

ESTIMATING THE NUMBER OF PEOPLE USING WI-FI MONITORING

A Thesis

by

HAI LI

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirement for the degree of

MASTER OF SCIENCE

Chair of Committee,	Jean-Francois Chamberland
Committee Members,	Gregory Huff
	Anxiao Jiang
	Tie Liu
Head of Department,	Miroslav M. Begovic

December 2016

Major Subject: Electrical and Computer Engineering

Copyright 2016 HAI LI

ABSTRACT

In various situations, there are demands for estimating the number of people in an specific area. This article focuses on estimating the number of device in a certain area based on Wi-Fi metadata. To accomplish this, four sensing devices are placed at the four points of a rectangular area respectively. The sensing devices can observe and record all local data packets under monitoring mode. For each sensing device, both directional and isotropic antennas are used to detect packets separately. Each sensing device retrieves the received signal strength indicators and the media access control addresses from the 802.11 frames packets transmitted by the active wireless devices nearby. The estimator takes received signal strength indicators as input and infers the number of active Wi-Fi devices inside the specific area. Two algorithms, bayesian and maximum-likelihood are employed for the schemes. We also compare the performances between directional antenna and isotropic antenna. The result shows by using the directional monitoring antennas, we can obtain better accuracy.

This dissertation work is dedicated to my advisors - Dr. J.F.Chamberland, my father, my mother, my friends for their support through out the process.

NOMENCLATURE

RSSI	Received Signal Strength Indicator
MAC	Media Access Control
OGAPS	Office and Graduate and Professional Studies at Texas A&M University
B/CS	Bryan and College Station
TAMU	Texas A&M University

TABLE OF CONTENTS

	Page
ABSTRACT	ii
DEDICATION	iii
NOMENCLATURE	iv
TABLE OF CONTENTS	v
LIST OF FIGURES	vii
LIST OF TABLES	viii
1 INTRODUCTION	1
2 RELATED WORK	3
3 PROBLEM FORMULATION	4
4 ESTIMATION SCHEMES	6
4.1 Bayes Estimation	6
4.2 Maximum likelihood estimation	7
5 NUMERICAL SIMULATION	9
6 EXPERIMENT IMPLEMENTATION	10
7 CONCLUSION	11
REFERENCES	12
APPENDIX A MISCELLANEOUS	14
A.1 Figures/Tables in Appendix	14

A.1.1	TEST1	14
A.1.2	TEST2	14
A.2	Random Pictures and Test	15
A.3	Misc Test	15
APPENDIX B SOURCE CODE		16
B.1	Misc Test 2	16
B.2	Misc Test 3	16
B.3	Resource Usage	16

LIST OF FIGURES

FIGURE	Page
--------	------

LIST OF TABLES

TABLE

Page

1 INTRODUCTION

The number of Wi-Fi access points and the number of Wi-Fi client devices have been increasing in recent years. According to Cisco Systems prediction in their Visual Network Index, 55 percent of total mobile data traffic will be offloaded onto fixed networks through Wi-Fi access points and femtocells by 2020 [1]. Modern smartphones with Wi-Fi module transmit Wi-Fi messages periodically. Therefore, it provides us a chance to estimate occupancy through Wi-Fi packets. By deploying Wi-Fi monitoring devices in an area of interest, it is possible to detect these Wi-Fi transmissions. Each Wi-Fi transmission contains MAC address and RSSI. The MAC address is the identifier of the device while RSSI indicates the physical distance between transmitter and receiver. In the research, we focus on occupancy estimation based on Wi-Fi packets and analyze the benefits using specific directional antenna. In this paper, we apply Bayes estimation scheme and maximum likelihood scheme to estimate occupancy. A numerical simulation based on these two schemes shows the performances corresponding to sensing devices with directional antennas and isotropic antennas. In addition, to evaluate the simulation results, we construct an experimental environment in 100j parking lot. In the experiment we use four sensing devices and deploy them in the four corners of area of interest, respectively. The rest of this paper is organized as below. In section 2, we explain related work. In section 3, we introduce the problem formulation. In section 4, we propose two estimation schemes. In section 5, we give the simulation results of performance. In section 6, we show the experiment result

along with description of experiment setup. Finally we conclude this paper in section 7.

2 RELATED WORK

There have been many researches for occupancy estimation. They are based either on camera or RF signals. For approaches based on camera, they used cameras to capture images and estimate the number of people in a crowded scene [2, 3]. However, in the camera-based approaches, the estimation accuracy is affected by many factors such as the brightness of the image resolution. In addition, the camera-based approaches is limited by high deployment cost. Other approaches based on RF signals attracted a lot of attention recently. Several methods are based on Bluetooth [4] or Wi-Fi signals [5]. However the short transmission range limits the performance of Bluetooth-based methods. A research compared the Wi-Fi with Bluetooth. The authors conclude that Wi-Fi has advantage over Bluetooth in monitoring people, due to shorter discovery time and higher detection rates[6]. According to their results, only five percent of all discovered unique devices at several locations are discovered via Bluetooth and over 90.

3 PROBLEM FORMULATION

Consider a scenario where several wireless clients randomly located nearby a rectangular region. Four monitoring devices are located at the four corners. Each monitoring device has information of its own location and orientation. The radiation pattern of antenna attached to each monitoring device is known as well. All of the monitoring devices are connected to the Internet. The wireless clients transmit data packets periodically and consequently detected by the monitoring devices. Since each wireless client has different MAC address, the packets transmitted from different clients can be distinguished. Here we use \mathcal{A}_t to represent the target area and \mathcal{A}_o to represent its complement. In this study, we assume the wireless clients are quasi-static and each client is equipped with isotropic antenna. For convenience, we use a single vector to denote the locations of the wireless clients.

$$\underline{\mathbf{U}} = (\mathbf{U}_1, \dots, \mathbf{U}_{n_a}). \quad (3.1)$$

where n_a is the number of the detected clients. We also assume that the signal captured by a monitoring device comes from a line-of-sight path. Therefore the signal strength obeys free-space transmission model. The received signal strength from client j to sensing device i can be expressed as

$$P_{ij}[\text{dBm}] = A + B \log_{10}(d_{ij}) + L_{ij} + G_i(\phi_{ij}) \quad (3.2)$$

where A and B are the mean decay parameters, d_{ij} is the Euclidean distance between the client j and sensing device i . L_{ij} is shadow fading parameter and $G_i(\cdot)$ antenna gain function of sensing device. ϕ_{ij} is the angle of the signal transmission direction. The shadow fading components $\{L_{ij}\}$ are assumed to be independent and identically log-normal distributed random variables. In the logarithmic domain, the probability density function is

$$f_{L_{ij}}(\ell) = \frac{1}{\sqrt{2\pi}\sigma_s} \exp\left(-\frac{\ell^2}{2\sigma_s^2}\right) \quad (3.3)$$

where σ_s is the standard deviation of shadowing. The observed information from the four sensing devices form a power matrix $\underline{\mathbf{P}} = (\mathbf{P}_1, \dots, \mathbf{P}_{n_a})$. The vector element $\mathbf{P}_j = (P_{1j}, P_{2j}, P_{3j}, P_{4j})$ contains signal strength of wireless client j detected by four sensing devices. We assume the number and locations of wireless clients located inside the area of interest form a Poisson point process with intensity λ_t . Therefore

$$\Pr(R_t = r_t) = \frac{(\lambda_t A_t)^{r_t}}{r_t!} e^{-A_t \lambda_t} \quad r_t = 0, 1, \dots$$

Where R_t is the number of clients inside. A_t is the are of the target region. Similarly, we get

$$\Pr(R_o = r_o) = \frac{(\lambda_o A_o)^{r_o}}{r_o!} e^{-A_o \lambda_o} \quad r_o = 0, 1, \dots$$

where R_o is the number of clients outside. A_o is the area of the complimentary of target region. λ_o is Poisson intensity parameter. The inference task is estimation the occupancy based on the Power matrix $\underline{\mathbf{P}}$.

4 ESTIMATION SCHEMES

4.1 Bayes Estimation

We assume the Poisson intensity parameters λ_t and λ_o are known. Our objective is to estimate the number of clients inside the target area based on observed data $\underline{\mathbf{P}}$. First we need to get the posterior distribution of R_t given $\underline{\mathbf{P}}$.

$$\begin{aligned} \Pr(R_t = r_t | \underline{\mathbf{P}} = \underline{\mathbf{p}}) &= \int_{\{\underline{\mathbf{u}}: R_t(\underline{\mathbf{u}}) = r_t, R_o(\underline{\mathbf{u}}) = r_o\}} f_{\underline{\mathbf{U}}|\underline{\mathbf{P}}}(\underline{\mathbf{u}}|\underline{\mathbf{p}}) d\underline{\mathbf{u}} \\ &= \int_{\{\underline{\mathbf{u}}: R_t(\underline{\mathbf{u}}) = r_t, R_o(\underline{\mathbf{u}}) = r_o\}} \frac{f_{\underline{\mathbf{P}}|\underline{\mathbf{U}}}(\underline{\mathbf{p}}|\underline{\mathbf{u}}) f_{\underline{\mathbf{U}}}(\underline{\mathbf{u}})}{f_{\underline{\mathbf{P}}}(\underline{\mathbf{p}})} d\underline{\mathbf{u}} \end{aligned} \quad (4.1)$$

Because the Poisson process are independent, the distribution of $\underline{\mathbf{U}}$ can be written as

$$\begin{aligned} f_{\underline{\mathbf{U}}}(\underline{\mathbf{u}}) &= \frac{1}{A_t^{R_t(\underline{\mathbf{u}})}} \frac{(\lambda_t A_t)^{R_t(\underline{\mathbf{u}})}}{(R_t(\underline{\mathbf{u}}))!} e^{-A_t \lambda_t} \\ &\quad \times \frac{1}{A_o^{R_o(\underline{\mathbf{u}})}} \frac{(\lambda_o A_o)^{R_o(\underline{\mathbf{u}})}}{(R_o(\underline{\mathbf{u}}))!} e^{-A_o \lambda_o} \\ &= \frac{\lambda_t^{R_t(\underline{\mathbf{u}})}}{(R_t(\underline{\mathbf{u}}))!} \frac{\lambda_o^{R_o(\underline{\mathbf{u}})}}{(R_o(\underline{\mathbf{u}}))!} e^{-A_t \lambda_t - A_o \lambda_o}. \end{aligned} \quad (4.2)$$

The distribution of the received power vector $\underline{\mathbf{P}}_j$ given a specific location $\underline{\mathbf{u}}_j$ is equal to

$$\begin{aligned} f_{\underline{\mathbf{P}}_j|\underline{\mathbf{U}}_j}(\underline{\mathbf{p}}_j|\underline{\mathbf{u}}_j) &= \prod_{i=1}^{n_s} f_{L_{ij}}(p_{ij} - A - B \log_{10}(d_{ij}) - G_i(\phi_{ij})) \\ &= \frac{1}{(2\pi\sigma_s^2)^{\frac{n_s}{2}}} \prod_{i=1}^{n_s} e^{-\frac{(p_{ij} - A - B \log_{10}(d_{ij}) - G_i(\phi_{ij}))^2}{2\sigma_s^2}} \\ &= (2\pi\sigma_s^2)^{-\frac{n_s}{2}} e^{-\frac{\sum_{i=1}^{n_s} (p_{ij} - A - B \log_{10}(d_{ij}) - G_i(\phi_{ij}))^2}{2\sigma_s^2}}. \end{aligned} \quad (4.3)$$

The conditional distribution of the $\underline{\mathbf{P}}$ given $\underline{\mathbf{U}} = \underline{\mathbf{u}}$, is

$$f_{\underline{\mathbf{P}}|\underline{\mathbf{U}}}(\underline{\mathbf{p}}|\underline{\mathbf{u}}) = \prod_{j=1}^{n_a} f_{\mathbf{P}_j|\mathbf{U}_j}(\mathbf{p}_j|\mathbf{u}_j). \quad (4.4)$$

With the conditional distribution of the $\underline{\mathbf{P}}$ given $\underline{\mathbf{U}} = \underline{\mathbf{u}}$ and distribution of $\underline{\mathbf{U}}$, we can compute the marginal distribution of $\underline{\mathbf{P}}$.

$$\begin{aligned} f_{\underline{\mathbf{P}}}(\underline{\mathbf{p}}) &= \int_{\{\underline{\mathbf{u}}: R_t(\underline{\mathbf{u}}) + R_o(\underline{\mathbf{u}}) = n_a\}} f_{\underline{\mathbf{P}}|\underline{\mathbf{U}}}(\underline{\mathbf{p}}|\underline{\mathbf{u}}) f_{\underline{\mathbf{U}}}(\underline{\mathbf{u}}) d\underline{\mathbf{u}} \\ &= \sum_{(r_t, r_o): r_t + r_o = n_a} \sum_{\{\mathbb{I} \subset [n_a]: |\mathbb{I}| = r_t\}} \frac{\lambda_t^{r_t} \lambda_o^{r_o}}{r_t! r_o!} e^{-A_t \lambda_t - A_o \lambda_o} \\ &\quad \times \prod_{j \in \mathbb{I}} \mathcal{J}_{\mathcal{A}_t}(j) \prod_{j \in \mathbb{I}^c} \mathcal{J}_{\mathcal{A}_o}(j). \end{aligned} \quad (4.5)$$

where the integral components are equal to

$$\mathcal{J}_{\mathcal{A}_t}(j) = \int_{\mathcal{A}_t} f_{\mathbf{P}_j|\mathbf{U}_j}(\mathbf{p}_j|\mathbf{u}_j) d\mathbf{u}_j \quad (4.6)$$

$$\mathcal{J}_{\mathcal{A}_o}(j) = \int_{\mathcal{A}_o} f_{\mathbf{P}_j|\mathbf{U}_j}(\mathbf{p}_j|\mathbf{u}_j) d\mathbf{u}_j. \quad (4.7)$$

Now we can the posterior distribution of R_t given $\underline{\mathbf{P}}$ by substituting (4.2),(4.4),(4.5) into (4.1).

Therefore the Bayes estimator is given by

$$\begin{aligned} \hat{R}_t(\underline{\mathbf{p}}) &= \mathbb{E}[R_t | \underline{\mathbf{P}} = \underline{\mathbf{p}}] \\ &= \sum_{r_t=0}^{n_a} r_t \Pr(R_t = r_t | \underline{\mathbf{P}} = \underline{\mathbf{p}}). \end{aligned} \quad (4.8)$$

4.2 Maximum likelihood estimation

In this section, we assume the Poisson intensity parameters λ_t and λ_o are unknown.

Under this assumption, the distribution of $\underline{\mathbf{U}}$ can be written as

$$f_{\underline{\mathbf{U}}}(\underline{\mathbf{u}}; \lambda_t, \lambda_o) = \frac{\lambda_t^{R_t(\underline{\mathbf{u}})}}{(R_t(\underline{\mathbf{u}}))!} \frac{\lambda_o^{R_o(\underline{\mathbf{u}})}}{(R_o(\underline{\mathbf{u}}))!} e^{-A_t \lambda_t - A_o \lambda_o}. \quad (4.9)$$

The likelihood function is a function with two parameters, λ_t and λ_o

$$\begin{aligned}\mathcal{L}(\lambda_t, \lambda_o; \underline{\mathbf{p}}, \underline{\mathbf{u}}) &= f_{\underline{\mathbf{p}}|\underline{\mathbf{u}}}(\underline{\mathbf{p}}|\underline{\mathbf{u}}; \lambda_t, \lambda_o) \\ &= f_{\underline{\mathbf{p}}|\underline{\mathbf{u}}}(\underline{\mathbf{p}}|\underline{\mathbf{u}}) f_{\underline{\mathbf{u}}}(\underline{\mathbf{u}}; \lambda_t, \lambda_o).\end{aligned}\tag{4.10}$$

By computing the integral over $\underline{\mathbf{u}}$, we get the marginal likelihood function

$$\begin{aligned}\mathcal{L}(\lambda_t, \lambda_o; \underline{\mathbf{p}}) &= \int_{\{\underline{\mathbf{u}}: R_t(\underline{\mathbf{u}}) + R_o(\underline{\mathbf{u}}) = n_a\}} f_{\underline{\mathbf{p}}|\underline{\mathbf{u}}}(\underline{\mathbf{p}}|\underline{\mathbf{u}}) f_{\underline{\mathbf{u}}}(\underline{\mathbf{u}}; \lambda_t, \lambda_o) d\underline{\mathbf{u}} \\ &= e^{-A_t \lambda_t - A_o \lambda_o} \sum_{(r_t, r_o): r_t + r_o = n_a} \frac{\lambda_t^{r_t} \lambda_o^{r_o}}{r_t! r_o!} \\ &\quad \times \sum_{\{\mathbb{I} \subset [n_a]: |\mathbb{I}| = r_t\}} \prod_{j \in \mathbb{I}} \mathcal{J}_{\mathcal{A}_t}(j) \prod_{j \in \mathbb{I}^c} \mathcal{J}_{\mathcal{A}_o}(j).\end{aligned}\tag{4.11}$$

This is a two-dimensional optimization, but we can simplify it to a one-dimensional optimization problem by following property.

$$\max_{\lambda_t, \lambda_o} \mathcal{L}(\lambda_t, \lambda_o; \underline{\mathbf{p}}) = \max_{\alpha} \mathcal{L}\left(\frac{n_a}{A_t} \alpha, \frac{n_a}{A_o} (1 - \alpha); \underline{\mathbf{p}}\right).\tag{4.12}$$

where α within the interval $[0, 1]$. Therefore, we rewrite the likelihood function in terms of α

$$\begin{aligned}\mathcal{L}\left(\frac{n_a}{A_t} \alpha, \frac{n_a}{A_o} (1 - \alpha); \underline{\mathbf{p}}\right) &= \sum_{(r_t, r_o): r_t + r_o = n_a} \frac{e^{-n_a} n_a^{n_a}}{r_t! r_o!} \left(\frac{\alpha}{A_t}\right)^{r_t} \left(\frac{1 - \alpha}{A_o}\right)^{r_o} \\ &\quad \times \sum_{\{\mathbb{I} \subset [n_a]: |\mathbb{I}| = r_t\}} \prod_{j \in \mathbb{I}} \mathcal{J}_{\mathcal{A}_t}(j) \prod_{j \in \mathbb{I}^c} \mathcal{J}_{\mathcal{A}_o}(j)\end{aligned}\tag{4.13}$$

Now we can use numerical methods to get the values of λ_t and λ_o maximizing the likelihood function. After we obtain λ_t and λ_o . The maximum likelihood estimator is calculated

$$\begin{aligned}\hat{R}_t(\underline{\mathbf{p}}) &= E_{\hat{\lambda}_t, \hat{\lambda}_o} [R_t | \underline{\mathbf{P}} = \underline{\mathbf{p}}] \\ &= \sum_{r_t=0}^{n_a} r_t \Pr(R_t = r_t | \underline{\mathbf{P}} = \underline{\mathbf{p}}; \hat{\lambda}_t, \hat{\lambda}_o)\end{aligned}\tag{4.14}$$

5 NUMERICAL SIMULATION

6 EXPERIMENT IMPLEMENTATION

7 CONCLUSION

REFERENCES

- [1] Cisco Systems, Inc., *Visual Networking Index: Global Mobile Data Traffic Forecast Update*, 2016.
- [2] Z. Ma and A. B. Chan, “Crossing the line: Crowd counting by integer programming with local features,” in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2013.
- [3] Y. Taniguchi, M. Mizushima, G. Hasegawa, H. Nakano, and M. Matsuoka, “Counting pedestrians passing through a line in crowded scenes by extracting optical flows,” *International Information Institute (Tokyo).Information*, vol. 19, pp. 303–316, 01 2016.
Copyright - Copyright International Information Institute Jan 2016; Document feature - Illustrations; Equations; Tables; Graphs; ; Last updated - 2016-07-30.
- [4] R. Nishide and H. Takada, “Exploring efficient methods to extract pedestrian flows on a mobile adhoc network,” *The 6th International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies*, 2012.
- [5] A. M. Depatla, Saandeep and Y. Mostofi, “Occupancy estimation using only wifi power measurements,” *IEEE Journal on Selected Areas in Communications*, 05 2015.
- [6] N. Abedi, A. Bhaskar, and E. Chung, “Bluetooth and wi-fi mac address based crowd data collection and monitoring : benefits, challenges and enhancement,” in *36th Aus-*

tralasian Transport Research Forum (ATRF), (Queensland University of Technology,
Brisbane, QLD), October 2013.

APPENDIX A

MISCELLANEOUS

A.1 Figures/Tables in Appendix

A.1.1 TEST1

Test subsection for toc display purpose only.

A.1.2 TEST2

Begin

If (shift == 1'b1)

Begin

sr[(N-1):1] <= sr[(N-2):0];

sr[0] <= sr_in;

end

else

begin

sr[(N-1):1] <= sr[(N-2):0];

sr[0] <= sr[(N-1)];

end

end

A.2 Random Pictures and Test

Section here is to test toc display purpose only.

A.3 Misc Test

Section here is to test toc display purpose only.

APPENDIX B

SOURCE CODE

****Some text here****

B.1 Misc Test 2

Section here is to test toc display purpose only.

B.2 Misc Test 3

Section here is to test toc display purpose only.

B.3 Resource Usage

Design Summary (This page shows how to use landscape format in Appendix.)

Design Summary:

Number of errors: 0

Number of warnings: 0

Logic Utilization:

Number of Slice Flip Flops: 3,899 out of 33,280 11%

Number of 4 input LUTs: 3,717 out of 33,280 11%

Logic Distribution:

Number of occupied Slices: 2,198 out of 16,640 13%

Number of Slices containing only related logic: 2,198 out of 2,198 100%

Number of Slices containing unrelated logic: 0 out of 2,198 0%

*See NOTES below for an explanation of the effects of unrelated logic.

Total Number of 4 input LUTs: 3,890 out of 33,280 11%

Table B.1: Summary of Equipment Used

NAME	NO.	COMMENT
Tektronix TDS7704B Scope	1	7GHz, 20GSa/s time-equivalent sampling oscilloscope
Tektronix P7240 Probe	2	4GHz Single Ended Active Probe(High Impedance)
Agilent 81130A Function Generator	1	2 CHs Signal Generator
Xilinx Spartan-3A DSP 1800A Demo Board	1	http://goo.gl/Svvpv

[By University Requirement, no text should be allowed here in this landscape table/picture page. **DON'T USE sideways table from rotating package, it cannot align landscape title to the left binding side.**]