

**COMP4336/9336 Mobile Data Networking
2022 Term 2**

**Individual Term Project: Due 5pm Friday 29 July 2022 (Week 9)
Assessment Weighting: 25%
Project Specifications and Marking Criteria (5 pages): Released 16 June 2022**

This is the complete specification of the term project. You are encouraged to discuss the project or any questions in the **Project Forum in Moodle**.

Topic

Location Identification with WiFi Fingerprinting

Aim

Distance-dependent path loss of wireless signals, together with dense deployment of WiFi in public spaces enable WiFi to be used as a tool for localising people in indoor environments. In this project the students will develop and implement algorithms that will enable location identification using signals available from in-situ public WiFi infrastructure.

Background on WiFi Fingerprinting

As GPS does not work densely deployed in public spaces indoors. There are many different techniques to use WiFi for localisation solution when indoors. However, the one that is widely pursued is based on a technique called WiFi Fingerprinting.

As we know, wireless signals are affected by distance (a multipath structure) reflecting objects in the environment. We have learned that the multipath structure is very sensitive to the location of the receiver, a small move can cause a small-scale fading, which will ultimately affect the received signal strength (RSS) due to constructive or destructive interference with the original signal. We have also learned and observed that Tx-Rx distance directly affects the RSS. Thus, if the receiver (WiFi Client) changes its location, the Tx-Rx distance may also change, causing changes in RSS. Consequently, WiFi RSS can help fingerprinting a location, i.e., we can potentially expect that a given location can be uniquely identified by a unique RSS value.

However, we have also learned that RSS is unstable and fluctuates a lot even for the same distance and location due to many random interferences in the environment. Thus, in practice, it's challenging to identify locations uniquely using RSS, especially if using a single WiFi AP as a reference for RSS measurements. The goal of this project is to explore the potential of using multiple WiFi APs for more reliable identification of locations, i.e., WiFi fingerprinting will be based on RSS data from multiple APs, instead of one. This is possible given the dense deployment of WiFi in urban environment. For example, it is common to receive beacons from tens of WiFi APs in a shopping mall or a university campus.

Hardware Requirements

This project can be completed using a laptop. WiFi RSS data can be collected using Wireshark as learned in the lab experiments.

Programming Requirements

Some basic programming is involved to process WiFi RSS data and implement algorithms that can identify locations from the WiFi RSS data. There are no restrictions on the choice of programming language (Python, MATLAB etc. are all OK).

Tasks Involved

You need to complete the following tasks:

- 1. WiFi RSS data collection:** Each student will choose a suitable indoor environment to conduct this task. It can be a university (e.g., UNSW) campus, a shopping centre, etc., which have plenty of public WiFi APs deployed. You need to use Wireshark to collect WiFi RSS data from the surrounding APs with your own device (e.g., laptop) in different locations in the indoor environment (*you may want to use some sort of system API to fetch WiFi RSS directly, which will be useful to build a demo of your localisation program in the following task*). Thus, the dataset is expected to be unique for every student. You should collect data from at least 10 different locations separated by at least 2-3 meters from each other for reliable location identifications. Collect plenty of RSS data for each location so that you have enough data for analysis despite the expected fluctuations in signal strength.
- 2. Design and implement a suitable algorithm to fingerprint a location:** Design and implement a suitable algorithm to fingerprint a location using the data you collected in Task 1. Use your algorithm/code to demonstrate that you can identify different locations with good accuracy. You are welcome to use any programming language if you have relevant background, but use of machine learning is preferred. Basic WiFi fingerprinting techniques, such as building a fingerprint database and then matching a new fingerprint to the database (as described in [1]) is also acceptable. [Note that this task requires some independent literature search/reading regarding WiFi Fingerprinting-based localisation. An initial reading list is included at the end of this document and a short lecture on WiFi Fingerprinting using a basic database matching approach will be covered in Lecture 3.]
- 3. Write a short report** (approx. 2000-3000 words plus figures/tables) explaining your experiments, algorithm design, and performance results. Do not forget to include a title, abstract, introduction, and conclusion.
- 4. Produce a short (less than 10-min and 100MB) demo video** demonstrating interesting parts of your project.

[1] Detecting Identity-Based Attacks in Wireless Networks Using Signalprints, by Faria and Cheriton, WiSe 2006. [<https://dl.acm.org/doi/10.1145/1161289.1161298>] [Note: Although the title says it is trying to identify attacks, actually the fingerprinting method described in this paper is quite applicable to location identification as well.]

You are welcome and encouraged to do your own research to explore suitable algorithms for locations detection using WiFi RSS. Note that **there is no restriction on the algorithms to be used.**

What to submit?

Submit the following 4 items using the **project submission link in Moodle**:

1. **Dataset:** The data set may have many data files, so you should give them appropriate labels. If machine learning is used, provide the training and testing datasets as well. Submit one ZIP file containing all your datasets.
2. **Code:** Submit the source code you have written for implementing the WiFi fingerprinting and location identification. You should also provide a clear description of how to compile/execute your code (specify the programming language, version, library, etc.). The code should also work with the dataset you submitted, so we can test it against the dataset. Submit one ZIP file for the code and any descriptions.
3. **Report:** Should be easy to read and comprehend. Maximum 20 MB. Submit in PDF.
4. **Demo Video:** Be creative. You have the complete freedom and flexibility to choose contents and style. The demo video should show how your program works in a real-world environment. Demonstrate the performance of your program's predictions across different locations in the indoor area and discuss interesting/important aspects of your work/design. Maximum 10-min, 100MB.

Assignment Project Exam Help

Marking Rubric

Dataset (8 Marks):

1. Data volume (5 marks)
 - a. The amount of data collected was adequate for detection training and testing. The relationship between volume and accuracy was shown clearly to establish that data collection was adequate. (3 marks)
 - b. The amount of data collected was adequate for detection training and testing. But impact of data volume on localization error (<10m) detection training and testing. But impact of data volume on detection accuracy was not shown explicitly. (2~3 marks)
 - c. The data volume was clearly not adequate. (1~2 marks)
 - d. No data is uploaded, or the data format does not meet the requirements. (0 mark)
2. Data quality (3 Marks)
 - a. The data set is of high quality, has clear labels, very easy to identify, and there is no redundant data. (2~3 marks)
 - b. The quality of the data set is average, with redundant or duplicate data. (1.5~2 mark)
 - c. The quality of the data set is poor, it is difficult to understand their labels and there are redundant or duplicate data. (1~1.5 mark)
 - d. No data is uploaded, or the data format does not meet the requirements. (0 mark)

Code (5 marks)

1. Complete working code with clear README instructions submitted, and the code compiles and executes properly. [4-5 marks]
2. Code not fully working and/or README file does not explain clearly how the code should be executed [2-3 marks]
3. No code submitted [0 marks]

Report (7 marks)

1. Algorithm design (2 Marks)

- Came up with an innovative idea/algorithm or borrowed/finetuned an existing algorithm from the literature; the idea/algorithm is presented clearly. (1.5~2 Marks)
- An existing idea without any innovation/improvement/adjustment. (1 ~ 1.5 Marks)
- An idea that has obvious logical problems or is incomprehensible. (0 ~ 1 Marks)

2. Experiments (2 Marks)

- The design and execution of the experiments to collect data were clearly explained with sufficient details so a reader can reproduce the experiments/data; the experiments were well designed to produce a valid/useful dataset for the task at hand (1.5-2 marks)
- The descriptions are not detailed enough to be reproduced (0-1.5 marks)

3. Performance (2 Marks)

- Excellent performance (e.g., locations are detected with high accuracies, e.g., median localization error $< 5m$) and there are detailed data, code, and explanations to validate the claims. (1.5~2 Marks)
- Average/medium accuracies, median localization error $< 10m$, and explanations are provided.
- Poor performance (locations are barely detected, median localization error $> 10m$), or there is no evaluation, or there are no results provided. (0~0.5 Marks)

4. Overall organization/presentation (1 Mark)

- The report is easy to read and understand; it has a succinct but clear abstract that reflects the contents of the report; proper conclusions were drawn at the end of the report (1 Mark)
- The report was challenging to read and understand; abstracts and conclusions do not capture the contents well (0-0.5).

Demo Video (5 Marks)

- Video is exciting to watch and clearly demonstrates how the proposed algorithm works possibly with real data and locations; video length and size within limit. (4-5 marks)
- Video is exciting to watch and clearly demonstrates how the proposed algorithm works possibly with real data and locations; video length and/or size not within limit. (3-4 marks)
- Video does not convey the message well and does not demonstrate any working system (0-2)

Reading List

Here are some initial reading materials for you to learn more about location identification from WiFi RSS. Note that some of these articles use different techniques than fingerprinting and some of them use other than RSS data, e.g., time of arrival, etc., but they all have the common goal of identifying locations using WiFi signals. You can read as much as you want/need. Use of WiFi for location identification is a hot topic of research and development and as such there are plenty of literature on this topic. You can also do your own literature search and read more widely or selectively.

Kotaru, M., Joshi, K., Bharadia, D. and Katti, S., 2015, August. Spotfi: Decimeter level localization using wifi. In *Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication* (pp. 269-282).

Vasisht, D., Kumar, S. and Katabi, D., 2016. {Decimeter-Level} Localization with a Single {WiFi} Access Point. In *13th USENIX Symposium on Networked Systems Design and Implementation (NSDI 16)* (pp. 165-178).

Yang, C. and Shao, H.R., 2015. WiFi-based indoor positioning. *IEEE Communications Magazine*, 53(3), pp.150-157.

Garcia-Valverde, T., Garcia-Sola, A., Hagra, H., Dopley, J.A., Callaghan, V. and Botia, J.A., 2012. A fuzzy logic-based system for indoor localization using WiFi in ambient intelligent environments. *IEEE Transactions on Fuzzy Systems*, 21(4), pp.702-718.

Sen, S., Lee, J., <https://eduassistpro.github.io/> multipath to revive inbuilding WiFi *International conference on Mobile systems, applications, and services* (pp.

Add WeChat edu_assist_pro

End of Project Specs and Marking Rubric
