Smart Phone Sensing

Report – 2; March 16, 2018

Bhavya Jain (4745507), Konrad Kleeberger (4748476)

Group – CCS, Phone – OnePlus 5, Android Version – 8.0.0, API – 26

MATLAB, org.apache.commons.math3.analysis.function.Gaussian, https://github.com/BSJAIN92/SPS

1. Data Collection

- In each iteration, we collected 10 samples with refresh time of 5s (50s/iteration)
- The recording was always done in one direction
- Recording was done on a quiet Friday afternoon (March 9) and busy Wednesday afternoon (March 14)

2. Data Processing

- The collected data was filtered for stable SSIDs, mainly those which are not expected to change during the course of the project.
    - Observed SSIDs = 15+, Selected SSIDs = 4
    - SSIDs Selected = "eduroam", "tudelft-dastud", "TUvisitor", "TRIBLER"
- In collected data, it was observed that many BSSIDs were observed very few times in a particular cell. To clean the collected data, BSSIDs were filtered for BSSIDs observed in at least 80% samples. 7% to 13% of observed BSSIDs were selected
- The data has been further processed via Matlab to make the gaussian fit.
    - Used "fitdist" function with the RSS values of a BSSID per cell as input and the parameter "normal" for normal distribution
    - For calculating the y value of a normal distribution we use the gaussian function of the library math3 of apache commons.

3. Radio Map

Figure 1 shows one BSSID which has been scanned in four cells. The graph for C14 and C15 are overlapping as well as for C12 and C13. Both of these pairs of cells are neighbors and hence, it's logical for them to have similar map. C12 and C15 are on the opposite site of the corridor and are quite different, so they hardly have any overlap in the normal distribution.
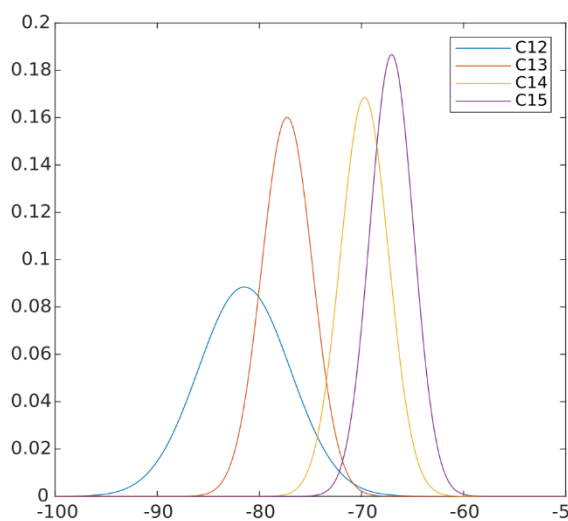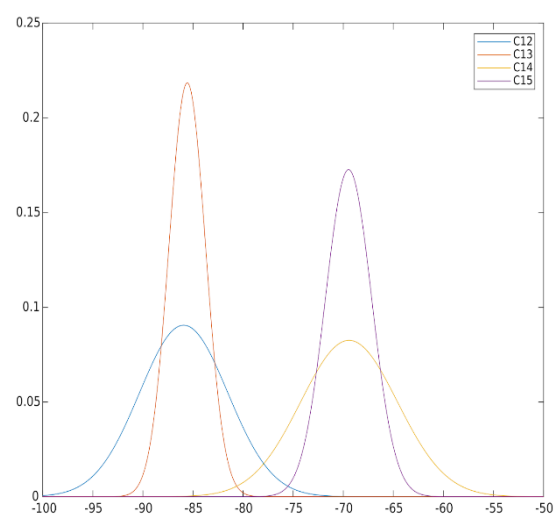

Figure 2

Figure 1

Figure 2 shows again that C12 and C15 are quite different and nearly have no overlapping of values. Interestingly, the adjacent cell pairs (12, 13 and 14, 15) have similar mu values, with much different sigma values. In this case, even with similar mu values, the cell with narrow distribution will always have higher probability, and will have more chance to win. This shows the variety of data that was observed and used for training.

4. Evaluation

Evaluation was done on Thursday morning (March 15). We used parallel and serial methods to determine the probability and filtered data for standard deviation values for less than 5, 4 and 3. Most accurate results were obtained using parallel method for standard deviation values less than 5. The following tables represent test results for parallel method to determine the probability with BSSIDs filtered for sigma values less than 3 and less than 5 respectively.

Sigma < 3	Cell 15	Cell 14	Cell 13	Cell 12
Cell 15	10			
Cell 14	5	4	1	
Cell 13			10	
Cell 12			10	0

Sigma < 5	Cell 15	Cell 14	Cell 13	Cell 12
Cell 15	9	1		
Cell 14	1	9		
Cell 13			10	
Cell 12	1		2	7

Actual	
Location	
Observed	
Location	

We decided to move ahead with parallel method to determine the probability and filtered the data to include BSSIDs with sigma less than 5.

Sigma < 5	Cell 15	Cell 14	Cell 13	Cell 12	Total
Success	90%	90%	100%	70%	87.5%

5. Discussion

The main difficulty was encountered while determining suitable method for localization. Serial and parallel methods both had their merits. Hence, we implemented both methods and tested the application based on both. This limited the amount of time we had to record data for cells. In order to account for complexity of localization for adjacent cells, we decided to test the application for cells 12, 13, 14 and 15. Since cells 13 and 14 are adjacent and with some Access Points in the hallway to East side rooms, testing for these cells provides a good validation of the method used. Another difficulty was encountered in data collection. This was due to high difference in heights of team members. Our initial recorded data was much more scattered due to data recording at different heights. We decided a point for recording data for each person to arrive at same height. Also, during application testing, different method of holding the phone yielded different results.

Wi-fi is scanned after a certain period of time. In order to obtain correct test results, "Locate" button is disabled after every localization and is enabled after 5s. Also, we observed impact of including BSSIDs with various sigma values to understand their impact. We tested the localization accuracy using BSSIDs with different sigma values (<3, <4, <5). Collected data was filtered and cleaned not only using SSIDs, but also for BSSIDs that were observed in at least 80% of the samples collected.

Following are screenshots for Data Collection, Localization (Please note the "LOCATE" button is disabled) and Localization with "LOCATE" button activated after 5s to allow for new scan of Wi-Fi.