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

ABSTRACT

In various situations, there are demands for estimating the number of people in a specific area. This article focuses on estimating the number of devices in a certain area based on Wi-Fi metadata. To accomplish this, four sensing devices are placed at the four corners 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 clients 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 estimation. We also compare the estimator performances between directional antenna and isotropic antenna. The result shows by using the directional monitoring antenna, we can obtain better accuracy.

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 crowed 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, more than 90% of all scanned unique MAC addresses in all places were Wi-Fi addresses and the popularity of using Wi-Fi devices is therefore significantly more than Bluetooth ones.

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 and send the captured data to a process center for inference. 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 $\mathscr{A}_{\mathrm{t}}$ to represent the target area and $\mathscr{A}_{\mathrm{o}}$ 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.

$$\underline{\mathbf{U}} = (\mathbf{U}_1, \ldots, \mathbf{U}_{n_{\mathrm{a}}}). \tag{3.1}$$

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

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

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 $G_i(\cdot)$ is 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

$$f_{L_{ij}}(\ell) = \frac{1}{\sqrt{2\pi}\sigma_{\mathrm{s}}} \exp\left(-\frac{\ell^2}{2\sigma_{\mathrm{s}}^2}\right) \tag{3.3}$$

where σ_{s} is the standard deviation of shadowing. The observed information from the four sensing devices form a power matrix $\underline{\mathbf{P}} = (\mathbf{P}_1, \ldots, \mathbf{P}_{n_{\mathrm{a}}})$. 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

$$\Pr(R_{\mathrm{t}} = r_{\mathrm{t}}) = \frac{(\lambda_{\mathrm{t}} A_{\mathrm{t}})^{r_{\mathrm{t}}}}{r_{\mathrm{t}}!} e^{-A_{\mathrm{t}}\lambda_{\mathrm{t}}} \quad r_{\mathrm{t}} = 0, 1, \ldots$$

Where R_{t} is the number of clients inside. A_{t} is the are of the target region. Similarly, we get

$$\Pr(R_{\mathrm{o}} = r_{\mathrm{o}}) = \frac{(\lambda_{\mathrm{o}} A_{\mathrm{o}})^{r_{\mathrm{o}}}}{r_{\mathrm{o}}!} e^{-A_{\mathrm{o}}\lambda_{\mathrm{o}}} \quad r_{\mathrm{o}} = 0, 1, \ldots$$

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 $\underline{\mathbf{P}}$.

$$\begin{aligned}
&\Pr\left(R_\mathrm{t} = r_\mathrm{t} | \underline{\mathbf{P}} = \underline{\mathbf{p}}\right) \\
&= \int_{\{\underline{\mathbf{u}}:R_\mathrm{t}(\underline{\mathbf{u}})=r_\mathrm{t},R_\mathrm{o}(\underline{\mathbf{u}})=r_\mathrm{o}\}} f_{\underline{\mathbf{U}}|\underline{\mathbf{P}}}\left(\underline{\mathbf{u}}|\underline{\mathbf{p}}\right) d\underline{\mathbf{u}} \\
&= \int_{\{\underline{\mathbf{u}}:R_\mathrm{t}(\underline{\mathbf{u}})=r_\mathrm{t},R_\mathrm{o}(\underline{\mathbf{u}})=r_\mathrm{o}\}} \frac{f_{\underline{\mathbf{P}}|\underline{\mathbf{U}}}\left(\underline{\mathbf{p}}|\underline{\mathbf{u}}\right) f_{\underline{\mathbf{U}}}(\underline{\mathbf{u}})}{f_{\underline{\mathbf{P}}}\left(\underline{\mathbf{p}}\right)} d\underline{\mathbf{u}}
\end{aligned} \tag{4.1}$$

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

$$\begin{aligned}
f_{\underline{\mathbf{U}}}(\underline{\mathbf{u}}) &= \frac{1}{A_\mathrm{t}^{R_\mathrm{t}(\underline{\mathbf{u}})}} \frac{(\lambda_\mathrm{t} A_\mathrm{t})^{R_\mathrm{t}(\underline{\mathbf{u}})}}{(R_\mathrm{t}(\underline{\mathbf{u}}))!} e^{-A_\mathrm{t}\lambda_\mathrm{t}} \\
&\quad \times \frac{1}{A_\mathrm{o}^{R_\mathrm{o}(\underline{\mathbf{u}})}} \frac{(\lambda_\mathrm{o} A_\mathrm{o})^{R_\mathrm{o}(\underline{\mathbf{u}})}}{(R_\mathrm{o}(\underline{\mathbf{u}}))!} e^{-A_\mathrm{o}\lambda_\mathrm{o}} \\
&= \frac{\lambda_\mathrm{t}^{R_\mathrm{t}(\underline{\mathbf{u}})}}{(R_\mathrm{t}(\underline{\mathbf{u}}))!} \frac{\lambda_\mathrm{o}^{R_\mathrm{o}(\underline{\mathbf{u}})}}{(R_\mathrm{o}(\underline{\mathbf{u}}))!} e^{-A_\mathrm{t}\lambda_\mathrm{t} - A_\mathrm{o}\lambda_\mathrm{o}}.
\end{aligned} \tag{4.2}$$

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

$$\begin{aligned}
&f_{\mathbf{P}_j|\mathbf{U}_j}(\mathbf{p}_j|\mathbf{u}_j) \\
&= \prod_{i=1}^{n_\mathrm{s}} f_{L_{ij}}(p_{ij} - A - B\log_{10}(d_{ij}) - G_i(\phi_{ij})) \\
&= \frac{1}{(2\pi\sigma_\mathrm{s}^2)^{\frac{n_\mathrm{s}}{2}}} \prod_{i=1}^{n_\mathrm{s}} e^{-\frac{(p_{ij}-A-B\log_{10}(d_{ij})-G_i(\phi_{ij}))^2}{2\sigma_\mathrm{s}^2}} \\
&= \left(2\pi\sigma_\mathrm{s}^2\right)^{-\frac{n_\mathrm{s}}{2}} e^{-\frac{\sum_{i=1}^{n_\mathrm{s}}(p_{ij}-A-B\log_{10}(d_{ij})-G_i(\phi_{ij}))^2}{2\sigma_\mathrm{s}^2}}.
\end{aligned} \tag{4.3}$$

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

$$f_{\underline{\mathbf{P}}|\underline{\mathbf{U}}}\left(\underline{\mathbf{p}}|\underline{\mathbf{u}}\right) = \prod_{j=1}^{n_\mathrm{a}} f_{\mathbf{P}_j|\mathbf{U}_j}(\mathbf{p}_j|\mathbf{u}_j). \tag{4.4}$$

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 $\underline{\mathbf{P}}$.

$$\begin{aligned} f_{\underline{\mathbf{P}}}\left(\underline{\mathbf{p}}\right) &= \int_{\{\underline{\mathbf{u}}:R_\mathrm{t}(\underline{\mathbf{u}})+R_\mathrm{o}(\underline{\mathbf{u}})=n_\mathrm{a}\}} f_{\underline{\mathbf{P}}|\underline{\mathbf{U}}}\left(\underline{\mathbf{p}}|\underline{\mathbf{u}}\right) f_{\underline{\mathbf{U}}}(\underline{\mathbf{u}}) d\underline{\mathbf{u}} \\ &= \sum_{(r_\mathrm{t},r_\mathrm{o}):r_\mathrm{t}+r_\mathrm{o}=n_\mathrm{a}} \sum_{\{\mathbb{I}\subset[n_\mathrm{a}]:|\mathbb{I}|=r_\mathrm{t}\}} \frac{\lambda_\mathrm{t}^{r_\mathrm{t}} \lambda_\mathrm{o}^{r_\mathrm{o}}}{r_\mathrm{t}! r_\mathrm{o}!} e^{-A_\mathrm{t}\lambda_\mathrm{t}-A_\mathrm{o}\lambda_\mathrm{o}} \\ &\quad \times \prod_{j\in\mathbb{I}} \mathscr{I}_{\mathscr{A}_\mathrm{t}}(j) \prod_{j\in\mathbb{I}^\mathrm{c}} \mathscr{I}_{\mathscr{A}_\mathrm{o}}(j). \end{aligned} \tag{4.5}$$

where the integral components are equal to

$$\mathscr{I}_{\mathscr{A}_\mathrm{t}}(j) = \int_{\mathscr{A}_\mathrm{t}} f_{\mathbf{P}_j|\mathbf{U}_j}(\mathbf{p}_j|\mathbf{u}_j) d\mathbf{u}_j \tag{4.6}$$

$$\mathscr{I}_{\mathscr{A}_\mathrm{o}}(j) = \int_{\mathscr{A}_\mathrm{o}} f_{\mathbf{P}_j|\mathbf{U}_j}(\mathbf{p}_j|\mathbf{u}_j) d\mathbf{u}_j. \tag{4.7}$$

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

$$\begin{aligned} \hat{R}_\mathrm{t}\left(\underline{\mathbf{p}}\right) &= \mathrm{E}\left[R_\mathrm{t}|\underline{\mathbf{P}} = \underline{\mathbf{p}}\right] \\ &= \sum_{r_\mathrm{t}=0}^{n_\mathrm{a}} r_\mathrm{t} \Pr\left(R_\mathrm{t} = r_\mathrm{t}|\underline{\mathbf{P}} = \underline{\mathbf{p}}\right). \end{aligned} \tag{4.8}$$

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

$$f_{\underline{\mathbf{U}}}(\underline{\mathbf{u}}; \lambda_\mathrm{t}, \lambda_\mathrm{o}) = \frac{\lambda_\mathrm{t}^{R_\mathrm{t}(\underline{\mathbf{u}})}}{(R_\mathrm{t}(\underline{\mathbf{u}}))!} \frac{\lambda_\mathrm{o}^{R_\mathrm{o}(\underline{\mathbf{u}})}}{(R_\mathrm{o}(\underline{\mathbf{u}}))!} e^{-A_\mathrm{t}\lambda_\mathrm{t}-A_\mathrm{o}\lambda_\mathrm{o}}. \tag{4.9}$$

The likelihood function is a function with two parameters,λ_{t} and λ_{o}

$$\begin{aligned}\mathscr{L}\left(\lambda_{\mathrm{t}},\lambda_{\mathrm{o}};\underline{\mathbf{p}},\underline{\mathbf{u}}\right) &= f_{\underline{\mathbf{P}},\underline{\mathbf{U}}}\left(\underline{\mathbf{p}},\underline{\mathbf{u}};\lambda_{\mathrm{t}},\lambda_{\mathrm{o}}\right) \\ &= f_{\underline{\mathbf{P}}|\underline{\mathbf{U}}}\left(\underline{\mathbf{p}}|\underline{\mathbf{u}}\right) f_{\underline{\mathbf{U}}}\left(\underline{\mathbf{u}};\lambda_{\mathrm{t}},\lambda_{\mathrm{o}}\right).\end{aligned} \tag{4.10}$$

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

$$\begin{aligned}&\mathscr{L}\left(\lambda_{\mathrm{t}},\lambda_{\mathrm{o}};\underline{\mathbf{p}}\right) \\ &= \int_{\{\underline{\mathbf{u}}:R_{\mathrm{t}}(\underline{\mathbf{u}})+R_{\mathrm{o}}(\underline{\mathbf{u}})=n_{\mathrm{a}}\}} f_{\underline{\mathbf{P}}|\underline{\mathbf{U}}}\left(\underline{\mathbf{p}}|\underline{\mathbf{u}}\right) f_{\underline{\mathbf{U}}}\left(\underline{\mathbf{u}};\lambda_{\mathrm{t}},\lambda_{\mathrm{o}}\right) d\underline{\mathbf{u}} \\ &= e^{-A_{\mathrm{t}}\lambda_{\mathrm{t}}-A_{\mathrm{o}}\lambda_{\mathrm{o}}} \sum_{(r_{\mathrm{t}},r_{\mathrm{o}}):r_{\mathrm{t}}+r_{\mathrm{o}}=n_{\mathrm{a}}} \frac{\lambda_{\mathrm{t}}^{r_{\mathrm{t}}}\lambda_{\mathrm{o}}^{r_{\mathrm{o}}}}{r_{\mathrm{t}}!r_{\mathrm{o}}!} \\ &\qquad \times \sum_{\{\mathbb{I}\subset[n_{\mathrm{a}}]:|\mathbb{I}|=r_{\mathrm{t}}\}} \prod_{j\in\mathbb{I}} \mathscr{I}_{\mathscr{A}_{\mathrm{t}}}(j) \prod_{j\in\mathbb{I}^{\mathrm{c}}} \mathscr{I}_{\mathscr{A}_{\mathrm{o}}}(j).\end{aligned} \tag{4.11}$$

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

$$\max_{\lambda_{\mathrm{t}},\lambda_{\mathrm{o}}} \mathscr{L}\left(\lambda_{\mathrm{t}},\lambda_{\mathrm{o}};\underline{\mathbf{p}}\right) = \max_{\alpha} \mathscr{L}\left(\frac{n_{\mathrm{a}}}{A_{\mathrm{t}}}\alpha, \frac{n_{\mathrm{a}}}{A_{\mathrm{o}}}(1-\alpha);\underline{\mathbf{p}}\right). \tag{4.12}$$

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

$$\begin{aligned}&\mathscr{L}\left(\frac{n_{\mathrm{a}}}{A_{\mathrm{t}}}\alpha, \frac{n_{\mathrm{a}}}{A_{\mathrm{o}}}(1-\alpha);\underline{\mathbf{p}}\right) \\ &= \sum_{(r_{\mathrm{t}},r_{\mathrm{o}}):r_{\mathrm{t}}+r_{\mathrm{o}}=n_{\mathrm{a}}} \frac{e^{-n_{\mathrm{a}}}n_{\mathrm{a}}^{n_{\mathrm{a}}}}{r_{\mathrm{t}}!r_{\mathrm{o}}!} \left(\frac{\alpha}{A_{\mathrm{t}}}\right)^{r_{\mathrm{t}}} \left(\frac{1-\alpha}{A_{\mathrm{o}}}\right)^{r_{\mathrm{o}}} \\ &\qquad \times \sum_{\{\mathbb{I}\subset[n_{\mathrm{a}}]:|\mathbb{I}|=r_{\mathrm{t}}\}} \prod_{j\in\mathbb{I}} \mathscr{I}_{\mathscr{A}_{\mathrm{t}}}(j) \prod_{j\in\mathbb{I}^{\mathrm{c}}} \mathscr{I}_{\mathscr{A}_{\mathrm{o}}}(j)\end{aligned} \tag{4.13}$$

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

$$\begin{aligned}\hat{R}_{\mathrm{t}}\left(\underline{\mathbf{p}}\right) &= \mathrm{E}_{\hat{\lambda}_{\mathrm{t}},\hat{\lambda}_{\mathrm{o}}}\left[R_{\mathrm{t}}|\underline{\mathbf{P}}=\underline{\mathbf{p}}\right] \\ &= \sum_{r_{\mathrm{t}}=0}^{n_{\mathrm{a}}} r_{\mathrm{t}} \Pr\left(R_{\mathrm{t}}=r_{\mathrm{t}}|\underline{\mathbf{P}}=\underline{\mathbf{p}};\hat{\lambda}_{\mathrm{t}},\hat{\lambda}_{\mathrm{o}}\right)\end{aligned} \tag{4.14}$$

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.