Errors in Determinstic Wireless Fingerprinting Systems for Localisation

Andrew G Dempster, Binghao Li, Ishrat Quader
School of Surveying and Spatial Information Systems
University of New South Wales
Sydney 2052 Australia
a.dempster@unsw.edu.au

***Abstract*—.Fingerprinting is a technique that records vectors of received power from several transmitters, and later matches these to a new measurement to position the new user. This paper examines data used in earlier fingerprinting experiments in a WiFi network to characterize the eventual positioning errors. The implied relationship between real distance and "vector" distance between fingerprints is tested and found to be poor. However, because fingerprinting algorithms use nearest neighbour techniques, these nearby fingerprints were examined and found to be better behaved.**

I. Introduction

Fingerprinting is a technique that has successfully been used for localisation in wireless networks. Unlike almost all other radio-navigation techniques, it is not geometrical. In other words the position solution does not rely on the angle to or distance from the transmitters. Instead, it requires a survey of signal strength vectors to be made ahead of the system's use for localisation. When positioning, the user's device records its own vector(s) of signal strength and matches it against the pre-recorded database of vectors. Location is then calculated based on good matches between the new and stored vectors. Deterministic [1] and probabilistic [2, 3, 4] algorithms have been used in WiFi [5] and mobile phone [6] networks. Orientation [7, 8] can be incorporated and WiFi applications can be implemented outdoors [9]. Several companies have already implemented systems based on received signal strength techniques, including Ekahau (www.ekahau.com), Skyhook (www.skyhookwireless.com), and Aeroscout (www.aeroscout.com).

Many geometric positioning systems have been described:

- Time-of-arrival (TOA) or *trilateration* systems, which record when a signal is received and subtract from it the time of transmit to give a distance to the transmitter. GPS is an example.
- Time-difference-of-arrival (TDOA), where times of arrival from different transmitters are differenced. An example is Loran-C.
- Angle-of-arrival (AOA), where angles are measured to transmitters and *triangulation* is used to solve for position, the classic surveying problem.

In these geometrical systems, a term that relates the measurement error with the eventual positioning error can be evaluated using statistical techniques. This term is known as *Dilution of Precision* (DOP) and can be calculated for TOA [10], TDOA [11] and AOA [12] systems. It can be used in a number of ways. First, it can be used predictively, i.e. given a network of transmitters (e.g. satellites for GPS), a value for DOP can be calculated that indicates whether a sound (i.e. low error) position is likely to be able to be calculated at a particular location. Similarly, it can be used when designing a network to place transmitters in locations that can ensure good positioning over a region. Secondly, a calculation of DOP that accompanies a particular position calculation can give an indication as to how much confidence a user can have in that calculation.

There has not been very much published in the area of estimating errors in fingerprinting systems. There are incomplete descriptions of how to estimate errors when using a probabilistic algorithm in [13] and [14]. In this paper, we investigate errors in deterministic fingerprinting systems, with the ultimate aim of evaluating something akin to DOP, noting all the while that because fingerprinting does not use geometry, an exact analogy to DOP cannot be achieved.

II. Deterministic Fingerprinting Background

First, some explanatory terminology:

- Access point (AP): These are the fixed transmitters from which power levels are measured. "Access point" is a term usually used in WiFi networks, but could also apply to a mobile phone base station for instance.
- Reference Point (RP): The reference points are the known positions at which fingerprints are recorded and stored in a database for later matching with measurements made at unknown locations.

- Test Point (TP): Points of at which fingerprints were recorded. These fingerprints are then input to various localisation algorithms to calculate location, which can be compared to the known TP location.
- Received Signal Strength Indication (RSSI): The raw measurement from which the fingerprints are generated. Typically, this is a vector of signal strengths in dBm, with one element of the vector associated with each AP that can be received.

When recording the database of fingerprints that are associated with RPs, many individual RSSIs are recorded, and these can vary significantly. A typical fingerprint is the average of the recorded RSSIs. The fingerprint can also include information about the distribution, either a histogram for each AP or a more simplified parameter such as variance.

Once the database of fingerprints exists, a user device can calculate position if a fingerprint is recorded and "matched" to the database. This matching process usually consists of measuring a "distance" between the recorded fingerprint and each RP fingerprint in the database. To avoid confusion later, we will refer to this distance as the "vector distance" which has units related to dBm (as opposed to the "geometric distance" in m between the TP and an RP). Simple vector distance measures are Manhattan and Euclidean, the L_1 and L_2 norms. Other measures have been examined, such as the L_i norms for i = 0.25 to 4, and Manhattan distance seems to give best results in an indoor WiFi application [15, p60] (which is handy because it is also easiest to calculate). Once this vector distance is calculated, an interpolation algorithm is used to provide location with respect to the RPs. "Nearest neighbour" simply selects the RP with shortest vector distance. A weighted average of nearest neighbours gives improved results as does the use of interpolation algorithms such as universal kriging [16].

If the propagation environment in which the system operates is known, then absolute distance measurements can be predicted to the APs. However, the point of fingerprinting is that it does not require knowledge either of the AP location, or the characteristics of the environment. Only the consequences of that environment are measured, the RSSIs.

III. ANALYSIS

A. Measurements at Reference Points

During the survey phase, where the RP fingerprint database is being created, the RSSI measurements are averaged. In a typical survey, these measurements include subsets facing in four directions. In early work, few measurements were taken, e.g. 12 [16]. Later we used software such as NetStumbler (for PC) and a pruned version for the PDA called MiniStumbler (for PDA) (http://www.netstumbler.com) to collect the RSSI data, which allowed hundreds of measurements to be easily taken, e.g. 180 in each direction equating to one per second for 3 minutes [8, 9].

Using 180 samples gives a very good estimate of the mean, and due to the central limit theorem, that estimate has a "normal" distribution with standard deviation $\sqrt{180} = 13.4$ times as small as that of the original distribution.

There are two problems with this method, however, both relating to the propagation environment. Taking measurements in the four directions recognises the first of these: that the user's body affects the RSSI. This can clearly be seen in Figure 1, where different levels are observed for the different directions. The only differences between "directions" are antenna orientation (and the assumption of a relatively isotropic antenna means this should not have a major effect) and the relative position of the user's body, which blocks signal differently in each direction. This problem is in part addressed by treating the directional databases separately [8], so the variation of the RSSI, visible in Figure 1, is greatly reduced. The second problem is to do with ongoing variation in the environment – movement of people and furniture, opening/closing doors etc. This is an important cause of variation in RSSI, but it is not one we are considering yet. One solution may be to take the database RP readings over the course of a day or a week (obviously a far greater burden than to take the measurements "at once" over a few minutes). So, the first problem we avoid by using directional databases and the second we are leaving for the moment, leaving us with our fingerprint database of good RSSI estimates.

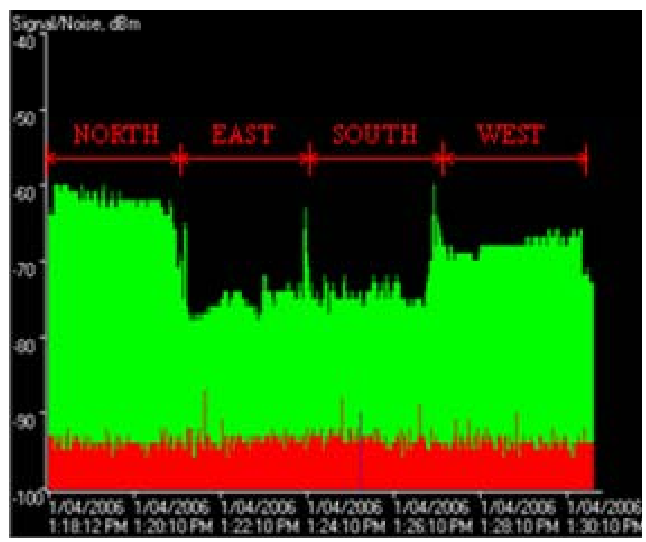

Figure 1 RSSI measurements for a single AP facing in each of 4 directions – 180 measurements per direction

Looking in more detail at the errors in the reference points, Figure 2 and Figure 3 show the distribution of variances that we measured in our earlier work [16, 8]. Making the (admittedly tenuous) assumption that power variance is the same at all RPs, the estimates of variance from these sets of samples are 13.8dBm^2 for all directions and 9.0dBm^2 for single directions. Means are taken from these distributions and used in fingerprints, as shown in Table 1.

Experiment	Power dist. variance (dBm^2)	Power samples	Power vector element variance (dBm^2)	Power vector element s.d. (dBm)
[16]	13.8	12	1.15	1.1
[8]	9.0	100+	0.09	0.3

Table 1 Measurement variances for RPs in [16] and [8]

B. Distance between Measured Power Vectors (Fingerprints)

The next step of the process, during the active phase, is where a fingerprint is taken at an unknown location. For many users, it will not be possible to take 3 minutes of measurements in a single location. The test fingerprint can thus be expected to be less accurate than the database fingerprints, i.e. its estimate of average RSSI will have greater variance.

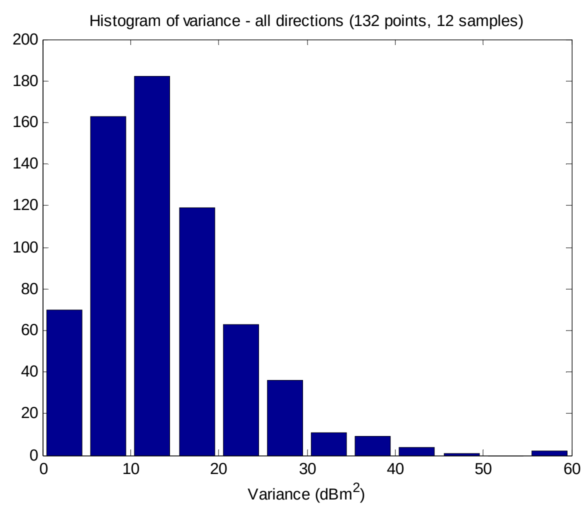

Figure 2 Histogram of all variances measured using 12 power level measurements from 5 APs at 132 separate RPs [16]

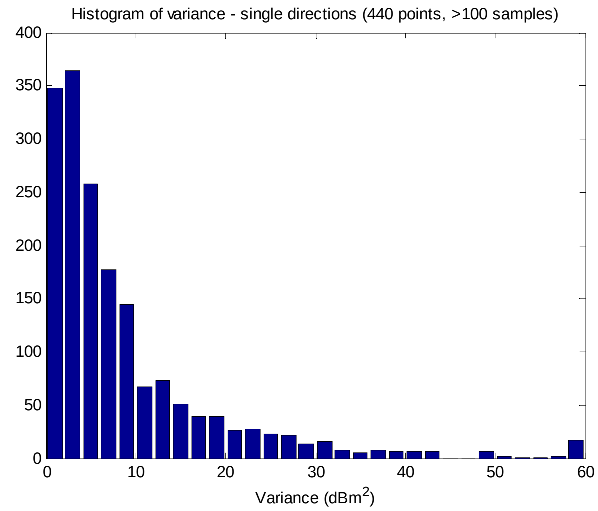

Figure 3 Histogram of all variances measured in single directions (N, S, E, W) using over 100 power level measurements from 5 APs at 110 separate RPs [8]

The Manhattan distance between a TP and RP can be written:

$$MH = \sum_{i=1}^{N} \left| RP(i) - TP(i) \right| \tag{1}$$

where N is the number of APs, i.e. the number of elements in the fingerprint vectors, RP is the RP fingerprint and TP is the TP fingerprint. If we make the simplifying assumption (not necessarily valid but it allows us to dismiss the absolute value in the sum) that the mean of $RP(i)$-$TP(i)$ is sufficiently greater than either RP or TP standard deviation, then the variance of MH can be estimated as:

$$\sigma_{MH}^2 = \sum_{i=1}^{N} \left(\sigma_{RP(i)}^2 + \sigma_{TP(i)}^2 \right) \tag{2}$$

In other words, the MH variance is the sum of the variances of all the fingerprint elements, both RP and TP. In the case of [16], these numbers were the same, but large RP sample sizes can drive down that contribution to the error. As it was, the data of Manhattan distances measured in [16] had similar variances for both RPs and TPs – about 1.1 dBm^2 (see Table 1), giving an overall variance of about 11 dBm^2 (because there are N = 5 APs), corresponding to a σ_{MH} = 3.3dBm. If many more RP measurements are made, then the distance variance will be dominated by the TP variances, and would be effectively halved.

C. Interpolation Algorithm

The next step of the process is to use an interpolation algorithm to determine the TP position. Implicitly, this relies on a relationship between the measured vector distance (e.g. the Manhattan distance discussed above) and the geometric distance. Interestingly, to the authors' knowledge, this relationship has not before been examined. There are plenty of studies into the relationship between geometric distance from an AP and the RSSI from that AP (e.g. see [17]). However, Figure 4 shows that the inferred relationship between the vector distance and the real distance is not as strong as we might like. By studying the Manhattan distance between the RP fingerprints used in [16] and comparing them with real distance, it can be seen that there is a definite trend, i.e. that they are related, and that a quadratic fit to the data seems good (a cubic fit being very similar). The sum-squared-error (SSE) of the quadratic fit gives a standard deviation of 15.7dBm. From the previous section, we expect only 3.3dBm of this to be due to variation in the power measurements at the RPs to produce the fingerprints. The remaining 12.5dBm must be accounted for in a different way: it is due to the power measurements varying much more erratically than would allow good prediction, due to signal fading and features such as walls between the RPs.

The experiment leading to Figure 4 relied on "known" locations for real distance, so it is relevant to examine the spread in the vector (Manhattan) distance as an outcome of that experiment. However, when using fingerprinting, the fingerprints and the distances between them are measured and from them, the location is inferred. It is therefore also useful to look at the spread in the real distance that occurs around the quadratic fit. As can be seen in Figure 4, the quadratic has a maximum value below the maximum value of the distribution and hence differences between measured vector distances and those predicted by the quadratic fit cannot be made. Hence the linear fit was used to indicate a spread of 4.1m.

This is significantly more than the errors noted in [16]. The reason for this is that while Figure 4 gives a good indication of how well Manhattan distance between RP fingerprints indicates real distance (not very well), the interpolation algorithms tend to match to *nearest neighbours* so behaviour of distant RPs is not relevant.

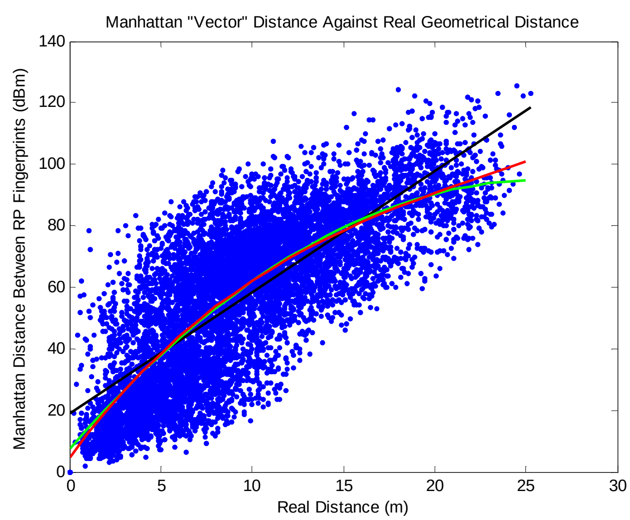

Figure 4 Distribution of Manhattan distances between 132 RPS of [16] versus the real distance between the RPs. Linear (black), quadratic (green) and cubic (red) regressions are also shown

The data of Figure 4 was analysed on a per-RP basis. The real nearest neighbour and that with the shortest Manhattan distance were found. For the 132 points in [16], only 26 (19.7%) matched. This explains why the several-nearest-neighbour algorithms work better than simple nearest neighbour. Figure 5 shows all nearest neighbour pairs (i.e. for a given RP, the nearest Manhattan distance fingerprint and the real nearest RP). Several things can be observed: the "worst mistake" made by the algorithm is less than 6 meters. Also, there is an example of two RPs 1.5 meters distant with fingerprints 80dBm apart! Numerical processing of the data shows that the error in the nearest neighbour measurement averages only 1.25m (note that this is the difference between the distance to the real nearest neighbour and the distance to the neighbour found by fingerprinting, *not* the distance between those two points), which is entirely consistent with the results in [16].

D. "Disclaimer" on Units

Throughout this discussion, units of dBm have been used. The validity of these units is not for discussion here, but it is worth making one point. Where dBm are differenced, the result should really be in dB. Where they are added, there isn't a convention. When working on dBm-based fingerprinting, it is often easier to ignore the physical underpinning of the system and simply treat the fingerprints as unitless numbers.

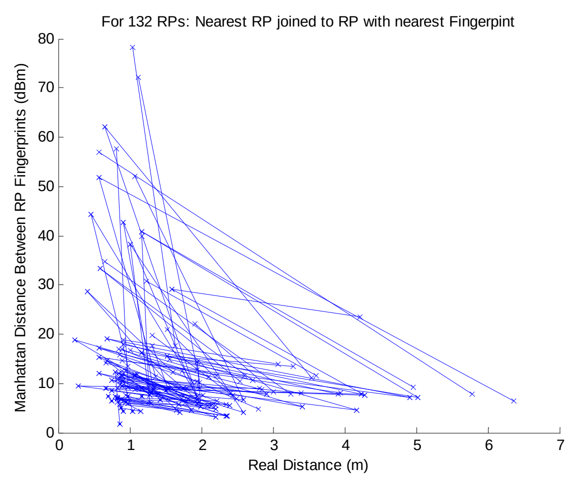

Figure 5 Pairs of data points joining the "nearest neighbour" found by measuring Manhattan distance between fingerprints and the neighbour that is physically nearest

IV. Conclusion

Existing data from fingerprinting experiments has been used to help gain some insight into the nature of errors arising in this process. The overall relationship between real distance and "vector" distance – the distance between fingerprints is investigated and found to be relatively poor. However, where real distances are short between fingerprints, this relationship improves so that several-nearest-neighbour algorithms are able to supply reasonable results.

Future work will include better characterization of the power statistics (for instance examination of the assumption of constant variance), examination of other vector distance measures such as Euclidean, examination of the effects when using a directional database, the effects of variation in the environment with time and development of an error prediction algorithm.

REFERENCES

[1] P. Bahl and V. N. Padmanabhan, “RADAR: An In-Building RF-based User Location and Tracking System,” Proc IEEE Conference on Computer Communications (INFOCOM), 2000

[2] T Roos, P Myllymaki, H Tirri, P Misikangas, J Sievanen, “A Probabilistic Approach to WLAN User Location Estimation”, Int J Wireless Info Networks, vol 9, no 3, July 2002, pp155-164

[3] Youssef, M, Agrawala, A, Shankar, AU “WLAN location determination via clustering and probability distributions”, IEEE Int. Conf. on Pervasive Computing & Communications (PerCom) 2003, Fort Worth, Texas, USA, 23-26 March, pp.143-150

[4] Li, B, Salter, J, Dempster AG, Rizos, C “Indoor Positioning Techniques Based on Wireless LAN”, First IEEE International Conference on Wireless Broadband and Ultra Wideband Communications, Sydney, Australia, 13-16 March, paper 113

[5] Li, B, Wang, Y, Lee, HK, Dempster, AG, Rizos, C “Method for Yielding a Database of Location Fingerprints in WLAN”, Communications, IEE Proceedings, Vol. 152, Issue 5, Oct 2005, pp. 580-586

[6] Li, B, Dempster, AG, Barnes, J, Rizos, C, Li, D “Probabilistic algorithm to support the fingerprinting method for CDMA location”, Int. Symp. on GPS/GNSS, Hong Kong, 8-10 Dec 2005, paper 9C-05, CD-ROM procs

[7] Xiang, Z, Song, S, Chen, J, Wang, H, Huang, J, Gao, X “A wireless LAN-based indoor positioning technology”, IBM J. Res & Dev, Vol 48 no. 5/6, Sep/Nov 2004

[8] Binghao Li, Jeffery Kam, Jonathan Lui, Andrew G. Dempster “Use of Directional Information in Wireless LAN based indoor positioning”, Proc IGNSS, Sydney, Dec 4-6 2007

[9] Ishrat J. Quader, Binghao Li, Wendi (Patrick) Peng, Andrew G. Dempster, “Use of fingerprinting in Wi-Fi based outdoor positioning”, Proc IGNSS, Sydney, Dec 4-6 2007

[10] Elliott D Kaplan and Christopher J Hegarty (eds), “Understanding GPS”, 2nd ed., Artech House, 2006

[11] D-H Shin and T-K Sung, “Comparisons of Error Characteristics between TOA and TDOA Positioning”, IEEE Trans Aerospace &Elect Sys, vol 38, no 1, Jan 2002, pp307-311

[12] A G Dempster, “Dilution of precision in angle-of-arrival positioning systems”, Electronics Letters, vol 42, no 5, 2 Mar 2006, pp291-292

[13] P Kontkanen et al, “Topics in probabilistic location estimation in wireless networks”, Proc Personal, Indoor and Mobile Radio Communications PIMRC 2004, 5-8 Sep 2004, vol 2, pp 1052- 1056

[14] P Myllymaki et al, “Error estimation concerning a target device’s location operable to move in a wireless environment”, US patent 7,209,752 B2, 24 April 2007

[15] B Li, “Terrestrial Mobile User Positioing using TDOA nad Fingerprinting Techniques”, PhD thesis, UNSW, Feb 2006

[16] Li, B., Wang, Y., Lee, H.K., Dempster, A.G. and Rizos, C. “A new method for yielding a database of location fingerprints in WLAN”, IEE Proc. Communications, vol. 152, no. 5, 2005, pp. 580-586

[17] Li, B, Dempster, AG, Rizos, C, Barnes, J, “Hybrid method for localization using WLAN”, Spatial Sciences Conference, Melbourne, Australia, 12-16 September, pp. 341-350, CD-ROM procs

.