

Practical Modeling and Prediction of Radio Coverage of Indoor Sensor Networks

Octav Chipara, Gregory Hackmann, Chenyang Lu, William D. Smart, Gruia-Catalin Roman

Department of Computer Science and Engineering
Washington University in St. Louis, USA
{ochipara, gwh2, lu, wds, roman}@cse.wustl.edu

ABSTRACT

The robust operation of many sensor network applications depends on deploying relays to ensure wireless coverage. *Radio mapping* aims to predict network coverage based on a small number of link measurements. This problem is particularly challenging in complex indoor environments where walls significantly affect radio signal propagation. Nevertheless, we show that it is feasible to accurately predict coverage through a two-step process: a propagation model is used to predict signal strength at a recipient node, which is then mapped to a coverage prediction. Through an in-depth empirical study, we show that complex models do *not* necessarily produce accurate estimates of signal strength: there is an important tradeoff between model accuracy and the number of parameters that must be estimated from limited training data. We find that the best performance is achieved by a family of models which classify walls based on their attenuation into a small number of classes and develop an algorithm to perform this classification automatically. Based on these insights, we build a novel *Radio Mapping Tool* (RMT) for predicting radio coverage in indoor environments. Experimental results demonstrate RMT's effectiveness in two buildings: RMT reduces the number of locations where coverage is erroneously predicted to exist by as much as 39% and 54% compared to the classic log-normal radio propagation model.

Categories and Subject Descriptors

C.2.1 [Network Architecture and Design]: Wireless communication; C.4 [Performance of Systems]: Modeling techniques

General Terms

Measurement, Performance

Keywords

Coverage, Wireless propagation models, Wireless Sensor Net-

works

1. INTRODUCTION

Sensor network applications involving mobile entities commonly require the deployment of wireless networks that cover a physical region. Examples of such applications include elderly care [22] and patient tracking and monitoring [5]. Our interest in this topic is motivated by a medical application which involves the collection of pulse and oxygenation readings from patients in general hospital units. Unlike patients in the intensive care units, the patients in general hospital units are often ambulatory. To support patient mobility, our system [4] requires enough relay nodes so that there is always at least one link from the patient to some relay. During the deployment of the system at Barnes Jewish Hospital, we became acutely aware of the lack of tools which would enable system managers to effectively assess the coverage of a deployed network. More specifically, we are interested in determining the reception coverage of a relay: i.e., the set of points (x, y) where a node would be able to transmit a packet to at least one relay with a PRR above a user-specified threshold¹.

The current best practice for assessing network coverage is to exhaustively measure link quality at numerous locations with deployed relays. This process is labor intensive and leads to significant deployment costs. Worse, physical changes (e.g., reconfiguring cubicles) or changes in the radio properties (e.g., switching radio frequency due to interference) may invalidate these measurements, leading to significant maintenance costs.

What is needed is a tool which can assess the coverage of a wireless network without an exhaustive survey. The key to assessing wireless coverage lies in effectively modeling radio propagation in the deployment environment, including obstacles that can attenuate the radio signal. Within the 802.11 networking community, there are a handful of tools which use ray tracing techniques to model signal propagation [6]. These tools require precise characterizations of the location and radio properties of objects that can significantly affect radio propagation, such as walls, bookshelves, or filing cabinets. In many indoor environments, such as office environments, these obstacles are numerous; for example, our

¹The techniques proposed in this paper are also applicable to the network's transmission coverage: i.e., the set of points that can receive transmissions from at least one relay. We focus on reception coverage in this paper, since our target application entails data collection. Henceforth, we use the term "coverage" to mean "reception coverage".

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

IPSN'10, April 12–16, 2010, Stockholm, Sweden.

Copyright 2010 ACM 978-1-60558-988-6/10/04 ...\$10.00.

1977m² indoor testbed contains 239 walls. Measuring each wall directly would impose an excessive burden on the user.

A less labor-intensive approach is to collect a set of link quality measurements from the environment. This training data can then be fit to some radio propagation models in order to estimate the value of each parameter. Indeed, this approach has proven effective in outdoor environments [16]. However, our empirical study shows that this approach is unsuitable for complex indoor environments. This occurs because obstacles, antenna orientation, and distance between sender and receiver affect signal propagation to different degrees indoors and outdoors. While complex models may be constructed to account for all these factors, there is an important tradeoff between model complexity and measurement effort: as radio models are made more complex, more data is necessary to accurately fit the additional parameters. Thus, in this paper, we consider the problem of how to effectively predict radio coverage in complex indoor environments from a *small* set of training data.

An empirical study in two office buildings shows that the best tradeoff between model realism and model complexity lies in automatically classifying obstacles into groups with similar attenuation. Using this knowledge, we divide the problem into two parts. We first predict the receive signal strength (RSS) at the relay from any point on the floor plan. Then, based on the RSS predictions and an RSS threshold for predicting good-quality links, we determine each relay's coverage.

This paper makes the following key contributions. First, we present an in-depth empirical study that characterizes the accuracy of RSS predictions based on several propagation models. The study shows the relative importance of modeling various aspects of wireless propagation such as antenna orientation, wall attenuation, and distance between sender and receiver. More importantly, the study shows that complex models do *not* necessarily produce accurate estimates of signal strength: the best performance is achieved by a family of models which classify walls based on attenuation into a small number of groups. We also propose an automatic process for selecting the best such model for the provided amount of training data, reducing errors by up to 9.7% compared to the classical log-normal radio propagation model [2]. Next, we develop a practical *Radio Mapping Tool* (RMT) which predicts the coverage of one or more relays. As a key component of RMT, we develop a novel automated wall classification algorithm to be used with the chosen radio model. We then characterize the accuracy of this tool in two different buildings with differing construction properties. We find that the combination of our chosen radio model with our wall classification scheme reduces the false positive rate (i.e., predicted coverage where the ground truth indicates otherwise) by as much as 54% compared to the log-normal model, based on a sampling density of only 0.01 samples/m².

The remainder of the paper is organized as follows. In Section 2, we discuss existing studies on characterizing wireless signal propagation. In Section 3, we overview several established radio models and discuss their applicability to indoor environments. In Section 4, we discuss methods to classify walls, including a computationally efficient algorithm that automatically performs this classification. The RSS prediction accuracy of different propagation models is assessed in Section 5. In Section 6, we present a radio mapping tool

built based on the insights gained from our empirical study. Section 7 evaluates the efficacy of our radio mapping tool through a case study. We then conclude in Section 8.

2. RELATED WORK

A key challenge in modeling radio properties is that low-power wireless links have complex, often probabilistic properties [7, 13, 15, 19, 20, 24]. The classical *log-normal* model [21, 25] models a node's transmission strength and signal decay over distance. As we show in Sections 5 and 7, the log-normal model is overly simplistic, resulting in significant prediction errors.

A deficiency of the log-normal model is that it does not capture the non-isotropic antenna pattern observed even with "omnidirectional" antennas [16, 19, 23]. [23] demonstrates that these non-regular radiation patterns can have a significant effect on routing performance in an outdoor wireless sensor network. [16] shows a similar effect for two outdoor Wi-Fi mesh networks. Both studies propose a *sectorization* approach that divides each node's signal into sectors, then attempts to independently model the signal properties of each sector. Our own study finds that, while the non-isotropy of antenna patterns also impacts radio propagation indoors, this effect is less significant than the attenuation caused by obstacles.

[16] expands the sectorization model to explicitly model non-isotropic antenna patterns. Exhaustive link data is collected at various points around each feature to individually estimate its attenuation. Our own study shows that modeling obstacle attenuation can also significantly improve coverage prediction indoors. However, our work differs from [16] in two key ways. First, as discussed above, we do not model antenna patterns; the impact of obstacles are more important in an indoor environment, and modeling non-isotropy introduces a large number of parameters that are difficult to estimate from a small number of samples. Second, [16] directly measures architectural features, which is impractical and labor-intensive in typical indoor environments such as offices, assisted living facilities, and hospitals. A novel feature of our work is that we leverage the fact that the walls in any given building can be classified into relatively few classes of similar attenuation, greatly reducing the amount of data needed to adequately estimate their attenuation. Moreover, we propose an algorithm which automatically classifies walls using a small set of training data, without requiring architectural knowledge or direct measurements of each wall.

At the other end of the complexity spectrum, researchers have proposed site-specific techniques involving ray tracing [14, 17]. [6] presents a tool for predicting signal strength of 802.11 access points at different locations. A fundamental limitation of these techniques is that they rely on the user to provide locations and attenuation coefficients for each partition or obstacle. Tables which provide the attenuation of different wall types [18] can alleviate this burden somewhat, though this still requires knowledge of the building's construction materials and may not capture the effect of objects like metal bookshelves that can alter a wall's attenuation. In contrast, our approach automatically estimates the attenuation of walls from training data.

Also closely related to our work are two recent papers which look at sensing coverage. [11] proposes a framework which uses Gaussian processes to model sensing and communication costs. A disadvantage of Gaussian processes

is that they cannot effectively model discontinuities such as those observed when a signal passes through walls. In contrast, our approach explicitly models wall attenuation, which our study in Section 5.5 shows to be significant. [9] proposes a method for determining a sensor’s sensing range through hierarchical sampling. This approach is complementary to our own, since it deals with efficient sampling strategies for refining coverage boundaries; our work focuses on processing the collected samples to predict coverage.

3. RADIO PROPAGATION MODELS

Propagation models optimized for different wireless technologies and environments have been proposed in literature [1, 8, 10]. We assume that nodes operate on a fixed frequency and transmission power. The models presented in this section focus on three characteristics which may significantly affect signal propagation in indoor environments: (1) the distance between the sender and receiver, (2) antenna orientation, and (3) the impact of walls. We note that the models considered in this section do not model multi-path propagation. While multi-path propagation may be modeled through ray-tracing techniques, such approaches are usually computationally demanding and require a precise characterization of the environment such as the materials of walls. Since our goal is to develop an interactive radio mapping tool that requires minimum knowledge about the environment, we opted to ignore these effects. Moreover, we show that the simple models proposed in this section may accurately predict coverage.

By their nature, these models’ parameters are estimated from an (ideally small) set of training data. The need for training data represents an important trade-off that we will revisit throughout this paper: while an overly simplistic radio model may not provide an accurate estimate of communication coverage, adding more complexity will not *necessarily* improve the model’s performance. As more parameters are added, more training data is needed to adequately estimate them — conversely, for a fixed training data size, the estimates for each parameter may degrade as more are added.

Thus, the challenge in creating a realistic radio model lies not only in identifying what factors can affect signal propagation, but also *which* of these factors are the most important to capture. Our goal is to identify the model with the best trade-off between prediction accuracy and the number of samples needed to estimate its parameters. This model will ultimately be used in our Radio Mapping Tool to generate signal strength predictions. The models presented in this section will be evaluated empirically in Section 5.

Log-Normal Shadowing: Under the log-normal model, signal strength decays exponentially as a function of distance. Let $d(s, r)$ be the distance between the sender node s and the receiver node r . The receive signal strength $P_r(s, r)$ at r from a sender s is given by: [2]

$$P_r(s, r) = \alpha - 10\beta \log_{10} d(s, r) + \sigma \quad (1)$$

Here, α represents the transmission power at a reference distance of 1m and β represents the pass loss exponent. σ models shadowing (i.e., the random signal variations between sender and receiver) and is usually considered to be a normally distributed random variable.

Prior empirical studies have shown that this model may accurately predict the receive signal strength of low-power

Log-Normal	Sector-Based	Per-Wall	Wall-Class
2	$NS * n + 1$	$ W + 2$	$ C + 2$

Table 1: Parameters per model

radios in outdoor environments [13] and in indoor environments where nodes have line-of-sight [13, 25]. However, this model does not account for the impact of walls, which commonly have major impacts on the coverage of sensor networks deployed indoors.

Sector-Based: Prior literature has extended the basic log-normal model to capture the fact that many low-power radios have non-isotropic radiation patterns [24]. That is, even when nodes are positioned at equal distances from the sender, they may observe significantly different receive signal strengths.

The receive signal strengths depend on the relative orientation of the sender and receiver. However, to simplify the problem, the relative position of the sender and receiver is commonly kept constant during data collection. In this case, non-isotropic behavior is accounted for by parametrizing α by the angle θ between the line connecting s and r and a fixed frame of reference:

$$P_r(s, r) = \alpha(s, \theta) - 10\beta \log_{10} d(s, r) + \sigma \quad (2)$$

$\alpha(\theta)$ may be a non-linear function [24]. As a result, non-linear optimization techniques would be necessary for fitting the model. To simplify fitting, the impact of antenna orientation may be captured by discretizing θ into a number of sectors. This enables us to use linear fitting to estimate all model parameters.

Per-Wall Attenuation: In indoor environments, walls may significantly attenuate wireless links. Hence, incorporating walls into the radio propagation model can improve its signal strength predictions.

An intuitive way of modeling wall attenuation is to assume that each wall $w_i \in W$ in the environment attenuates the signal by a constant factor γ_{w_i} . If we let $I_{s,r}$ be the set of all walls which intersect a virtual line between s and r , then the signal strength at r is:

$$P_r(s, r) = \alpha - 10\beta \log_{10} d(s, r) + \sum_{w \in I_{s,r}} \gamma_w \quad (3)$$

This model may also be modified to incorporate non-isotropic radio range by treating α as a function of θ as previously discussed.

Several measurements should be taken through each wall to accurately estimate γ . This may be a significant burden in some environments; for example, one building in our environment contained 128 walls in $1020m^2$ of floor space.

Wall-Class Attenuation: A pragmatic alternative to the per-wall scheme is to group walls into a few classes, reflecting the fact that only a few types of walls are used in construction. For example, the building shown in Figure 2 mainly uses two kinds of walls: cinder block and drywall. Given a set of classes C , a mapping $\Pi : W \rightarrow C$, and an attenuation coefficient Γ_{c_i} for each class $c_i \in C$, the signal strength at a node r is:

$$P_r(s, r) = \alpha - 10\beta \log_{10} d(s, r) + \sum_{w \in I_{s,r}} \Gamma_{\Pi(w_i)} \quad (4)$$

Table 1 summarizes the number of parameters used by each model. As later highlighted by the empirical results presented in Section 5, one of the key challenges of Radio

Mapping is selecting a model which achieves the best prediction accuracy given a number of measurements. Models with small number of parameters have the advantage of requiring a small number of measurements for determining their parameters. Moreover, when only a limited number of parameters are available, it is imperative to focus on the factors which have the most significant impact on signal propagation.

The log-normal model has only two parameters which need to be evaluated. In contrast, the sector-based model has as many as $NS * n + 1$ parameters, where NS and n are the number of sectors and relays, respectively. The two buildings used in our experiments contained 64 nodes and 28 nodes, respectively. Due to the multiplication between number of sectors and number of relays, such environments will generate models with numerous parameters. Similarly, the per-wall model accounts for the attenuation of walls; as a result, one expects it to be more accurate. The number of parameters used by this model is $|W| + 2$. $|W|$ may be as high as 100 in typical office buildings, resulting in a model with a significant number of parameters. Therefore, we expect the per-wall model to require copious measurements as training data. Moreover, obtaining good statistics for the attenuation of a wall potentially requires multiple measurements per wall. We hypothesize here (and show in Section 5) that models with numerous parameters require a significant amount of training data, making them impractical for our Radio Mapping Tool.

In contrast with the previously discussed models, the wall class model requires $|C|$ parameters. In our experience, typical values for $|C|$ are between 1–5 wall classes, significantly reducing the number of model parameters. Such a model is particularly attractive for our Radio Mapping Tool since it would require only a small number of measurements. However, it also creates a new problem: a mapping Π from walls to classes needs to be constructed. In the next section, we will present an efficient algorithm for constructing this mapping.

4. AUTOMATIC WALL CLASSIFICATION

One way to construct this wall classification is to manually classify walls based on their construction material. Linear regression may then be used to fit the remainder of the model’s parameters as described above. However, manual wall classification is labor-intensive and requires architectural information that may not be readily available to application developers or network managers.

Hence, we propose to classify each wall automatically. The problem of automatically classifying walls into classes may be addressed in the Expectation Maximization (EM) framework. The EM framework is best suited for finding the maximum likelihood estimate when the model depends on latent variables, which in our case are the wall classes. We propose the novel application of the EM framework to automatically classifying walls.

The input to the classification algorithm is based on link statistics collected by the user when located at a small number of measurement locations. Multiple packets are broadcast at each measurement location and the relay nodes record their RSS. For each link formed between a relay and a node positioned at a measurement location, we provide the median RSS as vector y and the Euclidean distance between the link’s endpoints as vector d . The set of walls and wall

```

 $[\alpha, \beta, \Gamma, \Pi] = \text{compute-parameters}(y, d, W, C):$ 
1: improvement = true;
2: for each wall  $w \in W$ :
3:    $\Pi(w) = \text{rand}(C)$ ;
4: while (improvement):
5:   improvement = false;
6:    $[\alpha, \beta, \Gamma] = \text{regress}(y, [d; \Pi])$ ;
7:   for each wall  $w \in W$  in random order:
8:      $\Pi_{new} = \Pi$  and  $c_{old} = \Pi(w)$ ;
9:     for each class  $c \in C$ :
10:       $\Pi_{new}(w) = c$ ;
11:       $\hat{y} = \alpha(s) - 10\beta \log_{10} d(s, r) + \sum_{w \in \mathcal{I}_{s,r}} \Gamma_{\Pi_{new}(w)}$ ;
12:       $SSE(c) = \sum_{i=1}^{|y|} (y(i) - \hat{y}(i))^2$ ;
13:       $c_{best} = \text{arg min}_c SSE(c)$ ;
14:      if ( $c_{old} \neq c_{best}$ ):
15:         $\Pi(w) = c_{best}$ ;
16:        improvement = true;
17:      break;

```

Figure 1: Wall classification algorithm

classes are provided as W and C , respectively.

Figure 1 presents the pseudocode of this algorithm. Initially, each wall is assigned to a random class. The algorithm then proceeds in two stages, repeating until changes in wall classification stop improving the sum of squared errors (SSE) between the predicted signal strengths (\hat{y}) and the actual signal strengths (y). In the first stage (line 6), the algorithm uses linear regression to fit the parameters α and β , as well as the attenuation coefficient Γ for each wall class. The second stage (lines 7–16) aims to improve the mapping of walls to classes with these values of α , β , and Γ fixed. This is done by considering each wall w in random order, computing the SSE when w is assigned to each class in C . If reassigning w results in a smaller SSE, then w ’s classification is updated accordingly and the algorithm goes back to executing the first stage with an improved wall classification. Otherwise, the algorithm considers the next wall. The algorithm terminates when no wall may be assigned to a new class that reduces the SSE. The values of the parameters α , β , and Γ are then returned along with the mapping Π of walls to classes.

This algorithm has two noteworthy features. First, it is much less computationally expensive than an exhaustive search. The wall-reassignment stage considers at most $|C| \times |W|$ potential assignments at each iteration. Thus, in practice, this algorithm can be executed in under two minutes on a modern laptop PC even when predicting coverage of relays spanning an entire building (tens of relay locations and about a hundred measurement locations).

Second, the algorithm is guaranteed to converge. This is because the algorithm reduces the squared error at each step until it terminates. There is no guarantee on the optimality of the solution, since it may get stuck in a local minimum. Because of the random initial assignment of walls to classes and the random ordering in which walls are reclassified, the algorithm may return different values each time it is run. Accordingly, we may further improve the squared error by repeating the algorithm several times and returning the parameters which resulted in the lowest squared error.

5. EMPIRICAL MODEL COMPARISON

In this section, we present an empirical study which aims to address three questions at the core of our Radio Mapping technique: (1) which factors affect signal propagation

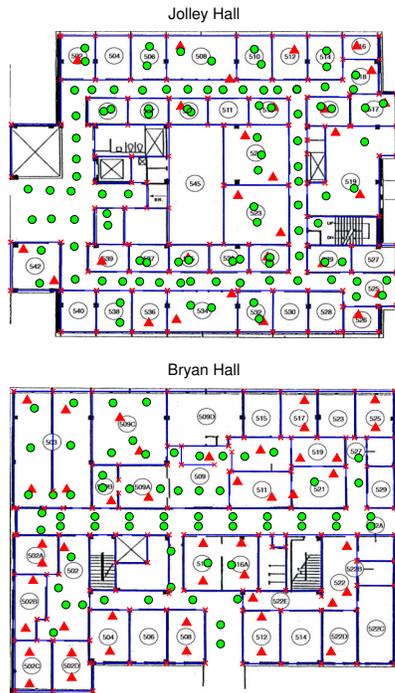


Figure 2: Test buildings

in an indoor environment, (2) how model accuracy affects the number of samples needed for parameter fitting, and (3) the robustness of different propagation models in indoor environments. In answering these questions, we provide guidelines for developing a practical radio mapping technique which can be used in complex indoor environments.

5.1 Experimental Setup

Our experiments were carried out in two indoor office buildings (see Figure 2; triangles represent relays and circles represent test positions) using TelosB nodes. These two buildings serve as good test cases because they were constructed in different years with different materials: for example, the walls in Bryan Hall contain steel rebars that attenuate wireless signals, while the walls in Jolley Hall do not. The nodes are equipped with CC2420 low-power radio chips, which provide an RSS indicator reading for each decoded packet. All nodes in our experiment were set to 802.15.4 channel 26, which does not overlap with the buildings' 802.11g network.

The experimental setup is motivated by our interest in supporting robust data collection from mobile users. Accordingly, we aim to ensure that at least one relay node is capable of receiving data from a user standing in any location. A total of 28 and 45 nodes were deployed close to the ceilings of Jolley and Bryan Halls, respectively, representing the locations where relays have been deployed. We used these nodes to record the link quality at when a test node is placed at numerous locations in the two buildings (104 locations in Jolley and 64 locations in Bryan). At each measurement location, a sender node broadcasted packets at the eight power levels available on the CC2420 radios. The Jolley dataset was collected with the sender placed 1.5m off the ground on a tripod, while the Bryan dataset used plas-

tic cups 15 cm off the ground. The relays recorded the RSS reading and sequence number of each successfully decoded packet, which was relayed to a central database through a wired back channel. Each data set was collected during the night over two consecutive days.

To compensate for errors in the CC2420's raw RSS readings, we calibrated the RSS data in a similar fashion as [3], using a calibration curve provided by the authors. After calibration, the collected data is divided into training and testing sets. To account for uneven spatial distribution in our measurements, we construct the training and testing sets as follows. A number of points are generated uniformly over the 2D floor plan, and the links with senders closest to these points are selected as part of the training set. The testing set is generated from the remaining links in a similar fashion. We vary the size of the training set by varying the number of randomly generated points, with densities ranging from 0.01 to 0.11 samples/m² in increments of 0.01.

Unless mentioned otherwise, the presented results are averages of 10 randomly generated training sets. We evaluate the performance of various models based on the 80th percentile of the absolute error between predicted and actual RSS values.

5.2 Effect of Walls

First, we evaluate the effectiveness of including walls in our radio propagation models by comparing the performance of models which incorporate topological information against the log-normal model. Figure 3 presents the error for these models, including three approaches that include wall attenuation: treating each wall as an independent variable (*per-wall attenuation*), assuming that walls are of the same construction material (*1 wall class*), and manually labeling the type of each wall based on architectural knowledge (*manual classification*). The 1-wall class and manual classification models consistently outperform the log-normal model in both environments, with 4.7%–8.4% lower error regardless of the amount of training data available.

The overall reductions in error are modest. However, we note that not all links are equally important for overall coverage predictions. Rather, the most important links are those close to the coverage boundary, which have an RSS in the transitional region. As discussed in detail in Section 6, the transitional region in our testbed occurs the RSS range of $[-87, -80]$ dBm. In this region, we have found that the models which model wall attenuation can significantly outperform models which do not. This effect is illustrated in Figure 3(c), which plots error as a function of predicted RSS for the Jolley dataset with a density of 0.1. Within the transitional region, the models which incorporate walls outperform the log-normal model by as much as 22.2%.

We also observed that adding more training data only slightly improves most of the models' performance. The per-wall attenuation model is the exception, improving by as much as 30% when more training data is provided. This is because the per-wall attenuation model has about 100 parameters that require large amounts of data to accurately estimate, whereas the other models have few parameters that can be fit well using relatively little data. We observe that the per-wall attenuation model may outperform the other wall models when given enough training data.

Summary: *The number of parameters in a model must be tuned to match the amount of available training data.*

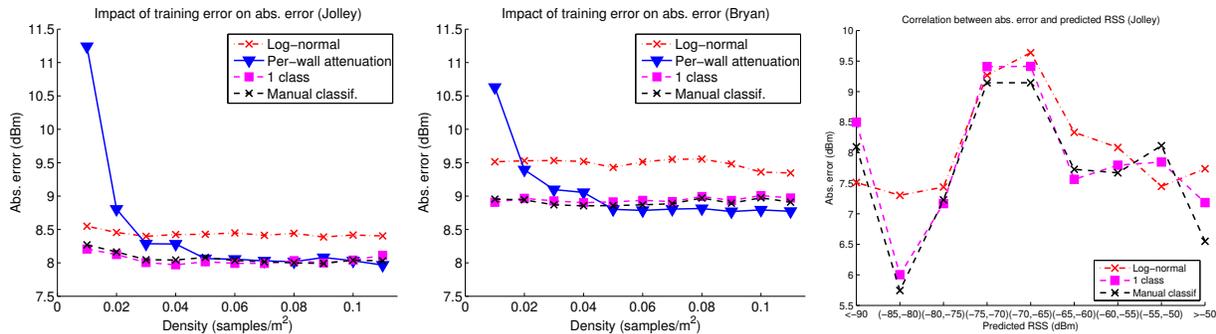


Figure 3: Comparison of propagation models

5.3 Automatic Wall Classification

The previous experiment showed the benefits of using wall information and the pitfalls of using models with numerous parameters for limited training data. In this section, we consider models which automatically classify walls into a small number of classes. By constraining the number of classes, we hypothesize that we can improve prediction accuracy without requiring large training sets.

Figure 4 compares the estimation accuracy of the automatic wall-classification model with 1–3 classes of walls. For comparison, we also include the per-wall model and the manual wall classification model. When little training data is available, using fewer wall classes improves the predicted accuracy for both data sets. At a density of 0.01, the 1-class model outperforms all other models, with 10.5% and 6.3% lower error in the two dataset than the 3-class model. In contrast, at a density of 0.10, the 3-class model achieves the lowest error, at 4.6% and 6.18% lower than the 1-class model. Both data sets indicate that additional wall classes are beneficial when more training data is available.

Summary: *Classifying walls into a few classes achieves the lowest error when the training data set is small; but as the amount of training data increases, more classes should be employed.*

We also note that the automatic wall classification scheme achieves lower error than the manual wall-classification scheme. In fact, on the Bryan data set, the automatic wall classification scheme has 6.2% lower error than the manual wall-classification scheme at the maximum sampling density. This is explained by the fact that the attenuation of a wall is partly determined by the additional shelving or furniture present in an office or in the room. Besides being more labor-intensive, a manual classification based purely on construction material would not capture this information.

Summary: *Automatic wall classification model achieves higher prediction accuracy without requiring the user to manually classify the walls.*

5.4 Boosting

Looking at the predictions from individual relays, we observed that the “best” model often depends on the location — the model with the lowest error in one room would not necessarily have the lowest error in another. Thus, we propose an approach we call *boosting*, which reduces error by combining different models’ predictions. Intuitively, boosting divides the map into regions (i.e., rooms or hallways) and finds the most accurate model on a per-region basis.

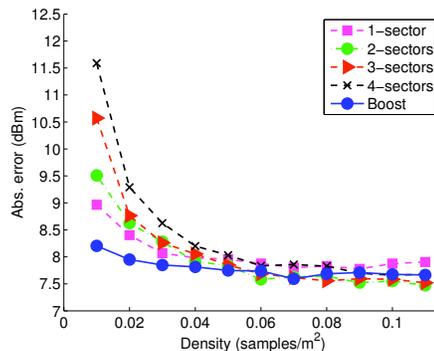


Figure 5: Sectorization Model

Formally, boosting combines the RSS predictions and training errors of multiple models as follows. For each region R and model M , we compute the average ($\mu_{R,M}$) and standard deviation ($\sigma_{R,M}$) training error. We then select the model M_{best} which minimizes $\mu_{R,M} + 2\sigma_{R,M}$ (i.e., the 95%-percentile of a normal distribution $\mathcal{N}(\mu_{R,M}, \sigma_{R,M})$). For regions which have no samples, the 1-class model is selected as a conservative choice, since it generally achieves good prediction accuracy with a small number of measurements.

As shown in Figure 4, the boosting procedure reduces the prediction error by as much as 8.8% over the log-normal model for the Jolley dataset, and as much as 9.7% for the Bryan dataset. More strikingly, its performance is consistently good across training sets of different sizes.

Summary: *The boosting procedure combines results from multiple models to achieve consistently good performance, independent of training data size.*

5.5 Impact of Sectorization

In the preceding models, we have ignored the effect of non-uniform radiation patterns. We will now explore the sectorization technique (Equation 3) that aims to improve the accuracy of the radio propagation model by modeling this effect when the automatic wall classification is used.

Figure 5 shows the prediction errors for models which consider both obstacles and directionality. In this figure, the number of wall classes is fixed at 2, and an increasing number of sectors are used. For comparison, we also include the results of the (nonsectorized) boost procedure described above.

We note that at densities lower than 0.06 samples/m²,

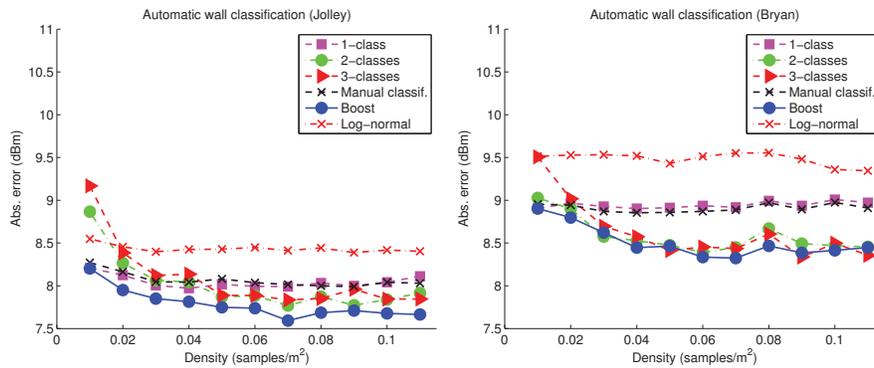


Figure 4: Automatic wall classification

adding more sectors *increases* the prediction error; the sectorization models outperform the boost approach only at densities > 0.08 samples/ m^2 . This phenomenon is caused by the greatly increased number of parameters needed for sectorization. Rather than solving for a single parameter α , it now necessary to solve for up to 112 values for Jolley Hall and up to 256 values for Bryan Hall². We emphasize that the lower densities are the most useful for radio mapping, since they represent less data that must be sampled.

Finally, we wish to explore how much decay, walls, and directionality actually contribute to wireless coverage. Figures 6(a) and 6(b) plot the impact of distance and walls for a representative wall class model. As the predicted RSS decreases, both distance and walls contribute more in absolute terms to the attenuation; this makes intuitive sense, since high-RSS links tend to be shorter and pass through fewer walls. This effect is even more striking when considering the relative contribution. For links with high RSS, the impact of walls may be as low as 5%. For the critical links close to the coverage boundary, walls contribute to up to 25% of the overall signal loss.

To evaluate the impact of directionality on RSS, we oriented the sender in each of the four compass directions and sent a number of packets to a fixed receiver. Figure 6(c) presents the difference in RSS relative to when the sender is pointed North. We observe differences in the 25th, 50th, and 75th percentiles of about 2.5 dBm, 5 dBm, and 8.5 dBm, which are consistent with those observed in [24]. For links in the critical transitional region, the impact of attenuation through obstacles has twice the impact of directionality. Thus, we choose models which ignore the impact of directionality; it is more effective to use the limited training data to better fit the wall classification parameters, which are fewer and have a greater impact on signal strength near the coverage boundary.

Summary: *Antenna orientation has a smaller impact than wall attenuation for links on the boundaries of coverage regions. Moreover, sectorization techniques are suitable only when large training data sets are available.*

6. RADIO MAPPING TOOL

In this section, we present our Radio Mapping Tool (RMT) for assessing network coverage. RMT is particularly

²In fact, at 5 sectors there would be more unknowns than experimental points.

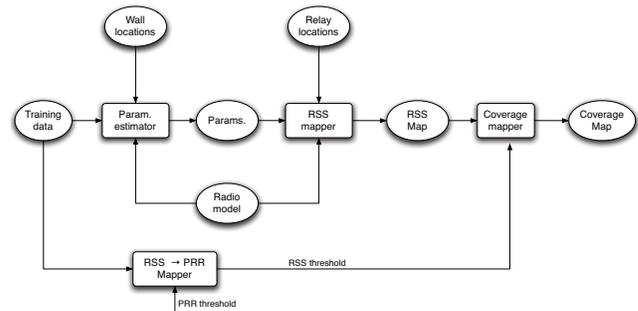


Figure 7: Radio Mapping Tool

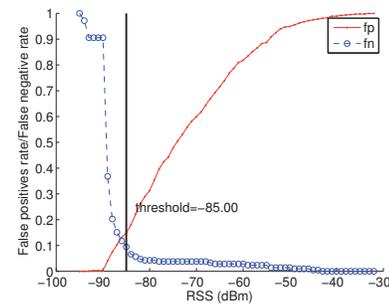


Figure 8: Selecting the RSS threshold

beneficial for applications which require a network to cover an entire physical area. The main use case of RMT is to evaluate the coverage of an already deployed network; this can be done by computing the union of the regions covered individually by each relay. RMT may also be used to assist during the initial network deployment: an overly dense network of relays is temporarily deployed to measure network coverage, and only those necessary to cover the area are permanently installed.

RMT has several salient features. (1) In contrast to ray-tracing techniques, RMT does not require the user to specify the attenuation coefficients or construction materials of walls. Wall locations may be extracted from readily available floor plans. (2) Based on our insights from the previous sections, RMT uses the wall classification models, which have been shown to provide accurate predictions even when few measurements are used for training. RMT combines

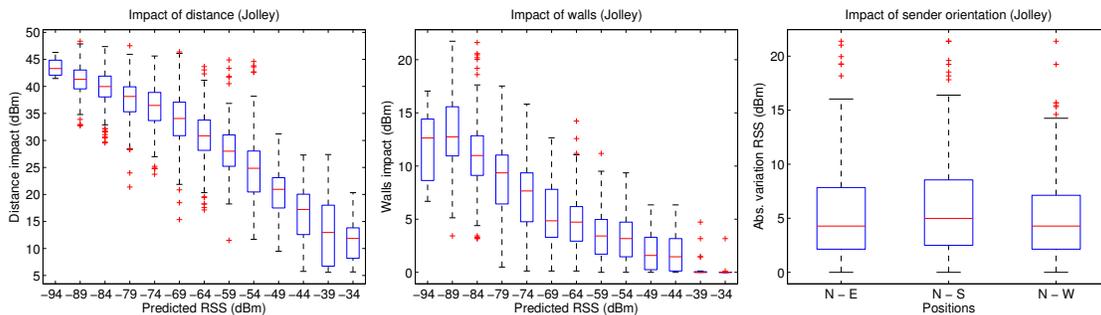


Figure 6: Impact of distance, wall attenuation, and antenna orientation

these results through the Boost procedure. (3) RMT uses the computationally efficient algorithm presented in Section 4 to classify each wall into a small number of classes and determine the model parameters.

As input from the user, RMT requires the physical locations of the relays whose coverage is being predicted, the locations of walls, a set of training data, and a PRR threshold (PRR_t) that determines “good” and “bad” links. RMT includes a TinyOS application to be deployed on a single “beacon” mote, which broadcasts beacon packets at each of the 8 power levels supported by the CC2420 at places where the user wishes to collect link quality data. By design, the user does not need to collect data exhaustively; our case study in Section 7 used as few as 0.01 samples/m². A corresponding TinyOS application on the relay nodes records the RSS and sequence number of all successfully decoded packets, which RMT uses to compute the PRR and average RSS of each link.

RMT has four main components, as shown in Figure 7: a *Parameter Estimator*, an *RSS Mapper*, an *RSS-to-PRR Mapper*, and a *Coverage Mapper*. The *Parameter Estimator* uses the computationally efficient algorithm described in Section 4 to estimate the model’s parameters. RMT fits models with 1–4 different wall types and merges their prediction using the Boosting procedure previously discussed. Based on the determined parameters, the *RSS Mapper* constructs a map of signal strength predictions on a dense 2D grid overlaid on the floor plan. A map is individually computed for each relay using the automatic wall classification model.

The *RSS-to-PRR Mapper* determines an RSS threshold which accurately separates the “good links” (with $PRR \geq PRR_t$) from the “bad links” (with $PRR < PRR_t$). To do so, we leverage the correlation between RSS and PRR previously observed in [20]. Any RSS threshold will necessarily have some false negative rate (i.e., links incorrectly predicted as poor-quality) and false positive rate (i.e., links incorrectly predicted as good-quality). An example of this phenomenon is shown in Fig. 8, where the PRR threshold has been set to 80% and the false positive and false negative rates have been calculated for each possible RSS threshold. In this example, an RSS threshold of -85 dBm offers the best tradeoff: the false positive rate is 9% and false negative rate is 15%. RMT allows the user to specify the maximum acceptable false negative and false positive rates; the *RSS-to-PRR mapper* then automatically locates the minimum RSS which satisfies both of these criteria in the training data, or reports an error when no such RSS threshold ex-

ists. We note that from the perspective of coverage prediction, it is desirable to have a low false positive rate, and it should arguably be configured conservatively. Nevertheless, an overly conservative threshold could result in overdeployment, which increases the monetary cost of the deployment and may increase channel contention. We note that the *RSS-to-PRR Mapper* is designed under the assumption that the noise levels observed during the measurements are representative of normal network operations. This could be improved with a better model of background noise, such as the one proposed in [12].

Finally, the *Coverage Mapper* uses this threshold to convert the RSS map into a binary coverage decision at each grid point; Figure 9 shows a sample output. By precomputing as much data as possible (e.g., the walls between each relay and grid location), RMT can train the models and make predictions within minutes.

7. EMPIRICAL EVALUATION OF RMT

In this section, we analyze the performance of the RMT on the data sets previously collected for our empirical study. We begin by assessing RMT’s performance through a case study which highlights RMT’s accuracy and the intuitive nature of the outputted radio maps.

We characterize the accuracy of RMT’s coverage predictions by its resulting false positive and false negative rates. In contrast to the previous section, the false positive and false negative rates discussed here refer to the prediction coverage rather than the RSS threshold. In this context, a false positive occurs when RMT predicts coverage where there is none; similarly, a false negative occurs when RMT predicts no coverage but ground truth data indicates otherwise.

7.1 Representative Example

The case-study is designed to emulate the use of RMT to predict the coverage of one relay in Jolley Hall. In order to illustrate the efficacy of our automatic wall-classification model, we present results with the normal RMT (which uses the automatic wall classification model with Boost) as well as with a version of RMT that has been modified to use the basic log-normal model. To highlight RMT’s accuracy when using only a small amount of training data, we choose a sampling density of 0.02 samples/m². The data is divided into training and testing sets through the same sampling strategy described in Section 5. For the purposes of this study, we define a “good link” to have a PRR higher than

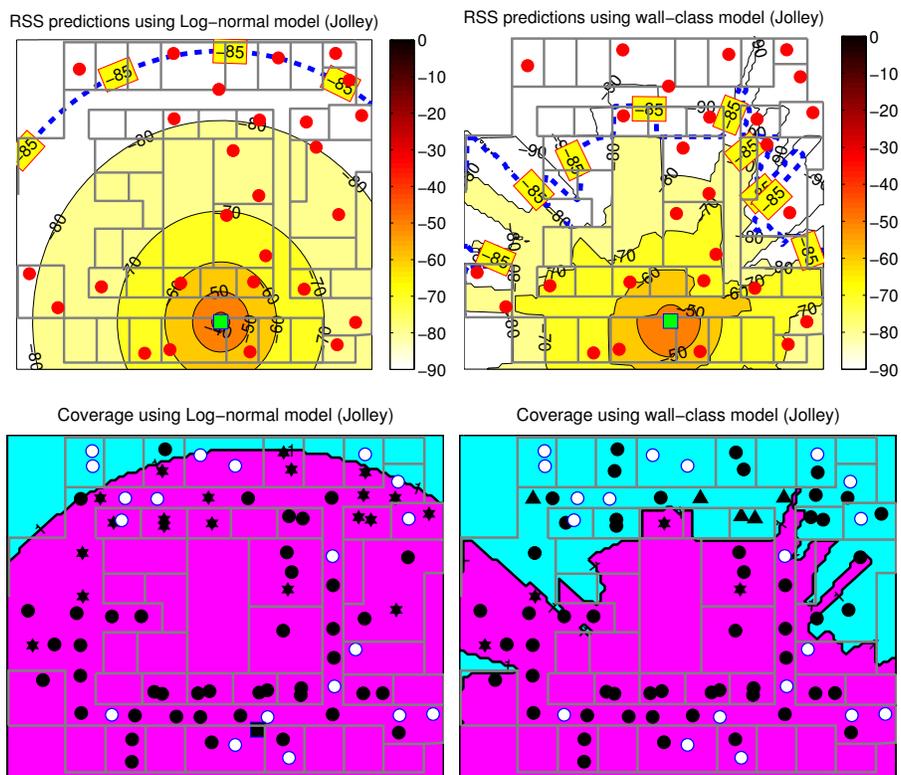


Figure 9: Example predictions using Radio Mapping Tool

80%. Using the RSS threshold selection technique previously discussed, a RSS threshold of -85 dBm was selected.

Figure 9(a) plots the RSS predictions when the log-normal model is used. Since the log-normal model does not account for wall attenuation, the contour graph consists of concentric circles. RMT also plotted the -85 dBm line that delineates the relay’s coverage area. It is worth highlighting the small number of samples which was used for training, shown as red dots. Moreover, we have only a few measurements in each room; other radio mapping techniques require every wall to be independently measured. Figure 9(c) shows the RSS predictions made by RMT using Boost. The predictions clearly indicate the strong impact of walls; the finger-like projections are caused by signals passing through different numbers of walls.

Figure 9(b) shows the coverage map for the log-normal model. Here, the white circles represent training data, the black circles indicate correct predictions, and the stars and triangles denote false positives and false negatives, respectively. Figure 9(d) plots the corresponding coverage map predicted when using Boost.

The log-normal model has 20 false positives, which are particularly disconcerting since they indicate coverage in regions where there is actually none. In contrast, RMT reduces the number of false positives from 20 to 4. This is particularly clear toward the top of the predicted coverage area, where the coverage area stops at the intersection with a wall. We note that most of these 4 false positive locations are close to the predicted coverage border. We expect that the coverage prediction could be further improved by targeted sampling near the border. This highlights the use of

RMT as an interactive tool to guide the user about where to collect additional coverage measurements.

Due to the log-normal model being overly optimistic, it predicts 1 false negative compared to 5 for the wall-based model. We note that some false negatives should be expected, since a threshold of -85 dBm leads to a 15% false negative rate when mapping RSS to PRR.

7.2 Detailed Empirical Results

Next, we statistically analyze RMT’s overall performance on both buildings’ data sets. For this analysis, we fix the PRR threshold to 80% and set the RSS false positive rate threshold to 10% for both data sets. This resulted in an RSS threshold of -85 dBm and -87 dBm for Jolley and Bryan, respectively. Additionally, we varied the sampling density and observed the impact on RMT’s performance.

Figure 10 plots the false negative and false positive rates for RMT when using the log-normal and Boost-based models. First, consider the results from the Jolley dataset. As seen in the case study, the log-normal model suffers from numerous false positives, with a false positive rate of 38%–46%. In contrast, RMT using the Boost procedure has a false positive rate between 23%–27%, representing a reduction in the false positive rate by up to 39% at a density of 0.01.

Again, the log-normal model achieves the lowest false positive rate (6.8%) by incurring a high false negative rate. The false negative rate for RMT was 10%–12%, which is comparable with the 10% false positive rate imposed on the RSS threshold. As previously noted, a moderate increase in false negatives may be acceptable, since it would result in

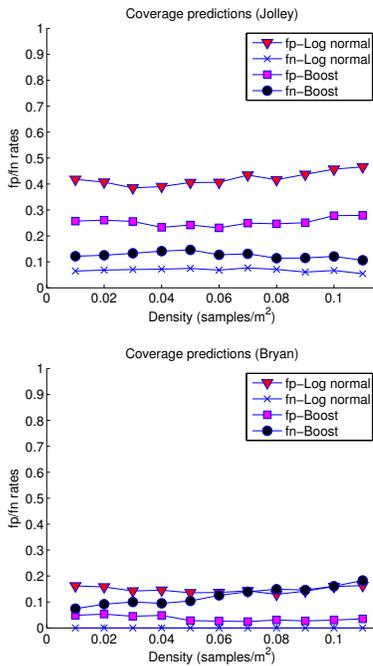


Figure 10: Coverage prediction accuracy

a slightly denser network.

The results from the Bryan data set paint a similar picture. Using the Boost procedure gives RMT a significantly lower false positive rate than the log-normal model. When the sampling rate is 0.01 samples/m², the Boost-based method reduced the false positive rate by as much as 54% (from 16.21% down to 7.42%). At the same sampling density, the Boost-based method has false negative rate of 16%, compared to a false negative rate of 0% for the log-normal model. Again, the false negative of zero occurs because the log-normal model significantly overestimates coverage area.

We conclude that our Boost-based approach may reduce false positive rates by as much as 54%, and achieved a false negative rate comparable to the user-specified constraints on the RSS threshold. Moreover, despite the two different radio propagation characteristics, RMT with Boost achieved consistently good performance across two different buildings.

8. CONCLUSION

Radio mapping is a challenging problem for real indoor environments due to signal attenuation through walls, complex signal propagation behavior, and the need to reduce the number of sampling measurements. This paper addresses this important challenge by developing a practical and effective radio mapping approach for indoor environments.

We first perform an in-depth empirical analysis of several signal propagation models in an office building. Our analysis shows the importance of balancing the accuracy of the model against the number of model parameters that need be estimated based on limited measurement. Our empirical results identify the wall-classification model family as the most practical and effective for indoor environments.

We then propose a practical algorithm to predict the RSS between different locations based on a small number of measurements. A key novelty of our algorithm lies in its ability

to automatically classify walls into a small number of classes with different degrees of signal attenuation, and to automatically select the best number of wall classes on a per-region basis. Empirical results show that our automatic wall classification scheme results in more accurate RSS prediction than a manual classification based on architectural knowledge.

We have developed a practical Radio Mapping Tool to predict the radio coverage of relay placements. RMT has several salient features. (1) It requires minimal information about the indoor environment. The only knowledge about the environment that RMT needs are the wall locations, which may be extracted from existing floor plans. (2) RMT can accurately predict radio coverage based on a small number of measurements, which can significantly reduce the cost of network deployment and maintenance. (3) RMT features computationally efficient algorithms that allow users to quickly assess and adjust the coverage of a potential relay placement.

An empirical evaluation in two office buildings shows that RMT achieves as much as 54% fewer false positives compared to the log-normal model based on a sampling density of only 0.01 samples/m². Our results demonstrate that RMT is a practical tool which can be used to facilitate the efficient deployment and robust operation of wireless sensor networks for indoor environments.

Acknowledgment

The project is supported by UL1RR024992 from the National Center For Research Resources. The content is the responsibility of the authors and does not necessarily represent the official views of the NCRP or the NIH. Additional funding provided by NSF NeTS-NOSS Grant CNS-0627126 and CRI Grant CNS-0708460.

9. REFERENCES

- [1] J. Andersen, T. Rappaport, and S. Yoshida. Propagation measurements and models for wireless communications channels. *IEEE Communications Magazine*, 1995.
- [2] R. Bernhardt. Macroscopic diversity in frequency reuse radio systems. *IEEE Journal on Selected Areas in Communications*, 1987.
- [3] Y. Chen and A. Terzis. Calibrating rssi measurements for 802.15.4 radios. Technical report, John Hopkins University, 2009.
- [4] O. Chipara, C. Brooks, S. Bhattacharya, C. Lu, R. Chamberlain, G.-C. Roman, and T. C. Bailey. Reliable real-time clinical monitoring using sensor network technology. In *AMIA*, 2009.
- [5] M. D’Souza, T. Wark, and M. Ros. Wireless localisation network for patient tracking. In *ISSNIP*.
- [6] S. J. Fortune, D. M. Gay, B. W. Kernighan, O. Landron, R. A. Valenzuela, and M. H. Wright. Wise design of indoor wireless systems: Practical computation and optimization. *IEEE Computational Science and Engineering*, 1995.
- [7] D. Ganesan, B. Krishnamachari, A. Woo, D. Culler, D. Estrin, and S. Wicker. Complex behavior at scale: An experimental study of low-power wireless sensor networks. Technical report, UCLA Computer Science Department, 2002.
- [8] H. Hashemi. The indoor radio propagation channel. *Proceedings of the IEEE*, 1993.
- [9] J. Hwang, T. He, and Y. Kim. Exploring in-situ sensing irregularity in wireless sensor networks. In *SenSys*, 2007.

- [10] M. Iskander and Z. Yun. Propagation prediction models for wireless communication systems. *IEEE Transactions on Microwave Theory and Techniques*.
- [11] A. Krause, C. Guestrin, A. Gupta, and J. Kleinberg. Near-optimal sensor placements: maximizing information while minimizing communication cost. In *IPSN*, 2006.
- [12] H. Lee, A. Cerpa, and P. Levis. Improving wireless simulation through noise modeling. In *IPSN*, 2007.
- [13] S. Lin, J. Zhang, G. Zhou, L. Gu, J. A. Stankovic, and T. He. ATPC: adaptive transmission power control for wireless sensor networks. In *SenSys*, 2006.
- [14] J. McKown and R. Hamilton. Ray tracing as a design tool for radio networks. *IEEE Network*, 1991.
- [15] N. Reijers, G. Halkes, and K. Langendoen. Link layer measurements in sensor networks. In *IEEE MASS*.
- [16] J. Robinson, R. Swaminathan, and E. W. Knightly. Assessment of urban-scale wireless networks with a small number of measurements. In *MobiCom*, 2008.
- [17] K. Schaubach, N. Davis, and T. Rappaport. A ray tracing method for predicting path loss and delay spread in microcellular environments. In *IEEE Vehicular Technology Conference*, 1992.
- [18] S. Seidel, T. Rappaport, M. Feuerstein, K. Blackard, and L. Grindstaff. The impact of surrounding buildings on propagation for wireless in-building personal communications system design. In *IEEE Vehicular Technology Conference*, 1992.
- [19] D. Son, B. Krishnamachari, and J. Heidemann. Experimental study of the effects of transmission power control and blacklisting in wireless sensor networks. In *IEEE SECON*, 2004.
- [20] K. Srinivasan and P. Levis. RSSI is under appreciated. In *EmNets*, 2006.
- [21] T. Stoyanova, F. Kerasiotis, A. Prayati, and G. Papadopoulos. Evaluation of impact factors on rssi accuracy for localization and tracking applications. In *MobiWac*, 2007.
- [22] A. Wood, J. Stankovic, G. Virone, L. Selavo, Z. He, Q. Cao, T. Doan, Y. Wu, L. Fang, and R. Stoleru. Context-aware wireless sensor networks for assisted living and residential monitoring. *IEEE Network*, 2008.
- [23] G. Zhou, T. He, S. Krishnamurthy, and J. A. Stankovic. Impact of radio irregularity on wireless sensor networks. In *MobiSys*, 2004.
- [24] G. Zhou, T. He, S. Krishnamurthy, and J. A. Stankovic. Models and solutions for radio irregularity in wireless sensor networks. *ACM Transactions on Sensor Networks*, 2006.
- [25] M. Zuniga and K. Bhaskar. An analysis of unreliability and asymmetry in low-power wireless links. *ACM Transactions on Sensor Networks*, 2007.