2012 International Conference on Indoor Positioning and Indoor Navigation, 13-15th November 2012

Positioning with Multilevel Coverage Area Models

Matti Raitoharju, Simo Ali-Löytty, Robert Piché
Tampere University of Technology
Tampere, Finland
matti.raitoharju@tut.fi

Marzieh Dashti
The University of Sheffield
Sheffield, UK
m.dashti@sheffield.ac.uk

***Abstract*—Fingerprinting techniques provide good indoor and urban user location estimates, but using them in large scale requires an enormous radio map (RM) database. To reduce the database size, we build a statistical model of the coverage area (CA) of each wireless communication node (CN) using "fingerprints" (FP), i.e. reception samples. In previous work we modeled each CA as a single ellipse, so only 5 parameters need to be stored in the RM for each CN. In this paper, we investigate the use of multiple CAs for every CN. FPs are grouped based on received signal strength (RSS) criteria and CA models are fitted to different FP groups. Different choices of RSS boundaries are examined with real data. We present a method for positioning using the proposed "multilevel coverage area radio map". The proposed method is applied on real data sets. The positioning results are compared with those of conventional single level CA positioning and a basic location fingerprint methods. The results show improvement of positioning accuracy compared with positioning with a single level CA. The improvement is due to better use of RSS level information in both the offline phase (constructing the CA radio map) and in the online phase (user positioning). The proposed multilevel CA positioning works with a much smaller RM database than the basic location fingerprint method, without degrading the positioning accuracy in indoor positioning.**

Radio map; Wireless LAN; RSS; Coverage area; Student-t distribution; Normal distribution; Fingerprint; Positioning

I. INTRODUCTION

Location fingerprinting is a well-known positioning technique that determines user's location using a database of radio signal measurements. A "fingerprint" (FP) contains the location of the user equipment (UE) and may include a set of radio characteristics records from a variety of radio networks, like received signal strength (RSS) and identifier of the transmitter e.g. identity of a communication node (CN). CN may be a radio station, a TV station, a cellular network base station, a WLAN access point or some other sensor node in a wireless network. In this work the CNs are WLAN access points. A UE may be a laptop, a mobile phone, or any other device connected to a wireless network. Location fingerprinting consists of two phases, an offline data-collecting phase and an online positioning phase. In the data-collecting phase, FPs are measured at various locations using positioning-capable UE [1–4]. The fingerprint database is processed and used to generate a radio map, which provides information about radio signal properties as a function of position. In the positioning phase, the UE samples measurements from CNs and computes user's location using the radio map [5].

Location fingerprinting takes into account the effects that buildings and environment have on radio signals. Hence, in contrast to many other positioning methods, it does not require line of sight conditions to ensure acceptable accuracy. This makes the location fingerprinting method often precise and reliable in complex environments such as indoor and urban environments. A drawback of the fingerprinting method is that, while accuracy may be good when the radio map is up-to-date, it degrades with time because the radio environment changes constantly [4, 6]. Moreover, performing extensive data collecting is needed and storing a huge database (e.g. covering an entire city or country) is costly.

To reduce the database size, we have used FP data to construct a statistical model of the coverage area of a wireless CN [5, 7]. Instead of raw FP data, the radio map consists of the parameters of the coverage areas (CA). Here, a CA means the region in the plane where signals from the CN can be received by the network user. The CAs are modeled as probability distributions whose parameters may be described using the mean and the covariance; this CA model may be visualized as an ellipse. Only five parameters have to be stored in the radio map for each CA. In the online positioning phase the CAs of the heard CNs are used to infer the position of the user.

In [5], the CA is modeled by computing a posterior distribution using Bayes' rule. The Bayesian prior models information about "typical" CAs. This information is especially important when the FP database contains only a few observations from a CN. In [5], CAs are fitted by modeling fingerprints as having a normal (Gaussian) distribution. However, the normal regression model is well known to lack robustness, i.e. outliers produce coverage areas that are too large. The Student-t distribution is an alternative to the Normal distribution that, due to its heavy tails, is better suited as a model of data that may contain outliers. In this paper, CAs are modeled using Student-t and normal distributions as explained in [7, 8].

In our previous works, only one coverage area for every CN was stored in the radio map [5, 7]. In this paper we consider the use of multilevel CAs for every CN in radio map. The FPs are classified into different sets based on their RSS level, and multilevel ellipse-shaped CAs are fitted to each set of FP data. In this paper, different criteria for classifying FPs into different sets are investigated.

A method to use the proposed so-called "multilevel coverage area radio map" for positioning is presented. The method is tested using real indoor and outdoor positioning data. Positioning results are compared with conventional single-level CA positioning and a basic location fingerprint method. The results indicate the enhancement of positioning accuracy compared to the positioning with one level CA. The proposed multilevel CA positioning requires much smaller radio maps than the basic location fingerprint method, without degrading the positioning accuracy in our indoor tests.

The remainder of this paper is organized as follow: The coverage area estimation model is presented in Section II. Section III describes the coverage area positioning. The test setup for evaluating positioning performance is described in Section IV. The positioning results using multilevel coverage areas are compared in Section V. Concluding remarks are given in Section VI.

II. Multilevel Coverage area models

A. One-level coverage area models

In this section, a method for fitting an ellipse-shaped CA to location FPs is presented. FPs are assumed to follow multivariate Student-t distribution. The method is less sensitive to outliers than existing smallest-enclosing ellipse and Normal-distribution based methods [8].

Here, each observation is modeled as a bivariate ($d = 2$) Student-t random vector z_n with mean μ, shape Σ, and ν degrees of freedom. When the degrees of freedom ν is fixed the mean and shape parameters for the multivariate Student-t distribution may be computed using Expectation-Maximization algorithm [8]:

Algorithm for modeling the CA from FPs

initialize $u_{1:N} \leftarrow$ ones
for $t = 1$ **to** T **do**
  $\mu \leftarrow \sum_n u_n z_n / \sum_n u_n$
  $\mathrm{S} \leftarrow \sum_n u_n (z_n - \mu)(z_n - \mu)^T$
  $\Sigma^{-1} \leftarrow (N + \tau - d - 1)\,(\mathrm{S} + \sigma \mathrm{I})^{-1}$
  for $n = 1$ **to** N **do**
    $u_n \leftarrow \frac{d+\nu}{\nu + (z_n-\mu)^T \Sigma^{-1} (z_n - \mu)}$
  end
end

In the algorithm τ is a weight parameter describing the ‘strength” of the prior and $\sigma = \tau r^2$ where r represents a typical range of a CA.

The relationship between covariance matrix P and the shape matrix Σ is

$$\mathrm{P} = \frac{\nu}{\nu - 2}\Sigma. \tag{1}$$

When degrees of freedom $\nu \to \infty$ the distribution approaches to normal and $\mathrm{P} = \Sigma$.

B. Multi-level coverage area models

To determine multi-level CAs of a specific CN, FPs are classified into groups and a CA ellipse is fitted to each set of FPs. Three different criteria for classifying the FPs are proposed as follows:

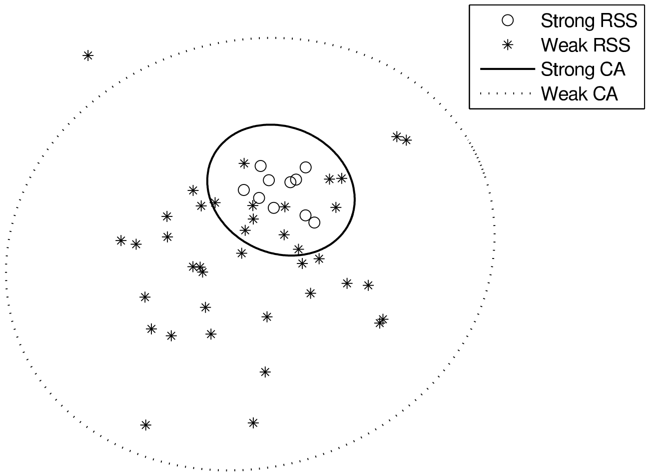

Figure 1: Two-level coverage areas

1) RSS-level: a presumed RSS threshold value is used to determine if a FP has strong signal strength.

2) n-strongest: the n fingerprints with the highest RSS values are classified to the strong set.

3) $x\%$-strongest: the $x\%$ strongest RSS-values of each FP to belong to the strong set.

We also investigate two different ways to deal with the “weak” area. In one case the RSS-values not considered to be strong are used for the weak area and in the other method all of the FPs are used to construct the ‘weak” area. Fig. 1 illustrates the basic idea of modeling two-level CAs, where the strong CA is constructed using strong FPs (circles) and weak CA using the weak FPs (asterisks).

III. Positioning

A. Positioning using CAs

Assuming a radio map containing multi-level CAs of CNs is constructed, the goal is to estimate the user's position using the radio map and information that the user receives; the identification codes of heard CNs and their RSS levels.

Let $\mu_1, \ldots, \mu_n$ be the means of the CAs of the CNs that are observed and $\mathrm{P}_1, \ldots, \mathrm{P}_n$ be the corresponding covariances. If the CAs are assumed to be independent measurements, then the best linear unbiased estimator (BLUE) of the user position is

$$x = \left(\mathrm{L}^T \mathrm{W} \mathrm{L}\right)^{-1} \mathrm{L}^T \mathrm{W} \left[\mu_1^T, \ldots, \mu_n^T\right]^T, \tag{2}$$

where

$$\mathrm{L} = [\mathrm{I}, \ldots, \mathrm{I}]^T \tag{3}$$

$$\mathrm{W} = \mathrm{diag}(\mathrm{P}_1, \ldots, \mathrm{P}_n)^{-1}. \tag{4}$$

This may be simplified into form

$$x = \left(\sum_i \mathrm{P}_i^{-1}\right)^{-1} \sum_i \mathrm{P}_i^{-1} \mu_i. \tag{5}$$

This is the same equation as derived in [5], where it was shown to be the Bayesian estimate when the coverage areas are modeled as Gaussians and the prior is uninformative.

As explained in Section II-B, the CAs were constructed using different sets of FPs. The positioning phase uses the same rules to choose which CA is used for each CN.

B. Positioning using fingerprints

As a reference method for positioning we used the weighted k-nearest neighbor method (WKNN) [4]. In WKNN a FP database is searched for k FPs that have the most similar RSS values of CNs and then the user position is computed as a weighted mean of the positions of FPs in the database. In our implementation we used $k = 5$, the similarity of all RSS values is computed using the 2-norm and the weight of a FP is proportional to the inverse of the 2-norm. If a FP in the database did not contain a CN that was in the positioning measurement it is assumed to have a weak RSS value (-105 dBm).

IV. Test setup

In our tests CA models were fitted using different parameter values. We tested all combinations of the following values of parameters: $\tau \in \{5, 10, 20\}$ and $r \in \{5, 10, 20, 40, 80\}$. Degree of freedom for Student-t was set to $\nu \in \{5, \infty\}$, a typical value for general-purpose robust regression and the normal model.

A. Indoor

These tests were done inside a university building. The CAs were fitted using 243 FPs and 331 CNs. Each CN is contained on average in 28 FPs.

Examples of two-level CA ellipses of three specific transmitters are shown in Fig. 2. The ground truth for our test route was manually marked using a laptop during the measurement session. The process of marking positions manually causes some error to the true route but it should be on the order of a meter or two.

B. Outdoor

In this test the data contained 26921 FPs collected mostly on streets in a suburban area. Fingerprints in the test route had 857 unique CN IDs that were found in the CA database. On average a CN had 57 FPs containing it. The CA models were constructed using the same rules and parameters as with the indoor data. The ground truth was determined using GPS, meaning that there is a couple of meters of error in the true route.

V. Results

Table I shows the positioning results on a route inside a university building for positioning using different rules for coverage areas. The CAs are constructed according to rules given in the table. In the positioning phase the CA corresponding to the first rule, which is true for the FP is used. The parameters shown in the table are those that had the smallest mean error for the rule. r_1 is the prior "range"

Figure 2: Three examples of two-level coverage area models fitted using normal and Student-t regression. In the first example the two strong FPs in left hand side affect strongly on normal model while the Student-t is not affected that much. The second and third examples show the difference of CAs if the strong FPs are included or excluded from the weak CA.

for the CA generated by Rule 1 and $r_{2,3}$ is for Rules 2 and 3. Mean, Median and 95% columns are the mean, median and 95% quantile of the positioning errors given in meters. The bold numbers are the smallest of each column.

Results show following

- Two CA models give better positioning performance than one
- Student-t ($\nu = 5$ always in Table I) outperform normal models with all rules
- It is better to use all FPs for the "weak" CA instead just "not strongs"
- In most of cases it was best to use small and weak prior for CAs ($\tau = 5, r = 5$)

The Student-t models gave on average 0.7 meters better accuracy than the normal models. We can also see that the use of the third CA for a CN does not improve the performance

Table I: Results for indoor positioning

CAs	Rule 1	Rule 2	Rule 3	τ	r_1	$r_{2,3}$	ν	Mean	Median	95%
1	all	-	-	20	5	5	5	11.6	9.8	23.0
2	1-strongest	all	-	5	5	5	5	10.9	10.2	18.8
2	1-strongest	not 1-strongest	-	5	5	5	5	11.0	10.5	19.1
2	3-strongest	all	-	5	5	5	5	9.7	9.6	**16.3**
2	3-strongest	not 3-strongest	-	5	5	5	5	9.9	9.9	16.5
2	5-strongest	all	-	5	5	5	5	9.5	8.9	17.4
2	5-strongest	not 5-strongest	-	5	5	5	5	9.8	9.3	18.4
2	7-strongest	all	-	5	5	5	5	10.2	9.5	23.0
2	7-strongest	not 7-strongest	-	5	5	5	5	11.4	9.9	28.4
2	10%-strongest	all	-	5	5	5	5	10.1	9.3	17.7
2	10%-strongest	not 10%-strongest	-	5	5	5	5	10.5	10.4	18.1
2	15%-strongest	all	-	5	5	5	5	9.6	9.1	18.4
2	15%-strongest	not 15%-strongest	-	5	5	5	5	10.3	10.6	18.2
2	20%-strongest	all	-	20	5	5	5	9.5	9.6	17.8
2	20%-strongest	not 20%-strongest	-	5	5	5	5	10.5	10.2	18.9
2	30%-strongest	all	-	20	5	5	5	9.9	10.0	18.8
2	30%-strongest	not 30%-strongest	-	20	5	40	5	11.0	10.0	24.3
2	40%-strongest	all	-	5	5	5	5	9.8	9.2	19.2
2	40%-strongest	not 40%-strongest	-	20	5	80	5	10.1	9.5	22.0
2	-75dBm	all	-	10	5	5	5	9.8	9.7	18.0
2	-75dBm	not -75dBm	-	5	5	5	5	10.7	10.5	19.3
2	-85dBm	all	-	5	5	10	5	9.6	9.2	19.4
2	-85dBm	not -85dBm	-	5	5	40	5	10.6	9.9	20.4
3	1-strongest	5-strongest	all	5	5	5	5	9.5	9.0	17.8
3	3-strongest	7-strongest	all	5	5	5	5	9.6	**8.7**	20.9
3	15%-strongest	30%-strongest	all	5	5	5	5	**9.3**	9.1	17.0
Fingerprinting								9.9	8.1	25.1

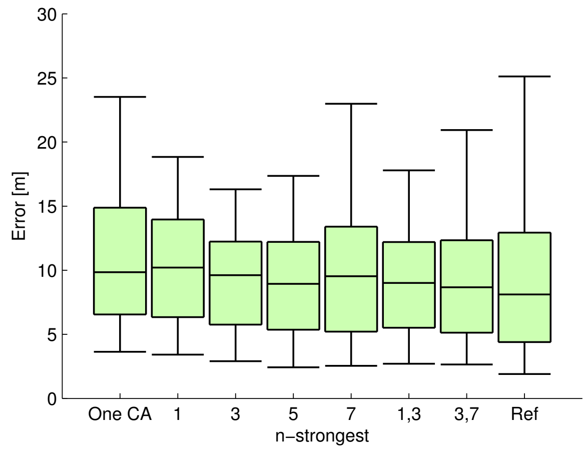

Figure 3: Indoor results with different rules for selecting the strong CAs

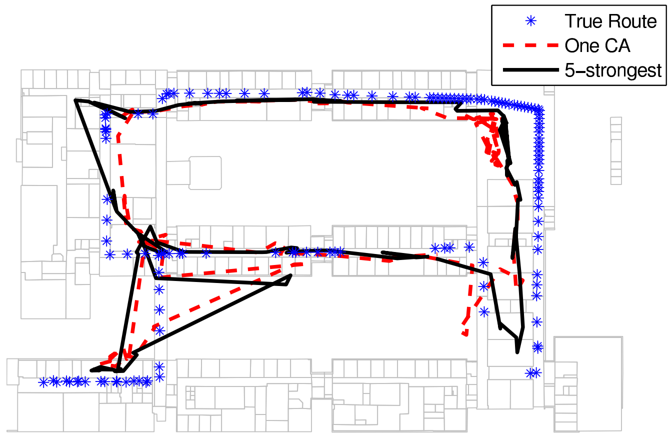

Figure 4: Comparison of routes

much and the 2- and 3-level positioning is comparable to fingerprinting in accuracy.

In Fig. 3 is a boxplot showing 5%, 25%, 50%, 75% and 95% error quantiles for different n-strongest rules in indoor positioning. In cases where two values are given in the x-axis we used three CA models. Used parameter values are $\nu = 5, r_1 = 5, r_2 = 5$ and $\tau = 5$, which seemed to the best for the indoor positioning. From this figure we can see that the median error of all the methods is almost the same, but the 75% and 95% error quantiles are smallest when using two level coverage areas with $n = 3$ or $n = 5$ and three level areas with limits 1 and 3. The reference fingerprinting method has the best 5% and 25% errors, but worse 75% and 95% error quantiles than the best multilevel models.

In Fig. 4 the routes given by our positioning algorithms are illustrated. Stars show the reference locations used in positioning. Dashed line presents the results computed using a single Student-t CA model and the solid line is the route computed using two CAs using 5-strongest rule. The 5-strongest rule is close to accuracy of best methods by all numbers in Table I.

The results for an outdoor scenario (Table II) show smaller improvement of use of the multiple coverage areas compared to the indoor tests. In Fig. 3 is a boxplot showing the error quantiles for different n-strongest rules in outdoor positioning.

Table II: Results for outdoor positioning

CAs	Rule 1	Rule 2	Rule 3	τ	r_1	$r_{2,3}$	ν	Mean	Median	95%
1	all	-	-	20	20	5	∞	51.9	45.7	103.4
2	1-strongest	all	-	20	20	20	∞	51.7	45.9	107.7
2	1-strongest	not 1-strongest	-	20	20	20	∞	52.0	46.4	107.9
2	3-strongest	all	-	20	5	10	∞	50.7	**39.5**	121.4
2	3-strongest	not 3-strongest	-	20	5	10	∞	52.0	42.0	121.9
2	5-strongest	all	-	20	10	20	∞	51.8	44.4	117.0
2	5-strongest	not 5-strongest	-	20	5	10	∞	51.5	43.7	116.9
2	7-strongest	all	-	20	5	20	∞	49.8	42.5	**96.5**
2	7-strongest	not 7-strongest	-	20	5	10	∞	**48.5**	41.7	103.1
2	10%-strongest	all	-	20	20	20	∞	51.6	45.1	111.2
2	10%-strongest	not 10%-strongest	-	20	20	20	∞	52.8	45.1	115.3
2	15%-strongest	all	-	20	10	20	∞	51.8	44.4	114.1
2	15%-strongest	not 15%-strongest	-	20	5	10	∞	51.1	42.1	113.1
2	20%-strongest	all	-	20	5	10	∞	50.7	43.7	113.2
2	20%-strongest	not 20%-strongest	-	20	5	10	∞	50.5	43.6	104.0
2	30%-strongest	all	-	20	10	20	∞	52.5	44.4	111.5
2	30%-strongest	not 30%-strongest	-	10	10	20	∞	52.7	44.8	107.5
2	40%-strongest	all	-	20	10	20	∞	51.5	43.7	102.7
2	40%-strongest	not 40%-strongest	-	20	10	40	∞	52.5	45.1	103.4
2	-75dBm	all	-	20	5	20	∞	51.0	45.7	102.7
2	-75dBm	not -75dBm	-	20	5	20	∞	52.8	46.8	111.8
2	-85dBm	all	-	20	20	20	∞	50.9	45.5	106.7
2	-85dBm	not -85dBm	-	20	20	40	∞	52.6	48.8	112.1
3	1-strongest	5-strongest	all	20	20	20	∞	52.0	45.8	110.4
3	3-strongest	7-strongest	all	20	5	10	∞	50.7	40.5	119.7
3	15%-strongest	30%-strongest	all	20	5	10	∞	52.2	46.3	120.2
Fingerprinting								43.4	39.0	85.4

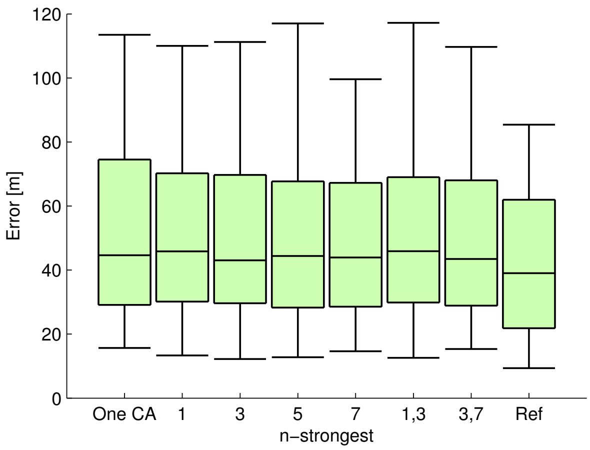

Figure 5: Outdoor results with different rules for selecting the strong CAs

Used parameter values are $\nu = \infty, r_1 = 10, r_2 = 20$ and $\tau = 20$, which seemed to be good values for the outdoor positioning. From this figure we can see that the use of multilevel coverage areas does not improve the positioning as much as in the indoor positioning case.

The normal distribution models outperform the Student-t and optimal values for τ and r are larger than in our indoor test. The mean difference between positioning error between Student-t and normal models is 2.6 meters. The positioning accuracy is somewhat worse than with the reference fingerprinting method. The reason why normal models are better than the Student-t models in our outdoor test is illustrated in Fig. 6. The dots show the locations of FPs that were used for generating the CAs with thick lines. There are a few FPs in lower right corner that are considered as outliers in the Student-t model. In the positioning phase this ellipse affects more to the estimate than the normal ellipse.

VI. Concluding Remarks

This paper examines the use of multiple CA models for a CN instead of one CA for positioning purposes. The proposed positioning method, using multilevel CA models, is compared with conventional CA positioning, using one-level CA models, and reference fingerprinting method.

In our tests with real data we got results showing that the use of multiple CA models for each CN improved the positioning results. The proposed method was tested using real indoor and outdoor positioning data. In indoor tests where the FPs covered the building well the proposed positioning method produced results that were even slightly better than the reference fingerprinting method.

Furthermore, the results show that the CAs constructed using Student-t regression provide better positioning results compared to the CAs constructed with normal regression indoors. In our test the use of three CA models did not give significant improvement compared to 2-level models.

In outdoors test the improvement was smaller and the positioning results were better for normal model. The normal model outperformed the Student-t method because the FP distribution was not uniform and some FPs that were inliers were considered as outliers in Student-t regression. To enhance the positioning accuracy of these cases a robust positioning algorithm should be used.

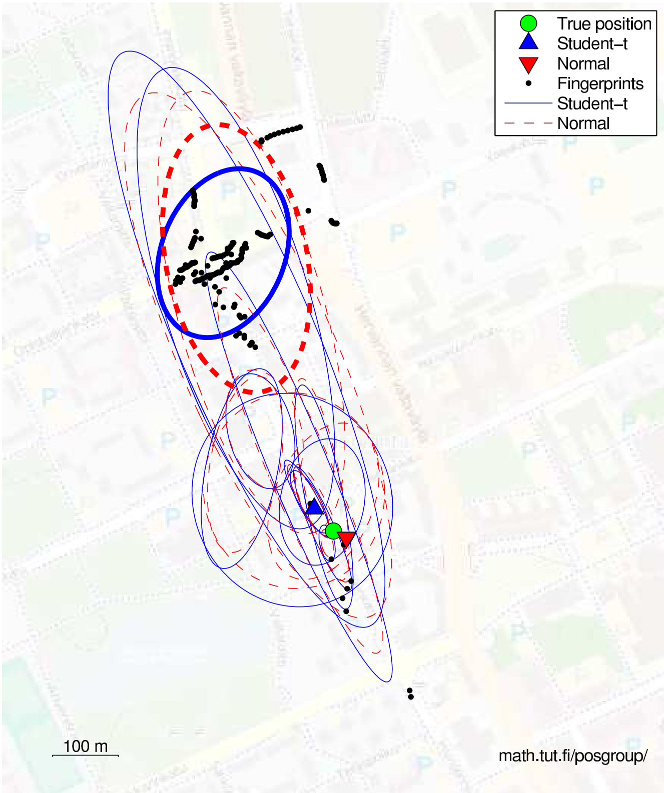

Figure 6: An outdoor positioning example

Our research shows that when building a multilevel radio map the following should be considered:

- Weak CAs should contain all FPs
- If collected data is nonuniform a larger τ and r values should be used compared to uniform data
- RSS-level rule should be avoided because it does not provide better accuracy than rules based on relative strengths of RSS values and the absolute RSS value is dependent of the UE used [9]

For the space requirements of multilevel CA models compared to fingerprinting methods we can consider our outdoor test case. There we had on average 57 FPs for each CN. For fingerprinting at least two numbers have to be stored in the FP for each CA (ID of the CA and the RSS). In our test case we have to store at least 114 numbers on average for a CN. The two level coverage areas require only 10 numbers for a CN, 4 numbers for two means and 6 for two covariance matrices. This simple calculation shows that a positioning database containing two-level CA models require lot less space than a fingerprinting database.

VII. Acknowledgment

This work was supported by Nokia, Inc.

References

[1] H. Laitinen, J. Lähteenmäki, and T. Nordström, “Database correlation method for GSM location,” in *IEEE 51st VTC*, Rhodes Greece, May 2001.

[2] S. Ahonen and H. Laitinen, “Database correlation method for UMTS location,” in *IEEE 57st VTC*, Rhodes Greece, April 2003.

[3] P. Bahl and V. N. Padmanabhan, “Radar: An in-building RF-based user location and tracking system,” in *Proceedings of the 19th International conference on Computer Communications (InfoCom). IEEE, vol. 2*, Dresden Germany, March 2000, pp. 775–784.

[4] V. Honkavirta, T. Perälä, S. Ali-Löytty, and R. Piché, “A comparative survey of WLAN location fingerprinting methods,” in *Proceedings of the 6th Workshop on Positioning, Navigation and Communication 2009 (WPNC’09)*, March 2009, pp. 243–251.

[5] L. Koski, R. Piché, V. Kaseva, S. Ali-Löytty, and M. Hännikäinen, “Positioning with coverage area estimates generated from location fingerprints,” in *Proceedings of the 7th Workshop on Positioning, Navigation and Communication 2010 (WPNC’10)*, Dresden Germany, March 2010, pp. 99–106.

[6] L. Wirola, I. Halivaara, and J. Syrjärinne, “Requirements for the next generation standardized location technology protocol for location-based services,” *Journal of Global Positioning Systems*, vol. 7, no. 2, pp. 91–103, 2008.

[7] M. Dashti, S. Ali-Löytty, L. Wirola, P. Müller, H. Nurminen, and R. Piché, “Robust Kalman filter for positioning with wireless BS coverage areas,” in *Proceedings of the 7th Workshop on Positioning, Navigation and Communication 2012 (WPNC’12)*, Dresden Germany, March 2012.

[8] R. Piché, “Robust estimation of a reception region from location fingerprints,” in *International Conference on Localization and GNSS*, Tampere, June 2011, pp. 31–35.

[9] K. Kaemarungsi, “Distribution of WLAN received signal strength indication for indoor location determination,” in *1st International Symposium on Wireless Pervasive Computing*, Jan. 2006.