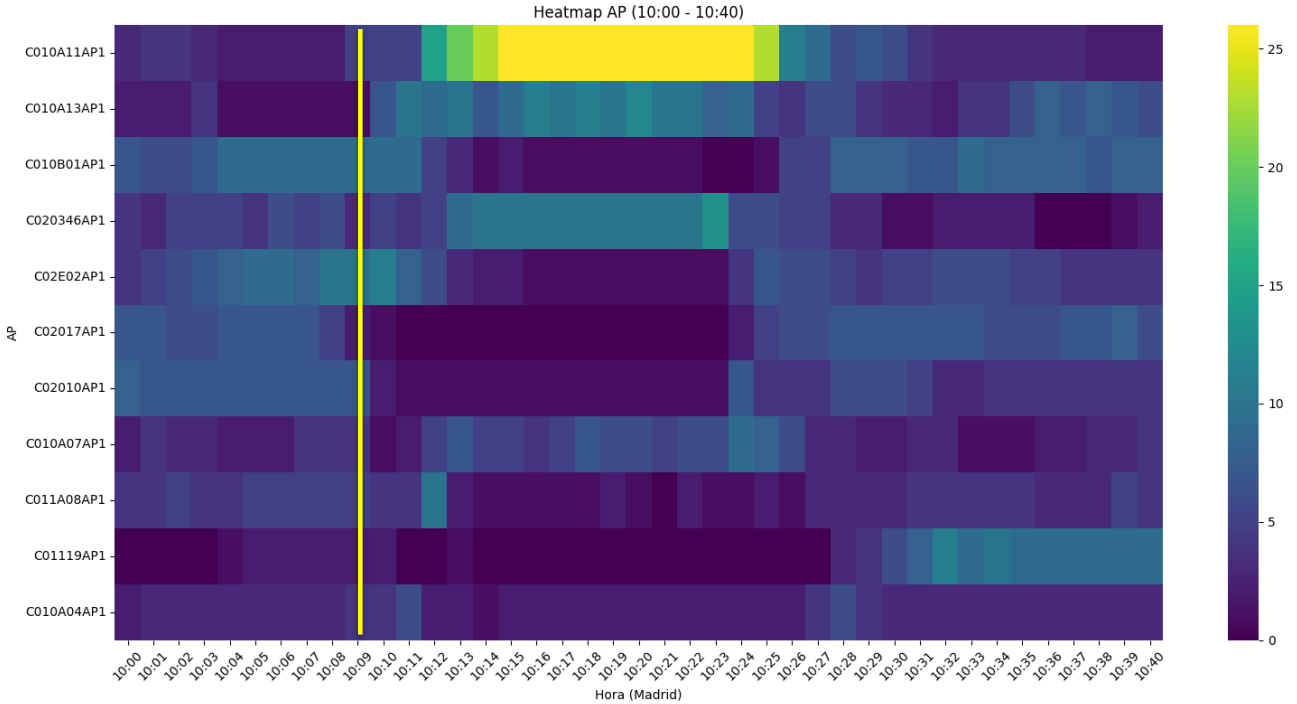
User distribution



Use Case: Fire drills (4/year)

- Colmenarejo campus (evacuation starts at 10:09)





Use Case: Class attendance

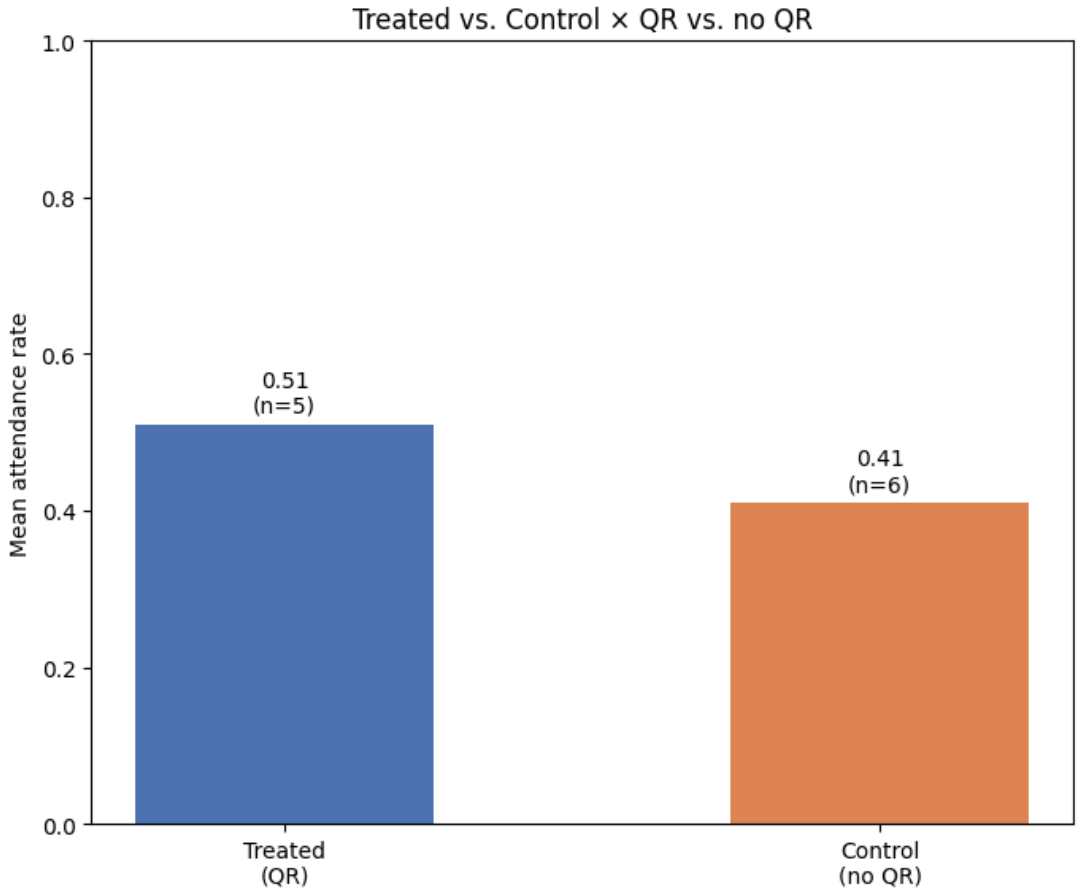
uc3m
Vice-Rectorate for Undergraduate Studies and Quality in Studies

- Techniques to improve the learning process
  - By improving class attendance
  - 12 macro groups of microeconomics, divided in 2 sets
- Set A: treatment
  - Class attendance registered with QR
  - No impact on grading
- Set B: control
  - No treatment
- We analyze
  - Real attendance (using WiFi traces)
  - Grade distribution per group (learning outcome)



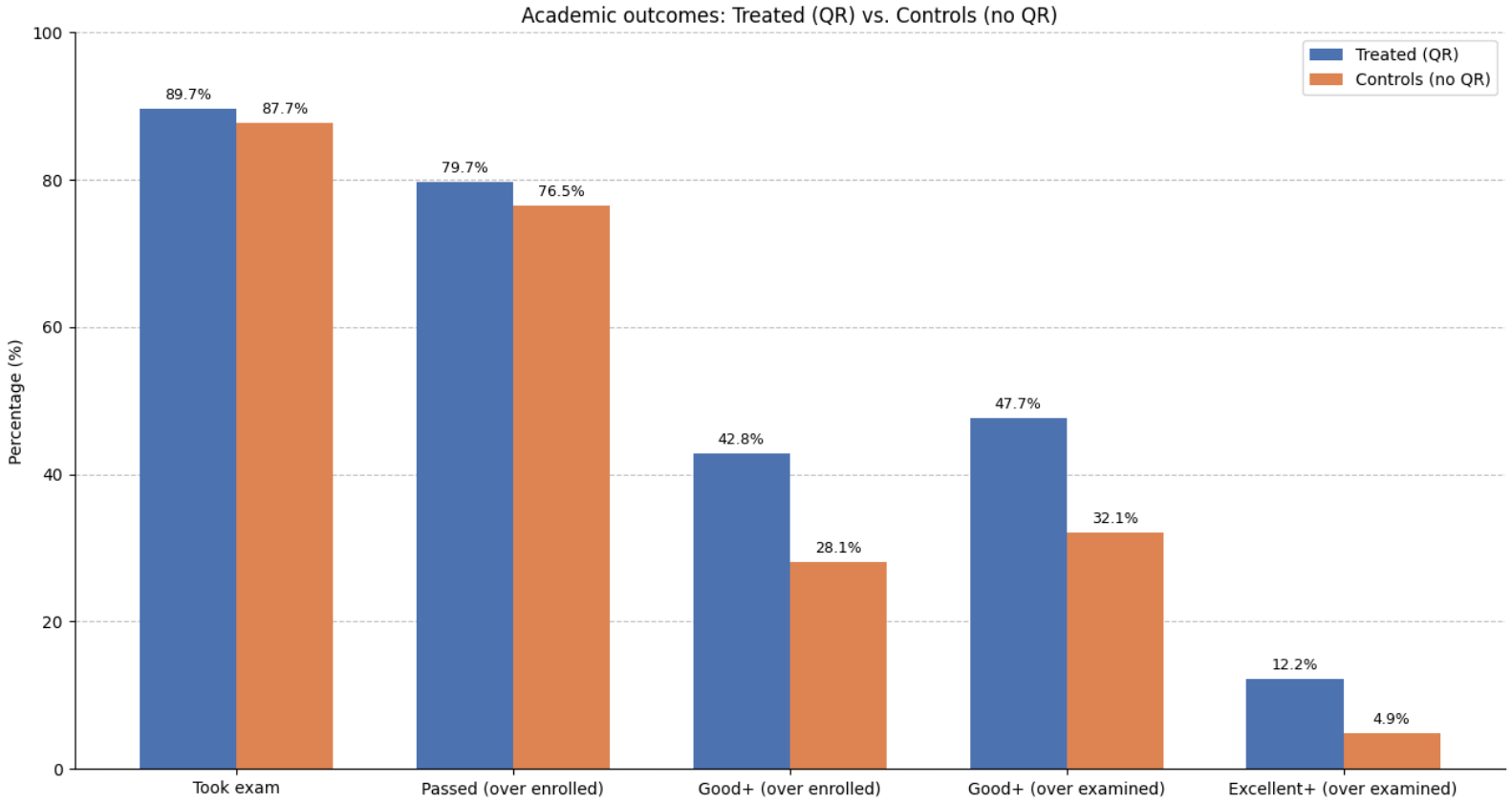
Attendance

- More attendance in Set A



Grades

- Better grades in Set A



PART 2: PRIVACY THREATS

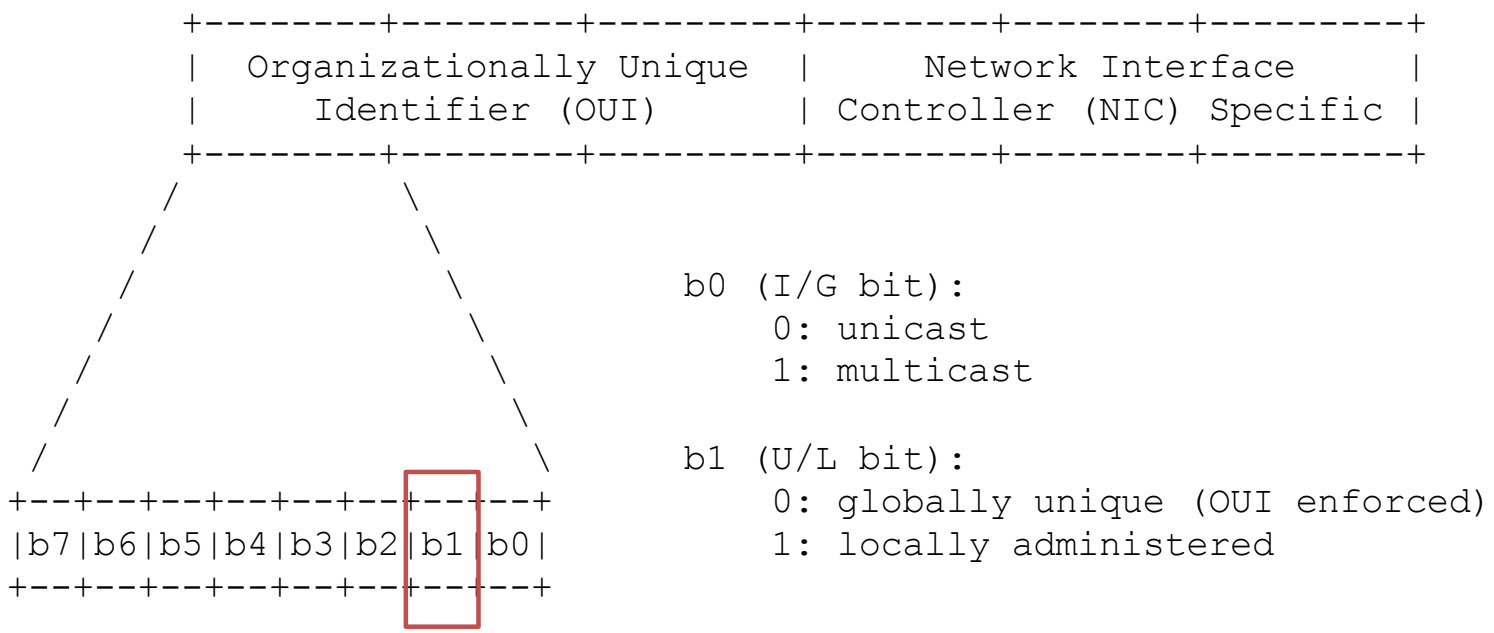
Privacy Risks in Mobile Networks

- Spatiotemporal traces are inherently identifying
  - Human mobility is regular and low-entropy: people return to the same places at predictable times [1, 2]
  - Four spatiotemporal points suffice to uniquely identify 95% of individuals in a large mobility dataset [1]

[1] de Montjoye, YA., Hidalgo, C., Verleysen, M. et al. Unique in the Crowd: The privacy bounds of human mobility. Sci Rep 3, 1376 (2013)
[2] Xu, F., Wang, Q., Moro, E. et al. Using human mobility data to quantify experienced urban inequalities. Nat Hum Behav 9, 654–664 (2025).

Random MAC address in 802.11

- IEEE 802.11aq (2018): randomize when unassociated (probes), generate another address to associate



RCM: Two possibilites

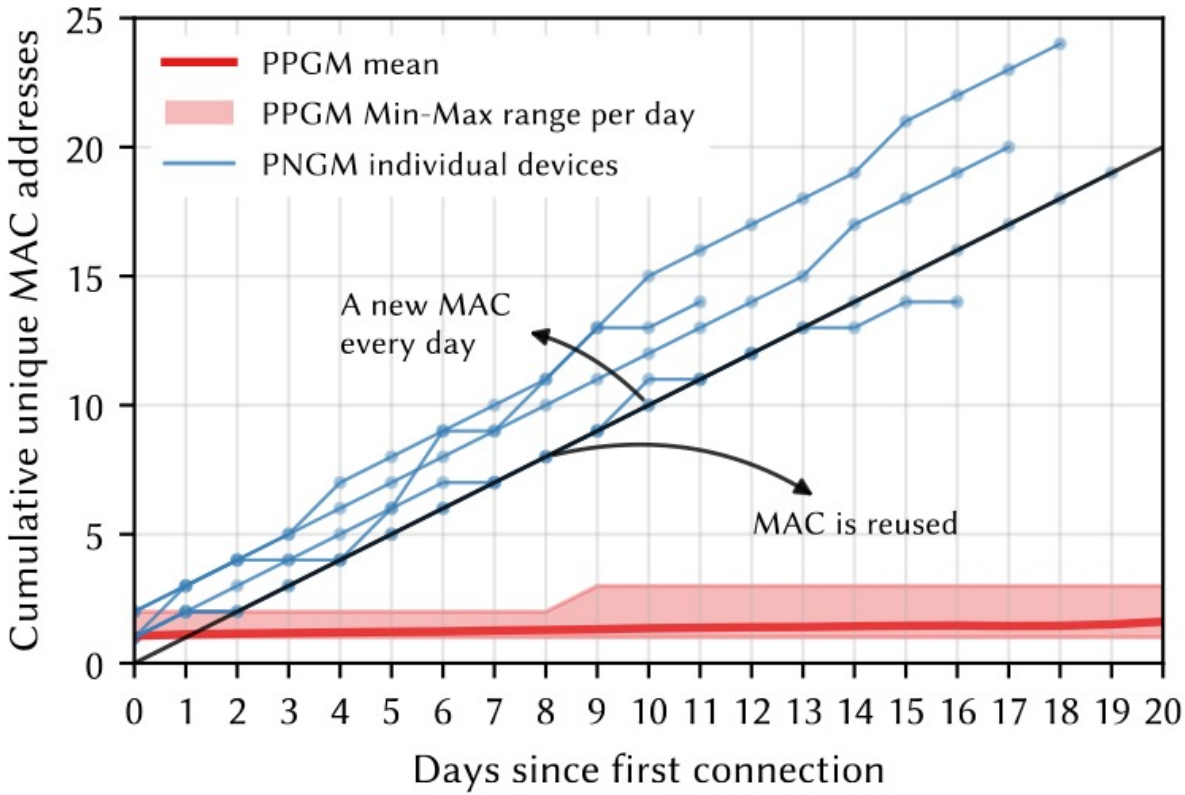
- Per-Network Generated MAC Address (PNGM)
  - One randomized MAC address per SSID
  - Default behavior in iOS, macOS, Android ≥ 10, Windows
- Per-Period Generated MAC Address (PPGM).
  - The device changes the address after a *sufficiently long* disconnection

Platform	Path	Options
iOS	Settings → Wi-Fi → [network] → Private Wi-Fi Address	Off / Fixed / Rotating
macOS	System Settings → Wi-Fi → Details → Private Wi-Fi Address	Off / Fixed / Rotating
Windows	Network & Internet → Wi-Fi → [network] → Random hardware addresses	Off / On / Daily
Android	Settings → Network & Internet → Wi-Fi → [network] → Privacy	Fixed / Randomized / Changing*

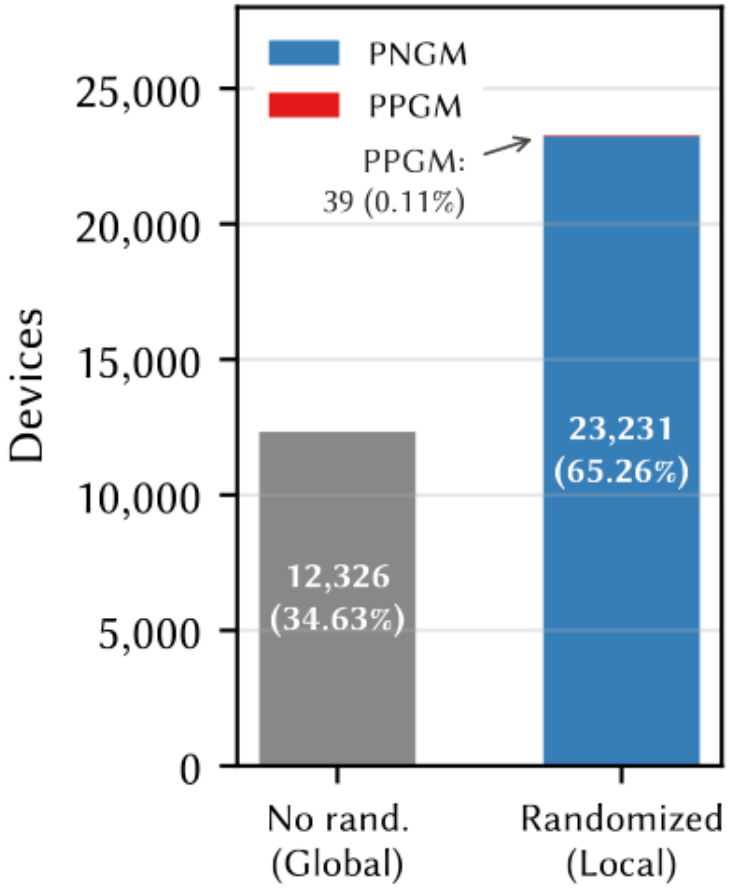
CURRENT STATUS OF MAC RANDOMISATION

Counting addresses

1. Analyze OUI field
   - Globally administered
   - Locally administered
2. For local addresses: count # addresses per user over time
   - PNGM: no change
   - PPGM: increase over time

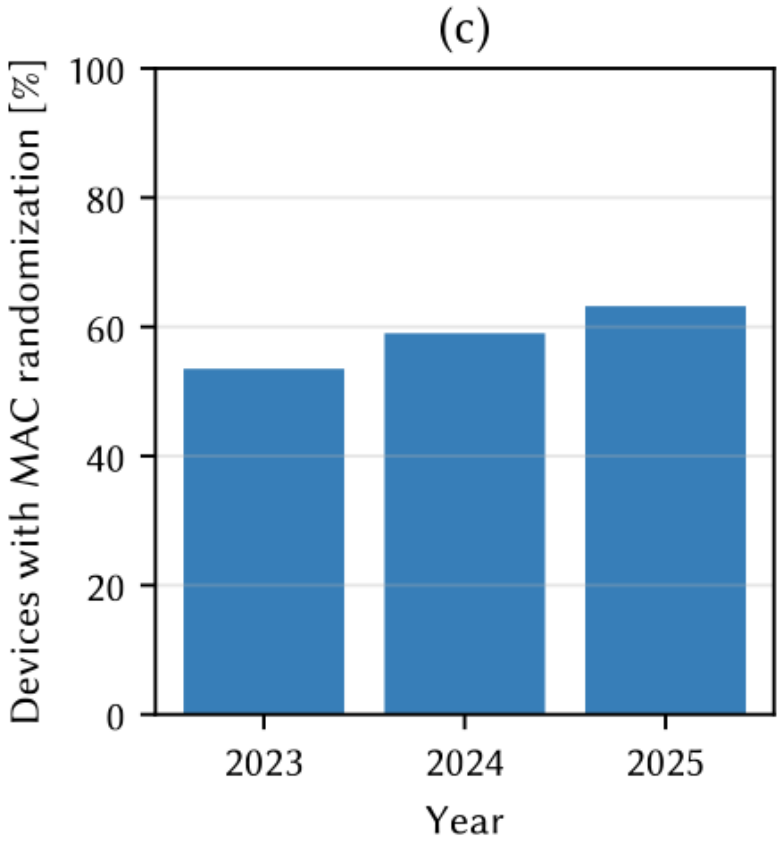


The status of MAC randomization



- One third: globally administered addresses
- Two thirds: one random address per SSID
- Only 39 devices use PPGM

PNGM more popular over time



- About 7% over the two-year observation period
- This may appear encouraging, but irrelevant from a privacy standpoint
  - Same MAC address

ATTACK MODEL AND RESULTS

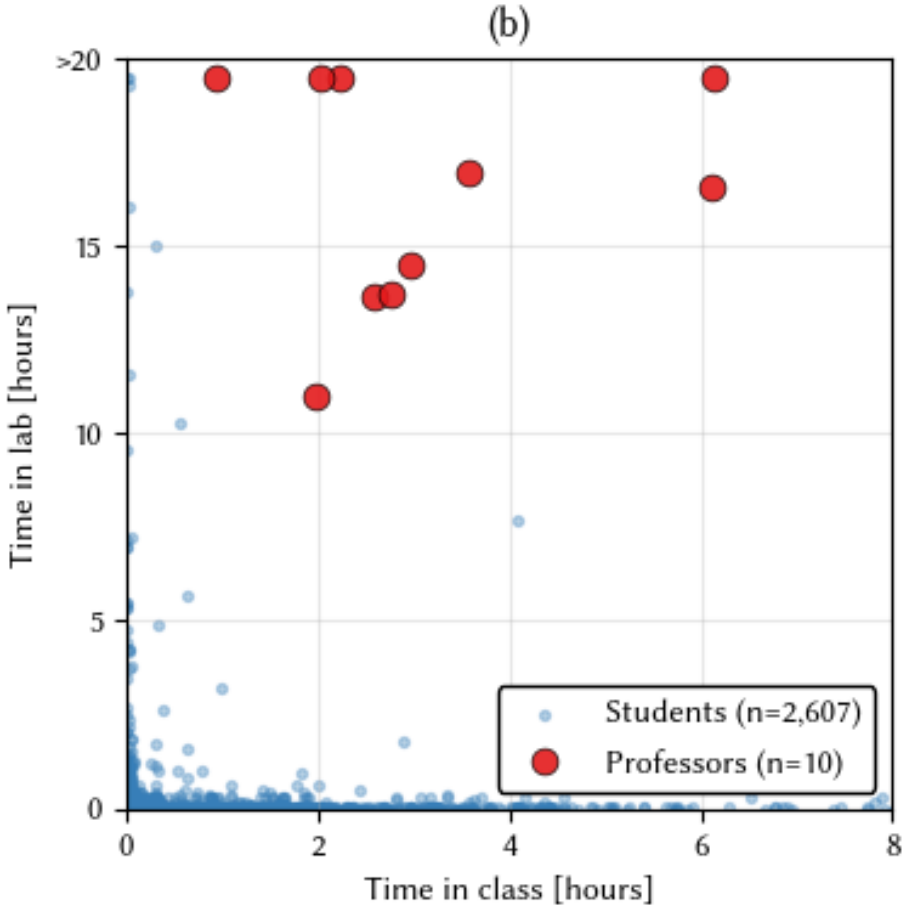
Threat

- Randomized MAC addresses should be un-linkable to physical persons (but for the sysadmin)
  - Under the GDPR, personal data is any information relating to an identified or identifiable natural person

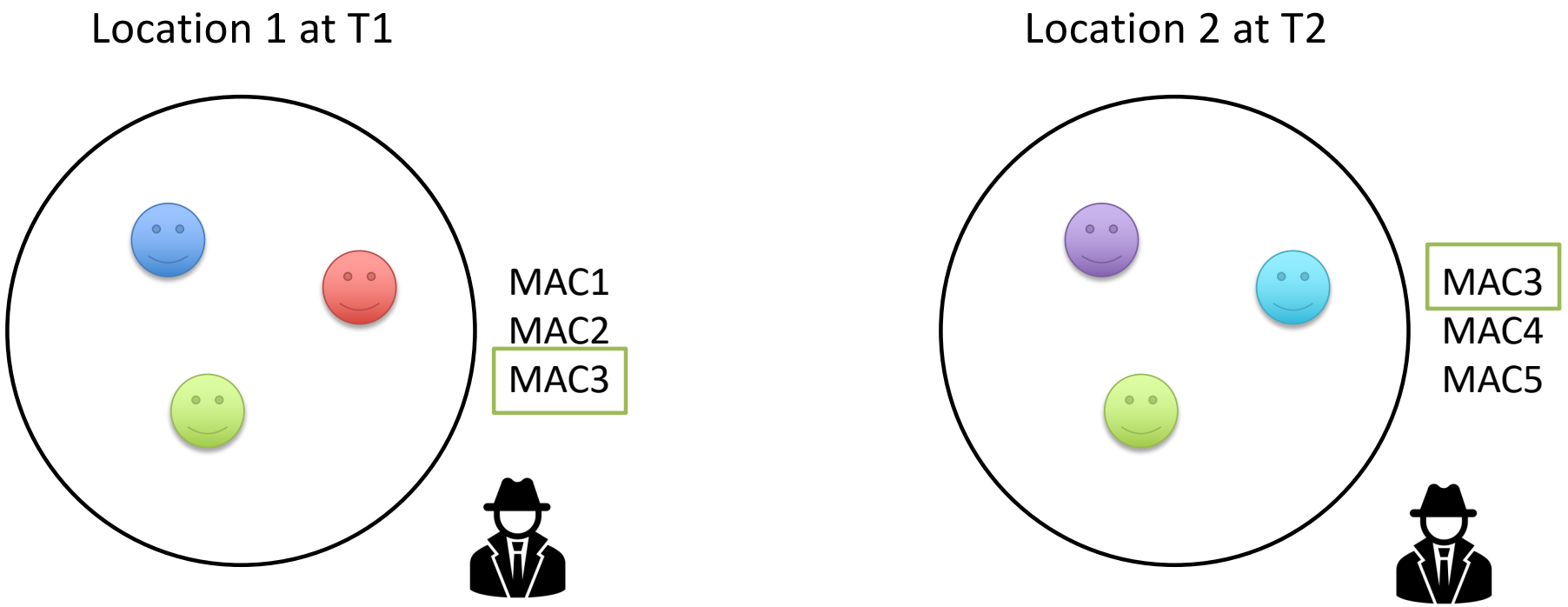
- How easy is to map humans to addresses if 99.89% of devices use the same address in the Campus?

De-anonymizing faculty MACs

- 10 **volunteers**
- Time connected to the office AP vs.
- Time connected to the lecture hall AP
- Over a week



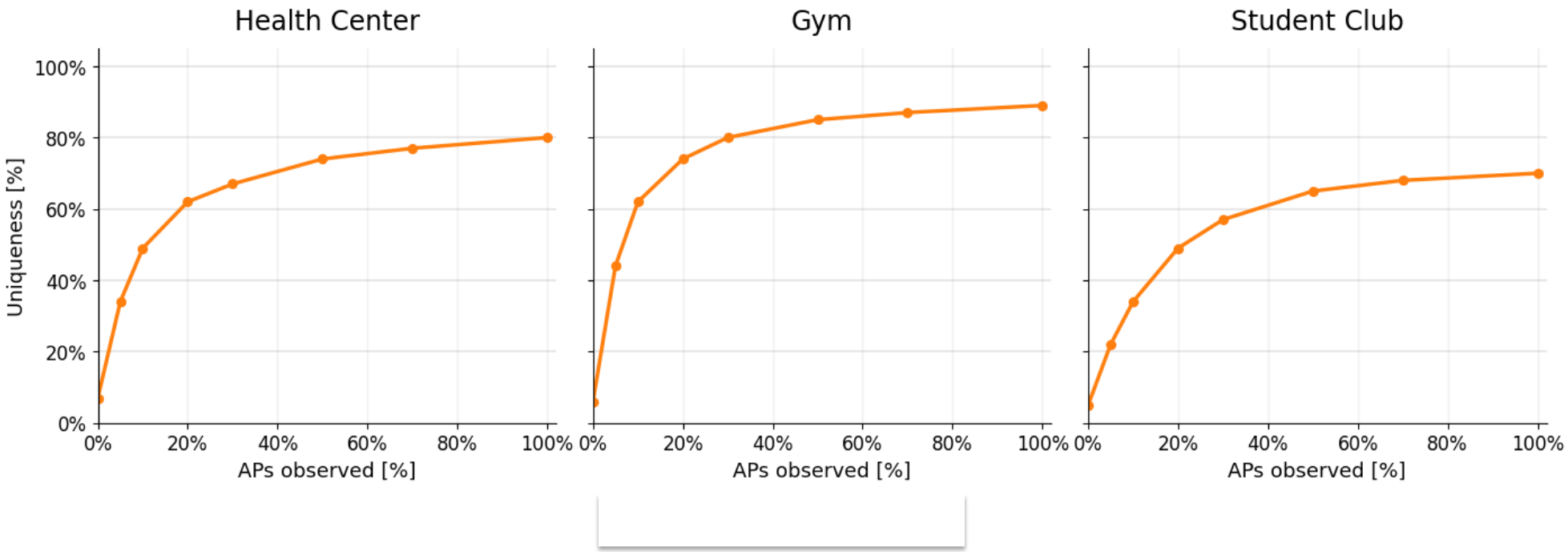
Attack Model - Device uniqueness



Device is **unique**: the only one with a given a combination of STPs

Sensible locations

- Sort APs by the number of devices not yet observed

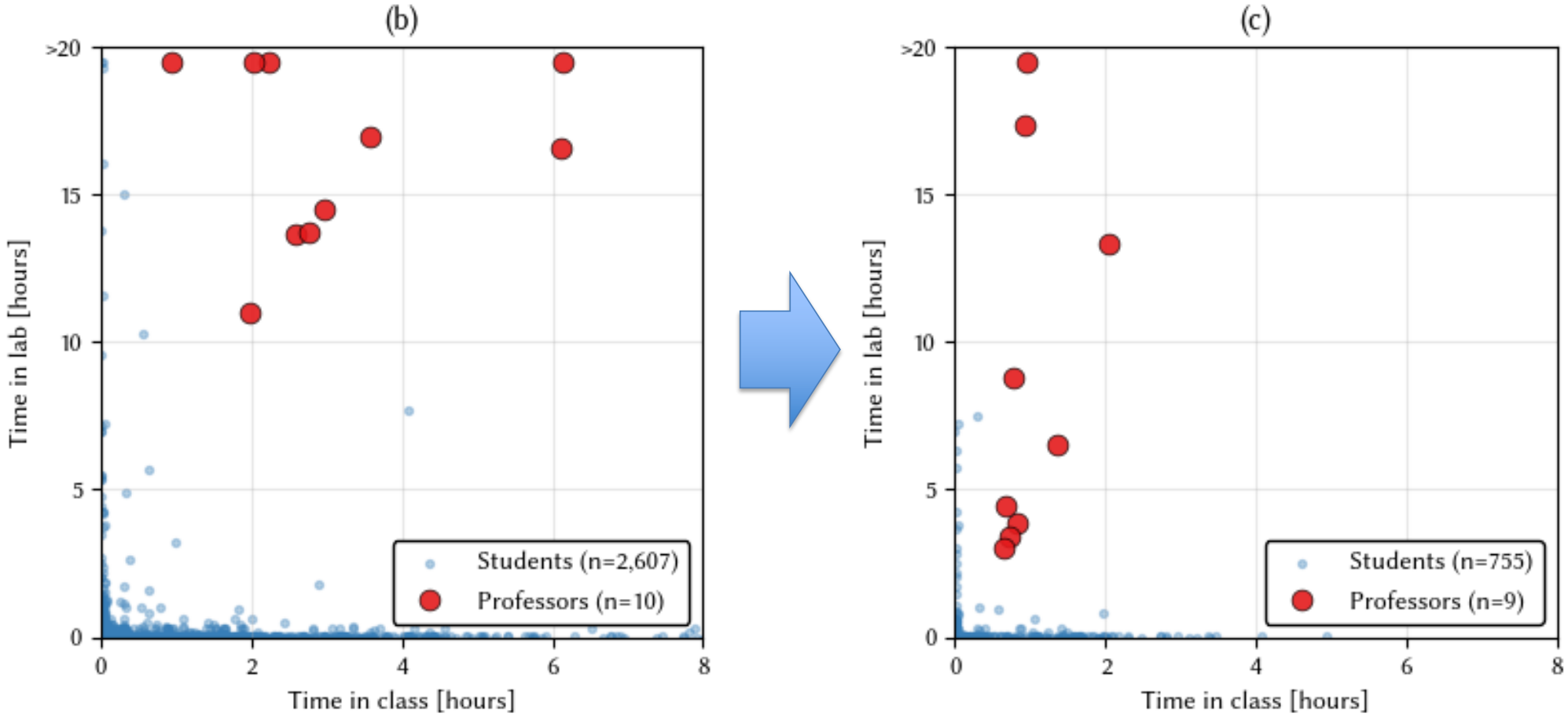


Benefits of PPGM?

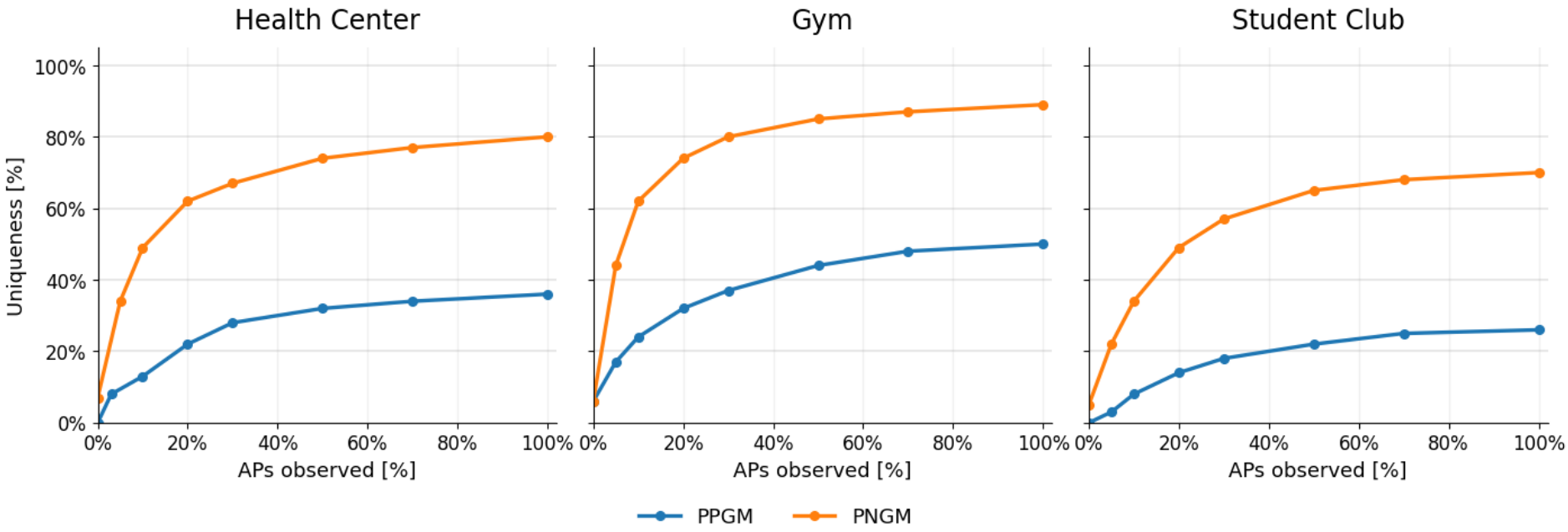
- Previous results with 99.89% of devices using Per-Network Generated MAC Address (PNGM).

- What if they used Per-Period Generated MAC Address (PPGM)?
  - Assumption: the device changes the address each day

Faculty – from 1 week to 1 day



Sensible locations – 1 day



Any location, all devices

- Similar to “unique in the crowd” (Four STP identify 95%)

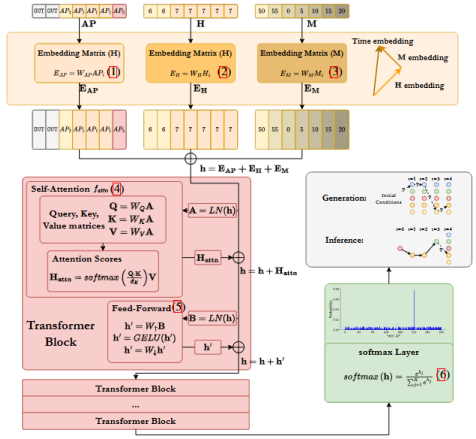
- In our case

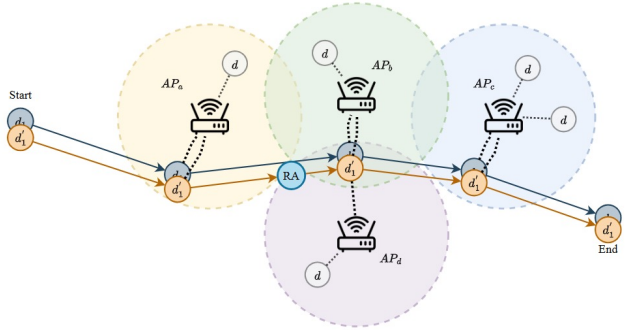
# STP (p)	1	2	3	4
% Unique	24.0	95.3	98.1	98.7

PART 3: DEFENCES

Defenses against Privacy Threats

- **Generation of synthetic traces**
  - Generate realistic traces that preserve utility while protecting privacy
  - Using transformers (LLMs)





Embeddings

Time embedding

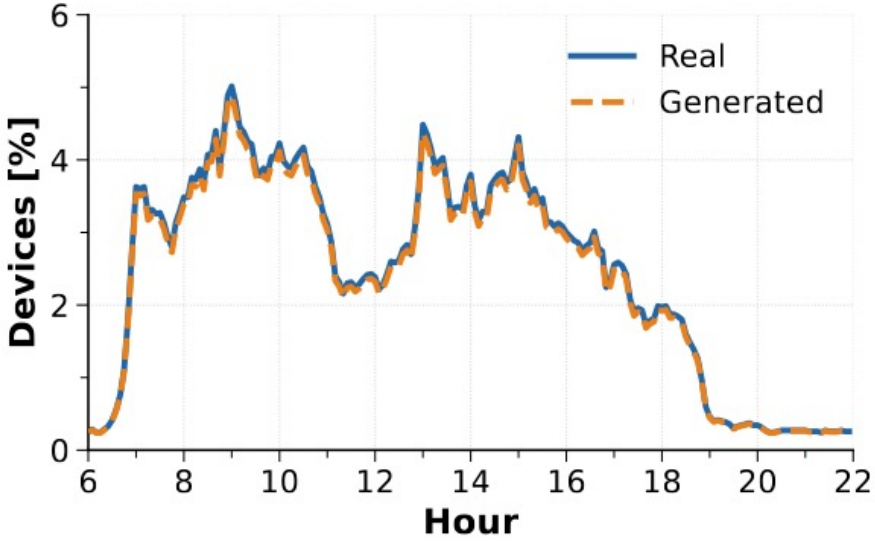
- Relative positions are OK in natural language
- Absolute positions are better for mobility
- Hour + Minutes
  - More scalable than minutes

Spatial embedding

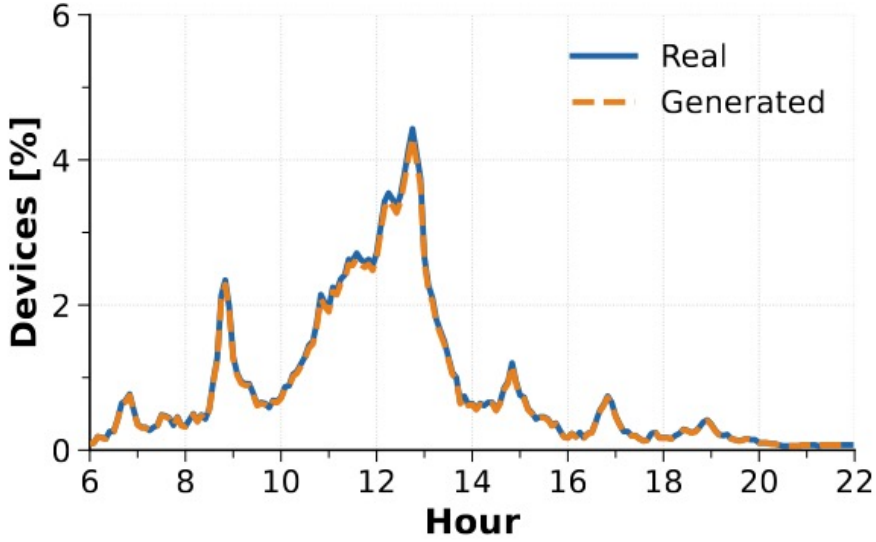
- Between 6 AM and 10 PM
- Discrete time, 5’ intervals
- 192 tokens
- Vocabulary
  - 279 Access Points
  - OUT

Results: HVAC systems

- Same performance as real traces



(b) Occupancy of Classrooms.



(c) Occupancy of Cafeteria.

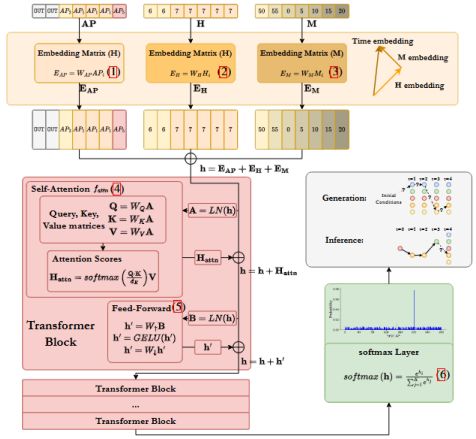
Privacy assessment

- Average probability of generating a real trace
  - How likely is the model to exactly reproduce a real trace?
- Closeness between real and synth dataset
  - Are synthetic traces similar to real ones without exposing snippets of specific devices?
- Membership inference
  - Can an attacker distinguish real traces from synthetic ones?

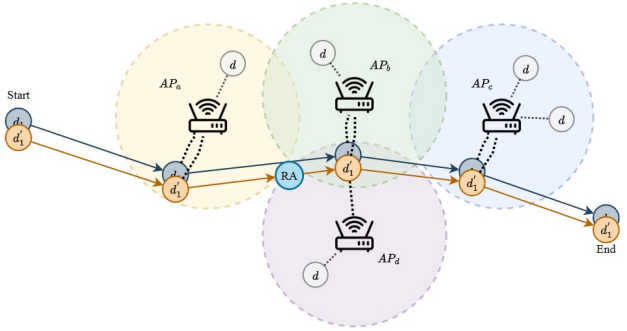
J. M. Montes-Lopez et al, “DiWi: A Transformer-Based Digital Twin for Wireless Mobility”, Elsevier Computer Networks, 2025

Defenses against Privacy Threats

- Generation of synthetic traces



- **Change client behavior**
  - Change randomization procedures so patterns become less identifiable



Two proposals to improve PPGM

- apPPGM: Mobility-based randomization
  - MAC address change when the device moves to a new AP
  - 802.11bh (2024): use a token to reconcile infrastructure continuity
- iPPGM: Interval-based randomization
  - The device changes its MAC address at regular intervals

Uniqueness

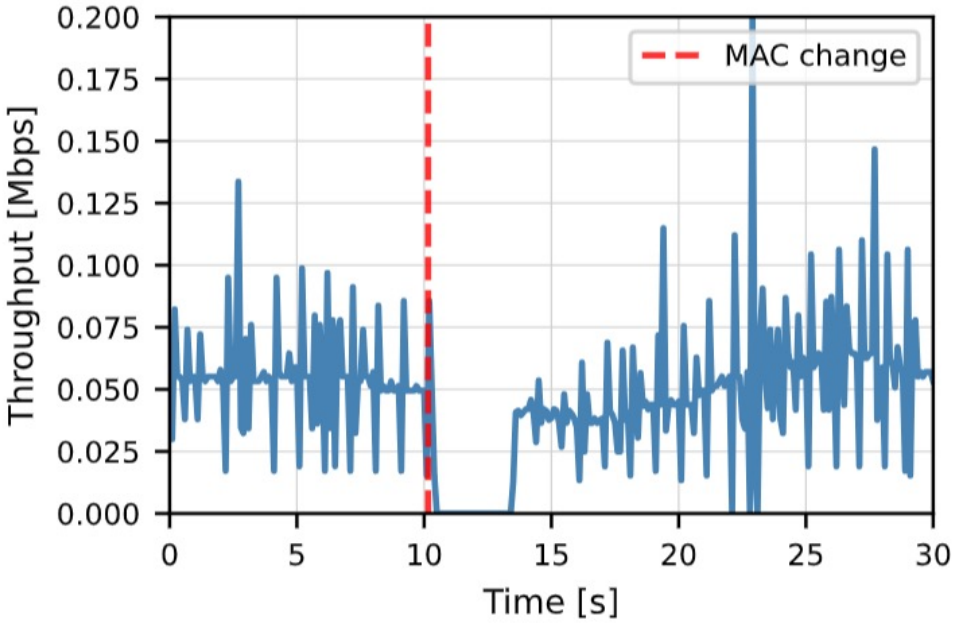
# STP (p)	1	2	3	4
PPGM (24h)	24.0	95.3	98.1	98.7

Implementation

Real-Time video

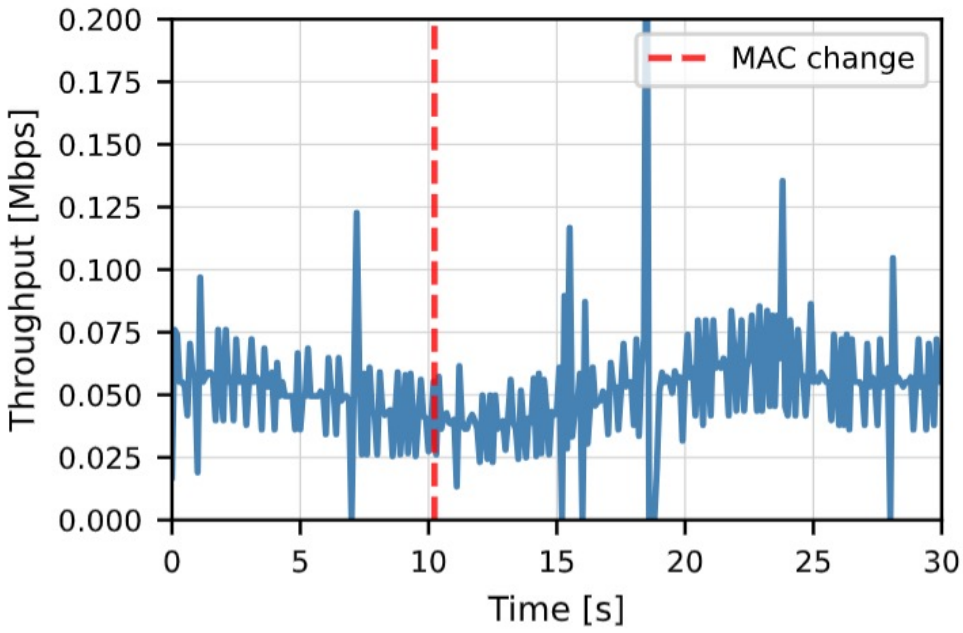
Single interface

- Interface goes down, changes MAC address, and reconnects.



Dual interface

- Traffic switches to one interface while the other rotates its MAC address.



CONCLUSIONS

TL; DR

Summary

- A campus WLAN is a continuous, passive recorder of human mobility
  - And this can be a good thing
- Deployed MAC address randomization offers no protection in these networks
  - Even when properly configured
- A better defense is possible (while we wait for 802.11bi)

Open Questions



- When is surveillance ethical, who decides?
  - WiFi tracking could save lives, occupancy data is useful
  - But we could also e.g. profile students by time @ library
- Does “informed consent” apply here? User awareness?
  - GDPR: accept or decline per service (vs. per Campus?)
- What happens when the network becomes an active sensor (JCAS / ISAC)?

The Trace You Leave Behind — and What We Can Do About It

Questions?

Pablo Serrano
pablo@it.uc3m.es

IMDEA Networks annual workshop 2026

uc3m | Universidad Carlos III de Madrid

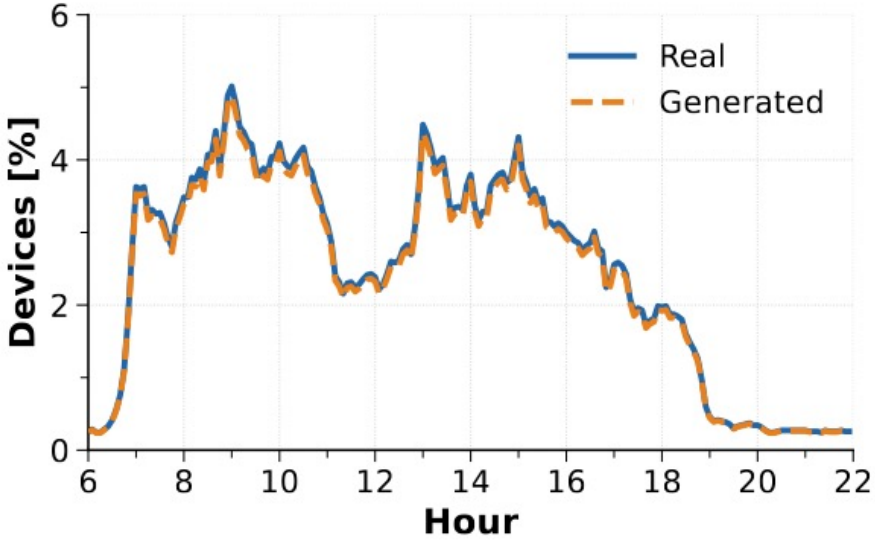
BACKUP SLIDES

Why?

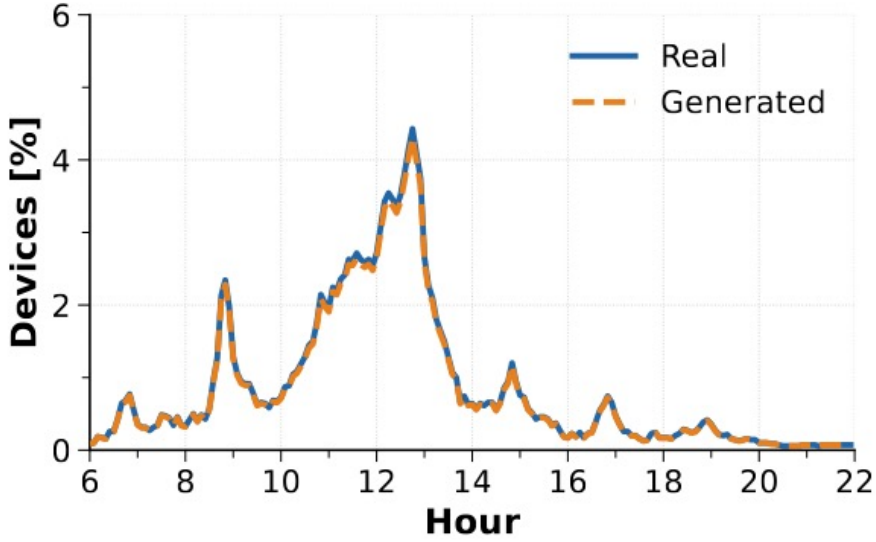
# STP (p)	1	2	3	4
PPGM (1 d)	9,810 / 9,810	9,750 / 9,810	9,682 / 9,810	9,610 / 9,810
PPGM (6 h)	19,300 / 19,300	18,725 / 19,300	18,146 / 19,300	17,547 / 19,300
PPGM (15 min)	150,011 / 150,011	65,732 / 150,011	41,290 / 150,011	31,545 / 150,011
PPGM (5 min)	217,644 / 217,644	63,635 / 217,644	39,515 / 217,644	28,082 / 217,644
pAPGM	295,166 / 295,166	32,875 / 295,166	22,243 / 295,166	16,108 / 295,166
PPGM+pAPGM (1 h)	314,481 / 314,481	45,462 / 314,481	28,873 / 314,481	17,164 / 314,481
PPGM+pAPGM (15 min)	377,857 / 377,857	37,032 / 377,857	5,045 / 377,857	0 / 377,857
PPGM+pAPGM (5 min)	419,934 / 419,934	0 / 419,934	0 / 419,934	0 / 419,934

Use Case 2: HVAC

- Heating, Ventilation, and Air Conditioning.

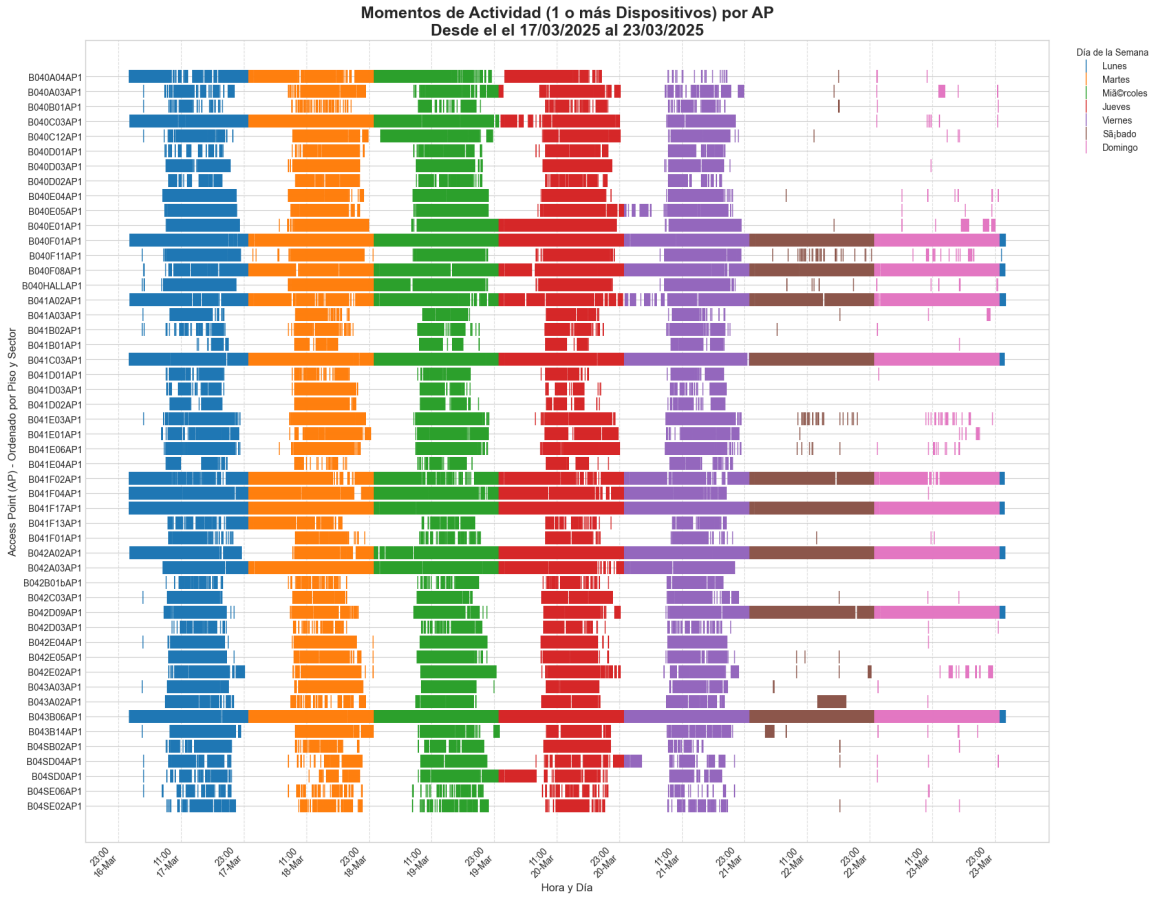


(b) Occupancy of Classrooms.

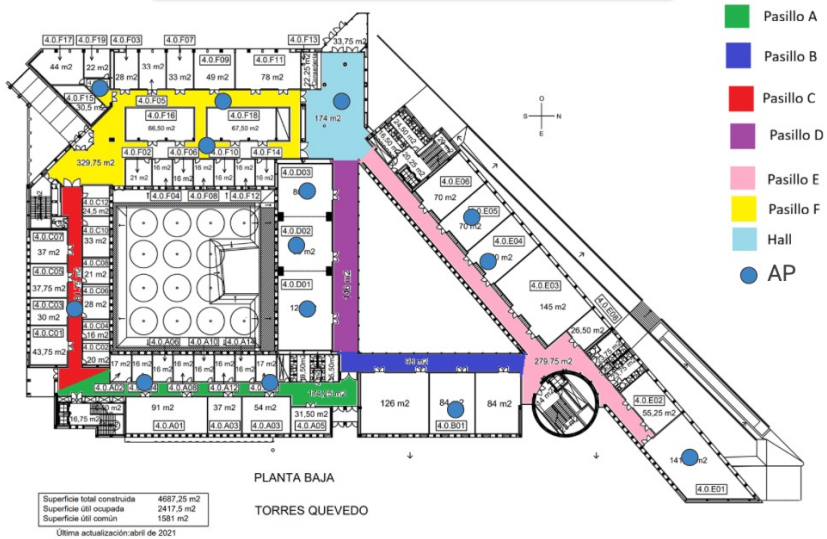


(c) Occupancy of Cafeteria.

Building 4: AP activity (> 0 STAs)



Designing thermal zones



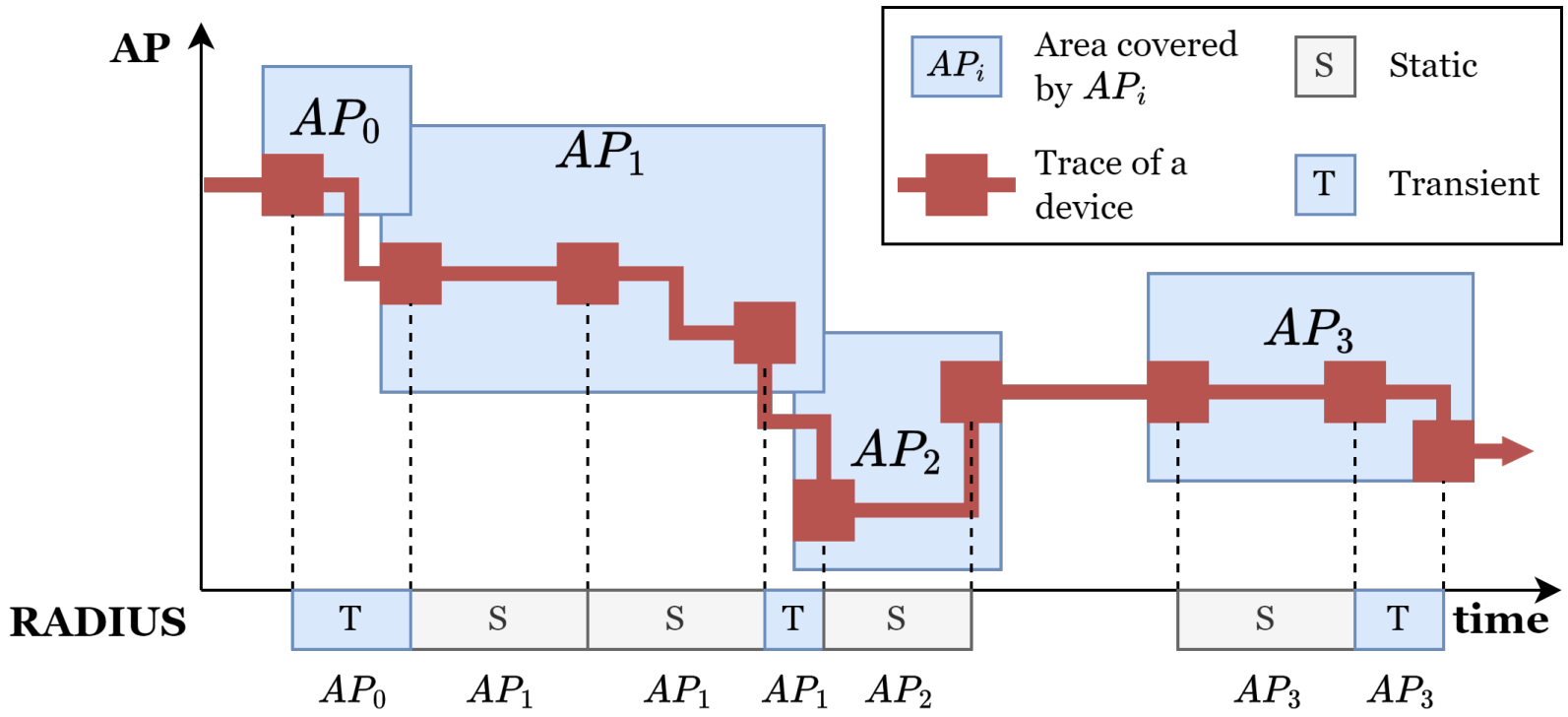


Maximum savings ~ 66%

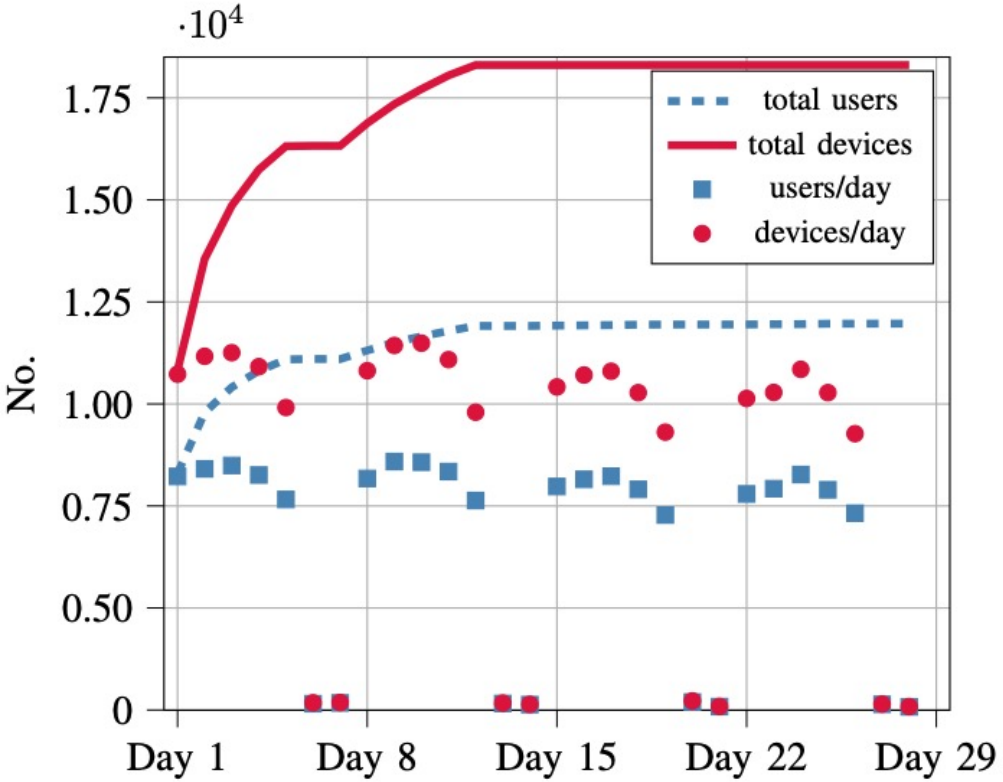
Adversary Model

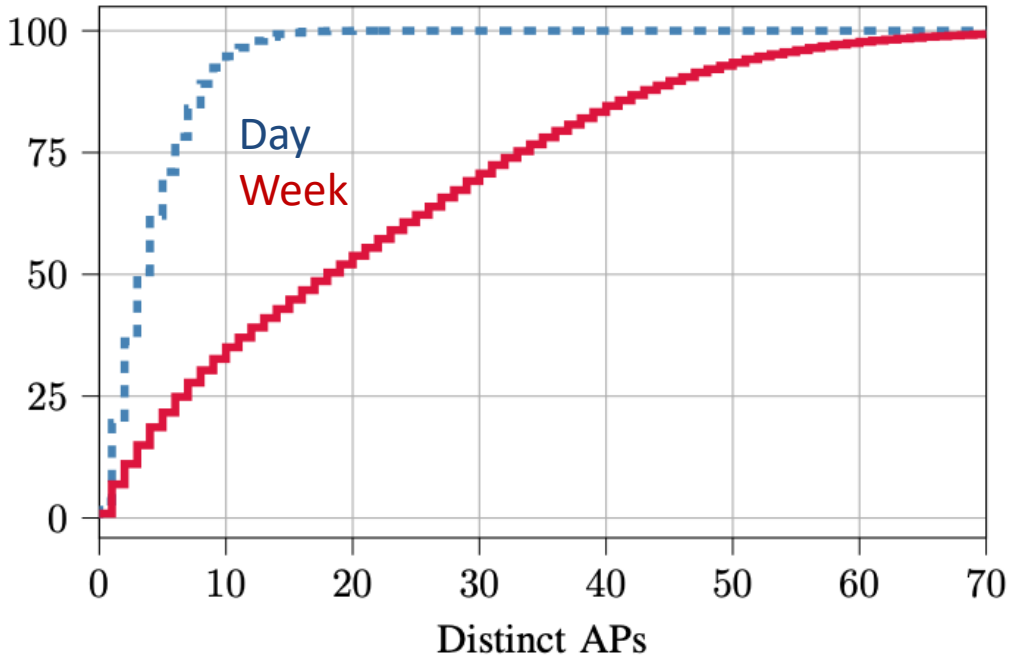
- Attacker: can observe spatiotemporal points (STPs), i.e., tuples of (MAC address, AP identifier, timestamp)
  - E.g., monitoring selected locations
- Exploit the fact that the observed MAC addresses are effectively persistent campus-wide identifiers (99.89%)
- We do not require full network coverage, only on observations from a subset of locations.
- Objective: find at least one “unique trait” of a device

Human Fingerprint



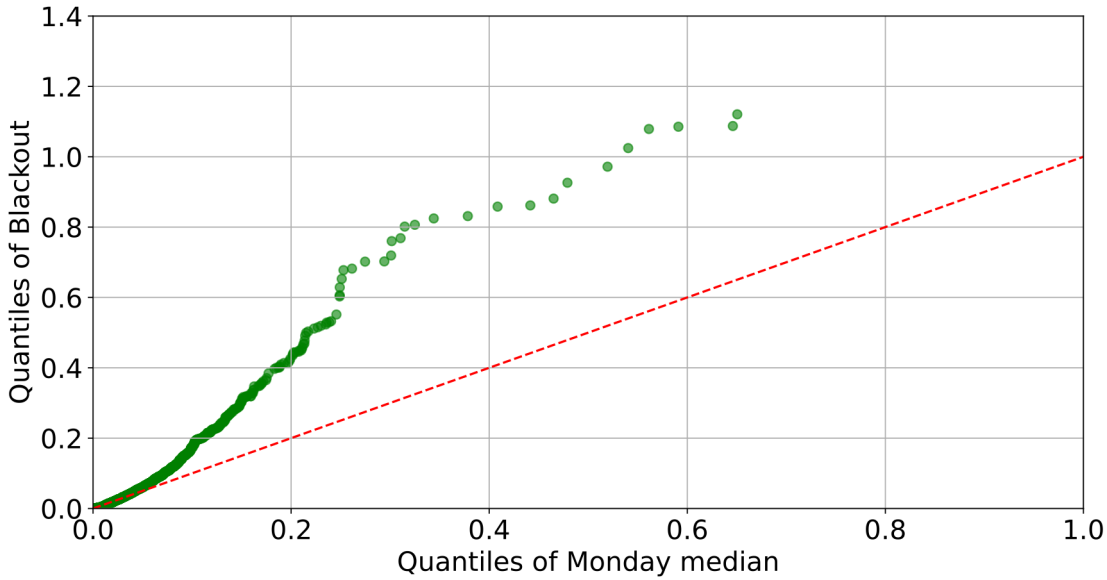
Some statistics

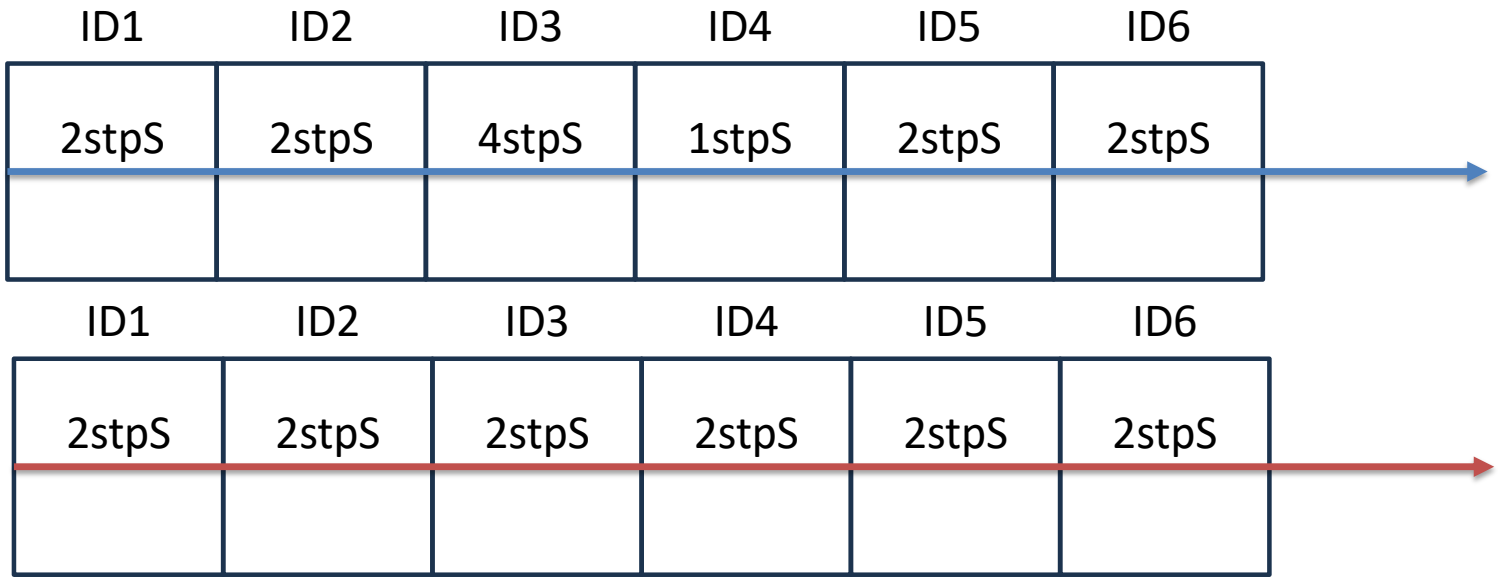




Generated traffic

- For devices present during the blackout and other Mondays
- Blackout traffic vs. median Monday traffic (Q-Q plot)





p	1	2	3
ELEGIBLE IDs	12/12	11/12	1/12
ELEGIBLE DEVICES	2/2	2/2	1/2

Como calculo ambos.
Elegible ids es el de al tabla y elegible devices es el de la figura

Privacy Risks in 802.11 Networks

- Additional attacks on proble requests
  - Unassociated devices actively broadcast Preferred Network Lists (PNLs) in probe requests, revealing previously joined SSIDs
  - PNL contents expose social links: shared or rare SSIDs can identify a device's owner and infer who they will meet [3]
  - Probe request templates (IEs, timing, periodicity) enable device fingerprinting even under MAC randomization [4, 5]
  - Passive roadside sniffing alone detects thousands of unique devices in hours and reconstructs trajectories [6]

[3] M. Cunche et al. "I know who you will meet this evening! Linking wireless devices using Wi-Fi probe requests," IEEE WoWMoM 2012
[4] D. Vogel et al. "I Know you were here: Leveraging Probe Request Templates for Identifying Wi-Fi Devices," WiMob 2024
[5] J. Qi and A. Sugiura, "An Improvement for Detection of Wi-Fi Device Utilizing Random MAC Address by Focusing on Inter-arrival Time and Periodicity of Probe Request Frames," IEEE GCCE 2024
[6] M. Cunha, R. Mendes, Y. -A. de Montjoye and J. P. Vilela, "WiFi-Based Location Tracking: A Still Open Door on Laptops," in IEEE OJCS, 2025

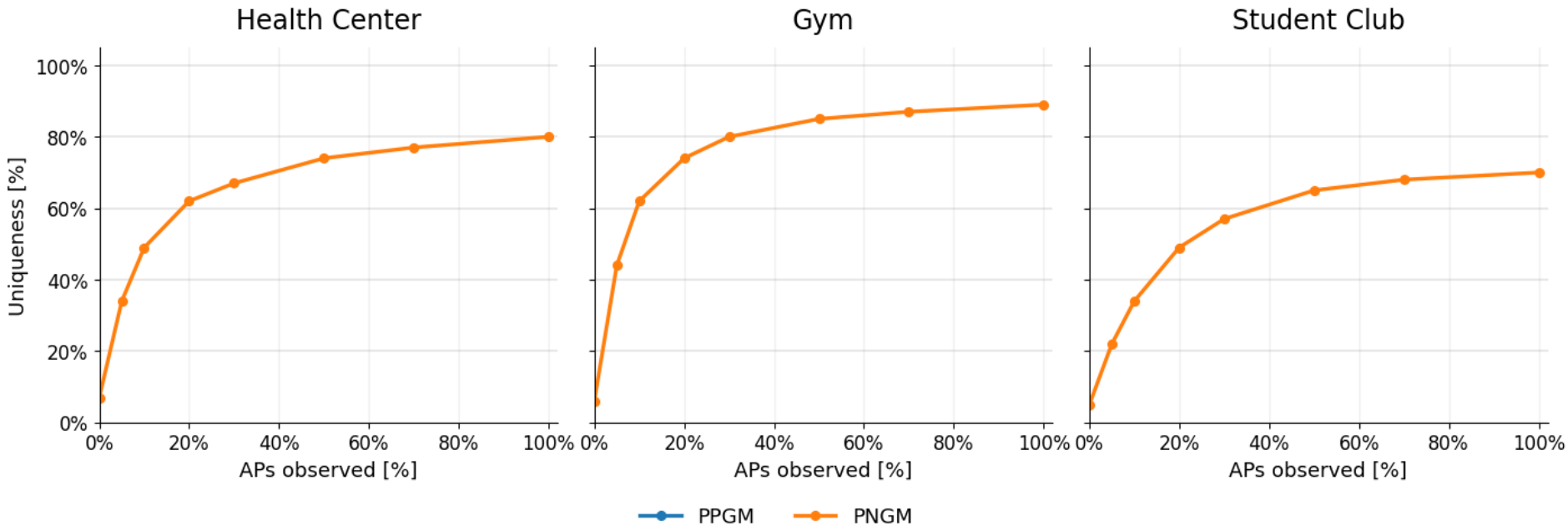
The 802.11 standardization response

- TODO [Local vs. global address]
- IEEE 802.11aq (2018): randomize when unassociated, then generate a random address
  - Per-SSID (Networks) or not
- 802.11bh (2024): use a token to reconcile infrastructure continuity

- 802.11bi (ongoing): advanced frame anonymization and encryption.

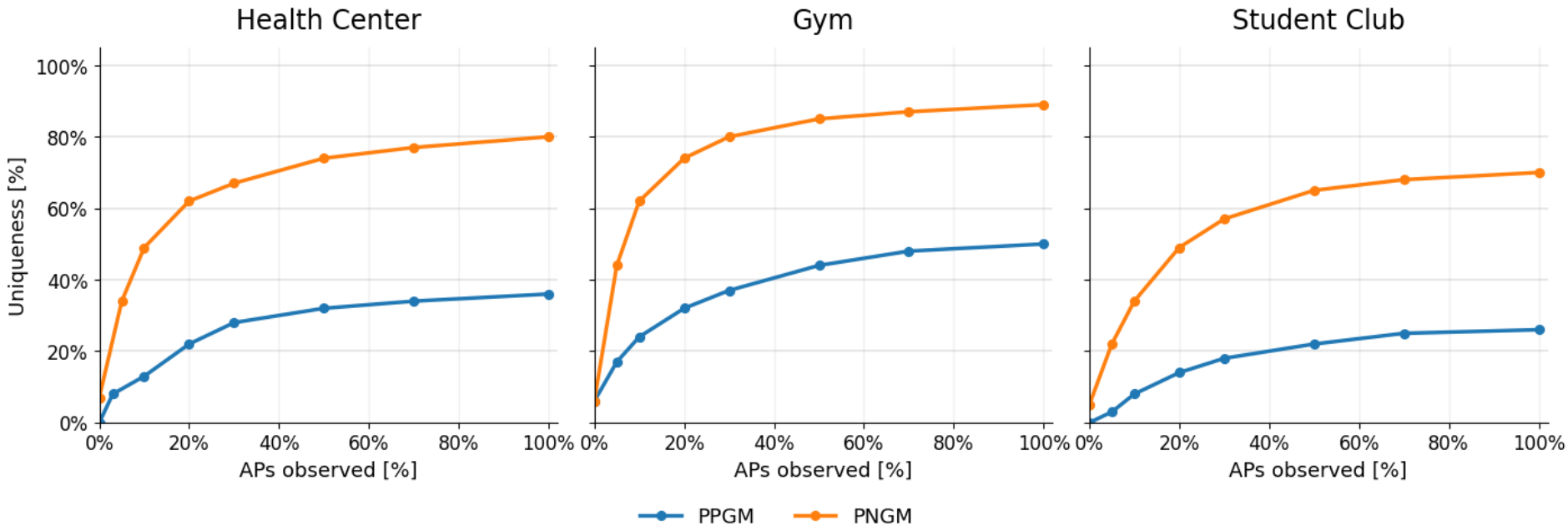
Any device – 1 fixed AP

- Sort APs by the number of devices not yet observed



Any device – 1 fixed AP

- Sort APs by the number of devices not yet observed



Use Case 1: April 28 Blackout

- 28 April 2025: Iberian Peninsula's power failure
  - Mobile networks collapsed
  - UC3M campus WiFi remained (mostly) operational
- Four questions we can now answer
  - Do people connect to eduroam?
  - How did campus occupancy evolve during the blackout?
  - Did per-user network intensity increase?
  - Did users redistribute spatially across zones?

Identifying the mobile device

- We define each device's mobility score as a weighted sum of mobility features x_i that capture relevant mobility traits:

$$s = \sum_{i=1}^{N} \omega_i x_i$$

- The cost function learns the weights by making each user's highest-scoring device stand out clearly from their other devices.

$$\max_{\omega \in \Delta^3} \sum_{k=1}^{K} \left(\max_i s_{ki}(\omega) - \frac{1}{N_k} \sum_{j=1}^{N_k - 1} \max_{j \neq \arg\max_i s_{ki}} s_{kj}(\omega) \right)$$

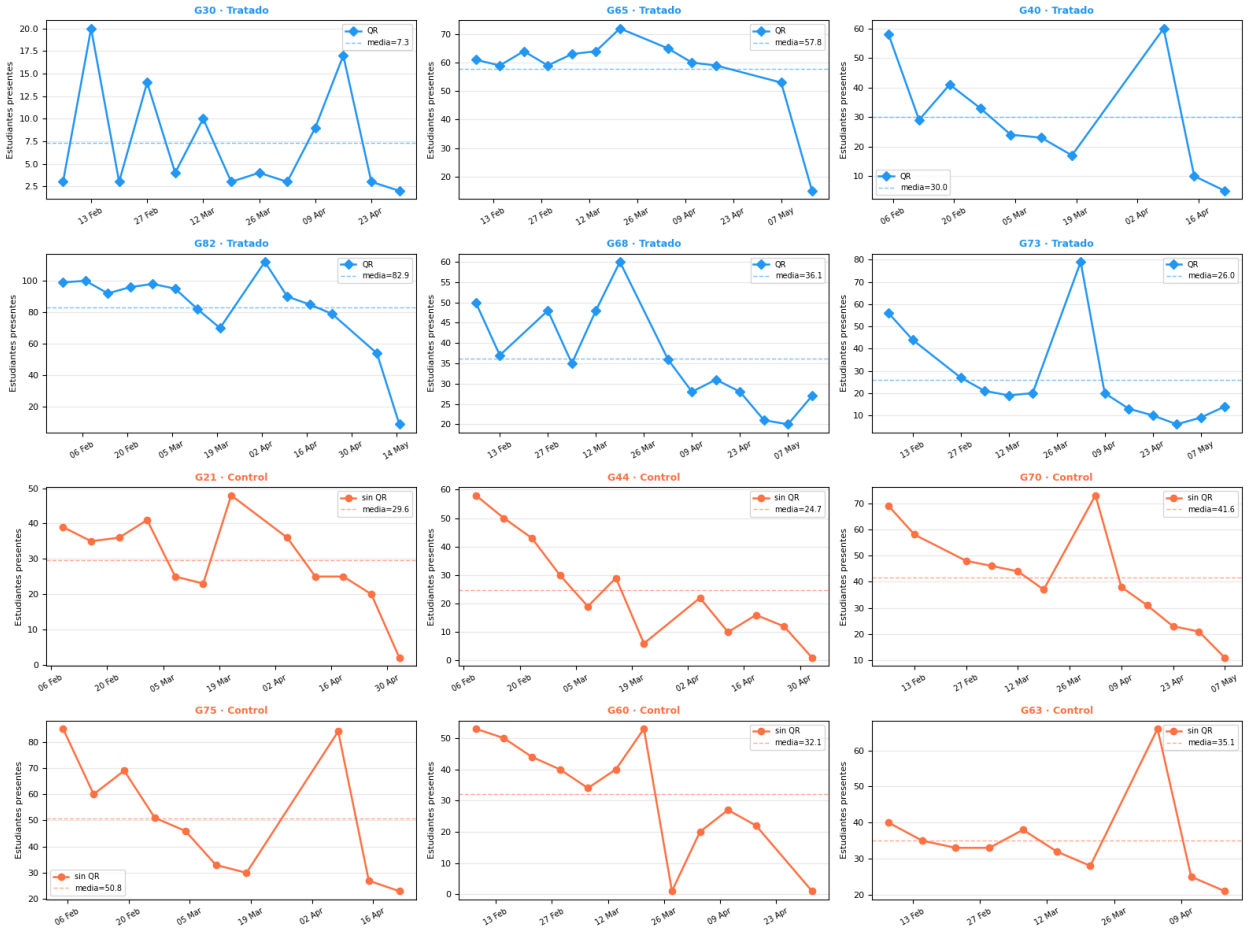
Estimating real attendance

- Alumno(s) líder: N o más días conectados a eduroam al menos una vez durante el horario de clase.

- Puntos de acceso de clase: aquellos en los que los alumnos líder pasan más tiempo conectados en horario de clase.
- Hasta un total de T tiempo acumulado (para no sumar todos)
- Se comprueba manualmente que están en la proximidad de clase

- Asistencia (para alumnos de la lista de clase): si pasa al menos C tiempo conectado a algún punto de acceso de clase durante el horario de clase

Tratamiento vs. control



Results

- 10% increase in attendance
- Better grade distribution
  - Grades 9-10/10: + 7% improvement
  - Grades 7-8/10: + 15% improvement

- The effect must be statistically very robust, as it emerges despite a small sample size
  - Unavailability of individual data

Next steps

- El estudio de posibles mejoras significativas en la docencia puede tener réditos importantes para la institución.

- Se puede medir el impacto de políticas que aumenten la asistencia a clase, del tamaño del grupo, de los horarios de clase de exámenes, de las metodologías docentes.
- Por ejemplo, cuando se hacían asignaciones a grupo por apellido (por tanto, aleatorias), y había tamaños muy diferentes de grupo, se pudo comprobar que variar tamaños de clase de 100 a 20 en magistral, o de 40 a 10 en reducido tenían efecto nulo en desempeño.
- También que los grupos de tarde tenían peor desempeño que los de mañana.

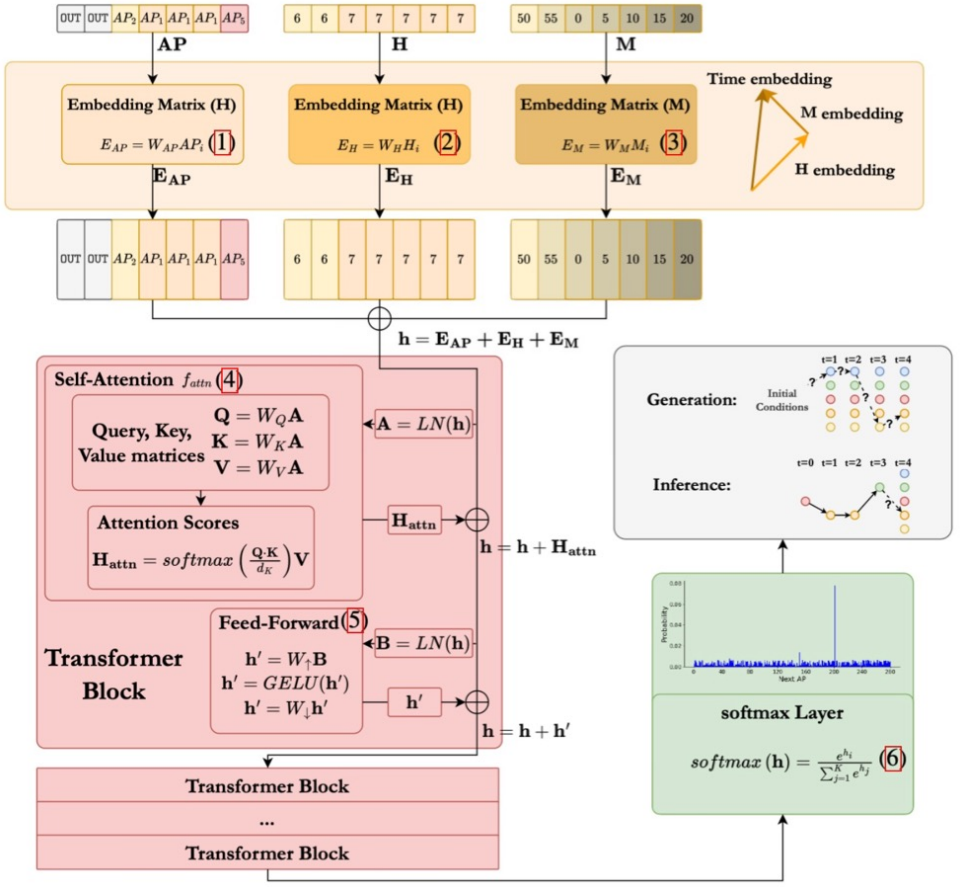
- Para desarrollar proyectos parecidos es necesario crear un cargo con autoridad para solicitar datos individuales de estudiantes a las unidades administrativas, un uso completamente legítimo según la ley / DPD.

Why transformers (LLMs)

- Goal: model the activity of users connected to the network
- Sequences of discrete spatiotemporal points
- Large Language Models (LLMs) learn from sequential data
- LLMs rely on a specific encoding of tokens (APs) and its relative position in the sequence.
  - These embeddings are then processed through self-attention layers, which learn the relationships between elements across time and space.

DiWi: overall architecture

- Time embedding
  - Absolute time
- Spatial embedding
  - OUT token
- Self-attention layers



Privacy Risks in 802.11 Networks

- Probe requests leak identity even before association
  - Unassociated devices actively broadcast Preferred Network Lists (PNLs) in probe requests, revealing previous SSIDs
  - PNL contents expose social links: shared or rare SSIDs can identify a device's owner and infer who they will meet [3]

[3] M. Cunche et al. "I know who you will meet this evening! Linking wireless devices using Wi-Fi probe requests," IEEE WoWMoM 2012

Formal Guarantees

- **Differential privacy**
  - Differential privacy ensures that the information inferred about any individual remains almost unchanged, whether or not they are included in the dataset.
  - In machine learning, this is achieved by adding controlled noise to the gradients during training.

