Computer Networks 180 (2020) 107422

Contents lists available at ScienceDirect

Computer Networks

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

Data Article

Wifi-based localisation datasets for No-GPS open areas using smart bins

Mohamed A. Nassar[a], Mahmud Hasan[a], Md Khan[a], Mirza Sultana[a], Md Hasan[a], Len Luxford[b], Peter Cole[a], Giles Oatley[c], Polychronis Koutsakis[a,*]

[a] *Information Technology, Mathematics and Statistics Discipline, Murdoch University, Australia*
[b] *Global Smart Cities T/A yStop, Australia*
[c] *School of Science, Engineering and Information Technology, Federation University, Australia*

ARTICLE INFO

Keywords:
Wifi datasets
Fingerprinting
Smart bins
Smart street furniture
Open area localisation

ABSTRACT

In recent years, Wifi-based localisation systems have gained significant interest because of the lack of Global Positioning System (GPS) signal in indoor and certain open areas. Over the past decade, many datasets have been introduced to enable researchers to compare different localisation techniques. Existing datasets, however, have failed to cover open areas such as parks in cases where GPS is still unavailable, and there is a lack of Wifi access points. Also, the existing datasets only focus on getting Wifi fingerprint collected and labelled by users. To the best of our knowledge, no dataset provides Received Signal Strengths (RSS) collected by Wireless Access Points (APs).

In this work, we offer two datasets publicly. The first is the *Fingerprint dataset* in which four users generated 16,032 accurate and consistently labelled WiFi fingerprints for all available Reference Points (RPs) in a central and busy area of Murdoch University, known as Bush Court. The second is the *APs dataset* that includes 2,450,865 auto-generated records received from 1000 users' devices, including the four users, associated with Wifi signal strengths. To overcome the Wifi coverage problem for the Bush Court, we attached our previously designed Wireless Sensor Nodes (WSNs) to existing garbage bins, enabling them to provide real-time environmental sensing and act as soft APs that sense MAC addresses and Wifi signals from surrounding devices.

Specifications Table

Subject	Wireless Networking
Specific subject area	Wifi-based localisation for No-GPS Open Areas using smart street furniture.
Type of data	Table
How data were acquired	Instruments: • Hardware (WSNs, Smartphones, Server) • Software (Wifi Analyser application [1], LAMP [2])
Data format	Raw
Parameters for data collection	• Four users were assigned dedicated smartphones for the entire trial. • Each user provided four fingerprints per RP (one fingerprint for each direction). • Each user scanned all available RPs row by row. • The four smartphones came with Wifi analyser application[1] with shortest scan interval. • WSNs were configured to send MAC addresses and RSSs received from surrounding devices every *sec* to the server.

(*continued on next page*)

* Corresponding author.
E-mail addresses: mohamedatifnassar.ali@murdoch.edu.au (M.A. Nassar), hemel.a0193@gmail.com (M. Hasan), ibrahim67778@gmail.com (M. Khan), azifarumy@gmail.com (M. Sultana), mesbahulhasanparag@gmail.com (M. Hasan), len@ystop.com.au (L. Luxford), p.cole@murdoch.edu.au (P. Cole), g.oatley@federation.edu.au (G. Oatley), p.koutsakis@murdoch.edu.au (P. Koutsakis).

https://doi.org/10.1016/j.comnet.2020.107422
Received 25 February 2020; Received in revised form 9 May 2020; Accepted 12 July 2020
Available online 15 July 2020
1389-1286/

(*continued*)

Subject	Wireless Networking
Description of data collection	**Fingerprint dataset** Each user installed Wifi analyser application [1] with the shortest scan interval. This application enabled users to get fingerprints with timestamp at each RP in CSV format. Users scanned RPs row by row during an one-month period. **AP dataset** Using our predesigned WSNs, garbage bins work as Wifi sniffers. They sent encrypted MAC addresses and RSSs received from surrounding devices to the server every *sec*. The data were stored in a database.
Data source location	Area: The Bush Court Institution: Murdoch University City/Town/Region: Perth Country: Australia
Data accessibility	Repository name: MurdochBushCourtLoC Direct URL to data: https://data.mendeley.com/datasets/rdhfvhyg5p/2
Related research article	• M.A. Nassar, et al., “The Current and Future Role of Smart Street Furniture in Smart Cities,” IEEE Communications Magazine, vol. 57, no. 6, 2019, pp. 68–73. • M.A. Nassar, et al., “Adaptive Low-Power Wireless Sensor Network Architecture for Smart Street Furniture-based Crowd and Environmental Measurements,” Proc. 2019 IEEE 20th International Symposium on" A World of Wireless, Mobile and Multimedia Networks"(WoWMoM), IEEE, 2019, pp. 1–9. • M.A. Nassar, et al., “Secure and Privacy-Preserving Real-time Dynamic Crowd Measurement System Using Smart Street Furniture,” Proc. Technology Surprise Forum as part of the Safeguarding Australia Conference, Canberra, 9 May 2019 (awaiting publication)
Related project(s)	The work of the first author as a doctoral candidate is jointly supported by: a) Murdoch University and Global Smart Cities (www.ystop.com.au), b) the Science Industry PhD Fellowship Program of the Department of Jobs, Tourism, Science and Innovation, Government of Western Australia.

Value of the data

- Existing datasets are generated using users’ smartphones only [3–5]. The process of generating them *requires users to participate using smartphones which makes the aforementioned datasets limited in terms of size and scalability*. Our datasets are the first ones produced using two different Wifi sources for the same RPs for an open area. The first source is the four users’ smartphones which is limited compared with the second source. The second source is APs which generated a huge amount of data for all users in the bush court including the four users.
- These datasets allow networking/communications researchers to study the signal variation between devices’ fingerprints and AP RSSs for the same devices.
- These datasets allow data scientists to provide different visualisations and analyses to test homogeneity between the two datasets.
- Using the two datasets, more efficient models can possibly be developed [6].
- These datasets open the door for providing new techniques to label the very large amount of auto-generated data received from APs using the limited amount of labelled data.

1. Data

1.1. *fingerprint.csv*

This dataset contains 2 MB of RSSs collected from 20 APs by four users. It has 16,032 records, and each record consists of 20 RSSs values (fingerprint), timestamp, RP location(x,y) and user ID where:

- RSSs are 20 RSS values received from 20 APs and values vary from −31 to −100 (the bigger value indicates closer proximity to a given AP and −100 indicates out-of-range RSS).
- timestamp format is YYYY-MM-DD-HHmmSS.
- x value varies from 0 to 30 (31 rows).
- y value varies from 0 to 32 (33 columns).
- user ID value varies from 1 to 4 (four users).

1.2. *APs.csv*

This dataset contains 72 MBs of auto-generated RSSs. It has 2,450,865 records automatically generated by 20 APs. Each record consists of AP ID, user ID, RSS and timestamp where:

- AP ID contains a unique ID for each AP.
- User ID is a unique number for each unique MAC address received from users’ smartphones by APs. User IDs from 1 to 4 are assigned to the four users who created fingerprint.csv.
- timestamp format is YYYY-MM-DD-HHmmSS.
- RSSs are RSS received from users’ smartphones.

2. Experimental design, materials, and methods

This paper introduces the first Wifi datasets for No-GPS open area based on Wifi. We nominated Bush Court of Murdoch University to be the area of study because it is central to the campus and can be considered as a hub for staff and students given that it is surrounded by many facilities including the main library and restaurants. We tackled the problem of APs coverage in Bush Court by scaling our previously designed Wireless Sensor Nodes (WSNs) across all existing garbage bins. Fig. 1. shows the distribution of garbage bins and their IDs in which 20 bins provide excellent coverage for Bush Court. ArcGIS is used to visualise bins locations [7]. We created the first dataset, **Fingerprint dataset**, with the support of four users to achieve the labelling. The second dataset, **APs dataset**, includes Received Signal Strengths (RSS) of many users identified by MAC addresses. We conducted different analyses and statistical tests to validate the **Fingerprint dataset** and test whether the two datasets are homogenous.

Before starting to deploy our WSNs across the garbage bins, we simulated the coverage of the Wifi signal. Wifi can cover more than 100 m distance, however, we limited the Wifi coverage of each bin to be only 30 m to avoid excessive signal decay due to blocking or noise problems. Fig. 2. shows a heatmap that represents the number of bins that cover all locations within the Bush Court. Each location is covered by up to 10 APs, hence the garbage bins provide excellent coverage for the Bush Court. We deployed 20 WSNs inside the garbage bins. Fig. 3. shows a bin after adding our WSN.

The Bush Court is a relatively big area (around $100 \times 95\text{m}^2$) and the ***Fingerprint dataset*** generation was a very time-consuming task which

Fig. 1. Distribution of garbage bins across the Bush Court of Murdoch university.

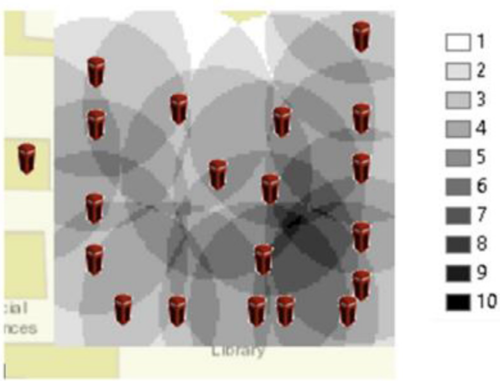


Fig. 2. Heatmap showing Wifi coverage.

Fig. 3. A bin after adding a WSN.

lacked scalability, i.e. it was required that each user scanned RPs and labelled them. On the other hand, we have auto-generated a huge amount of data at the server side reported by APs. Accordingly, we generated the two datasets as follows.

3. Fingerprint dataset

Using our WSNs, users' smartphones receive and record RSSs from garbage bins. Our main objective for the ***Fingerprint dataset*** is to produce an unbiased dataset i.e. same number of fingerprints per user per RP with 100% coverage of the available area. Four users supported us

Table 1
Users' details.

User ID	Smartphone Type	Operating System
1	PRIMO GH7I	Android 8.1
2	LG G6	Android 7.0
3	Samsung S8	Android 7.0
4	Oppo F1 Plus	Android 6.0

Access Points																				timestamp	RP		user
11	12	13	15	17	18	19	10	21	22	3	33	37	38	39	4	6	7	8	9		x	y	
-80	-79	-100	-100	-63	-100	-100	-72	-100	-100	-100	-100	-100	-100	-100	-79	-100	-67	-67	-85	2019-09-25-184140	0	0	1
-80	-79	-100	-100	-63	-100	-100	-72	-100	-100	-100	-100	-100	-100	-100	-79	-100	-67	-67	-85	2019-09-25-184142	0	0	1
-83	-83	-100	-100	-100	-100	-100	-81	-100	-100	-100	-100	-85	-100	-100	-100	-100	-67	-78	-86	2019-09-25-184152	0	0	1
-100	-100	-100	-100	-71	-100	-100	-81	-100	-85	-100	-100	-100	-100	-100	-100	-100	-67	-85	-100	2019-09-25-184159	0	0	1
-84	-80	-86	-100	-68	-85	-100	-75	-100	-87	-100	-78	-88	-100	-100	-71	-89	-76	-81	-88	2019-09-25-184209	0	1	1
-83	-85	-89	-100	-60	-83	-100	-86	-100	-86	-82	-100	-84	-100	-100	-70	-100	-85	-73	-86	2019-09-25-184315	0	1	1
-100	-86	-100	-100	-65	-100	-100	-84	-100	-100	-88	-100	-100	-100	-100	-78	-87	-73	-81	-100	2019-09-25-184321	0	1	1
-100	-100	-100	-100	-73	-100	-100	-81	-100	-82	-89	-100	-100	-100	-89	-85	-100	-100	-72	-100	2019-09-25-184329	0	1	1
-86	-83	-100	-100	-65	-100	-100	-87	-100	-84	-81	-100	-100	-100	-100	-77	-100	-84	-81	-100	2019-09-25-184337	0	2	1
-83	-100	-100	-100	-51	-83	-90	-76	-100	-83	-100	-100	-85	-100	-80	-76	-91	-65	-64	-100	2019-09-25-184432	0	2	1
-100	-78	-100	-100	-51	-84	-100	-100	-100	-100	-79	-100	-100	-88	-100	-100	-89	-65	-67	-80	2019-09-25-184440	0	2	1
-87	-100	-82	-100	-100	-100	-100	-85	-87	-82	-100	-100	-100	-100	-100	-77	-100	-78	-71	-100	2019-09-25-184447	0	2	1
-82	-78	-84	-100	-100	-87	-100	-79	-100	-100	-77	-100	-81	-100	-100	-73	-84	-74	-61	-100	2019-09-25-184455	0	3	1
-81	-100	-100	-100	-64	-100	-100	-73	-100	-100	-76	-100	-100	-100	-100	-68	-83	-71	-69	-79	2019-09-25-184532	0	3	1
-77	-84	-100	-100	-64	-100	-100	-100	-86	-85	-76	-100	-86	-88	-100	-70	-100	-69	-70	-78	2019-09-25-184539	0	3	1
-100	-81	-75	-100	-64	-100	-100	-80	-100	-83	-100	-74	-84	-100	-100	-100	-100	-75	-74	-86	2019-09-25-184549	0	3	1

Fig. 4. A sample from the fingerprint database.

to generate the ***Fingerprint dataset***. Table 1 lists the users' details. We divided the area into a grid of 31 rows and 33 columns. Accordingly, we have 1023 RPs and each RP is 3×3 m^2. Each user installed Wifi analyser application [1] with the shortest scan interval. This application enables users to get fingerprints with timestamp at each RP, which can then be saved in CSV format. Users scanned RPs row by row during an one-month period. Each user started at the middle of a RP and recorded four fingerprints for the four directions (North, South, East, West). Among 1023 RPs, 21 RPs cannot be visited because they have huge trees. These RPs can be seen in Fig. 5 and their indices are as follows: (7,21), (8,31), (12,28), (14,19), (15,22), (16, 17), (16,18), (17,17), (17,18), (18,17), (23, 11), (21, 21), (21, 22), (22, 21),(22,22), (23,21), (23,22), (23,23), (24,21), (24, 22) and (22,23).

Each user generated 4008 fingerprints in the form of CSV files, which were subsequently consolidated into one single CSV file. Fig. 4 shows some records from the dataset. RSS values are from −31 to −100 where the bigger value indicates closer proximity to a given AP and −100 indicates out-of-range RSS. The dataset contains 16,032 fingerprints.

This dataset is novel in that users cover all available RPs. Moreover, they provide the same number of fingerprints per RP. Accordingly, any prediction model will not suffer from a data bias problem toward certain users or certain RPs [8]. Fig. 5 shows the number of fingerprints across the Bush Court as a heatmap. It shows exactly 16 fingerprints for each RP except for the 21 unavailable RPs (grey area). Fig. 6 is another way to show the consistent number of fingerprints provided by users across different dates.

4. AP dataset

Using our WSNs, garbage bins work as Wifi sniffers because they act as APs and they collect MAC addresses and RSSs received from surrounding devices and send them to the server periodically. It is worth mentioning that MAC addresses are encrypted before being sent to the server using one-way encryption method to maintain users' privacy. Our design supports a two-way communication model between WSNs and the server, hence we can configure the frequency of sending data (encrypted MAC addresses and RSSs) remotely. For the ***APs Dataset***, we configured WSNs to communicate the data to the server every 1 s. ***APs dataset*** contains 2,450,865 records collected during the same period as the ***Fingerprint dataset***. It contains the data gathered for all users in the

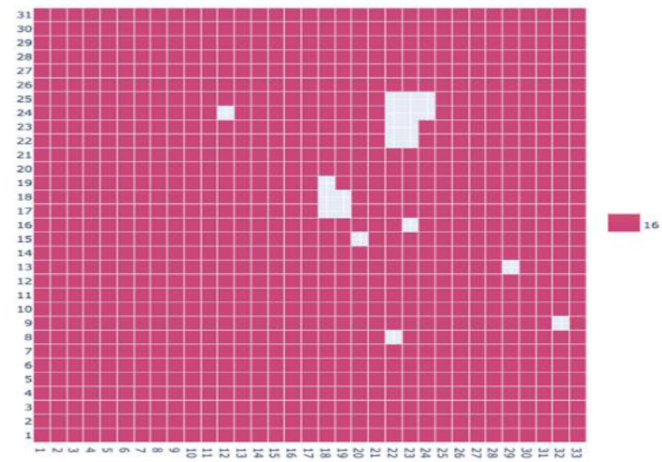


Fig. 5. No. of fingerprints per Reference Point.

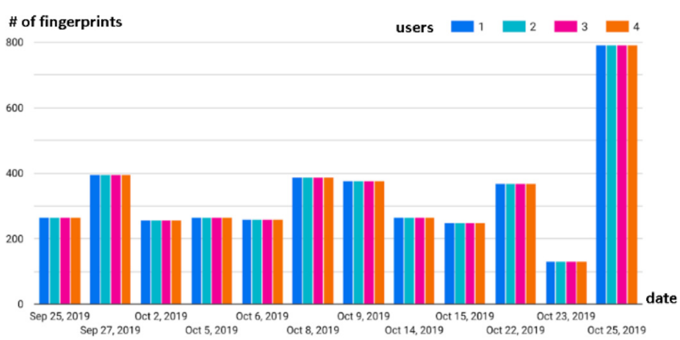


Fig. 6. No. of fingerprints per users across different dates.

bin	user	rss	time_stamp
13	1	90	25/09/2019 18:42:09
10	1	83	26/09/2019 18:42:09
22	1	89	27/09/2019 18:42:09
4	1	79	28/09/2019 18:42:09
17	1	72	29/09/2019 18:42:09
7	1	84	30/09/2019 18:42:09
12	1	90	1/10/2019 18:42:09
8	1	86	2/10/2019 18:42:09
11	1	90	3/10/2019 18:42:09
4	1	81	4/10/2019 18:42:09

Fig. 7. A sample from APs dataset.

Bush Court including the four users. Fig. 7 shows a sample from the ***APs dataset***.

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.

Acknowledgments

The work of the first author as a doctoral candidate is jointly supported by: a) Murdoch University and Global Smart Cities (www.ystop.com.au), b) the Science Industry PhD Fellowship Program of the Department of Jobs, Tourism, Science and Innovation, Government of Western Australia.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.comnet.2020.107422.

References

[1] Netspot for Android, Check the Best WiFi Analyzer Apps For Android, 2019 https://www.netspotapp.com/wifi-analyzer-apps-android.html.
[2] IBM Cloud Learn Hub, “LAMP Stack”, 2019 https://www.ibm.com/cloud/learn/lamp-stack-explained.
[3] J. Torres-Sospedra et al., “UJIIndoorLoc: a new multi-building and multi-floor database for WLAN fingerprint-based indoor localization problems”, in Proceedings of the 2014 IEEE international conference on indoor positioning and indoor navigation (IPIN), pp. 261–270.
[4] R. Montoliu et al., “IndoorLoc platform: a public repository for comparing and evaluating indoor positioning systems”, in Proceedings of the 2017 IEEE International conference on indoor positioning and indoor navigation (IPIN), pp. 1–8.
[5] E.S. Lohan, et al., “Wi-Fi crowdsourced fingerprinting dataset for indoor positioning”, Data 2 (4) (2017) 32.
[6] M. Mohammadi, “Semisupervised deep reinforcement learning in support of IoT and smart city services”, IEEE Internet Things J. 5 (2) (2017) 624–635.
[7] “ArcMap | ArcGIS Desktop,” 2019, https://desktop.arcgis.com/en/arcmap/.
[8] A. Torralba, A.A. Efros, “Unbiased look at dataset bias”, in: Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), 2011.

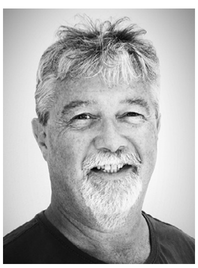
Mohamed A. Nassar is a Ph.D. candidate at Murdoch University. His research interests include machine learning, big data analysis, Spatio-temporal data mining, and recommender systems. He acquired his Master and Bachelor degrees in Computer Engineering in 2011 and 2007 respectively from Alexandria University, Egypt. He has published ten papers in international journals and conferences. Mohamed has had more than 13 years of experience in Software/Data Engineering. He has successfully managed and developed many projects including the street furniture recommender system for yStop clients, as well as centralized course content and new architecture/design for the Learning Management System for eight distributed branches of Arab Open University. He currently works as a Consultant at the Servian company.

Mahmud Hasan is a postgraduate student of Information Technology at Murdoch University. His research interests include pattern recognition, online data security, machine learning, data security policies.

Md Ibrahim Khan is a postgraduate student at Murdoch University. His research interests include network security, vulnerability detection and prevention, IoT technology.

Mirza Azifa Sultana is a postgraduate student at Murdoch University. Her research interests include artificial intelligence, machine learning, security and privacy, augmented reality and virtual reality, wireless and sensor systems, and big data analysis.

Md Mesbahul Hasan is a postgraduate student at Murdoch University. His research interests focus on the practical implementation and evaluation of scientific theories.



Len Luxford is Director of yStop, a Perth based, furniture technology company. The company has installed a network of smart-ready furniture throughout Perth councils including WA's largest - the City of Stirling. The latest project is Australia's first free (local business funded) community smart-bin system in the Town of Mosman Park. This solar powered network has many unique IoT integrations including patented waste sensing, air quality sensing, LoRa comms, people counting, full Dashboard and a community Recommender system. yStop has also just completed the replacement of all the City of Joondalup's seat furniture stock (the first in 38 years) with self-reporting units that will - amongst other features - report on vandalism (tilt angle) and local environmental conditions including temperature, soil moisture level and air quality.

Peter Cole is a Senior Lecturer at Murdoch University, Perth, Australia. He is a Fellow of the Australian Computer Society and a Fellow of the Australian Institute of Management. Peter was formerly the Dean of the School of Information Technology at Murdoch University for over 8 years and served at the president of the Australian Council of Deans of ICT. Peter is the Australian Representative and Senior Assessor on the Standards and Accreditation Council (SAC) of the International Federation of Information Processing, International Professional Practice Partnership (IP3). His research interests are in Computer Vision and Intelligent Systems, Computer Languages and Communications Technology. Peter actively sources and supervises innovative projects with industry and academia.

Giles Oatley has worked for 20 years in the area of data mining, decision support systems and modelling human behaviour. This includes an interest in spatiotemporal analysis, public health surveillance and sensor systems, currently applied to modelling populations in a smart city context, leading to this reported project. Dr Oatley's research has always contained a strong applied component, and in the UK he worked on various projects (sponsored by DTI, ERDF) related to taking smart/intelligent systems to industry partners, and his interest has continued in Australia including work with yStop, the industrial partners in this project.

Polychronis Koutsakis received his Ph.D. in Electronic and Computer Engineering from the Technical University of Crete, Greece. From July 2006 till December 2008 he was an Assistant Professor at the Electrical and Computer Engineering Department of McMaster University, Canada. In January 2009 he joined the School of Electronic and Computer Engineering of the Technical University of Crete, where he received tenure as an Associate Professor in 2014. In January 2016 he joined Murdoch University. He is a Senior Member of the IEEE and has been honored three times as Exemplary Editor of the IEEE Communications Society, for his work as an Editor of the IEEE Communication Surveys and Tutorials Journal. Dr. Koutsakis has served as the general chair of the IEEE WoWMoM 2018. He has authored more than 120 peer-reviewed papers, most of them on the design and performance evaluation of computer networks, and is the co-inventor of 1 US patent acquired by Blackberry Ltd.