Lab 2: WiFi Signal

EE447 - Mobile Network, 2020 Spring.

* 姓名: 陆昱昊　学号: 517021910533　Email: dtassassin@sjtu.edu.cn

The whole project has been pushed to my github, along with a demo.

Q & A

1. Q: Why is necessary to record all the measured value rather than only the average value? Please give your own explanation.

   A: Recording all measured values allows the feasibility of further inspection. For example, we can use the data to debug the program, or manually label valid/invalid data to select AP with trusted data as source. Moreover, we can study the distribution of the data, like standard deviation, or even use other distributions to model the signal strength and perform MLE to estimate real signal strength.

2. Q: Besides the WiFi signal strength, what other information of the Routers can be got in the test?

   A: We can get a `List` of `ScanResult` by invoking `WiFiManager.getScanResults()`. In the given test, it uses (1) `level`: The detected signal level in dBm, also known as the RSSI. (2) `SSID`: The network name.

   According to Android doc, we can also get:

Field Name	Explanation
`BSSID`	The address of the access point
`capabilities`	Describes the authentication, key management and encryption schemes supported by the access point.
`centerFreq0`	Not used if the AP bandwidth is 20 MHz. If the AP use 40, 80 or 160 MHz, this is the center frequency (in MHz). If the AP use 80+80 MHz, this is the center frequency of the first segment (in MHz).
`centerFreq1`	Only used if the AP bandwidth is 80+80 MHz. If the AP use 80+80 MHz, this is the center frequency of the second segment (in MHz).
`channelWidth`	AP channel width.
`frequency`	The primary 20 MHz frequency of the channel over which the client is communicating with the access point.
`operatorFriendlyName`	Indicates Passpoint operator name published by access point.
`venueName`	Indicates venue name published by access point.

3. Q: Why does the scanning need to be operated in the thread `scanThread`?

   A: Because the scan process is a time consuming task, each scan takes at least `ScanningTime` $\times$ `TestTime`. If not operated in thread, it will block the UI (main) thread, and no events can be dispatched. From the user's perspective, the application appears to hang. More details in doc. BTW, the given implementation still blocks the UI thread due to the `while` loop in the `onClickListener` of `changactivity` (SCAN button). To address this, I think all the manipulations in this listener should be operated in a thread.

Indoor Positioning System Design

Our design is based on RSS-based triangulation. In brief, use the RSS to estimate the distance between the device and the APs. For the relation between distance and strength loss, we adopt the path loss model only with a constant environment medium loss as follows (taken from textbook):

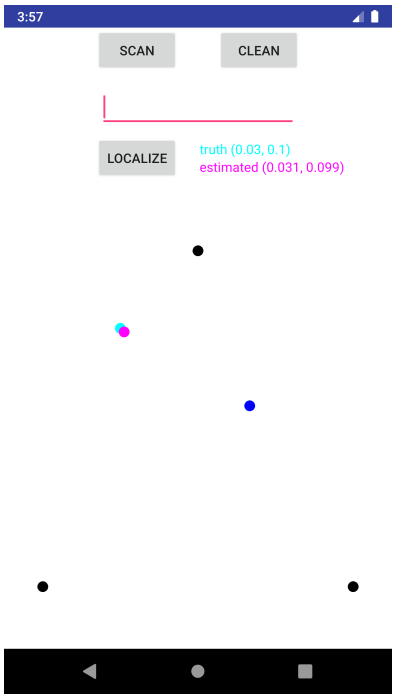
$$L = 32.45 + 20 \lg f_c + 20 \lg d + \beta$$

where L is the loss of strength measured in dBm, f_c is the frequency measured in MHz, d is the distance measured in km and β is the environment medium loss. In different scenarios, it is necessary to introduce a medium loss rather than use the free space path loss directly. In most cases, we can assume the environment is uniform (the indoor environment does not change dramatically in a certain range), so that β is a constant w.r.t. a certain environment.
Some basic assumptions are:

1. Three known APs A_i for positioning and an additional one for estimating β with their locations (x_i, y_i), transmitting power a_i and frequency f_i, $i = 1, 2, 3, 4$.
2. There exists observation noise $\varepsilon \sim N(0, \sigma)$. Therefore, the measured RSS $\hat{a}_i$, $i = 1, 2, 3, 4$ and device position (x, y) subjects to

$$a_i - \hat{a}_i = 32.45 + 20 \lg f_i + 20 \lg \sqrt{(x - x_i)^2 + (y - y_i)^2} + \beta + \varepsilon$$

3. For simplicity, without loss of generality, let A_1 be the origin and $\overrightarrow{A_1 A_2}$ be the direction of x axis. Thus, we re-denote the coordinates of A_1, A_2, A_3, A_4 to be $(0, 0), (x_2, 0), (x_3, y_3), (x_4, y_4)$



The positioning process is

1. Measure and calculate the average RSS $\bar{a}_i, i = 1, 2, 3, 4$ to determine the environment medium loss $\hat{\beta}$. Detailed solving process in Appendix A.
2. Using the estimated $\hat{\beta}$ and the j-th measurement of the RSS from A_i, $\hat{a}_{ij}$ to perform a maximum likelihood estimation to estimate the position $(\hat{x}, \hat{y})$. Detailed solving process in Appendix B.

Left is a demo of the positioning result. Black is the three main APs, while the blue dot is the forth auxiliary AP. Cyan dot is the real position and the magenta one is the positioning result. We can see that the error is at the scale of meter.
The main entrance to the positioning logic is the method `Positioning.position()`.
The basic settings, including environment medium loss β, APs' positions $x_2, (x_3, y_3), (x_4, y_4)$ and device position (x^*, y^*) in Line $35 - 56$ respectively in `Positioning.java`.

We also implement real localization. The only difference is to substitute simulated scan results with real scan results. To perform real localization, make sure (1) `FileNameGroup` in `SuperWifi.java` contains at least four APs, because we use the first 4 APs to perform estimating beta and positioning. (2) Set `simulated` to `false` in Line 25, `MainActivity.java`. (3) First click SCAN then click LOCALIZE. (4) Set the positions (in km) of APs in Line $37 - 56$ in `Positioning.java`. Also do not change x_1, y_1, y_2. But now, the truth point is not truth anymore unless you set it correctly.

Through analysis, we find out that the unit of d, i.e. km makes the real scene test very tricky. A very small terbulance in RSS can result in a drastic change in distance because of $20 \lg d$, e.g. if the RSS change by 1 (because of noise), then d changes by 1.1 km. We can change the unit of d to meter and add a constant to compansate for logarithm. However, the effect of logarithm on unit is a problem to us. In this case, if the RSS change by 1, then d only changes by 1.1 m. For a real scenario, when the distance between APs are less than 50 meters, we think the unit should be meter rather than kilometer. However, this might require a new formula to characterize RSS.

Appendices

A Estimating Environment Medium Loss

According to our assumptions, three fixed known APs A_i, with coordinates $(0,0),(1,0),(x_3,y_3)$, transmitting power a_i, frequency f_i and average measured RSS $\bar{a}_i, i = 1,2,3$ We have

$$\begin{aligned} a_1 - \bar{a}_1 &= 32.45 + 20 \lg f_1 + 10 \lg \left(x^2 + y^2\right) + \beta & (1)\\ a_2 - \bar{a}_2 &= 32.45 + 20 \lg f_2 + 10 \lg \left((x - x_2)^2 + y^2\right) + \beta & (2)\\ a_3 - \bar{a}_3 &= 32.45 + 20 \lg f_3 + 10 \lg \left((x - x_3)^2 + (y - y_3)^2\right) + \beta & (3) \end{aligned}$$

Eqn.(2) − Eqn.(1), Eqn.(3) − Eqn.(1)

$$\begin{aligned} a_2 - a_1 - \bar{a}_2 + \bar{a}_1 &= 20 \lg \frac{f_2}{f_1} + 10 \lg \frac{(x - x_2)^2 + y^2}{x^2 + y^2} & (4)\\ a_3 - a_1 - \bar{a}_3 + \bar{a}_1 &= 20 \lg \frac{f_3}{f_1} + 10 \lg \frac{(x - x_3)^2 + (y - y_3)^2}{x^2 + y^2} & (5) \end{aligned}$$

Let

$$\begin{aligned} \lg k_1 &= \frac{a_2 - a_1 - \bar{a}_2 + \bar{a}_1 - 20 \lg f_2 + 20 \lg f_1}{10}\\ \lg k_2 &= \frac{a_3 - a_1 - \bar{a}_3 + \bar{a}_1 - 20 \lg f_3 + 20 \lg f_1}{10} \end{aligned}$$

Eqn.(4)(5) can be simplified to

$$\begin{aligned} (k_1 - 1)x^2 + 2x_2 x + (k_1 - 1)y^2 - x_2^2 = 0 & (6)\\ (k_2 - 1)x^2 + 2x_3 x + (k_2 - 1)y^2 + 2y_3 y - x_3^2 - y_3^2 = 0 & (7) \end{aligned}$$

$(k_1 - 1)\times$Eqn.(7)$-(k_2 - 1)\times$Eqn.(6)

$$2\left((k_1 - 1)x_3 - (k_2 - 1)x_2\right)x + 2(k_1 - 1)y_3 y - (k_1 - 1)(x_3^2 + y_3^2) + (k_2 - 1)x_2^2 = 0 \quad (8)$$

According to Eqn.(8), let

$$\begin{aligned} m &= -\frac{k_1 x_3 - x_3 - k_2 x_2 + x_2}{(k_1 - 1)y_3}\\ n &= \frac{(k_1 - 1)(x_3^2 + y_3^2) - (k_2 - 1)x_2^2}{2(k_1 - 1)y_3} \end{aligned}$$

Plug in $y = mx + n$ into Eqn.(6), we can get

$$(k_1 - 1)(1 + m^2)x^2 + \left(2mn(k_1 - 1) + 2x_2\right)x + (k_1 - 1)n^2 - x_2^2 = 0 \quad (9)$$

This is a typical quadratic equation. We can obtain two solutions.

$$
\begin{gathered}
a = (k_1 - 1)(1 + m^2) \\
b = 2(mn(k_1 - 1) + x_2) \\
c = (k_1 - 1)n^2 - x_2^2 \\
x_{(1)} = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \qquad x_{(2)} = \frac{-b - \sqrt{b^2 - 4ac}}{2a}
\end{gathered} \tag{10}
$$

By $y = mx + n$, we get two possible positions $(x_{(1)}, y_{(1)}), (x_{(2)}, y_{(2)})$, but our goal is to determine β. Plug the two positions into Eqn.(1), we can get two solutions for β.

$$
\begin{aligned}
\beta_{(1)} &= a_1 - \bar{a}_1 - 32.45 - 20 \lg f_1 - 10 \lg(x_{(1)}^2 + y_{(1)}^2) \\
\beta_{(2)} &= a_1 - \bar{a}_1 - 32.45 - 20 \lg f_1 - 10 \lg(x_{(2)}^2 + y_{(2)}^2)
\end{aligned}
$$

By empirical evaluation, $\beta_{(1)} \neq \beta_{(2)}$ and both of them satisfies Eqn.(1)(2)(3). Therefore, we need to introduce a forth AP (x_4, y_4) to finally determine β. By the RSS equation, we have

$$
\begin{aligned}
\beta'_{(1)} &= a_4 - \bar{a}_4 - 32.45 - 20 \lg f_4 - 10 \lg \left((x_{(1)} - x_4)^2 + (y_{(1)} - y_4)^2\right) \\
\beta'_{(2)} &= a_4 - \bar{a}_4 - 32.45 - 20 \lg f_4 - 10 \lg \left((x_{(2)} - x_4)^2 + (y_{(2)} - y_4)^2\right)
\end{aligned}
$$

Then our estimated $\hat{\beta}$ is derived from

$$
\hat{\beta} = \arg\min_{\beta} |\beta - \beta'|, \quad \beta \in \{\beta_{(1)}, \beta_{(2)}\}, \beta' \in \{\beta'_{(1)}, \beta'_{(2)}\}
$$

Fig.1 is a simulation.

B Positioning

Denote $\mu_{ij}(x, y) = a_i - 32.45 - 20 \lg f_i - 20 \lg \sqrt{(x - x_i)^2 + (y - y_i)^2} - \hat{\beta}$, we have

$$
p(\hat{a}_{ij} | x, y) = N\left(\mu_{ij}(x, y) - \varepsilon \mid \mu_{ij}(x, y), \sigma\right) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\left(\hat{a}_{ij} - \mu_{i,j}(x,y)\right)^2}{2\sigma^2}}
$$

Given n scan is performed w.r.t. each AP, the likelihood function is

$$
\mathcal{L} = \prod_{i=1}^{3} \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\left(\hat{a}_{ij} - \mu_{ij}(x,y)\right)^2}{2\sigma^2}}
$$

To solve this, we take logarithm to get the log-likelihood

$$
\ell = -\frac{1}{2\sigma^2} \sum_{i=1}^{3} \sum_{j=1}^{n} \left(\hat{a}_{ij} - \mu_{ij}(x, y)\right)^2 + C
$$

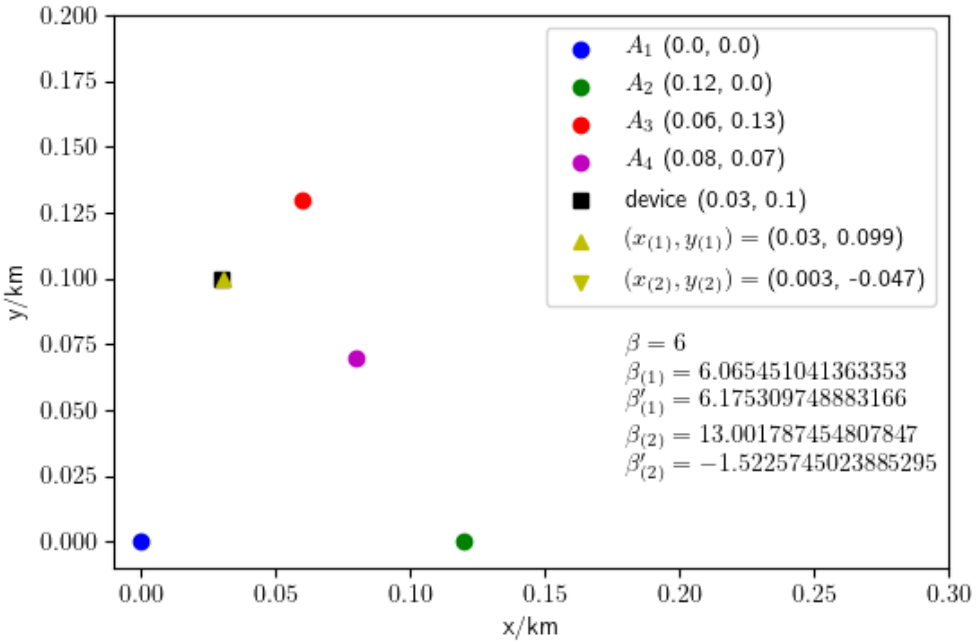


Fig. 1: The ground truth $\beta = 6$, the estimated position $(x_{(1)}, y_{(1)}), (x_{(2)}, y_{(2)})$ is rounded to the scale of meter. We can see that the positioning (although our aim is not to position at this time) error between $(x_{(1)}, y_{(1)})$ and device is at the scale of meter, given the distance between APs at the scale of $10 \sim 10^2$ meters. By the verification of A_4, we can see that $|\beta_{(1)} - \beta'_{(1)}|$ is much smaller than $|\beta_{(2)} - \beta'_{(2)}|$, actually itself is very small. Therefore, our estimated environment medium loss $\hat{\beta} = \beta_{(1)}$.

Although $\mu_{ij}(x,y)$ is a convex function, $(\hat{a}_{ij}-\mu_{ij}(x,y))^2$ is not. So, we use gradient descent to get the optimal (x,y). Let $10c_{ij}=(\hat{a}_{ij}-a_i+32.45+20\lg f_i+\hat{\beta})\ln 10$

$$\max_{x,y}\ell\equiv\min_{x,y}\sum_{i=1}^{3}\sum_{j=1}^{n}\left(\ln\left((x-x_i)^2+(y-y_i)^2\right)+c_{ij}\right)^2=\min_{x,y}\hat{\ell}$$

$$\frac{\partial\hat{\ell}}{\partial x}=\sum_{i=1}^{3}\sum_{j=1}^{n}\frac{4(x-x_i)\left(\ln\left((x-x_i)^2+(y-y_i)^2\right)+c_{ij}\right)}{(x-x_i)^2+(y-y_i)^2}$$

$$\frac{\partial\hat{\ell}}{\partial y}=\sum_{i=1}^{3}\sum_{j=1}^{n}\frac{4(y-y_i)\left(\ln\left((x-x_i)^2+(y-y_i)^2\right)+c_{ij}\right)}{(x-x_i)^2+(y-y_i)^2}$$

$$\nabla_{x,y}\hat{\ell}=\left(\frac{\partial\hat{\ell}}{\partial x},\frac{\partial\hat{\ell}}{\partial y}\right)$$

First, initialize the position$(\hat{x},\hat{y})$ arbitrarily. While the loss $\hat{\ell}$ does not converge, compute the gradient $\nabla_{x,y}\hat{\ell}$ and update the position by $(\hat{x},\hat{y})\leftarrow(\hat{x},\hat{y})-\alpha\nabla_{x,y}\hat{\ell}$. Fig.2 is a simulation.

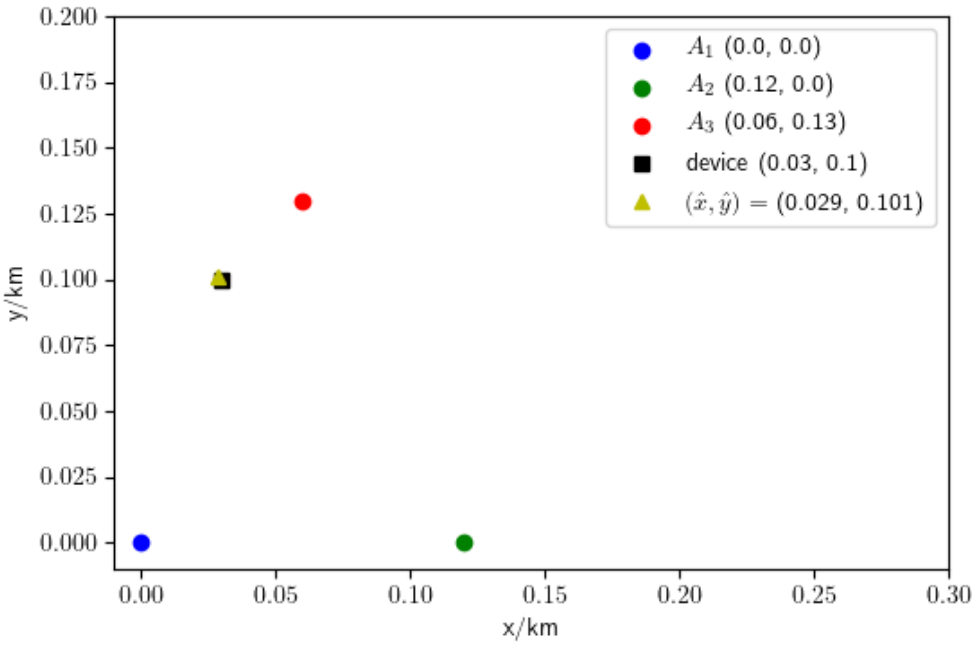


Fig. 2: A simulation where $\alpha=0.0001$ and the loss converges within 400 iterations. Empirically, the optimization terminates in 10^2 iterations which takes $0.2\sim0.4$ second in Python.