Journal of Applied Sciences

ISSN 1812-5654

Journal of Applied Sciences 12 (1): 78-83, 2012
ISSN 1812-5654 / DOI: 10.3923/jas.2012.78.83

Location Determining Technique Using Signal Correlation Method in Suburban Environment

K. Singh, S. Sulaiman, M. Ismail and K. Jumari
Department of Electrical, Electronics and System Engineering, Faculty of Engineering
Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor DE, Malaysia

Abstract: In this study, a Location Determining Technique (LDT) to determine the approximate position of a cellular mobile user in a suburban environment is presented. This LDT uses Signal Correlation Method (SCM) where the Received Signal Strength (RSS) of one Node B is used to approximate the location of the 3G data user in a suburban environment. A new training technique called Unique Sample Undefined Collection (USUC) is introduced for SCM which uses Neural Network to match the current RSS with stored signal strength of Node Bs detected around the User Equipment (UE). With this new technique, the Cumulative Distribution Function (CDF) for location estimation at 67% error distance between the simulated samples and the actual positions of UE was significantly reduced from 379 to 107.4 m, while at 95% of the simulated samples, the error distance remains approximately the same, at 379 m. SCM-USUC may not be suitable for emergency location search but should be utilized by Universal Intelligent Positioning System (UIPS) for tracking location of UE when data or voice services are used especially in larger cell areas, such as in suburban environment.

Key words: Location determining technique, location based services, mobile positioning

INTRODUCTION

Universal Intelligent Positioning System (UIPS) was introduced by Singh and Ismail (2005) in order to provide the different levels of Quality of Positioning (QoP) based on the request search categories. The search categories could be defined as emergency search, navigation search or for Location Based Services (LBS). In the US, FCC (Federal Communication Commission) has requested cellular operators to comply with the US Emergency 911 standards, where the location (position) of 95% of calls made to E911 number, should be able to be searched within 300 m and 67% of the location search should fall within 100 m. Therefore, emergency type of location search will require the highest QoP from UIPS's LDT module. UIPS's LDT module consists of various location techniques such as based on OTDOA (Observed Time Difference of Arrival) from at least three unique Node Bs, Signal Correlation Method (SCM) of one Node B (Singh *et al*., 2008a) and a few other collection of techniques (Singh *et al*., 2008b). UIPS will decide based on the criteria received which LDT should be used in order to determine the location of the mobile user. For example, if only one Node B is hearable by the UE and the user is in urban area (where cell sizes are small), UIPS will check if there are any stored information on RSS from the detected Node B. Furthermore, if the service requires high QoP, then UIPS will utilize SCM with LEAN (Learn Another) technique (Singh *et al*., 2008a) in order to determine the proximity of the mobile user. This technique when simulated, estimated 95 and 67% of location search accuracy falls within FCC's E911 standards. However, in suburban areas, where there are fewer cells and the sectored cell size coverage is large, SCM LEAN could not accurately estimate locations that could fulfill emergency search standards. In this situation, techniques such as Database Correlation Method (DCM) as utilized by Laitinen *et al.* (2001) and Ahonen and Eskelinen (2003) could provide better location estimation, where more Node B's signals' are matched with several received Node Bs signals. This will increase the chances of matching more stored samples data in order to reduce the location estimation error. Zhu and Durgin (2005) also utilized the received signal strength values in their research in order to determine the mobile user's location.

However, in this study, when only one Node B's RSS measurement is achievable, UIPS would need to utilize SCM with a new technique that accommodates signal matching with larger cell size for data or voice bearer, within suburban environment.

Corresponding Author: K. Singh, Department of Electrical, Electronics and System Engineering, Faculty of Engineering, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor DE, Malaysia

MATERIAL AND METHODS

The overview of methodology in Singh *et al.* (2008a) was described where in the beginning drive test data is collected using Telco's drive test equipment or self developed tools (application on phone to collect RSS of detected Node B, Location Area Code (LAC), Cell ID of Node B, GPS coordinates, speed of vehicle, etc.). Secondly data is analyzed using statistical analysis tools, with concentration on data correlation, mean and deviations with respect to base station distances from each other or geographical placements. For areas where the coordinates of Node B (base station) are not know, some estimation of distance to the actual test point coordinate may be required (the test point coordinate will be measured using GPS). Then a simulator is build for UIPS where an initial technique is utilized to predict location based on the stored values and some simulated range of values for RSS, Cell ID and LAC of Node B. Then the result is analyzed in terms of 67 and 95% location estimation error based on E-911 standards. If the result does not meet this benchmark for emergency location search, then further studies are conducted in order to find ways different to collect data and to improve the matching technique between stored data and received signal strength data of the one Node B. This process will go through a feedback loop for several months until a new technique for learning and training the samples is developed. Finally, this process will be integrated as part of the collection of LDTs for UIPS.

According to Singh *et al.* (2008a) at least three drive tests were necessary to be conducted on the same route in urban Kuala Lumpur for data collection. After data was stored, the simulation for SCM using the LEAN technique was able to meet the 67 and 97% of location accuracy for E-911 standards in urban. But SCM LEAN fails to meet E-911 standards for suburban when tested because, the cells in suburban areas are larger and the grid size used for storing drive test data of urban samples for SCM-LEAN was 100 by 100 m. Furthermore, in urban Kuala Lumpur, SCM-LEAN is successful to be utilized even though with one Node B because 95% of the Node B are within 1796 m. In suburban and rural areas, Node Bs could be up to several km (Kupper, 2005).

Data collection for suburban using UMTS data service: For suburban environment, the grid size was increased to 200 by 200 m since Node B base stations cover larger areas. The route chosen was in the town of Cheras, about 20 km from Kuala Lumpur, Malaysia. Figure 1 illustrates the drive test route conducted for this experiment during the data collection process in the same day during the night h. The speed of the vehicle is maintained around 60 km h^{-1}.

The tools used for data collection are as following:

- Nokia N95 phone with built in GPS (Global Positioning System) receiver
- An application developed in Symbian Operating System that is capable to grab RSS, Cell ID and LAC information of the detected Node B and coordinate of GPS (in longitude and latitude) when samples are collected every interval of 5.

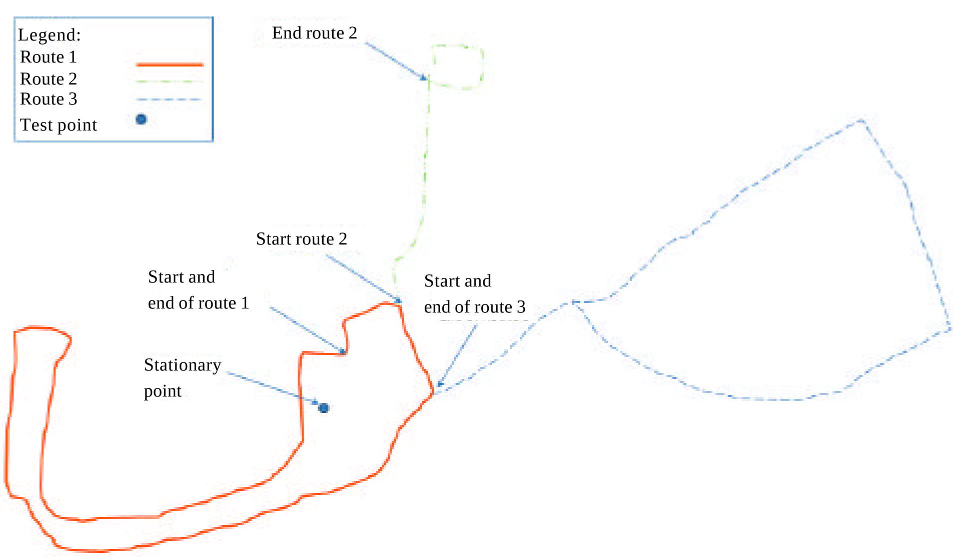

Fig. 1: Drive test routes in suburban Cheras with 3 different routes and one stationary test point

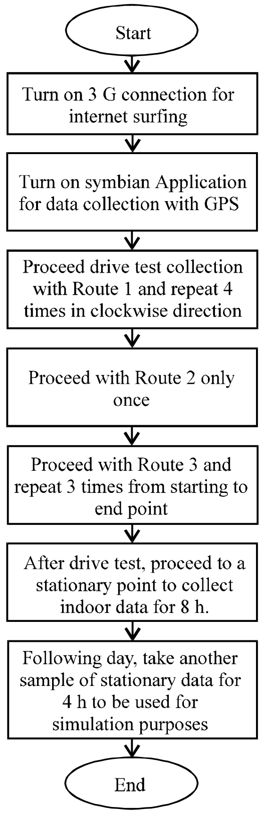

Fig. 2: Drive test and stationary point data collection process for SCM in suburban

Each sample collected for a route consists of the following information:

R_{ij}= [RSS, Cell ID, LAC, Longitude, Latitude], where, i represents which route is being collected and j represents the sample number for that particular route. RSS from the serving cell of Node B will be represented in dBm while Cell ID will be in decimal format.

The method to collect data is as described in Fig. 2, where the phone was set to view http sites using 3G internet access point of Telco's data service during the entire duration of the drive test.

The GPS of the Nokia N95 phone is turned on and calibrated with another standalone unit of GPS. It is then compared with the starting point's coordinate as viewed from Google Map (http://maps.google.com.my/). Then the Symbian application is turned on, where the collection process starts for Route 1. Route 1 is about 6.15 km in distance and is repeated for 4 times. This is because the three times data of the same route will be used for storing purposes and the other one time for simulation process to estimate the mobile location. Route 2 is about 1.3 km and is only done once because it is a short route, while Route 3, 4.6 km is repeated 3 times. After the drive test was completed, the phone was kept in an indoor location for 8 h to record stationary test point's RSS, Cell ID and LAC. This data will be used by SCM for storing, while another 4 h of recording was done the following day in order to be utilized by SCM's simulation to estimate the drive test and stationary collection data from the previous day.

The phone was on active 3G data service, where in cases it will be automatically downgraded to GPRS in the event where cells do not provide coverage for 3G data. After the data collection process is completed, the next step would be to analyze the data and develop a technique suitable to learn and train the stored data.

Data analysis and the development of SCM-USUC for suburban environment: It was assumed during the data collection, Telco did not perform any GPRS or 3G network optimization on the same night and the next day the data samples were collected. In general Telco's personnel would perform network optimization work several times in a year, where cell numbers, transmit power and antenna's direction could be changed. First of all the sample points within the drive test route and the stationary data point is analyzed. The whole area is gridded into 200 by 200 m, where there could be between one to four samples represented within the 200 m square grid area and at times where the drive test route is far from the housing area, no sample are represented for the housing area. For example, the stationary test point data did cover similar Node B's Cell ID as the drive test samples from Route 1 since it is quite close to Route 1 (on the left of the test point). But now the next challenge is to determine which cells are dominant and which range of RSS should be stored for each grid. The issue is, if there are two much similarity between Grid A and the neighboring Grid B, the SCM's correlation method may not be able to select which location of grids the mobile user will be located in. For example, when data was collected for 8 hours for the stationary test point (let say Grid X), Cell 101 collected RSS values from -80 till -88 dBm, while Cell 104 collected RSS values from -92 till -98 dBm, Cell 110 collected values from -88 till -95 dBm, Cell 120 collected values from -82 till -90 dBm and Cell 122 collected RSS value of -80 dBm. Only unique samples of RSS occurrence between CDF of 5 to 95% of the serving cells are stored within SCM for the representation of the suburban area. For example, the mean, the deviation and the SS range for Cell 101 is determined and followed by the rest of the cells detected within Grid X. Then the dominant RSS ranges (for example RSS value -83-84 dBm and -85 dBm are dominant and will be stored for Cell 101) for all the pertaining

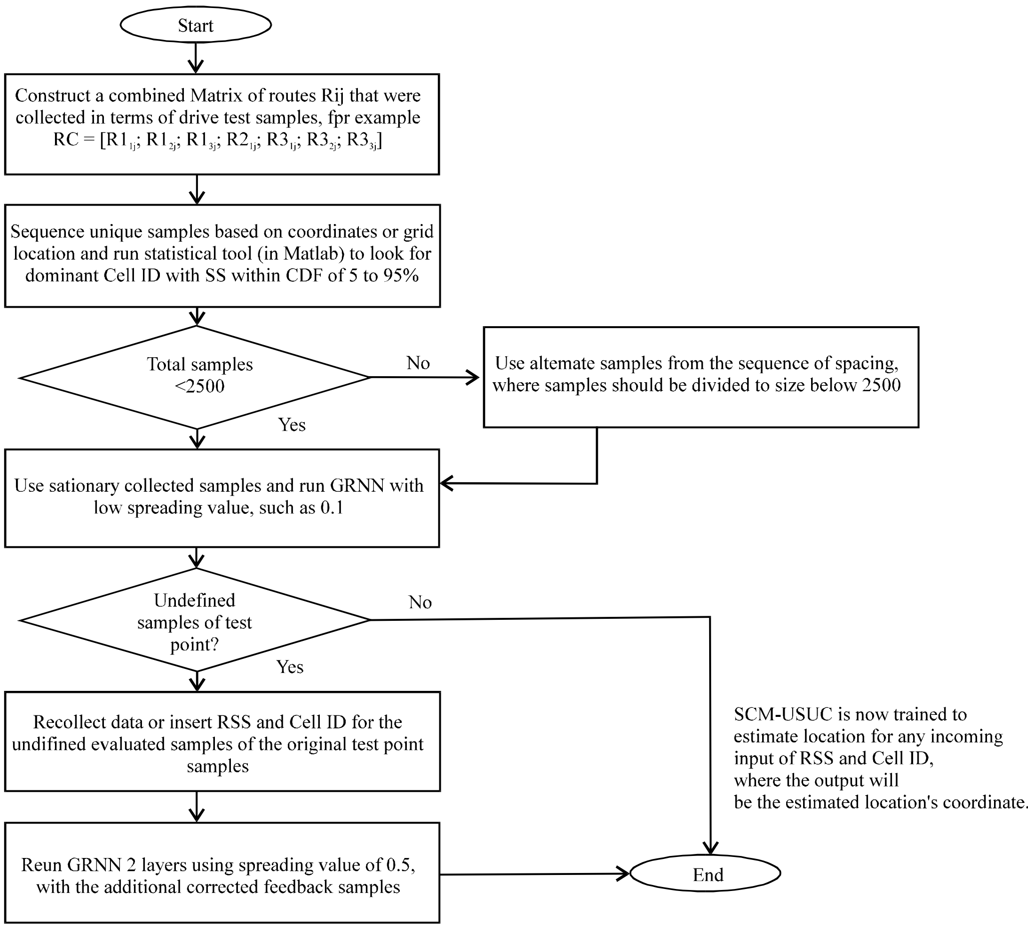

Fig. 3: SCM-USUC stored data to be utilized for location estimation

cells are stored for Grid X. After unique data is stored, SCM uses Generalized Regression Neural Network (GRNN) with low spreading value such as 0.1 for matching stored data with evaluated sample data consisting of RSS value and Cell ID value. The output will be the coordinate of the evaluated sample's location in longitude and latitude. This process uses Matlab tools for Neural Network (Demuth *et al*., 2008).

Figure 3 describes the process used for developing SCM-USUC samples which are stored and ready to be utilized for any location estimation request by UIPS based on LAC surrounding the suburban area.

Firstly, the drive test data routes will be combined to represent the entire suburban area's data collection. The uniqueness of this data is being analyzed based on taking dominant cells with frequent RSS values. If the samples are too big, the processing power will be reduced for a given area being evaluated. Therefore, from experiments, 2500 samples should be the optimum number of samples to cover a small suburban town. Then this unique drive test samples are stored in order to be matched with the stationary test point's samples. SCM is simulated using GRNN with small spreading value. The purpose of using this small spreading value is to test the robustness of the unique drive test samples in covering housing areas (stationary test point) within the roads being drive tested. GRNN will match the stored data with the stationary point's collected data (RSS and Cell ID) in order to predict the location of the mobile user. With low spreading value, when RSS and Cell ID are badly correlated with the drive test stored data, GRNN will not produce any output. The sample being evaluated would be defined as undefined collection and thus the program would reinsert the value of RSS and Cell ID into the unique drive test samples to increase the robustness of the entire samples, representing the suburban area. Sometimes

Table 1: Simulation parameters for SCM-USUC

Parameters	Descriptions
427 unique samples	Drive test samples for route 1, 2 and 3 collected for suburban area
177 unique samples	Indoor stationary test point data collected around same night of drive test
793 samples	New indoor stationary test point data collected the following day will be used for location estimation
GRNN with spreading value of 0.1	Simulation with SCM-US: using stored drive test data (427 samples) with indoor stationary data (177 samples)
GRNN with spreading value of 0.5	Simulation of SCM-USUC: using missing undefined samples added to the original 427 drive test samples and compared to the new 793 samples of indoor stationary test point data

only one Cell ID with one RSS could improve the location estimation process for the entire suburban area. Therefore, more stationary points should be located within the other routes to represent the different housing areas that may be far away from the roads being drive tested on. Finally, the unique combined drive test stored data is reinserted with the corrected feedback samples (additional Cell ID with RSS values from the stationary test point) and rerun with GRNN with regular spreading value of 0.5. SCM-USUC (Singh, 2009) is now trained for location estimation of actual or simulated inputs.

Simulation for location estimation: Simulation is done using the parameters as listed in Table 1. There were 427 Unique Samples (US) stored in SCM based on the drive test data for the suburban area in Cheras. Then the indoor stationary test point samples were used by GRNN (spreading 0.1) to match with the stored unique drive test data. The simulation of SCM-US will generate the location estimation errors and also list the undefined samples. Finally, the undefined samples' Cell ID and RSS are added to the original drive test data samples, where the simulation of SCM USUC (GRNN with spreading 0.5) will be conducted in order to estimate the location of the new indoor stationary test point data (793 samples).

RESULTS AND DISCUSSION

Before SCM-USUC, was performed, it was necessary to determine the undefined collection. Therefore, the simulation to estimate the location of the 177 samples of the known indoor test point with the stored drive test data samples was necessary. GRNN was able to match 161 samples, where the 67% error (location accuracy) was at 226 m, 95% error was at 431 m and the maximum error was at 1307 m. Since GRNN was set to low spreading value (0.1), with the purpose to determine the robustness of the drive test samples in collecting housing areas' data (which is normally away from the routes being collected), would generate some unmatched samples. There were 16 samples that are unable to be estimated by GRNN. The cause of the 16 unmatched samples, were programmatically determined, where from the drive test samples (427 samples), 3 samples' information were uncollected (missing) from the housing area where the indoor stationary test point is located. For examples, the 3 samples that are obtained from the indoor stationary samples are as following:

- Cell ID 8862 with RSS of -99 dBm
- Cell ID 8871 with RSS of -95 dBm
- Cell ID 8857 with RSS of -95 dBm

These 3 samples that had caused the undefined matching by GRNN for the 16 estimated indoor samples simulated earlier would then be added to the original 427 samples. The hybrid 430 samples (427 drive test samples and 3 undefined collected samples) would now be stored for the SCM-USUC. With new 793 indoor test point samples, collected at the same stationary point the next day, would be used as input to GRNN (spreading of 0.5), in order to estimate the location of the samples. Table 2 shows the CDF results for location estimation accuracy of 793 samples using SCM in the suburban area with spreading value of 0.5. The same 793 samples were used to estimate location accuracy when matching with SCM-US (427 drive test samples) and matching with SCM –USUC (430 hybrid samples). There was only one sample's location that was unable to be detected by GRNN out of the 793 samples. This was due to the phone inability to detect measurement at that time frame and caused an erroneous positive RSS value reading. GRNN would not be able to match this positive RSS value to all the other negative RSS values stored by SCM. From the result, SCM-USUC definitely improves the 50 and 67% of location estimation accuracy compared to SCM-US for the suburban area. With, SCM-USUC the 67% of location accuracy has been significantly improved from 379 m to 107.4 m and this 67% performance is also better than techniques that uses two GSM base stations attenuation difference in urban Taipei (Lin *et al*., 2004), where the 67% error is at 190 m.

Table 2: CDF for location estimation

CDF	50% (m)	67% (m)	95% (m)	Max error (m)
SCM-USUC	107.4	107.4	379	597
SCM-US	379	379	380	597

Even though, SCM-USUC is not suitable to be used for FCC Emergency 911 search (67% at 100 m and 95% at 300 m) but it should be used by UIPS for LBS or NBS (navigation based services). More indoor test point location should be placed between Route 2 and also Route 3 to represent the housing areas within those routes or that are further away from the drive test routes. This will assist to obtain some more unique stationary samples that could be added to the drive test samples in order to make the hybrid samples more robust in representing the suburban area. Storing too much of irrelevant SCM-USUC samples could also cause competition for GRNN's criteria to match and approximate the location of the input samples being evaluated. Therefore, it is important to strategically plan the indoor samples pertaining to each Route, where the purpose is to store quality samples (small number of unique samples) that could represent the bigger suburban area.

CONCLUSION

In this research, only three samples of undefined collection were required to improve the 50 and 67% of location estimation accuracy in the suburban area. SCM-USUC has provided accuracy within 107.4 m at 67% of the experiment, compared to using SCM-UC which provided accuracy within 379 at 67% of the CDF. In this research the data simulated for location estimation and the data stored for SCM-USUC, is based on 3G/GPRS data service. It should be noted that the samples used for estimation during voice or data service should only be matched with stored data with similar type of services. For example, data service samples should not be compared with stored data from test call of voice services samples collected during drive test or from indoor stationary voice samples.

Working cooperation with local Telcos would be advantageous in the event that network optimization and changes that are done by them would be known and, therefore new set of data could be updated for SCM-USUC by performing drive test for the areas that have been optimized.

For further studies, the new technique of SCM-USUC will be integrated as part of UIPS's LDT collections. It will be further researched, when an actual input for location search is requested, UIPS will check for LAC and Cell ID before finalizing which SCM techniques such as LEAN, US or USUC to utilize in urban or suburban areas' categorization as stored in UIPS's database. In the event, where the area's information are not stored, other LDTs such as timing technique will be utilized to estimate the location of the mobile user with voice or data service.

ACKNOWLEDGMENT

This research is supported by UKM's Center of Research Institute and by the Ministry of Science Technology and Innovation, Malaysia under the Science Fund Grant 01-01-02-SF0344 and UKM-MI-OUP-2009.

REFERENCES

Ahonen, S. and P. Eskelinen, 2003. Mobile terminal location for UMTS. IEEE Aerosp. Electron. Syst. Maga., 18: 23-27.

Demuth, H., M. Beale and M. Hagan, 2008. Neural Network Toolbox 5 User's Guide. The Math Works, Inc., Natick, MA.

Kupper, A., 2005. Location-Based Services: Fundamentals and Operation. 1st Edn., John Wiley and Sons, England, ISBN: 978-0-470-09231-6.

Laitinen, H., J. Lahteenmaki and T. Nordstrom, 2001. Database correlation method for GSM location. Proceedings of the IEEE 53rd Vehicular Technology Conference, March 6-9, 2001, Piscataway, NJ., USA., pp: 2504-2508.

Lin, D.B., R.T. Juang and H.P. Lin, 2004. Robust mobile location estimation based on signal attenuation for cellular communication systems. Electr. Lett., 40: 1594-1595.

Singh, K. and M. Ismail, 2005. OTDOA location determining technology for universal intelligent positioning system (UIPS) implementation in Malaysia. Proceedings of 13th IEEE International Conference on Networks and 7th IEEE Malaysia International Conference on Communications, November 16-18, 2005, Kuala Lumpur, pp: 1057-1061.

Singh, K., M. Ismail and K. Jumari, 2008a. A new technique using signal correlation of one node b to estimate mobile location. Int. J. Comput. Sci. Network Security, 8: 133-139.

Singh, K., M. Ismail, K. Jumari, M. Abdullah and K. Mat, 2008b. Development of universal intelligent positioning system techniques in universal mobile telecommunications system networks. J. Appl. Sci., 8: 2412-2419.

Singh, K., 2009. Location determination techniques using timing and signal correlation for universal intelligent positioning system in cellular networks. Ph.D. Thesis, Universiti Kebangsaan Malaysia, Malaysia.

Zhu, J. and G.D. Durgin, 2005. Indoor/outdoor location of cellular handsets based on received signal strength. Electr. Lett., 41: 24-26.