

Location Determination based on WiFi Infrastructure

Yiming Lin, Daokun Jiang, Roberto Yus, Sharad Mehrotra, Nalini Venkatasubramanian

Department of Computer Science

University of California, Irvine

{yiminl18,daokunj,ryuspeir}@uci.edu, {sharad,nalini}@ics.uci.edu

ABSTRACT

PVLDB Reference Format:

Ben Trovato, G. K. M. Tobin, Lars Thørväld, Lawrence P. Leipuner, Sean Fogarty, Charles Palmer, John Smith, Julius P. Kumquat, and Ahmet Sacan. A Sample Proceedings of the VLDB Endowment Paper in LaTeX Format. *PVLDB*, 12(xxx): xxxx-yyyy, 2019.

DOI: <https://doi.org/10.14778/xxxxxxx.xxxxxxx>

1. INTRODUCTION

Locating people within buildings is important for variety of application context, e.g., occupancy based HVAC control, building analytic services that characterize how buildings are used, and location-based services such as where an employee is, etc. For many of these applications, locating up to the room level is critical for accuracy. As an instance, locating people in room level enables fine-grained occupancy computation, leading to more precise HVAC control [1].

While several indoor localization technologies exist, they either require the active participation of devices (active localization [2]), such as WiFi fingerprinting based systems [3], or require extra hardware, such as Bluetooth [4], RFID [5], Visible Light [6], Ultra Wideband [7], acoustic signal [8] based, etc. However, they fall short in at least one of the following aspects, which prevent them from being widely adopted. First, hardware deployment can be prohibitive in terms of cost in large deployments and/or may not be possible due to physical restrictions of space; Second, users are not always willing to cooperate in the localization process as it requires the execution of additional services on their devices which, in turn, might incur additional energy consumption. In addition, users may not bother installing special software either due to inertia or since they may not value the services offered to be important; Third, the platform and device heterogeneity prevent active localization from being applied to all platforms [9]. For instance, Apple does not publish programming APIs for WiFi scanning on iOS-based devices [10] hence, making active WiFi localization hard to achieve for third-party applications on iOS-based devices.

This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>. For any use beyond those covered by this license, obtain permission by emailing info@vldb.org. Copyright is held by the owner/author(s). Publication rights licensed to the VLDB Endowment.

Proceedings of the VLDB Endowment, Vol. 12, No. xxx
ISSN 2150-8097.

DOI: <https://doi.org/10.14778/xxxxxxx.xxxxxxx>

In contrast WiFi Access Point (AP) is passive, i.e., WiFi connectivity data is generated using SNMP traps over existing WiFi routers and access points. Note that all the passive localization work so far requires extra equipment, such as WiFi monitors [11, 12, 13, 14] and thus it owns similar drawback as the above analysis.

In this paper, we study a method that uses WiFi data to get location data up to room level, which do not require any extra hardware or explicit cooperation of the user device. Since WiFi connectivity is ubiquitous, if we can use it to locate up to room level reliably, it can serve as a vehicle for several applications.

Exploiting WiFi AP connectivity, however, is challenge for several reasons: a) devices only connect sporadically. A missing data might not actually mean user not there; b) A WiFi AP covers several rooms, and thus a connection to an AP only tells you the several candidate rooms but not exact one. c) WiFi connectivity data often comes at large volume and high speed (stream), and often involves great number of devices, thus efficiency and scalability have to be taken into account. We postulate the problem as that of *WiFi data cleaning* problems. We conclude WiFi connectivity data as dirty for LBS in the following two aspects. First, there exists some time interval where no connectivity is captured by WiFi APs. We need to identify if the device is out of the building or disconnected but in the building, which is formulated as xxx. ¹; Second, it is hard to generate fine-grained location information, i.e., room-level location. Considering a typical scenario, when a device is connected to some WiFi access point (AP), we can conclude that the device is in the region covered by that AP, but no idea of which exact room as a region often covers multiple rooms. The goal is find the exact room among candidates, which could be viewed as a *value disambiguation* problem. As an illustrating example, WiFi connectivity data comes in the format of {Mac Address, WiFi AP, Timestamp}. As shown in Table 1, for device with Mac Address *5f0b*, it is captured by *AP₁* at “14:16:12” and “14:25:36”, but we do not know the location information between them. Another device with Mac address *258e* connected to WiFi AP *2* at time “2018-03-14 12:03:35”. Now we can draw that this device must be in one of the rooms covered by *AP₂*, which are {2002, 2004, 2014, 2012, 2008, 2066, 2051, 2059, 2052, 2054, 2056, 2058} (From Fig 1) Our goal is to decide the exact room the device is in.

¹daokun: to summary dirty data from the aspect of gaps
Question: Can we call gap dirty data?

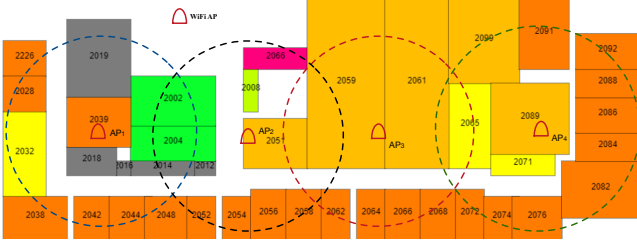


Figure 1: Example of a space with different types of rooms and four WiFi APs.

Table 1: Sample connectivity data from our experiments.

MAC Address	WiFi AP	Timestamp
258e...	2	2018-03-14 12:03:35
258e...	3	2018-03-14 14:15:53
5f0b...	1	2018-03-14 14:16:12
5f0b...	1	2018-03-14 14:25:36

To address the above challenges in developing fine grained location based on AP data, our approach exploits contextual information that is present in the connectivity data and/or metadata about users. In particular, we leverage knowledge from historical connectivity data such as the APs to which a user device typically connects, which other devices are connected to the same APs at the same time, and the affinity of a user to specific areas in the space. This knowledge can be extracted from spaces where their inhabitants have predictable patterns (e.g., an office or a university building), which are the main scope of our approach. Leveraging such models, our system is able to calculate the probability of a device being present in the space but disconnected from the network. Furthermore, for such events, our approach can compute a prediction on which AP the disconnected device would be connected to based on the context such as the day and time. Then, it predicts the room in which the device would be located given the AP to which it is connected or predicted to be connected. This is done based on a metric of affinity among devices and between devices and rooms based on which we propose a graph-based algorithm for prediction.

Sensor data cleaning has been considered in the past, such as RFID data cleaning [15, 16].²

In summary, we make the following contributions. a) We first formally define the WiFi data cleaning problem and split it into three subtasks in building, region and room level; b) we propose a hybrid location determination algorithm that combines graph-based and rule-based methods to exploit contextual information to disambiguate the room a device is located in; c)³; d) We optimize the above techniques and make it scale to large number of devices by proposing on-line location determination; e) We show experiments to validate our approach using WiFi connectivity data captured from the 64 WiFi APs in our six-story building over a period of one year.

The rest of paper is structured as follows. Section 2 defines problems formally. Section 3 and 4 describes the techniques to clean WiFi data in building, region and room levels. Scal-

ability issues are discussed in Section 5. Section 6 evaluates our methods, Section 7 presents the related work and Section 8 concludes paper.

2. PROBLEM DEFINITION

In this section, we describe our problem formulation more precisely and overview our methods. We begin by first describing notations we will need.

Space Model. We model space as a hierarchy of *buildings*, *regions*, and *rooms* though our approach can easily be generalized to other spatial models. Regions, as will become clear, are defined by the coverage of a WiFi AP. Further, rooms are classified into two types: *public*, which are shared by multiple users generally (e.g., meeting rooms, lounges, kitchens, food courts, etc.); and *private*, which are owned by a specific user(s) (e.g., a person’s office). This classification of rooms is metadata that is commonly available and it will be exploited by our approach. In the example of Fig. 1 the building is “Donald Bren Hall building” at UCI and it contains, among others, the rooms in the figure (e.g., 2065, 2072, etc.). Notice that rooms have been color coded based on their type – dark orange rooms are offices (which we classify as private), whereas yellow, light orange, and green are conference rooms, laboratories, and rest rooms (which are considered public).

WiFi AP Connectivity Data For Localization. We assume that a given space is covered by several *WiFi APs* that are used to provide users with Internet connectivity. For example, in Fig. 1, there are four WiFi APs covering the space. Each WiFi AP has an associated coverage [17] that models the region of the space that the AP serves - coverage of two WiFi APs may overlap.

We assume that WiFi APs log connectivity events when a device connects to a certain WiFi AP like the ones in Table 1. Such events can be captured by current WiFi APs using SNMP traps, for instance. A connectivity event generally includes information such as the MAC address of the device, the MAC address (or id) of the WiFi AP, and the timestamp of the event. Since WiFi APs are fixed in location, such an event localizes the connected device to the region associated with the corresponding WiFi AP.

3. ROOM DETERMINATION

4. BUILDING DETERMINATION

5. SCALABILITY

6. EVALUATION

7. RELATED WORK

8. CONCLUSION

9. REFERENCES

- [1] A. Afram and F. Janabi-Sharifi, “Theory and applications of hvac control systems—a review of model predictive control (mpc),” *Building and Environment*, vol. 72, pp. 343–355, 2014.

²daokun: explain how our problem differs with them in concise words

³daokun: summary contributions of your part

- [2] G. Deak, K. Curran, and J. Condell, "A survey of active and passive indoor localisation systems," *Computer Communications*, vol. 35, no. 16, pp. 1939–1954, 2012.
- [3] M. Kotaru, K. Joshi, D. Bharadia, and S. Katti, "Spotfi: Decimeter level localization using wifi," in *ACM SIGCOMM Computer Communication Review*, vol. 45, no. 4. ACM, 2015, pp. 269–282.
- [4] J. J. Diaz, R. d. A. Maues, R. B. Soares, E. F. Nakamura, and C. M. Figueiredo, "Bluepass: An indoor bluetooth-based localization system for mobile applications," in *Computers and Communications (ISCC), 2010 IEEE Symposium on*. IEEE, 2010, pp. 778–783.
- [5] A. A. N. Shirehjini, A. Yassine, and S. Shirmohammadi, "An rfid-based position and orientation measurement system for mobile objects in intelligent environments," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 6, pp. 1664–1675, 2012.
- [6] Y.-S. Kuo, P. Pannuto, K.-J. Hsiao, and P. Dutta, "Luxapose: Indoor positioning with mobile phones and visible light," in *Proceedings of the 20th annual international conference on Mobile computing and networking*. ACM, 2014, pp. 447–458.
- [7] S. Gezici, Z. Tian, G. B. Giannakis, H. Kobayashi, A. F. Molisch, H. V. Poor, and Z. Sahinoglu, "Localization via ultra-wideband radios: a look at positioning aspects for future sensor networks," *IEEE signal processing magazine*, vol. 22, no. 4, pp. 70–84, 2005.
- [8] W. Huang, Y. Xiong, X.-Y. Li, H. Lin, X. Mao, P. Yang, Y. Liu, and X. Wang, "Swadloon: Direction finding and indoor localization using acoustic signal by shaking smartphones," *IEEE Transactions on Mobile Computing*, vol. 14, no. 10, pp. 2145–2157, 2015.
- [9] "Apple developers: Available: <https://developer.apple.com/>."
- [10] D. Zeinalipour-Yazti and C. Laoudias, "The anatomy of the anyplace indoor navigation service," *SIGSPATIAL Special*, vol. 9, no. 2, pp. 3–10, 2017.
- [11] M. Youssef, M. Mah, and A. Agrawala, "Challenges: device-free passive localization for wireless environments," in *Proceedings of the 13th annual ACM international conference on Mobile computing and networking*. ACM, 2007, pp. 222–229.
- [12] C. Xu, B. Firner, R. S. Moore, Y. Zhang, W. Trappe, R. Howard, F. Zhang, and N. An, "Scpl: indoor device-free multi-subject counting and localization using radio signal strength," in *Proceedings of the 12th international conference on Information processing in sensor networks*. ACM, 2013, pp. 79–90.
- [13] M. Seifeldin, A. Saeed, A. E. Kosba, A. El-Keyi, and M. Youssef, "Nuzzer: A large-scale device-free passive localization system for wireless environments," *IEEE Transactions on Mobile Computing*, vol. 12, no. 7, pp. 1321–1334, 2013.
- [14] A. Musa and J. Eriksson, "Tracking unmodified smartphones using wi-fi monitors," in *Proceedings of the 10th ACM conference on embedded network sensor systems*. ACM, 2012, pp. 281–294.
- [15] S. R. Jeffery, M. Garofalakis, and M. J. Franklin, "Adaptive cleaning for rfid data streams," in *Proceedings of the 32nd international conference on Very large data bases*. VLDB Endowment, 2006, pp. 163–174.
- [16] A. I. Baba, M. Jaeger, H. Lu, T. B. Pedersen, W.-S. Ku, and X. Xie, "Learning-based cleansing for indoor rfid data," in *Proceedings of the 2016 International Conference on Management of Data*. ACM, 2016, pp. 925–936.
- [17] J. Tervonen, M. Hartikainen, M. Heikkilä, and M. Koskela, "Applying and comparing two measurement approaches for the estimation of indoor wifi coverage," in *New Technologies, Mobility and Security (NTMS), 2016 8th IFIP International Conference on*. IEEE, 2016, pp. 1–4.