A weighting system for building RSS maps by crowdsourcing data from smartphones

Hosam Alamleh Ali Abdullah S Alqahtani Jean Gourd Hatwib Mugasa

***Abstract*— Mobile devices can sense different types of radio signals. For examples, broadcast signals. These broadcasted signals allow the device to establish a connection to the access point broadcasting it. The received signal strength (RSS) of these signals can be aggregated to form RSS maps. RSS maps show the values of RSSs at different locations. They are useful for several applications such as location fixing, navigation, access control, and evaluating network coverage and performance. In this paper, we propose a method to build RSS maps by crowdsourcing RSS data, GPS locations, from participating mobile devices. The proposed system gives different weights to each data point provided by the participating devices based on the data source's trustworthiness. Our tests showed that the different models of mobile devices return GPS location with different location accuracies. Consequently, when building the RSS maps our algorithm assigns a higher weight to data points coming from devices with higher GPS location accuracy. This allows accommodating a wide range of mobile devices with different capabilities in crowdsourcing applications.**

Keywords*— *Crowdsourcing, RSSI, Map, Smartphone, Wi-fi, Location.

I. Introduction

RSS maps contain the received signal strength information from different access points operating in an area. RSS maps are useful as they can be used for different applications including:

- Location and navigation: In this application, they can be used to calculate a device's location; for example, through pattern matching (e.g., RSS patterns at the mobile device are matched with the RSS patterns on the RSS map to determine the location of a device). Furthermore, they are used to aid in navigating indoors; for example, robots navigating through a maze or in some underground tunnel.
- Access control: In this application, they can be used as proof of location in, for example, access control applications. RSSs can be used to derive a location proof used for authentication and for proximity detection.
- Network performance: In this application, they can be used to monitor network coverage and performance. Since the RSS is an indication of the access point's signal strength at different locations, higher RSS values mean closer proximity – which correlates to better coverage.

Typically RSS maps are built for the purpose of evaluating network coverage and performance by dedicating a number of devices to go around an area of interest and gather the RSS data for the access points in that area. In this paper, we propose a method to build RSS maps by crowdsourcing RSS data from a group of participating mobile devices operating normally.

Crowdsourcing is a process through which a problem is solved or a project is completed by a group of diverse participants. It is a joint process and a problem-solving technique that requires participation from a network of elements. Take for example, the Defense Advanced Research Projects Agency (DARPA) Network Challenge that was launched in 2009[3]. In this challenge, teams competed to locate ten red balloons placed around the United States. The teams were then to report the balloons' coordinates to DARPA. The winning team crowdsourced the problem by recruiting participants via social media [source]a portion of the prize money was used as an incentive). The challenge showed the general effectiveness of using crowdsourcing techniques to solve problems. The DARPA program managers were surprised by how quickly the challenge was completed. The concept of crowdsourcing data from mobile devices developed with the growth of smartphones and the mobile Internet. Now it is used to solve more complicated problems (e.g., Google Maps' traffic conditions feature). In this paper, RSS data is crowdsourced from mobile devices with a specialized application installed on them. This application automatically captures the RSS data of Wi-Fi access points in an area and the device's GPS location. In the proposed system, the collected RSS data is correlated to a location (on a map) through the GPS location provided by the device.

GPS accuracy in mobile devices has improved in the past few years. Most smartphones today have a GPS accuracy of 10 meters in open areas (i.e., suburban settings). It is anticipated to be accurate within one foot in some of the more recent smartphone models [7]. Older devices tend to provide poorer GPS accuracy. In the proposed system, data points provided by a participating device are given a weight based on the trustworthiness of the location provided. This is discussed later in more detail.

II. Related work

First, There have been several research efforts that utilize RSS measurements for localization purposes [6]. The authors (in [2]), for example, propose an indoor positioning system by using matching algorithms [2]. RSS measurements are used for navigation [in 11] and road traffic information [in 4]. The Radar system (in [1]) uses RSS measurements and propagation models to construct radio maps that are used for locating and tracking users inside buildings. The authors (in [9]) proposed to combine a collection of available sources of RF signals crowdsourced from participating smartphones to build signal maps that can be used for localization. The authors use GPS signals, along with other sources such as NFC and QR tags, to determine the location of a device. However, their systems does not take into consideration the variety of the capabilities of the participating phones. Furthermore, it requires auxiliary infrastructure (e.g., NFC and QR tags). Similarly, the system proposed in [4] constructs radio maps using foot-mounted inertial measurement units (IMUs) and GPS position. However, their proposed system requires extra hardware (i.e., IMUs). Kovalev in [8] proposed an indoor positioning system using Wi-Fi and GPS signals. The author's method uses the Naive Bayes classifier to calculate room level positioning. However, it requires extended user interaction, which takes time and effort from the user and negatively affects user convenience. The authors in [12] propose using crowdsourcing to collect GPS location along with RSS, and use it for positioning based on clustering by utilizing the k-means clustering algorithm. However, they don't consider the diversity of the phones' capabilities, as practical crowdsourcing systems usually include wide range of users with a wide range of different devices with different capabilities. Supporting a wide range of devices allow more users to participate, and consequently increases the amount of data collected. The authors (in [10]) present five methods for the generation of WLAN maps for indoor positioning using crowdsourced fingerprints. A fingerprint is assumed to contain identifiers that take into consideration the variety of capabilities of the participating phones. However, their system assumes that GPS position is exact. In a real-world, practical application, this is not typically the case.

Our proposed system does not require any additional infrastructure and takes into consideration the capabilities of different devices used by assigning a weight for each data point provided by these devices. In fact, this is what is unique about the proposed work, and furthermore, what sets it apart from existing methods. Assigning weights to data points increases the robustness of the crowdsourcing system because it decreases the impact of less accurate data on the system. Furthermore, it enables more users with different type of devices to participate in the crowdsourcing application. Meanwhile, as shown later in this text, any device weaknesses can be considered by using weighting factors..

III. System Model

The proposed system architecture consists of the participating devices (which are GPS capable devices loaded with a custom mobile application) and an application server, where participating devices upload their unique ID, GPS location, and the scanned Wi-Fi access points along with their corresponding received signal strength indicator (RSSI) values. Then, the application server arrange the received data and stores it in a database. It then pre-process the data and feeds results to the map building engine. This engine uses the received data from the participating devices to build the RSS map. As previously mentioned, a special application is installed on each participating device. As shown in Figure 1, the mobile application uploads the RSS values from different access points in an area and their corresponding MAC addresses, along with the GPS location observed by the smartphone to the application server using a regular network connection.

In the proposed system, the RSS map is divided into a grid consisting of square-shaped clusters. A cluster is the smallest unit in the map, and it carries the RSS information for that location in the map. The map building engine updates clusters based on the received data from participating devices. The

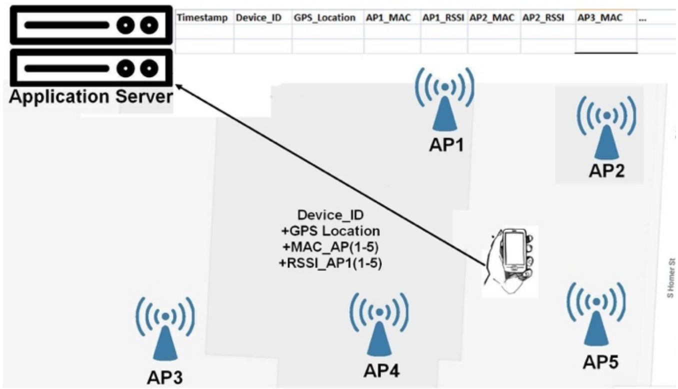

Figure 1: The system's overview.

crowdsourced data is provided automatically by the device (GPS location and RSS data). The proposed system operates as follows:

1. The application installed on the participating smartphones uploads the GPS location along with the MAC address of nearby access points (including their corresponding RSSI values). The upload process is done repeatedly, every defined period of time. This period can be optimized by the user. The user can decide the period based on some factors, including processing power requirements, mobile data requirements, and battery life.
2. The server is configured with parameters for each device ID. These parameters are:
   a) Effective radius: The radius that is used for a data point to update the RSS map. This value depends on the accuracy of the location source (i.e., GPS). The larger the accuracy, the smaller the effective radius. Therefore, data points originating from more accurate location sources result in fewer clusters than data points originating from less accurate participating devices.
   b) Weighting factor: This factor depends on the trustworthiness of the location data source. Data provided by newer smartphones have a greater weighting factor than data provided by older ones. In general, the higher the location accuracy the larger the weighting factor.
3. The map building engine uses the device ID to fetch the pre-configured values of the effective radius and weighting factors for a device.
4. The map building engine calculates the distance between each of the centers of the clusters on the map and the location of a data point. This is done using Vincenty's formula for calculating the distance between two points. Specifically, this function is an iterative method that calculate the distance between two points on the surface of a spheroid. If the distance equals or is less than the effective radius, then the map engine updates the RSSI value for the access points in the cluster. This is done by employing the weighting factor as follows:

$$\mathrm{RSSIm} = \frac{wu * RSSIu + wm * RSSIm}{(wu + wm)}$$

where:

RSSIm is the current RSSI value stored at the RSS map for a specific access point.

RSSIu is the RSSI value provided by a user for the specific access point,

wu is the weighting factor of this data point,

wm is the sum of all weighting factors that updated this cluster in the past.

Then, the map engine updates wm as follows:

$$wm=wm+wu$$

Employing the weighting factors allows more accurate data to have a larger effect on the RSS map than less accurate data. Thus, allowing more diverse range of devices to participate in this crowdsourcing application. This is very useful in the practical cases, where different users use a variety of smartphones' models.

IV. TESTING AND RESULTS

In this research, we tested a number of smartphones manufactured between 2010 and 2018. The test was conducted at six different open-to-sky outdoor locations and ten different indoor locations (inside a two-story building). The location service settings in the phones were set to "device only mode" which relies solely on the GPS radio signal generated by the GPS chip built into the phone, unlike the high accuracy mode which uses a combination of GPS, Wi-Fi, Bluetooth, and/or cellular networks. The GPS yields were compared to the ground truth (which are the points of the test with known coordinates). The accuracy was calculated as the average of the distances between the GPS yield and the ground truth for each of the devices. Results, as shown in Table 1, show that different phones have different GPS location accuracies. Furthermore, it shows that newer devices, relatively, have better GPS location accuracies.

Table 1: GPS location accuracy for different smartphones.

Device Model	Release Year	Indoor accuracy (m)	Outdoor accuracy (m)
Nexus	2010	24.45	31.74
LG G3 Vigor	2014	15.74	29.05
Alcatel Flash	2014	8.67	21.45
Motorola Moto Z	2016	10.83	20.97
Oneplus 3	2016	9.32	16.82
Samsung Galaxy S9	2018	5.94	12.45

In our experiments, the six Android-based smartphones shown above were used. Location service settings were set to the device-only such that the location is generated solely based on the GPS integrated circuit (i.e., without correlating it with Wi-Fi, Bluetooth, etc.). An android application was designed and installed on these devices as shown in Figure 3. This application performs the simple function of uploading the GPS data and RSS data for scanned Wi-Fi access points every predetermined period of time. In this experiment this period of time was set to five seconds (to accelerate data collection in order to collect large amounts of data for analysis). However, in non-experimental applications, this period of time can be set to a minute to avoid overwhelming the resources of participating smartphones.

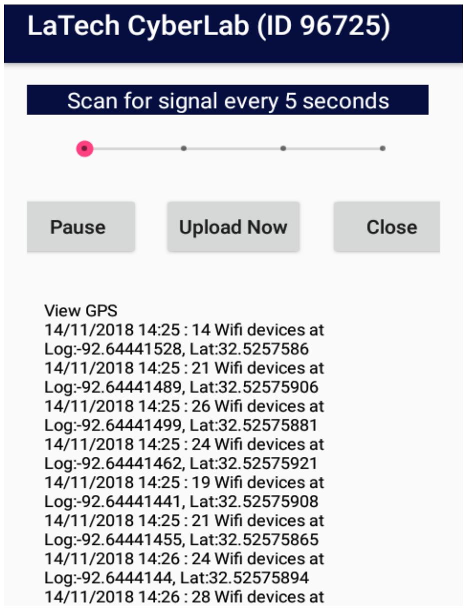

Figure 3: The Android application running on a smartphone

Tests were conducted in a two-story building, where data was collected from the ground floor only (in this experiment, altitude was ignored). Six different devices were used for data collection. The RSS map size was 100 meters by 100 meters, which covers one campus building and its surroundings. The cluster size in this experiment was set to one square meter. This size was chosen according to the GPS accuracy of the participating devices (the most accurate device in the experiment has GPS accuracy of 9.53 meters).

The application was used to collect approximately 6,000 data points divided equally between the participating devices. Approximately 800 unique access points were scanned. The data points were processed by the map building engine on the application server to build the RSS map. The collected data points were fed to the map engine, and the RSS map was built using the criteria above.

In order to test the generated RSS map, 100 additional data points were collected at locations with known coordinates. The testing is done by comparing the value of the RSSs of the test data points versus the clusters in the map with the closest values of RSSs. The location of that cluster was then compared with the ground truth that was recorded during the test. This was used to calculate the distance between the true location and the location returned by the RSS map. The results indicate that large location errors are less likely when a high weighting factor is present. This shows how data from less accurate devices can be balanced with the data from more capable ones allowing. This solves the problem of diverse devices in crowdsourcing for practical cases, and allows a wide range of users to participate in the crowdsourcing applications. The results shown that the average location accuracy using RSSI was 13.69 meters compared to 18.35 meters representing the average of the GPS accuracy of the participating devices. This shows slight improvements due to the implementation of the weighing factor. Figure 4 shows the relationship between weighing factor and the location accuracy of test points. As can be seen from the figure, it is more likely to get more accurate locations when the weighing factor for a grid in the map is higher.

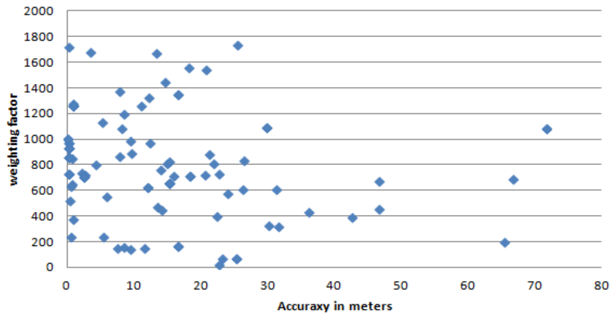

Figure 4: The relationship between weighing factor and accuracy.

V. CONCLUSION

The proposed methodology achieves an effective approach to generate RSS maps because it does not require dedicated devices and personnel to collect RSS data manually. As discussed previously, the use of crowdsourcing saves many resources and offers clear performance advantages. The proposed system takes into consideration the diversity of mobile devices used by users and does not require any extra infrastructure. This work is different from current work in that it supports a wide range of devices. This fits well in real-life applications, where there is a wide range of devices' models with different manufacturers and capabilities. The proposed system addresses this by allowing a wide range of users to participate in the crowdsourcing application without affecting the performance of the system. This is achieved by employing weighting factors for each of the data points based on the trustworthiness of the source. The most trustworthy devices have a larger effect on the system than less trustworthy ones.

REFERENCES

[1] P. Bahl and V. N. Padmanabhan, "Radar: An in-building RF-based user location and tracking system," in INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies, vol. 2, no. 10, March 2000, pp. 775-784.

[2] Z. Chong, H. Yu, "Indoor matching location algorithm based on WIFI signal strength", International Conference on Communications Circuits and Systems IEEE, pp. 116-119, 2013.

[3] Defense Advanced Research Projects Agency. "DARPA Network Challenge Project Report". 2012

[4] Y. Gu, C. Zhou, A. Wiese, Z. Zhou "Trajectory estimation and crowdsourced radio map establishment from foot-mounted IMUs, Wifi fingerprints and GPS positions" DOI 10.1109/JSEN.2018.2877804, IEEE Sensors

[5] D. Gundlegard J.M. Karlsson "The smartphone as enabler for road traffic information based on cellular network signaling" Intelligent Transportation Systems - (ITSC) 2013 16th International IEEE Conference on pp. 2106-2112 6-9 Oct. 2013.

[6] V. Honkavirta, T. Perala, S. Ali-Loytty, R. Piché, "A comparative survey of WLAN location fingerprinting methods", Proc. IEEE WPNC, pp. 243-251, 2009.

[7] J. Kastrenakes. "GPS will be accurate within one foot in some phones next year" https://www.theverge.com/circuitbreaker/2017/9/25/16362296/gps-accuracy-improving-one-foot-broadcom

[8] M. Kovalev "Indoor positioning of mobile devices by combined Wi-Fi and GPS signals " 978-1-4673-8054-6

[9] P. Mirowski T. K. Ho S. Yi M. MacDonald "SignalSLAM: Simultaneous localization and mapping with mixed WiFi Bluetooth LTE and magnetic signals" Proc. Int. Conf. Indoor Position. Indoor Navig. (IPIN) pp. 1-10 2013.

[10] M. Raitoharju, T. Fadjukoff, S. Ali-Loytty, and R. Piche, "Using unlocated fingerprints in generation of WLAN maps for indoor positioning," in Proc. of Position Location and Navigation Symposium, pp. 23-26, 2012

[11] K Wu, J Xiao, Y Yi, "CSIBased Indoor Localization[J]", Parallel & Distributed Systems IEEE Transactions, vol. 24, no. 7, pp. 1300-1309, 2013.

[12] Y. Zhang, S. Zhang, R. Li, D. Guo, Y. Wei, Y. Sun "WiFi fingerprint positioning based on clustering in mobile crowdsourcing system" Electronic ISSN: 2473-9464