

# Towards Practical Probabilistic Location Inference for Indoor Environment

Yi-Hsuan Chiang<sup>1</sup>, Ying-Chih Chen<sup>1</sup> and Polly Huang<sup>1, 2, 3</sup>  
 Graduate Institute of Communication Engineering<sup>1</sup>,  
 Graduate Institute of Networking and Multimedia<sup>2</sup>,  
 Department of Electrical Engineering<sup>3</sup>,  
 National Taiwan University, Taipei, Taiwan  
 {b92901134, r97942100}@ntu.edu.tw, phuang@cc.ee.ntu.edu.tw

## Abstract

In this work, we highlight the truncation effect in Received Signal Strength Indication (RSSI) distributions. The effect is often overlooked when applying probabilistic methods such as Multivariate Gaussian Inference (MGI), for location inference. Towards practical and accurate probabilistic location inference, we propose Multivariate Truncated Gaussian Inference (MTGI) to deal with the beacons with persistent or sporadic packet losses due to signal decay or collisions.

## 1. Background

Location inference based on Received Signal Strength Indication (RSSI) fingerprint of wireless signals is gaining popularity for indoor environment. The benefit of the approach is that the RSSI fingerprint captures not only the shadowing but also the multipath effect commonly seen in wireless signals transmitted in obstacle-rich space. In general, RSSI-fingerprint-based positioning consists of two phases: (i) the training phase where a mapping between the RSSI fingerprint and the corresponding position is established and (ii) the tracking phase where the RSSIs of an unknown position are measured and the position is estimated by identifying an RSSI fingerprint that is ‘closest’ to the measured RSSIs. .

In an RSSI-fingerprint-based location system, the sensing area is typically subdivided into smaller cells and RSSI readings are taken from several fixed beaconing nodes. The most likely cell is selected by determining which cell the fingerprint fits the best to the measurement. The methods used to determine the best fit are generally divided into two classes: deterministic and probabilistic.

In common deterministic methods, a system, for example [1], would represent the RSSI fingerprint as a vector and calculate the Euclidean distance between the measured and the fingerprint vectors to estimate their similarity. Assuming that the RSSI vectors measured at the same position should remain close, systems implementing the method often pick K vectors with the least distance (K Nearest Neighbors) and calculate the average of the corresponding K positions as the localization result.

The deterministic methods are easy to implement and efficient in computation time. The probabilistic methods, on the other hand, are considered more robust handling the inherent instability in the RSSI readings and can potentially provide consistent and precise location estimations. In most probabilistic methods, one would take long-term measurement of RSSIs and construct an RSSI fingerprint distribution at each cell. In the tracking phase, a ‘likelihood’ of the measured RSSIs fitting in a

certain RSSI fingerprint distribution is computed. Often times, those cells with the highest likelihood are selected and the average position of these cells is returned as the final result.

The probabilistic methods differ in the distribution used to model the RSSI fingerprint distributions, as well as the algorithm used to compute the likelihood. For instance, the location system presented in [2] models the RSSI fingerprint distributions as Gaussian processes. The likelihood function takes in multiple variables, i.e., the RSSI values coming from multiple beacon nodes. It outputs the probability the combination of RSSIs belongs to the fingerprint distribution per cell. Despite the precision and robustness, probabilistic methods, such as Maximum Likelihood (ML) by Bayesian inference [3,4,5] or Hidden Markov Model (HMM) [6], require more time to collect training data, as well as to complete the computation.

## 2. Multivariate Gaussian Inference

Multivariate Gaussian Inference (MGI) is the simplest of all the probabilistic methods. Consider a simple case where there is only one beacon in the system. Gaussian inference assumes the distribution of received RSSIs from the beacon at a particular cell follows the Gaussian distribution. Given a new RSSI measurement from the beacon, the probability density function of the Gaussian distribution implicates the chance of the RSSI value taken at the cell.

Consider multiple beacons in the system. MGI is the generalization of the one-dimensional Gaussian inference. The probability density function is formulated in equation (1).

$$p(X) = \frac{1}{\sqrt{(2\pi)^N |\Sigma|}} \exp\left(-\frac{1}{2}(X - \mu)^T \Sigma^{-1} (X - \mu)\right) \dots\dots(1)$$

- ◆  $X$  is the RSSI vector:  $X \equiv (d_1 \dots d_N)^T$
- ◆  $d_i^{(j)}$  is the  $j^{\text{th}}$  RSSI reading from beacon  $i$ . There are  $N$  beacons in localization system.
- ◆  $\mu \equiv (\bar{d}_1 \dots \bar{d}_N)^T$ ,  $\mu$  is the vector of the fingerprint distribution, composed of the mean RSSI value measured from all beacons.
- ◆  $|\Sigma|$  denotes the determinant of the covariance matrix:

$$\Sigma = \begin{bmatrix} \sigma_{11}^2 & \dots & \sigma_{1N}^2 \\ \vdots & \ddots & \vdots \\ \sigma_{N1}^2 & \dots & \sigma_{NN}^2 \end{bmatrix} \text{ with } \sigma \text{ as the standard deviation, and}$$

$\sigma_{lm}^2$  the covariance between  $l$  and  $m$ , i.e.

$$\sigma_{lm}^2 = \frac{\sum_i (d_l^{(i)} - \bar{d}_l)(d_m^{(i)} - \bar{d}_m)}{K - 1}, \text{ for } K \text{ received readings.}$$

- ♦ And  $\Sigma^{-1}$  is the inverse of covariance matrix  $\Sigma$

In MGI, the correlations of the RSSI values from different beacons are captured by taking into account the covariance matrix. As a dominant part in MGI, the covariance matrix  $\Sigma$  can be divided into diagonal and non-diagonal terms. The diagonal terms are essentially the variance of RSSIs from each beacon. The higher the variance is, the less stable the signals are. Its inverse, as a result, gives  $d_i - \bar{d}_i$  a lower importance, meaning the difference between the measured value and the sample mean is less credible. Similarly, the non-diagonal terms indicate the cross-covariance between different beacons, reflecting the correlation among beacons and the importance of the  $l$ th beacon to the  $m$ th and vice versa.

Instead of treating RSSIs from each beacon as an independent Gaussian distribution, MGI tracks the subtle correlation between the neighboring beacons, as well as other unexpected inter-dependent beacons due to the multipath effect commonly observed in the indoor environment.

### 3. Test and Implementation

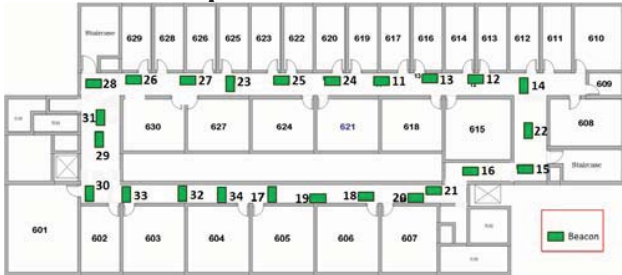


Figure 1. MGI Location System Testbed

The testbed is built on the 6<sup>th</sup> floor of Barry Lam Hall at NTU. We deployed 24 beacons which periodically transmit short packets containing the beacon ID. The beacon nodes are Telos-like modules [7] equipped with TI MSP430 microcontrollers and CC2420 802.15.4 radio. The software is implemented on TinyOS, and the default MAC, a CSMA/CA-like mechanism, is on for all beacon packet transmissions.

Figure 1 shows the floor plan. The smaller rooms, numbered 611 to 629, are faculty offices and the remaining are graduate assistant laboratories. The 24 Telos-like beacon nodes are small boxes distributed evenly along the corridor. We divide the sensing area into a number of small cells. Each cell is approximately 30cm apart. During the training phase, a receiving node is held by a staff member. The staff member would walk along the corridor and stay at each cell to collect approximately 150 RSSIs per beacon to form the RSSI fingerprint distributions. In the tracking phase, the receiving node takes the RSSI values from the beacons and the MGI computation is performed to estimate the location of the receiving tag.

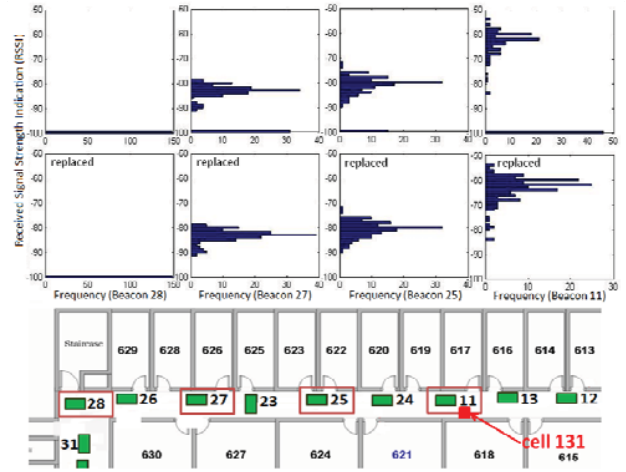


Figure 2. Distributions of RSSIs from Different Beacons

### 4. Are RSSI distributions Gaussian?

Whether MGI will work effectively estimating the receiving tag's position depends greatly on whether the RSSI distributions are indeed Gaussian. Selecting a particular cell (cell #131 in Figure 2), we examine the measured RSSIs of beacons from the far end of the corridor to the nearest. The 4 columns of plots in Figure 2 are histograms of the RSSIs from beacon #28, #27, #25, and #11 respectively. One can observe that the RSSIs concentrate at level -60, -80, and -83 for beacon, #11, #25, and #27. Although not perfectly Gaussian, the distributions are in the ball park. In addition, one can also see that the closer the beacon is, the higher the RSSIs are.

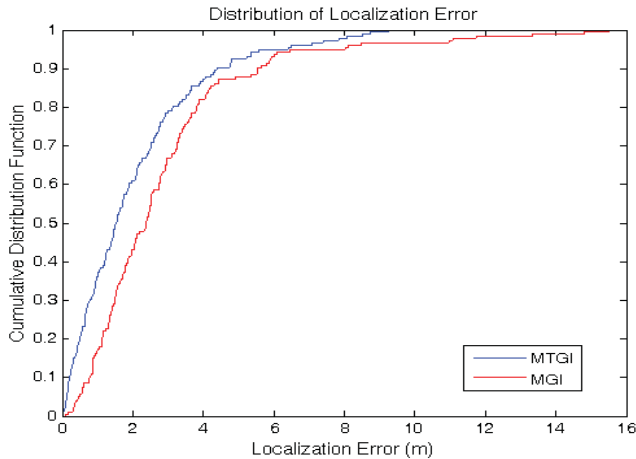
One unusual data point in the histograms is the -100 bin, the defined minimum RSSI level. Lost beacon packets, due to signal strength decay or collision, are all accounted into the -100 bin in the plots. We can see that the amount of losses is significant, although the losses of beacon #11 are more likely due to collision and those of beacon #28 due to signal decay. How we account the lost beacon packets in MGI computation will be crucial to the localization accuracy.

In beacon #11's case, one can replace the missing RSSIs with the value generated from the Gaussian distribution fitted by the calculating mean and variance of the received ones. With this method, we can recover the few beacon losses given the majority is well received and further improve the accuracy of MGI. See the histograms of the replaced data in the second row of Figure 2.

In beacon #27 and #28's cases, the RSSIs are close to the minimum. The resulting distributions are respectively a partially and a fully truncated Gaussian. It will not be robust to recover the many beacon losses using the distribution of the minority. Especially for those beacons that cannot receive any beacon packet at all, using the same minimum values, all -100 for example, to represent the RSSI distribution will result in a singular value while calculating the inverse of covariance matrix. The variance and cross covariance involving the beacon will all be zero. Adding very weak white noise may avoid the singular point. However, the auto-correlation in artificially inserted noise may distort the computation of MGI. Hence, classifying measured RSSIs from different beacons received and then processing with

different methods accordingly would be the keys towards accurate probabilistic location inference.

## 5. Handling Truncated Gaussian



**Figure 3. The distribution of localization error of MGI and MTGI**

The approach we are proposing, called Multivariate Truncated Gaussian Inference (MTGI), consists of two components to be applied to the training and tracking phases of the location system respectively: (1) Packet Reception Rate (PRR) inspection (2) Multivariate Truncated Gaussian Inference (MTGI).

In the training phase, the data collected at each cell are used not only to form the fingerprint distribution, but also to calculate the PRR of each beacon. An additional data structure introduced to the system is a bit map in which a bit is set to 1 when the PRR is larger than 0.9, otherwise, set to 0. In the tracking phase, the covariance matrix is constructed by only the beacons mapped to 1. In that, the mean and covariance are calculated using only the RSSIs from which beacon packets are well received.

Figure 3 depicts the cumulative distribution function (CDF) of the localization errors applying the naïve MGI and the proposed MTGI on the traces taken for Section 4. For MTGI, the median is less than 2 meters and the 80-percentile is bounded within 3 meters, as opposed to naïve MGI whose median and 80-percentile are approximately 3 meters and 4 meters. The result indicates that there is a significant degree of improvement using MTGI over naïve MGI. The potential is the highest keeping the worst-case error low.

## 6. Summary and Outlook

Although MTGI performs better than MGI, the localization error is still high. This might be the result of RSSI signature

instability over time. In this preliminary study, each RSSI signature distribution is collected in a short 30-second interval. The signature distribution derived might not be really characteristic of the beacon. To further reduce the localization error, especially for the worst case, we are looking forward to relaxing the Gaussian assumption by substituting the distribution to bi-peak models (beacon #27) or empirical models characterized from long-term observations of RSSI distributions.

## 7. References

- [1] Bahl, P.; Padmanabhan, V.N., "RADAR: an in-building RF-based user location and tracking system," *INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol.2, no., pp.775-784 vol.2, 2000
- [2] Seshadri, V.; Zaruba, G.V.; Huber, M., "A Bayesian sampling approach to in-door localization of wireless devices using received signal strength indication," *Pervasive Computing and Communications, 2005. PerCom 2005. Third IEEE International Conference on*, vol., no., pp. 75-84, 8-12 March 2005
- [3] Youssef, M.A.; Agrawala, A.; Udaya Shankar, A., "WLAN location determination via clustering and probability distributions," *Pervasive Computing and Communications, 2003. (PerCom 2003). Proceedings of the First IEEE International Conference on*, vol., no., pp. 143-150, 23-26 March 2003
- [4] Ito, S.; Kawaguchi, N., "Bayesian based location estimation system using wireless LAN," *Pervasive Computing and Communications Workshops, 2005. PerCom 2005 Workshops. Third IEEE International Conference on*, vol., no., pp. 273-278, 8-12 March 2005
- [5] Swangmuang, N.; Krishnamurthy, P., "Location Fingerprint Analyses Toward Efficient Indoor Positioning," *Pervasive Computing and Communications, 2008. PerCom 2008. Sixth Annual IEEE International Conference on*, vol., no., pp.100-109, 17-21 March 2008
- [6] Andrew M. Ladd, Kostas E. Bekris, Algis Rudys, Lydia E. Kavraki, Dan S. Wallach, Guillaume Marceau, "Robotics-based location sensing using wireless Ethernet," *Proceedings of the 8th annual international conference on Mobile computing and networking*, September 23-28, 2002
- [7] J. Polastre, R. Szewczyk, and D. Culler. Telos: enabling ultra-low power wireless research. In *Proceedings of IPSN '05*, page 48, Piscataway, NJ, USA, 2005. IEEE Press.