Honeycomb: Indoor location estimation based on Wi-Fi signal strength

by

Michael Linder, B.A.

REPORT

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science in Engineering

THE UNIVERSITY OF TEXAS AT AUSTIN

May 2015

Honeycomb: Indoor location estimation based on Wi-Fi signal strength

APPROVED BY

SUPERVISING COMMITTEE:

Christine Julien, Supervisor

William Bard

Dedicated to my wife, Dana, whose support made this possible.

Honeycomb: Indoor location estimation based on Wi-Fi signal strength

Michael Linder, M.S.E.
The University of Texas at Austin, 2015

Supervisor: Christine Julien

This paper presents Honeycomb, an indoor location estimation product based on Wi-Fi signal strength. Wireless Local Area Networks are ubiquitous today, and most people carry Wi-Fi capable devices in their pocket. This existing infrastructure can thus be leveraged for purposes of location estimation. Using Wi-Fi signal strength fingerprinting, Honeycomb harnesses existing Wi-Fi infrastructures as a means to track the movements of individuals through an indoor space. Fingerprinting is a method by which Wi-Fi signal strengths are mapped at regular intervals in a bounded space. Once a space is fingerprinted, a given node must simply sample Wi-Fi signal strengths as it moves through the same space and Honeycomb's algorithm will determine the node's path in an offline manner. Because Honeycomb only requires nodes to passively measure Wi-Fi signal strengths rather than send out its own beacon, it prevents malicious third parties from gaining access to any real time data, and thus maintains the security and privacy of the user. By performing location

estimations on the data collected on an independent platform, and not on the device itself, it saves the user from spending the computing power, and thus the device's battery. We believe Honeycomb to be a product unlike any other, which is suitable for deployment in multiple real world scenarios.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

In recent years wireless LAN technology has become ubiquitous. Wi-Fi access points have become virtually trivial to install, and nearly everyone carries a Wi-Fi capable mobile device in their pocket. It is also the case that much research has been done on various methods of location estimation. It follows, then, that location estimation that leverages Wi-Fi would be a valuable topic, and in fact much research has already been done in the space, including [9], [11], [10], [7], and [12].

The benefits of using Wi-Fi for location estimation are manifold. For instance, while the Global Positioning System is in many ways the premier method for location estimation in the world [4], GPS signals are often unreliable indoors [19], making it a poor choice for any indoor location estimation. Location systems that use other mechanisms such as RFID [17], radio waves, ultrasound [13], or geomagnetism [5] are difficult to setup, require specialized hardware, and ultimately can only be used for a single purpose. Wi-Fi based location estimation solves all of these problems. Wi-Fi signals are readily available indoors. Wi-Fi is relatively cheap and easy to setup, and in many cases existing access points can be leveraged.

1.1 Definitions

There are a few terms that will be used throughout this paper that it is important to define early. Understanding these definitions will help make clear the purpose of this paper and its contributions.

Signal Strength vs. RSSI Much of the research involving location estimation with Wi-Fi signal strength refers to the measured power present in the radio signal as the Received Signal Strength Indicator (RSSI). While in general terms this moniker is good enough, in truth, the IEEE 802.11 specifications [8] do not indicate a specific relationship between RSSI and the actual power level as measured in either milliwatts (mW) or Decibel-milliwatts (dBm). As such manufactures are free to provide their own arbitrary units, and RSSI measurements are generally found to be integer values greater than 0. Because of this inconsistency, Honeycomb does not use RSSI, favoring instead what we refer to as simply "signal strength". For our purposes, signal strength is a measure of the power present in the Wi-Fi signal as measured in dBm. dBm is a measurement relative to 1 mW of power, where 0 dBm is equal to 1 mW. Because dBm measurements are made on a logarithmic scale, we find our measurements to be negative integers between 0 and 100, where the measured value is the exponent in the logarithm. So, where 0 dBm is equal to 1 mW, -10 dBm is equal to .1 mW, -20 dBm is equal to .01 mW, and so on. Measuring signal strength in this way allows Honeycomb to maintain consistent tracking across signal strength measurement platforms, and thus makes Honeycomb a more diverse and viable product.

Fingerprint Throughout this paper will we refer to fingerprints. In this context, a fingerprint is a set of Wi-Fi signal strength measurements taken from a set of Wi-Fi access points at a specific point in a given space.

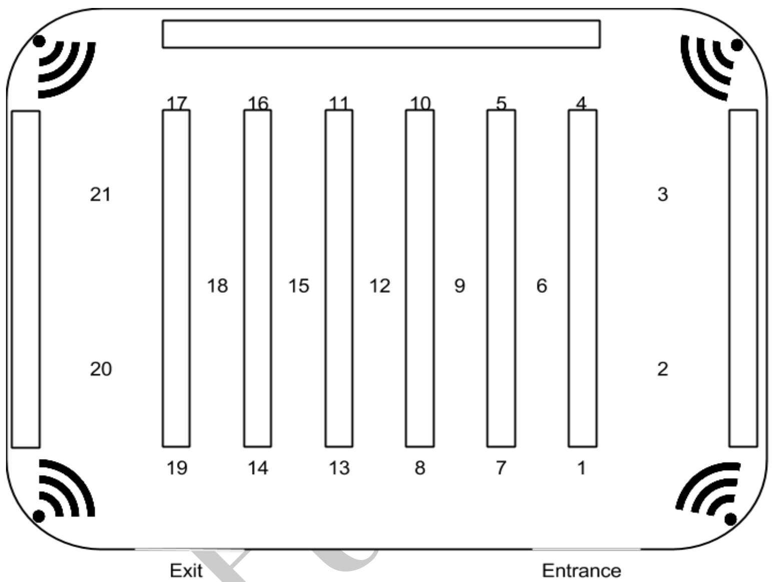

Figure 1.1: An example of a location with labeled measurement points

Fingerprint Session While the concept of the Wi-Fi fingerprint is relatively common, here we introduce a new concept that will enable a Honeycomb installation to maintain its value over time: the fingerprint session. A fingerprint session is the complete set of fingerprints from all points in a space at any given time. Because the internal layout of any given space can change over time, causing issues with blocking and reflection of Wi-Fi signals, Honeycomb has built in the ability to re-fingerprint the entire location at any time so that the measurements can be as precise as possible. Figure 1.1 shows an example location with aisles like that of a grocery store. It includes four Wi-Fi access points and numbered labels for points that may be considered relevant for location estimation. A single fingerprint session in this space would be the complete set of Wi-Fi signal strength measurements from all four access points at every numbered location.

User Track A user track differs from a fingerprint in that it represents a single user's movement through the space. Thus, a user track is composed of a set of timestamps, each of which is associated with a set of signal strength measurements for each of the access points in the space. Figure 1.2 shows an example of a user track. In this example, the small dots represent the user's actual path through the space, while each large dot represents a timestamped set of signal strength measurements. Thus it can be seen that in this example that the user walked at a relatively constant pace through the space, slowing down four times near locations 5, 7, 14, and 17. For location estimation purposes, Honeycomb will compare a given user track to the most recent fingerprint session for the given space.

1.2 Motivation

Wi-Fi based location estimation is a well researched topic [11]. The goal of Honeycomb is to leverage that research and build an indoor location

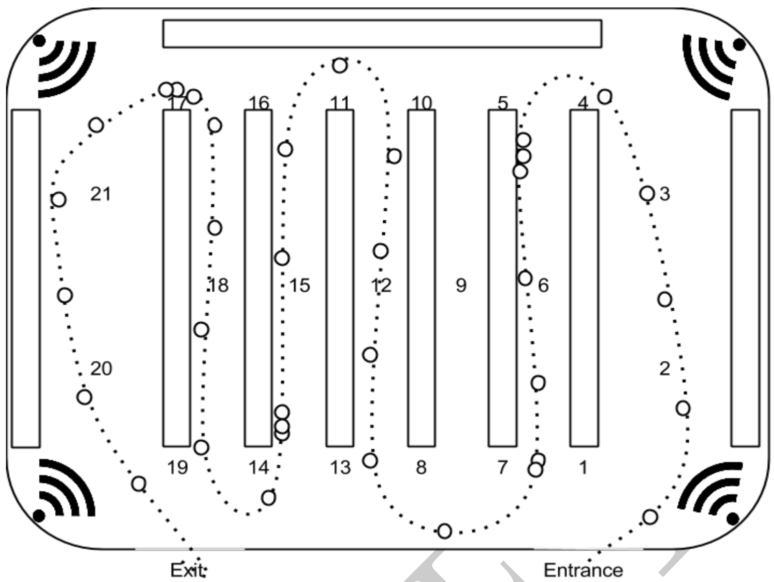

Figure 1.2: An example user track

tracking system which is suitable for deployment in a real world scenario. As such, Honeycomb includes an Android application capable of fingerprinting a space, and an API which is deployed to the web for uploading both fingerprints and user track data. The web application also executes the location estimation algorithm, and provides a user interface for browsing the user track data. Honeycomb itself remains agnostic of the mechanism used to gather the user's Wi-Fi signal strength data. By decoupling Honeycomb in this way, we allow Honeycomb to be used in multiple scenarios in which a specialized user track gathering mechanism is desired.

User privacy is also a major motivating factor in our work. By only performing location estimation based on signal strength and timestamp data passively gathered on the Wi-Fi capable device, we have avoided the pitfalls of systems that require the mobile device to send out a beacon [9] [19] which can be intercepted by malicious third parties. Additionally, the offline nature of the location estimation algorithm greatly simplifies the entire process, as real time location estimation is still a highly volatile field [18]. It also helps provide an additional layer of privacy protection for the user, as the data can easily be decoupled from any identifying information before processing.

1.3 Contribution

While there has been much research done in the space, to the author's knowledge, there does not yet exist a product on the market that achieves true location estimation via Wi-Fi signal strength measurements. Honeycomb is such a product. Honeycomb provides tools on the web for site administrators to manage their locations and view individual user tracks. It also includes an API through which fingerprints and user tracks can be uploaded, and an Android application capable of doing the fingerprinting and uploading the results to Honeycomb through the API.

1.4 User Stories

We envision Honeycomb being deployed in multiple different scenarios. Essentially, wherever there is a desire to track a person's movements through a bounded space, we believe Honeycomb to be part of a viable solution. In this section, we describe two such scenarios.

1.4.1 The Grocery Store

The canonical example, and the one to which we will refer throughout this paper, is the retail establishment that wishes to track customer movements through their space. In this case, we use the example of the grocery store. The grocery store lends itself well to this scenario due to the fact that stores are generally rather large in size and that there is a general expectation that its customers will spend most of their time moving around the space. In this scenario, we see two major benefits of customer location tracking. Although we've chosen the grocery store for this scenario, these same benefits could be applied to similar scenarios, such as large conferences with multiple rooms and displays. In this scenario, we see two major benefits to the grocery store:

Visibility High visibility of products is a valuable commodity in any retain environment. Each store can use aggregate data about its customers movement through the space to identify key, high traffic areas, and sell shelf space or ad space accordingly. Additionally, [16] shows that customers respond to engaging store layouts, which can be facilitated by customer movement data. Similarly, a conference can identify high traffic areas and place sponsor ads, or other information valuable to attendees, at the site.

Flow Control Data about how people move through a space can be used to identify bottlenecks or other poorly designed traffic areas and improve them in order to provide a better user experience for patrons. In the context of a grocery store, this could result in a generally happier clientele, which means more repeat business [16]. At a conference, this data could be used to identify popular booths, and rearrange them in such a way that will cause traffic to flow in desired patterns, either to eliminate bottlenecks or to direct traffic flow past more sponsors.

1.4.2 Security Guards

For security companies, a critical component of their service is often regular patrolling of the space by a human being. For this reason, it is crucial for the security company to make absolutely sure that the security guard actually goes on their patrols. This is often accomplished via RFID stations or QR codes located throughout the space that the guard must scan in order to prove that they were there. However, this scanning requires the security guard to be both mentally and visually distracted for the length of time required to make the scan, and therefore creates a weak point in their security that can be exploited. Passive tracking of the security guard via Wi-Fi signal strength polling eliminates this distraction, while still maintaining the necessary tracking.

Note that in the above examples, the method by which the polling data is transferred from the individual's Wi-Fi capable device into Honeycomb may be dramatically different. In the case of the grocery store, there may be some desire on the part of store management to evaluate the data before transferring it to Honeycomb, for example to credit the customer's account for their incorporated rewards system, which may be necessary as a motivation for the user to allow themselves to be tracked. A grocery store's general patterns of ingress and egress provide a natural place for the data to be collected, possibly over Wi-Fi itself, so as to be as unobtrusive to the customer as possible. Conversely, in the example of the security guard, there may not be a convenient area in which to place a data collector, since you may be tracking multiple security guards through multiple spaces, and it is not worth setting up a data collector for one individual in a given space. Additionally, obtrusiveness is not an issue, since reporting their position data is part of the security guard's job. It is for this reason that Honeycomb remains agnostic of the user data gathering mechanism, in order to provide benefit in a wider variety of areas.

1.5 Structure Of This Report

The goal of this report is to provide context for the value of a Wi-Fi signal strength based indoor location tracking system and to describe the particular implementation of Honeycomb. In Chapter 2 we discuss the state of Wi-Fi based location tracking and explain why we feel that the methods we chose were the best choices for Honeycomb. In Chapter 3 we present BumbleBee. Because Honeycomb remains agnostic of user track gathering mechanisms, we needed to choose a product that is capable of gathering user track data. BumbleBee is an independent Wi-Fi signal strength measurement tool used to collect user signal strength measurements, and was co-written by the author of this paper. In Chapter 4 we discuss the architecture of

Honeycomb and the technologies on which it was built. In Chapter 5 we present the testing procedures that were implemented and their results. In Chapter 6 we discuss the results of our tests and the future of Honeycomb as a product.

Chapter 2

Background and Related Work

Several key decisions were made in designing Honeycomb. Chief among these were the basing of our location estimation system on Wi-Fi signal strength, the fingerprinting of the space, and the subsequent offline location estimation algorightm. In this section, we review the state of Wi-Fi location estimation and explain why we made the choices that we made.

2.1 High Level Location Estimation Schemes

Liu [11] categorizes three high level location estimation schemes: triangulation, proximity, and scene analysis. In triangulation schemes, nodes are tracked based on the time of arrival (TOA) or roundtrip time of flight (RTOF) if Wi-Fi signals. These methods, while potentially extremely precise, require knowledge of the locations of access points themselves, as well as the distances between them. We consider the near plug-and-play ability of Honeycomb to be a benefit to its adopters, and thus consider this requirement to be a significant blocker to adoption. Additionally, in order to gather precise TOA and RTOF measurements, a line of sight must be maintained between the access point and the mobile node, which is not possible in the scenarios in which we expect Honeycomb to be deployed. Thus triangulation schemes were rejected immediately.

Proximity schemes generally consist of a dense array of antennas which are capable of detecting mobile nodes, and the location of the mobile node is considered to be whichever antenna detects it. These schemes thus require a lot of extra infrastructure, which we believe would be a barrier to entry for Honeycomb's expected customers. Additionally, these schemes require the mobile node to send out a beacon for the antenna to detect, which we consider to put the mobile node carrier's security and privacy in jeopardy.

Thus we are left with scene analysis schemes. Scene analysis schemes generally consist of two phases: a training phase and an estimation phase. In the training phase the location is mapped, usually via fingerprinting, and in the estimation phase a node gathers its own measurements which are then compared to the fingerprints. There are many methods for doing this comparison, which we will discuss in later sections. While scene analysis schemes are not without their own overhead, they are far preferable to both triangulation schemes and proximity schemes for Honeycomb's intended uses. For these reasons, we chose a scene analysis scheme, which we will describe further here.

2.2 Deterministic and Probabilistic Approaches

Scene analysis schemes can be generally divided into two categories: probabilistic and deterministic approaches.

Both types can require a map as a training phase, but Probabilistic approaches often require knowledge of location of access points. Probabalistic Approaches are more computationally intensive Most approaches require division of the space into equal cells, but mine doesn't. You can put the points anywhere you want them for the amount of accuracy you want.

Probabilistic: [9] [15] [7]

Deterministic: [1] [14] [12]

Unsure, but need to cite: [13] [19] [20] [3] [2] [18]

2.3 Fingerprinting

Define again. Why choose this method? something here about density of fingerprints Benefits and drawbacks: fingerprinting requires a map algorithmic determinations require knowledge of distances between access points, which is harder to get than a fingerprint map, and is not as forgiving of signals bouncing off stuff

2.4 Offline Location Estimation

As opposed to real time offline is more secure (and less creepy)

Overview: [11] Three typical location estimation schemes: triangulation, scene analysis, and proximity

Triangulation: requires knowledge of location of access points, which is more complicated and less "plug and play". Relies on Time Of Arrival or

Roundtrip Time Of Flight Requires Line Of Sight Suffers from the multipath effect, due to reflection of signals Much more computationally intensive.

Scene Analysis requires knowledge of the scene (fingerprinting) two phases: training and estimation (actually calls them ”offline” and ”online” but eff that) ”There are at least five location fingerprinting-based positioning algorithms using pattern recognition technique so far: probabilistic methods, Formula-nearest-neighbor (Formula NN), neural networks, support vector machine (SVM), and smallest M-vertex polygon (SMP).” For probabilistic methods, see below

Proximity: dense grid of antennas, mobile node is collocated with the one that detects it

[9] probabilistic in nature estimation is done on the device, specifically to keep the data out of the hands of third parties. (not suitable for our purposes) uses pre-observation

[15] training phase and estimation phase requires training by making a map of wifi signal strenghts throughout the space (just like mine) more computationally intensive

[7]

[1] [14] [12]

[13] [19] [20] [3] [2] [18]

Chapter 3

BumbleBee

3.1 About BumbleBee Here

Chapter 4

Tech Overview

4.1 Components

4.2 Technologies

4.3 Architecture

Something about averaging measurements to get the fingerprint Something about euclidean distance algorithm

Chapter 5

Testing and Results

5.1 Testing Setup

5.2 Test Variants

5.3 Results

Chapter 6

Discussion

6.1 Interpretation of Results

6.2 Future Work

Do more with the track data instead of just viewing individual tracks better user interface more security

Bibliography

[1] Jagoba Arias, Aitzol Zuloaga, Jesús Lázaro, Jon Andreu, and Armando Astarloa. Malguki: an RSSI based ad hoc location algorithm. *Microprocessors and Microsystems*, 28(8):403–409, 2004.

[2] Paramvir Bahl and Venkata N Padmanabhan. RADAR: An in-building RF-based user location and tracking system. In *INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, volume 2, pages 775–784. Ieee, 2000.

[3] Paramvir Bahl, Venkata N Padmanabhan, and Anand Balachandran. Enhancements to the RADAR user location and tracking system. Technical report, technical report, Microsoft Research, 2000.

[4] Rashmi Bajaj, Samantha Lalinda Ranaweera, and Dharma P Agrawal. GPS: location-tracking technology. *Computer*, 35(4):92–94, 2002.

[5] Jaewoo Chung, Matt Donahoe, Chris Schmandt, Ig-Jae Kim, Pedram Razavai, and Micaela Wiseman. Indoor location sensing using geomagnetism. In *Proceedings of the 9th international conference on Mobile systems, applications, and services*, pages 141–154. ACM, 2011.

[6] Prabal K Dutta and David E Culler. System software techniques for low-power operation in wireless sensor networks. In *Proceedings of the 2005 IEEE/ACM International conference on Computer-aided design*, pages 925–932. IEEE Computer Society, 2005.

[7] S Hotta, Y Hada, and Y Yaginuma. A robust room-level localization method based on transition probability for indoor environments. In *Indoor Positioning and Indoor Navigation (IPIN), 2012 International Conference on*, pages 1–8. IEEE, 2012.

[8] IEEE 802.1: Wireless LANs. `http://standards.ieee.org/about/get/802/802.11.html`.

[9] Seigo Ito and Nobuo Kawaguchi. Bayesian based location estimation system using wireless LAN. In *Pervasive Computing and Communications Workshops, 2005. PerCom 2005 Workshops. Third IEEE International Conference on*, pages 273–278. IEEE, 2005.

[10] Nobuo Kawaguchi. WiFi location information system for both indoors and outdoors. In *Distributed Computing, Artificial Intelligence, Bioinformatics, Soft Computing, and Ambient Assisted Living*, pages 638–645. Springer, 2009.

[11] Hui Liu, Houshang Darabi, Pat Banerjee, and Jing Liu. Survey of wireless indoor positioning techniques and systems. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 37(6):1067–1080, 2007.

[12] Tomotaka Nagaosa and Hironori Iguchi. Performance evaluation of a Wireless LAN positioning system using spot information. In *ITS Telecommunications (ITST), 2012 12th International Conference on*, pages 512–516. IEEE, 2012.

[13] Nissanka Bodhi Priyantha. *The cricket indoor location system*. PhD thesis, Massachusetts Institute of Technology, 2005.

[14] Michael Quan, Eduardo Navarro, and Benjamin Peuker. Wi-Fi Localization Using RSSI Fingerprinting. 2010.

[15] Vinay Seshadri, Gergely V Zaruba, and Manfred Huber. A bayesian sampling approach to in-door localization of wireless devices using received signal strength indication. In *Pervasive Computing and Communications, 2005. PerCom 2005. Third IEEE International Conference on*, pages 75–84. IEEE, 2005.

[16] Paurav Shukla and Barry J Babin. Effects of consumer psychographics and store characteristics in influencing shopping value and store switching. *Journal of Consumer Behaviour*, 12(3):194–203, 2013.

[17] Edip Toplan and Cem Ersoy. RFID based indoor location determination for elderly tracking. In *Signal Processing and Communications Applications Conference (SIU), 2012 20th*, pages 1–4. IEEE, 2012.

[18] Daniel Turner, Stefan Savage, and Alex C Snoeren. On the empirical performance of self-calibrating wifi location systems. In *Local Computer*

Networks (LCN), 2011 IEEE 36th Conference on, pages 76–84. IEEE, 2011.

[19] Jie Xiong and Kyle Jamieson. Towards fine-grained radio-based indoor location. In *Proceedings of the Twelfth Workshop on Mobile Computing Systems & Applications*, page 13. ACM, 2012.

[20] Jie Xiong and Kyle Jamieson. ArrayTrack: A Fine-Grained Indoor Location System. In *NSDI*, pages 71–84, 2013.

[21] Rong Xu, Zhiyuan Li, Cheng Wang, and Peifeng Ni. Impact of data compression on energy consumption of wireless-networked handheld devices. In *Distributed Computing Systems, 2003. Proceedings. 23rd International Conference on*, pages 302–311. IEEE, 2003.

Vita

TODO: VITA

Permanent address: mplinder@utexas.edu

This report was typeset with LaTeX[†] by the author.

[†] LaTeX is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's TeX Program.