Author(s): Mondal, Riaz; Turkka, Jussi; Ristaniemi, Tapani

Title: An efficient cluster-based outdoor user positioning using LTE and WLAN signal strengths

Year: 2015

Version:

An Efficient Cluster-Based Outdoor User Positioning Using LTE and WLAN Signal Strengths

Riaz Uddin Mondal[1], Jussi Turkka[2], Tapani Ristaniemi[1]
[1]Department of Mathematical Information Technology, University of Jyvaskyla, Jyvaskyla, Finland
[2]Magister Solutions Ltd, Tampere, Finland
riaz.u.mondal@student.jyu.fi, jussi.turkka@magister.fi, tapani.ristaniemi@jyu.fi

Abstract— **In this paper we propose a novel cluster-based RF fingerprinting method for outdoor user-equipment (UE) positioning using both LTE and WLAN signals. It uses a simple cost effective agglomerative hierarchical clustering with Davies-Bouldin criterion to select the optimal cluster number. The positioning method does not require training signature formation prior to UE position estimation phase. It is capable of reducing the search space for clustering operation by using LTE cell-ID searching criteria. This enables the method to estimate UE positioning in short time with less computational expense. To validate the cluster-based positioning real-time field measurements were collected using readily available cellular mobile handset equipped with Nemo Handy software. Output results of the proposed method were compared with a single grid-cell layout based RF fingerprinting method. Simulation results show that if a single LTE and six WLAN signal strengths are used then the proposed method can improve positioning accuracy of 35% over the grid-based RF fingerprinting.**

Keywords-component; LTE cell-ID; Grid-based RF fingerprinting; Hierarchical Clustering; Minimization of Drive Tests.

I. INTRODUCTION

Location is a vital component in consumer services like social media, search, advertising and navigation. For authorities, mobile location is mandatory for emergency-call location, and can also be used for road-traffic management and machine-to-machine purposes. GPS-based consumer navigation devices have reached mass market status, benefits are undeniable: it is ubiquitous and always available, high-accuracy positioning [1]. However, GPS has two major drawbacks: the signals broadcast by the satellites are too weak to be received indoors in places such as shopping malls, and in dense urban environments not enough satellites are visible to obtain positioning fixes in a reasonable time. The popularity of IEEE 802.11 infrastructures, their low deployment cost, and the advantages of using them for both communication and positioning, make them an attractive choice. Therefore, authors in [2] have proposed a portable positioning system, that utilizes both GPS and Wi-Fi-based pattern matching

methods to estimate the position [2][3]. To improve this combined GPS and Wi-Fi-based Pattern Matching Method, in [4] authors proposed to assign weights to different weather conditions, determined the position of the mobile terminal by the Euclidean distance, and adjusted the weights according to the environment. Wi-Fi positioning system based on fingerprinting was evaluated in the Sydney CBD area where Wi-Fi APs are densely deployed and test results show that it works well for outdoor localization with errors in the tens of meters [4]. Also in [5] authors have carried out experimental analysis for outdoor fingerprinting system, implemented over the WLAN and demonstrated that it is feasible to perform outdoor positioning with reasonable accuracy using 802.11-based positioning. A three-phase methodology (measurement, calibration and estimation) for locating mobile stations (MS) in an indoor environment using wireless technology was proposed in [6] where combination of fingerprint and cluster based positioning system was developed to overcome the problem of the relative effect of doors and walls on signal strength and the system is independent of the hardware technology manufacturer. A new algorithm was proposed in [7] for enhancing the performance of adaptive enhanced cell-ID (AECID) fingerprint positioning in LTE, where clustering was employed to increase the accuracy of the polygon computation scheme of the AECID algorithm. The basic positioning method in most cellular communication systems is the cell-identity (cell-ID) method which has the advantage of short response time and thus it fulfills the time to first fix (TTFF) requirement for E-911 emergency positioning in the North American market which is specified to be below 30s. This method is applicable in all situations where there is cellular coverage.

One major requirement of RF fingerprint based positioning is to create and maintain the big correlation database in order to update the training fingerprints with surrounding structural and environmental changes. Operators usually conduct extensive and expensive periodical drive test campaigns to fulfill this requirement. The operational expenditure (OPEX) associated with traditional drive tests can be removed by a feature introduced in 3GPP Release 10, known as Minimization of Drive Tests (MDT) which enables operators to utilize users' equipment to collect radio measurements and associated location information [8]. MDT provides a framework for gathering user reported location-aware radio

measurements from commercial mobile phones that can be used for creating and maintaining such training database. This procedure allows operators to collect radio measurements, i.e. received signal strength and quality, with UE location information and a time stamp [9].

In our previous work grid-cell based RF fingerprinting (GRFFP) has shown good positioning accuracy in dense urban scenario using MDT samples obtained from a dynamic LTE system simulator [10][11]. To improve the user equipment (UE) positioning accuracy using grid-based RF fingerprinting, weighted Kullback-Leibler Divergence based overlapping grid-cell layout method was proposed in [12]. However GRFFP delivers good positioning when two requirements are fulfilled: (i) training signatures need to be updated in regular interval of time, (ii) an optimal grid-cell layout needs to be chosen for different cellular network scenario and for the amount of available training MDT samples.

In this study we propose a simple cluster-based RF fingerprinting (CRFFP) method which does not go through any training phase to estimate UE position. It uses MDT samples comprising of both LTE and Wi-Fi signals; we refer to this as *generalized MDT* (GMDT). The CRFFP takes advantage of the LTE serving cell-ID based searching technique to deliver UE positioning in short time. Here we also analyze the UE positioning accuracy with three different combinations of LTE and Wi-Fi signals and results were compared with the traditional GRFFP method.

The following section contains a brief description of the GMDT field measurements used in this study and then the conventional GRFFP method is explained. In Section III, first description of the proposed CRFFP is given, then test results obtained with GRFFP and CRFFP positioning methods are presented, and finally concluding remarks are given.

II. GRID-BASED RF FINGERPRINTING

A. Generalized MDT

According to 3GPP specifications MDT enables the operation, administration, and maintenance (OAM) system to collect radio measurements from the UE, together with location information if available when the measurements are taken [8]. Here we propose a GMDT that is capable of collecting Wi-Fi signal strengths along with the LTE and UE location information. There are two main reasons behind this: (i) RF fingerprinting gives very good positioning accuracy using Wi-Fi signal strengths in outdoors (especially in dense urban areas), (ii) to decrease the search space in CRFFP positioning which also shortens the operation time. In order to create the GMDT database we have used Samsung Galaxy S3 (LTE capable) which was installed with a handheld drive test software application- Nemo Handy. This application is very suitable for performing measurements both outdoors and indoor spaces while the device being simultaneously used as a regular mobile phone [13]. LTE reference signal received power (RSRP) and WLAN received signal strength indicator (RSSI) measurements were recorded from a residential urban area in Tampere, Finland during September

2014 as shown in Figure 1. Two measurement campaigns were done for 800 and 1800 MHz LTE bands, in the 1800 MHz case, inter-frequency measurements were also reported according to the measurement configuration provided by the network. Hence in this study we have used GMDT samples from LTE 1800 MHz measurements.
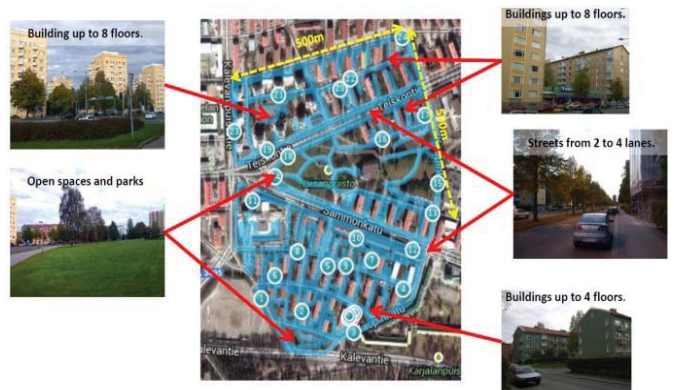

Figure 1: GMDT field measurement area in Tampere, Finland

More than 150 kilo-metres of measurements were collected by feet, bicycle and car covering approximately an area of 0.33 square kilo-metres. In all measurements, the route was repeated at least twice to ensure that enough measurement samples are collected for each grid unit.
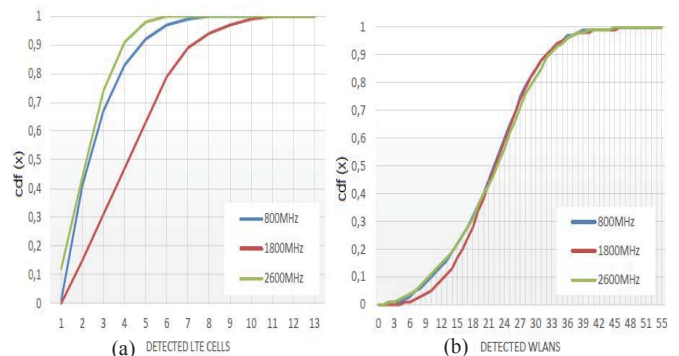

Figure 2: (a) Number of detected LTE cells and (b) WLAN APs per measurement sample.

As we can find from Figure 2 that every GMDT sample contains at least 1 serving LTE base-station (BS) signal and 98% of the samples comprises of more than 5 WLAN access points (AP). A study conducted in [14], measured the significance of Wi-Fi APs for UE position estimation where good results were obtained by limiting the Wi-Fi AP number to seven for all analyzed samples. In [15] the selection of APs was done based on the largest signal strength values recorded at each location. Hence we were motivated to use seven signals in total including both LTE BSs and WLAN APs. So, every GMDT contains the serving LTE BS ID of the recording mobile handset. Both LTE and WLAN signals were sorted in descending order of signal strength values. We have also used three different sets of GMDT samples by choosing different combinations of LTE and WLAN signals from the total database.

A set of GMDT measurement can be defined by:

$$M_j = \{s_{j,1}, s_{j,2}, \ldots, s_{j,N}\} \qquad (1)$$

where, j=1, 2 and 3 referring different GMDT set, N is the total number of measurement samples of a particular set. The nth GMDT sample of a set can be presented by a row vector:

$$s_{j,n} = \{LW_{ID}, RSS_{LW}, P_{XY}\} \qquad (2)$$

where, LW_{ID} denotes the LTE BS IDs and WLAN AP IDs, RSS_{LW} stands for the corresponding RSRP and RSSI values, and P_{XY} contains the x-y coordinates of the UE obtained from GNSS position information.

B. A Simple Grid-based RF Fingerprinting Method

Here we have used a single grid-cell layout based RF fingerprinting method by segmenting the whole geographical area of interest with square grid-cell units (GCU). We have used Euclidean distance to measure the statistical difference between training fingerprints and test samples since it has previously been used for outdoor RF fingerprinting in order to obtain good UE positioning accuracy [16].

Training Phase of GRFFP: In conventional GRFFP method multiple training signatures are formed within a single GCU [10][11]. To reduce the searching time to find the best match training signature for a test sample and also to reduce the related computational cost, a single training signature ($Train_{Sig}$) is created from all the training GMDT samples (GMDTs) that belong to a one particular GCU. The $Train_{Sig}$ formed from all the GMDT samples of ith GCU ($GMDTs_i{}^{All}$) is define as follows:

$$Train_{Sig}^i = \{TS_{ID}^{LW}, RSS_{TS}^{LW}, P_{Ref}^{XY}\} \qquad (3)$$

where, TS_{ID}^{LW} contains all unique LTE BS IDs and WLAN AP IDs obtained from $GMDTs_i{}^{All}$, RSS_{TS}^{LW} is a vector of the corresponding LTE RSRP and WLAN RSSI values, and P_{Ref}^{XY} is the reference x-y coordinate calculated from the mean values of x and y coordinates of $GMDTs_i{}^{All}$.

An example training signature creation and test phase matching of GRFFP method is illustrated in Figure 3. Here for simplicity only two GCUs are shown, the blue dots inside a

GCU represent its GMDT samples and the small red triangle depicts the corresponding reference position. We can find from Figure 3 that GCU 2 has two samples: $GMDT_2^1$ and $GMDT_2^2$ represented by two row vectors. Here black squares containing LW_{ID} (L1 indicates LTE BS ID number 1, W1 is for WLAN AP ID-1), the corresponding RSS_{LW} values (S_L^1 indicates RSRP of LTE BS ID-1, S_W^1 is RSSI of WLAN AP ID-1) are within the green squares and P_{XY} are inside the blue squares. For GCU 2 a single training signature- $Train_{Sig}^2$ has been created from $GMDT_2^1$ and $GMDT_2^2$. It has three parts: (i) TS_{ID}^{LW} contains all unique LTE and WLAN IDs. (ii) RSS_{TS}^{LW} comprises of mean RSRP and RSSI values for common LTE and WLAN IDs, otherwise the RSRP values are copied from either $GMDT_2^1$ or $GMDT_2^2$. And (iii) P_{Ref}^{XY} gives the reference x-y coordinates calculated from the mean x and y coordinates of $GMDT_2^1$ and $GMDT_2^2$.

Test Phase of GRFFP: To test a GMDT we first compare its LTE and WLAN IDs with all the training signatures available and select those signatures which meet the least matching threshold. For example in Figure 3, $Train_{Sig}^2$ contains four matching IDs: L1, L4, W2 and W5 which are common to the Test sample ($Test_{Sam}$), hence we have 57% of ID match between $Test_{Sam}$ and $Train_{Sig}^2$. Now if the minimum matching threshold is set to 50% then $Train_{Sig}^2$ is selected for distance measurement. As shown by red dotted arrows in Figure 3, only common RSRP and RSSI values are used to calculate the Euclidean distance between $Test_{Sam}$ and $Train_{Sig}^2$. A simplified Mahalanobis distance equation is used for distance calculation where the inverse covariance matrix is replace by an identity matrix:

$$d(Test_{Sam}, Train_{Sig}) = \sqrt{(\boldsymbol{u_{Te}} - \boldsymbol{u_{Tr}})^{\boldsymbol{T}} \boldsymbol{I}(\boldsymbol{u_{Te}} - \boldsymbol{u_{Tr}})} \quad (4)$$

where, $\boldsymbol{u_{Te}}$ and $\boldsymbol{u_{Tr}}$ denotes the RSRP and RSSI values of the $Test_{Sam}$ and a selected $Train_{Sig}$ respectively and $\boldsymbol{I}$ is the identity matrix. After separate calculation of all the distances between a $Test_{Sam}$ and the selected training signatures; the $Train_{Sig}$ corresponding to the smallest Euclidean distance is chosen for positioning purpose. The estimated position of that $Test_{Sam}$ is given by P_{Ref}^{XY} of the chosen $Train_{Sig}$.
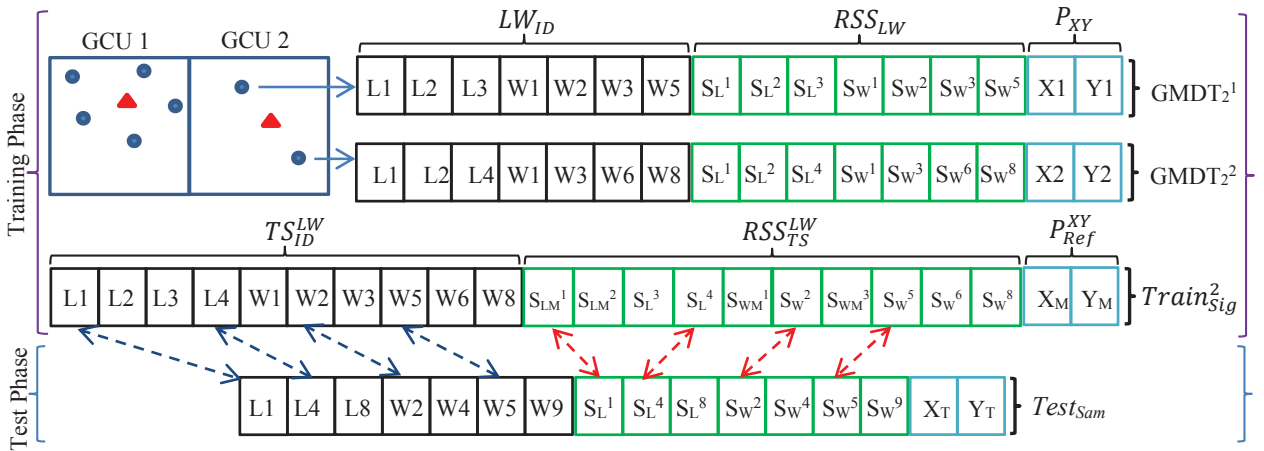

Figure 3: Training and test phases of grid-cell based RF fingerprinting

III. Cluster-based RF Fingerprinting

A. An Efficient Cluster-based Positioning Algorithm

At first the GMDT samples of the total data-base are sorted into different GMDT groups according to the serving LTE BS ID. For testing a sample the group that matches the serving LTE BS ID of that $Test_{Sam}$ is selected. From this selected group GMDT samples are selected which fulfill the least matching threshold- the matching is similar to the one described in section II (B), the only difference is that here matching is between $Test_{Sam}$ and a GMDT sample of the selected group. Now for the clustering purpose, the RSS_{TS}^{LW} values of the $Test_{Sam}$ and the selected GMDTs are put together in the same pool. We have used a simple agglomerative hierarchical clustering with Davies-Bouldin criterion to select the optimal cluster number [17]. This criterion is based on a ratio of within-cluster and between-cluster distances. The Davies-Bouldin index (DB) is defined by the follow equation:

$$DB = \frac{1}{k} \sum_{i=1}^{k} max_{j\neq i}\{D_{i,j}\} \qquad (5)$$

where, $D_{i,j}$ is the within-to-between cluster distance ratio for the ith and jth clusters. $D_{i,j}$ is given by,

$$D_{i,j} = \frac{(d_i^- + d_j^-)}{d_{i,j}} \qquad (6)$$

where, d_i^- is the average distance between each point in the ith cluster and the centroid of the ith cluster. d_j^- is the average distance between each point in the ith cluster and the centroid of the jth cluster. $d_{i,j}$ is the Euclidean distance between the centroids of the ith and jth clusters. The optimal cluster

number is obtained between 1 to 6 clusters using the smallest Davies-Bouldin index value. After multiple clusters are formed, clustering criteria (CC) is checked: the cluster which contains the $Test_{Sam}$ must have two GMDTs. If CC is met, then the cluster that contains the $Test_{Sam}$ is selected and $Test_{Sam}$ UE position is calculated from the mean x-y coordinates of all GMDTs of that cluster. If CC is not fulfilled the matching threshold is reduced and clustering is performed again in order to analyze the $Test_{Sam}$. Thus CRFF method does not need any prior training before the test phase; it utilizes the cell-ID advantages in reducing the search space thereby reduces position estimation time. Hence it offers a computationally less expensive RF fingerprinting method which can be implemented in real-time using GMDT. CRFFP positioning method is described in Figure 4.
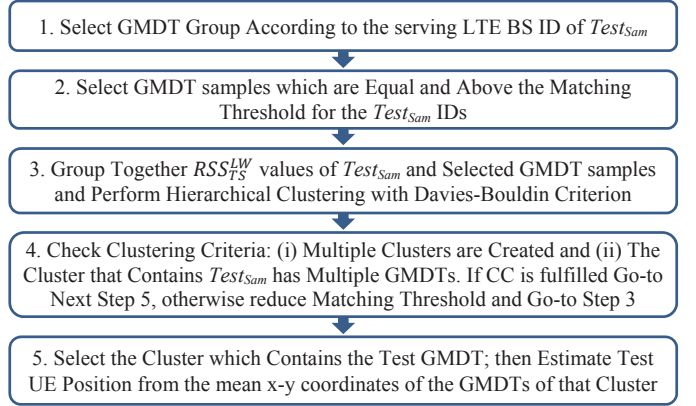

Figure 4: Block-diagram of the CRFFP Positioning Method

B. Experimental Results: Outdoor UE positioning

In the total GMDT data-set we have merged multiple samples into a single one which contain similar LTE BS ID and WLAN AP ID and were recorded from the same x-y coordinate. In order to avoid over-optimal results consecutive GMDTs have been grouped into chunks of 20 samples in sequence. Training and test data-sets were created by randomly choosing such data chunks.

TABLE I: RESULTS OF GRFFP AND CRFFP METHODS USING LTE AND WLAN SIGNALS

LTE and WLAN Combination	Matching Thres-hold	Using Total Test GMDT Samples						Using Common Test GMDTs between GRFFP and CRFFP				
		GRFFP			CRFFP			GRFFP		CRFFP		Comm. Test GMDT (%)
		68% PE (m)	95% PE (m)	Test GMDT (%)	68% PE (m)	95% PE (m)	Test GMDT (%)	68% PE (m)	95% PE (m)	68% PE (m)	95% PE (m)	
LTE BS:3 & WLAN AP:4	80%	16.91	45.41	85.86	9.40	33.74	21.17	14.21	39.77	9.40	33.75	21.15
	60%	17.65	47.78	97.68	10.39	35.84	43.60	14.02	40.35	10.40	35.85	43.59
	40%	17.88	48.99	99.96	14.16	51.17	61.77	15.59	42.57	14.16	51.17	61.77
LTE BS:2 & WLAN AP:5	80%	15.16	42.69	85.36	7.58	26.39	21.94	11.98	36.04	7.61	26.45	21.83
	60%	16.04	44.76	97.51	7.94	27.59	46.04	11.81	36.03	7.95	27.60	46.02
	40%	16.33	45.92	100	9.14	33.88	62.37	12.66	38.32	9.14	33.88	62.37
LTE BS:1 & WLAN AP:6	80%	14.29	40.31	85.43	7.33	22.80	24.59	11.47	33.69	7.36	22.81	24.38
	60%	15.09	42.54	97.02	7.24	20.56	51.35	10.60	31.76	7.24	20.58	51.32
	40%	15.35	43.42	99.97	7.80	24.34	69.21	11.17	33.89	7.80	24.34	69.21

In the simulations we have used 23080 training GMDTs and 2565 samples were tested. The GRFFP method uses a 10m-by-10m grid-cell layout which was chosen from several square grid-cell layouts according to the delivered positioning accuracy. Tenfold cross-validation method was used to obtain positioning results for both GRFFP and CRFFP methods with three different GMDT data-sets as shown in Table I. For the 1st GMDT data-set each sample is constructed with maximum 3 LTE BS signals (LTE_{BS}) and 4 WLAN AP signals ($WLAN_{AP}$); in 2nd GMDT data-set there are 2 LTE_{BS} and 5 $WLAN_{AP}$; and the 3rd set comprises of only 1 LTE_{BS} and 6 $WLAN_{AP}$. In Table I the second column indicates the different matching threshold used in both of the methods. After that UE positioning error (PE) results (68%-ile and 95 %-ile values) along with the analyzed test sample percentages are given when both methods use all the test samples. From these results we can find that when 80% matching threshold was used the 68 %-ile and 95 %-ile PEs of CRFFP are lower than that of respective GRFFP PE values, but the GRFFP method has analyzed more test samples than CRFFP. Hence on right side of Table I we have PE results considering only those test samples which were analyzed in both methods. This help to compare the proposed method with the GRFFP in the best possible way. It is found from the simulation results with common test results that when 1st data-set was used, CRFFP has given lower PE in both 68%-ile and 95 %-ile values for 80% and 60% of matching threshold. However when the threshold is lowered to 40%, GRFFP has given better positioning in 95 %-ile than that of CRFFP. With the 2nd data-set CRFFP outperforms GRFFP in both percentile values and also for all three matching thresholds used. The best positioning accuracy given by CRFFP is with the 3rd data-set and for 80 % of threshold: it has shown 35% improvement in positioning accuracy for both the 68%-ile and 95%-iles as compared to that of the GRFFP. It is clear from the results that CRFFP offers better positioning than GRFFP when matching threshold is high and 5 or 6 WLAN signals are used. Hence in dense urban areas where multiple WLAN signals can be detected CRFFP is capable of providing good outdoor UE positioning with GMDT samples.

IV. CONCLUSION

In this paper we propose a novel cluster-based RF fingerprinting method for outdoor UE positioning which uses LTE and WLAN signals. It provides better positioning accuracy as compared to that of grid-based RF fingerprinting. The benifit of the cluster-based approach is that it uses simple clustering method and no prior training phase for estimating test UE positioning. During cluster operation it reduces the searching space by utilizing LTE cell-ID; thus delivers output result in short time. The proposed method is capable of providing good positioning by using only serving LTE BS signal and six WLAN AP signals. Hence the present research outcome suggests that the next MDT functionality should include WLAN signals into consideration; which would benefit cellular operators to develop cost-effective solutions for developing real-time positioning systems.

References

[1] T. S. Rappaport, J. H. Reed, and D. Woerner, "Position location using wireless communications on highways of the future," *IEEE Communications Magazine*, vol. 34, no. 10, pp. 33–41, 1996.

[2] Son H. C., J. G. Lee and G. I. Jee, "Mobile station location using hybrid GPS and a wireless network," The 57th IEEE Semiannual Vehicular Technology Conference 2003, VTC 2003-Spring, vol. 4, pp. 2716 – 2720, April 2003.

[3] Sheng-Cheng Yeh, Wu-Hsiao Hsu, Ching-Hui Chen and Rung-Huei Liang," A Context-aware Blogging and Learning Service in Mobile IPv6 Networks", IEEE Wireless Communications and Networking Conference, March 2007.

[4] Li B., I. J. Quader, A. G. Dempster, "On outdoor positioning with Wi-Fi," *Journal of Global Positioning Systems*, Vol. 7, No. 1, pp. 18-26, 2008.

[5] Liu X., S. Zhang, J. Quan, X. Lin, "The Experimental Analysis of Outdoor Positioning System Based on Fingerprint Approach," 12th IEEE International Conference on Communication Technology (ICCT), 2010.

[6] Mengual L., O. Marbán, S. Eibe, "Clustering-based location in wireless networks," *Expert Systems with Applications,* 37, pp. 6165–6175, 2010.

[7] Wigren, T., "Clustering and Polygon Merging Algorithms for Fingerprinting Positioning in LTE," 5th International Conference on Signal Processing and Communication Systems (ICSPCS), December 2011.

[8] Johansson J., Hapsari W.A, Kelley S. and Bodog G., "Minimization of drive tests in 3GPP release 11", *in IEEE Communications Magazine*, Vol. 50, No. 11, pp. 36-43, November 2012.

[9] Wuri A. et al., "Minimization of Drive Tests Solution in 3GPP", *in IEEE Communications Magazine*, Vol. 50, No. 6, June 2012.

[10] Mondal R., Turkka J., Ristaniemi T. and Henttonen T., "Performance Evaluation of MDT RF Fingerprinting Framework", in *Proc. of 7th International Conference on Mobile Computing and Ubiquitous Networking (ICMU2014)*, Singapore, January 2014.

[11] Mondal R., Turkka J., Ristaniemi T. and Henttonen T, "Positioning in Heterogeneous Small Cell Networks using MDT RF Fingerprints", in *Proc. of the First IEEE International Black Sea Conference on Communications and Networking*, Batumi, Georgia, July 2013.

[12] Mondal, R. U., J. Turkka, T. Ristaniemi, "An Efficient Grid-based RF Fingerprint Positioning Algorithm for User Location Estimation in Heterogeneous Small Cell Networks", International Conference on Localization and GNSS (ICL-GNSS), Helsinki, Finland, June 2014.

[13]Web-link:http://www.anite.com/businesses/network-testing/products/nemo-handy-world%E2%80%99s-most-widely-used-handheld-drive-test-tool#.VT_jNSGeDRY

[14] E. Laitinen, E. S. Lohan, J. Talvitie and S. Shrestha, "Access point significance measures in WLAN-based location", In proc of. Workshop on Positioning, Navigation and Communication, March 2012.

[15] M. Youssef, A. Agrawala, and A. U. Shankar, "WLAN location determination via clustering and probability distributions," in *Pervasive Computing and Communications*, pp. 143–150, 2003.

[16] Kim Jae-H., K. SikMin,2 and W. Y. Yeo, "A Design of Irregular Grid Map for Large-Scale Wi-Fi LAN Fingerprint Positioning Systems," *The Scientific World Journal*, Vol 2014, Article ID 203419, 2014.

[17] Davies, D. L., and D. W. Bouldin, "A Cluster Separation Measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-1, No. 2, pp. 224–227, 1979.