

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

The pervasiveness of WiFi infrastructure enables a large number of services based on it. In this paper we focus on Location Based Service (LBS). LBSs use the location of a device (e.g., a smart-phone) and hence of the owner to offer personalized information. Thousands of interesting applications have been built based on such localization techniques. [1] However, WiFi data is far from enough to provide location information with high quality if we use it directly.

We conclude WiFi connectivity data as dirty for LBS in the following two aspects. First, there exist some time interval where no connectivities are captured by WiFi APs. We need to identify if the device is out of building or disconnected but in 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

¹daokun: to summary dirty data from the aspect of gaps

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>

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.

The cleaning of dirty WiFi data poses several challenges. First,²; Second, value disambiguation for room-level location is very tricky only based on WiFi connectivity data as a-priori knowledge is limited. Third, since WiFi data comes on streams, the technique has to be very efficient to support online location determination; Finally, the number of devices captured by WiFi APs within a building could be large, so that we need to address scalability issue very carefully to be applicable in large systems.

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.

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.

²daokun: describe challenges in your part

³daokun: summary contributions of your part

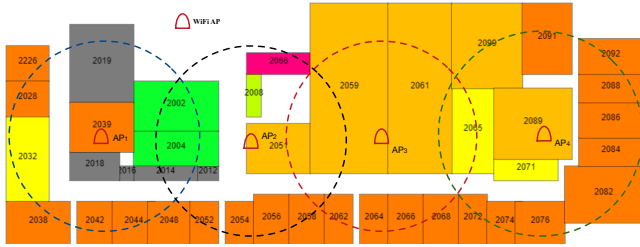


Figure 1: Example of a space with different types of rooms and four WiFi APs.

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

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. Scalability 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

3. ROOM DETERMINATION

4. BUILDING DETERMINATION

5. SCALABILITY

6. EVALUATION

7. RELATED WORK

8. CONCLUSION

9. REFERENCES

- [1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE communications magazine*, vol. 40, no. 8, pp. 102–114, 2002.