k- NEAREST NEIGHBOR ALGORITHM FOR IMPROVING ACCURACY IN CLUTTER BASED LOCATION ESTIMATION OF WIRELESS NODES

Muhammad. A[1,2], Mazliham.M.S[3], Patrice Boursier[2,4], Shahrulniza.M[5]

[1] Institute of Research and Postgraguate Studies, UniKL(BMI), Kuala Lumpur, Malaysia,
[2]Laboratoire L3i Université de La Rochelle, La Rochelle France,
[3]Malaysia France Institute (UniKL MFI), Bandar Baru Bangi, Malaysia,
[4]Faculty of Computer Science & IT Universiti Malaya, Kuala Lumpur, Malaysia,
[5]Malaysian Institute of Information Technology (UniKL MIIT), Kuala Lumpur Malaysia
[1,2]muhammad.unikl@gmail.com, {[3]mazliham,[5]shahrulniza}@unikl.edu.my , [2]patrice.boursier@univ-lr.fr, [4]patrice@um.edu.my

ABSTRACT

This research is focusing on the precise location estimation of mobile node by using k - nearest neighbor algorithm (k-NN). It is based on our previous research findings in which we divided the geographical area into thirteen clutters/terrains based on the behavior of radio waves. We calculated the point-to-point distance from the antennas to mobile node by using receive signal strength and available signal strength information. A C# prototype was developed by using WiFi (IEEE 802.11 b/g standard) to record data points in different clutters at every 2 meter distance. We derived Clutter based Enhanced Error Rate Table (CERT) for precision. Although CERT minimizes errors due to the atmospheric considerations but in highly attenuated clutters the error rate was still high. The k-NN algorithm is used to minimize the error ranging from 1-3 meters to 0.4- 1.2 meter only, depending on the clutter we are dealing with. Current research is divided into three steps. First we calculate the P2P distance in all different clutters by using available signal strength and receive signal strength at five different time intervals ($t_0 – t_4$). In step 2 we construct three triangles at random by using the data gathered in step 1 and calculate mean value of predicted locations. Finally based on locations calculated in step 2 we apply the k-NN algorithm to minimize error in the estimated location. Results show that the k-NN can produce from 217% - 289% better results compare to the famous triangulation method. The purpose of this research is to reduce errors in order to achieve estimated position near to accurate.

Keywords: ***k-NN, Available Signal Strength, Receive Signal Strength, Location Estimation, Triangulation, Clutters/terrains, CERT.***

1.0 INTRODUCTION

Location estimation of mobile/wireless node is becoming a very famous research area from past few decades. It is not only limited for the Global Positioning System (GPS) [1], vehicle tracking system or other application of GPS [2]. It is also used in Mobile Computing [3], Wireless Local Area Network (WLAN) [4], robot tracking [5], VANET [6], Wireless Sensor Network (WSN) [7] etc. Because of the rapid growth in the wireless services, advancement in the mobile services and architecture and in the usage of handheld devices, the hand phone originating calls are also increasing very rapidly [8]. Recent study shows that almost 50% of rescue and response calls are originated by the cell phones [9]. Therefore precise location estimation in mobile/wireless devices is becoming key information for crime investigation agencies, rescue services and disaster management authorities [3] in order to reach the receiver. On the other hand location estimation in GSM or similar networks is not very accurate. The use of cell Identifier (CI) and Location Area Identifier (LAI) for the positioning of cell phone always comes with error [10].

Location estimation are divided into four categories [11, 12], which are statistical techniques, geometrical techniques, satellite based techniques and mapping techniques. Each of these techniques has advantages and disadvantages. In order to achieve accuracy sometime researchers combine two or more techniques [13]. Radio waves are the only parameter used for the location estimation, but geographical and atmospheric attenuations can distort the radio waves. The attenuated signals are used for location estimation leads towards the inaccurate estimation of this position. This is the area which is not extensively considered by researchers. Behavior of radio waves can be categorized as [14, 15]

- Reflection
- Refraction
- Diffraction

- Scattering
- Absorption

In our previous research work [16, 17], we proposed thirteen different terrains based on the geographical areas and the above factors. These terrains are divided into four environment groups which are [16, 17]

Urban Terrains/Clutters
- Low Dense Urban (LDU)
- Medium Dense Urban (MDU)
- High Dense Urban (HDU)

Rural Terrains/Clutters
- Low Dense Rural (LDR)
- Medium Dense Rural (MDR)
- High Dense Rural (HDR)

Plantation
- Agriculture/Field (A/F)
- Low Forest/ Plantation (LF/P)
- Dense Forest (DF)

Highways & Water Bodies
- Highway/Motorway (H/M)
- River/Lake (R/L)
- Sea

(We propose desert as a 13^{th} category in the terrain division but has not been considered in this research because of the unavailability of desert in the Malaysian peninsula).

Triangulation is very famous technique for location estimation. By constructing three triangles we can have GPS like scenario which leads towards the accurate position. If we built triangulation method by considering clutters/terrains atmospheric behavior then the results will be more accurate because of the atmospheric considerations. Current research is focusing on the implementation of k-NN algorithm for the precise location estimation by using clutter based data points and its comparison with the triangulation based calculated position. The proposed architecture is divided into three steps. In step one, we record clutter based ASS and RSS values. Based on ASS and RSS we then calculate P2P distance between antenna and the mobile node at every 2 meter distance with the repetition of experiment from t_0- t_4. In step two we then construct three triangles and calculate three locations (L_1, L_2, L_3). We then took average from all three locations. Finally we use k-NN algorithm on fifteen locations which were found from triangulation (L_{10} , L_{11}, L_{12}, L_{13}, L_{14}, L_{20}, L_{21}, L_{22}, L_{23}, L_{24}, L_{30}, L_{31}, L_{32}, L_{33}, L_{34}) as

$$(t_0 - t_4) \times (L_1, L_2, L_3) = 15$$

This research article is divided into four sections. Section 2 is dedicated for literature review, k nearest neighbor brief description and discussion review is also added in this section. Section 3 is focusing on the proposed architecture, data collection, experimental work and result analysis. Finally result comparison and conclusion are added in section 4.

2.0 LITERATURE REVIEW

This research paper discuss about k-NN based precise location estimation while considering the clutters/terrains affects. Therefore literature survey is divided in to four sections. Section 2.1 is focusing the description of K- Nearest Neighbor algorithm. Section 2.2 is focusing on the terrain/clutter based location estimation and the terrain impairments. Section 2.3 is focusing on the precise location estimation. Finally discussion and review is added in section 2.4.

2.1 K-Nearest Neighbor Algorithm

The k-Nearest Neighbor algorithm (k-NN) is one of the easiest and simplest machine learning algorithms [18]. On the other hand it is also known as “lazy algorithm” as it is slow and “eager algorithm” [19] as it uses all data elements for training. Its selection criteria are based on closest set of data from the point of focus. The value of k varies based on the set of data elements available for training. Each data elements behave as a central point until frequency of occurrence is obtained for all data set. Weights are further multiply with frequency to calculate weighted mean. Weighted mean leads towards the estimated position from the data set.

$$K\text{-weighted mean} = [\sum fw / \sum f] + 1$$

where ***f*** is frequency of occurrence and ***w*** is the weight of each data element.

We use k-NN because it is simple and consider all data sets which help to calculate precise location of a mobile node.

2.2 Terrains/Clutters Discussion

In [16, 17] we considered terrains error rate for location estimation. We divided the region in to thirteen categories based on the geographical area. WiFi (IEEE 802.11b/g) prototype was used to collect real time data of different terrain after repeating the experiment 5 times at each point of every terrain. By comparing the result they came up with three zones, which explained the signal strength and distance relation in each terrain. Finally we proposed an Error Rate Table (ERT) [16] and CERT [17] which helped to calculate error in each terrain/clutter. Results showed that error can be reduced up to 3m in low attenuation terrains.

In [20], authors divide the projected geographical area in eight terrains/clutters for the deployment of the GSM network architecture in a specific Indian terrain. Categorization of available terrains were low density urban, village/low density vegetation, medium density vegetation, high density vegetation, agriculture, open/quasi open area, water bodies and river/sea. Authors introduced the concept of clutter loss which they named as CL. Based on the Receive Signal Strength (RSS) measurement they proposed a tower location prediction for the region. Based on the path losses simulation results they concluded that the proposed tower prediction method will better work in the remote Indian area (Manipur).

In [21], authors considered terrains/clutters impairments for the purpose of power management of mobile nodes. They proposed Power Management Algorithm (PMA) based on the clutter based location estimation. Authors considered clutters/terrains impairment knowledge to decide that how much power could be required for the cell phone in order to maintain communication. In the first phase they proposed eight clutters then they used Google earth facility for the mapping of clutters. Finally based on the clutters knowledge they proposed power management algorithm for the purpose to use the mobile phone battery power more efficiently. The simulation showed that the PMA could save battery power up to 40% if clutter knowledge was used efficiently.

In [22], authors studied the forest terrain. They studied the propagation losses which were due to antenna height, antenna gain, depolarization, humidity effect and few other factors. They discussed that the propagation losses can increase or decrease due to the above in forest terrain. They also discussed that the external effects such as rain can cause unexpected losses in the communication links.

2.3 Precise Location Estimation Discussion

In resembling MDS [23], authors proposed a method resembling **M**ulti-**D**imensional **S**caling for the tracing of GSM node. They used subsequent signal strength measurements to different base stations for the purpose. As their research was focused on vehicular network therefore they used direction and velocity of vehicle. Based on the testing they concluded that the average deviation in the location of a node is 60 m, whereas a maximum deviation in velocity is 10m/s. Based on the testing researchers concluded that resembling MDS was applicable as a decision support for the assignment of channels in macro and micro cells in hierarchical networks.

In [24], author focused on statistical location estimation techniques. He used propagation prediction model for the precise location estimation. Author used Receive Signal Strength (RSS) and Available Signal Strength (ASS) as signal parameters. Propagation delay was considered as a random variable. This variable parameter was statistically dependent on the receiver location, the transmitter location and the propagation environment (terrains/clutters information). Author discussed that statistical location estimation approaches had certain type of flexibilities. This flexibility approach allows the fusion of different types of measurement results such as Receive Signal Strength and the timing advance. Author concluded if signal propagation environment differs significantly from the ideal condition then the angle measurement and distance will be unreliable.

In predicted and corrected location estimation [13], authors combined the advantages of geometric and statistical location estimation techniques. Authors divided their research in four steps. In step I (estimation) they used Receive Signal Strength and the Available Signal Strength as a primary parameter for distance calculation. After estimation they used variance for the filtering of calculated location points. In step III (prediction) they used Bayesian decision theory for error correction. Finally in fusion they combined two statistical techniques to achieve accuracy in estimated location. Authors used Kalman Filter's prediction and correction cycle for the purpose. Based on their test results they claimed that in an indoor environment their approach can achieve the accuracy level from 1.5 m to 1m.

2.4 Discussion and Review

The transmitted signal (Tx) and the received signal (Rx) are two of the most important parameters used for location prediction of wireless node. Tx is used to calculate Available Signal Strength (ASS) and Rx is used to calculate Receive Signal Strength (RSS). These ASS and RSS are used to calculate the distance between the sending node and the receiving node. As radio signal propagates in free space therefore attenuated factors such as reflection, refraction, diffraction, scattering and absorption play their role depending on the terrain to weaken the signal. When these attenuated parameters are used to calculate distance, they introduce error in estimated location. On the other hand the most famous triangulation method is also not very accurate for location prediction. When error of triangulation and atmospheric error are combined they multiply error. In this research we are using k-NN rule by considering clutter errors to predict location which is near to accurate.

3.0 PROPOSED ARCHITECTURE, EXPERIMENTAL WORK AND RESULTS ANALYSIS

k-NN algorithm is famously know as " eager algorithm" as it trail all available data elements. This nature of k-NN helps to minimize error in location prediction of a mobile node. We use three triangulated location with five different times (L_{10}- L_{14}, L_{20} – L_{24}, L_{30} – L_{34}). All 15 locations are used by the k-NN rule to estimate the best possible location. Result comparison shows that k-NN is from 217% - 289% better than famous triangulation method if we also consider clutter information.

Proposed architecture is divided into three steps. In step one, we collect the terrain based Available Signal Strength (ASS) and Receive Signal Strength (RSS) at five different time intervals (t_0 to t_4). By using the ASS and RSS the clutter based distance is calculated (refer Table 1) [25]. In step two, we construct triangles to calculate estimated location of a mobile node. Triangulation results are further used for precise location estimation by using averaging and k-NN. In averaging we simply calculate the average of the locations estimated by using triangulation. As we are considering clutter error rate therefore the error rate is lower as compare to the results of other researchers [17, 26]. Finally in step three k- Nearest Neighbor rule is used to estimate precise location. As we have three triangles and five times therefore every clutter is producing 15 estimated locations. k-NN process all estimated positions and select most appropriate position as shown in fig.

3.1 Step 1: Clutter Based ASS and RSS and P2P distance Calculation

Data collection from all different clutters/terrains at t_0-t_4 with the regular interval of 2 meters is presented in table 1 [25]. Corresponding value in front of distance is signal strength in percentage.

Table1: Distance and signal strength calculation in all available clutters/terrains [25]

Distance in meters (t_0-t_4)	Urban Clutters/Terrains			Rural Clutters/Terrains			Plantation Clutters/Terrains			Highway & Water Bodies Clutters/ Terrains		
	(Signal Strength in %)											
	LDU	***MDU***	***HDU***	***LDR***	***MDR***	***HDR***	***A/F***	***LF/P***	***DF***	***R/L***	***SEA***	***H/M***
10	**100.0**	**100.00**	**96.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**92.00**	**100.00**	**92.0**	**96.00**
12	96.24	94.96	92.24	96.64	96.08	95.52	97.92	97.84	88.96	97.68	88.2	92.96
14	92.56	90.24	88.56	93.36	92.32	91.28	95.68	95.36	85.84	95.12	84.5	89.84
16	88.96	85.84	84.96	90.16	88.72	87.28	93.28	92.56	82.64	92.32	80.9	86.64
18	85.44	81.76	81.44	87.04	85.28	83.52	90.72	89.44	79.36	89.28	77.4	83.36
20	**82.00**	**78.00**	**78.00**	**84.00**	**82.00**	**80.00**	**88.00**	**86.00**	**76.00**	**86.00**	**74.0**	**80.00**
22	78.64	74.56	74.64	81.04	78.88	76.72	85.12	82.24	72.56	82.48	70.6	76.56
24	75.36	71.44	71.36	78.16	75.92	73.68	82.08	78.16	69.04	78.72	67.3	73.04
26	72.16	68.64	68.16	75.36	73.12	70.88	78.88	73.76	65.44	74.72	64.1	69.44
28	69.04	66.16	65.04	72.64	70.48	68.32	75.52	69.04	61.76	70.48	64.0	65.76
30	**66.00**	**64.00**	**62.00**	**70.00**	**68.00**	**66.00**	**72.00**	**64.00**	**58.00**	**66.00**	**58.0**	**62.00**
32	64.38	62.33	61.62	69.21	66.67	65.04	71.30	62.62	50.06	64.65	51.0	60.33
34	62.99	60.67	59.55	68.25	65.75	64.47	70.47	61.41	46.26	63.25	47.4	58.67
36	61.81	59.04	56.89	66.01	63.22	62.99	68.57	60.27	44.72	61.66	45.5	57.04
38	60.82	57.48	54.27	63.10	61.95	60.35	67.28	59.14	44.22	59.89	44.6	55.48
40	**60.00**	**56.00**	**52.00**	**62.00**	**60.00**	**58.00**	**66.00**	**58.00**	**44.00**	**58.00**	**44.0**	**54.00**
42	57.50	54.62	50.19	60.02	56.86	56.12	63.92	56.83	43.68	56.10	43.4	52.62
44	55.54	53.35	48.82	58.58	54.66	54.69	62.12	55.64	43.09	54.28	42.7	51.35
46	54.07	52.18	47.73	57.54	53.27	53.63	60.57	54.43	42.22	52.64	41.8	50.18
48	52.95	51.07	46.84	56.74	52.47	52.78	59.22	53.22	41.15	51.20	40.9	49.07
50	**52.00**	**50.00**	**46.00**	**56.00**	**52.00**	**52.00**	**58.00**	**52.00**	**40.00**	**50.00**	**40.0**	**48.00**
52	51.08	48.93	45.10	55.19	51.61	51.16	56.84	50.79	38.88	49.01	39.0	46.93
54	50.07	47.83	44.07	54.21	51.10	50.16	55.58	49.58	37.88	48.18	38.2	45.83
56	48.90	46.65	42.88	53.01	50.36	48.96	54.49	48.38	37.06	47.45	37.4	44.65
58	47.53	45.38	41.51	51.59	49.32	47.57	53.27	47.19	36.44	46.75	36.6	43.38
60	**46.00**	**44.00**	**40.00**	**50.00**	**48.00**	**46.00**	**52.00**	**46.00**	**36.00**	**46.00**	**36.0**	**42.00**
62	44.34	42.51	38.39	48.30	46.45	44.32	50.72	44.81	35.68	45.14	35.3	40.51
64	42.64	40.94	36.73	46.57	44.78	42.61	49.45	43.62	35.42	44.13	34.6	38.94
66	40.96	39.29	35.08	44.90	43.07	40.94	48.22	42.42	35.11	42.93	33.8	37.29
68	39.40	37.63	33.49	43.36	41.45	39.38	47.06	41.21	34.66	41.54	32.9	35.63
70	**38.00**	**36.00**	**32.00**	**42.00**	**40.00**	**38.00**	**46.00**	**40.00**	**34.00**	**40.00**	**32.0**	**34.00**
72	36.82	34.45	30.62	40.86	38.77	36.83	45.04	38.79	33.05	38.35	30.9	32.45
74	35.85	33.04	29.35	39.93	37.80	35.87	44.18	37.58	31.76	36.64	29.7	31.04
76	35.09	31.81	28.18	39.18	37.05	35.12	43.40	36.38	30.16	34.97	28.4	29.81
78	34.50	30.79	27.08	38.56	36.48	34.52	42.69	35.18	28.22	33.40	27.2	28.79
80	**34.00**	**30.00**	**26.00**	**38.00**	**36.00**	**34.00**	**42.00**	**34.00**	**26.00**	**32.00**	**26.0**	**28.00**
82	33.50	29.42	24.91	37.43	35.52	33.49	41.31	32.83	23.59	30.82	24.9	27.42
84	32.93	29.02	23.76	36.77	34.95	32.90	40.58	31.65	21.07	29.87	23.9	27.02
86	32.19	28.72	22.56	35.99	34.20	32.16	39.79	30.47	18.56	29.14	23.1	26.72
88	31.22	28.42	21.29	35.06	33.22	31.20	38.92	29.26	16.17	28.55	22.5	26.42
90	**30.00**	**28.00**	**20.00**	**34.00**	**32.00**	**30.00**	**38.00**	**28.00**	**14.00**	**28.00**	**22.0**	**26.00**
92	28.54	27.32	18.75	32.88	30.61	28.56	37.05	26.66	12.13	27.33	21.5	25.32
94	26.89	26.26	17.63	31.80	29.15	26.94	36.11	25.22	10.61	26.36	21.0	24.26
96	25.15	24.70	16.75	30.88	27.77	25.21	35.25	23.64	9.45	24.91	20.3	22.70
98	23.47	22.60	16.18	30.26	26.67	23.52	34.53	21.91	8.61	22.81	19.3	20.60
100	**22.00**	**20.00**	**16.00**	**30.00**	**26.00**	**22.00**	**34.00**	**20.00**	**8.00**	**20.00**	**18.0**	**18.00**
102	20.86	19.11	15.21	29.09	25.33	20.77	33.66	17.94	7.51	16.54	16.2	17.41
104	20.12	18.35	14.07	28.30	24.93	19.92	32.42	15.77	6.98	12.72	14.2	16.05
106	19.67	17.43	12.98	27.11	23.95	19.39	32.07	13.60	6.28	9.19	12.3	15.43
108	19.20	16.89	12.43	25.55	22.87	18.94	31.16	11.59	5.29	7.03	11.1	14.67
110	**18.00**	**16.00**	**12.00**	**24.00**	**22.00**	**18.00**	**30.00**	**10.00**	**4.00**	**8.00**	**12.0**	**14.00**

Distance in meters (t_0-t_4)	**Urban Clutters/Terrains**			**Rural Clutters/Terrains**			**Plantation Clutters/Terrains**			**Highway & Water Bodies Clutters/ Terrains**		
	(Signal Strength in %)											
	LDU	***MDU***	***HDU***	***LDR***	***MDR***	***HDR***	***A/F***	***LF/P***	***DF***	***R/L***	***SEA***	***H/M***
112	14.47	12.77	8.94	21.71	19.58	15.32	29.36	9.36	3.83	6.38	11.4	11.49
114	11.61	9.98	6.49	19.59	17.10	12.73	28.64	8.61	3.49	4.99	10.6	9.23
116	9.30	7.62	4.57	17.63	14.63	10.28	27.84	7.78	3.05	3.81	9.85	7.23
118	7.46	5.63	3.10	15.77	12.25	8.03	26.96	6.90	2.53	2.82	8.93	5.49
120	**6.00**	**4.00**	**2.00**	**14.00**	**10.00**	**6.00**	**26.00**	**6.00**	**2.00**	**2.00**	**8.00**	**4.00**
122	4.84	2.69	1.21	12.30	7.93	4.23	24.96	5.11	1.48	1.34	7.08	2.76
124	3.91	1.66	0.67	10.65	6.07	2.75	23.84	4.24	1.00	0.83	6.20	1.75
126	3.16	0.90	0.31	9.09	4.46	1.55	22.64	3.43	0.58	0.45	5.38	0.96
128	2.53	0.35	0.11	7.50	3.10	0.63	21.36	2.68	0.25	0.18	4.65	0.39
130	**2.00**	**0.00**	**0.00**	**6.00**	**2.00**	**0.00**	**20.00**	**2.00**	**0.00**	**0.00**	**4.00**	**0.00**
132	1.53	0.00	0.00	4.57	1.16	0.00	18.56	1.41	0.00	0.00	3.45	0.00
134	1.09	0.00	0.00	3.23	0.57	0.00	17.04	0.92	0.00	0.00	2.98	0.00
136	0.69	0.00	0.00	1.99	0.20	0.00	15.44	0.52	0.00	0.00	2.60	0.00
138	0.33	0.00	0.00	0.91	0.03	0.00	13.76	0.21	0.00	0.00	2.28	0.00
140	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**12.00**	**0.00**	**0.00**	**0.00**	**2.00**	**0.00**
142	0.00	0.00	0.00	0.00	0.00	0.00	10.16	0.00	0.00	0.00	1.73	0.00
144	0.00	0.00	0.00	0.00	0.00	0.00	8.24	0.00	0.00	0.00	1.45	0.00
146	0.00	0.00	0.00	0.00	0.00	0.00	6.24	0.00	0.00	0.00	1.09	0.00
148	0.00	0.00	0.00	0.00	0.00	0.00	4.16	0.00	0.00	0.00	0.63	0.00
150	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**2.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**

a. External affects like humidity, heavy rain and bright sunlight are not considered in this specific research.

b. All clutters are considered in the tropical region of Malaysian peninsula

3.2 Step 2: Implementation of Triangulation

By taking distance at random we construct three triangles in order to fulfill triangulation method's basic needs and calculate 3 locations at five different time intervals. Same experiment is repeated for all clutters/terrains.

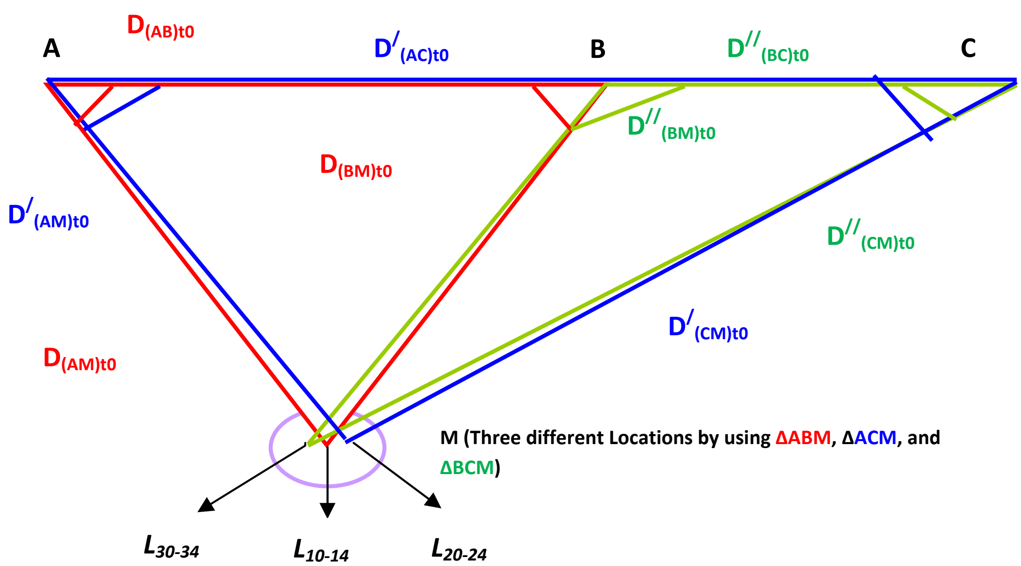


Fig.1: Triangulation implementation for selecting a precise position

where first digit in location (L) is either 1,2 or 3 and second digit is ranging 0-4.

First digit: 1 is representing ΔABM, 2 is representing ΔACM, 3 is representing ΔBCM and
Second digit: Number 0 -4 are representing data collection at different timestamp

Therefore L_{10} means data collection by using ΔABM at t_0, similarly L_{24} means data collection by using ΔACM at t_4.

As there are 15 calculated locations in triangulation therefore we calculate mean value in order to get a precise location.

$$Loc\ M_{to} = [Loc\ M_{(\Delta ABM)t0} + Loc\ M_{(\Delta ACM)t0} + Loc\ M_{(\Delta BCM)t0}] \ / \quad 3$$

$$Loc\ M_{t1} = [Loc\ M_{(\Delta ABM)t1} + Loc\ M_{(\Delta ACM)t1} + Loc\ M_{(\Delta BCM)t1}] \ / \quad 3$$

$$Loc\ M_{t2} = [Loc\ M_{(\Delta ABM)t2} + Loc\ M_{(\Delta ACM)t2} + Loc\ M_{(\Delta BCM)t2}] \ / \quad 3$$

$$Loc\ M_{t3} = [Loc\ M_{(\Delta ABM)t3} + Loc\ M_{(\Delta ACM)t3} + Loc\ M_{(\Delta BCM)t3}] \ / \quad 3$$

$$Loc\ M_{t4} = [Loc\ M_{(\Delta ABM)t4} + Loc\ M_{(\Delta ACM)t4} + Loc\ M_{(\Delta BCM)t4}] \ / \quad 3$$

Finally calculate the average of all above mean values

$$Loc\ M = [Loc\ M_{to} + Loc\ M_{t1} + Loc\ M_{t2} + Loc\ M_{t3} + Loc\ M_{t4}] \ / \quad 5$$

3.3 Step 3: Implementation of k Nearest Neighbor (k-NN) Algorithm

k-NN algorithm is used to calculate best estimated point out of 15 locations. Fig 2. is representing the k-NN calculated location and the triangular position.

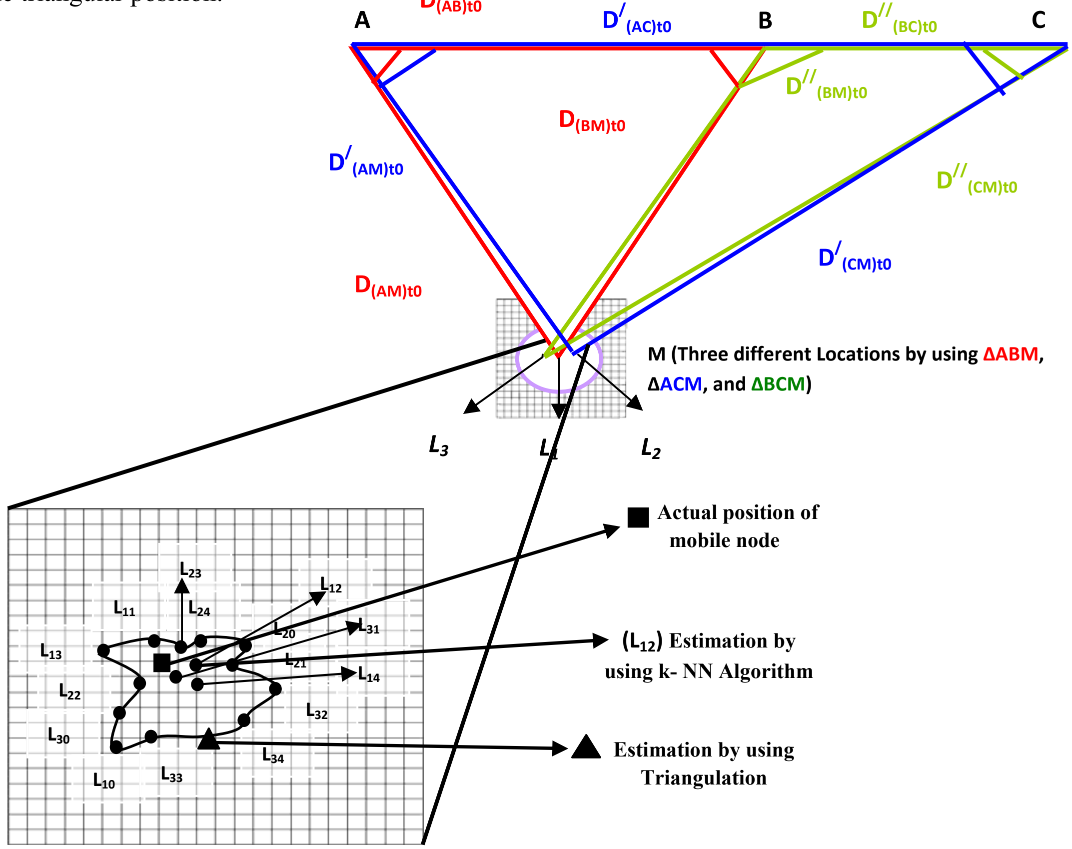


Fig. 2: k-NN implementation on triangulated estimated positions.

Table2 : k-NN repeated neighbor calculation

Estimated Location	**K^{th} nearest Neighbor (where K= 5)**	**Frequency of Occurrence**
L_{10} (t_0-t_4)	L_{33}, L_{11}, L_{32}, L_{14}, L_{31}	2
L_{11}	L_{32}, L_{33}, L_{14}, L_{34}, L_{31}	6
L_{12}	L_{31}, L_{24}, L_{22}, L_{14}, L_{23}	9
L_{13}	L_{30}, L_{32}, L_{12}, L_{14}, L_{31}	3
L_{14}	L_{11}, L_{31}, L_{32}, L_{33}, L_{34}	10
L_{20}	L_{22}, L_{23}, L_{24}, L_{12}, L_{31}	4
L_{21}	L_{23}, L_{24}, L_{34}, L_{31}, L_{20}	4
L_{22}	L_{20}, L_{12}, L_{31}, L_{23}, L_{24}	5
L_{23}	L_{24}, L_{21}, L_{20}, L_{12}, L_{31}	7
L_{24}	L_{31}, L_{12}, L_{22}, L_{23}, L_{21}	7
L_{30}	L_{13}, L_{32}, L_{22}, L_{12}, L_{14}	3
L_{31}	L_{14}, L_{12}, L_{24}, L_{11}, L_{23}	14
L_{32}	L_{13}, L_{30}, L_{31}, L_{14}, L_{12}	6
L_{33}	L_{11}, L_{14}, L_{34}, L_{10}, L_{31}	5
L_{34}	L_{21}, L_{33}, L_{11}, L_{14}, L_{31}	5

Respective frequencies of each location are as under

F_2 (L_{10})
F_3 (L_{13}, L_{30})
F_4 (L_{20}, L_{21})
F_5 (L_{22}, L_{33}, L_{34})
F_6 (L_{11}, L_{32})
F_7 (L_{23}, L_{24})
F_9 (L_{12})
F_{10} (L_{14})
F_{14} (L_{31})

$$\text{K Weighted Mean} = \frac{(2 \times 1) + (3 \times 2) + (4 \times 2) + (5 \times 3) + (6 \times 2) + (7 \times 2) + (9 \times 1) + (10 \times 1) + (14 \times 1)}{15} + 1$$

K Weighted Mean = 7

As the frequency of occurrence of L_{12}, L_{14}, L_{23}, L_{24}, L_{31} are greater than or equal to 7 therefore all five locations are selected as a best possible location.

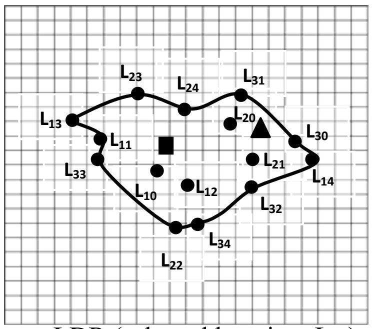
Table 3: k –NN repeated neighbor calculation for single neighbor

Estimated Location	**K^{th} nearest Neighbor**	**Frequency of Occurrence**
L_{12}	L_{31}, L_{24}	5
L_{14}	L_{31}, L_{12}	2
L_{23}	L_{24}, L_{12}	2
L_{24}	L_{12}, L_{23}	3
L_{31}	L_{14}, L_{12}	3

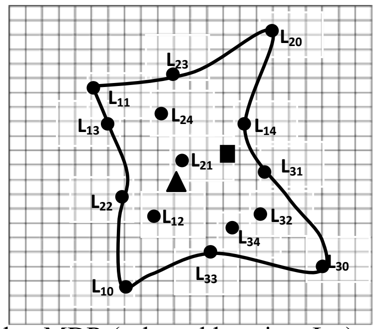
Frequencies of occurrence are

F_2 (L_{14}, L_{23})
F_3 (L_{24}, L_{31})
F_5 (L_{12})

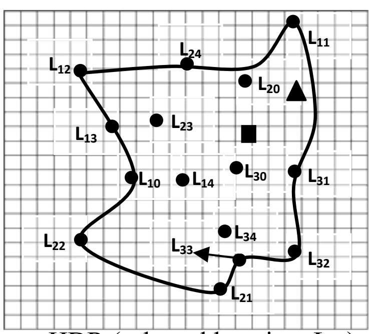
As KWM = 4 is calculated, therefore locations having frequency of occurrence of 4 and above will be selected. In the table above only $\mathbf{L_{12}}$ is a calculated location with frequency of occurrence of 5. By using k-NN algorithm L_{12} is the best estimated position for the Location prediction. k-NN implantation of other clutters/terrains is as under.



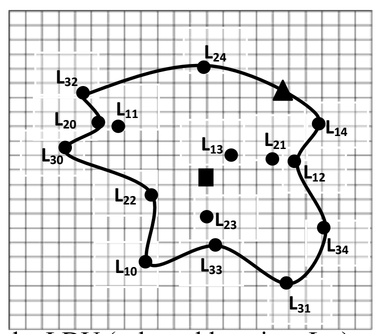
a. LDR (selected location: L_{24})



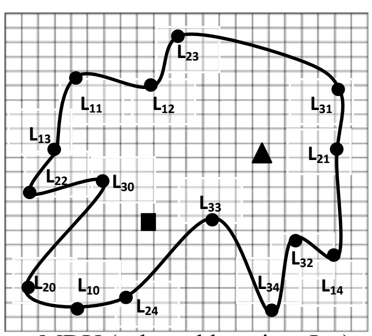
b. MDR (selected location: L_{31})



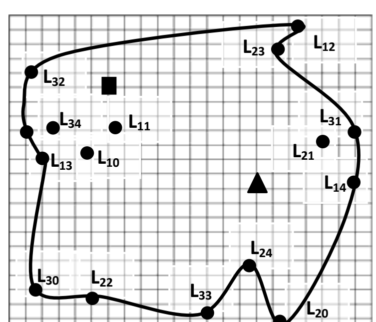
c. HDR (selected location: L_{14})



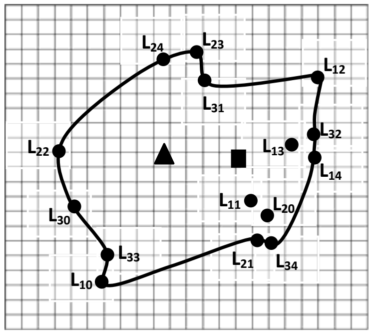
d. LDU (selected location: L_{13})



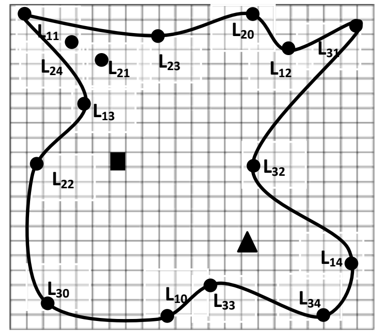
e. MDU (selected location: L_{24})



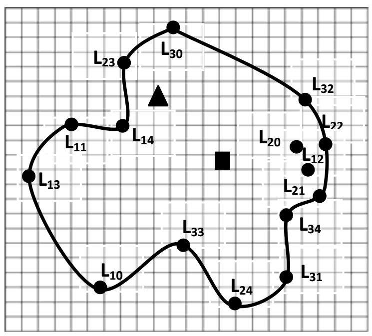
f. HDU (selected location: L_{10})



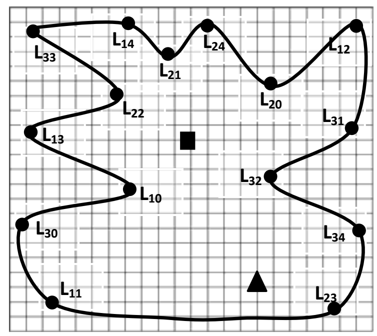
g. LF/P (selected location: L_{20})



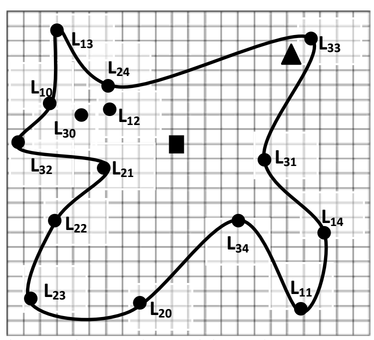
h. DF (selected location: L_{21})



i. R/L (selected location: L_{12})



j. Sea (selected location: L_{21})



k. H/M (selected location: L_{30})

Fig.3: k-NN implementation in all available clutter/terrains.

In Fig 3 triangle is representing calculated location by using triangulation method and square is representing the actual position of a mobile node. k-NN results in all available clutters are mentioned at the bottom with the clutter type. Highly attenuated clutters such as High Dense Urban (HDU), Dense Forest (DF), Sea and Highway/Motorway (H/M) still have more errors because of atmospheric impairments but results are still improved from 217% - 289% from triangulation location estimation. As in above mentioned terrains signal may reflect, refract, diffract, scatter or absorb especially in sea terrain absorption plays a vital role. Tracking architecture of k-NN Algorithm is presented in fig. 4.

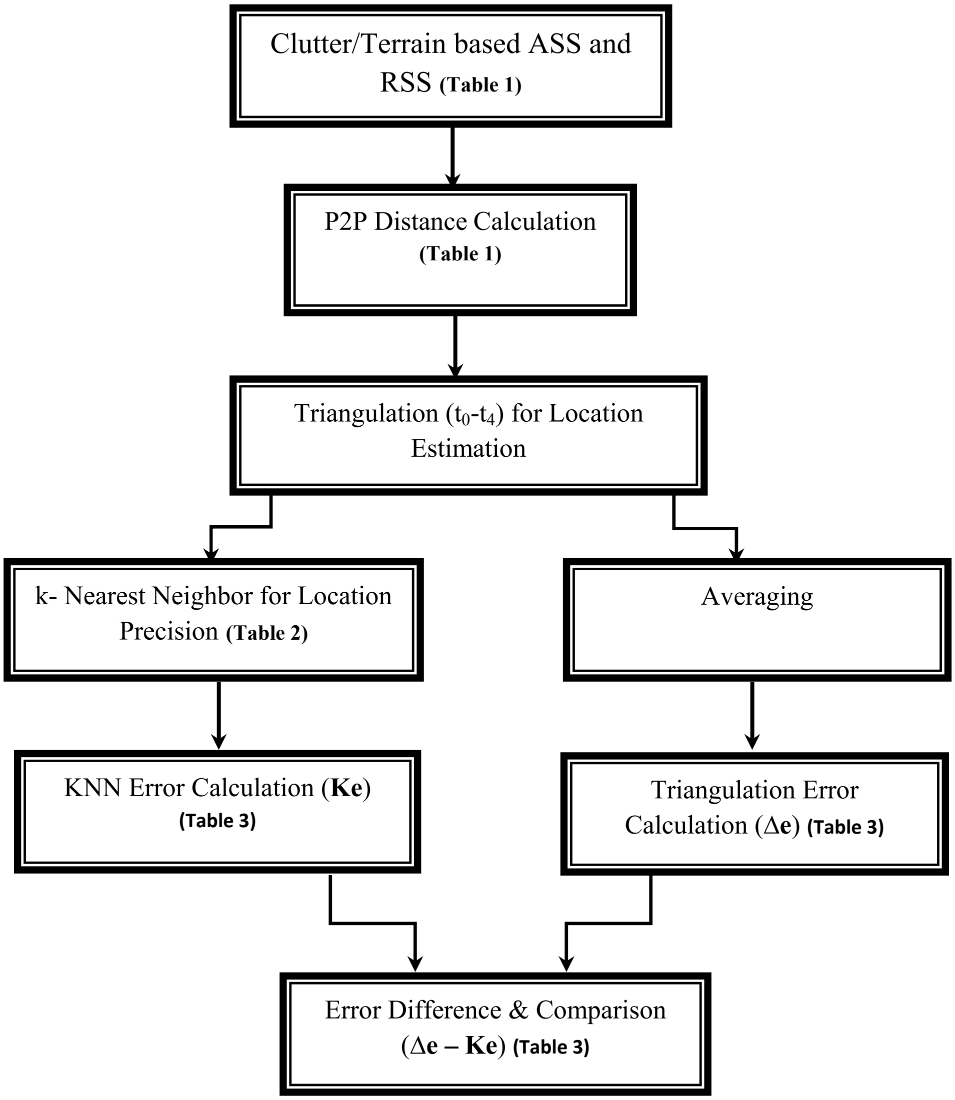


Fig 4: Tracking Architecture of k- NN implementation for precise location estimation

Fig 4 is presenting the step by step procedure for location estimation of a mobile node by using the k-nearest neighbor algorithm.

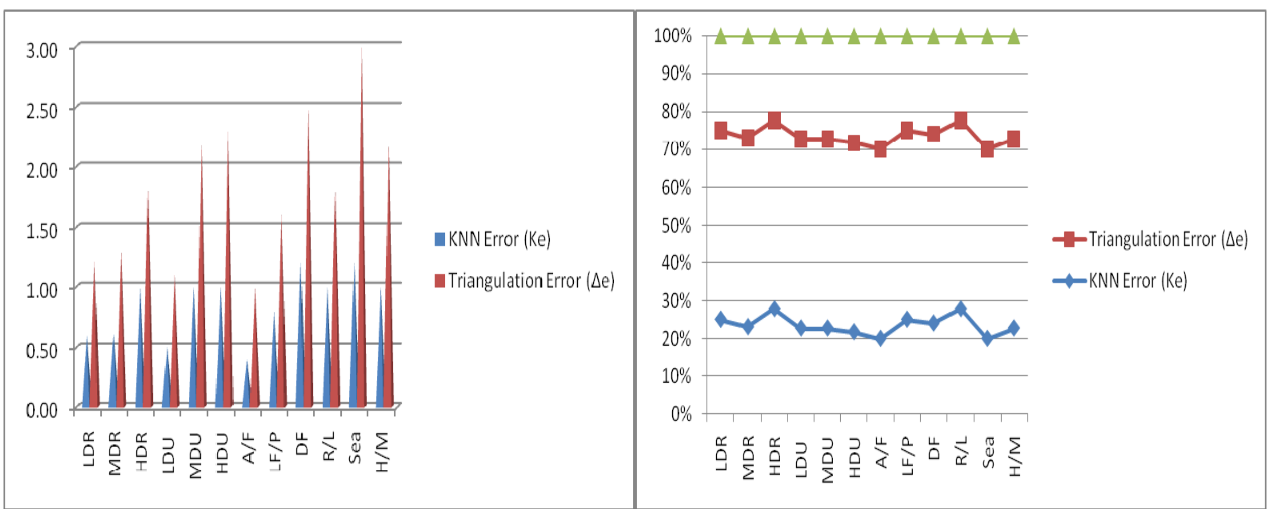
4.0 RESULTS COMPARISON AND CONCLUSION

Although k-NN produces better result compare with triangulation method, but it still have some errors as recorded in table 4. It is also recorded that clutters with high impairments have more error as compare to less attenuated clutters.

Table 4: k-NN error (Ke) and triangulation error (Δe) with the actual location.

Terrain/Clutter Type	KNN Error (Ke)	Triangulation Error (Δe)	Δe - Ke
LDR	0.6	1.59	0.99
MDR	0.8	2.02	1.22
HDR	1.0	2.89	1.89
LDU	0.7	1.95	1.25
MDU	1.0	2.17	1.17
HDU	1.3	3.3	2
A/F	0.4	1.02	0.62
LF/P	0.8	1.94	1.14
DF	1.3	3.59	2.29
R/L	1.2	3.03	1.83
Sea	1.6	4.01	2.41
H/M	1.3	3.03	1.73

Although High Dense Urban (HDU), Dense Forest (DF), Sea and Highway/Motorway (H/M) still have more errors because of atmospheric impairments but results are still improved from 217% - 289% from triangulation location estimation. Fig 5(a) is presenting the error comparison between the k-NN and the triangulation and fig 5(b) is presenting the actual error in percentage which varies from clutter to clutter.



(a) (b)

Fig 5. Error rate comparison between k-NN and triangulation method

Agri/Field (A/F) is the lowest attenuated clutter as its ***Ke*** is only 20% whereas in the same clutter the Δ***e*** is 70% (refer fig 5b) similarly other clutters ***Ke*** is falling between 20% - 30% whereas Δ***e*** is ranging from 70% - 80%.

REFERENCES

[1] Elliott Kaplan and Christopher Hegarty, Understanding GPS: Principles and Applications, Second Edition ISBN: 1-58053-894-0

[2] Muruganandham , P.R.Mukesh “Real Time Web based Vehicle Tracking using GPS” World Academy of Science, Engineering and Technology 61, 2010.

[3] Muhammad. A, Mazliham M.S, Patrice Boursier, Shahrulniza. M, Jawahir Che Mustapha @ Yusuf. “Location Estimation and Power Management of Cellular Nodes for rescue operation” ICET 2009, Dec 08- 10, Kuala Lumpur, Malaysia.

[4] Teemu Roos, Petri Myllymäki, Henry Tirri, Pauli Misikangas and Juha Sievänen “A Probabilistic Approach to WLAN User Location Estimation” International Journal of Wireless Information Networks Volume 9, Number 3, 155-164, DOI: 10.1023/A:1016003126882.

[5] Elliott Kaplan and Christopher Hegarty “Understanding GPS: Principles and Applications” Second Edition ISBN 1-58053-894-0

[6] Laurendeau, C. and Barbeau, M. (2008), Insider attack attribution using signal strength-based hyperbolic location estimation. Security and Communication Networks, 1: 337–349. doi: 10.1002/sec.35

[7] Patwari, N.; Hero, A.O., III; Perkins, M.; Correal, N.S.; O'Dea, R.J.; “Relative location estimation in wireless sensor networks” Signal Processing, IEEE Transactions Aug. 2003 Volume: 51 Issue: 8 on page(8): 2137-2148 ISSN: 1053-587X

[8] http://www.guardian.co.uk/world/2008/feb /27 / nokia .mobile phones

[9] EU Institutions Press Release. (2003). “Commission Pushes for Rapid Deployment of Location Enhanced 112 Emergency Services” , DN: IP/03/1122, Brussels.

[10] Axel Küpper, Location-Based Services: Fundamentals and Operationed 2005 John Wiley & Sons, Ltd. ISBN: 0-470-09231-9 (HB)

[11] Mohamed Khalaf-Allah, “A Novel GPS-free Method for Mobile Unit Global Positioning in Outdoor Wireless Environments” Wireless Personal Communications Journal, Volume 44, Number 3, February, 2008

[12] Sinan Gezici “A Survey on Wireless Position Estimation” Wireless Personal Communications: An International Journal, Volume 44, Issue 3 (Feb 2008) ISSN: 0929-6212

[13] Muhammad Alam, Mazliham Muhammad Suud, Patrice Boursier, Shahrulniza Musa, Jawahir Che Mustapha Yusuf “Predicted and Corrected Location Estimation of Mobile Nodes Based on the Combination of Kalman Filter and the Bayesian Decision Theory” Mobile Wireless Middleware, Operating Systems, and Applications, Lecture Notes of the Institute for Computer Sciences,Social Informatics and Telecommunications Engineering, 2010, Volume 48, Part 7, 313-325, DOI: 10.1007/978-3-642-17758-3_24

[14] P.K.Dalela, M.V.S.N Prasad, A.Mohan. “A New Method of Realistic GSM Network planning for Rural Indian Terrains” IJCSNS International Journal of Computer Science and Network Security, VOL.8 No.8, August 2008.

[15] Lecture notes on RF fundamentals Universitas Bina Darma http://images.ilmanzuhriyadi.

[16] Muhammad, A.; Mazliham, M.S.; Boursier, P.; Shahrulniza, M.; Mustapha, J.C., “Terrain/Clutter Based Error Calculation in Location Estimation of Wireless Nodes by using Receive Signal Strength” 2nd International Conference on Computer Technology and Development (ICCTD), 2010 Cairo, Egypt. November 2-4, 2010 ISBN: 978-1-4244-8844-5

[17] Muhammad, A.; Mazliham, M.S.; Boursier, P.; Shahrulniza, M.; Mustapha, J.C., “Clutter based Enhance Error Rate Table (CERT) for Error Correction in Location Estimation of Mobile Nodes” International Conference on Information and Computer Networks, ICICN 2011, 26-28 January 2011Guiyang, China IEEE Catalog Number: CFP1145M-PRT ISBN: 978-1-4244-9514-6

[18] http://saravananthirumuruganathan.wordpress.com/2010/05/17/a-detailed-introduction-to-k-nearest-neighbor-knn-algorithm/

[19] http://people.revoledu.com/kardi/tutorial/KNN/index.html

[20] P.K.Dalela, M.V.S.N Prasad, A.Mohan. “A New Method of Realistic GSM Network planning for Rural Indian Terrains” IJCSNS International Journal of Computer Science and Network Security, VOL.8 No.8, August 2008.

[21] Muhammad.A, Mazliham.M.S, Shahrulniza.M. “Power Management of Portable Devices by Using Clutter Based Information”. IJCSNS, International Journal of Computer Science and Network Security, VOL.9 No.4, April 2009, pp 237-244. ISSN: 1738-7906.

[22] Y. S. Meng, Y. H. Lee, B. C. Ng, “Study Of Propagation Loss Prediction In Forest Environment”, Progress In Electromagnetics Research B, Vol. 17, 117-133, 2009

[23] Martin Hellebrandt, Rudolf Mathar, and Markus Scheibenbogen, "Estimating Position and Velocity of Mobiles in a Cellular Radio Network". IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, VOL. 46, NO. 1, FEBRUARY 1997.
[24] Teemu Tonteri, M.Sc Thesis "A Statistical Modeling Approach to Location Estimation" Department of Computer Science University of Helsinki 25th May 2001.
[25] Muhammad. A, Mazliham M.S, Patrice Boursier, Shahrulniza.M "Modeling Interpolated Distance Error for Clutter Based Location Estimation of Wireless Nodes" The International Conference on Digital Information and Communication Technology and its Applications DICTAP 2011 June 21-23, DIJON, France Publisher: SpringerLink (LCNS) (accepted)
[26] Moustafa Youssef. "PinPoint: An Asynchronous Time-Based Location Determination System", ACM Proceedings of the 4th international conference on Mobile systems, applications and services. Pages: 165 – 176. Uppsala, Sweden 2006.