Computer Networks 243 (2024) 110329

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

Data Article

Space and time user distribution measurements dataset in a university campus

Olga Ruiz, Juan Sánchez-González, Jordi Pérez-Romero *, Oriol Sallent, Irene Vilà

Universitat Politècnica de Catalunya (UPC), Barcelona, Spain

ARTICLE INFO

Keywords:
User distribution
Wireless networks
University campus
Network optimisation
Radio resource management
Wifi

ABSTRACT

Radio Resource Management (RRM) strategies are essential components in mobile wireless networks, such as Long Term Evolution (LTE) or 5G New Radio (NR)[1]. The design of RRM algorithmic solutions is an area of research that has received a lot of attention for decades (see e.g. surveys [2–4] and references therein). Given that RRM strategies need to handle traffic dynamicity in the Radio Access Network (RAN), the availability of realistic user distributions in space and time can be very useful to support a more realistic performance assessment of RRM solutions. For this purpose, this article introduces a dataset containing real measurements of the number of users connected to the different Wifi Access Points (APs) at the Campus Nord facilities of the Universitat Politècnica de Catalunya (UPC) in Barcelona. The Wifi network is composed of 247 APs. To characterize the temporal variations, the data was collected for each AP every 1000 s (approximately) during 62 days. Besides the number of users connected to each AP (including users connected to both 2.4 GHz and 5 GHz bands), the dataset also contains information about the theoretical coverage area of each AP, so that the number of users connected to each AP can be associated to a specific geographical area. In this way, the dataset captures the spatio-temporal variations of users in the Campus at different times of the day, different days of the week and different periods of the academic year.

Specifications table

Subject	Wireless Communications
Specific subject area	Wireless user distribution measurements given in terms of the number of users with an active Wifi connection in a university campus area.
Type of data	Tables
How data were acquired	Data were acquired from measurements taken at 247 Wifi access points (APs) by means of Cisco Prime Infrastructure software [5] and filtered using a Matlab script to extract only the users connected to each AP.
Data format	Filtered
Parameters for data collection	The measurements were carried out continuously during 62 days from Tuesday 18th April 2023 00:07:30 to Sunday 18th June 2023 23:47:22.
Description of data collection	The data was collected by means of Cisco Prime Infrastructure software [5], which is a network management software that has access to the information generated by all the Wifi APs of the campus and it generates csv files with different measurements per AP. The dataset includes measurements for a total of 62 days with a periodicity of approximately 1000 s. In each measurement the number of users connected to each AP is collected. Then, to characterize the spatial distribution, this number of users is associated with a geographical region that corresponds to the theoretical coverage area of the AP.
Data source location	Institution: Universitat Politècnica de Catalunya (UPC) City: Barcelona Country: Spain Latitude and longitude (and GPS coordinates) for collected samples/data: The UPC campus area where the measurements have been taken comprises a rectangle with vertices (41.387939 N, 2.110101E), (41.38994 N, 2.11289E), (41.38923 N, 2.1181E), (41.38731 N, 2.11107E).
Data accessibility	Data is available in a public repository. Repository name: Mendeley Data Data identification number: O. Ruiz, J. Sánchez-González, J. Pérez-Romero, O. Sallent, I. Vilà, "Space and Time User Distribution in a University Campus", Mendeley Data, V1, February 2024. doi: 10.17632/55vx86j8wf.1

(continued on next page)

* Corresponding author.
E-mail address: jordi.perez-romero@upc.edu (J. Pérez-Romero).

https://doi.org/10.1016/j.comnet.2024.110329
Received 22 February 2024; Accepted 9 March 2024
Available online 11 March 2024

(*continued*)

	Direct URL to data: https://data.mendeley.com/datasets/55vx86j8wf/1
Related research article	–
Related project(s)	Project name: AI-powered eVolution towards opEn and secuRe edGe architEctures (VERGE) Funding body: European Union Type of grant: HORIZON-JU-SNS-2022-STREAM-A-01–05 (GA nº 101,096,034) Project duration: 1st January 2023 to 30th June 2025 Project website: https://www.verge-project.eu/ Project name: Beyond 5 G Artificial Intelligence Assisted Energy Efficient Open Radio Access Network (BeGREEN) Funding body: European Union Type of grant: HORIZON-JU-SNS-2022-STREAM-A-01–01 (GA nº 101,097,083) Project duration: 1st January 2023 to 30th June 2025 Project website: https://www.sns-begreen.com/ Project name: smART radIo acceSs with inTegration of user devices (ARTIST) Funding body: Spanish Ministry of Science and Innovation Type of grant: MCIN/AEI/10.13039/501,100,011,033 (ref. PID2020–115104RB-I00) Project duration: 1st September 2021 to 31st August 2024 Project website: https://artist.upc.edu/

1. Value of the data

- Why are these data useful?
  This dataset provides real distributions of the number of connected users to the Wifi APs of a university campus in space and time. This information is useful to accurately model the space/time user distribution in system level simulations of wireless networks (e.g. in Wifi networks, in a 5G Radio Access Network (RAN) [1], etc.), allowing for the performance assessment of different algorithmic solutions (e.g. Radio Resource Management, Self-Organizing Networks) for these networks under realistic testing conditions.
- Who can benefit from these data?
  Having real user distributions can be very useful for researchers working in wireless networks, particularly in areas such as network optimisation, radio resource management, network planning or self-organizing networks (see e.g. [2,3,4,6]). Real user distribution data can be useful for example to identify spatial heterogeneities so that developed solutions can place the focus on the regions with higher number of users. Similarly, having a large amount of data including several days can also be helpful for developing dynamic solutions that adapt to the evolution of the number of users.
- How can these data be used for further insights and development of experiments?
  The dataset of real user distribution measurements can be used for performing experiments or simulations of wireless networks which will allow developing efficient and usage-oriented solutions. As an example, the given data is being currently used in the context of the abovementioned VERGE and BeGREEN projects to obtain realistic dynamic user distributions used by simulations reproducing the campus scenario. These simulations intend to identify regions with poor propagation conditions that at the same time have a significant number of users. Based on this identification, the current research is trying to solve the coverage problems by means of the usage of relays located at strategic points, following a methodology like the one in [7], which is being extended to consider this dataset.
- What is the additional value of these data?
  The dataset provides an accurate representation of the user distribution in a university campus, as it was collected during a sufficiently comprehensive period that covers different situations of the academic year, including regular class days, holidays, etc. so that different user patterns can be derived for each situation. Besides that, the dataset captures the evolution of the spatial user distribution in different periods of the day. This is especially useful for developing and testing dynamic and adaptive solutions.

AP_id	Time	Users
1	Tue Apr 18 08:57:39 CEST 2023	0
1	Tue Apr 18 09:14:17 CEST 2023	10
1	Tue Apr 18 09:30:53 CEST 2023	17
1	Tue Apr 18 09:47:28 CEST 2023	18
1	Tue Apr 18 10:04:01 CEST 2023	13
1	Tue Apr 18 10:20:35 CEST 2023	19
1	Tue Apr 18 10:37:07 CEST 2023	20
1	Tue Apr 18 10:53:43 CEST 2023	7

Fig. 1. Sample of a file of the dataset.

2. Data

The provided dataset contains the measurements of the number of users connected to each of the 247 Wifi indoor APs deployed in the campus of *Universitat Politècnica de Catalunya* in Barcelona. The measurements cover a total of 62 days. Within each day, measurements are taken approximately every 1000 s, although this periodicity is not constant and can vary across different days. In addition to the number of users, the dataset also includes the information of the theoretical coverage area of each AP, so that the number of users connected to each AP can be associated to a specific geographical area. Based on these considerations, the dataset consists of two types of files, namely the files including the temporal measurements of the number of connected users per AP and the files that include the information with the theoretical coverage area of each AP. The total size of the dataset is 44.5 MB and it is stored in the commonly used csv format, with comma separated fields. The following subsections describe in detail the two types of files.

- **User distribution measurement files**:

The dataset includes a total of 62 csv files each one containing the time evolution of the number of connected users per AP for one day in the period from Tuesday 18th April 2023 00:07:30 to Sunday 18th June 2023 23:47:22. These dates cover different periods of the academic year including regular class days, weekends, examination periods and holidays. Each file is named following the format *campus_users_yyyymmdd.csv* (e.g. *campus_users_20,230,418.csv*) where dd, mm and yyyy represent, respectively, the day, month and year.

Each file contains a table with the following three columns and the first line of the file contains the names of these columns:

- *AP_id* – The ID of the Wifi AP corresponding to the given data. The IDs are integers ranging from 1 to 247.
- *Time* - The timestamp of each measurement. This field is specified in the following format: *DayOfTheWeek Month Day hh:mm:ss TimeZone Year* (e.g. Tue Apr 18 23:53:03 CEST 2023). Typically, there are between 85 and 87 measurements per AP in each day.
- *Users* – The total number of connected users to the AP at the specified time.

Fig. 1 shows a sample of one of the files included in the dataset.

As an example, Fig. 2 shows the evolution of the number of users measured by AP 20 during the 18th of April 2023. This AP's coverage area corresponds to a lecture room in the campus, so the number of users is concentrated within the school timetable, which comprises from 8 h to 20 h.

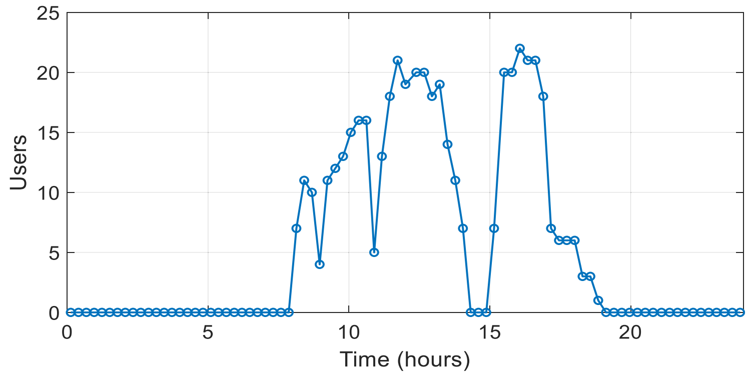

Fig. 2. Evolution of the number of users measured by AP 20 during one day.

Fig. 3. UPC Campus Nord.

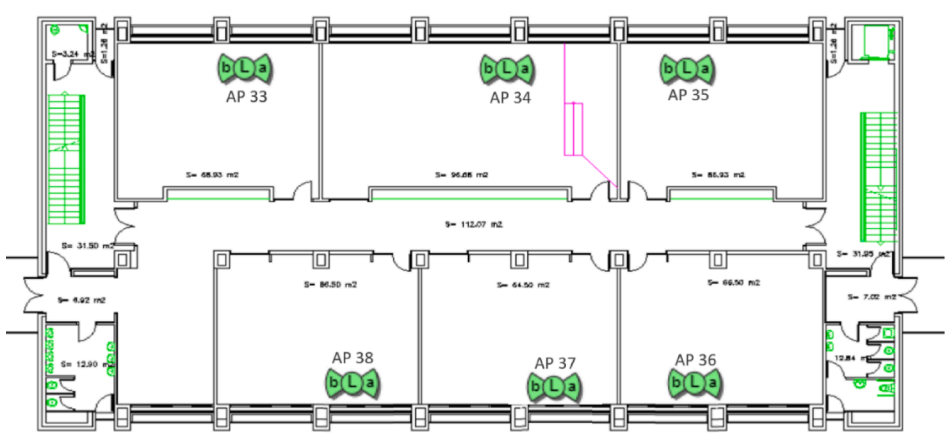

Fig. 4. Example of one floor and the deployed APs in one of the buildings in the campus.

- **Spatial regions files**:

The university campus is a rectangular area of 335 m x 125 m (see Fig. 3). It comprises 24 buildings that are 3 floors high. The APs are distributed in the different floors of the buildings. For illustrative purposes, Fig. 4 shows an example of the AP distribution in one of the floors of one building, taken from Cisco Prime Infrastructure software. Based on the real positions of the APs, a postprocessing is carried out to determine the theoretical coverage area of each AP. Specifically, the campus scenario is divided in pixels of 1 m x 1 m to form a matrix of 335 rows and 125 columns per floor, and a Voronoi tessellation is assumed, where each pixel takes the value of the identifier of the closest AP. Pixels not associated to any AP (e.g., those representing outdoor positions) take the value 0. The information of the matrix of each floor is provided in three files, named as *regions_map_floor_0.csv*, *regions_map_floor_1.csv* and *regions_map_floor_2.csv*, corresponding, respectively, to the ground floor, the first floor and the second floor. Each one of these files includes the 335×125 matrix for all the pixels in the campus so that the element located at the position (1,1) designates the upper left corner of the campus (see Fig. 3), and the rows and columns of the csv file correspond to the X and Y axes respectively.

3. Experimental design, materials, and methods

The data included in this article was collected by means of Cisco Prime Infrastructure software [5], which is a network management software used for the management of the Wifi network deployed in UPC. This software has access to information generated by all the Wifi APs of the different UPC campuses and it collects a multiplicity of Wifi-related metrics including, among them, the number of users connected to each AP distinguishing between users connected to the 2.4 GHz band and the users connected to the 5 GHz band.

The csv reports containing all the metrics for all the APs in each measurement time generated by the software were filtered by means of a Matlab code. This code generates csv files containing only the timestamps and the number of users corresponding to the APs of the considered UPC campus. The selection of these APs was done manually by checking the maps provided by Cisco Prime Infrastructure software. The number of users in each AP was obtained as the aggregate of the users connected to the 2.4 and 5 GHz bands. The coordinates with the position of each AP were also obtained from this software and, based on them, the theoretical coverage area of each AP was determined as the set of pixels having this AP as the nearest one. In this way, the above-mentioned files *regions_map_floor_0.csv, regions_map_floor_1.csv* and *regions_map_floor_2.csv* were created reproducing the campus scenario as a matrix in which each element corresponds to a 1 × 1 meter square pixel.

CRediT authorship contribution statement

Olga Ruiz: Data curation, Writing – original draft, Writing – review & editing. **Juan Sánchez-González:** Data curation, Writing – review & editing. **Jordi Pérez-Romero:** Conceptualization, Writing – original draft, Writing – review & editing. **Oriol Sallent:** Conceptualization, Funding acquisition, Writing – review & editing. **Irene Vilà:** Conceptualization, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

This is a dataset paper and the data is available at Mendeley Data repository.

Acknowledgments

This work is partially funded by the Smart Networks and Services Joint Undertaking (SNS JU) under the European Union's Horizon Europe research and innovation programme under Grant Agreements No. 101096034 (VERGE project) and No. 101097083 (BeGREEN project) and by the Spanish Ministry of Science and Innovation MCIN/AEI/10.13039/501100011033 under ARTIST project (ref. PID2020–115104RB-I00). Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them. The work of Irene Vilà has been funded by the European Union-NextGenerationEU, the Spanish Ministry of Universities and the Plan for Recovery, Transformation and Resilience, through the call for Margarita Salas Grants of the Universitat Politècnica de Catalunya (ref. 2022UPC-MSC- 94079).

References

[1] E. Dahlman, S. Parkvall, J. Skold, 5G NR: The Next Generation Wireless Access Technology, 1st ed., Academic Press (Elsevier), 2018.
[2] T.O. Olwal, K. Djouani, A.M. Kurien, A survey of resource management towards 5G radio access networks, IEEE Commun. Surv. Tutor. 18 (3) (2016) 1656–1686, https://doi.org/10.1109/COMST.2016.2550765.
[3] I.A. Bartsiokas, P.K. Gkonis, D.L. Kaklamani, I.S. Venieris, ML-based radio resource management in 5G and beyond networks: a survey, IEEE Access 10 (2022) 83507–83528, https://doi.org/10.1109/ACCESS.2022.3196657.
[4] S. Manap, K. Dimyati, M.N. Hindia, M.S.A. Talip, R. Tafazolli, Survey of radio resource management in 5G heterogeneous network, IEEE Access 8 (2020) 131202–131223, https://doi.org/10.1109/ACCESS.2020.3002252.
[5] Cisco Prime Infrastructure, https://www.cisco.com/c/en/us/support/cloud-systems-management/prime-infrastructure/series.html (accessed 21 February 2024).
[6] H. Fourati, R. Maaloul, L. Chaari, M. Jmaiel, Comprehensive survey on self-organizing cellular network approaches applied to 5G networks, Comput. Netw. 199 (2021), https://doi.org/10.1016/j.comnet.2021.108435.
[7] J.J. Hernández-Carlón, J. Pérez-Romero, I. Vilà, O. Sallent, F. Casadevall, On the detection and solution of coverage holes in 5G networks through relay user equipment: a combined DBSCAN and deep-Q network approach, in: Proceedings of the IEEE 97th Vehicular Technology Conference, 2023, https://doi.org/10.1109/VTC2023-Spring57618.2023.10199254.

Olga Ruiz received a B.E. degree in Telecommunication Systems Engineering in 2019 and a M.S. degree in telecommunications engineering in 2022 from Universitat Politècnica de Catalunya and is currently a research assistant at the Mobile Communications Research group of the Signal Theory and Communications department at UPC. Her research interests include mobile and wireless communications systems.

Juan Sánchez-González received the Telecommunication Engineer degree in May 2002 and the Doctor Engineer in Telecommunication degree in December 2007 from the Universitat Politècnica de Catalunya (UPC), Spain. He is currently associate professor in the Escola d'Enginyeria de Telecomunicació i Aeroespacial de Castelldefels (EETAC). He joined the Mobile Communications Research group of the Signal Theory and Communications department in 2002. His research interests are in the field of mobile and wireless communications systems, especially in radio resource and quality of service management and optimization of heterogeneous wireless networks. He has been involved in different European Projects as well as in projects for private companies. He is currently working in the context of the applicability of Artificial Intelligence/Machine Learning techniques for the optimization of heterogeneous wireless networks.

Jordi Pérez-Romero received a degree in telecommunications engineering and a Ph.D. degree from the Universitat Politècnica de Catalunya (UPC), Barcelona, Spain, in 1997 and 2001, respectively. He is currently a Professor with the Department of Signal Theory and Communications of UPC. He is working in the field of wireless communication systems, with a particular focus on 5 G and beyond cellular systems, including radio resource management and network optimization. He has been involved in different European projects with different responsibilities, such as researcher, work package leader, and project responsible, has participated in different projects for private companies and has contributed to the 3GPP and ETSI standardization bodies. He has coauthored more than 300 papers in international journals and conferences. He has also coauthored three books and has contributed to seven book chapters. He is an Associate Editor for the IEEE Vehicular Technology Magazine.

Oriol Sallent is a Professor at the Universitat Politècnica de Catalunya, Barcelona, Spain. He has participated in a wide range of European and national projects, with diverse responsibilities as a principal investigator, coordinator, and work package leader. He regularly serves as a consultant for a number of private companies. He has contributed to standardization bodies such as 3GPP, IEEE, and ETSI. He is the coauthor of 13 books and has authored or coauthored 300+ papers, mostly in high-impact IEEE journals and renowned international conferences. His-research interests include 5 G RAN (Radio Access Network) planning and management, artificial intelligence-based radio resource management, virtualisation of wireless networks, cognitive management cognitive radio networks and dynamic spectrum access and management, among others.

Irene Vilà received a B.E. degree in Telecommunication Systems Engineering in 2015, an M.S. degree in Telecommunication Engineering in 2017, and a Ph.D. degree in Signal Theory and Communications in 2022, all from Universitat Politècnica de Catalunya (UPC), Barcelona, Spain. She is currently a postdoctoral researcher at the Department of Signal Theory and Communications (TSC) at UPC, funded by a Margarita Salas grant from the Ministry of Universities with funds from the European Union NexGenerationEU. She has been involved in different national and European research projects founded by both public and private organizations, and she has contributed to ETSI. She has coauthored more than 20 papers in international journals and renewed conferences, as well as organized a conference tutorial, and contributed to a book chapter. Her research interests focus on the field of mobile communications, particularly on radio resource management and network optimization, the application of artificial intelligence to radio access network management, and edge computing.