2012 International Conference on Computer Technology and Science (ICCTS 2012)
IPCSIT vol. 47 (2012) © (2012) IACSIT Press, Singapore
DOI: 10.7763/IPCSIT.2012.V47.40

Analysis of Fingerprint Data in Cellular Networks: An Android Application Case Study

Dr. SRN Reddy, Akriti Bagaria, Divya Aggarwal, Namrata Arora

Department of Computer Science and Engineering,
Indira Gandhi Institute of Technology, Kashmere Gate, Delhi

Abstract. Positioning systems that use location fingerprinting techniques are gaining popularity in recent years due to their cost effectiveness and accuracy as compared to other infrastructure based localization techniques. Fingerprinting technique records the vectors of received signal strength from several mobile base stations in a database and later matches these to a new measurement to get the location of the user. Data analysis of received signal strength is essential for understanding and predicting the behavior of fingerprints under different physical situations. Such analysis enables the designer to glean insights into location dependent features of location fingerprints. This would facilitate the system designers in accurately modeling an efficient system with improved positioning performance. In this paper we present the analysis of the radio map of test area extracted from the GSM network with various metrics like time, orientation etc. and try to infer the relationships, dependencies among them.

Keywords: Location fingerprinting, GSM, received signal strength

1. Introduction

Location based services contribute a major share to the applications on smart phones in recent developments. The performance of such services depends solely on the underlying localization method used. Many techniques have been explored the most advent and popular being GPS. Methods based on measurement of angle and time metrics have been also devised like Angle of Arrival (AOA), Time of Arrival(TOA),E-OTD in cellular networks[1]. Extracting location from radio map which is collection of time stamped signal samples more abundantly referred to as fingerprinting has also been considered [2]. Here location sensitive parameters like signal strength, Cell-ID, LAC and others can be exploited to position the device. Fingerprinting involves construction of a database of samples and subsequent matching process with unknown samples to determine the location.

Cell-ID based positioning is the cheapest, simplest of procedures but lacks in the area of accuracy. Thus a variety of modifications have been suggested in researches which boost its capabilities. Fingerprinting GSM network data requires building a radio map of received signal strength.

2. Background

2.1. Location fingerprinting systems

In location fingerprinting systems the radio database is populated with set of records . Each record is a vector R=[S(x), x] where ‘x’ represents geographical location coordinates in terms of Latitude and Longitude. S(x) represents a measurement vector of signal strengths received from one main base station and a set of neighbouring base stations. For a particular area A and at a specific location L the radio fingerprint is stored in database as {id, s, n_1, s_1, n_2,s_s, …., n_i, s_i, L_x} w here ‘id’ is the cell id of the main base station and ‘s’ is it’s received signal strength at location L . The symbols n_1, n_2…..n_i represents the cell id’s of i neighboring base stations and s_1,s_2,…..s_n are their respective received signal strengths.

The radio database is then given by R = $\{r_n\}$n=l...N. The positions in X may be picked up according to a regular geographical pattern, e.g. a uniform square grid ([3]).During the localization phase the mobile performs a measurement at location x . The estimated position according to the basic nearest neighbour method is obtained by calculating Euclidean distance between observed fingerprint and fingerprints present in database.

Not much has been seen in statistical analysis of GSM fingerprints. But work in [4] discusses model for probability distribution of fingerprint selection. It puts into use Voronoi diagrams and graphs to study structure of fingerprints in WLAN network. Research carried out in [5] explores evaluation of four different localization algorithms (MinMax, Maximum Likelihood, Ring Overlapping Circle RSSI and k-Nearest Neighbor) in signal strength based localization. Effects of grid resolution,temporal variations have been observed in the analysis of radio map [5]. In [5] authors take measure of the ramification of the user's body, orientation and dependency between various RSS vectors.

3. Android overview

Android is an open source software stack for mobile devices. The Android OS is based on modified linux kernel. The android platform is coded in C for its core, C++ for third party libraries and Java for user interface. Android software stack runs on a java based, object oriented application framework. It operates over Java core libraries running on Dalvik Virtual Machine (VM). Prior to execution Android Applications are converted into Dalvil executables(DEX) format, rendring them suitable for portable devices with memory and processing speed constraints. It is a register based architecture.

3.1. Architecture

The Android platform is composed of 4 layers: Applications at the top, an Application Framework layer that provides services to applications, e.g., controlling activities or providing data access, a Library/VM layer, and at the bottom, the Linux kernel. Applications run at the very top of the platform. Services for applications, e.g., the Activity Manager, which controls activities for each application, or Content Providers which load the content provider defined by each application while restricting data accessibility across applications are located in the Application Framework layer. The Library/VM layer contains static libraries and the Android runtime. environment. Static libraries provide common system and application libraries for applications. The Android runtime environment is composed of core runtime libraries and the Dalvik virtual machine (VM)—an optimized Android-specific Java virtual machine. Finally, the Linux kernel completes the OS and the software stack. Each Android application runs with a unique user ID, in its own copy of the Dalvik virtual machine, which ensures separation between applications and provides protection. Android applications can be composed of four component categories: Activity, Broadcast Receiver, Content Provider and Service. Activities are focused windows in which the user interaction takes place; only one activity can be active at a time. Each activity is a class in the source code and should perform according to events generated by users and system. Services run in the background, e.g., an email client may check for new mails while users are running another application. A Content Provider manages data for a certain application and controls the accessibility of the data; for example, an email client may make email addresses in its database accessible to other applications. Broadcast Receivers listen and react to broadcast announcements. For example, an email client may receive a notification that the battery is low and, as a result, proceed to saving email drafts.

3.2. Positioning using android API

Android.telephony package is used to extract and monitor various cellular network information like cell-ids and received signal strengths of current and six neighbouring cells. The locationManager system service in android.location package has been accessed for determining current GPS coordinates and timestamps.

3.3. Database

Android provides full support for SQLite databases. The package android.database.sqlite . management to manage database. The data stored in the database is Cell-id, RSS of serving BS and up to 6 neighboring cells, GPS readings and Timestamp.

Table 1: Radio Map Structure

Sno.	CID	RSS	NCID1	NRSS1	NCID2	NRSS2	NCID3	NRSS3	LAT	LON
1	2082	-61	44202	-71	2084	-71	2081	-59	28.6642	77.2328
2	2082	-59	44202	-69	2084	-67	2081	-61	28.6642	77.2328
3	2082	-57	44202	-65	2084	-63	2081	-61	28.6642	77.2328

3.4. Test Area

A ground (115m * 62.5m) was chosen as the experimental Field, which is situated in Indira Gandhi Institute of Technology campus. It is depicted in the following Figure 1. Pedometer was used to measure step count. Single step size was approximated to be 0.69 m. Cell size of 30X30 meters was taken. Readings were logged at centre of each cell every 45 seconds. GPS was turned on at this time to get latitude, longitude values. Different orientations of the phone were considered at each predefined position. Then the fingerprinting database for positioning user's location in this ground was set up.

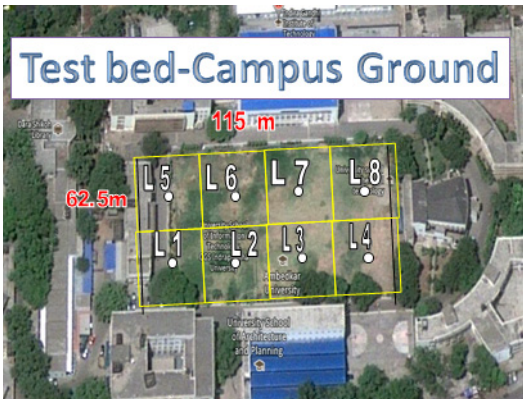

Figure 1: Test Bed

4. Analysis of fingerprints

Data analysis of received signal strength is essential for understanding and predicting the behavior of fingerprints under different physical situations. Such analysis enables the designer to glean insights into location dependent features of location fingerprints. Radio database acts as a key element for location fingerprinting systems. Performance of such systems is highly dependent upon the quality of data in the radio database. The knowledge gained by the analysis process can assist the designer in accurately modelling a positioning system, improving positioning performance, and efficiently designing such a system. .This study investigates extensively through measurements, the features of the received signal strength. The results of the statistical data analysis help in identifying a number of phenomena that affect the precision and accuracy of indoor positioning systems.

4.1. Received signal strength properties

Behavior of received signal strength is difficult to predict because of the dense multipath environment and propagation effects such as reflection, diffraction, and Scattering. Understanding the statistical properties of the location fingerprint is important for the design of positioning systems for several reasons. It can provide insights into how many neighboring base stations are needed to uniquely identify a location with a given accuracy and precision, whether pre- processing of the RSS measurements can improve the accuracy. In our RSS measurements, the values are integers ranging between 0 to -110 dBm. The location of our experiment is campus ground.

4.2. Effect of User Orientation

Depending upon the orientation of user the received signal strength may vary due to direction of mobile device's antenna. During the previous works, an observation was made that the user's orientation could cause a variation in RSSI level of up to 5 dBm. However, no analysis of the RSSI data was provided. Different orientations of user and mobile device could change the mean values of RSSI at a location. Some studies referred to this phenomenon as radio irregularity. Radio irregularity is caused by two categories of factors: devices and the propagation media. Device properties include the antenna type (directional or omni-directional), the sending power, antenna gains (at both the transmitter and receiver), receiver sensitivity, receiver threshold and the Signal-Noise Ratio (SNR). Media properties include the media type, the background noise and some other environmental factors, such as the temperature and obstacles within the propagation media. In general, the radio irregularity is caused by the non-isotropic properties of the propagation media and the heterogeneous properties of devices.

To study the impact of user orientation we performed measurements in four different orientations of mobile device (North, East, West, South) at eight different Locations L1,L2,L3,L4,L5,L6,L7,L8 as shown in Table 2. For each orientation sample finger prints were collected over a duration of 45 seconds at each location .The mean and variance are then calculated for each sample set. The results are shown in Table 2. It can be observed that average variance is least for west direction. Hence west direction can be chosen for construction of radio database. The effect of user's orientation is significant and the orientation should be recorded in the database.

Table 2: Mean and Variance values for different orientations

NORTH	**L1**	**L2**	**L3**	**L4**	**L5**	**L6**	**L7**	**L8**
Mean	-57	-55	-53	-55	-67.67	-71.33	-73.67	-69
Variance	7.11	8	6	6.67	1.67	16.67	5.867	17.6
WEST								
Mean	-53.4	-53.67	-53.28	-55.25	-58.55	-69	-72.5	-71
Variance	2.24	4.266	4.5	2.785	2.778	0	.857	5.11
SOUTH								
Mean	-59	-54.14	-58.81	-53.25	-61	-69.4	-72	-70.4
Variance	9	9.14	8.36	5.07	2	10.8	14.9	6.28
EAST								
Mean	-54.5	-57	-60	-55.44	-68.14	-69.4	-72	-74
Variance	4.9	3.2	4.4	5.7	21.14	2.28	3.42	5.714

4.3. Time dependency of RSS

Our studies show that RSS is time dependent .To analyse the time dependency we performed various sets of experiments. These experiments were aimed at determining time dependency over hours of a day and days of a week. All measurements were done at location L1. The sample fingerprints were collected for a duration of 45 seconds at various hours of a day .For each set of sample fingerprints mean is calculated and histograms are plotted as shown in Figure 2. To understand the property better, we require more measurements and this is part of our ongoing research work.

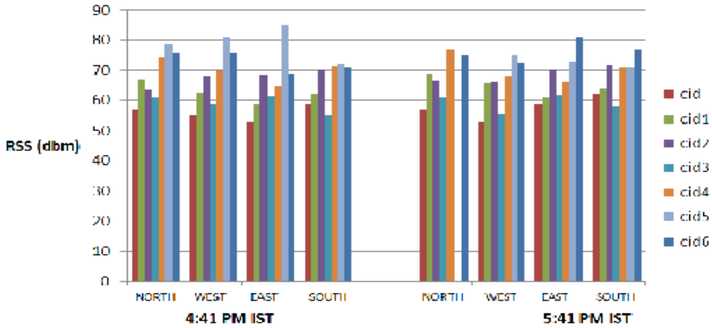

Figure 2: Time dependency

4.4. Statistical properties

To observe the statistical properties if RSS we collected the 15 minute long samples on location L1. Then the frequency of each signal strength value is observed and probability distribution curves are plotted .We analysed the curves and found that practically the curves are not symmetric about mean.

Ideally the RSS should be log normally distributed which is symmetric around the mean value but our observations show that the RSS distributions are asymmetric and non Gaussian. Some distributions are left skewed and some were observed to b right skewed. However, some distributions with a weak mean RSSI could be approximated by the log-normal distribution. The plots are shown in Figure 3. The randomness of RSSI patterns is clearly described by its probability distribution function (PDF) or its distribution. To understand the cause of the error, we need to understand the nature of the randomness of the RSSI[7]. Three different RSSI distributions are shown for comparison purpose. The distribution of the RSS is not usually Gaussian, it is often left-skewed and the standard deviation varies according to the signal level. Because of the complexity of radio propagation the distribution of RSS is difficult to model and fit to well known distributions[6]. System designers will be greatly benefitted if could find the approximate distribution of underlying RSS process. This is a part of our ongoing research work.

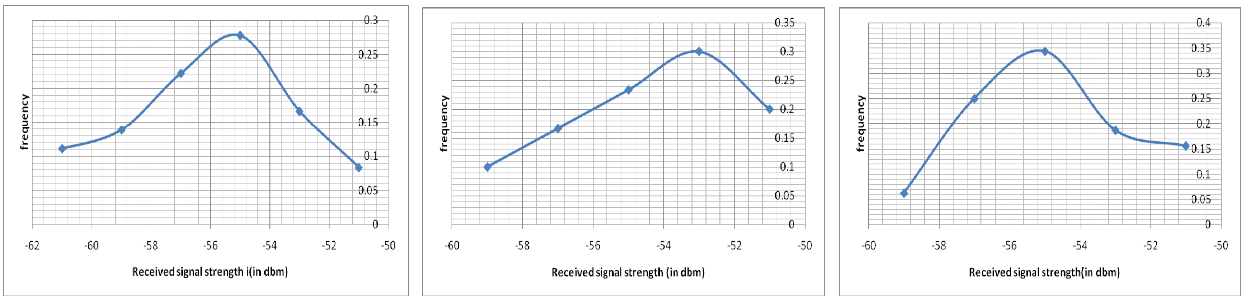

Figure 3: Probability distributions

5. Conclusions

We presented the data analysis of received signal strength values. Our studies show that user presence should be taken into account while collecting the sample fingerprints the effect of user orientation is significant and should be recorded in the fingerprint database. Although the mean usually stays nearly around the same value the variance could shift with large amount along with each orientation. . Changes in environment such as a human movement could also change the mean RSS .We also analyzed the probability distribution curves for sampling process and we found that distribution of RSS is not usually Gaussian. The average RSSI is usually modelled by a log-normal distribution which is symmetric around a mean value, but our measurement results show that most distributions are often left-skewed. Another most influential effect is the time dependency, which should be included in the design of positioning systems and when collecting the location fingerprints. Due to the time dependency property of RSSI, the location fingerprint collection process should be done at different periods of day for best results[6].Since the standard deviation or variance of RSSI is the most important factor which should be included when forming a location fingerprint beside the average of RSSI, a suitable pattern classification that should provide better location determination performance should include both mean and variance of RSSI into its consideration.

6. References

[1] E .Trevisani and A. Vitaletti, “Cell-ID location technique, limits and benefits: an experimental study,” Proceedings of the 6th IEEE Workshop on Mobile Computing Systems and Applications, pp. 51-60, 2004.

[2] H. Laitinen, J. Lahteenmaki, and T. Nordstorm, “Database correlation method for GSM location,” Proceedings of the IEEE Vehicular Technology Conference, pp. 2504-2508, 2001.

[3] A. Arya, P. Godlewski, P. Melle, "A hierarchical clustering technique for radio map compression in location fingerprinting systems," in Proc. International Conference on Vehicular Technology, May 2010, pp. 1-5

[4] Nattapong Swangmuang, Prashant Krishnamurthy: Location Fingerprint Analyses Towards Efficient Indoor Positioning. PerCom 2008: 100-109

[5] Azin Arya, Philippe Godlewski: An analysis of radio fingerprints behavior in the context of RSS-based location fingerprinting systems.PIMRC 2011: 536-540

[6] Kamol Kaemarungsi, Prashant Krishnamurthy: Analysis of WLAN's received signal strength indication for indoor location fingerprinting.Pervasive and Mobile Computing 8(2): 292-316 (2012)

[7] Kamol Kaemarungsi, Prashant Krishnamurthy: Properties of Indoor Received Signal Strength for WLAN Location Fingerprinting.MobiQuitous 2004: 14-23