



Universitat
de les Illes Balears

Analysis of wireless network access logs for a hierarchical characterization of user mobility (summary)

Francisco Talavera, Isaac Lera,
Carlos Guerrero



carlos.guerrero@uib.es



<https://acsic.uib.es/cguerrero/>



@siesinformatica



ARTICLE INFO

Article history:
Received 9 November 2021
Revised 8 March 2022
Accepted 11 March 2022
Available online 29 March 2022

Keywords:
User mobility
Mobility modeling
User behavior simulation
Fog computing

ABSTRACT

This paper presents a method that generates a hierarchical user mobility model from the analysis of the data available from Wi-Fi connections. The data obtained from the Wi-Fi infrastructure is defined in terms of the coverage areas of the access points that the users move through. These access points are recursively grouped into different levels of granularity based on their geospatial features. The track of a user is defined as a sequence of Wi-Fi access points, which is enough to simulate user mobility in, for example, fog scenarios. The hierarchical definition of the region under study is proposed to reduce the complexity of the model in high-scale scenarios and to increase the adaptability between scenarios with different geospatial features. The model creation is based on a user profiling method that uses a clustering algorithm and each user type is defined with a transition matrix between coverage areas and a time length vector for the areas. The method is applied to the case of the campus of the University of the Balearic Islands. From the analysis of the mean square error of the results, we deter-

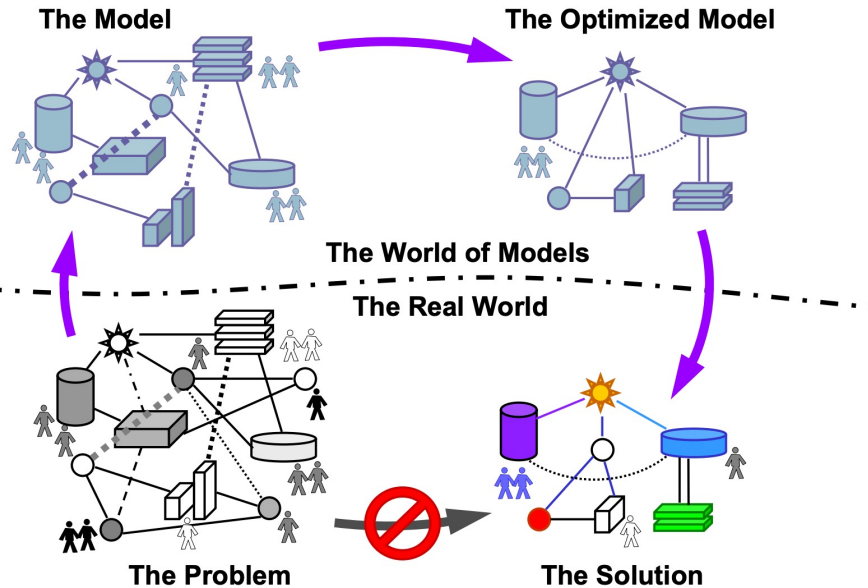
Proyecto PID2021-128071OB-I00 financiado por:



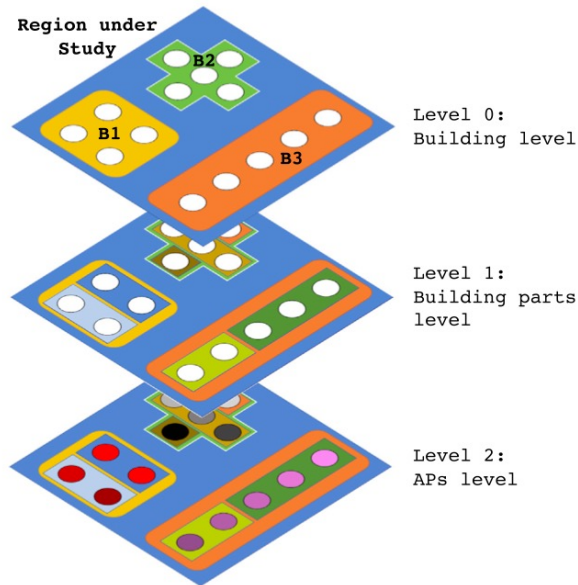
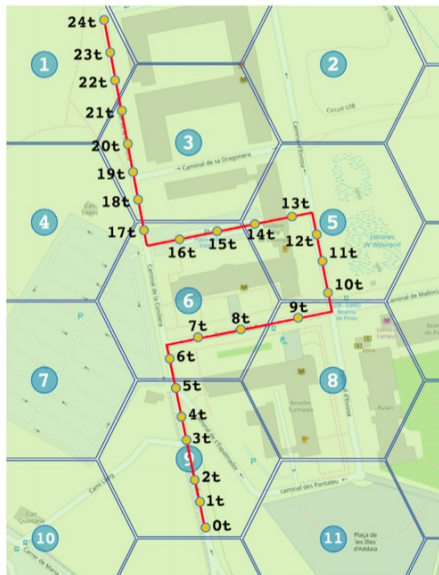
Contextualización



Universitat
de les Illes Balears



- **X** Difícil acceso a datos reales
- **X** Datos de casos particulares. Dificultad de generalización
- **X** Escenarios Fog de tamaños muy elevados
- **X** Complejidad de los modelos
- **X** Experimentación basada en datos no reales

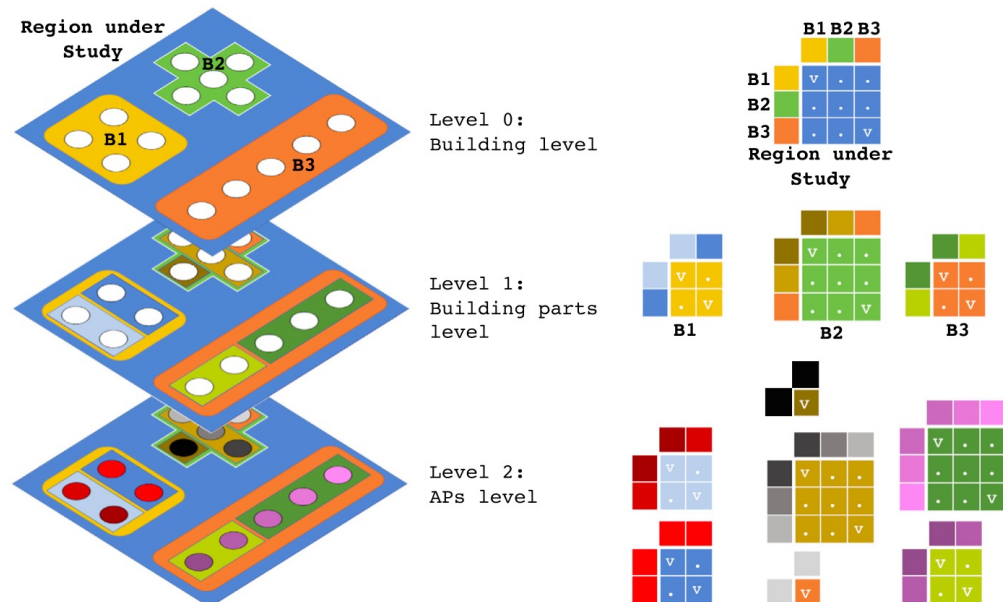
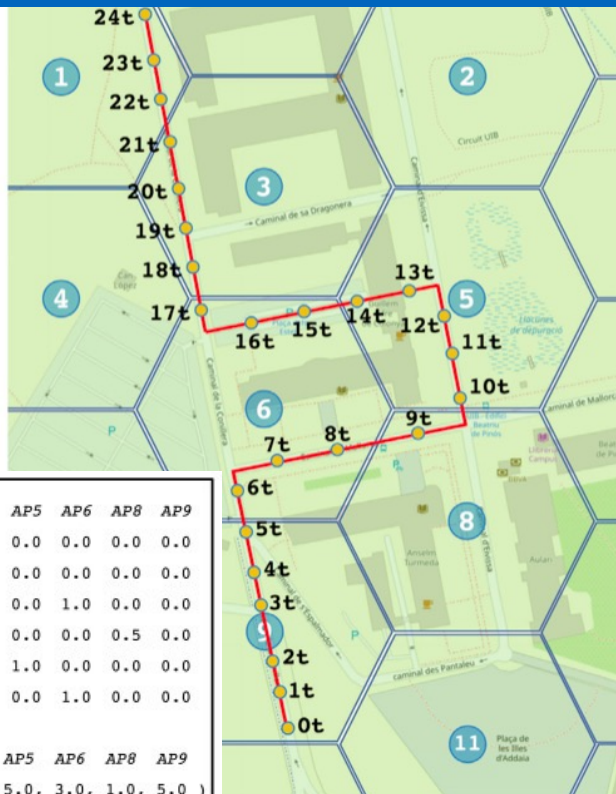


- Uso de datos obtenidos desde redes Wi-Fi

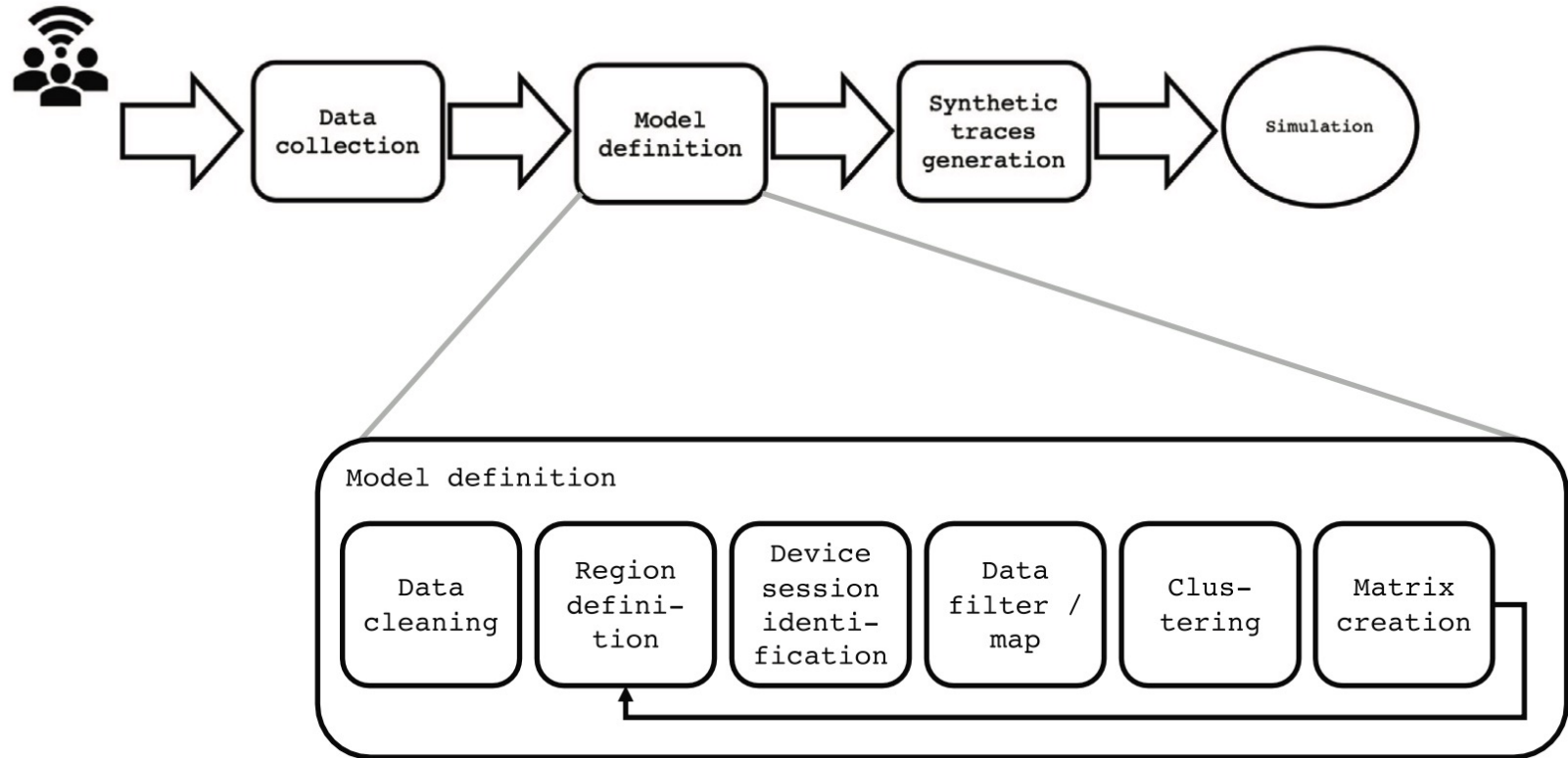
- Definir un modelo jerárquico para la definición del movimiento de los usuarios

Propuesta:

Modelo jerárquico basado en matrices de transición + vector de tiempos



Método propuesto



Data collection

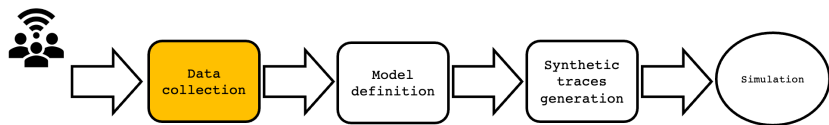


Universitat
de les Illes Balears

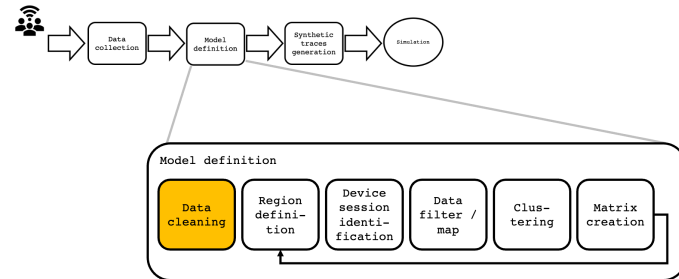


```
0t - dev0 - AP9
1t - dev0 - AP9
2t - dev0 - AP9
3t - dev0 - AP9
4t - dev0 - AP9
5t - dev0 - AP9
6t - dev0 - AP6
7t - dev0 - AP6
8t - dev0 - AP6
9t - dev0 - AP8
10t - dev0 - AP5
11t - dev0 - AP5
12t - dev0 - AP5
13t - dev0 - AP5
14t - dev0 - AP5
15t - dev0 - AP6
16t - dev0 - AP6
17t - dev0 - AP6
18t - dev0 - AP3
19t - dev0 - AP3
20t - dev0 - AP3
21t - dev0 - AP3
22t - dev0 - AP1
23t - dev0 - AP1
24t - dev0 - AP1
```

Wireless Sessions Trace Log



- Específico de cada caso de estudio
- Ejemplos:
 - Usuarios con conexiones a un único AP
 - Conexiones a un AP en concreto
 - Conexiones desde cierto tipo de dispositivos
 - Conexiones durante cierto periodo de tiempo
 - ...

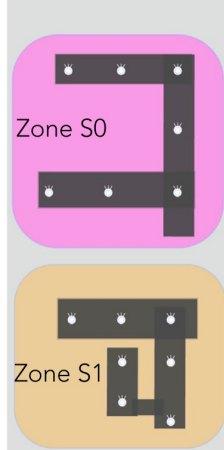


Hierarchical region definition



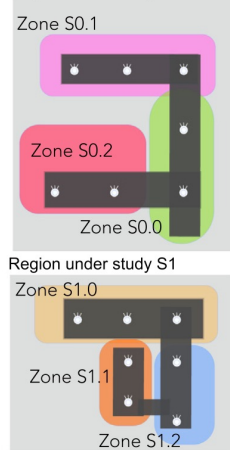
Level 0
Buildings

Region under study



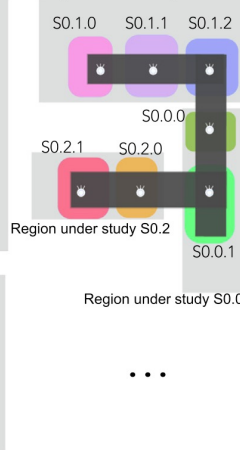
Level 1
Winds (buildings)

Region under study S0

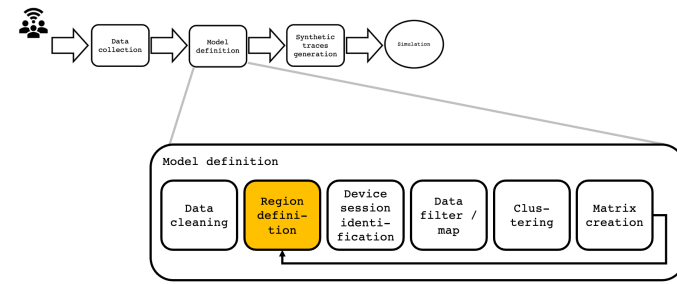


Level 2
APs

Region under study S0.1



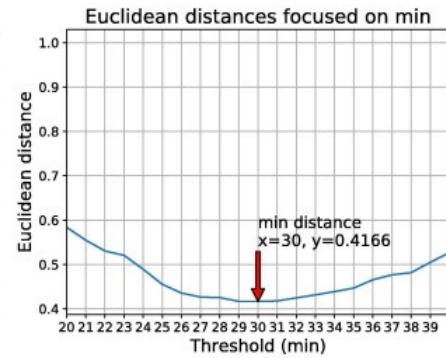
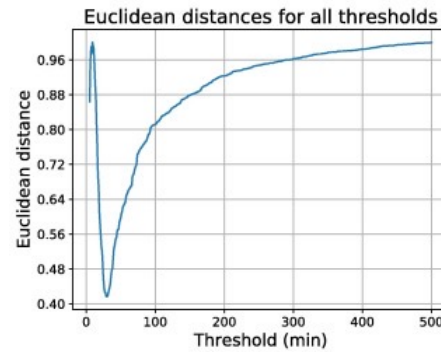
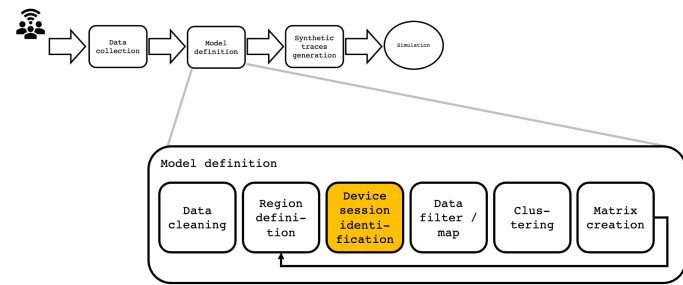
...



- Basado en la topología o en la distribución geográfica
- Niveles y zonas determinado por cada caso particular
- De forma recursiva:
 - Nivel $\leftrightarrow$ Matriz de transición
 - Matriz de transición = Zonas X Zonas
 - Cada zona $\rightarrow$ se repite el proceso recursivamente
- Número de zonas:
 - Alto $\rightarrow$ aumenta complejidad del modelo
 - Bajo $\rightarrow$ disminuye la precisión del modelo

Device session identification

- MAC address $\leftrightarrow$ Usuario
- Sesión de usuario $\rightarrow$ cada vez que un usuario entra y sale de una región bajo estudio
 - Cuando se detecta la entrada/salida de una región
 - No se detectan conexiones una vez superado un tiempo límite (requiere de un estudio previo y que balancee el número total de conexiones y la longitud de las sesiones)

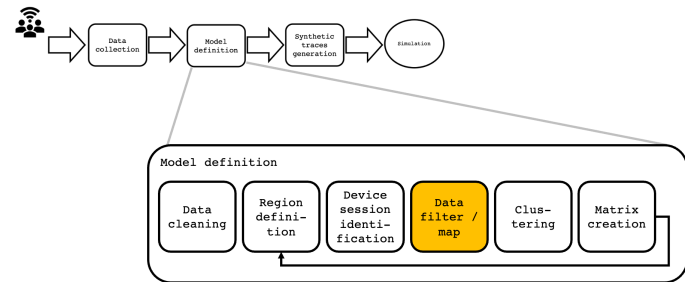


Data filter/map



0t	—	dev0	—	AP9
1t	—	dev0	—	AP9
2t	—	dev0	—	AP9
3t	—	dev0	—	AP9
4t	—	dev0	—	AP9
5t	—	dev0	—	AP9
6t	—	dev0	—	AP6
7t	—	dev0	—	AP6
8t	—	dev0	—	AP6
9t	—	dev0	—	AP8
10t	—	dev0	—	AP5
11t	—	dev0	—	AP5
12t	—	dev0	—	AP5
13t	—	dev0	—	AP5
14t	—	dev0	—	AP5
15t	—	dev0	—	AP6
16t	—	dev0	—	AP6
17t	—	dev0	—	AP6
18t	—	dev0	—	AP3
19t	—	dev0	—	AP3
20t	—	dev0	—	AP3
21t	—	dev0	—	AP3
22t	—	dev0	—	AP1
23t	—	dev0	—	AP1
24t	—	dev0	—	AP1

Wireless Sessions Trace Log

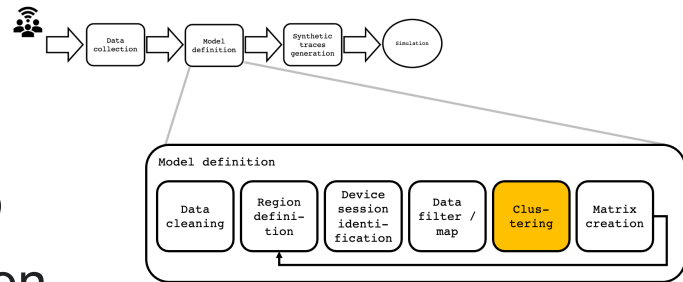


0t	—	dev0	—	AP9
6t	—	dev0	—	AP6
9t	—	dev0	—	AP8
10t	—	dev0	—	AP5
15t	—	dev0	—	AP6
18t	—	dev0	—	AP3
22t	—	dev0	—	AP1
24t	—	dev0	—	AP1

Handoff Trace Log

User profiling: clustering

- Diferenciar usuarios según su comportamiento
- Criterio de clustering: el porcentaje de tiempo en las zonas
- Algoritmos de clustering: k-means



	AP_0	AP_1	...	AP_{n-1}
UserSession ₀	$\frac{time_{AP_0}}{session_length_0}$	$\frac{time_{AP_1}}{session_length_0}$		$\frac{time_{AP_{n-1}}}{session_length_0}$
UserSession ₁	$\frac{time_{AP_0}}{session_length_1}$	$\frac{time_{AP_1}}{session_length_1}$		$\frac{time_{AP_{n-1}}}{session_length_1}$
...				
UserSession _{n-1}	$\frac{time_{AP_0}}{session_length_{n-1}}$	$\frac{time_{AP_1}}{session_length_{n-1}}$		$\frac{time_{AP_{n-1}}}{session_length_{n-1}}$

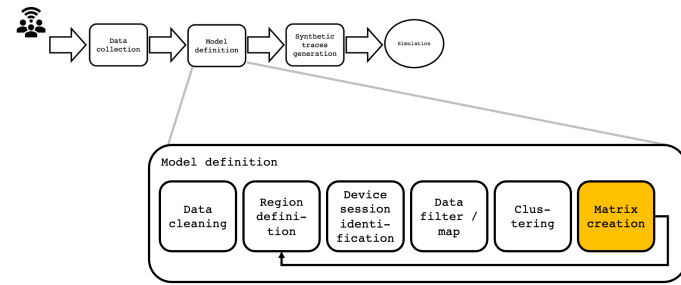
Transition matrix and time vector creation

Transition matrix

- $|\text{Número de zonas} + 2| \times |\text{Número de zonas} + 2|$
- Zonas IN y OUT
- $a_{i,j}$ = probabilidad de ir a la zona j cuando el usuario está en la zona i
- $a_{IN,i}$ = probabilidad de iniciar la sesión en la zona i
- $a_{i,OUT}$ = probabilidad de finalizar la sesión en la zona j
- $a_{i,j} = \frac{\text{número de saltos } zona_i \rightarrow zona_j}{\text{número de saltos totales}}$

Time vector

- $|\text{Número de zonas}|$
- a_i = tiempo medio de estancia en la zona i
- $a_i = \frac{\sum \text{tiempo visitas } zona_i}{\text{número de visitas } zona_i}$



$TM(lv1, bldgAT, -) =$

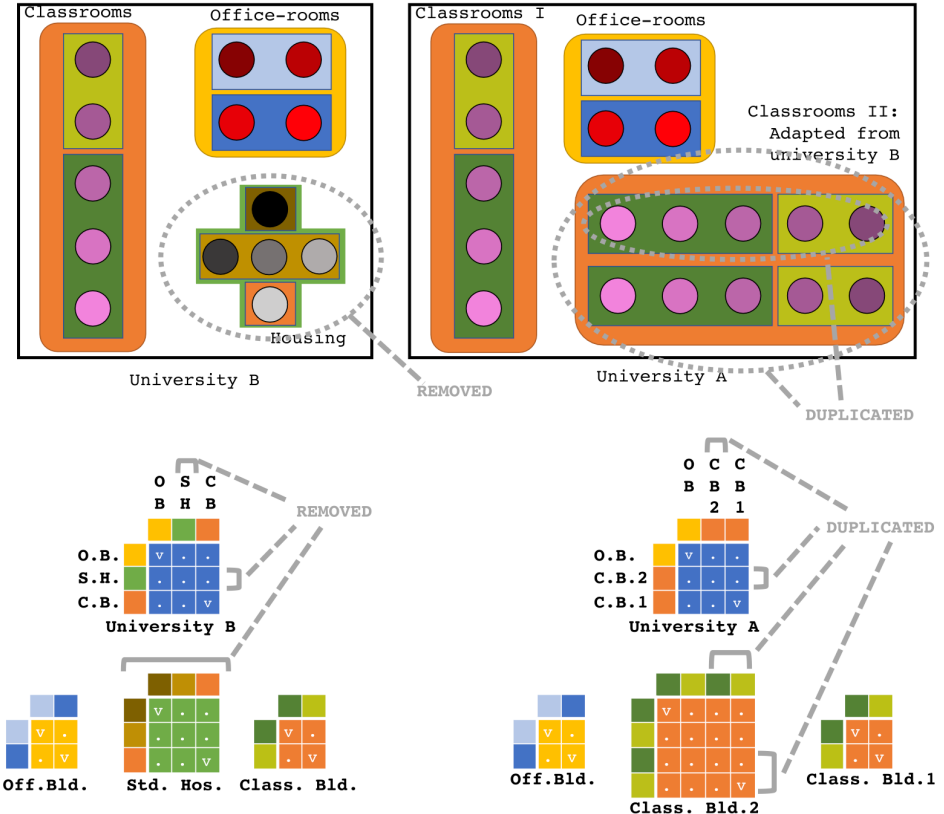
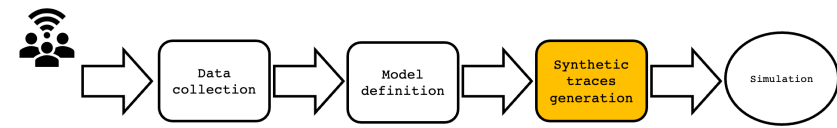
	IN	OUT	1st_fl_East	1st_fl_North	0_fl_East	0_fl_North	2nd_fl
IN	0	0	0.0364	0.0182	0.0364	0	0.909
OUT	0	0	0	0	0	0	0
1st_fl_East	0	0.4	0	0	0	0	0.6
1st_fl_North	0	0.286	0	0	0	0	0.714
0_fl_East	0	0.25	0	0	0	0	0.75
0_fl_North	0	0.5	0	0	0	0	0.5
2nd_fl	0	0.568	0.243	0.081	0.081	0.027	0

$TV(lv1, bldgAT, -) =$

1st_fl_East	P1N	0_fl_East	0_fl_North	2nd_fl
[6.92	15.03	14.93	6.23	51.85]

Generation of synthetic traces

- Trazas sintéticas que pueden (o no) seguir el patrón de los datos iniciales
 - Eliminar/replicar tipos de usuarios/zonas/niveles, ...
 - Carga de usuarios (tasa de llegada de usuarios, *i.e.*, número de usuarios)
- Salida:
 - `<time_stamp, user_id, ap_id>`



Case study: Universidad Illes Balears



Universitat
de les Illes Balears

- 425 APs
- 18 edificios
- Aruba ALE (Analytics and Location Engine)
- Una semana (9-15 noviembre 2022) → 122 MB de trazas wifi

- Evaluación:
 - Datos reales → Modelo datos reales → Trazas sintéticas → Modelo datos sintéticos
 - Evaluar error cuadrático medio (RMSE) de las matrices de transición y vectores de tiempo entre modelo datos reales vs. Modelo datos sintéticos
 - Matrices de transición: 1,5% - 7,7%
 - Vectores de tiempo: 4,5% - 16,8%

- Error aceptable en las trazas sintéticas generadas
- Tiempos de ejecución menor
 - Modelo jerárquico 250s vs. modelo no jerárquico 15.000s
- Adaptabilidad de los modelos
 - Se probó con éxito extender el modelo de un único edificio a tres edificios con distintas “configuraciones de plantas y alas”



Universitat
de les Illes Balears

Analysis of wireless network access logs for a hierarchical characterization of user mobility (summary)

Francisco Talavera, Isaac Lera,
Carlos Guerrero



carlos.guerrero@uib.es



<https://acsic.uib.es/cguerrero/>



@siesinformatica



Proyecto PID2021-128071OB-I00 financiado por:

