Dataset: User side acquisition of People-Centric Sensing in the Internet-of-Things

Chenguang Liu, Jie Hua, Tomasz Kalbarczyk, Sangsu Lee, and Christine Julien
The University of Texas at Austin
{liuchg,mich94hj,tkalbar,sethlee,c.julien}@utexas.edu

ABSTRACT

We envision a world where people-centric sensing and personalized services can be achieved without centralized data collection and processing or the reliance on video-based surveillance. To help assess the feasibility of building context-aware applications while allowing users to fully control their potentially sensitive data, we studied the feasibility of people-centric sensing using only user-side data acquisition. Using only off-the-shelf, energy restricted sensor kits in a smart environment together with an energy-efficient message exchange scheme implemented on top of Bluetooth Low Energy, the collected dataset provides insight on a continuous cyber-physical view from users' individual perspectives. The availability of the dataset encourages further studies of users' activities, for instance to perform distributed inference on users' social interactions and activity trajectories.

CCS CONCEPTS

• **Human-centered computing → Ubiquitous and mobile computing**; • **Networks** → *Network properties*; • **Computer systems organization** → *Embedded and cyber-physical systems.*

KEYWORDS

Datasets, People-Centric Sensing, Internet of Things

ACM Reference Format:

Chenguang Liu, Jie Hua, Tomasz Kalbarczyk, Sangsu Lee, and Christine Julien. 2019. Dataset: User side acquisition of People-Centric Sensing in the Internet-of-Things. In *The 2nd Workshop on Data Acquisition To Analysis (DATA'19), November 10, 2019, New York, NY, USA.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3359427.3361914

1 INTRODUCTION

In people-centric sensing [3], individuals utilize sensors to learn and share information about themselves and their environments to enable personalized digital services or contribute to a social good. With technological advances in embedded systems and wireless communication, today's low power Internet-of-Things (IoT) devices enable numerous, formerly impossible opportunities for sensing and sharing information about the ambient environment. In this work, we are interested to how user-side data acquisition can stimulate the fusion of personal, public, and social sensing in a privacy-preserving way.

Wireless fingerprinting, in particular via Bluetooth technology, is known to be useful to provide distance estimations in indoor localization [5, 11]. Recent years have also seen an increasing interest in leveraging Bluetooth beacons as a tool for device interaction in many application-specific contexts, including tourism [12, 17], device-to-device collaboration [8, 10], social interactions [2], security [13], smart buildings [16], and smart cities [1]. However, most publicly available datasets are collected from service-side devices (e.g., devices connected on unmetered networks). There is a lack of datasets directly acquired from the user side with context information that can correlate to device-to-device or human-to-device interactions. In this work we focus on the idea of people-centric sensing and provide a dataset containing *continuous* views of users' digital surroundings with rich contextual information. We introduce the setup of our data acquisition in Section 2. Section 3 provides details of the presented dataset. Lastly we discuss the potential usage of this dataset in Section 6.

2 HARDWARE SETUP AND CONTINUOUS COLLECTION

To keep our study focused on data collection and make it easy to repeat in other environments, our hardware setup uses off-the-shelf Nordic Thingy52 IoT sensor kits [15]. The device is equipped with eight on-board sensors, Bluetooth Low Energy (BLE) for wireless connectivity and a rechargeable lithium polymer battery. BLEnd [6] is implemented on top of the BLE stack as the communication substrate. In the deployment, the devices continuously transmit beacons that can be received by other devices in range. In particular, we parameterize the BLEnd protocol on each device to achieve a 95% guarantee that each neighboring device receives that device's beacon every four seconds.

In our deployment, we programmed and deployed 55 IoT sensor kits during the 6-week data collection period. Of these kits, 48 were placed as anchor nodes at 24 locations inside a university academic building[1]. These stationary beacons (illustrated in Figure 1) were programmed as pairs; at any time, one device in a pair was deployed at the anchor location, while the other was pulled for battery charging. The remaining 7 sensor kits were carried by the human participants in the data collection. All devices beaconed as described above. Each human participant also carried an Android device that also collected any overheard beacons and stored the data from all received beacons into a local SQLite database. Finally, the deployment also included three Android tablets installed at three separate building locations (the two doors to an interior space

[1]http://www.ece.utexas.edu/about/facilities/eerc

DATA'19, November 10, 2019, New York, NY, USA

ACM ISBN 978-1-4503-6993-0/19/11…$15.00
https://doi.org/10.1145/3359427.3361914

Table 1: Sample Data Entries in the Beacon Table

HostId	*Timestamp(ms)*	*Bluetooth Address*	*RSSI*	*Sound avg.*	Sound max	...	*Temp.*	*Humidity*	*Air Pressure*	eCO^2*(ppm)*	*tVOC(ppb)*
4	1554330245495	C1:DA:6A:2A:4E:D5	-67	269	669	...	22.96	48%	994.88	427	4
1	1556731374232	D1:DE:0B:60:CA:95	-99	1215	2193	...	27.70	73%	996.43	400	0
0	1557160392903	F0:55:C1:1B:5D:6B	-98	306	669	...	23.29	86%	993.97	408	1

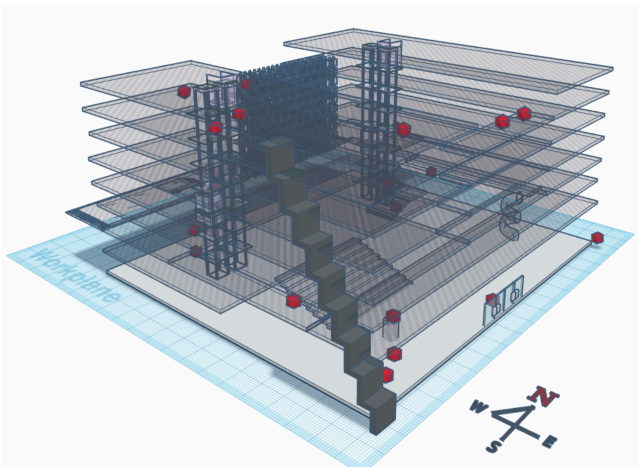

Figure 1: Static Beacon Deployment

and an open cubicle work area). The deployment used these tablets to collect participants' explicit "check-ins," which served as ground truth participant location information.

3 DATA DESCRIPTION

We collected the beacon data from the local databases on the participants' Android devices into a single merged database. Within this database, each "row" contains information about one received beacon, as shown in Table 1. These are indexed by the identifier of the participant device that received the beacon (i.e., *HostId*, a value in the range [0..6]). Each row contains information about the received beacon: a description of the received beacon including the timestamp, the received signal strength indicator (RSSI), the sender's Bluetooth address, and the receiver's id. The data also includes the summary of the sensor information read from the sender's onboard sensors, including sound level indicators (peak and average), temperature, air quality measurements (eCO^2 particles per million and volatile organic compound particles per billion), air pressure, and humidity. During the data collection period, it was possible for a device to receive more than one beacon from a given neighboring device within a short time period. Because the phenomena being sensed were unlikely to change at a very high frequency, we down sampled the beacons received to one second (i.e., each device stores at most one beacon from any other device each second). Our participants also carried the Android devices for 24 hours; we pruned any data collected off-campus for privacy reasons. The main table in our final dataset contains 20,612,286 entries. As described above, we used tablet check-in locations to collect additional ground truth labels for the data set; we stored these events in a second table as a time series of *ParticipantId*, check-in location, and timestamp.

4 USES

The data set has a variety of potential use cases.

Composition: This dataset leverages user-side acquired beacons that carry a rich set of ambient context information that can be useful, for example, for mining the correlation between context snapshots and human activity recognition and prediction [4, 14]. Context snapshots can be derived from the beacon entries in the main *beacon table* at any aggregation level as needed. Inter-human encounters can be inferred from the senders' Bluetooth addresses, which are cross-referenced in the *device description table*. While the relative location of the hosts could be estimated using the *RSSI* of the beacons sent from the stationary nodes (category provided as part of device description), the *check-in table* offers explicit human-to-machine interactions.

Beacon Carried Context: The sensed attributes contained in each beacon are sampled with the same interval from the device (identified by its Bluetooth address). Note that time consumption for sampling can vary between different types of sensors. Most sensors can return the readings back to the controller with negligible delay. The air quality sensor and microphone, however, need more sampling time. The accuracy of the embedded sensors are referred to the hardware specification in [15]. Depending on the use case, certain pre-processing or noise reduction should be applied when extracting higher-level contexts from the raw data. For instance, characteristics of *RSSI* (e.g., multi-path fading) [18] need to be considered when deriving proximity information.

Example Usage: The presented dataset can be explored from both user and environment perspectives. The nature of user-side acquisition makes it straightforward to navigate the context changes for a given user through time. For instance, one of the things the first row in Table 1 can tell us is that *host* #4 is relatively close to a device d_i(C1:DA:6A:2A:4E:D5) at time t. Then we can do a simple filter on the *HostId* and timestamp columns to figure out what other devices *host* #4 had encountered in an arbitrary look back time(e.g. last five minutes). Alternatively, we could make predictions on what *host* #4 would encounter in the next five minutes and the subsequently assess the accuracy of these predictions. From the environment perspective, we have leveraged the dataset to develop an approach for *continuous authorization* in smart buildings [7]. The dataset allows us to use the frequency of beacons within a given interval of time to reflect the "presence" of a device from the perspective of a *host*. The density of beacons in the dataset is sufficient for us to generate useful access control rules at the smart building scale, based on attributes derived from this "presence" information. Existing datasets do not provide the necessary granularity of beacons necessary to support these types of applications.

5 DATA ACCESS

The data collection can be found at [9].

6 CONCLUSION

The presented dataset provides a new perspective on studying people-centric sensing that is essential in continuous context-aware applications. The people-centric nature of the dataset lends itself to the development of privacy conscious applications, where the sensors take a passive role and the users are in control of the sensor aggregation. In the future, the proximity based message exchanges captured in the dataset can be used to evaluate applications or scenarios to exploit social or personalized services without the assumption of a central collector.

7 ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. CNS-1703497.

REFERENCES

[1] J. Adkins, B. Ghena, N. Jackson, P. Pannuto, S. Rohrer, B. Campbell, and P. Dutta. 2018. The signpost platform for city-scale sensing. In *Proc. of IPSN*. ACM/IEEE, 188–199.

[2] A. Biri, P. Pannuto, and P. Dutta. 2019. Demo Abstract: TotTernary-A Wearable Platform for Social Interaction Tracking. In *Proc. of IPSN*. ACM/IEEE, 346–347.

[3] A. T. Campbell, N. D. Lane, E. Miluzzo, R. A. Peterson, H. Lu, X. Zheng, M. Musolesi, K. Fodor, S. B. Eisenman, and G.-S. Ahn. 2008. The rise of people-centric sensing. *IEEE Internet Computing* 4 (2008), 12–21.

[4] D. De, P. Bharti, S. K. Das, and S. Chellappan. 2015. Multimodal wearable sensing for fine-grained activity recognition in healthcare. *IEEE Internet Computing* 19, 5 (2015), 26–35.

[5] R. Faragher and R. Harle. 2015. Location fingerprinting with bluetooth low energy beacons. *IEEE journal on Selected Areas in Communications* 33, 11 (2015), 2418–2428.

[6] C. Julien, C. Liu, A. L. Murphy, and G. P. Picco. 2017. Blend: practical continuous neighbor discovery for bluetooth low energy. In *Proc. of IPSN*. ACM/IEEE, 105–116.

[7] T. Kalbarczyk, C. Liu, J. Hua, and C. Julien. 2019. LAD: Learning Access Control Polices and Detecting Access Anomalies in Smart Environments. In *Proc. of MASS*. IEEE.

[8] C. Liu, J. Hua, and C. Julien. 2019. Scents: Collaborative sensing in proximity IoT networks. In *Proc. of PerCom Workshops*. IEEE, 189–195.

[9] C. Liu, J. Hua, T. Kalbarczyk, S. Lee, and C. Julien. 2019. Dataset: User side acquisition of People-Centric Sensing in the Internet-of-Things. https://doi.org/10.5281/zenodo.3450690

[10] C. Liu, C. Julien, and A.L. Murphy. 2018. PINCH: Self-organized context neighborhoods for smart environments. In *Proc. of SASO)*. IEEE, 120–129.

[11] D. Lymberopoulos, J. Liu, X. Yang, R.R. Choudhury, V. Handziski, and S. Sen. 2015. A realistic evaluation and comparison of indoor location technologies: Experiences and lessons learned. In *Proc. of IPSN*. ACM/IEEE, 178–189.

[12] C. Martella, A. Miraglia, M. Cattani, and M. Van Steen. 2016. Leveraging proximity sensing to mine the behavior of museum visitors. In *Proc. of PerCom*. IEEE, 1–9.

[13] M. Miettinen, S. Heuser, W. Kronz, A.-R. Sadeghi, and N. Asokan. 2014. ConXsense: automated context classification for context-aware access control. In *Proc. of ASIACCS*. ACM, 293–304.

[14] S. Nath. 2012. ACE: exploiting correlation for energy-efficient and continuous context sensing. In *Proc. of MobiSys*. 29–42.

[15] Nordic. [n. d.]. Nordic Thingy52 Sensor Tag. https://www.nordicsemi.com/eng/Products/Nordic-Thingy-52.

[16] M.S. Rahaman, H. Pare, J. Liono, F. D. Salim, Y. Ren, J. Chan, S. Kudo, T. Rawling, and A. Sinickas. 2019. OccuSpace: Towards a Robust Occupancy Prediction System for Activity Based Workplace. In *Proc. of PerCom Workshops*. IEEE.

[17] M. Saloni, C. Julien, A.L. Murphy, and G.P. Picco. 2017. LASSO: A Device-to-Device Group Monitoring Service for Smart Cities.

[18] R. Wu, Y. Lee, and Y. Jan H. Tseng, and M. Chuang. 2008. Study of characteristics of RSSI signal. In *2008 IEEE International Conference on Industrial Technology*. IEEE, 1–3.