AN ANALYSIS OF THE DOMESTIC POWER LINE INFRASTRUCTURE TO SUPPORT INDOOR REAL-TIME LOCALIZATION

A Thesis
Presented to
The Academic Faculty

by

Erich P. Stuntebeck

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Electrical and Computer Engineering

Georgia Institute of Technology
August 2010

AN ANALYSIS OF THE DOMESTIC POWER LINE INFRASTRUCTURE TO SUPPORT INDOOR REAL-TIME LOCALIZATION

Approved by:

Dr. Gregory D. Abowd, D.Phil.,
Advisor
College of Computing
Georgia Institute of Technology

Dr. Douglas Blough, Ph.D.
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Santiago Grijalva, Ph.D.
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Shwetak Patel, Ph.D.
Department of Computer Science and Engineering
University of Washington

Dr. George Riley, Ph.D.
School of Electrical and Computer Engineering
Georgia Institute of Technology

Date Approved: 23 June 2010

In memory of my grandfather, Adolf von Hake.

ACKNOWLEDGEMENTS

First and foremost I would like to thank my advisor, mentor, and friend Dr. Gregory Abowd. No one could ask for a better person to fill any of these rolls. I would also like to thank the rest of my thesis committee — Dr. Douglas Blough, Dr. Santiago Grijalva, Dr. Shwetak Patel, and Dr. George Riley — for their support and guidance that has shaped my work for the better.

None of this work would have been possible without my various sources of funding along the way. For that, I would like to thank my advisor, Gregory Abowd, as well as Marie Thursby, Margi Berbari, and Anne Rector of the TI:GER program. I would also like to thank the Georgia Tech Research Institute.

I would like to thank my collaborators in this work — Shwetak Patel, Tom Robertson, Matt Reynolds, Yi Han, Sidhant Gupta, Mayank Goel, Rahul Rajan, and Ramakrishnan Chandrasekharapuram. I also wish to thank the rest of the UbiComp Group, both past and present members, especially Mario Romero, Tae Jung Yun, Tracy Westeyn, Jay Summet, and Julie Kientz, for their friendship and support.

I have been fortunate to have had many wonderful internships over the course of my graduate career and to have worked with many great people in the process. I would especially like to thank John Davis and Maria Ebling at IBM Research, as well as Iqbal Mohomed and Doug Terry at Microsoft Research for their guidance.

This work has required a considerable amount of equipment to accomplish, much of which was generously loaned to me along the way. For that I would like to thank Eyal de Lara and the Computer Systems & Networks group at the University of Toronto, as well as the Georgia Tech Research Institute. I also wish to thank Brian Jones at the Georgia Tech Aware Home for providing me with the space necessary

for the experiments that are the basis of this thesis.

Last, but not least, I would like to thank my family for their love and support along the way and for always believing in me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

PLP — PowerLine Positioning

WPLP — Wideband PowerLine Positioning

RTLS — Real-time Locating System

IFRTLS — Indoor Fingerprinting-baed Real-time Locating System

CDF — Cumulative Distribution Function

ADC – Analog-To-Digital Converter

NCO – Numerically Controlled Oscillator

CIC – Cascaded Integrator-Comb

FFT – Fast Fourier Transform

DFT – Discrete Fourier Transform

KNN – K-Nearest-Neighbors

RSSI – Received Signal Strength Indicator

TDOA – Time Difference of Arrival

ISM – Industrial, Scientific, and Medical

LPU – Location Processing Unit

AoA – Angle of Arrival

SNR – Signal-to-Noise Ratio

HMM – Hidden Markov Model

SUMMARY

The vision of ubiquitous computing is to seamlessly integrate information processing into everyday objects and activities. Part of this integration is an awareness on the part of a system of its user's context. Context can be composed of several variables — such as a user's current activity, goals, or state of mind — but location (both past and present) is almost always a key component.

Determining location *outdoors* has become quite simple and pervasive with today's low-cost handheld Global Positioning System (GPS) receivers. Technologies enabling the location of people and objects to be determined while *indoors*, however, have lagged due to their extensive infrastructure requirements and associated cost. Just as GPS receivers utilize radio signals from satellites to triangulate their position, an indoor real-time locating system (RTLS) must also make use of some feature of the environment to determine the location of mobile units. Since the signal from GPS satellites is not sufficiently strong to penetrate the structure of a building, indoor RTLS systems must either use some existing feature of the environment or generate a new one. This typically requires a large amount of infrastructure (e.g. specialized RF receivers, additional 802.11 access points, RFID readers, etc.) to be deployed, making indoor RTLSs impractical for the home. While numerous techniques have been proposed for locating people and objects within a building, none of these has yet proven to be a viable option in terms of cost, complexity of installation, and accuracy for home users.

This dissertation builds on work by Patel *et al.* in which the home power lines are used to radiate a low-frequency wireless RF signal that mobile tags use for location fingerprinting. Leveraging the existing power line permits this system to operate on

far less additional infrastructure than existing solutions such as cellular (GSM and CDMA), 802.11b/g, and FM radio based systems.

The contributions of this research to indoor power line-based RTLS are threefold. First, I examine the temporal stability of a power line based RTLS system's output. Fingerprinting-based RTLS relies upon some feature of the environment, such as the amplitude of an RF signal, to be stable over time at a particular location (temporal stability), but to change in space (spatial differentiability). I show that a power line-based RTLS can be made much more resistant to temporal instability in individual fingerprint components by utilizing a wide-band RF fingerprint. Next, I directly compare the temporal stability of the raw features used by various fingerprinting based indoor RTLSs, such as cellular, 802.11b/g, and FM radio. In doing so, I show that a power line based indoor RTLS has an inherent advantage in temporal stability over these other methods. Finally, I characterize the power line as a receiving antenna for low-powered wireless devices within the home, thus allowing the power line to not only transmit the RF signals used for fingerprinting, but also to receive the sensed features reported by location tags. Here, I show that the powerline is a viable receiver for these devices and that the globally available 27.12 MHz ISM band is a good choice of frequency for communications.

CHAPTER 1

INTRODUCTION AND MOTIVATION

The vision of ubiquitous computing is to seamlessly integrate information processing into everyday objects and activities [91]. Part of this integration is an awareness on the part of a system of its user's context. Context can be composed of several variables — such as a user's current activity, goals, or state of mind [54] — but location (both past and present [23]) is almost always a key component.

The number of compelling location-based services and context-aware applications is seemingly infinite. GPS-based automatic routing and navigation is probably the best known and most widely used today. Location-based search is now possible on most current smart phones — searching Google for "pizza" on an iPhone automatically returns the closest pizza restaurants, not just generic search results containing the term. Some phones are now offering augmented reality-based applications that label live images of the surrounding environment, identifying, for example, restaurants within view along with their rating. The ability to determine location *outdoors* has become quite simple and pervasive with today's low-cost handheld Global Positioning System (GPS) receivers, which has prompted widespread adoption of these applications.

Location-based applications are not limited to outdoor environments, however. In the home, one can imagine a system that automatically finds lost objects such as car keys or remote controls. Another possibility is automated lighting and HVAC control based on user presence. For the elderly or those with disabilities, indoor localization could be utilized to gather mobility data that may be of use to healthcare professionals [72]. In commercial buildings, interactive building maps showing a person's

current location within the building could be used to navigate unfamiliar spaces. The location of high value assets could be automatically tracked for inventory and / or security purposes. Nursing homes could track the location of patients suffering from dementia. Unfortunately, determining location *indoors* is not as simple as outdoors, which has all but prevented applications like these from being adopted. GPS, which makes outdoor localization simple, inexpensive, and ubiquitous, cannot operate indoors as the satellite signals it relies on are not sufficiently strong to penetrate building structures. Thus, alternate methods are required for indoor localization.

Indoor positioning has been a very active area of research in ubiquitous computing over the past decade [43]. Some of the first indoor solutions introduced new infrastructure to support localization [41, 42, 74, 75, 76, 89, 90]. Just as GPS receivers utilize an environmental feature to determine their location — radio signals from satellites — an indoor real-time locating system (RTLS) must make use of some feature of the environment to determine the location of mobile units [60]. The deployed infrastructure serves to generate this feature. Despite some success, as indicated by commercialized products [13, 14, 19, 34], the cost and effort of installation are a major drawback to wide-scale deployment, particularly in domestic settings. Thus, technologies enabling the location of people and objects to be determined while *indoors*, as well as their associated applications, have experienced poor adoption.

Given this, a trend in more recent location-based systems research has been reuse of existing infrastructure to ease the burden of deployment and lower the cost. The earliest demonstrations leveraged WiFi access points [27, 59], and shortly afterwards researchers began to explore Bluetooth [65] and wireless telephony infrastructure, such as GSM [38, 58, 69]. A concern of these methods, however, is that individuals may not have control over the characteristics of this infrastructure and the operational parameters of the infrastructure may change without warning. This would result in the need to recalibrate, which can be a time consuming process. Additionally,

the existing infrastructure, particularly in the case of WiFi, may be insufficient to support localization. Individuals may opt to install additional WiFi infrastructure in the home, however this can be impractical due to the number of base stations required for effective localization being much greater than the one or two that cover a typical home for wireless data.

The desire to both control the infrastructure and to scale inexpensively by utilizing as minimal an infrastructure as possible inspired Patel *et al.* to develop an indoor positioning system called PowerLine Positioning (PLP) that leveraged the power line infrastructure in a home [73]. The system requires only two transmitter modules to be installed and compares favorably to GSM and WiFi-based localization systems. PLP was the first truly practical indoor localization system for a domestic environment. The appeal of this approach, wherein existing infrastructure in the home was utilized in order to reduce the system's deployment burden and cost, inspired me to obtain a deeper understanding of the power line-based localization approach and to address limitations of the initial work.

1.1 Purpose of Research and Thesis Statement

The purpose of this research is to explore the use of in-wall electrical wiring in a home as the enabling infrastructure for a simple-to-deploy indoor real-time locating system (RTLS). This dissertation builds on Patel *et al.*'s PLP system, in which the home power lines were used to radiate a low-frequency wireless RF signal that mobile tags use for location fingerprinting [73]. Location fingerprinting is a localization approach which relies upon some feature of the environment, such as the amplitude of an RF signal, to be stable over time at a particular location (temporal stability), but to change in space (spatial differentiability). A signal map is made of the deployment space post-deployment either empirically through a measurement process known as the *site survey*, or via models, such as an RF propagation model. PLP suffers from

two major drawbacks that may prevent its acceptance by domestic users — temporal instability of the fingerprint features and the need for a complex mobile tag with multiple radios. The overall objective of this research is to address these drawbacks.

1.1.1 Thesis Statement

This dissertation addresses these concerns in the context of the following thesis statement:

> *A domestic fingerprinting-based indoor RTLS using the amplitude of power line radiated low-frequency RF signals as its fingerprint will: (1) benefit in accuracy, reduced sensitivity to short-term noise, and increased long-term temporal stability through the use of a wideband fingerprint as compared to a narrow-band frequency pair approach; (2) exhibit superior temporal stability of its underling fingerprint components as compared to WiFi, cellular, and FM based solutions; and (3) be able to utilize the power line as a receiving antenna as well as a transmitting antenna, thus allowing both forward and back-channel communications to occur over the power line.*

1.2 Research Questions

In this dissertation, I address the following research questions:

- Can the accuracy, sensitivity to short-term noise, and long-term signal map instability of the PowerLine Positioning indoor fingerprinting-based real-time locating system (IFRTLS) all be improved in a domestic environment through the use of a multi-frequency wideband fingerprint rather than a narrow-band fingerprint?
- Of the four primary wireless technologies around which fingerprinting-based indoor RTLSs have been built (WiFi, cellular, FM, low-frequency power line

RF), which technology's underlying signals exhibit the greatest temporal stability over time as defined by the lowest standard deviation in the measured feature (e.g. signal strength)?

- Can the power line, which already acts as a *transmitting* antenna to generate the features used by PowerLine Positioning, also serve as a *receiving* antenna for low-powered RF transmissions originating within a home, thus eliminating the need for a separate wireless backchannel over which the mobile location tags communicate with the central Location Processing Unit (LPU)?

1.3 Dissertation Overview

This dissertation makes three primary contributions to the field of indoor real-time localization. I begin in Chapter 3 by revisiting the results of the PLP system in an environment that experiences considerable electrical and RF noise — the Georgia Tech Aware Home. Fingerprinting-based RTLS relies upon some feature of the environment, such as the amplitude of an RF signal, to be stable over time at a particular location (temporal stability), but to change in space (spatial differentiability). I show that in noisy environments like the Aware Home, the performance of PLP, both initially and over time, can be much worse than expected based on the original results due to temporal instability of the underlying fingerprint components. I also show that due to the transient and unpredictable nature of RF noise, it can be difficult, if not impossible, to select *a priori* a pair of RF frequencies for the fingerprinting features of PLP that will remain stable over time. I then identify a wideband fingerprint as the solution to transient and long-term instability in the fingerprint features, and show that a wideband fingerprint both improves the accuracy of the system's location estimates both initially, as well as several months post-deployment.

In Chapter 4, I directly compare the temporal stability of the raw features used by various fingerprinting-based indoor RTLSs, such as cellular, WiFi, and FM radio.

Week-long measurements were made at three different locations — two single-family homes, and one apartment. I present an analysis of the fingerprint stability over the entire test period and show that certain technologies, particularly WiFi and GSM, exhibit large fluctuations in the features being fingerprinted, which translates directly into errors in a location system's estimates. I then present a grid-level analysis for each technology, wherein the sensitivity to noise is tested across an entire spatial grid for each. The results of these experiments and analyses show that a power line-based indoor RTLS has an inherent advantage in temporal stability over other systems which utilize alternate features.

Finally, in Chapter 5, I characterize the power line as a receiving antenna for low-powered wireless devices within the home, thus allowing the power line to not only *transmit* the RF signals used for fingerprinting, but also to *receive* the sensed features reported by location tags. In the initial PLP work, location estimates were computed at a central LPU which stored the signal map. Mobile location tags sensed the power line radiated RF signals and reported the sensed features wirelessly to the LPU using a short-range RF protocol such as ZigBee. This required the tags to have both a power line radio as well as a secondary radio for the backchannel. Utilizing the power line as a receiver allows these tags to use a single radio for both sensing the features needed for fingerprinting, as well as transmitting those sensed features to the LPU. This saves on the cost of these tags, reduces their size, and improves power efficiency. This research has much broader implications, however, due to the observed efficiency of the power line as a receiver of low-powered in-home wireless communications. I show that not only is the power line a viable receiver for transmissions from mobile location tags, but also that it is extremely efficient at doing so, thus enabling communication at practical ranges with extremely low powered radios never before seen. I show that the globally available 27.12 MHz ISM band is a good choice of frequency for communications.

CHAPTER 2

BACKGROUND AND RELATED WORK

Given the plethora of compelling applications that could utilize indoor location data, there is a large body of existing work on indoor localization. The space has remained a largely unsolved problem, however, with respect to a reliable, cost-effective, and easy to deploy system. This is especially true for indoor localization in a domestic environment. This chapter provides a brief background on localization and an overview of work to date in the field, focusing primarily on fingerprinting-related work. I begin in Sec. 2.1 with a discussion of various techniques for localizing objects, the types of labels that a location system can utilize for its output, and the various levels of location granularity that are possible. I then discuss in Sec. 2.2 several "legacy" indoor RTLSs, which I define as systems not utilizing a fingerprinting technique. In Sec. 2.3, I introduce the concept of fingerprinting in more detail. Finally, in Sec. 2.4, I discuss some of the many existing fingerprinting-based indoor RTLSs. For additional detail, the reader is referred to the more extensive surveys by LaMarca and de Lara [60], Varshavsky and Patel [55], and Hightower and Borriello [43].

2.1 Background on Localization

This section describes the techniques that can be used for localization, the types of output — or labels — that a location system can provide, and the different levels of location granularity that are possible. The terminology introduced in this section will be used in Sections 2.2 and 2.4 to classify existing indoor RTLSs.

2.1.1 Location Representation

2.1.1.1 Granularity

The granularity of a location system is the smallest change in location the system is able to reliably detect. For example, civilian GPS units are generally accurate to within slightly over five meters [60]. While it is generally desirable for indoor location systems to be more accurate — for example, an error of 5 m indoors could place a person in either the kitchen or the bedroom, two rooms with very different associated contexts — it is important to keep the end application in mind when considering performance. For example, a home automation system utilizing a user's location to control lighting and climate based on presence within a room need only know what room a person is in. For this application, a *room-level* symbolic output location system would be adequate. More complex applications, however, such as an augmented reality based game, may require more precise location data (e.g. *meter-level*, or even *sub-meter-level*) as well as orientation. Other applications, such as finding lost objects within a building, can benefit from an intermediate level of accuracy, such as *sub-room-level* wherein a room is broken up into several regions.

2.1.1.2 Labeling

The output of a location system can be absolute, relative, or symbolic. GPS is an example of a system that gives absolute coordinates as output — e.g. (41.702° N, 86.248° W). *Absolute* coordinates define a specific position on the Earth. *Relative* coordinates are specified relative to another location — systems that use coordinates specified on a grid, wherein locations are all relative to the arbitrarily selected origin point of the grid, are an example. Finally, a system with *symbolic* output defines locations with arbitrarily selected symbolic labels — e.g. kitchen, living room, bedroom, etc.

2.1.2 Localization Techniques

Determining the location of a person or object, whether indoors or outdoors, is most commonly accomplished via triangulation, scene analysis, or proximity. Dead reckoning is another method, although it requires knowledge of a starting location. This section provides a brief overview of each — the reader is referred to Hightower and Borriello for a more comprehensive description of triangulation, scene analysis, and proximity [43]. For additional details on dead reckoning, the reader is referred to Woodman [94].

2.1.2.1 Triangulation

Triangulation computes the location of an object by measuring either the angle or the distance between it and several fixed reference points of known locations. At least three non-collinear reference points are necessary to compute the position of an object in two dimensions; four reference points are necessary for three-dimensional positioning. Triangulation with *distance* measurements is referred to as lateration. Distance between the unknown location can be measured in several ways, including a direct measurement, measurement of the time-of-flight of a wireless signal, or the attenuation of a wireless signal. Triangulation can also be performed by measuring the *angle* between three or more points. This procedure is referred to as angulation, and is typically achieved by measuring the angle-of-arrival (AoA) of signals emanating from the reference points. Triangulation is the most commonly known method as it is the method used by the GPS system.

Although triangulation works well outdoors with line-of-sight (unobstructed view between the transmitter and receiver) radio propagation, it is generally unsuitable for indoor localization due to the unpredictable non-line-of-sight propagation within a building. While an approximation of radio wave propagation outdoors can be relatively easily obtained with the free-space path loss model [79], indoor propagation

is difficult to predict due to the complex interactions radio waves have with the building structure, furniture, people in the environment, etc. This produces effects such as reflection, refraction, and scattering which can be extremely difficult to model.

2.1.2.2 Proximity

Proximity to a device of known location can also provide location information. This can include, for example, physical contact with a device such as a touch sensor, logging in at a computer terminal, or scanning one's ID card at an access control system. Proximity can also include sensing the presence or absence of wireless access points or Bluetooth beacons. A simple example of a proximity-based localization system would be if one's laptop were programmed to detect "Home" and "Office" locations based on the differently named wireless access points at these locations. A proximity-based system's precision is a function of the range at which beacons can be sensed, making solutions such as the home/office one largely unsuitable for precise indoor localization within a single building.

2.1.2.3 Scene Analysis

Scene analysis is a method of determining location by observing the features of a scene. The observed "scene" could be visual [32, 36], auditory [29], or even electromagnetic [27]. Scene analysis can be either *static*, in which location is determined by looking up the observed features in a pre-determined dataset that maps features to location, or *differential*, in which location is tracked by observing the difference between successive scenes [43]. Static scene analysis is often referred to as *fingerprinting*. Additional details on fingerprinting are provided in Sec. 2.3.

2.1.2.4 Dead Reckoning

Dead reckoning is a process by which a person or object's location is tracked via continual updating of a known initial location through speed and heading information.

Table 1: Overview of several "legacy" indoor RTLSs.

Name	Technology	Location Type	Indoor Accuracy	Infrastructure Requirements	Ref.
ActiveBadge	Infrared	Symbolic	Room-level	IR sensors and custom tag	[89]
CRICKET	Ultrasound and RF	Absolute	3 cm @ 90%	Ultrasound transmitters and receivers	[74]
ActiveBat	Ultrasound and RF	Absolute	3 cm @ 90%	Ultrasound transmitters and receivers	[90]
UbiSense	Wideband RF	Absolute	15 cm @ 90%	Wideband RF sensors and tags	[13]
MERIT	RF	Symbolic	Room-level	900 MHz wireless sensors	[61]
AirBus	Air pressure	Symbolic	Room-level, 88%	Pressure sensor in HVAC unit	[71]
SpotON	RFID	Relative	3 m	RFID readers	[45]
LANDMARC	RFID	Relative	1.75 m @ 95%	RFID readers, reference tags	[64]

A widely used application of dead reckoning is in automotive GPS navigation units, which utilize this approach when driving through a tunnel and the satellite signal is lost. Positions given by dead reckoning are necessarily relative to the starting position, which must either be known or determined in some way. Since dead reckoning can only calculate relative locations, the method is subject to substantial error over time since errors are cumulative [94].

2.2 Legacy Indoor RTLSs

In this section, I describe several *legacy* RTLSs, which I consider to be those systems not utilizing a fingerprinting-based approach to indoor localization. Table 1 summarizes the systems I discuss and provides accuracy information. This is by no means an exhaustive list — for additional details, the reader is referred to the more extensive surveys by LaMarca and de Lara [60], Varshavsky and Patel [55], and Hightower and Borriello [43].

One of the first indoor RTLSs was the Active Badge system, which is a proximity-based system that utilizes infrared badges and ceiling-mounted receivers [89]. The system provided room-level symbolic localization for each room where a receiver was installed. Badges emitted a unique infrared signature that allowed them to be individually identified. The Cricket system [74] operates on a similar principle, but uses ultrasound instead of infrared and uses the fixed infrastructure to transmit the ultrasound signal. Ubisense, a commercially available RTLS, uses triangulation via ultrawideband (UWB) RF signals [13]. Since the system requires precise time synchronization amongst each of the fixed UWB emitters, Ubisense requires the installation of timing cables.

The SpotON [45] and LANDMARC [64] systems used active RFID tags for localization. SpotON used a triangulation approach based on the signal strength of a mobile tag at multiple RFID readers (basestations). LANDMARC used a similar approach, however it introduced reference tags into the environment at known locations to aid in location estimation.

The MERIT system — a mesh of RF sensors for indoor tracking — utilized a multi-hop mesh network of wireless sensor nodes for location tracking [61]. This system provides room-level symbolic localization based on signal strength measurements from nearby sensor nodes acting as RF beacons. A proximity-based approach is utilized,

wherein the mobile node is considered to be at the location of the group of fixed-location sensors detecting it with the greatest signal strength. Although MERIT is easily installed, it can be difficult to maintain due to the large number of battery-operated sensor nodes.

All of these early indoor RTLSs suffered from the same problem — extensive infrastructure requirements. Infrastructure is costly, time consuming to deploy, and, for domestic environments, unsightly. Reducing the amount of required infrastructure was one of the major motivators for the latest method of indoor location tracking — fingerprinting.

2.3 Fingerprinting-based Indoor Localization

Fingerprinting, which has emerged as the leading method for indoor localization, relies upon some feature of the environment to be unique to each particular location — *spatial differentiability* — and to remain constant over time — *temporal stability*. Features observed by a mobile location tag are compared against a database of *fingerprints* — also called the *signal map* or, in the case of fingerprinting based on the properties of radio waves, the *radio map* — which maps feature properties to a particular location. Fingerprinting-based RTLSs (IFRTLSs) can vary in three primary ways:

- **Features** – the types of features utilized to construct the signal map.
- **Signal Map Creation** – the manner in which the signal map is generated.
- **Estimation Method** – the method by which locations are estimated from the signal map and live fingerprint samples.

I now discuss each of these in detail. The reader is referred to Kjærgaard for a more extensive taxonomy for classifying IFRTLSs that consists of eleven different taxons [52].

2.3.1 Fingerprinting System Variables

2.3.1.1 Features

Fingerprinting systems can vary in the types of features utilized to construct the signal map. A variety of environmental features can be used for localization, including visual features and features of the radio environment, such as the amplitude of a particular radio signal. RADAR [27], for example, uses the received signal strength indicator (RSSI) of WiFi basestations, while CILoS [85] uses the time offset of CDMA basestation pilot signals. Other possible features include basestation identifiers (e.g. the name of one or more WiFi basestations), signal-to-noise ratio (SNR), or packet response rate. Although signal strength has been the most popular feature to date, any feature which is temporally stable and spatially differentiable is a candidate for use in constructing the signal map of a RTLS.

2.3.1.2 Signal Map Creation

Once a feature is selected for use, the signal map of that feature can be constructed either *empirically* or with a *model*. Empirical signal map construction involves taking measurements at various locations throughout the deployment space. This process is referred to as the *site-survey*. Empirical measurements can be pre-processed in various ways to improve the precision and spatial resolution of the system [62], including methods such as interpolation between measured samples, aggregation of multiple samples at a single location, and removal of outliers. Model-based signal map construction utilizes signal propagation models to predict the properties of the feature of interest at each location within the deployment space. The signal map is estimated from these models. Although this method avoids the time-consuming site survey process required by empirical signal maps, it is generally significantly less accurate due to the complexities in accurately predicting indoor signal propagation.

2.3.1.3 Estimation Method

Fingerprinting-based systems can also differ in the method they utilize to estimate the position of a client. Systems can be either *deterministic* or *probabilistic*. Deterministic systems, such as PowerLine Positioning [73], consider only the value of the feature being used. In the case of PLP, this is the amplitude of two power line radiated RF signals. Probabilistic systems consider the value of a feature as part of a random process [31, 56, 96]. This typically involves measuring the variability in the feature at each location during the site survey. Nearly any machine learning technique or estimation method can be applied to location estimation. A popular choice for deterministic systems is the nearest-neighbor machine learning classifier [67], wherein the Euclidian distance in signal space is calculated between a live testing point and each point in the signal map. The point with the lowest distance is then selected as the estimated location. The K-Nearest-Neighbors (KNN) [67] approach can also be utilized, wherein the physical location of the K closest points in signal space to the testing point is averaged to produce a location estimate. Popular choices for probabilistic systems include Hidden Markov Models (HMMs) or Bayesian models.

2.3.2 Addressing Temporal Variations

One of the problems plaguing IFRTLSs is temporal variation in the signal map. If the signal map changes sufficiently, the accuracy of the system begins to degrade. The level of degradation depends on the type and number of features used, as will be discussed in more detail in Chapters 3 and 4. Signal map temporal variations can be either large-scale (across the entire deployment space) and prolonged — such as changes between day and night — or small-scale and transient — such as a change caused by a person walking through the space.

Several attempts have been made by the research community to date to address the issue of signal map temporal variation. One approach that has been examined is

the use of different radio maps for different times of day, such as for an office building that contains a large number of people during the day but is empty at night [28]. This of course requires repeating the time consuming site-survey process multiple times at different times of day. Another approach that has been examined is the use of adaptive radio maps based on monitoring reference points [95]. Although reference point monitoring can detect changes at the location of the reference point, it cannot address changes at other points. Since signal map perturbations may be localized, extrapolating changes at a reference point across the entire signal map is not necessarily an accurate approach. Probabilistic estimation approaches make use of variation in the signal, however the variation is typically small and around a fixed mean — these methods still cannot accommodate significant variation or prolonged shift in mean in the signal map.

2.4 Existing Fingerprinting-based RTLSs

Fingerprinting has been a primary focus of indoor RTLS research since the seminal work of Bahl *et al.* [27, 28]. The original inspiration for fingerprinting-based systems such as RADAR was to reduce or eliminate the need for infrastructure installation by utilizing some already existing feature of the environment. In the case of the RADAR system, this was 802.11b (WiFi) wireless access points. Bahl and Padmanabhan found the received signal strength (RSSI) of WiFi access points to be relatively constant in time at a single location, but to vary sufficiently in space so as to be useful information in determining the location of a WiFi device. They tested two approaches to signal map generation: empirical gathering of samples, and calculated samples based on an indoor radio propagation model. Three models for indoor radio propagation were tested: Rayleigh fading [40], Rician fading [77], and the Floor Attenuation Factor model [78]. They found that a modified version of the Floor Attenuation Factor

model, which they call the Wall Attenuation Factor model, provided the best performance. Using this model, they were able to achieve a median accuracy of 4.3 m, compared with 2.54 m using empirical samples for the signal map. Fingerprinting with an empirical signal map enables improved accuracy by taking into account the effects that buildings, solid objects, or people may have on an RF signal, such as reflection and attenuation. They utilized the K-Nearest-Neighbors (KNN) algorithm to match unknown fingerprint samples to samples in the signal map.

Since RADAR, there has been a considerable body of additional research on WiFi fingerprinting. LaMarca *et al.* [26] looked at reducing the site-survey burden through self-mapping based on sporadic availability of alternative localization methods such as GPS or Bluetooth/WiFi beacons. Ji *et al.* [46] also looked at reducing the site-survey burden, however they utilize a sparse initial signal map combined with a detailed floor plan of the environment. Lemelson *et al.* addressed this issue through a sparse empirical signal map collection combined with various fingerprint weighting schemes [62]. Various probabilistic approaches to WiFi-based localization have also been studied [68, 81]. Roos *et al.* achieve an accuracy of 2.76 m @ 90% using their probabilistic method and 10 WiFi access points [81]. Woodman and Harle looked at combining dead reckoning through foot-mounted accelerometers with WiFi-based fingerprinting [93].

Other work has looked at fingerprinting cellular telephone signals via both the amplitude of cellular towers [69, 86, 87, 88] as well as the time offset of signals from different towers [85]. Otsason *et al.* noted the benefit of a "wide" fingerprint utilizing the signal strength of up to 35 different GSM cells. Although cellular network based localization requires no additional infrastructure other than the mobile "tag", the approach suffers from the fact that the user/operator of the RTLS system is reliant upon a third party — the cellular telephone network operator — for the functionality of their system. Should the cellular towers change at all, the RTLS may be rendered

inaccurate and a new site survey may be required.

Recent work by Papliatseyeu *et al.* [70] uses the signal strength of portable low-powered FM radio modulators as the fingerprint for an indoor RTLS. They achieve accuracy of 4.5 meters at a confidence level of 95%.

The infrastructure mediated sensing (IMS) approach of Patel *et al.* seeks to combine the benefits of fingerprinting — low infrastructure requirements — with the benefit of the RTLS system operator having full control over the infrastructure. Their PowerLine Positioning (PLP) system [73] uses the power lines in a home as a transmitting antenna for distributing several low-frequency RF signals throughout a home. The amplitude of these signals comprises the fingerprint sensed by mobile tags. Signals are generated by two or more plug-in modules that inject into the power line. As I show in Chapter 3 however, this system can suffer from accuracy-reducing temporal instability in the fingerprint. These systems and others are summarized in Table 2.

Table 2: Overview of several existing fingerprinting-based indoor RTLSs.

Name	Technology	Feature	Location Type	Indoor Accuracy	Ref.
RADAR	WiFi (802.11b)	RSSI	Symbolic	4–5m @ 90%	[27, 28]
Probabilistic WiFi	WiFi (802.11b)	RSSI	Symbolic	2.76m @ 90%	[81]
GSM	GSM Cellular	RSSI	Symbolic	8–10m @95%	[69, 86, 87]
CILoS	CDMA	Time offset	Symbolic	18-23m @ 90%	[85]
FINDR	FM Radio	RSSI	Symbolic	4.5m @ 95%	[70]
PLP	Low-frequency power line RF	RSSI	Symbolic	3m @ 87–95%	[73]
WiFi-assisted Dead Reckoning	Foot-mounted accelerometers & WiFi fingerprinting	RSSI (WiFi)	Relative	0.73 m @ 95%	[93]

CHAPTER 3

WIDEBAND POWERLINE POSITIONING

In this chapter I present the Wideband PowerLine Positioning (WPLP) system, inspired by the PowerLine Positioning system of Patel *et al.* [73]. I analyze the performance of WPLP in terms of both its percentage accuracy for room, sub-room, and sub-meter grid-level localization, as well as the distance length of errors made by the system. I show that the temporal stability of a simple frequency pair fingerprinting approach used in the PLP work can be much worse than the data presented in the original publication. I then propose the use of a wideband fingerprint instead, and show that this not only improves the system's accuracy at all granularities of localization upon initial installation, but also makes the system more robust to changes in the fingerprint over time that can cause severe performance degradation in narrowband PLP.

3.1 Motivation

Ever since the seminal indoor fingerprinting-based real-time locating system (IFRTLS) work of RADAR [27], many examples of RF fingerprinting have shown how existing infrastructure for mobile communications, primarily WiFi and GSM, can be leveraged to provide location information. As discussed previously in Chapter 2, Patel *et al.*, motivated by the downsides of relying on infrastructures over which individuals typically have no control for a domestic IFRTLS, proposed fingerprinting of RF signals injected into and radiated by in-wall residential power lines [73]. In their PowerLine Positioning (PLP) work, they propose that the in-wall electrical infrastructure of a home acts as an antenna, and that artificially-generated signals injected into the power line and transmitted over this antenna create the spatially differentiable and

temporally stable signal map that is necessary for an IFRTLS to work.

While spatial differentiability of the features used for fingerprinting is clearly critical in an IFRTLS, temporal stability is just as important, and often is not examined as thoroughly. Changes over time in the fingerprint captured at a given location during the post-deployment site-survey result directly in inaccuracies when locating mobile tags. If the systems's accuracy becomes poor enough, its maintainer is forced to repeat the tedious and time-consuming site-survey process wherein fingerprints are gathered at various locations throughout the indoor space. This will severely limit any such system's acceptance in the home environment.

Temporal instability is a concern not only of PLP, but of all IFRTLSs. To date, most fingerprinting-based indoor localization work has been carried out with evaluation datasets obtained within a short time—hours to a few days—of the training data set. Accuracy is typically analyzed using a leave-one-out cross-validation approach not reflective of real-world usage of these systems [27, 69, 70, 85]. The work in this chapter began as an effort to evaluate the temporal stability of PLP on a longer-term basis than was performed in the original work. In this chapter, I consider the implications of time separating the training and evaluation data sets up to 78 days. As will be discussed, this led to the finding that a system using *any* two-frequency narrowband fingerprint, not just the two frequencies used by PLP, is susceptible to accuracy degradation over time due to noise sources which affect the fingerprint.

3.2 Experimental Procedure

In this section, I describe the location where experiments were performed, as well as the equipment and procedures used to do so. I also discuss the processing technique by which a fingerprint is extracted from the raw signals that were captured, and the classification method used to map unknown samples to a location within the surveyed space.

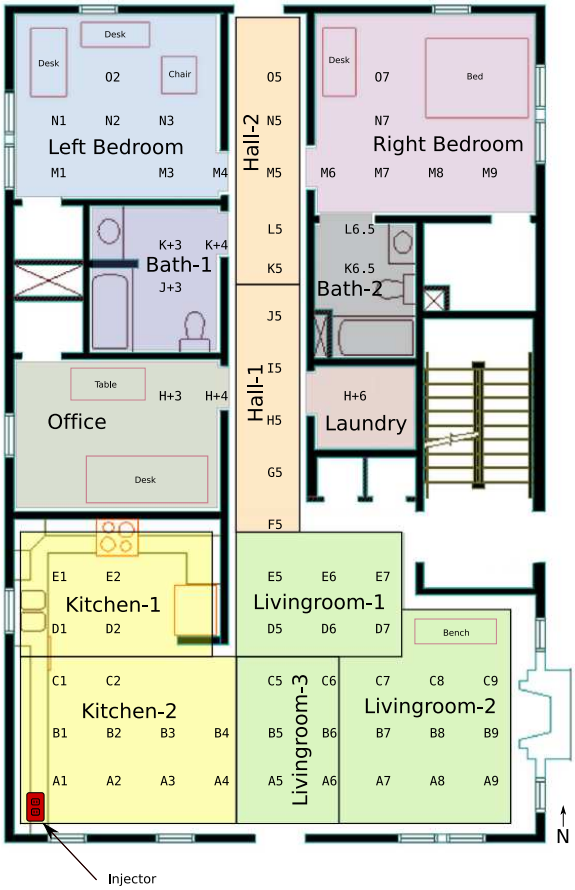

Figure 1: Floor plan and associated room-level (indicated by color), sub-room-level (textual labels), and grid-level (A1, A2, etc.) labels of the first floor of the Georgia Tech Aware Home, where tests were conducted.

3.2.1 Location

Experiments were conducted on the first floor of the Georgia Tech Aware Home [50], which contains two floors with identical layouts—each with a kitchen, living room, two bedrooms, two bathrooms, and an office. The space was divided into a 0.9-meter by 0.9-meter (3-foot by 3-foot) grid, shown in Fig. 1, with a total of 66 grid points throughout the first floor. Fingerprint features (amplitude of a signal at differing frequencies) were sampled at each point on the grid. Each point was identified by a letter (A–O) for the row and a number (1–9) for the column. Occasionally, due to the layout of the floor, surveyable points fell halfway between major grid points.

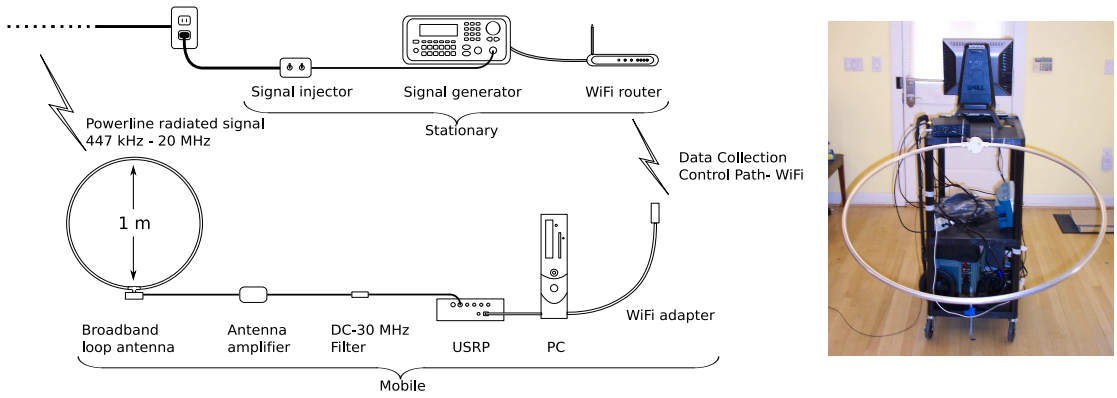

Figure 2: Equipment used to transmit and receive the wideband fingerprint.

These points are designated by a '.5' in the column number for a point falling halfway between two columns, and a '+' after a row letter for a point falling halfway between two rows.

Since PowerLine Positioning utilizes a fingerprinting-based approach, it is not necessary to know anything about the layout of the electrical infrastructure of the test environment. The only purpose of the power line is as a conduit to radiate the signal used for fingerprinting throughout the space of interest.

3.2.2 Equipment Setup

The PowerLine Positioning system consists of two components—one or more plug-in modules which inject a signal into the power line infrastructure, and one or more mobile 'tags' which receive the injector's signals as radiated by the power line infrastructure. A prototype system composed of research equipment designed to simulate these two components was used for the experiments in this chapter.

The signal injector consisted of an Agilent 33220A 20 MHz function generator [1]. Injected signals were pure unmodulated sine waves at various frequencies and at an amplitude of 10 V peak-to-peak. Forty-four different frequencies were used in all—447 kHz, 448 kHz, 600 kHz, and 601 kHz (Patel *et al.* PLP frequencies), and 500 kHz to 20 MHz in 500 kHz steps. The output of the signal generator was connected to a

custom-built power line injector box that allowed the high frequency signals produced by the signal generator to enter the power line while isolating the signal generator from the high voltage 60 Hz power signal. The injector box, which connected to the power line in differential mode, had the capability to inject the signal on the hot and neutral, hot and ground, or neutral and ground wires. To compare my results as closely as possible with the original PLP work, I chose to inject the signal on the hot and neutral wires. A circuit diagram of the injector box is shown in Fig. 45, of Appendix A.

The equipment used to simulate the mobile location tag consisted of an amplified broadband antenna, a software radio, and a standard laptop computer. The antenna, a Wellbrook Communications ALA-1530+ loop antenna, had a frequency range of 150 kHz–30 MHz [16]. The software radio, an Ettus Research Universal Software Radio Peripheral (USRP) [4], contained a 64 MHz 12-bit analog-to-digital converter (ADC). The ADC's input was connected to the antenna through a low-pass antialiasing filter and the Ettus Research LFRX daughterboard [3]. This daughterboard uses a differential amplifier to connect its input to the USRP's ADC input and operates from DC to 30 MHz.

A large battery and a power inverter were used to power the mobile location tag equipment (receiver) in order to both isolate it from the power line and to make it electrically similar to the small battery-powered mobile tags that indoor localization systems typically utilize. All components of the receiver (laptop, antenna, and software radio) were placed onto an electrically isolated plastic cart to ensure no signal was coupled through the cart. The equipment setup is shown in Fig. 2. Although the size of the current receiver, which represents a mobile location tag, is clearly impractical for a real-world system, the necessary components should be able to be easily miniaturized into a small device — the current prototype was designed for flexibility while conducting this research.

3.2.3 Data Collection

Before each dataset was captured, the USRP's output was calibrated by using the function generator to produce a signal of known amplitude at each of the 44 frequencies that were tested. This eliminated any variation in the USRP hardware that may have occurred between dataset captures, for example due to temperature changes in the environment. Each dataset was then captured with the signal generator and injector box connected to the power line via an outlet in the kitchen, the location of which is illustrated in Fig. 1. The injector remained stationary throughout the experiments. The receiver cart was moved in succession to each point on the grid (marked with tape on the floor), where a marker attached to the cart was used to align it consistently across data captures. At each of the 66 grid points, samples of all 44 frequencies were captured using the software radio, which digitizes the analog waveform at the output of the antenna. The raw, digitized waveform was transferred to the laptop computer from the USRP over a USB connection, where it was stored for later processing. For each frequency, a sample was taken with the signal generator output both on and off, allowing me to capture the ambient RF noise in the environment as well as the signal being transmitted. It was observed that the amplitude of the injected signal was above the amplitude of the noise floor for all samples taken.

Collection of this data over the entire grid took approximately three hours. Seven complete datasets were collected over a two-month period—an initial training dataset, and subsequent testing datasets at 14, 42, 49, 58, 58.5, and 78 days separation from the training dataset.

3.2.4 Data Processing

Once the raw digitized signal was captured from the USRP and stored on the laptop, the data required further processing to extract the amplitude of each signal—which is the feature PLP and WPLP use for fingerprinting. Understanding the processing

methods used on this data requires a bit of background knowledge on the USRP. The USRP uses a 64 $Megasample/second$ ADC which samples at a resolution of 12-bits. ADC samples are first processed by an FPGA on the USRP which performs downconversion of the signal (if requested) via a CORDIC numerically controlled oscillator (NCO). This is necessary since although the ADC samples at 64 $Megasample/second$, the USB bus over which the USRP communicates is not capable of transferring data at this high bandwidth. To accommodate the USB bus bandwidth, signals must be downconverted and then decimated to fit within the effective bandwidth of 8 $Megasamples/second$. Decimation is also done in the USRP's FPGA using a cascaded integrator-comb (CIC) filter. Following the USRP's onboard processing, each ADC sample is output as two 16-bit integers representing in-phase and quadrature (I & Q) samples.

For these experiments, the USRP's decimation was set to 16, giving an effective sample rate of 4 $Megasamples/second$. This means that each sample captured a 4 MHz-wide "chunk" of radio spectrum, centered at the specified downconversion frequency. For all frequencies tested that were 1.5 MHz or greater, the downconversion frequency was set to 1 MHz below the frequency of interest. For those frequencies 1 MHz and below, downconversion was not necessary. For each of the 44 frequencies at each of the 66 grid points, $50,000$ I & Q samples were captured using the GNU Radio software [6] on the laptop. These captures were then loaded for processing into GNU Octave [5], an open-source MATLAB clone. It was observed that the first $1,999$ samples of each capture were consistently corrupt, and so they were dropped. A discrete Fourier transform (DFT) was then computed on the remaining 48,001 complex samples of each capture using Octave's built-in FFT algorithm. A 48,0001 point FFT was calculated, giving a frequency resolution of approximately 83.33 Hz. The magnitude of the complex-valued FFT was then computed and converted to a logarithmic scale. The amplitude of the signal of interest was then found by calculating

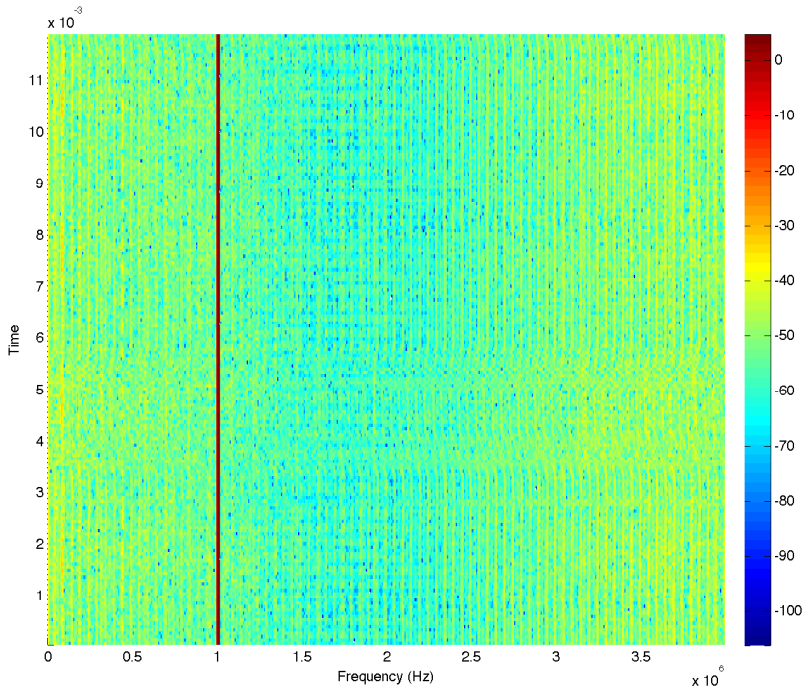

Figure 3: A spectrogram of the 16.0 MHz signal captured at grid point A1 in the initial training dataset. The USRP used a downconversion factor of 15.0 MHz, making the 16.0 MHz signal present at 1.0 MHz in this plot. Note that the signal radiated by the powerline is easily distinguished from the background noise. The amplitude shown is in dB.

the appropriate FFT bin number where the signal would fall and examining this bin. This amplitude was then adjusted by a calibration factor calculated during the initial calibration procedure, which allowed me to map the ADC's output values to a dBm scale.

A sample of one of these captures is shown in Fig. 3, which shows the 16.0 MHz signal captured at grid point A1 during the initial training dataset. The USRP was set to downconvert 15.0 MHz to baseband, meaning the 16.0 MHz signal falls at 1.0 MHz in this plot. This plot shows that using this processing method, the power line radiated 16.0 MHz signal is clearly distinguishable from background noise. The other 43 frequencies produced similar results.

3.2.5 Classification

Once the amplitudes of each of the 44 frequencies were calculated at each of the 66 grid points, the data was converted to the ARFF format for use in Weka, a Java-based machine-learning toolkit [15]. Each of the complete datasets was represented by a single ARFF file containing the 44 features (amplitudes of each of the 44 frequencies tested) at each of the 66 surveyed grid points. Using Weka, various machine learning techniques could then be tested on the data. I utilized the nearest-neighbor algorithm, which calculates the Euclidian distance in signal space between an unknown sample and each of the samples in a training dataset. The algorithm then classifies the unknown sample as the training point least far away in signal space. This algorithm was found to produce the best results amongst those available in Weka and is the same algorithm utilized by Patel *et al.* in PLP.

3.3 Revisiting PowerLine Positioning

I began by attempting to reproduce the results of the original system [73]. Since PLP utilized two injectors, each transmitting a single unique frequency, I simulated it by selecting two frequencies from the larger set of 44 that were tested. To replicate the original system as closely as possible, I began by selecting the original frequencies used by PLP (447 kHz and 600 kHz). Note that although Patel *et al.* utilized 33 kHz rather than 600 kHz, 33 kHz was outside the operating range of my antenna, and discussions with the authors of this work revealed that in subsequent experiments they had switched to 600 kHz and obtained results similar to those previously published.

Another difference from PLP is that my experiments used a single injection point, rather than two physically-separated injection points. Physically separating the injectors can serve to add information useful to the classifier that may not be present if both signals are transmitted from the same location, although this is not necessarily the case. A measure of the amount of additional information provided by using a

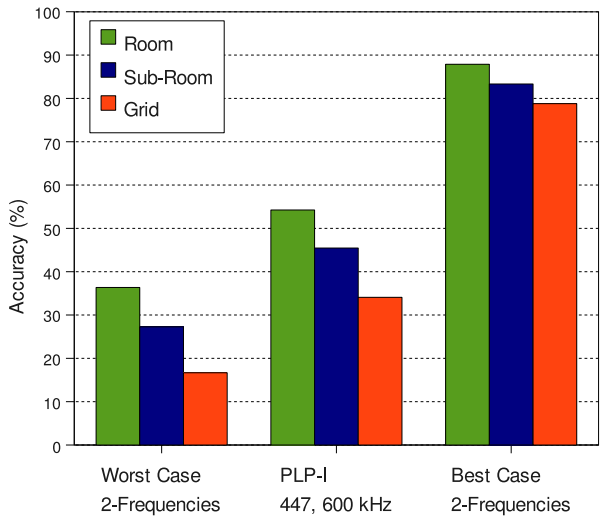

Figure 4: Localization accuracy for the two PLP frequencies, and also the best and worst possible accuracies with two frequencies chosen from the set of 44 that were tested. Sets of two frequencies generating the best and worst accuracies were chosen independently for each of the three localization granularities.

second frequency (over using just one for classification) is the correlation between the amplitudes of the two frequencies. In my data, the correlation between 447 kHz and 600 kHz ranged from 0.35–0.42. A correlation of 1 indicates that the second frequency provides no additional information, and a correlation of 0 means that the two frequencies share no information. Patel *et al.* found correlations between the two injectors ranging from 0.05–0.6. The correlation values for my data fall within this range, and I therefore believe the use of a single injection point for multiple frequencies serves as a valid comparison. This shows that sufficient fingerprint spatial variability comes from differences in how different frequencies interact with the power line, not necessarily where they are injected into the power line.

3.3.1 Initial Post-Deployment Results

To evaluate the performance of PLP in the test environment, I took the amplitude data at 447 kHz and 600 kHz captured at each of the 66 surveyed points, and used

this data to perform room-level, sub-room-level, and grid-level classification of a test dataset. The classification was performed using nearest-neighbors method on the signal space using the received signal amplitudes. Thus, each physical location on the grid (as shown in Fig. 1) had associated with it the amplitude of the two signals (447 kHz and 600 kHz) as sensed at that particular location, as well as the symbolic label given to that grid point. The test and training datasets were completely independent, captured several hours apart.

Room-level accuracy ranged from 48.5%–60% depending on which of the two datasets was used for training and which was used for testing, which is significantly lower than the 78%–100% accuracy achieved by Patel *et al.* in their experiments. Sub-room-level performance ranged from 39.4%–51.5%, again much lower than the 87%–95% achieved in the original work. Grid-level accuracy ranged from 30.3%–37.9%. The average of the two accuracies is shown in Fig. 4 for each localization granularity. The lower accuracies observed can be attributed to both ambient RF noise as well as electrical noise on the power line due to the large amount of electrical equipment operational in the test environment.

The construction of the electrical infrastructure in the Aware Home may also be a factor. Although the layout of the test environment appears to be that of a home, it is actually a laboratory and as such was constructed according to commercial building standards. Substantial ambient noise, possibly caused by the large amount of electrical equipment operating within the test environment, was observed in the 400-600 kHz range. This phenomenon was not observed in the deployment of PLP in residential spaces. This may also be due to the electrical wiring being deployed within metal conduit, as described earlier. Note that the hot and neutral wires on which the signals are injected are completely contained within this conduit, which can serve to reduce the amount of signal radiated. The experiments conducted for Patel *et al.*'s PLP work were in homes with residential grade wiring without conduit. The results

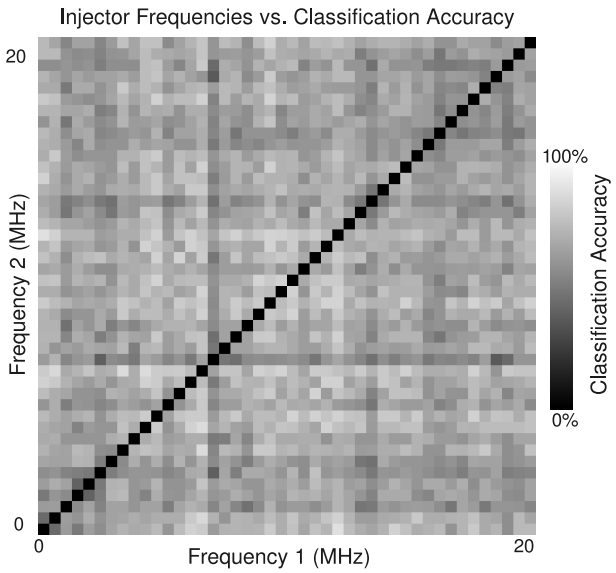

Figure 5: Room-level localization accuracy using two frequencies for each of the 1,892 possible combinations of two frequencies from the 44 that were tested. The diagonal represents no data since the injector can not transmit two signals simultaneously on the same frequency.

obtained in this environment, therefore, would likely be significantly better if these experiments were carried out in the more traditional domestic environments where PLP was tested.

3.3.2 Optimal Frequency Pair Selection

Next, I performed an exhaustive search of all possible combinations of two frequencies from the 44 that were captured in order to determine the best accuracy achievable with a two frequency fingerprint. The idea was that if a pair of frequencies could be found that consistently provide the best accuracy across each of the three classification granularities, PLP could be adapted to use these frequencies rather than the original 447 kHz and 600 kHz. The classification was again performed using the KNN algorithm with a K value of 1.

The results of a room-level classification of each of the 66 samples taken throughout the floor are illustrated in Fig. 5 for all 1,892 possible combinations of frequencies.

Lighter pixels represent better accuracy, and the diagonal of black pixels represents no data where the injector would be required to transmit the same frequency twice, which is impossible. The frequency scale starts with 447 kHz, 448 kHz, 500 kHz, 600 kHz, 601 kHz, 1 MHz, and continues in 0.5 MHz steps to 20 MHz. The best possible room-level accuracy was 87.88% (achieved with 447 kHz and 11.5 MHz), the best sub-room-level accuracy was 83.33% (achieved with 8.5 MHz and 11 MHz), and the best grid-level accuracy was 78.79% (achieved with 8.5 MHz and 9.0 MHz). These results, along with those of the PLP frequencies and the worst-case two frequencies, are summarized in Fig. 4.

This exhaustive search leads to two important results. First, significant improvements in accuracy can be made simply by using frequencies other than the 447 kHz and 600 kHz signals used by PLP. Room-level accuracy improves by 33.6%, sub-room-level accuracy by 37.9%, and grid-level accuracy by 44.7% simply by selecting a more optimal alternative set of two frequencies. The second result of this search is that the frequencies that provide the best possible accuracy are not consistent across the three classification granularities. Not only are they inconsistent, but they also span much of the range tested, starting with the lowest frequency tested of 447 kHz all the way to 11 MHz. Additionally, Fig. 5 fails to demonstrate any obvious pattern that would lead to a formula for *a priori* selection of two frequencies to be used in a PLP localization system. This is a major reason why I propose a higher dimensional fingerprint.

3.3.3 Noise Sensitivity

For a fingerprinting-based localization system to work, the fingerprint space must exhibit both temporal stability and spatial differentiability. The previously discussed results, with room-level accuracy of up to 88% for two frequencies based on training and testing sets taken the same day, demonstrate that the amplitude of these signals

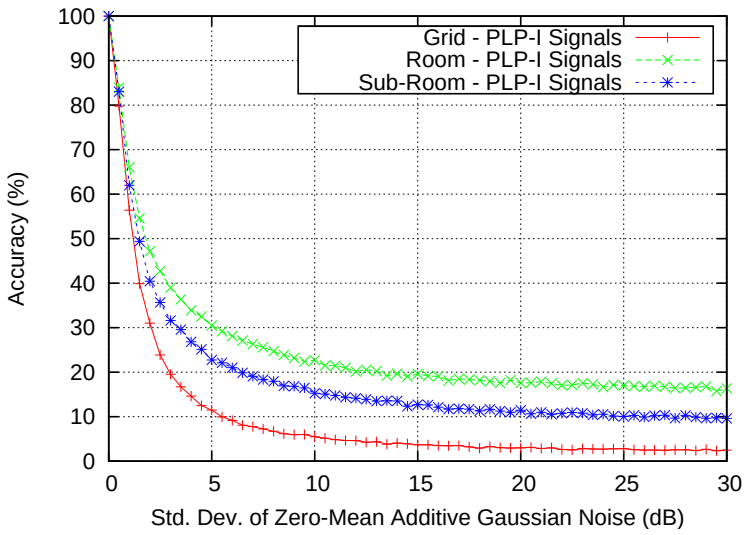

Figure 6: Localization accuracy vs. standard deviation of zero-mean additive Gaussian noise when using the two PLP frequencies (447 kHz, 600 kHz). This shows that the classifier has a high sensitivity to noise for a localization system based on the two-frequency PLP method.

is indeed spatially differentiable. I thus set out to examine both the stability of the signal amplitudes in time, as well as the effect that any instability has on localization results.

To understand the impact of any temporal instability in the signal amplitudes, I ran tests for each dataset where the original dataset was used for training, and an artificially generated noisy version of that same dataset was used as the subsequent testing set. The noisy version was generated by adding independently-generated samples of zero-mean Gaussian noise to the amplitude of each of the PLP frequencies at each of the 66 locations in the dataset. The impact of noise with standard deviation ranging from 0 dB to 30 dB was examined, and is shown in Fig. 6. The accuracies shown represent the mean over 100 independent generations of the noisy dataset versions, as well as over all collected datasets. These results show that with just 1 dB of noise, the localization accuracy drops below 50% for all classification granularities when using the PLP frequencies of 447 kHz and 600 kHz.

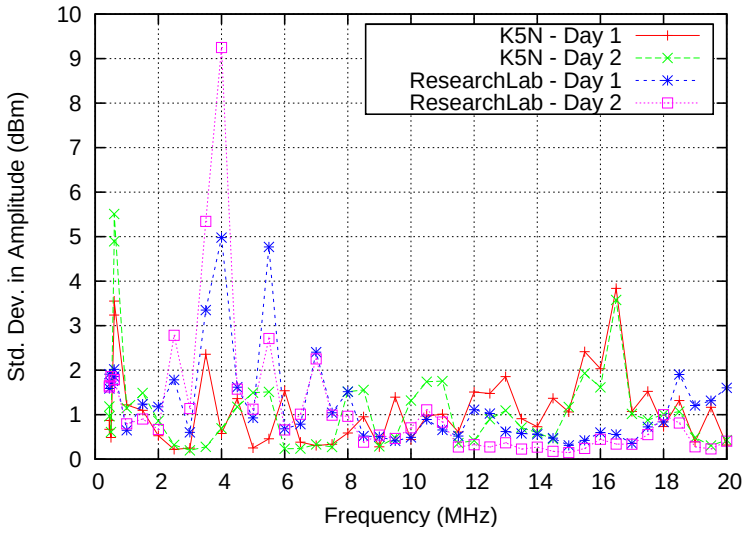

Figure 7: Standard deviation in amplitude vs. frequency for long-term tests at two locations. This provides an indication of the variability of the amplitude of each frequency over the course of a day.

3.3.4 Long-Term Temporal Stability

The next question then is what amount of temporal instability in the signal amplitudes exists in the real-world. To answer this, I took measurements over four separate 24-hour periods at two locations. Approximately 950 samples at each of the 44 frequencies were obtained across each 24-hour period. One of these locations was at point K5 on the grid, and the other was in a laboratory space. During these tests the cart containing the equipment remained completely stationary, eliminating any change in signal amplitude that may be present in the datasets taken over the entire grid due to inconsistent cart placement between measurements. The results of these long-term tests are illustrated in Fig. 7, which shows the mean standard deviation in amplitude at each of the 44 frequencies across the 24 hours of each test.

Figure 7 leads to two important conclusions. First, the mean standard deviation across all four tests and all 44 frequencies is 1.17 dB, which is above the 1 dB that was determined earlier to cause classifier accuracy to drop below 50%. Three out of the four 24-hour periods had 21 frequencies with standard deviations above 1 dB,

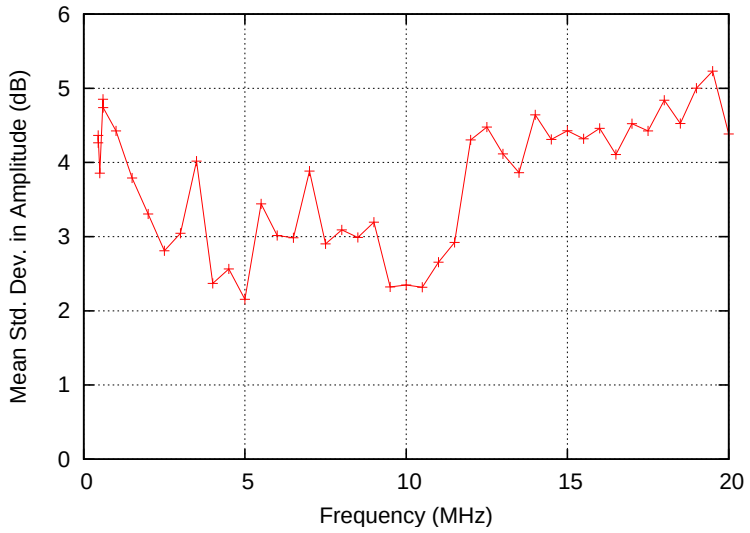

Figure 8: Standard deviation in amplitude vs. frequency. The standard deviation is found using all six independent datasets captured over the two-month period at each of the 66 surveyed points, and all points are then averaged.

and one had 15 frequencies above 1 dB. Of these same tests, two had five frequencies with standard deviations above 2 dB, one had six, and one had three. Figure 6 shows that at 2 dB of standard deviation, localization accuracy ranges from 25% to 35% depending on localization granularity. Second, although some frequencies show low standard deviation ($<$ 1 dB) in one location, they may experience high standard deviation ($>$ 2 dB) in another location. Therefore, it is not clear how to to select *a priori* two frequencies that will produce consistently good localization results over both time and space.

A real-world deployment will obviously last much longer than the 24-hour tests I described, and will also suffer from additional 'noise' due to inconsistent placement of the receiver between training and testing samples. To get an indication of what real-world standard deviation in the signal amplitude might look like, I calculated the same value as before (standard deviation in amplitude at each of the 44 frequencies) across all six datasets. These values are shown in Fig. 8. Values were computed by finding the standard deviation among the amplitude samples at each frequency for

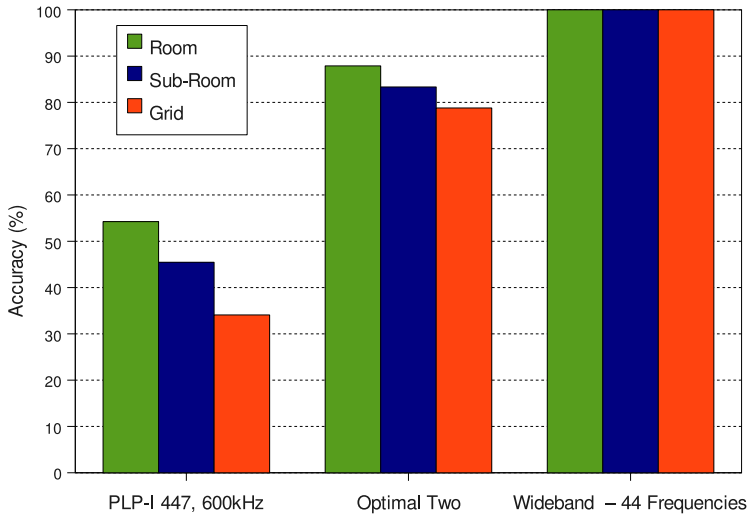

Figure 9: Localization accuracy when the classifier is trained and tested on independent datasets taken on the same day, several hours apart.

each of the 66 surveyed locations in each independent dataset, and then taking the mean over all 66 locations. For the six datasets I collected over a two month period, the mean standard deviation in amplitude across all 44 frequencies was 3.74 dB, and the minimum was 2.16 dB. Recall that the localization accuracy of a two-frequency system is less than 35% with noise of greater than 2 dB (shown in Fig. 6).

3.4 Wideband PowerLine Positioning

I have demonstrated that a general rule for selection of two frequencies for PLP does not seem to exist. Additionally, even if the optimal two frequencies are selected initially, their performance may degrade over time. I thus set out to examine the effect of using a more complex feature space in the system. In this section, I demonstrate that the use of a wideband fingerprint consisting of the amplitude of signals at 44 distinct frequencies not only improves initial localization accuracy, but also better maintains that accuracy over time by mitigating some of the effects of temporal instability in each of the constituent signals. I refer to a system using this wideband fingerprint as Wideband PLP (WPLP).

3.4.1 Initial Post-Deployment Results

The use of a 44-frequency wideband fingerprint can significantly improve localization accuracy. Figure 9 shows localization performance for a training and testing set taken several hours apart, which should yield excellent localization results since changes in the physical environment that could affect the RF signals will be minimal. However, as seen earlier in Fig. 4, this is not necessarily the case when using only two frequencies for localization. In fact, the PLP system achieves only 54% room-level accuracy with 447 kHz and 600 kHz in this environment, even for these relatively ideal conditions. A wideband fingerprint provides additional features to the classifier, allowing it to make better decisions. In this case, the wideband fingerprint leads to 100% accuracy at room, sub-room, and grid level. This represents a gain of 46%–66% over the accuracy achieved with the PLP frequencies. Figure 9 also shows that even with the optimal selection of two frequencies (with each of the three localization granularities having its optimal frequencies selected independently), the wideband fingerprint provides gains in accuracy of 12%–21%.

3.4.2 Noise Sensitivity

Having observed the effect that a wideband fingerprint has on initial localization results, I set out to quantify its effect on sensitivity to noise. Noise sensitivity of the classifier was first shown in Fig. 6, which presented results for a system utilizing the two PLP frequencies of 447 kHz and 600 kHz. Figure 10 illustrates this information for the 44-frequency wideband fingerprint. As before, this information was obtained by adding independently generated zero-mean Gaussian noise samples with varying standard deviation to the amplitude of each of the 44 frequencies sampled at each of the 66 data-points. The classifier was then trained on the initial uncorrupted dataset, and tested on the noisy version of that same dataset. Figure 10 represents the mean over 100 independent generations of a noisy version of all six datasets.

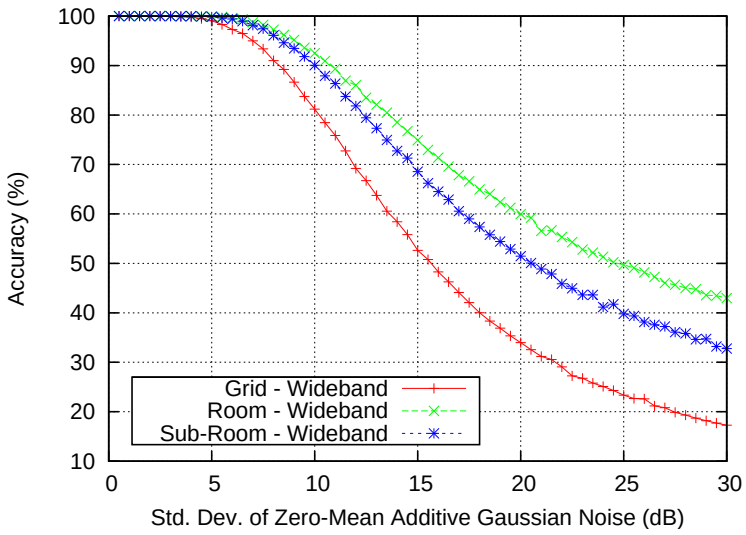

Figure 10: Localization accuracy vs. standard deviation of additive zero-mean Gaussian noise. The classifier can tolerate significantly more noise in a wideband fingerprint before accuracy degrades, and the degradation is more graceful than the two-frequency localization approach of PLP, shown in Fig. 6.

Figure 10 shows that a wideband fingerprint is capable of tolerating noise of up to 5 dB in standard deviation before classifier accuracy drops below 100%. In contrast, Fig. 6 demonstrates that the PLP two-frequency signal has accuracy of less than 25% for all localization granularities at this level of noise. In addition, Fig. 10 shows that as the noise increases, classifier accuracy degrades much more gracefully than in the PLP two-frequency signal.

3.4.3 Long-Term Temporal Stability

With results showing that a wideband fingerprint can improve temporal stability by increasing the noise tolerance of the classifier, I now discuss the results of a real-world experiment with a 44-frequency wideband fingerprint for training and testing data taken two months apart. The results for the two-frequency PLP signal, the optimal two frequency signal (with optimal selection being independent for each level of localization granularity), and the 44-frequency wideband fingerprint are shown in Fig. 11. Two important effects can be observed here, and in Fig. 12, which illustrates the

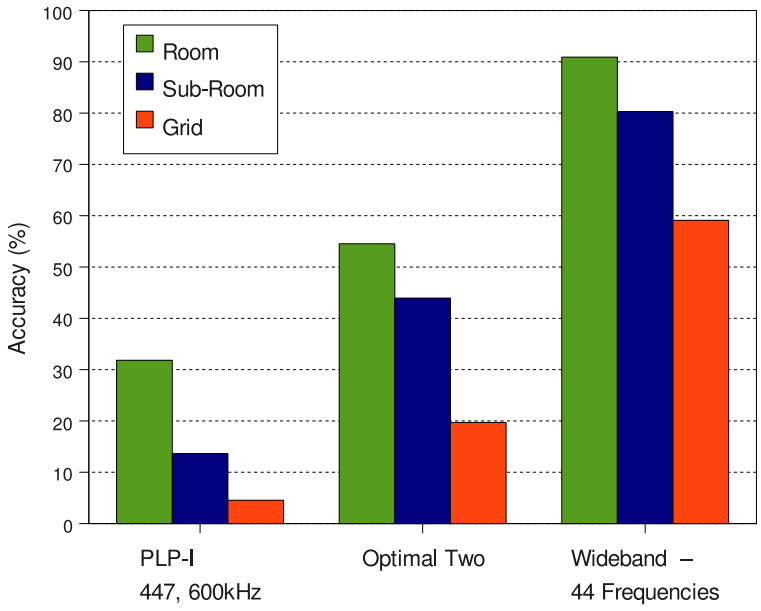

Figure 11: Localization accuracy when the classifier is trained and tested on independent datasets taken two months apart. Accuracy is shown for the two-frequency PLP signal (447 kHz and 600 kHz), the optimal two frequencies from all possible combinations, and the 44-frequency wideband fingerprint.

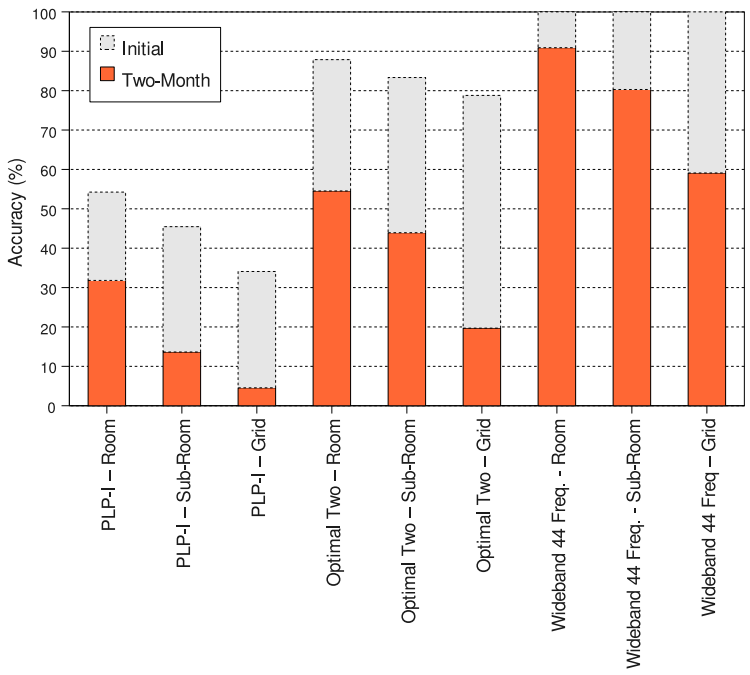

Figure 12: Classifier performance degradation between closely temporally spaced training and testing data (several hours), and a two-month separation between training and testing data.

decrease in performance for each of the three types of signals from initial deployment (Fig. 9) to two months later (Fig. 11). First, although a two-month temporal separation between training and testing data has reduced the accuracy of the 44-frequency wideband fingerprint to 59%–91% depending on the localization granularity, it has maintained significantly better accuracy than either the two-frequency PLP signal or the optimal two-frequency signals across all three localization granularities. Second, the decrease in performance with the 44-frequency wideband fingerprint for all localization granularities (except grid-level) is lower than with the two-frequency PLP signal and the two-frequency optimal signal. Thus, not only does a wideband fingerprint deliver better initial accuracy, it also serves to better maintain that accuracy over time.

The 'initial' and 'two-month' accuracy results of Fig. 12 are based on the comparison of three independently collected datasets—two separated by several hours, and

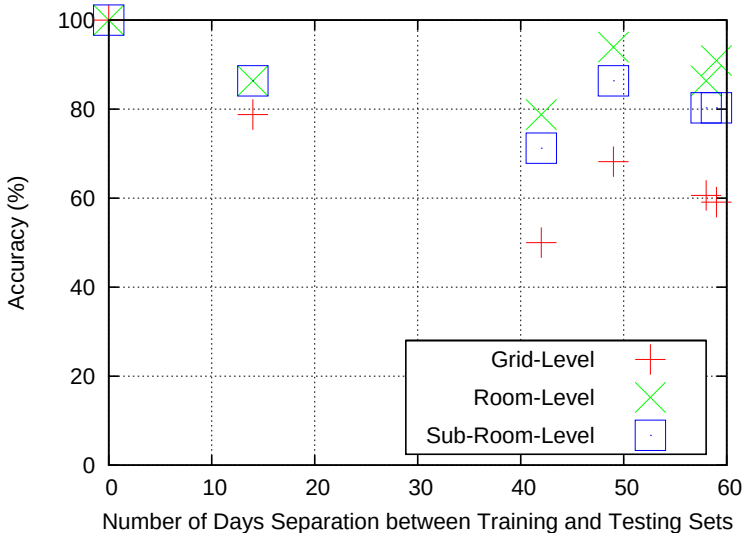

Figure 13: Localization accuracy vs. number of days separation between the training and testing data for a 44-frequency wideband fingerprint.

one separated from the initial set by two months. The earliest dataset has been used as the training set, which simulates the initial site survey necessary upon deployment of a fingerprinting-based localization system. Over the two-month period, I collected three additional datasets in the interim. Using these interim datasets as additional testing data for a classifier trained on the initial site-survey can provide insight into how localization accuracy degrades over time, and is shown in Fig. 13. The interesting result here is that there does not appear to be a pattern in accuracy degradation over the two-month period. It appears that any significant temporal separation between training and testing datasets will degrade accuracy, but that accuracy degradation is not necessarily correlated with increasing time after a point.

3.4.4 Wideband Fingerprint Size

Evidence clearly shows the benefit of the wideband fingerprint over a two-frequency signal, although it is unclear if a wideband fingerprint with fewer than 44 constituent frequencies may provide the same accuracy. Using the fewest number of signals possible to obtain the demonstrated benefits of a wideband fingerprint is desirable since

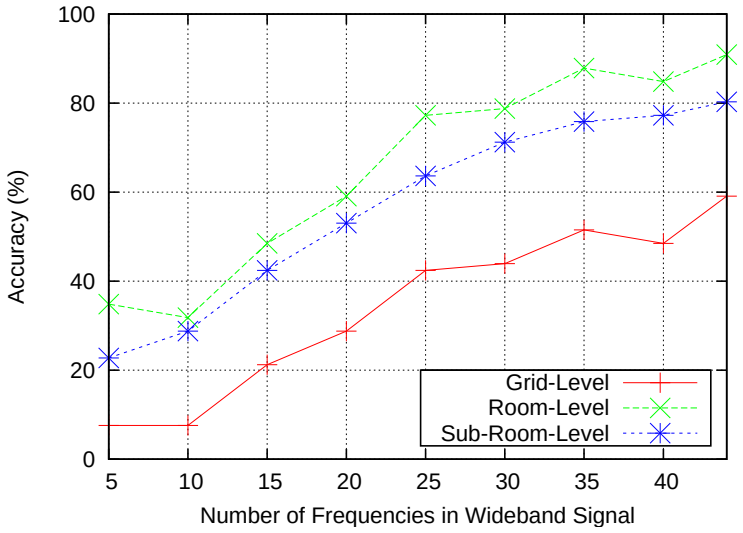

Figure 14: Localization accuracy vs. number of frequencies in the wideband fingerprint when trained and tested on data separated by two months. The use of additional frequencies provides additional features for the classifier and generally improves accuracy. Frequencies were added from the set of 44 in order of decreasing signal-to-noise ratio (SNR).

additional frequencies may increase the cost, complexity, and size of receivers designed for a real-world practical localization system based on the wideband approach.

To determine the effect of additional frequencies on localization performance, I examined the performance of a wideband fingerprint ranging in size from five frequencies all the way to the complete set of 44 frequencies on the two-month separated datasets. Since a key benefit of the use of a wideband fingerprint is resistance to performance degradation over time, examining the two-month separated data makes more sense than the datasets captured closer together in time. To determine the order in which frequencies were added to the signal, they were ranked in order of decreasing signal-to-noise ratio (SNR) based upon the training dataset. Although other orderings could be used, and may produce different results, the problem space is too large to perform an exhaustive analysis. Adding frequencies by decreasing SNR is a logical approach since frequencies closer to the ambient noise level are more likely to be affected by noise, and will thus have a higher standard deviation in amplitude. Recall Figs. 6 and 10,

which showed that this adversely affects localization performance. Frequencies with higher SNR's will thus likely provide better localization accuracy. Localization accuracies for the wideband fingerprint ranging in size from five to 44 frequencies are shown in Fig. 14. The general trend is that the use of additional frequencies provides additional localization accuracy, although it appears that the marginal utility of each additional frequency begins to decrease past 35 frequencies.

3.5 Understanding Classification Errors

Thus far, the accuracy of WPLP has been considered only in terms of the percentage of correctly classified samples. In this section, I analyze the system's accuracy in terms of the physical distance between the true location of a sample and the classified location. Examining this aspect of performance is critical to understanding WPLP's real-world performance—an error classifying a sample adjacent to its true location has much different implications than one classifying it at the opposite end of the space.

3.5.1 Single Initial Site-Survey Training Dataset

Figure 15 illustrates the average error distance for the six testing datasets when tested against a classifier trained on the one initial "site-survey" dataset. Recall that each dataset, both training and testing, consists of one sample of the 44-frequency wideband fingerprint at each of the 66 distinct physical locations in Fig. 1. Thus, the classifier was tested on 66 samples for each of the testing datasets shown here. As with Fig. 13, which examined the system's percentage accuracy for room, sub-room, and grid levels, there does not appear to be a pattern of accuracy degradation correlated with time separation between training and testing data. In fact, the four datasets spaced furthest from the training data in time exhibit lower average distance errors than the closest two datasets. Mean distance error for the six testing datasets ranged from 0.56 m–1.41 m, with an overall average of 0.9012 m.

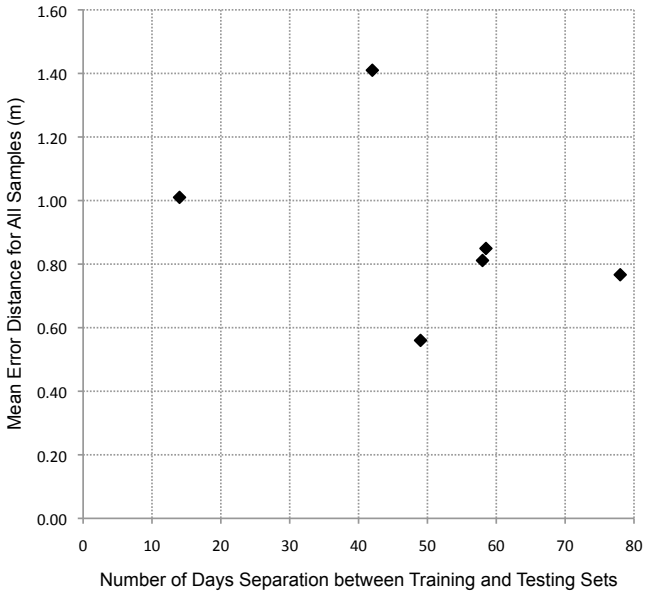

Figure 15: Mean error distance in meters vs. number of days separation between the training and testing data for a 44–frequency wideband fingerprint. Error distance is averaged across all classified samples—including those classified correctly with zero error distance.

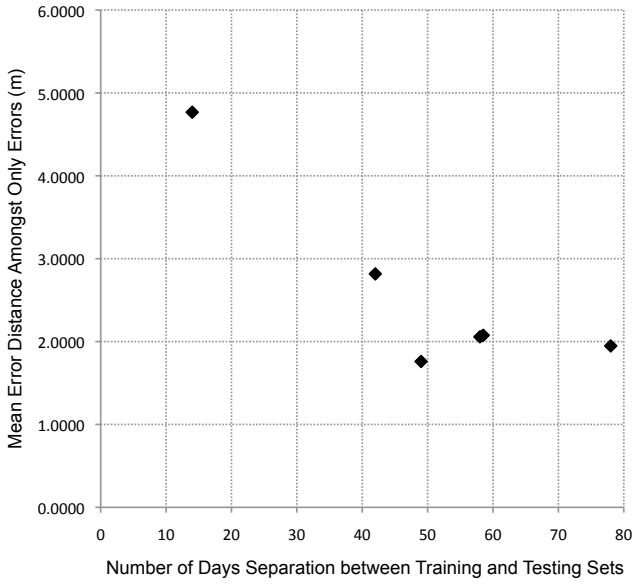

Figure 16: Mean error distance in meters vs. number of days separation between the training and testing data for a 44–frequency wideband fingerprint. Error distance is averaged across only incorrectly classified samples. This represents the mean error for those samples where an error occurred.

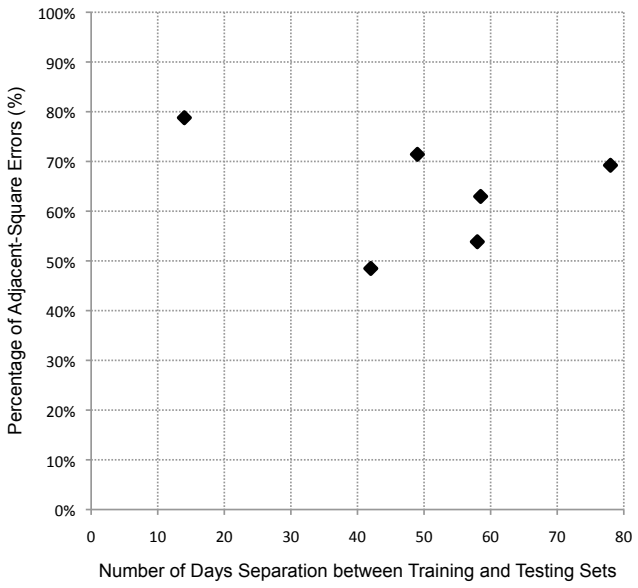

Figure 17: Percentage of errors where the erroneously classified cell is adjacent to the true cell vs. number of days separation between the training and testing data for a 44-frequency wideband fingerprint.

While Fig. 15 characterizes the overall accuracy of WPLP in terms of physical distance, this data averages errors over all samples, including correct samples with zero error. To understand the characteristics of errors, it is helpful to look at the average distance error for only those samples that were incorrectly classified—this data is illustrated in Figs. 16 and 17, which show the mean error for only those samples with errors and the percentage of errors that map to an adjacent square. The mean distance of an error for the six testing datasets ranges from 1.76–4.76 m, and the percentage of adjacent square errors ranges from 48.5%–78.8%. Thus, typically greater than 50% of the errors the classifier makes map to a grid point adjacent to the true point. Interestingly, the average length of an error seems to decrease with an increasing time separation between the training and testing datasets.

It can also be insightful to look at the errors on a map of the physical space. The PLP-Viz tool [39], discussed in Appendix B, makes this easy to do. As an example,

a mapping of the errors in a dataset captured 49 days after the initial training site-survey dataset is shown in Fig. 18. In this figure, bright green blocks without a line originating in the center represent samples that were correctly classified. Blocks with darker colors represent errors, and the line from that block points to the location to which the system classified that sample. The color varies from green to red with increasing physical distance of the error. This same type of plot is shown for all other testing datasets in Appendix A.

A common method of analyzing the distance error of a localization system is with a cumulative distribution function (CDF) plot, which is shown in Fig. 19. For any given point along the line, the y coordinate represents the percentage of samples with distance error less than or equal to the corresponding value of the x-axis. Figure. 19 shows the CDF for each of the six testing datasets when classified against the initial testing dataset using the 44-frequency wideband fingerprint. A CDF of all errors from all six datasets is also plotted. Looking at the CDF line representing all errors, the accuracy of WPLP is 3 meters or better for 90% of samples, and 5 meters or better for 95% of all samples.

3.5.2 Leave-One-Out Cross-Validation

Thus far, all results have been based on testing subsequent "online" datasets against an initial site-survey dataset. I believe that this method represents the most fair assessment of the system's real-world accuracy as it directly reflects how the system would be used—a single site-survey is taken post-deployment to use as training data. To date however, the most common method of analyzing a system's performance in the literature has been leave-one-out cross-validation wherein one dataset is tested on a classifier trained on all other datasets. The process is repeated for each dataset and the results are averaged. This method, which does not reflect real-world usage of an IFRTLS, can artificially inflate performance data. I include this data for WPLP and

Figure 18: Classification errors resulting from training on one initial dataset and classifying a dataset captured 49 days later.

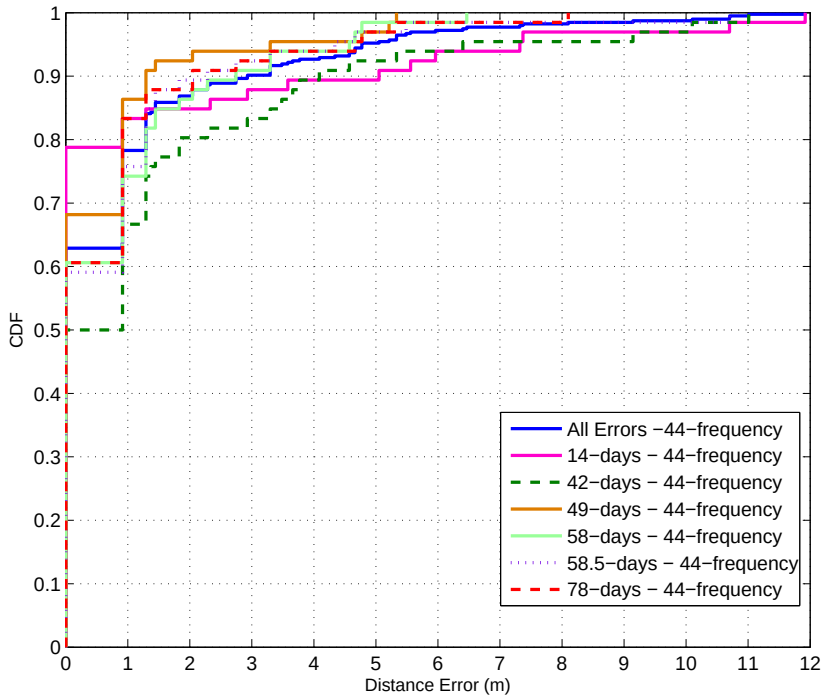

Figure 19: Cumulative distribution function for the error distance of the six testing datasets when trained on the first dataset and using a 44-frequency wideband fingerprint. For all six datasets combined, 90% of tested samples show an error of less than 3 m, and 95% show an error of less than 5 m.

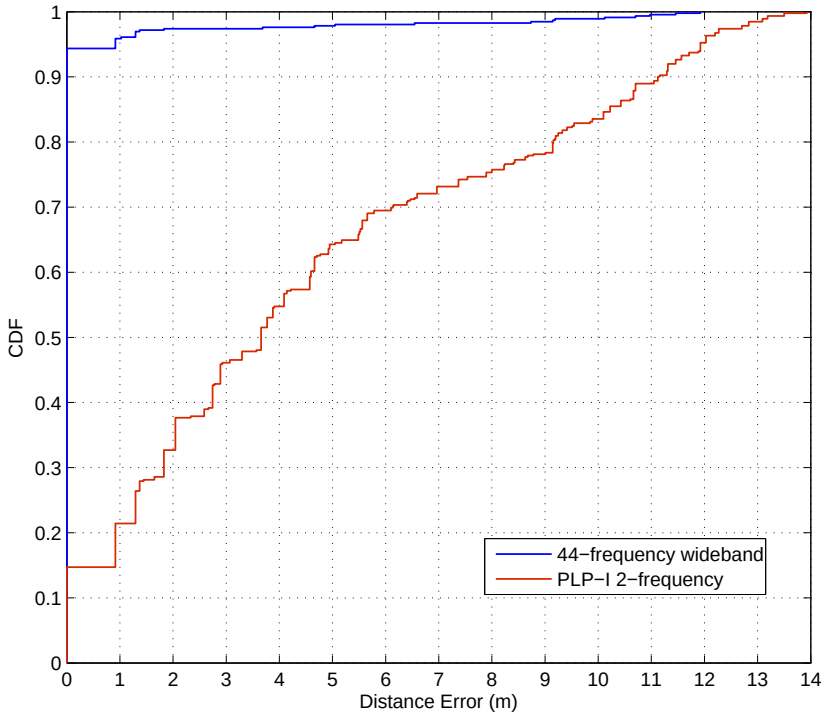

Figure 20: Cumulative distribution function for the error distance of all seven datasets when using leave-one-out cross-validation — this is the method most commonly used to report results in the existing literature. Results are shown for both a 44-frequency wideband fingerprint and the PLP two-frequency fingerprint. WPLP shows an error distance of 1 m or less for 95% of all samples; PLP shows an error distance of 12 m or less for 95% of all samples.

PLP in Fig. 20, however, so that a more direct comparison can be made between these systems and existing IFRTLSs in the literature. Using leave-one-out cross-validation, WPLP with a 44-frequency fingerprint has an accuracy of 0.914 meters or better for 95% of all samples, and the original 2-frequency PLP system has an accuracy of 11.92 meters or better for 95% of all samples. This plot clearly illustrates the superior performance of a wideband fingerprint approach in distance accuracy.

3.6 Summary of Contributions

In this chapter, I have made the following contributions:

- I presented evidence that an IFRTLS utilizing a two-frequency narrowband fingerprint is susceptible to both accuracy degradation over time, as well as lower than expected initial accuracy in certain spaces due to variability in the fingerprint. Variability comes in the form of ambient RF noise, changes over time in the transmitting and receiving hardware, and changes in the electrical infrastructure, amongst other sources.

- I introduced the Wideband PowerLine Positioning system. WPLP is an IFRTLS inspired by the PowerLine Positioning system that uses a wideband fingerprint consisting of the amplitude of 44 different signals ranging in frequency from 447 kHz to 20 MHz.

- I showed that the use of a wideband fingerprint consisting of the amplitudes of 44 different frequencies ranging from 447 kHz to 20 MHz can both improve the initial post-deployment accuracy of the system, and also mitigate some performance degradation over time.

- I showed that, at least up to the maximum of 44 frequencies tested, increasing the size of the wideband fingerprint serves to reduce accuracy degradation over time.

- I performed an analysis of the errors made by the WPLP system in the test environment and showed that generally 50% or more of the errors made by the system map to a grid point adjacent to the true grid point. I also showed that the system has an average accuracy of 0.9 meters.

CHAPTER 4

WPLP TEMPORAL STABILITY VS. COMPETING IFRTLSS

As discussed in Chapter 2, fingerprinting of radio frequency (RF) signals has emerged as the current leading approach to real-time indoor location tracking. A form of scene analysis, fingerprinting relies on one or more observable features of the environment to be spatially differentiable yet temporally stable [43]. The objective of this chapter is to compare the WPLP system discussed in Chapter 3 against the other leading RF-based IFRTLSs — WiFi, GSM, CDMA, and FM. This comparison is done in two parts. First, I examine the long-term temporal stability of the underlying features of each technology. Second, I perform a grid-level analysis to test the spatial differentiability of these technologies.

4.1 Motivation

Temporal instability represents a significant problem for fingerprinting-based indoor RTLSs that has been largely overlooked. Continuous tests of stability in the existing literature (see Table 3) are typically on the order of several hours, which can be insufficient to observe long-term temporal instability. Exacerbating the problem is the standard practice in the literature of gathering testing and training samples at the same time and later splitting the dataset for evaluation of the localization potential of a given feature. Datasets generated in this manner, where each physical point in the space is observed for both testing and training within seconds or minutes of each other, are almost guaranteed to miss the problem of long-term temporal instability. In this chapter, I provide the results of over 16 days of continuous sampling for each

Table 3: The five RF technologies we examined and a summary of relevant previous work using these technologies for indoor positioning.

Technology	Feature	Previous Work	Temporal Stability Tests
WiFi (802.11)	Amplitude	[27, 28, 96]	3-hours
GSM Cellular	Amplitude	[69, 86, 88]	3-hours
CDMA Cellular	Time Delay	[85]	8-hours
FM Radio	Amplitude	[70]	4-hours
Low-Frequency Power Line RF	Amplitude	[73, 80]	4 24-hours periods continuous sampling, instantaneous samples with 2 months separation

of the five RF technologies that have been proposed as the basis of a fingerprinting localization system: WiFi, GSM, CDMA, FM, and low-frequency power line. My tests indicate that the problem of temporal instability is a very real one that must be solved if these systems hope to ever achieve real-world use. The chapter begins with a brief reminder of the relevant related work, much of which has been discussed in more detail in Chapter 2.

4.2 Prior Temporal Stability Analyses

A considerable body of work has been published on RF fingerprinting-based approaches to constructing an indoor RTLS. RADAR [27], which used the ambient signal strength from WiFi access points, was the first to introduce this method. Numerous other WiFi-based works followed with various improvements upon the original approach, but all make use of the signal strength of access points as the feature for fingerprinting [28, 51, 57, 62, 96]. The longest continuous study of the temporal stability of a WiFi access point's signal strength that is available in the literature lasted for only three hours. Interestingly, this experiment was performed by Otsason *et al.* as a comparison with their GSM-fingerprinting-based system [69].

Once the potential of RF fingerprinting-based indoor location systems became known through the initial WiFi-based works, researchers began to explore other types

of RF signals that might be used for fingerprinting. One of these is the signal from GSM cellular basestations [69, 87]. Otsason *et al.* demonstrated the feasibility of using the signal strength of ambient pilot signals from nearby GSM basestations as the fingerprinting feature for an indoor RTLS. They found the signal strength of the three strongest GSM cells detected to be relatively stable over a three-hour window. A single dataset, consisting of 10 consecutive samples 5-seconds apart, was gathered for both training and testing their localization algorithm. Other works have explored GSM-based localization in an outdoor setting, however I focus my discussion here on indoor RTLSs.

CDMA is another form of cellular network whose ambient signals have been explored for use in constructing a fingerprinting-based indoor RTLS. The CILoS work of ur Rehman *et al.* is the seminal research in this space [85]. The authors of that work point out that unlike the strengths of pilot signals from GSM basestations, those of CDMA basestations are not guaranteed to be stable as the signal strength can be adjusted for network load balancing. CDMA networks are tightly time synchronized, however, and track a feature known as the PN delay for each nearby basestation. Each of 512 possible basestations is assigned a specific time slot within a window for pilot signal transmissions. The mobile device monitors the time each of these pilot signals is received and calculates the difference between when it expected to receive the pilot from a particular basestation and when it actually did. This time difference is calculated in "chips" of the 32,768 chip pseudo-random sequence that is the pilot signal. The CILoS work found the PN delay of a basestation to be reasonably stable over time as well as spatially differentiable, and thus to be a suitable feature for the basis of a fingerprinting-based indoor RTLS. The reader is referred to this work for additional details on the operation of CDMA networks. ur Rehman *et al.* monitored the PN delay stability of a CDMA basestation for four 2-hour windows of continuous sampling on four separate days. A single dataset, consisting of 120 consecutive

samples at an unspecified interval, was gathered for both training and testing their localization algorithm.

Although previous work has explored the use of FM radio signals for *outdoor* positioning [53], only more recently has their potential for *indoor* localization been examined — the FINDR work of Papliatseyeu *et al.* is the first to do so [70]. Unlike previous work on outdoor positioning with ambient FM radio signals, FINDR uses inexpensive FM modulators designed for short-range transmission of the audio signal from an MP3 player or other audio source for personal listening. These transmitters are typically very low powered and have a range on the order of several meters. Papliatseyeu *et al.* examine the use of three of these transmitters deployed within a single room for fingerprinting-based localization. They found the signals to be relatively stable over a 4-hour window. Two independent datasets, with an unspecified time separation, were collected for testing and training the system.

PowerLine Positioning (PLP) and Wideband PowerLine Positioning (WPLP) use the existing electrical wiring of a building to transmit low frequency (447 kHz – 20 MHz) RF signals throughout a building [73, 80]. Using the power lines as a signal distribution medium allows a single "injector module" plugged into a standard wall outlet to cover a typical home with the signals needed by PLP and WPLP location tags. When these signals are injected, the power line acts as a transmitting antenna and radiates them as RF energy. Location tags monitor the signal strength of these power line radiated signals as the fingerprint. This method combines the benefits of a low/no infrastructure indoor RTLS with control over the hardware generating the feature that is the basis of the radio map. The original PLP work looked at the stability of these signals over four 3-hour continuous windows on four separate days. The WPLP work looked at stability over four 24-hour continuous windows on four separate days. WPLP used multiple independent datasets for training and testing that were gathered with up to two months separation in time.

4.3 Experimental Procedure

The goal of these experiments was to gain empirical data on all of the discussed RF technologies in order to assess the suitability of their spatial and temporal properties for use in an IFRTLS. In this section, I discuss the experimental procedure utilized in gathering and processing this data. This includes the location, equipment utilized, and process by which data was analyzed.

4.3.1 Data Collected

My experimental setup was designed to continually record relevant features from WiFi (802.11b/g), GSM, CDMA, FM, PLP, and WPLP signals. For all *but* CDMA, the relevant feature is the amplitude of the signal from various transmitters. For CDMA, the relevant feature is the time offset of the arrival of a basestation's pilot signal from its expected arrival time.

4.3.2 Location

Experiments were performed at three different locations: two single-family homes, and a a one-bedroom apartment. Details on each test location are provided in Table 4. Tests of long-term temporal stability for all technologies discussed above were performed at each of these locations. The data collection equipment was placed in a central location in each of the three test environments and remained stationary throughout the long-term stability tests. In `house-1` and `apartment-1`, data was gathered in the living room. In `house-2`, data was gathered in the kitchen.

Subsequent to the long-term stability tests, two 0.9–by-0.9 m grid-level site-surveys were performed at `house-1` in order to compare grid-level spatial differentiability and temporal stability. A total of 30 grid-points were laid out on the first floor of `house-1` for these site-surveys.

Table 4: Locations where tests were conducted.

Name	Style	Floors	Size (sq. m.)	Year Built	Test Length (hrs)	Occupants
house-1	1 Family Home	3	371	2003	106.2	1 Adult
house-2	1 Family Home	3	464	2001	169.8	2 Adults/3 Children
apartment-1	1 Bed Apartment	1	37	1969	119.5	1 Adult

4.3.3 Time Period

Long-term temporal stability tests were conducted for a total of over 16 days of continuous sampling across each of the three locations. The two grid-level site-surveys at `house-1`were conducted roughly 24 hours apart and represent approximately one minute of data capture at each grid-point.

4.3.4 Equipment Setup and Data Collection Procedure

Each technology required its own specialized equipment to monitor the relevant feature. For technologies such as GSM and CDMA that monitor ambient signals, I needed only a receiver. For the other technologies, I also deployed transmitters in the test environment. All receiving equipment was co-located on a mobile plastic cart, shown in Fig. 22. A diagram of the entire experimental setup is shown in Fig. 21. I now discuss the specific setup for each of the five technologies I tested.

4.3.4.1 WiFi (802.11b/g)

I created a Windows XP application using C# that logs the signal strength (RSSI) of all visible WiFi access points at a one-second interval. The application also gathers other relevant information about an access point such as MAC address, channel, and SSID. Data is recorded along with a timestamp in a comma-separated text file. I ran this application on an IBM ThinkPad laptop with a built-in 802.11b/g card based on an Atheros chipset. The application uses the Windows Network Driver Interface

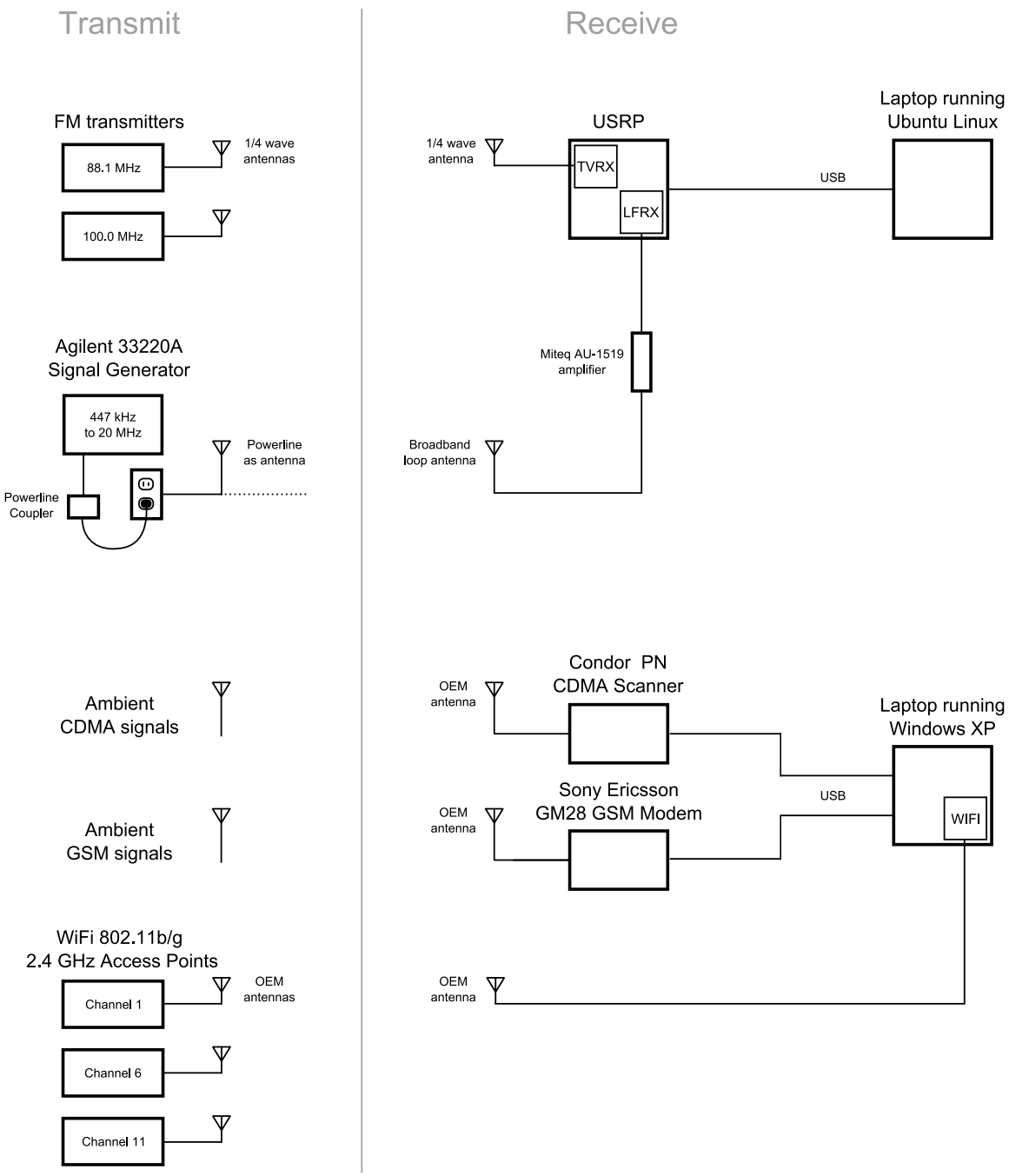

Figure 21: Diagram of the equipment setup for all five technologies that were tested. All receiver equipment was co-located on a mobile plastic cart (see Fig. 22). Transmitter equipment was distributed throughout the test environment.

Figure 22: The mobile plastic cart containing all receiver equipment.

Specification (NDIS) and DeviceIOControl API call to force the network card to perform a scan for available access points every second and to update the RSSI for each.

I deployed three 802.11b/g access points at each test location. One was tuned to channel 1, one to channel 6, and one to channel 11 — out of the 11 total channels (in the US) these are the only three channels with non-overlapping spectrum in the WiFi standard. This ensured that my application would at least have these three access points to monitor. I also collected data on ambient access points.

4.3.4.2 GSM

Ambient GSM cellular signals were monitored using a Sony/Ericsson GM28 GSM modem — the same modem used in previous work on GSM-based indoor positioning [69]. I used an AT&T Wireless SIM card with the modem. The GM28 exposes two API calls for gathering the relevant data: *cellsAPI* and *channelsAPI*. *cellsAPI* returns signal strength data for the six strongest GSM cells that the modem detects.

channelsAPI returns signal strength data for up to 35 channels, six of which are typically associated with the six strongest cells. To communicate with the GSM modem, I created an application that ran on Windows XP. This application issued the relevant API calls to the modem over a USB serial interface, and recored the response, along with a timestamp, in a comma-separated text file. Data for both *cellsAPI* and *channelsAPI* was continually recored at five-second intervals.

4.3.4.3 CDMA

Ambient CDMA cellular signals were monitored using a Condor CDMA scanner made by Berkeley Varitronics — the same scanner used in the CILoS work [85]. The Condor was connected via USB to a laptop running Windows XP. Once tuned to a particular channel, software provided with the Condor can continually record the PN delay (the relevant fingerprinting feature for CDMA-based localization) and Ec/Io (signal strength) of all basestations on that channel. To find active channels to monitor, I used a Motorola E815 mobile phone that operates on the Verizon Wireless network, a CDMA network operator in the United States. The E815 can be put into an engineering mode that exposes the current channel the phone is using. Given that one of the motivations for using CDMA fingerprints for an indoor RTLS is to utilize existing mobile phones already carried by people as a location tag, I felt that using the channel picked by an actual phone, rather than scanning through all channels with the Condor, was the most appropriate approach. Channels were chosen independently at each test location, and the Condor remained on the same channel at each location throughout the data logging process. Samples were recorded approximately once per second. The goal was to recreate the data gathering process employed in the CILoS work [85].

4.3.4.4 FM

FM signals were monitored using a Universal Software Radio Peripheral (USRP) made by Ettus Research connected via USB to a laptop running Linux. The first-generation USRP I used has four 64 Megasample/second analog-to-digital converters (ADCs), meaning that signals of up to 32 MHz can be directly sampled by the ADC without any additional front-end radio hardware. Since the 88–108 MHz band used by FM radio is well above this 32 MHz limit, I required additional radio hardware to first down-convert the signal to baseband before the ADC samples it. For this, I used the Ettus Research TVRX card, which is a 50–860 MHz receiver daughter-card for the USRP based on a commercially available TV tuner module [3]. I used a simple $\lambda/4$ length of wire as the TVRX's antenna.

The TVRX card was tuned to approximately 12 kHz lower than the signal of interest. Thus, in the baseband converted signal, the FM signal of interest was observable 100 FFT bins to the right of the center frequency with a 32,768-point FFT and a 4 Megasample/second sample rate. Each sample was an average of 122 contiguous 32,768-point FFTs — almost exactly one second of data at a 4 MHz sample rate. Incoming data from the USRP was processed by my custom GNURadio block and recorded to a log file along with a timestamp.

The goal was to recreate the FINDR work, and so I deployed my own FM transmitters at each location rather than monitoring ambient FM radio station signals [70]. I deployed two transmitters, off-the-shelf FM modulators designed for use with portable audio players, at each location. One was a Belkin TuneCast with four possible frequencies: 88.1, 88.3, 88.5, and 88.7 MHz. The other was an Inland 668A, which was tunable to any frequency from 87.5–108.0 MHz. The transmitters were initially designed to be battery operated, and so to permit long-term unattended testing I modified them to run off power from a standard AC transformer plugged into the wall. Rather than connecting an audio source to each transmitter as designed, I

soldered a $\lambda/4$ long piece of wire to the connectors in order to extend the transmitters' range. Small FM modulators like these are designed for very low power short range transmission and typically will not be detectable throughout a standard house without a modification such as this to make the antenna more efficient.

Since the same USRP was used for monitoring both FM signals and WPLP signals, the sample rate for both was limited by the time to sample and switch between the 24 frequencies that were tested in total between FM and WPLP. The overall sample rate for each FM signal monitored was approximately 1 sample / 38.4 seconds. Although the FINDR work [70] that inspired us to explore the temporal stability of FM signals used a Nokia N800 tablet with a built-in FM tuner as the receiver, the authors of that work point out that a limitation of the N800 is that its FM tuner senses received signal strength as only 16 discrete levels. Additionally, in their experiments the N800 ran out of memory after only four hours of continuous testing, preventing them from obtaining longer-term temporal stability data. Using the USRP as a receiver instead allowed much more precise measurements and allowed us to utilize the greater storage capacity of a standard computer.

One of the drawbacks of using FM radio signals for a fingerprinting-based indoor RTLS is that the bandwidth of interest can be very crowded with licensed users (radio stations), particularly in urban environments. I discovered that finding an unused frequency without much interference from commercial radio stations was definitely a challenge at my three test environments, all within the city of Atlanta. I monitored a number of frequencies and eventually settled on two for use in all three locations: 88.1 MHz and 100.0 MHz. Thus, the Belkin TuneCast was tuned to 88.1 MHz and the Inland to 100.0 MHz. The two transmitters were placed in separate rooms at each test location, away from the receiver cart. I focused long-term temporal stability analysis on the amplitude of the signal from these two in-home transmitters. For the grid-level analysis however, I also monitored the amplitude of seven different ambient

commercial FM radio stations: 92.9 MHz, 94.1 MHz, 96.1 MHz, 98.5 MHz, 99.7 MHz, 101.5 MHz, and 103.3 MHz. As will be shown, the presence of these stations may not be a drawback as this provided additional features that improved the grid-level accuracy of an FM-based localization scheme.

4.3.4.5 PLP and WPLP

The low-frequency power line radiated RF signals used by PowerLine Positioning [73] and Wideband PowerLine Positioning [80] were monitored using the USRP, similar to the method for monitoring FM signals. The procedure and equipment setup were modified slightly from the procedure used in Chapter 3 in order to reduce the time necessary to capture a full fingerprint.

Since the signals of interest for PLP and WPLP are below the 32 MHz Nyquist frequency for the USRP's 64 MHz ADCs, a tuner module is not necessary to down-convert these signals. I thus used the Ettus Research LFRX USRP daughter-card, which couples its input directly to the ADC's input through a differential amplifier. I used a custom-built loop antenna as the receiving antenna, which was connected to the LFRX card through a Miteq AU-1519 low-noise amplifier.

As with the FM signals, the signal was sampled at 4 MHz then down-converted such that the signal of interest was present at 12 kHz. A 32,768-point FFT was used to measure the strength of the signal of interest. Down-conversion of the signal to baseband was done via software in the USRP's FPGA for these signals (rather than via an actual tunable crystal like that of the TVRX card). This is possible since the signals of interest are all below 32 MHz.

To generate the signals to be transmitted, I used an Agilent 33220A 20 MHz signal generator to produce unmodulated carrier signals of between 8 V and 10 V peak-to-peak (the 33220A's capabilities vary by frequency) at 447 kHz, 600 kHz and 1 MHz–20 MHz in 1 MHz steps. This represents half the frequencies used in the

WPLP work. I chose to reduce the number of frequencies sampled to increase the amount of data gathered for each frequency since generally only one frequency can be monitored at a time. The arbitrary waveform capabilities of the 33220A were utilized to produce a signal consisting of three signals at once — for example, 1 MHz, 2 MHz, and 3 MHz. This allowed the time to capture a fingerprint to be cut significantly. The signal generator injected these signals into the power line using a coupling box designed to isolate the signal generator from the 60 Hz US power signal. The coupling box and the signal generator were the same units discussed earlier in Chapter 3. In `House-1` and `Apartment-1`, the signal injector was located in the same room as the receiver. In `House-2`, the signal injector was located in an office on the opposite site of the floor from the receiver.

4.4 Long-Term Temporal Stability

In this section, I analyze the extensive data that was collected for each of the five RF technologies. The research objective was to determine whether one of the tested RF technologies stands out as having inherently more stable features. To determine this, I present the standard deviation of the key features that were captured for each technology, the average standard deviation across all features for each technology, and plots of the raw data for each feature across the entire multi-day datasets.

When examining plots of raw data, note that the range of the y–axes of these plots is fitted to the data for each plot and must be taken into account when comparing plots. I chose to present the data in this manner to give the most close-up view possible of the temporal variations in each signal. These plots show both the raw data (as dots) as well as a 1-minute time averaged version of the raw data. This averaged version was meant to test the hypothesis that while instantaneous samples may be unstable, the mean over short periods of time is stable in the long-term. Unfortunately, most of these plots illustrate that this is not the case. Even if it were

the case, however, an IFRTLS that needs to time-average fingerprint features will be limited in its location update rate by that time-averaging period. This means that if the feature a system uses is stable only over one-minute periods, the system could only determine the location of a mobile tag once each minute, and would require the tag to be motionless during that time. This will severely limit the usefulness of these systems for many applications.

4.4.1 WiFi

Although I captured data on all WiFi access points visible to the scanning laptop, I focus the analysis here on the three access points I deployed in each environment that were tuned to the three non-overlapping WiFi channels. Table 5 shows the standard deviation (σ) in the RSSI for each of these access points at each of the three test locations across the entire length of the test. The length of the test at each location was shown in Table 4 previously. I also show the standard deviation of a time-averaged version of the signal where each sample represents an average of one minute of data (60 samples in this case since the sample rate of WiFi in my tests was once per second). Figure 23 shows both the raw samples as well as the time-averaged samples for each access point at each location over the entire sample period.

The plots in figure 23 illustrate that temporal stability does not hold for the signal strength of WiFi access points over the long term. Note the significant dispersion of the raw samples (shown by the blue dots), which is on the order of 20 dB or more for each of these. Although many of these signals are reasonably stable on the order of hours, when we look at the entire data set shown in this plot we see that this same stability does not hold on the order of days. Even `(Apartment-1, Channel-1)`, which shows a relatively low standard deviation of 1.58 dB has a dynamic range of about 15 dB for its raw samples (even with removing the visible outliers in the plot). If we look at the time-averaged version of this plot, where each point represents the average

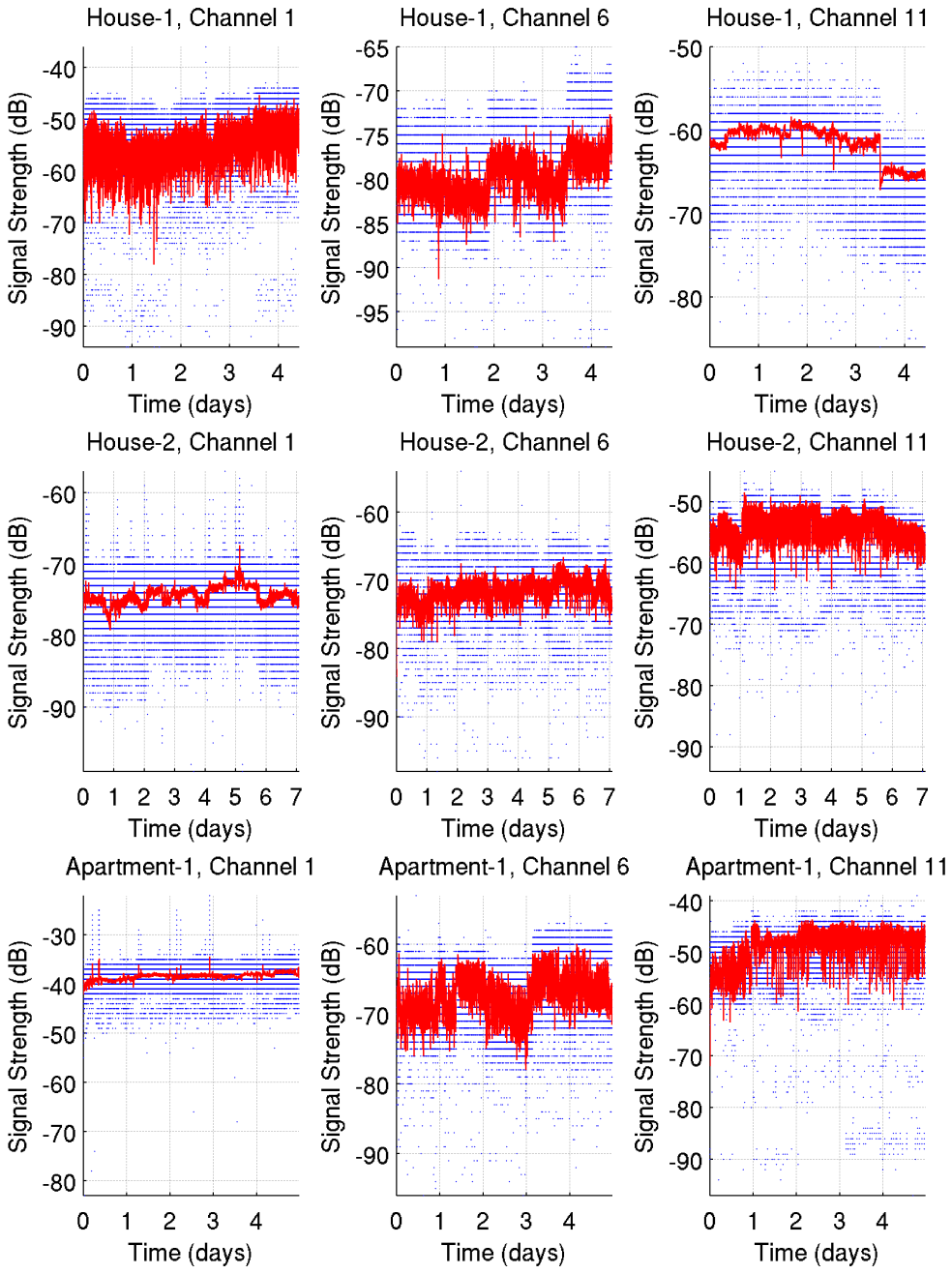

Figure 23: Plots of the raw data for all three WiFi access points that were deployed at each location. The channel assignment for each physical access point remained constant across all locations. Samples were taken at a one second interval. Dots in blue represent the raw data samples and the line in red represents a 1-minute time average of the raw samples.

Table 5: Standard deviation of the amplitude of all three deployed WiFi access points, by location. The overall average (last row) is weighted by the different number of samples at each location and for each access point.

Location	AP Channel	σ RSSI (dB)	1-Minute Time Averaged σ RSSI (dB)	Test Duration (hrs)
house-1	1	6.09	3.50	106.2
	6	2.99	2.22	
	11	2.55	2.06	
house-2	1	2.31	1.20	169.8
	6	2.28	1.45	
	11	3.27	1.93	
apartment-1	1	1.58	0.67	119.5
	6	4.12	2.82	
	11	4.39	3.12	
Average	—	3.18	2.02	131.83

of 1-minute of samples (60 samples in this case), the stability looks a bit better for `(Apartment-1, Channel-1)`, however the other locations and channels still exhibit significant temporal variation on the order of a 15 dB dynamic range. The more troubling issue these plots illustrate though is not the dispersion of the data, which can be reduced by time averaging if the mean is consistent, but the inconsistent mean over time. For example, `(House-1, Channel-11)` shows a noticeable and sudden drop of about 5 dB in the mean of the RSSI around day 3.5 that holds through the end of the dataset. Interestingly, this same shift in mean is also present for `(House-1, Channel-6)` at the same time, although this access point *increases* in mean RSSI rather than decreasing as with channel 11, suggesting that an environmental factor may have caused this shift. Similar shifts in mean can be seen at the other two test locations.

4.4.2 GSM

Table 6 shows the standard deviation of the top six strongest GSM cells at each location, as reported by the *cellsAPI* API of the GSM modem. Figure 24 shows plots

Table 6: Standard deviation of the amplitude of the top six GSM cells for each location. The overall average is weighted by the different number of samples at each location and for each cell ID.

Location	Cell ID	σ RSSI (dB)	1-Minute Time Averaged σ RSSI (dB)
house-1	0x7BFD	3.39	2.87
	0x7BFF	5.28	4.82
	0x7BFE	3.17	2.85
	0x7C03	3.90	3.30
	0x7C05	3.14	2.60
	0x780D	1.89	1.51
house-2	0x7BDF	2.80	2.20
	0x7BFF	2.52	1.94
	0x7789	2.21	1.94
	0x780D	1.81	1.23
	0x7D52	2.39	1.55
	0x7BE5	2.08	1.62
apartment-1	0x7789	2.80	2.30
	0x0000	2.61	2.16
	0x778A	2.56	2.30
	0x778F	3.33	2.74
	0xEF39	2.37	2.10
	0x778B	1.94	1.54
Average	—	**2.77**	**2.29**

of the raw (5-second intervals) and 1-minute (12 sample) time averaged data for the top two strongest cells at each location. As with WiFi, the plots illustrate temporal instability of the amplitude of these signals, although the standard deviation of the raw signals for GSM is slightly less than that of WiFi. As with WiFi, the most concerning result here is not the general dispersion of the samples, which could be reduced by sufficiently time-averaging, but the trend in the mean over time. Cell `0x7BFD` at `house-1`, for example, shows periods where the mean shifts by 5 dB or more and remains there for a period of time.

I also observed that a consequence of the GSM modem only returning the top six cells is that the set of six cells is prone to change over time as some of the weaker cells

Table 7: Statistics for the five CDMA basestations that were found with an Ec/Io strength sufficient to obtain reliable PN delay data from amongst all CDMA data collected over the three test locations. Samples were taken approximately once per second.

Location	Basestation	σ PN Delay (chips)	σ Ec/Io (dB)	1-Minute Time Averaged PN Delay (chips)	1-Minute Time Averaged σ Ec/Io (dB)
house-1	160	9.66	2.06	1.68	1.08
house-2	304	5.17	2.63	1.25	1.64
	248	8.96	2.58	2.28	1.43
	144	9.68	2.20	1.89	1.17
	148	11.24	2.06	2.31	1.1
Average	–	**8.74**	**2.34**	**1.88**	**1.32**

move down in rank (whether temporarily or permanently) and cells previously not on the six-strongest list move up. Thus, the set of features exposed by the modem did not remain constant over the entire test duration. I speculate that this is why Otsason *et al.* [69] had significantly improved results for their GSM fingerprinting-based localization system by using a wide fingerprint that included up to 35 channels from the *channelsAPI* as well as the six strongest cells.

4.4.3 CDMA

Of the five technologies I chose to monitor, CDMA is the only one that does not use amplitude as a fingerprint.

As with the other RF technologies, I recorded CDMA measurements continuously at each of my three test locations. When post-processing the data, I found that only five basestations in total (out of a potential 511 on each channel) produced usable PN delay data — one at location `house-1` and four at location house-2. The signals from the remaining basestations were too weak to be of use.

I chose basestation 304 at `house-2`, the one amongst the five with the most stable delay, for a more thorough analysis. Figure 25 shows the raw PN delay values and Ec/Io over the entire seven-day sample period. The vertical gaps in the data represent

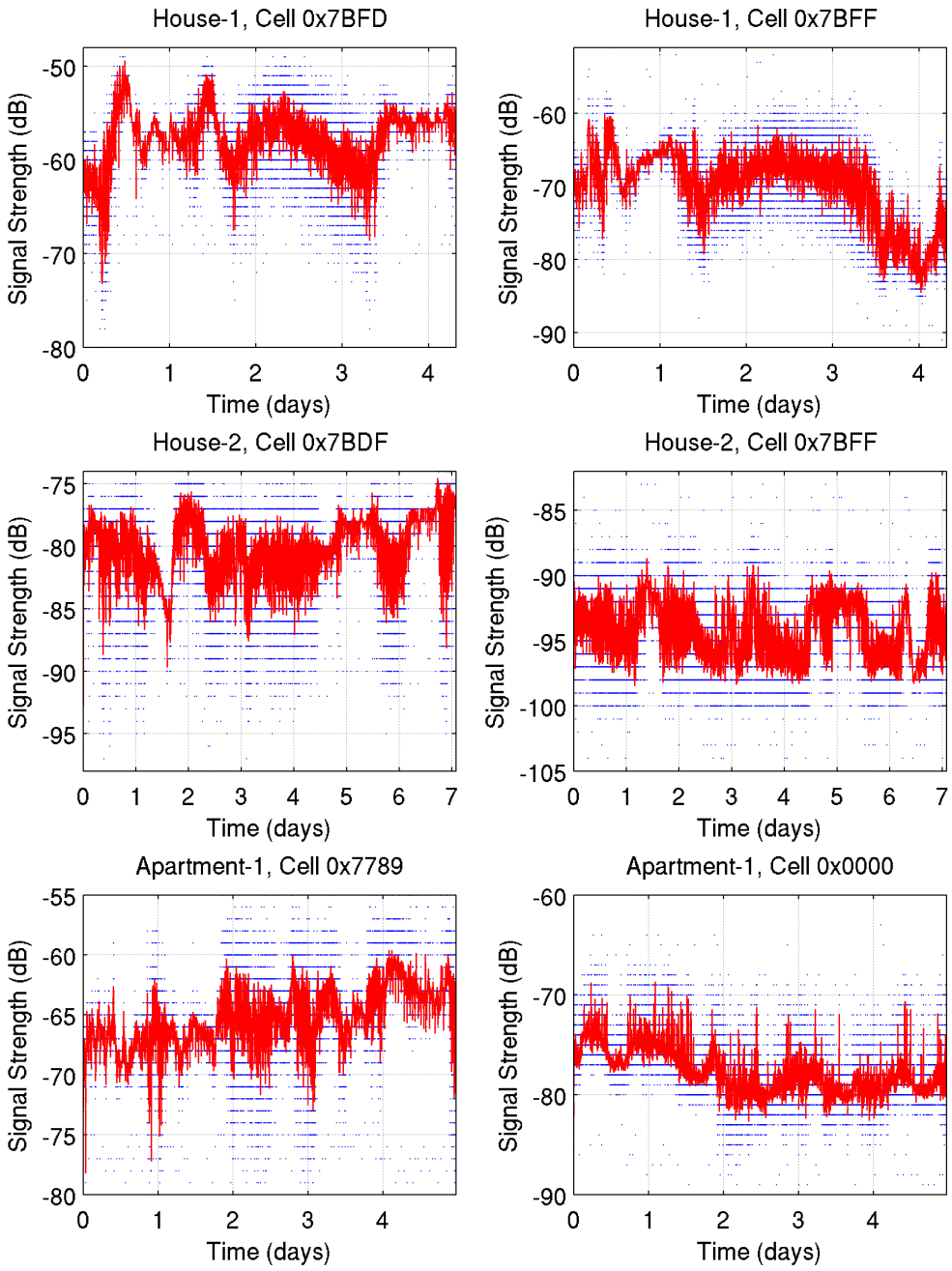

Figure 24: Plots of the raw data for the top two strongest GSM cells at each location. Samples were taken at five second intervals. Dots represent the raw data samples and the line represents a 1-minute time average of the raw samples.

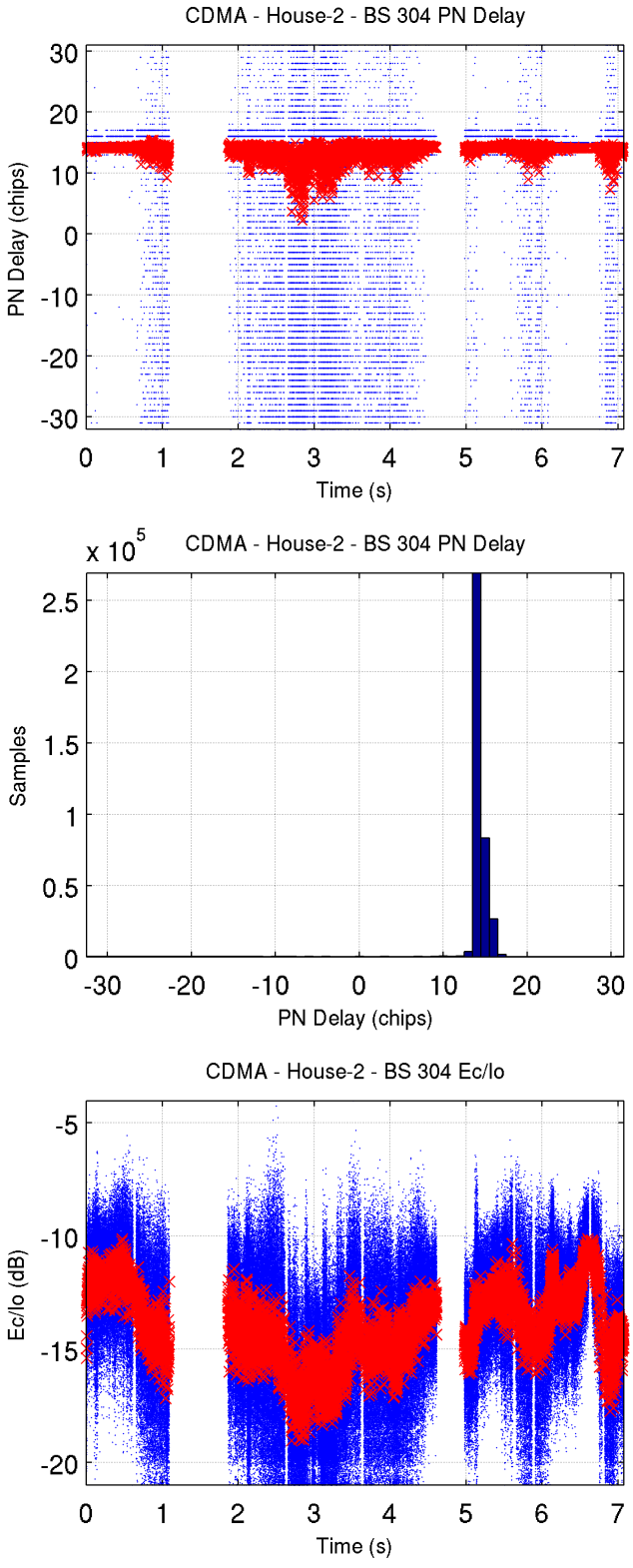

Figure 25: CDMA basestation ID 304 at house-2. This was the strongest observed basestation at this location. The top plot is raw PN delay values. The bottom plot is the raw Ec/Io values. Only samples with an Ec/Io above -21 dB are considered.

Table 8: PN delay accuracy degradation as a function of Ec/Io for basestation 304. The correct PN delay is considered to be the most common delay observed for the entire dataset for this basestation – 14.

Ec/Io (dB)	Samples with Ec/Io ≤	Correct Samples	Percentage Correct	Hmys of Bad Data
-19	26,853	5,327	19.8%	5.98
-20	13,449	1,376	10.2%	3.35
-21	7,005	446	6.4%	0.12

times when the proprietary Condor data logging software crashed while recording — the proprietary software had difficulty handling the large log files that were generated during long-term testing and required periodic attention to stop and re-start logging in a new file. The topmost plot shows the raw PN delay values for basestation 304. Samples with an Ec/Io of less than -21 dB were disregarded due to their high potential for inaccurate PN delay readings. Table 8 demonstrates this, as only 6.4% of the data with Ec/Io below -21 dB were correct. The middle plot of Fig. 25 shows that despite the fluctuations we see in the above plot, there is a stable and consistent PN delay of 14 from basestation 304. This could be better detected through windowing and filtering as done by ur Rehman *et al.*, but there needs to be a large enough sample size to accurately do so (about 20 readings [85]). The other four basestations at both homes produce similar results.

Although Ec/Io (signal strength of a basestation's pilot signal divided by the total power in the channel) is not itself a stable feature, I found it to be a good indicator of the stability of the PN delay samples for each basestation — an Ec/Io of at least -20 dB seems to be the cutoff for reasonable stability. Table 7 shows that PN delay stability is not a step function once Ec/Io reaches -20 dB though — note that the standard deviation of the delay steadily decreases as the mean Ec/Io increases.

While the relevant feature of the one basestation observed does appear to be

Table 9: Standard deviation of the amplitude of each FM signal monitored, by location. The overall average is weighted by the different number of samples at each location.

Location	Frequency (MHz)	σ Amp. (dB)
house-1	88.1	0.94
	100.0	1.16
house-2	88.1	1.16
	100.0	1.94
apartment-1	88.1	1.32
	100.0	0.50
Average	—	**1.22**

reasonably stable, CDMA-based localization has its own unique challenges, such as a lack of availability of a sufficient number of basestations to fingerprint. At the three test locations, one location did not detect any usable basestations, and another detected only one.

4.4.4 FM

The two FM signals of 88.1 MHz and 100.0 MHz exhibit the best temporal stability of any of the technologies discussed thus far. The standard deviation of each is shown in Table 9. Plots of the raw data are shown in Fig. 26. Note that for FM, I do not provide a time-averaged version since samples are only taken every 38.4 seconds due to sharing the USRP with WPLP. Each sample represents one second of data.

Looking at Fig. 26, I see that even the raw un-averaged data looks more stable than any of the features from WiFi or GSM. Although definite shifts in the mean of some of these signals can be observed (such as with `(Apartment-1, 100.0 MHz)` at the end of day one and the first half of day three), I see that these shifts are only one or two dB compared to the 5 dB shifts seen previously.

An interesting observation is that while both 88.1 MHz and 100.0 MHz are stable at `house-1` and `apartment-1`, they exhibit greater variance at `house-2`. `House-2`

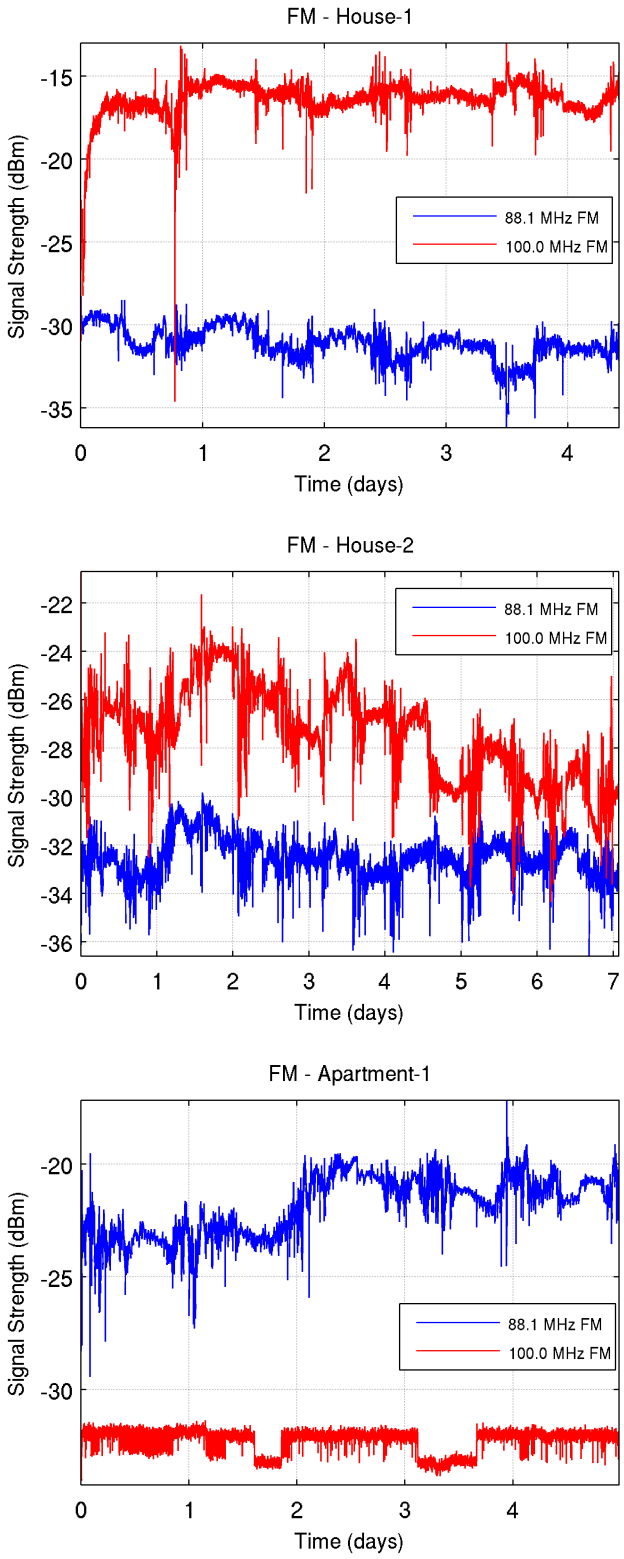

Figure 26: Raw signal strength of the two FM transmitters at each of the three test locations. Samples were taken approximately once every 38 seconds. Each sample represents the mean amplitude over exactly 1 second of data.

Table 10: Standard deviation of the amplitude of the 22 WPLP frequencies that were tested, by location. The overall average is weighted by the different number of samples at each location.

	House-1		House-2		Apartment-1	
Frequency (MHz)	σ Amp. (dB)	Avg. Amp. (dBM)	σ Amp. (db)	Avg. Amp. (dBM)	σ Amp. (dB)	Avg. Amp. (dBM)
0.447	0.32	-20.15	1.51	-40.57	0.72	-32.24
0.600	0.31	-16	1.25	-33.79	0.71	-25.1
1.0	**0.25**	-8.13	1.91	-37.94	0.71	-16.67
2.0	0.31	-1.72	1.33	-33.66	**0.69**	-11.69
3.0	0.35	-2.1	**5.11**	-20.43	0.73	-12.2
4.0	0.31	2.3	3.8	-24.41	1.00	-18.37
5.0	0.31	5.78	2.02	-13.61	0.72	-7.87
6.0	0.38	-3.12	3.71	-28.58	0.80	-16.76
7.0	0.31	2.26	1.59	-18.49	0.88	-17.65
8.0	0.37	-2.35	3.78	-26.88	0.81	-7.36
9.0	0.37	-2.09	3.12	-27.69	0.74	-13.98
10.0	0.38	-0.36	3.36	-33.48	1.50	-16.53
11.0	0.42	-0.77	1.76	-22.23	1.03	-9.83
12.0	0.34	1.78	1.55	-22.57	0.96	-14.07
13.0	0.43	1.64	1.32	-26.52	0.84	-20.1
14.0	0.56	-5.66	1.59	-31.96	0.84	-22.74
15.0	0.35	-8.83	1.32	-27.91	1.26	-18.03
16.0	0.47	-10.78	**0.86**	-23.71	1.24	-20.55
17.0	0.83	-11.2	1.74	-11.99	**2.17**	-15.06
18.0	1.53	-17.09	1.9	-20.71	1.02	-13.82
19.0	0.74	-20.2	1.05	-20.39	1.07	-19.71
20.0	**4.17**	-14.74	4.27	-24.19	0.94	-19.6
Average	**0.63**	—	**2.22**	—	**0.98**	—

was the most active of the three locations tested, with five occupants, three of which are children. I speculate that some of this variance may be due to human activity within the space, however evaluation of this hypothesis is left to future work.

4.4.5 Wideband PowerLine

The standard deviation in amplitude of the 22 PLP and WPLP frequencies that were tested is the lowest of any of the five technologies tested. The standard deviation of each is shown in Table 10 by location. The highest and lowest standard deviation

frequencies for each location are highlighted in bold, and plots of these are shown in Fig. 27. Again, it is important to note the scale of the y-axes on these plots as compared to previous plots from less stable technologies. For the three lowest standard deviation frequencies, most of the raw data points fall within a range of just 2 dB. As with FM, each sample represents one second of data, and I do not provided time-averaged data for PLP since samples were taken 38.4 seconds apart.

A number of interesting observations can be extracted from Fig. 27. First, I see that while the lowest standard deviation frequencies (left column of plots) are quite stable, the high standard deviation frequencies are also reasonably stable except for certain periods. For 3.0 MHz at `house-2` in particular, I see a very time-dependent pattern of instability. The amplitude consistently drops by about 10 dB at nearly the same two points in each day. Interestingly, as with FM, `house-2` has the highest average standard deviations amongst the three locations tested. I again speculate that the higher average standard deviations and the time-dependent instability in certain frequencies is an artifact of the higher level of human activity at this test location, but leave a thorough examination of this to future work.

4.5 Effect of Signal Variation on Location Classification

The previous section examined the long-term temporal stability of the raw RF signals that constitute the features for fingerprinting of an IFRTLS. In this section, I examine the implication of the observed variation on location classification, similar to the analysis performed for PLP and WPLP in Chapter 3. These experiments were performed at `house-1`, where a 30-point grid was laid out on the first floor with spacing of 0.9 m between points. Two independent site surveys were then conducted approximately 24 hours apart. During these site surveys, data for PLP, WPLP, FM, GSM, and WiFi was captured at each grid point for approximately one minute. CDMA was not tested on the grid since in the previous long-term stability experiments only one

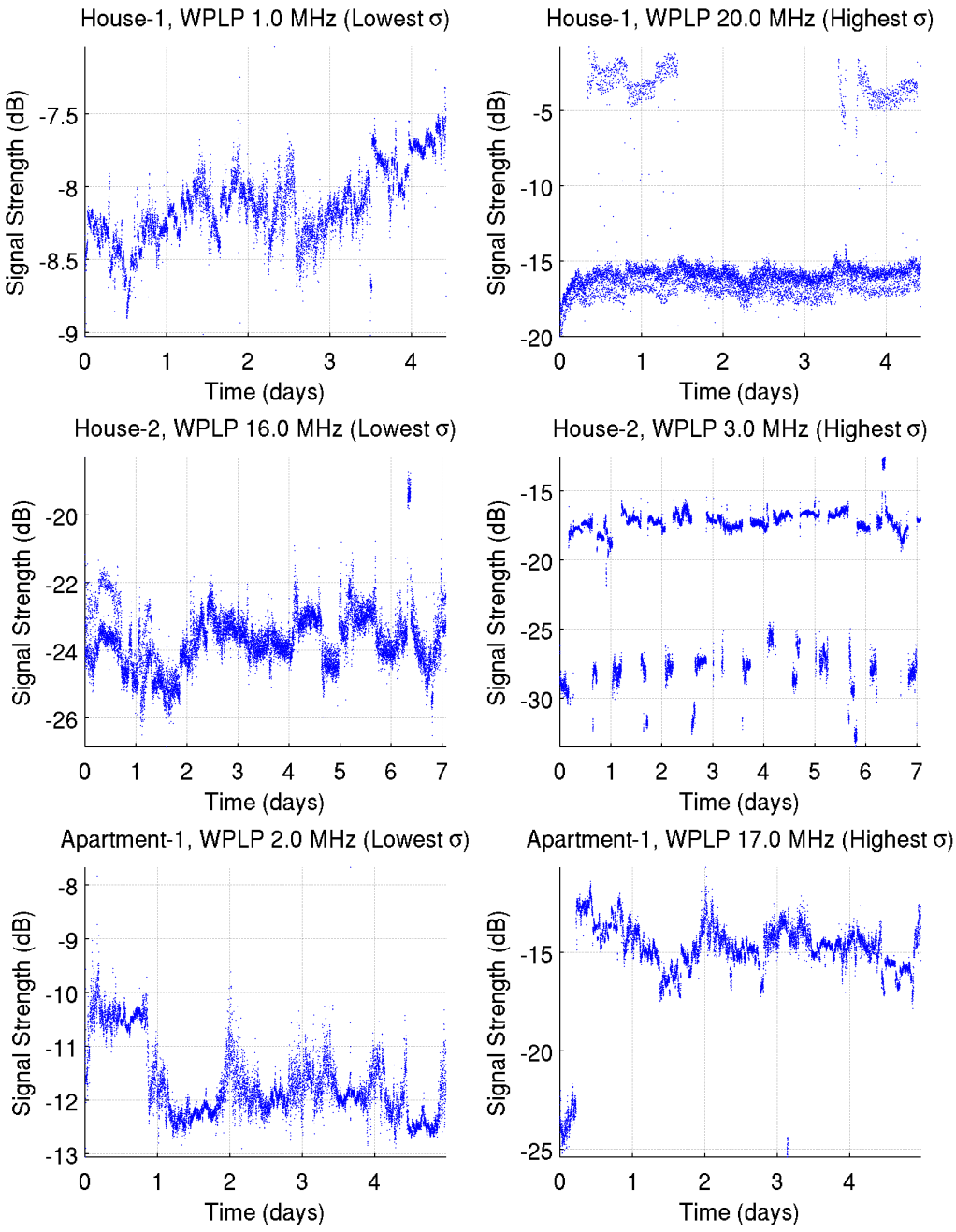

Figure 27: Plots of the raw data for the WPLP frequency with the lowest and the frequency with the highest standard deviation in amplitude for each location.

CDMA feature was observable at `house-1` — at least two features are necessary to form the fingerprint of an IFRTLS.

Data was entered into a SQL database which was used to extract the mean and standard deviation of the amplitude of each feature at each grid point. For PLP, WPLP, and FM, two samples were collected in the roughly one minute capture at each grid point. WiFi samples were taken approximately once each second and GSM once every 30 seconds, so the fingerprints for these technologies in this section show more time-averaging than those for PLP, WPLP, and FM.

4.5.1 Feature Selection

Chapter 3 demonstrated the value of a wideband fingerprint in both improving classification accuracy immediately post-deployment of an IFRTLS as well as in maintaining that accuracy over time. In this chapter, my objective is instead to perform a more direct comparison of the various RF technologies that have been used for fingerprinting. To do this, I utilized the feature selection technique of Otsason and Varshavsky *et al.*, wherein a greedy feature selection algorithm was used to prune the fingerprint [69, 87]. The initial "wideband" fingerprint for each technology, consisting of all features that were observed for that technology, was pruned by removing from the fingerprint the feature that produced the largest accuracy gain when removed. If none of the features produced an *increase* in accuracy when removed, then any feature that did not *decrease* the accuracy when removed would be selected instead. This process was repeated until removing any feature resulted in an accuracy reduction. This method of greedy feature selection was generally able to slightly improve accuracy, and typically showed a significant reduction in the number of features in a fingerprint. The results of a grid-level classification for each technology are shown in Table 11. The average error distance for each technology is shown in Fig. 28. In this particular test environment WPLP is able to produce 100% accuracy, followed

Table 11: Number of features and grid-level accuracy for each of the five technologies with and without feature selection (labeled as "FS" in the columns). Fingerprinting schemes utilizing only amplitude, only standard deviation of the amplitude, and a hybrid of the two are shown for each. The best scheme for each technology is shown in bold.

		# of Features		0.9x0.9m Grid-level Accuracy (%)	
		Un-pruned	FS	Un-pruned	FS
FM_all	amplitude	9	7	60.00	73.33
	σ	9	5	6.67	16.67
	hybrid	18	10	46.67	76.67
WPLP	**amplitude**	23	8	90.00	100
	σ	23	14	16.67	33.33
	hybrid	46	21	63.33	86.67
GSM_channels	amplitude	7	5	3.70	14.81
	σ	7	3	7.41	11.11
	hybrid	14	2	7.40	22.22
WiFi	**amplitude**	6	4	16.67	26.67
	σ	6	4	10.00	13.33
	hybrid	12	8	10.00	20.00
PLP	**amplitude**	2	—	53.30	53.33
	σ	2	—	6.67	6.67
	hybrid	4	2	26.67	53.33

by FM with 73.33% accuracy and an average distance error of 0.62 m.

For each technology, three different types of fingerprints were tested — amplitude only, standard deviation of the amplitude only, and a hybrid of both. The obtainable accuracy for each of the schemes is shown in Table 11. For GSM, I also had the option of using data from only the `cellsAPI` or the `channelsAPI`. As Otason *et al.* previously reported, a wider fingerprint utilizing both performed best. I therefore only present the results of a fingerprint utilizing both cells and channels (which I refer to as `GSM_channels`). For FM, I tested fingerprints based only on the two indoor receivers, only the seven captured commercial broadcast FM stations, and a hybrid of both (all the signals). The hybrid approach performed best, and I therefore only present the results of this approach. I refer to this as `FM_all`. As illustrated in Table 11, an amplitude only approach worked best for PLP, WPLP, and WiFi. A

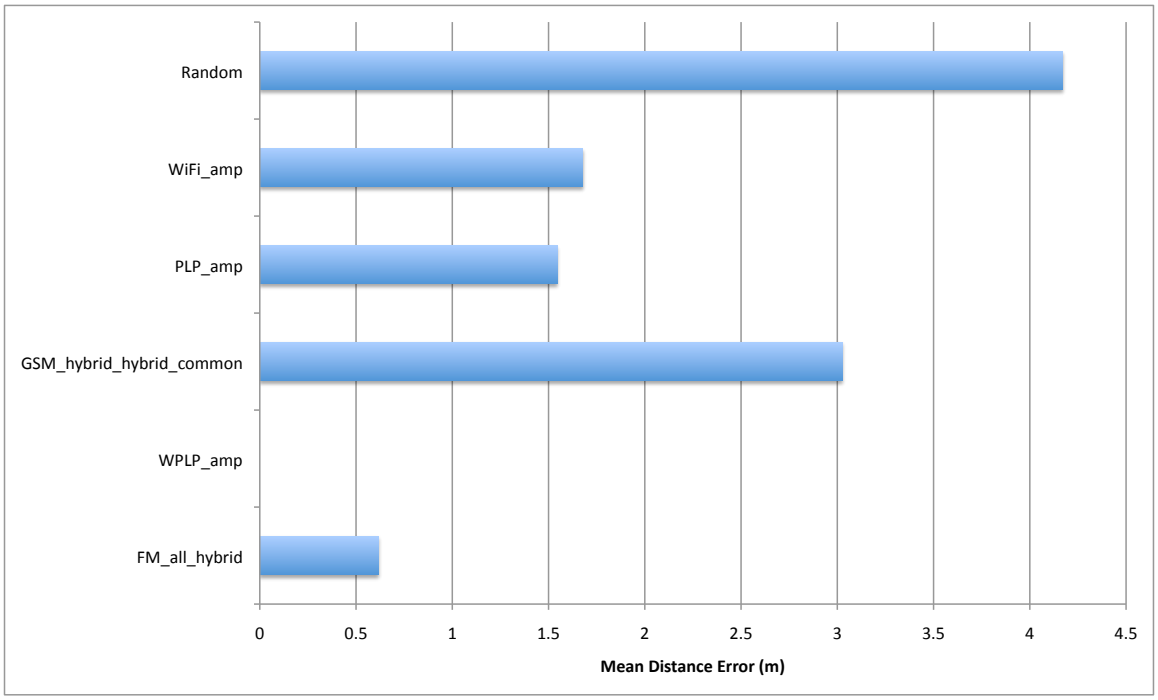

Figure 28: Mean distance error for the optimally pruned configuration of each of the five fingerprinting schemes when trained and tested on a dataset taken one day apart. The datasets were captured on the first floor of `house-1`.

hybrid amplitude / standard deviation of the amplitude approach worked best for FM and GSM. This is a new result that has not been previously discussed in the FM and GSM IFRTLS literature.

4.5.2 Sensitivity Analysis

As was done in Chapter 3 for WPLP, I performed a sensitivity analysis of each of the technologies to see how an IFRTLS based on each will degrade in accuracy with noise in the fingerprint. The results of this analysis are shown in Figs. 29 and 30. These figures show the grid-level accuracy as well as the average distance error for each of the five technologies when a training site-survey is corrupted by varying degrees of Gaussian noise. The classifier was trained on the original dataset and tested on the corrupted version. This analysis tests both the spatial differentiability of these systems, as well as their potential for temporal stability through resistance to change in fingerprint features over time. These figures show that an IFRTLS based on WPLP or FM shows superior resistance to noise compared to WiFi and GSM. This results in both increased grid-level accuracy, as well as lower distance errors when an error does occur. These figures also show the results of a random classifier for comparison. In the test environment, a random classifier will on average lead to an error of 4.16 m. This was found through 5,000 trials of each grid point where another random grid point was selected and the distance between the two was calculated. Note that the discrepancy between WPLP's noise sensitivity in Fig. 29 and that presented in Chapter 3 is due to the narrower fingerprint used for WPLP in these experiments to make a more direct comparison of the temporal stability and spatial differentiability of each underlying feature. Noise resistance can be increased by increasing the number of frequencies used in the WPLP fingerprint; the same is true for the other technologies as well.

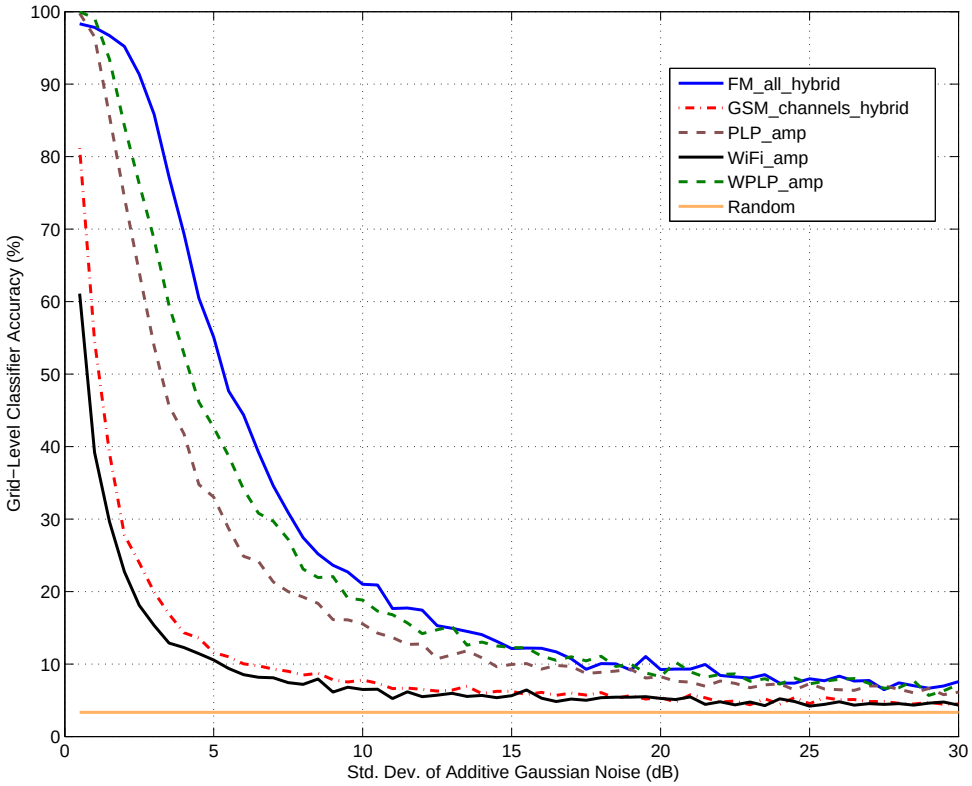

Figure 29: Localization accuracy vs. standard deviation of zero-mean additive Gaussian noise for each of the five technologies. Accuracy is calculated by producing a "noisy" version of a site survey and testing the "noisy" version against the original data, and then averaging results over 100 trials. A line representing accuracy when a point is randomly selected is also shown. This shows that for this environment, FM is able to tolerate the most noise, followed by WPLP, PLP, GSM, and WiFi.

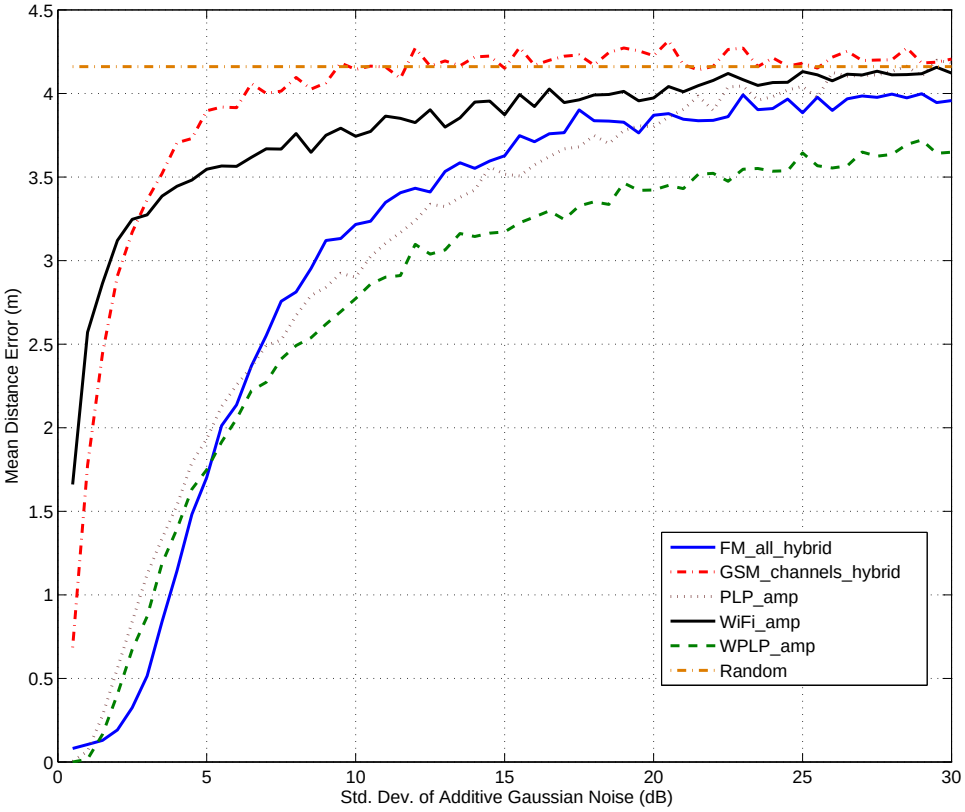

Figure 30: Average distance error vs. standard deviation of zero-mean additive Gaussian noise for each of the five technologies. Accuracy is calculated by producing a "noisy" version of a site survey and testing the "noisy" version against the original data, and then averaging results over 100 trials. A line representing distance error when a point is randomly selected is also shown. This shows that for this environment, WPLP has the lowest distance error past 5 dB of noise, and is close to that of FM below 5 dB. GSM and WiFi produce the highest distance errors.

Table 12: Grid-level accuracy and average distance error based on the sensitivity analysis in Figs. 29 and 30 for the actual deviation observed in the long-term temporal stability analysis previously.

	Observed σ (dB)	Accuracy (%)	Distance Error (m)
FM_all	1.22	97.0	0.16
GSM_channels_hybrid	2.77	22.0	1.75
PLP_amp	0.57	90.0	0.28
WiFi_amp	3.18	15.0	2.50
WPLP_amp	1.28	90.0	0.15

4.5.3 Revisiting Temporal Stability

Figs. 29 and 30 allow us to now interpret the standard deviation measurements performed in Sec. 4.4 and to answer the question of what are the implications of the observed amount of instability on grid-level accuracy and average distance error. The observed instability from Sec. 4.4 is summarized in Table 12, along with the associated grid-level accuracy and average distance error (both found from Figs. 29 and 30). We see that the instability in GSM and WiFi causes extremely low grid-level accuracy, and relatively high distance errors. WPLP, PLP and FM all exhibit high accuracy and low distance errors with the observed instability.

4.5.4 Understanding Classification Errors

The CDF shown in Fig. 31 illustrates the type of errors that can be expected from the different systems in a real-world setting. This represents the distance errors made by the various systems when trained and tested on site surveys taken roughly 24-hours apart. WPLP is represented simply by a vertical line at the origin since it obtained 100% accuracy. FM is the next best system, followed by PLP, WiFi, and GSM. Table 13 shows the median and 95th percentile accuracy for each system, based on the CDF in Fig. 31.

Table 13: Median and 95th-percentile accuracy for the grid-level tests at `house-1`.

	50th-percentile (m)	95th-percentile (m)
FM_all	0.00	4.00
GSM_channels_hybrid	3.16	8.90
PLP_amp	0.00	6.00
WiFi_amp	1.00	6.70
WPLP_amp	0.00	0.00

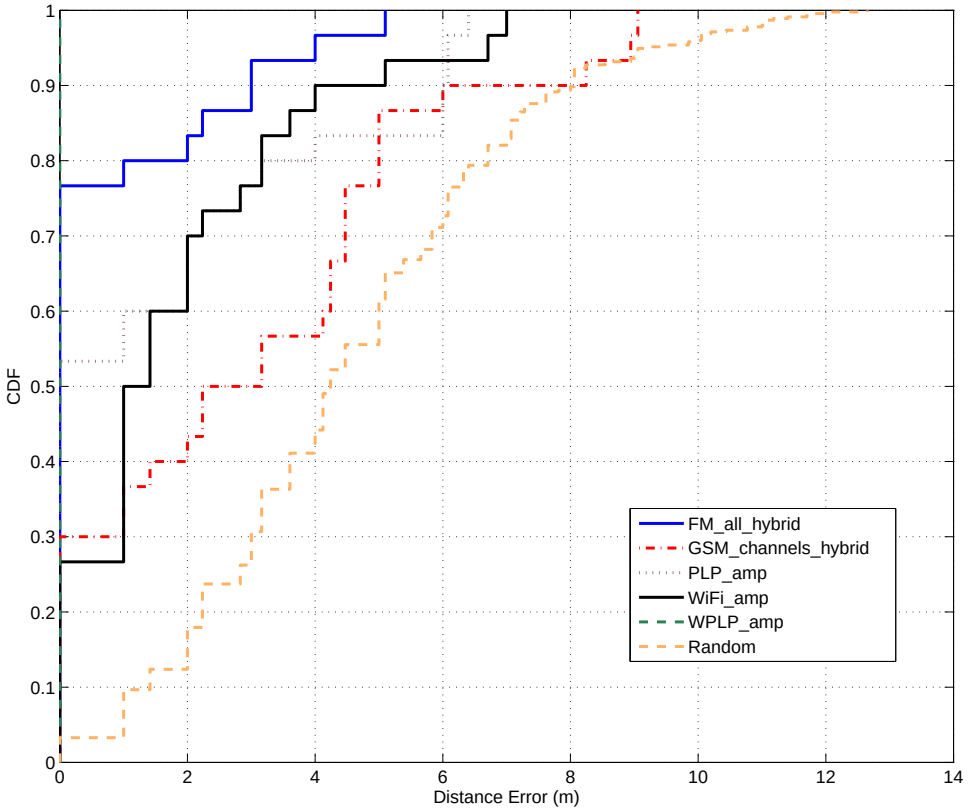

Figure 31: Cumulative distribution function for the error distance for the five different technologies, as well as a random classifier. Data is for the two grid-level site-surveys taken 24-hours apart at `house-1`. Note that the WPLP data is a vertical line at zero distance error since 100% of the samples had no error.

4.6 Summary of Contributions

While fingerprinting has been shown to be a promising approach to indoor localization over the short term, long-term temporal stability of fingerprinting-based systems has been largely ignored. The typical long-term study of a feature proposed for fingerprinting has previously been on the order of hours, not the days, weeks, and months over which a real-world system must remain stable. To address this, in this chapter I have made the following contributions:

- Over sixteen total days of continuous sampling across three locations were performed for the five leading RF technologies used in fingerprinting-based RTLSs, which lends new insight into the temporal variations that can be expected over the long term in real-world deployments. I believe that this characterization of the real-world long-term variability can be of use to future research in the space that seeks to find methods for accommodating the observed instability.

- I have shown that of the tested technologies, WPLP and FM have the most stable features and are therefore the best options on which to base an IFRTLS.

- I have shown that of the tested technologies, WPLP and FM exhibit the best grid-level accuracy for a 0.9–by-0.9 m grid and lowest mean distance error.

- I have shown that including in the fingerprint the standard deviation of the amplitude for FM and GSM can improve the accuracy of these systems over amplitude-only fingerprints.

CHAPTER 5

ULTRA-LOW-POWERED WIRELESS COMMUNICATION IN THE HOME

In Chapter 3, I showed the value of utilizing a wideband fingerprint for a power line-based IFRLTS. In Chapter 4, I showed that this system offers the best temporal stability and spatial differentiability amongst IFRTLSs with fingerprints based on alternate RF technologies. In this chapter, I address one of the deployment concerns of WPLP (backchannel communication) by demonstrating that the domestic power line is capable not only of *transmitting* RF signals as with PLP and WPLP, but also of *receiving* them as well. This allows mobile location tags to communicate with the central location processing unit without the need for an additional backchannel radio on the tag or additional infrastructure deployment in the environment to support an alternate backchannel. I also discuss how this phenomenon has much broader implications for sensor network deployment in the home by enabling extremely power efficient wireless communication at whole-home range.

5.1 Motivation

A drawback of RTLSs, particularly for use in domestic environments, is their need for the installation of extensive, expensive, and potentially unsightly infrastructure. Perhaps the best example of this is the previously discussed Active Badge system [89], which requires an infrared receiver attached to the ceiling of every room in which the system is intended to operate. One of the benefits of a fingerprinting approach to real-time localization is the potential to use existing features of the environment as the means of localizing mobile tags, thus reducing or eliminating the need for new

infrastructure deployment. Unfortunately, sufficiently temporally stable and spatially differentiable features are rarely available without being generated specifically for the purpose of an IFRTLS [73]. Thus, the potential for reducing or eliminating infrastructure through fingerprinting is rarely realized.

The original PLP work [73] sought to address this by generating a fingerprint feature in a manner that utilized existing infrastructure, thereby combining the benefits of fingerprinting (minimal infrastructure) with the benefits of full control over the infrastructure (not being dependent upon a third party for operation of one's RTLS). Although this work demonstrated the power line infrastructure to be an efficient distribution medium for RF signals forming the basis of a fingerprint, the backchannel aspect of an IFRTLS was largely ignored. The site-survey database is typically not stored on mobile location tags, but rather in a central location processing unit (LPU) which handles all computations involved in locating mobile tags. This means that mobile tags that sense the RF environment to form a location fingerprint, such as those in PLP and WPLP, rely on a backchannel to communicate that fingerprint to the LPU. Since PLP tags form their fingerprint by sensing RF signals emanating from the power line, unlike with WiFi-based IFRTLSs in which tags can utilize their WiFi radio to both receive (to sense the fingerprint features) and transmit (to forward the sensed fingerprint features to the LPU), PLP tags in the original system design were required to utilize some alternate wireless communications scheme (ZigBee) since the power line was never explored as a receiver. This dual-radio scheme increases the cost, complexity, and power requirements of tags. The contribution of this component of the research is to demonstrate the feasibility of utilizing the existing home power line as not only a *transmitting* antenna as with PLP and WPLP, but also as a *receiving* antenna for receiving the fingerprint from mobile tags. This allows tags to operate solely with a power line-based low-frequency low-powered radio.

Although reducing the complexity and power requirements for mobile tags in a

power line-based IFRTLS was the original motivation for this work, I believe the implications of the work to be broader. A consistently emerging theme in ubiquitous computing for the home is that while it is possible to think of the home as a sensor-rich environment, there are practical roadblocks to widespread deployment. Though the research community and others have begun to define more and more compelling applications of sensor-enabled environments, there remain too many obstacles to adoption by average householders. One constant concern is that of power requirements, particularly for wireless sensor networks. Wireless sensors are appealing because they can be placed in a variety of interesting locations, but the argument against them is that their current power consumption, largely attributed to wireless communication, limits the battery life to several months. Deploying tens and hundreds of these sensors in a home will result in weekly or daily requirements to change batteries.

In the rest of this chapter, I develop the idea of using the domestic power line as a conduit for in-home wireless transmission of data. In Section 5.2, I explore how the power line can be used as a conduit for HF, VHF and UHF radio transmissions and examine how frequencies in the unlicensed spectrum can be exploited for direct coupling communication via the power line. In Section 5.3, I establish why 27.12 MHz is the best alternative for exploiting wireless coupling from a transmitter to the power line. I provide empirical evidence from a home testbed that helps to characterize the range and limitations of this approach. In Section 5.4, I provide evidence for how this power line-assisted wireless transmission can act as a platform for an IFRTLS backchannel or in-home wireless sensing, providing a comparison of power and data transmission efficiency compared to commercial alternatives and exploring application opportunities and security concerns.

5.2 The Power Line as a High-Frequency Signal Conduit

In this section, I explore the home power line as a transmission line (no wireless communications, both transmitter and receiver directly connected) for signals at frequencies significantly higher than the power signal they were designed for. In order to act as a conduit for very low power wireless sensors, the power line must be reasonably efficient at conducting signals of the frequency used by the sensor network. This section presents experimental data to demonstrate the feasibility of transmitting HF (3–30 MHz), VHF (30–300 MHz), and UHF signals (300 MHz–3 GHz) over the power line. I begin with a brief introduction to existing power line communications technologies, and then explore the feasibility of using specific frequencies within these higher frequency ranges (HF, VHF, and UHF). This largely empirical process will lead to a small set of frequencies to explore within the radio spectrum for unlicensed communication.

5.2.1 Background

5.2.1.1 Power Line Communications

The home power line, despite being designed to transmit low-frequency electrical power (typically at 50–60 Hz depending on the country), has been used successfully for communication at a variety of higher frequencies for various purposes. Home automation with X10, and more recently Insteon, has been a popular use of power line communication. These standards utilize carrier frequencies of 120 kHz for X10 and 131.65 kHz for Insteon [8, 17]. Commercial home networking solutions, such as the Netgear XET1001 Ethernet-over-power line system, show the practicality of the power line as a high-speed data transmission medium. Many of these solutions, including the XET1001, operate in accordance with the HomePlug standard, providing data rates up to 200 Mbps over the power line and operating in the standard's defined 2–28 MHz range [7].

A benefit of using higher frequency signals on the power line is their propagation range. In the US and Europe, homes are typically configured with a multi-phase power system which separates the home power line into several isolated branches. This creates problems for X10 signals, since a controller may be plugged into one phase while a device it intends to control may be on another. The solution to this is to install a phase coupler either at the circuit breaker box or at a high voltage outlet (such as for a clothes dryer) where the phases come together. This allows the low frequency 120 kHz signals to propagate across the phases. Higher frequency signals ($>$ 1 MHz) do not have this problem, however, as they are able to wirelessly couple between the phases at points where wiring from both phases runs close to one another (such as at the circuit breaker box). This is why Ethernet-over-power line devices are able to cover an entire house without the use of a phase coupler.

5.2.1.2 Unlicensed Wireless Communications

I now motivate the frequencies at which I chose to explore the power line as a signal conduit with a discussion on regulations regarding unlicensed wireless communications. The majority of the radio spectrum is reserved for licensed use. This ensures that wireless services such as mobile telephones, television and radio broadcasting, and radio-navigation services are generally free from interference. The International Telecommunication Union (ITU), a United Nations agency responsible for coordination of the radio spectrum on a global level, has specified several areas of radio spectrum that local governments should make available for unlicensed devices. These spectrum areas are commonly referred to as the Industrial, Scientific, and Medical (ISM) bands. Common ISM devices include WiFi (IEEE 802.11a/b/g) devices, Bluetooth devices such as wireless headsets, and wireless keyboards and mice with proprietary radio protocols. A list of the ITU suggested ISM spectrum is shown in Table 14.

Table 14: ITU specified Industrial, scientific, and medical (ISM) spectrum up to 2.4 GHz[9, 20].

Center Frequency	Range (MHz)	FCC Regulated Max. Field Strength	FCC Max. EIRP (dBm)
6.78 MHz [a]	±0.015	15 $\mu V/m$ @ 30 m	-51.71 dBm
13.56 MHz	±0.007	15,848 $\mu V/m$ @ 30 m	+8.77 dBm
27.12 MHz	±0.163	10,000 $\mu V/m$ @ 3 m	-15.23 dBm
40.68 MHz	±0.020	1,000 $\mu V/m$ @ 3 m	-35.23 dBm
433.92 MHz [b]	±0.870	—	—
915.00 MHz [c]	±13.000	—	+36.00 dBm [d]
2.45 GHz	±50.000	—	+36.00 dBm [d]

[a]Subject to approval by the local regulatory body in the country of interest.
[b]ITU Region 1 only (Europe, Africa, the Middle East west of the Persian Gulf, the former Soviet Union, and Mongolia).
[c]ITU Region 2 only (The Americas, Greenland, and some eastern Pacific Islands).
[d]For digital modulation schemes having bandwidth of at least 500 kHz.

Although in some countries, such as the United States, use of spectrum outside of the ISM regions is permitted (for example, under Part 15 of the US Federal Communications Commission regulations), I chose to focus my exploration on the ISM spectrum to make the work more globally applicable. Given this, I also chose to avoid utilizing the 433.92 MHz and 915.00 MHz bands since they are not globally available. Additionally, a large number of consumer electronic devices, such as cordless telephones and garage door openers, already operate in these bands and might cause interference for low-powered wireless sensors. Given this, I chose to explore five frequencies on the power line: 6.78 MHz, 13.56 MHz, 27.12 MHz, 40.68 MHz, and 2.45 GHz.

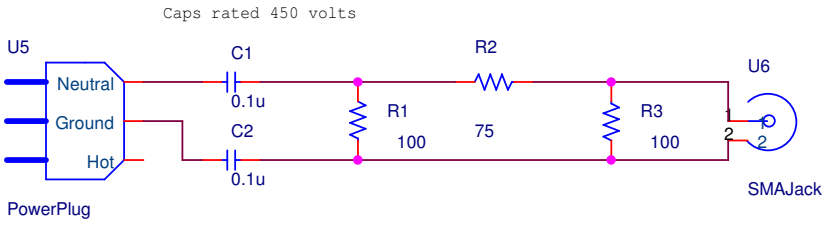

Figure 32: The custom-built power line coupling box. This box is used to connect a high-frequency receiver to the power line while isolating it from the 50-60 Hz power signal.

5.2.2 Testing the Power Line at ISM Frequencies

To test the power line as a transmission line at the selected frequencies, I built a high-frequency coupling box that allowed me to connect a spectrum analyzer to the power line. This coupling box isolates the test equipment from the low-frequency, high-voltage power signal, but allowed high frequency signals to pass through. A circuit diagram of this box is shown in Fig. 32. I utilized two of these coupling boxes: one to inject a high-frequency signal into the power line using an Agilent E4433B signal generator, and one to measure the received signal with a Rhode & Schwarz FSH8 spectrum analyzer. Note that this box was configured to couple to the power line using the neutral and ground wires, which differs from the box used for PLP and WPLP in Chapters 3 and 4, which utilized the hot and neutral wires. The best reception of wireless devices within the home was found to occur on neutral and ground.

Our experiments, the results of which are presented in Table 15, were performed in a 3-story, 371 square meter home built in 2003. A floorplan of the test environment is shown in Fig. 33. This is the same home referred to as `house-1` in Chapter 4. Data was obtained by using the signal generator to inject a 0 dBm (1 mW) signal into the power line at each of the five frequencies. The signal generator was connected to the

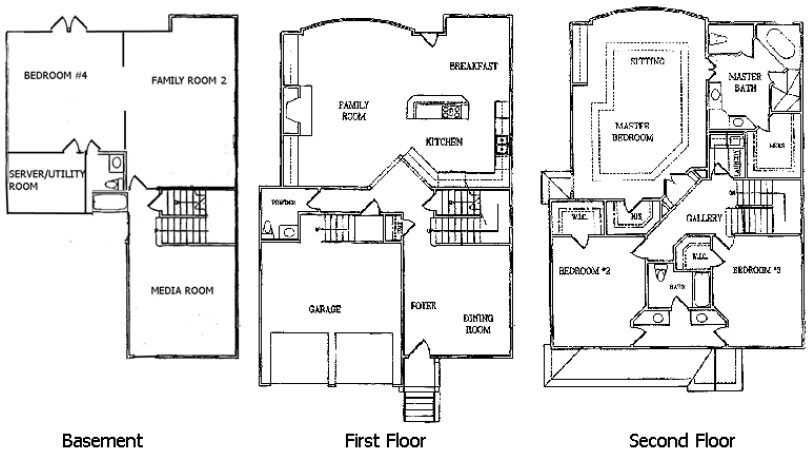

Figure 33: Floorplan of the home that was utilized as one of the test environments. The home was constructed in 2003 and is 371 square meters.

power line via the coupling box plugged into an outlet in the dining room. During the tests the signal generator was powered by a UPS to isolate any signal leakage from its power supply. The received signal strength was then measured using the spectrum analyzer by connecting it to a single, representative outlet in each of nine listed rooms on the three floors. Average attenuation for the ISM frequencies from 6.78 MHz through 40.68 MHz ranged from 47–53 dB. At 2.45 GHz, no signal could be detected over the power line in any of the rooms. Given this, I limited further exploration to 6.78 MHz, 13.56 MHz, 27.12 MHz, and 40.68 MHz.

Since, as noted in Section 5.2.1.1, Ethernet-over-power line devices utilize spectrum from 2–28 MHz, I tested interference caused by these devices at the three ISM bands that fall within this range. To do so, I utilized two Netgear XET1001 HomePlug-based Ethernet-over-power line devices, a laptop to generate traffic over the power line-based network, and the spectrum analyzer to detect interference caused by the XET1001. I noted that the device generated significant interference at 6.78 MHz and 13.56 MHz when transmitting data, but that the 27.12 MHz band was clear.

Table 15: Received signal strength of a signal directly injected into the power line, as sensed in each of nine rooms of the test home. The signal was injected in the dining room (see Fig. 33) at a power of 0 dBm, thus the numbers in this table represent the amount of attenuation caused by the power line. No signal was detected over the power line in any room for 2.45 GHz. The measured loss induced by the two power line coupling boxes at each frequency was subtracted from these figures to give only the loss caused by the power line.

	Signal Strength (dBm)								
	Basement			1st Floor			2nd Floor		
Frequency (MHz)	Media	Family #2	Bed #4	Dining	Kitchen	Family #1	Master Bed	Bed #2	Bed #3
6.780	-41.2	-57.5	-44.9	-19.2	-51.6	-50.8	-69.0	-72.0	-45.5
13.56	-56.8	-60.8	-61.1	-33.3	-53.1	-55.9	-57.1	-54.3	-50.2
27.12	-43.3	-53.5	-61.6	-21.7	-55.9	-61.6	-58.1	-50.4	-40.7
40.68	-38.3	-46.2	-66.7	-35.5	-44.6	-55.2	-53.9	-51.6	-31.1
2450	—	—	—	—	—	—	—	—	—

5.2.3 Summary

I now summarize my exploration of the ISM bands initially considered. 6.78 MHz and 13.56 MHz were eliminated due to interference from HomePlug-based Ethernet-over-power line networks. 433.92 MHz and 915.00 MHz were eliminated since neither band is globally available and each is crowded with consumer devices that might cause interference. 2.45 GHz was eliminated since signals of this frequency did not transmit over the power line. This leaves two ISM bands to consider: 27.12 MHz and 40.68 MHz. I now explore the performance considerations of wireless transmission coupling at those frequencies.

5.3 Exploring The Power line as a Wireless Receiving Antenna

My exploration of the power line as an antenna for receiving wireless signals is motivated by the PLP and Wideband PowerLine Positioning WPLP work described in the previous chapters. PLP and WPLP utilize the power line to wirelessly transmit

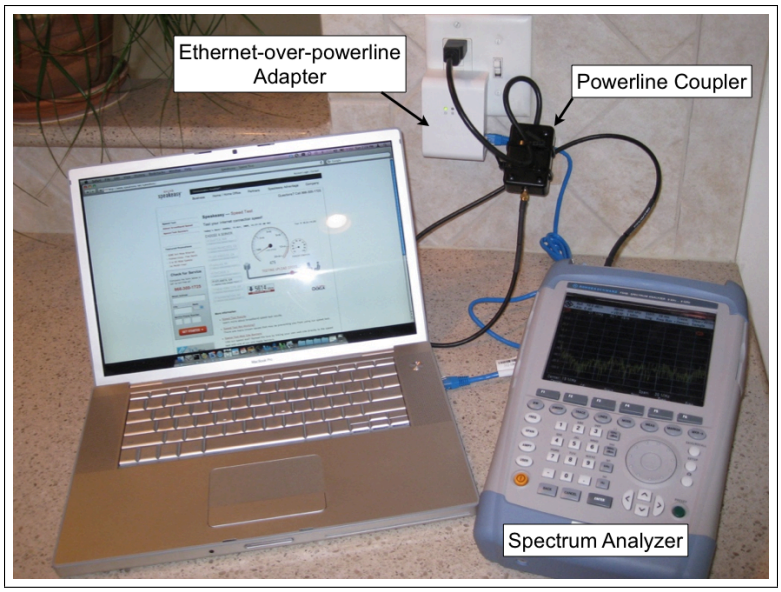

Figure 34: Experimental setup used to check for interference that HomePlug-based Ethernet-over-power line devices may cause in the 6.78 MHz, 13.56 MHz, and 27.12 MHz ISM bands. The laptop was used to generate traffic on the Ethernet-over-power line network, and the spectrum analyzer, connected to the power line through the coupling box shown in Fig. 32 was used to monitor the signals generated by the HomePlug adapter.

signals, ranging in frequency from 400 kHz–20 MHz and produced by one or more signal injector modules plugged into an outlet throughout a home. The principle of reciprocity in antenna design, that is, any given antenna is equally good at receiving and transmitting [66], suggests that the power line would be a good receiver for at least the 400kHz–20MHz range that was explored in the earlier work. Additionally, inexpensive AM and FM clock radios often utilize a line cord antenna [35, 83], which is comprised of a transformer and several capacitors to couple the radio's input to the power line (similar to my coupling box in Fig. 32). I verified the capability of my test home's power line to receive signals in the US AM radio broadcast band by connecting the spectrum analyzer to the power line and monitoring the 520–1610 kHz band (see Fig. 35). Line cord antennas have also been utilized for over-the-air VHF and UHF television reception [84].

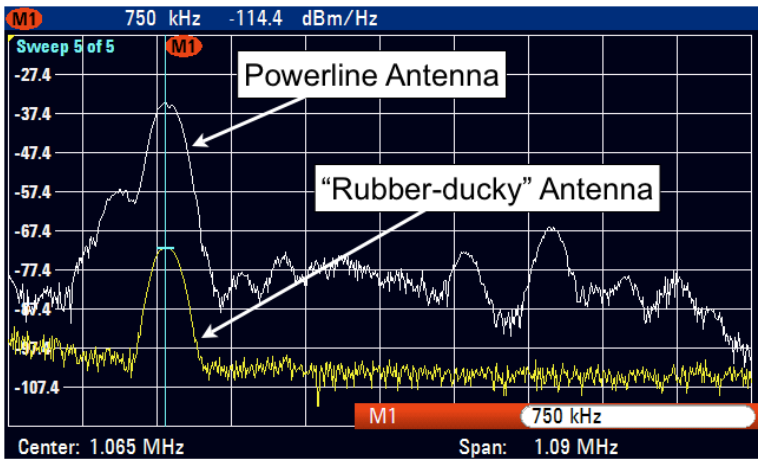

Figure 35: The AM radio broadcast spectrum (520–1610 kHz in the US) as sensed by the spectrum analyzer when connected to the power line through the coupling box of Fig. 32 and a standard "rubber ducky" antenna typical of hand-held radio scanners. Note that AM radio stations are received with much greater signal strength over the power line antenna than with the rubber ducky antenna. A marker is placed at 750 kHz, which is a local AM radio station.

5.3.1 Selecting a Frequency for In-Home Power Line-based Wireless Communication

I begin my exploration of the power line as a receiving antenna for low-power wireless signals originating within the home by making a final selection of a frequency to explore, between 27.12 MHz and 40.68 MHz, as discussed in Section 5.2. Since the objective is to use the power line for receiving very low-power wireless transmissions, we need to select a frequency band that is relatively quiet on the power line. Figure 36 illustrates the power line background noise from 100 kHz–42 MHz and specifically points out the two frequencies still in consideration (27.12 MHz and 40.68 MHz). Both of these fall in relatively quiet regions of the power line, and so background noise levels did not influence my selection of one of these two.

Another consideration in selecting a frequency is the size of an efficient antenna at that frequency. For example, although the power line has proven to be a reasonably reliable conduit for signals in the 100 kHz range (X10 and Insteon), frequencies this low are not practical for wireless communication. Although both X10 and Insteon

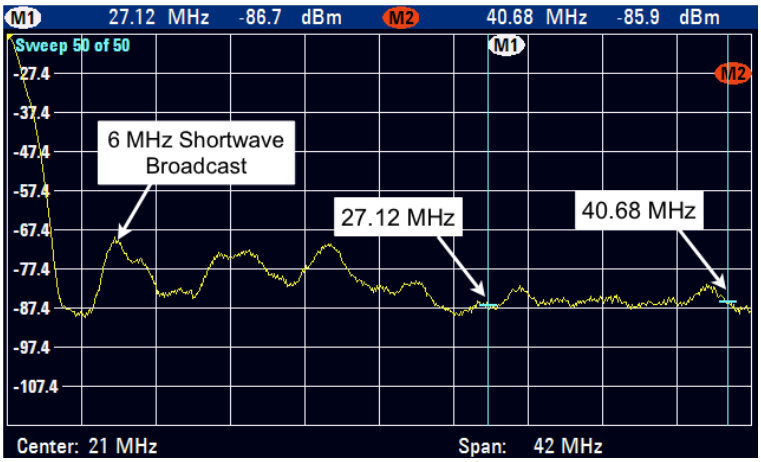

Figure 36: Noise floor of the power line from 100 kHz–42 MHz using the coupling box shown in Fig. 32 and the R&S FSH-8 spectrum analyzer. The sample was taken in the kitchen of the test home, as shown in Fig. 33. The peaks generally represent various radio broadcasts from outside the home. The second from left, for example, is a 6 MHz shortwave-radio AM broadcast. Note that the noise floor is artificially raised by several dB when sampling such a large chunk of spectrum at once.

offer wireless remote controls for operation of any X10 or Insteon-enabled device, these wireless remotes operate at higher frequencies (310 MHz in the US, under FCC Part 15 regulations) and do not directly couple with the power line — a bridge between the high-frequency wireless channel and the low-frequency power line channel is needed. There is a simple explanation for this. The size of an efficient antenna at any given frequency is proportional to the wavelength of the frequency. A common antenna design is the $\frac{1}{2}\lambda$ or $\frac{1}{4}\lambda$ dipole, where λ represents the wavelength of the frequency the antenna is designed to operate at. At 120 kHz, a $\frac{1}{4}\lambda$ dipole antenna would be 625 m long. In contrast, the higher frequencies utilized by HomePlug devices (2–28 MHz) are much more applicable to wireless communication. As a comparison, a $\frac{1}{4}\lambda$ dipole antenna at 28 MHz is 2.7 m, vs. 625 m at 120 kHz. The shorter wavelength of these higher frequencies is important in that the transmitting antenna for wireless sensors can be smaller, and also in that a typical home will contain numerous segments of power line in the walls on the order of several meters, but certainly no segments of hundreds of meters.

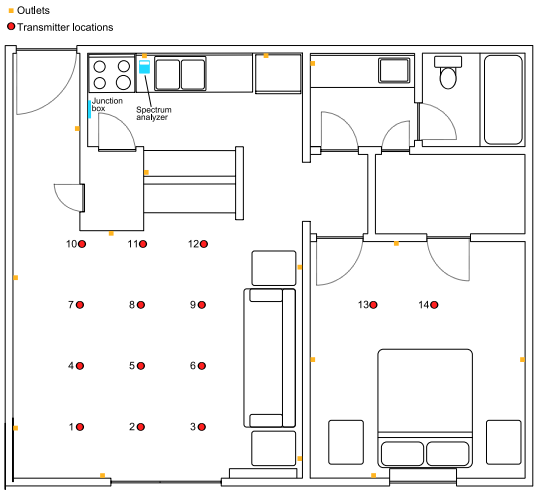

Figure 37: Floorplan of the apartment used for tests. The building was constructed in approximately 1969 and the apartment is approximately 37 square meters. Numbered dots indicate test points for the heat-map shown later.

Table 16: ITU specified Industrial, scientific, and medical (ISM) spectrum up to 2.4 GHz[9, 20].

Center Frequency	Range (MHz)	Potential for Power Line-based Low-power Wireless Communication
6.78 MHz	±0.015	Rejected: HomePlug devices interfere.
13.56 MHz	±0.007	Rejected: HomePlug devices interfere.
27.12 MHz	±0.163	Selected.
40.68 MHz	±0.020	Rejected: FCC regulated maximum output power too low (-35.23 dBm).
433.92 MHz	±0.870	Rejected: Not globally available.
915.00 MHz	±13.000	Rejected: Not globally available.
2.45 GHz	±50.000	Rejected: Experiments show power line does not carry signals of this frequency.

I ultimately selected 27.12 MHz as the optimal frequency for power line-based low-power wireless signal reception due to regulatory constraints. Revisiting Table 14, the third and fourth columns now become important. Within each frequency band, the FCC has specified a maximum power at which devices may operate [20]. This is typically specified as a maximum field strength in $\mu V/m$ at some distance from the transmitter. Equation 1 expresses the electric field as a function of the transmitter power (P_T), the antenna gain (T_G), and the distance from the transmitting antenna (r) [18].

$$E = \frac{\sqrt{30 P_T G_T}}{r} \quad Volts/meter \tag{1}$$

$$P_T = \frac{E^2 r^2}{30} \tag{2}$$

Power in radio systems is often quoted as EIRP (Effective Isotropic Radiated Power), which refers to the power output of the system from a hypothetical isotropic antenna (a point-source which radiates equally in all directions). I can calculate EIRP from Equation 1 by setting G_T to 1 and rearranging to solve for P_T, giving Equation 2. These equations are how the fourth column of Table 14 was calculated. This column provides us with a good estimate of the maximum output power a wireless transmitter can have at each of these frequencies and still be within regulations. 40.68 MHz was ultimately eliminated from consideration based on this band's very low output power constraint of -35.23 dBm.

Table 16 now reflects my reasoning that leads to a selection of 27.12 MHz as the frequency for a very low-power wireless sensor network.

5.3.2 Receiving 27.12 MHz Signals Within the Home Using the Power Line

I first wanted to establish that strictly over-the-air communication was not possible at 27.12 MHz throughout my test home. I configured a transmitter, which consisted

of a signal generator, producing a 27.12 MHz carrier signal at -35 dBm (0.32 μW) and AM modulated with a 550 Hz tone, connected to a 27.12 MHz Citizens' Band antenna that was placed close to an outlet in the dining room. The receiver consisted of a hand-held radio scanner (a RadioShack PRO-97) tuned to 27.12 MHz and used to listen to the 550 Hz tone. I did a survey of the test home using the "rubber-ducky" antenna that came with the scanner.

Where the 550 Hz tone can be heard clearly, the 27.12 MHz carrier signal is considered to be strong enough; where the tone cannot be heard indicates that the signal is too weak to be detected there. I quickly observed that outside of the dining room/foyer area where the transmitting antenna was placed, the signal was generally too weak to be detected by the scanner. I noticed, however, that when bringing the scanner's antenna close to an outlet, wall switch, or other plugged-in electrical device, the signal could be heard faintly. This indicates that the 27.12 MHz signal coupled from the transmitting antenna onto the power line and was being radiated, albeit weakly, by the power line at various other points in the home.

In practice, the receiver will be connected directly to the power line, so the next step was to connect the scanner to the power line via the coupling box of Fig. 32. I did this at a variety of outlet locations throughout the home and noted that the signal could be easily detected in every room of the home by directly connecting to the power line. Not every outlet resulted in good reception, but the vast majority of outlets were satisfactory.

Given this empirical evidence suggesting that the 27.12 MHz signal coupled to the power line over the air, I wanted to see how far away the transmitting antenna could be from the power line and still be sensed by a power line connected receiver at most outlets within the home. To do this, I moved the antenna away from an outlet in increments of 2.5 cm and used the power line coupled scanner to check for the signal at various outlets throughout the home. I found that at a transmitter

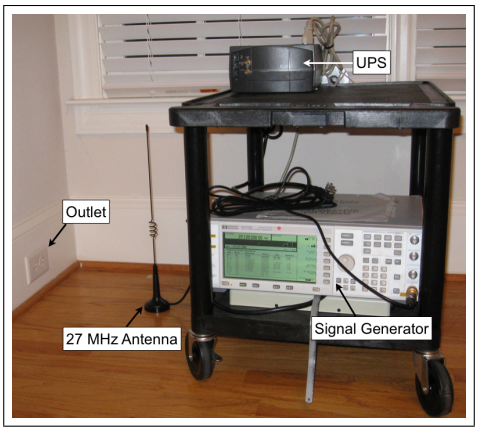

Figure 38: Setup for initial testing of the power line as a receiving antenna. The signal generator was electrically isolated by being placed on a plastic cart and powered by a UPS. A 27.12 MHz antenna (a Cobra HG A1000 designed for Citizens' Band radio use) was used to transmit a 27.12 MHz signal AM modulated at 550 Hz and at a power of -35 dBm (0.32 μW).

output power of -35 dBm, the antenna could be up to 60 cm away from the outlet before the signal was too weak to be detected by the receiving scanner. Generally, the further the antenna was away from the power line, the weaker the signal detected by the scanner. Coupling is not limited to outlets, however. I noted that a similar effect could be observed by bringing the transmitting antenna close to other electrical wiring, such as a wall switch.

As a final test, I observed that some of the outlets in the test home served as particularly good coupling locations for the receiver. One of these, located in the basement, is near to the circuit breaker for the entire house, and so is in close proximity to every circuit in the home as result. Connecting the receiver to this outlet allowed me to test the range of the transmitting antenna in a best-case scenario. Fig. 39 shows a heat-map of the received signal strength along the first floor of the test home, indicating that the antenna's signal was strong enough throughout almost the entire first floor. I also repeated this test in a small apartment, as shown in Fig. 40. While there are some important limitations to this test (I only took measurements from a fixed height of approximately one foot off the ground), it does reveal great promise

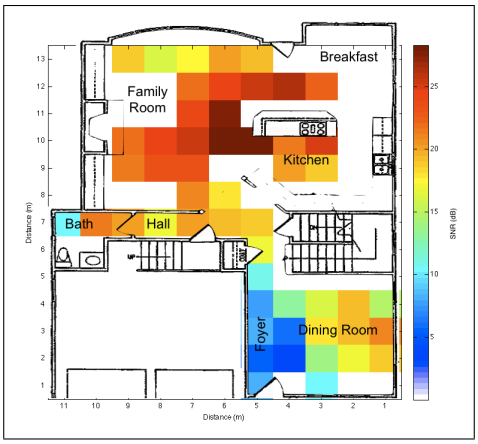

Figure 39: A heat-map of the test home first floor indicating the Signal-to-Noise Ratio (SNR) for a 27.12 MHz wireless transmitter at each indicated location as received by a fixed power line coupled receiver in the basement. The transmitter utilized an output power of -25 dBm (3.2 μW) and a custom-built 27.12 MHz loop antenna.

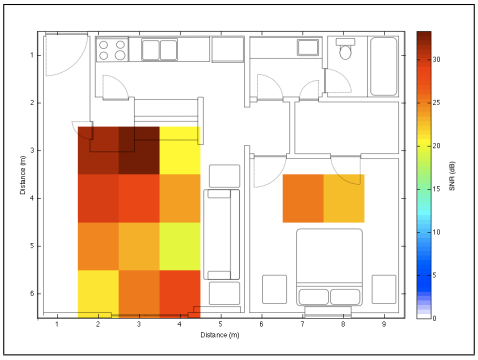

Figure 40: A heat-map of the test apartment indicating the Signal-to-Noise Ratio (SNR) for a 27.12 MHz wireless transmitter at each indicated location as received by a fixed power line coupled receiver in the kitchen. The transmitter utilized an output power of -35 dBm (0.32 μW) and a custom-built 27.12 MHz loop antenna.

for using this solution to support a whole-house wireless sensing solution. I explore the larger systems issues of whole-house, low-power sensing in the next section.

5.4 Whole-House, Low-Power Sensing

5.4.1 A Platform For Power Line-Based Sensing

Given my success in detecting an AM modulated 27.12 MHz carrier signal from a wireless transmitter over the power line, I wanted to build a real platform for sensing around this phenomenon. I used a Texas Instruments MSP430 microcontroller [12]

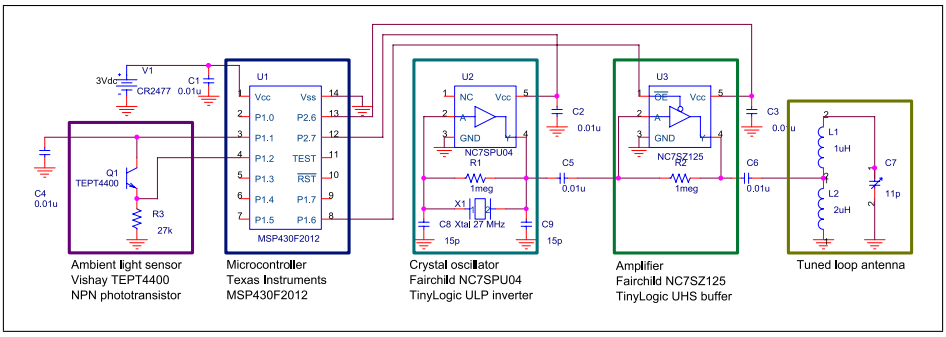

Figure 41: A circuit diagram of my custom-built power line-based wireless sensor.

with an attached light sensor and a custom-built 27.12 MHz radio. The radio consisted of a 27.12 MHz crystal with a low-power amplifier. The crystal and amplifier could be turned on and off by adjusting the voltage on a control pin, which was driven by the MSP430's serial bus output.

I chose On-Off-Keying (OOK) as the modulation scheme for transmitting data from the sensor. On-Off-Keying — wherein the transmitter is on to transmit a one and off to transmit a zero — is a very efficient modulation scheme for low-power devices as the transmitter does not expend any energy to transmit a zero. The sensor transmitted a packet of 16 bits once per second at a bit rate of 62.5 kilobits per second. The 16 bits consisted of five bits for the sensor's ID, followed by ten bits for the value of the light sensor, followed by one stop bit — if more than 32 sensors are needed per home, or if more granularity is desired in the sensor reading, additional bits can easily be added to the packet as appropriate.

Since the objective of the sensor node is to transmit readings from its sensor, I elected not to include the ability for the node to receive data. This keeps the hardware cost and complexity down, as well as reduces power consumption of the node. The sensor, radio, and antennas are shown in Fig. 42, and a circuit diagram of the sensor and radio is shown in Fig. 41.

The custom 27.12 MHz radio operates on just 1.5 mA at 1.2 V (this is in addition to the 165μA used by the MSP430 microcontroller). I programmed the microcontroller

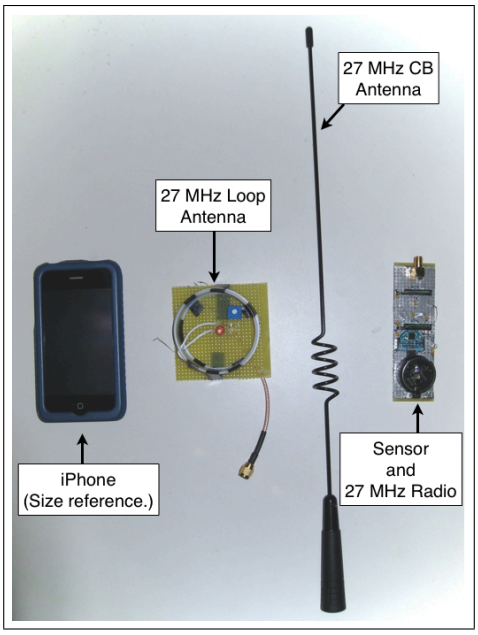

Figure 42: Our wireless sensor platform along with two 27.12 MHz antennas.

to transmit a sensor reading once per second and to shut off the oscillator during the interim sleep period, making the sensor's duty cycle 0.941 milliseconds. With the amplifier and oscillator I used, the transmit power was actually more than sufficient to be sensed everywhere in the house, and so the power draw could actually be made less by utilizing more efficient oscillators and amplifiers.

I tested two types of antennas with the sensor, both pictured in Fig. 42. One was a CB antenna measuring 43 cm tall and the other a custom-built loop antenna measuring 6.5 cm long x 6.5 cm wide x 2 cm tall. The transmit power necessary for sensordata reception is a function of the efficiency of the antenna. Clearly the loop antenna has a size advantage in terms of enabling a compact form factor for in-home sensors, however both of these antennas are relatively small compared to the 11.06 m wavelength at 27.12 MHz, meaning that both are reasonably inefficient at this frequency. The efficiency of the power line as a receiving antenna at 27.12 MHz largely makes up for this, but since I desire to operate the sensors at as low a power output as possible so as to increase their lifespan, a more efficient antenna will allow a

lower transmit power. Thus, longevity of the sensor is a direct tradeoff with antenna size. I found that a sensor built around the larger 43 cm CB antenna was able to transmit at a power level of about 2 dB less on average in order to be detected with the same SNR as with the loop antenna.

5.4.2 Range & Power Efficiency–PL Sensing vs. Existing Technologies

Power line-based wireless sensors should be evaluated against existing wireless sensing technologies on two fronts — range and power efficiency. Communication range of wireless sensors is an important consideration since sensors must be able to communicate with a basestation to transmit readings. The primary model of enabling low power wireless sensor networks to date has been multi-hop mesh networking [22], wherein if a sensor cannot directly communicate with a basestation, it forwards its readings through other sensors within range until the packet reaches the basestation. Although this potentially allows sensors to operate at a lower transmit power than if they had to directly reach the basestation, it also requires other sensors to receive data and then re-transmit it. Interestingly, the Texas Instruments CC2420, a popular 2.4 GHz ZigBee-compliant RF transceiver used in the Sun Microsystems SunSPOTs and Crossbow MicaZ wireless sensors, uses more power in receive mode (19.7 mA) than in transmit mode (17.4 mA @ 0 dBm) [2, 11].

To make a direct comparison between my power line-based sensors and existing sensor network technology, I tested the communication range of two SunSPOT wireless sensors within the test home [10]. One SunSPOT continually transmitted packets at a power of -25 dBm, while the other continually listened for packets and flashed an LED when it received them. -25 dBm was chosen since a power line connected receiver was able to detect a 27.12 MHz transmitter at that power level throughout the house (in fact, -35 dBm was sufficient power to sense a transmitter within 60 cm of a power line when using a power line connected receiver anywhere in

the home). I left the transmitting SunSPOT in a fixed location (to simulate a fixed wireless sensor) and walked around the test home with the receiving SunSPOT (to test possible locations for the sensor basestation). I found that at -25 dBm, while the power line-based sensors were able to be detected at any room in the home by connecting a receiver to the outlet, the SunSPOTs could generally only communicate within the same room. I also tested whether the SunSPOTs' range could be extended by coupling to the power line (despite Sec. 5.2 showing that 2.4 GHz signals do not transmit over the power line) by placing both SunSPOTs near an outlet. This actually *decreased* their range. To give the SunSPOTs' whole-house coverage, I needed to increase their transmit power to 0 dBm (1 mW).

Besides utilizing the power line to extend the range of my low-power sensors, an additional source of power efficiency is the low complexity of my data transmission protocol. As mentioned earlier, my power line-based sensors are designed only to transmit sensor readings periodically and do not need to receive data or forward data for other sensors. This allows my data transmission protocol to be extremely simple. Instead of the 256 bits of overhead in every ZigBee packet [30], my protocol needs only six bits of overhead (five bits for the sensor address and one stop bit). Thus, although my sensor has a lower data transmission rate than Zigbee (62.5 kbps vs. 250 kbps), the time the transmitter must be active to transmit a 10 bit sensor reading is actually less for my power line-based sensor, thus requiring less energy.

A direct comparison metric for the two technologies is milliamp-hours per bit transmitted (mAh / bit). This can be calculated as:

$$\frac{NumberOfBits}{bits/sec} \cdot RadioPower \cdot \frac{1hr}{3600sec} \tag{3}$$

I provide this data for my power line-based 27.12 MHz wireless sensors and for 2.4 GHz Zigbee sensors in Table 17. This table shows two configurations for Zigbee: no mesh and mesh. No mesh is a more direct comparison with my technology since

Table 17: 27.12 MHz power line-based wireless sensing compared with 2.4 GHz Zigbee. Power requirements for the CC2420 Zigbee radio can be found in [2].

	Power Line	Zigbee (no mesh)	Zigbee (mesh)
Total bits for a 10-bit Sensor Reading	16	272	544
Bit Rate (kbps)	62.5	250	250
Transmit Current (mA)	1.5	17.4	17.4
Receive Current (mA)	—	—	19.7
mAh / bit	$1.07 \cdot 10^{-7}$	$52.6 \cdot 10^{-7}$	$164.7 \cdot 10^{-7}$
%age of Power line-sensing	100%	4,930%	15,442%

here I only account for the power that a Zigbee-based sensor must use to transmit a 10-bit sensor reading with similar coverage to power line-based sensing. The mesh Zigbee configuration, however, considers that Zigbee may need to forward packets in a mesh network configuration from sensors not within direct communication range of the basestation. These values account for a sensor that must receive a packet from one of these out-of-range sensors and then forward it on to the basestation, in addition to sending one of its own packets with a sensor reading for each forwarded packet. These calculations show that my power line-based wireless sensors use just 2% – 7% of the power that a Zigbee-based wireless sensor requires for similar coverage. In fact, the actual results are potentially better since this calculation does not account for the fact that on-off-keying uses no power when transmitting a zero and represents the power necessary to transmit 16 ones.

5.4.3 Multiple-Access Protocol

At the current data rate of 62.5 kilobits per second, each sensor needs 0.03% of a second to transmit its data. When each sensor transmits a 16-bit packet once per second, this leads to a theoretical maximum of 3,906 sensors per home. Our system currently utilizes what can be considered a uni-directional version of the ALOHA protocol for multiple access on the channel [21]. Since sensors are not capable of

receiving, I rely upon the various sensors within a home to transmit at different times so as not to interfere. This can be done reasonably well without synchronization when the number of sensors does not approach the theoretical maximum by having the sensors add a randomized delay each time to the one-second interval between transmitting packets.

5.4.4 Security

Although my current data transmission protocol does not use encryption, the nature of the system has a basic level of security built-in. By utilizing extremely low transmit powers on the sensors, the reception range is greatly limited for anyone attempting to overhear the sensors from outside the home. Although, as with any wireless technology, these devices are susceptible to snooping with high-gain antennas, this is generally expensive and impractical.

The power line also has a natural security mechanism built-in — transformers. Electric power is distributed at a higher voltage than is utilized in the home. Transformers convert this higher voltage to the voltage carried on the home power line. Transformers work well for low frequency signals like the 50-60 Hz power signal, but their high inductance causes them not to pass higher frequency signals like 27.12 MHz. Since one transformer can typically serve just a few homes, they act as a natural barrier to the sensors' signals propagating too far along the power line and being sensed by neighbors from within their homes.

I concede that these sensors are susceptible to snooping via physical access to outlets on the outside of the home, but believe that the security implications of being able to snoop such low-level sensors as this system was designed for are quite limited.

5.5 SNUPI: Sensor Nodes Utilizing Power Line Infrastructure

Subsequent to the development of the platform in Section 5.4.1, Cohn *at al.* utilized the powerline as receiver phenomenon to develop the Sensor Nodes Utilizing Power Line Infrastructure (SNUPI) platform [37]. This work, which further validates the research that is the basis of this chapter, not only demonstrates utilization of the domestic power line for reception of wireless data transmissions at 27 MHz, but also achieves whole-home range with the lowest power radio to date. The SNUPI radios demonstrated in this work operate on just 50 μW, as compared with the 1800 μW used by the radio in Sec. 5.4.1.

5.6 Summary of Contributions

The objective of this component of the research was to utilize the home power line as a backchannel for PLP and WPLP tags to communicate with the central location processing unit. In the process, this work has also effectively take power consumption for wireless data transmission in sensor networks out of the lifespan equation for a sensor by making it negligible compared to power consumption for sensors and microcontrollers. To do so, this chapter has:

- Shown that the home power line is capable of carrying much higher frequency signals than the 60 Hz power signal it was designed for.
- Demonstrated a new model for building in-home wireless sensor networks that leverages the naturally occurring phenomenon of the existing home power line infrastructure acting as an antenna for 27.12 MHz devices. This allows wireless sensors to transmit at a much lower output power than would otherwise be necessary to cover a typical house.
- Demonstrated a custom-designed sensor platform built around this phenomenon

that used a light sensor as an example. The 27.12 MHz radio in this device was able to communicate with a power line connected receiver anywhere in the home while using just 6.3 μA on average, with a peak current draw of 1.5 mA.

I believe that leveraging this phenomenon will finally enable us to overcome the power constraints that have always been the roadblock in realizing the ubicomp community's vision of ubiquitous sensing.

CHAPTER 6

CONCLUSION AND FUTURE DIRECTIONS

I begin this chapter in Sec. 6.1 by revisiting each of the three claims of the thesis statement introduced in Sec. 1.1.1. I conclude the dissertation in Sec. 6.2 by discussing opportunities for future work in this space.

6.1 Revisiting the Thesis Statement

In Sec. 1.1.1, I introduced the following thesis statement:

> *A domestic fingerprinting-based indoor RTLS using the amplitude of power line radiated low-frequency RF signals as its fingerprint will: (1) benefit in accuracy, reduced sensitivity to short-term noise, and increased long-term temporal stability through the use of a wideband fingerprint as compared to a narrow-band frequency pair approach; (2) exhibit superior temporal stability of its underling fingerprint components as compared to WiFi, cellular, and FM based solutions; and (3) be able to utilize the power line as a receiving antenna as well as a transmitting antenna, thus allowing both forward and back-channel communications to occur over the power line.*

In the following three subsections, I summarize how this dissertation has addressed each of the three claims of that statement.

6.1.1 Wideband Fingerprinting

The first thesis claim was that a domestic fingerprinting-based indoor RTLS using the amplitude of power line radiated low-frequency RF signals as its fingerprint will benefit in *accuracy*, *reduced sensitivity to short-term noise*, and *increased long-term*

temporal stability through the use of a wideband fingerprint as compared to a narrowband frequency pair approach. I supported this claim in Chapter 3, beginning by identifying both initial accuracy and longer-term temporal stability concerns with the PowerLine Positioning IFRTLS. I showed that PLP can suffer from both short-term and long-term temporal instability of the signal map in particularly "noisy" RF environments or environments with a considerable amount of electrical equipment.

Testing of the system was performed in the Georgia Tech Aware Home over a 66-point grid with 0.9 x 0.9 m spacing. Initial experiments sought to reproduce the results of the original two-frequency narrowband fingerprint PLP work in this noisy environment, and showed that the accuracy of the system was significantly lower initially post-deployment than was experienced at the homes tested in the original work. I also showed that this accuracy was subject to significant degradation over time with increasing temporal spacing between the training signal map and live testing data. Additionally, I showed that the two-frequency narrowband fingerprint approach suffers from high sensitivity to noise — in the tested environment, just 1 dB of noise causes the system's accuracy in terms of correctly classified samples to fall below 50% for room-level, sub-room-level, and grid-level accuracy.

After identifying the temporal stability concerns of the two-frequency narrowband approach, I showed that a 44-frequency wideband approach could address these issues. The wideband approach both improved the initial post deployment accuracy of the system — to 100% — as well as reduced the amount of degradation in accuracy over time with increasing separation between the initial training site survey and live testing samples. I also showed that the 44-frequency wideband fingerprint exhibits significantly less sensitivity to noise in the individual fingerprint components — up to 5 dB of noise can be accommodated before the system's accuracy drops below 100%. Finally, I analyzed the types of errors a wideband fingerprint-based system makes and showed that the mean distance error of the system is typically below 1 meter.

6.1.2 Temporal Stability

The second thesis claim was that a domestic fingerprinting-based indoor RTLS using the amplitude of power line radiated low-frequency RF signals as its fingerprint will exhibit superior temporal stability of its underling fingerprint components as compared to WiFi, cellular, and FM based solutions. I supported this claim in Chapter 4 by performing an evaluation of the temporal stability of several other IFRTLS's discussed in the literature — this included systems with fingerprints based on WiFi, cellular signals (both GSM and CDMA), and FM radio signals, as well as a variant of WPLP with a slightly reduced fingerprint size. I gathered data for a period of several days at a single point within each of three test locations — two single-family homes and one apartment. At one of the single-family homes, I also gathered two independent datasets consisting of grid-level samples on a 0.9 x 0.9 meter grid.

The underlying motivation of this component of the dissertation was to identify which of the popular RF technologies used as the fingerprint feature of an IFRTLS exhibits the greatest temporal stability. To do this, I presented plots of the raw sample data at each location across the entire test period. I also analyzed stability in terms of the standard deviation of the raw data, as well as of a one-minute time-averaged version of the feature since averaging has been used in the literature as a method to improve the accuracy of an IFRLTS. My results showed that WiFi and GSM exhibit significantly more temporal *instability* as compared to FM signals or the low-frequency power line radiated RF signals used by PLP and WPLP. Although CDMA exhibited reasonable temporal stability, I found that there were an insufficient number of features available at the test locations for it to be used as the basis of an IFRTLS.

Not only was the temporal instability higher for WiFi and GSM than for PLP/WPLP and FM, but these technologies also exhibited much greater distance errors when tested on the grid. While FM, PLP, and WPLP all showed mean errors of less than

0.5 meters, WiFi and GSM both showed average errors of greater than 1.5 meters. Finally, I showed that these systems also exhibited the greatest sensitivity to noise in their fingerprint components.

6.1.3 The Power Line as a Receiving Antenna

The final thesis claim was that a domestic fingerprinting-based indoor RTLS using the amplitude of power line radiated low-frequency RF signals as its fingerprint can utilize the power line as a *receiving* antenna as well as a *transmitting* antenna, thus allowing both forward and back-channel communications to occur over the power line. The motivation behind this claim was to address the practicality of deployment for PLP/WPLP. I discussed that one of the drawbacks of the original PLP system was its need for a separate wireless backchannel to allow the mobile location tags to transmit their observed fingerprint to the central LPU. To address this, I supported this claim in Chapter 5 by showing that the home power line not only acts as a *transmitter* to distribute the RF signals used as the fingerprint features for PLP/WPLP, but also as a *receiver* for responses from the mobile location tags.

I began the power line as an RF receiving antenna experiments by testing the power line's ability to act as a wired signal conduit through the home at various frequencies. Since the goal of the work was a practically deployable system, I focused my exploration on frequencies globally available for unlicensed use. I found that the power line can act as a signal conduit for the four frequencies tested below 41 MHz, however it was not capable of carrying the higher frequency signals. Following this, I examined the power line's ability to act as an RF receiving antenna at the four frequencies still of interest. I showed that signals at the globally available 27.12 MHz band are capable of wirelessly coupling to the power line and being detected by a centrally located receiver in a home that is directly connected to the power line. Finally, I showed the SNUPI platform for power line-based RF communications within

a home, which is capable of transmitting with whole-home range using just 50 μW, representing a significant improvement in power efficiency over existing home RF technologies.

6.2 Limitations and Future Directions

Although this dissertation has made a significant contribution to the field of indoor fingerprinting-based real-time locating systems, it does have limitations, and it suggests several opportunities for future work. The following are open problems:

- **WPLP Tag and Injector Miniaturization** – Research to date has been performed using a large prototype to represent the WPLP mobile location tag. Although this equipment should be easily miniaturized with a custom circuit design, this is left to future work. Additionally, a WPLP injector capable of outputting the entire 44-frequency fingerprint continuously needs to be designed. The current prototype generates the fingerprint components in a sequential manner.

- **Height / Orientation Experiments** – The experiments carried out in this dissertation were based on fingerprints gathered at a constant height off the floor and at a constant orientation. Changing the height and/or orientation of a mobile location tag may change the observed fingerprint. Future work must determine if this is the case, and find ways to mitigate any accuracy degradation related to this.

- **Site Survey Complexity** – Although this dissertation has somewhat addressed the problem of a time-consuming site-survey by reducing the need for repeat site-surveys due to signal map degradation, the initial site-survey is still necessary. Future research should address potential methods for reducing the complexity of a WPLP site survey in a domestic environment. A potential

solution is a more sparse survey than the one carried out in this dissertation at 0.9 m intervals. Although this may reduce the accuracy of the system, the observed accuracy may still be sufficient for a given application.

- **Further Temporal Stability Analysis** – This dissertation studied the temporal stability of WPLP over the course of months, but not years. Future research should examine the stability of this system over longer periods.

- **Exploration of Larger Fingerprints** – Although this dissertation established that a wideband fingerprint improves the accuracy and temporal stability of WPLP, only fingerprint sizes of up to 44 features were tested. Future research should test larger fingerprints to determine if there is additional benefit.

- **Additional Testing Environments** – The experiments in this dissertation were carried out in two homes, an apartment, and a research laboratory designed to look like a home. WPLP needs to be tested further in additional environments, such as commercial spaces.

- **Hybrid IFRLTSs** – The research comparing PLP/WPLP with other RF technologies used for fingerprinting identified FM as another promising fingerprint feature. An interesting future direction would be a hybrid WPLP / FM solution. A hybrid scheme with WiFi is another possibility. Future research should address whether hybrid schemes improve accuracy, temporal stability, and sensitivity to noise.

- **Battery-less Sensing Applications** – The SNUPI platform presents numerous opportunities for future work by enabling a new class of applications for the home through ubiquitous, inexpensive, and potentially battery-less wireless sensors. An example is in-wall sensors deployed when a home is built that can detect the presence of mold. Future research should explore this application

space, as well as address the potential of harvesting energy for these sensors from the power line radiated signals of WPLP.

APPENDIX A

ADDITIONAL FIGURES FOR WIDEBAND PLP

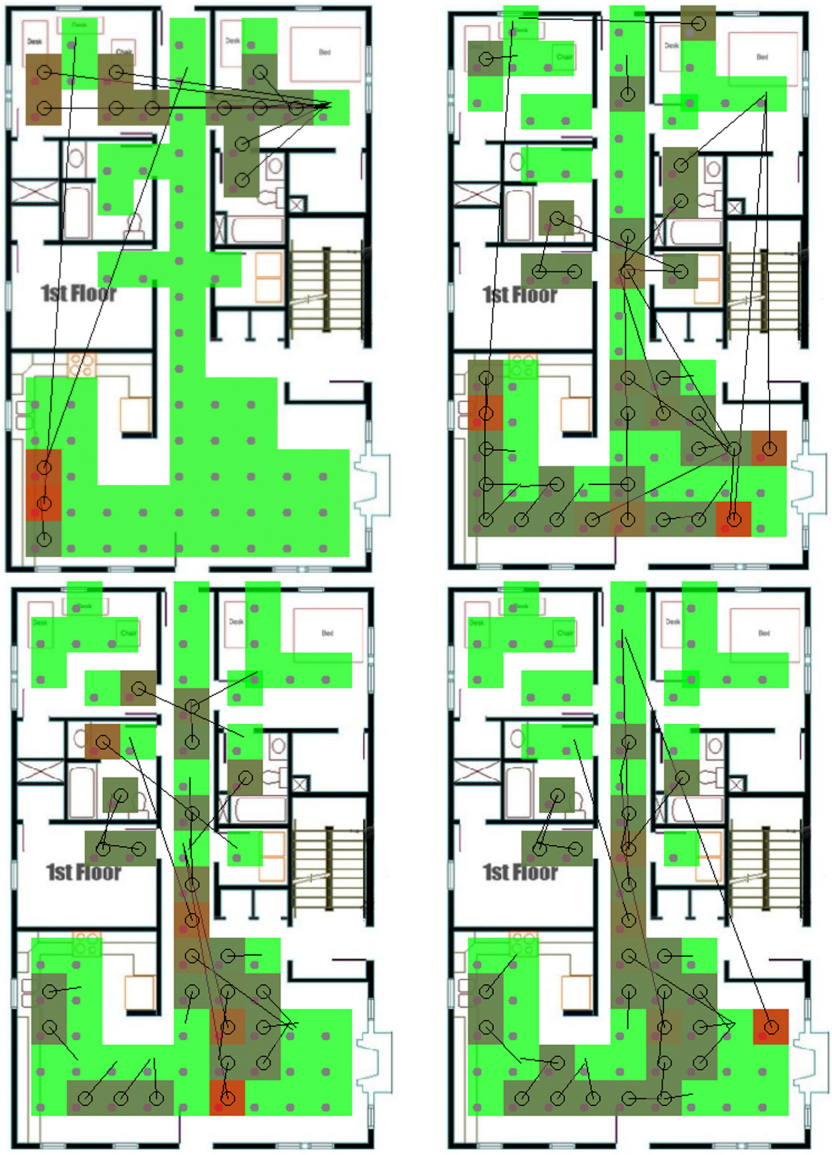

Figure 43: Classification errors resulting from training on one initial dataset and classifying a dataset captured 14 days later (upper left), 42 days later (upper right), 58 days later (bottom left), and 58.5 days later (bottom right).

Figure 44: Classification errors resulting from training on one initial dataset and classifying a dataset captured 78 days later

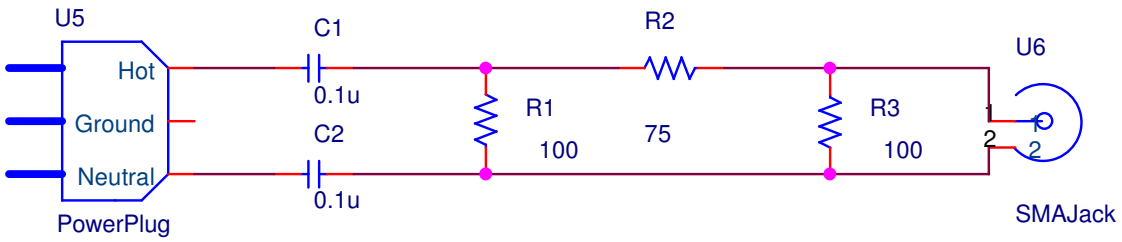

Figure 45: Circuit diagram of the coupling box used to connect the output of the signal generator to the power line.

APPENDIX B

PLP-VIZ DATASET VISUALIZATION TOOL

For the convenience of the reader, this appendix reproduces in its entirety the paper "A Visual Analytics System for Radio Frequency Fingerprinting-based Localization" by Yi Han, Erich P. Stuntebeck, John T. Stasko, and Gregory D. Abowd. This paper, which was published in the proceedings of the 2009 IEEE Symposium on Visual Analytics Science and Technology (VAST 2009) [39], describes the PLP-Viz tool that was built for analysis of the datasets captured during the WPLP experiments.

B.1 Introduction

Tracking the location of people and objects inside of buildings has been an active area of research for some years. The traditional means of accomplishing this outdoors — GPS satellites — is unavailable indoors since buildings block the satellite signals. One approach researchers have taken in solving this problem is generating their own indoor radio frequency (RF) signal(s) as a type of local GPS signal. Small tags, which can be thought of as "indoor GPS receivers" track some aspect of these locally generated RF signals and use this information to locate themselves within the building.

Outdoor GPS receivers operate by triangulating their position based on the time of arrival of signals from multiple GPS satellites. There is typically line-of-sight between the GPS satellites and the receivers, allowing predictable RF signal propagation. Indoors, RF signal propagation is very difficult to predict due to phenomena such as multi-path propagation, wherein the signal can propagate from transmitter to receiver via multiple paths by bouncing off walls and furniture. Small movements

in physical space can produce large differences in the signal since the multiple paths may constructively or destructively interfere at any given position. These phenomena are nearly impossible to predict *a priori*.

To address this problem, researchers have developed the method of radio frequency location fingerprinting. RF fingerprinting relies on measurements of relevant features of the signals at various discretized locations. These measurements are taken when the system is initially deployed. Later, when the system is in use, live measurements taken by the mobile tags are matched to the fingerprinted measurements to calculate the location of the tags. Since the year 2000, there have been over thirty fingerprinting-based localization systems proposed by researchers around the world [52, 73, 80].

An RF fingerprint consists of a set of features of the available RF signals at a particular location. A commonly used feature is the received signal strength of a signal at a particular frequency, illustrated in Figure 46. RF fingerprinting requires that the chosen features vary in space so as to be able to differentiate the locations, but remain constant in time, so that the off-line fingerprint measurement phase does not need to be continually repeated. A location fingerprint is normally built from multiple sets of samples in order to tolerate some degree of noise in the features. Each set of samples collected from all the surveyed locations that constitutes a dataset is called a site survey. Site surveys can be gathered with some time in between to observe how temporally stable the fingerprints are.

One of the most important challenges for RF fingerprinting, therefore, is to select features of the RF signal for the fingerprint that will produce reliable location estimates of the tags. Too few features selected for the fingerprint may not give sufficient information to differentiate the various locations of interest, while too many features may include bad features that are unstable in time, causing the system to produce poor results.

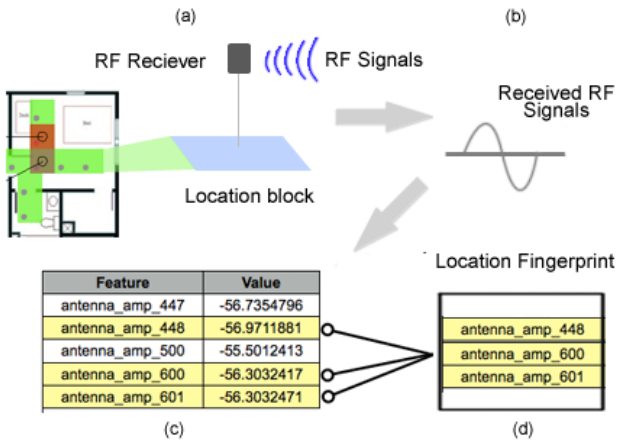

Figure 46: Generation of a location fingerprint. (a) An RF receiver receives the RF signal at a location block. (b) the received signal data are parsed and preprocessed. (c) The sampled signal data are the potential features. (d) The location fingerprint is a subset of these collected features that is unique to this location block.

To aid with the process of RF fingerprinting-based localization system development, we present a visual analytics system for viewing the quality of the fingerprinting data collected during a site survey. By utilizing heat maps to display different perspectives of the features used in the location fingerprints, developers of these systems can not only visually inspect the geospatial feedback for the location classification results, but also be able to select the features to use by visually finding those that are temporally stable and spatially differentiable in a high dimensional feature space. When necessary, developers can even explore lower level details of any individual feature to see raw values and relationships to others through a multivariate visualization. Using our system, developers of localization systems can tell whether their datasets collected are capable of building good RF location fingerprints that can enable accurate location estimates over time.

The contribution of this work is to show how visual analytics can support the development and practical deployment of fingerprinting-based localization systems. We feel that this tool is a particularly good example of visual analytics because the most effective way to find a good location fingerprint is to combine the computational

data analysis with an interactive geospatial visualization interface.

B.2 Related Work

The RADAR system proposed by Bahl and Padmanabhan in 2000 was the earliest RF fingerprinting-based localization system [27]. The researchers were able to achieve a median of 2–3 meters accuracy indoors using WiFi signals. Since then, researchers have reported over thirty systems using different RF signals or classification algorithms [52]. However, although these localization systems are easy to deploy, the initial setup and calibration process for generating the fingerprints is tedious and time consuming [63]. They can also be less reliable when the features used for the fingerprints are not spatially differentiable and stable over time. Kaemarungsi and Padmanabhan studied the properties of WiFi location fingerprints using received signal strength and learned that even the presence of a human body can make a significant difference on the fingerprints [49]. Therefore, it is crucial to identify and remove unstable features in the generated fingerprints to maintain the reliability of the localization system over time.

Visualizing RF signals on a geospatial map using heat maps is prevalent in 802.11 WLAN site survey tools for optimizing WiFi network coverage. Ekahau Site Survey took a step forward to not only visualize the propagation of WiFi signals but also integrate the output to power their Real-Time Location Tracking System [34]. Nevertheless, this consumer-facing site survey tool cannot support more advanced visual debugging functions on feature selection and location fingerprint classification.

Spectrum analyzers for identifying physical locations of signal sources also require visualizing signals on a geospatial map and classifying them. Tektronix's RFHawk Signal Hunter identifies potential malicious RF signals by singling them out from known signals [82]. The malicious signal will then be documented on a geospatial map with a color-coded wave form or signal strength icon for later reference. However, as

the tool did not aim to support location fingerprinting, the wave form icons on the map have little power to show the individual feature differences for building spatially differentiable location fingerprints.

Andrienko and Andrienko used interactive cartographic visualization to output results of the C4.5 classification learning algorithm for knowledge discovery [25]. Their work suggested that interactive visual facilities that allow an analyst to manipulate variables and immediately observe the resulting changes in a map is effective for geospatial data analysis. Our visual analytics system took a step further for the K-nearest neighbor classification algorithm as to even visualize the intermediate steps of the algorithm for indoor localization.

Our system was developed with data from the PowerLine Positioning localization system (PLP) [73]. PLP injects an RF signal into the power lines of a residential building and uses the power lines as a giant antenna for propagating the signals. The mobile wireless tag can then use this signal's characteristics as the feature set to fingerprint locations within the area where the power lines can reach. The latest revision of this system utilizes a feature set that samples 44 different frequencies of the amplitudes of the signal for location fingerprinting [80]. All the illustrations shown in this paper are either using the original data of this system or a modified version of it. The data of the system was gathered in a residential laboratory on a university campus. This lab has a similar layout and electrical infrastructure as a common residential house. We marked out one meter by one meter blocks on the floor, producing 66 different locations for our site survey.

In the next section, we will discuss the current problems in building a good location fingerprint with the existing, analytic text-based machine learning approaches. In Section 4, we will briefly provide an overview of the visualization interface. We will present a scenario that demonstrates how our visual analytics system works in Section 5. More details and example uses of the visualization will be discussed in Section 6.

B.3 Radio Frequency Fingerprinting-based localization

B.3.1 System Development Procedure

The procedure to build a location fingerprinting system can be roughly decomposed into three steps. The system presented here is focused on supporting the last two steps.

1. The first step is to gather the datasets and feature sets that can be potentially used to generate a location fingerprint database. This requires a tedious site survey that maps where the RF signals are gathered in the real world.

2. The second step is to find the right set of RF signal characteristics for the fingerprints. This step involves feature selection and building the fingerprints with the selected features.

3. The last step is to test the collected fingerprints with RF signals received at random locations in the surveyed area (random fingerprints). The signal data will be input to the localization system to see if it can accurately find the true locations of these random fingerprints through classification algorithms.

B.3.2 Problems and Challenges for Building Location Fingerprints

The generation of the location fingerprint database on a radio map requires a site survey in advance. This survey normally requires a user to manually tell the system where they are so that the system can learn the RF signal pattern at that specific location. This process can be very tedious and time-consuming. For example, in the PowerLine Positioning system, the time to survey each location with the full 44 features can take around 2 minutes. It takes about 2 hours to survey 66 locations in practice. If the location fingerprints are unstable over time, users might need to conduct the site survey again later to calibrate the system.

One major challenge is how to find the best features that can be used for building a set of good location fingerprints. In practice, we would like to use as few features as possible to build the fingerprints. There are two reasons for this. First, the fewer the features means that the training time and classification time for the machine learning algorithm can be shorter. For real-time localization, this can be very crucial. Second, fewer numbers of features for a fingerprint can result in a shorter time required for the site survey data collection process. Half the number of features needed means half the time for this tedious preprocessing procedure. However, the fewer the features used, the less likely individual fingerprints will be unique, resulting in higher overall classification error. So the technical challenge is how to find a balancing point where a smaller set of features can be used while the system is still capable to accurately classify a certain area of interest.

B.3.3 Problems with the Current Approach

The current approach used by localization developers to prove these required properties of the location fingerprints are achieved is by running machine learning algorithms with the fingerprints gathered at different times. The outputs of this approach are the text-based classification accuracy and misclassified locations when they test the fingerprints. There are several problems with this approach:

First of all, it is not easy to tell how each feature composed in a fingerprint is contributing to the overall classification results. For practical applications, one might have a few locations that are more important to be always classified with high accuracy while other locations are fine to be occasionally incorrect. There are many feature selection algorithms to analyze how each feature can build up the overall accuracy. However, different features may improve the classification accuracies of different areas on the radio map while they all improved the same overall accuracy.

Additionally, if there are a few locations that are always misclassified by the algorithm, it is very difficult to dig down into the multi-dimensional raw feature sets to identify the problem. Is it caused by a problematic training data set gathered or is the current fingerprint just not unique enough to correctly classify this location? If this kind of debugging cannot be performed, it is very hard for a location fingerprinting system to be practically deployed with the desired accuracy for any specified area of interest. Moreover, during the site survey process, sometimes there are RF interferences. These interference events can jeopardize the reliability of the produced fingerprints that should be mostly accurate for the most common cases. Moreover, it is not easy to find extreme cases when dealing with multidimensional data.

The requirements can be summed up in two major questions that need to be answered:

1. How do we effectively find a set of location fingerprints that are good enough for certain areas of interest?

2. If there are some locations that consistently receive inaccurate classifications, how do we find the problem?

To answer these two questions, several capabilities are required.

1. Test new unknown fingerprints with a preview of classification results on a map.

2. Test different subsets of features that can be used to compose the location fingerprints.

3. Examine the raw data of each individual feature for the fingerprints at different locations and its temporal stability.

4. Examine the spatial variance between locations in the high dimensional feature space of the fingerprints.

Figure 47: The four panels of our visual analytics system interface.

The design of the visual analytics system directly addresses these questions and targets these tasks. However, in subsequent use of the system, several unexpected interesting insights of the datasets and features were also discovered.

B.4 Visual Analytics System Overview

B.4.1 Interface Overview

The interface of the system contains four main panels as shown in Figure 47.

1. **Dataset selection** – This panel allows developers to select the datasets to be viewed or used. Datasets can be selected individually or with others according to the operation context. For example, multiple datasets should be selected when one attempts to calculate the standard deviation between them whereas only a single selection is needed when one attempts to view the raw feature value of a specific dataset. To the right of the dataset selection combo box is a timeline that shows when the datasets selected were collected. When a dataset is selected, the oval symbol representing it will be highlighted in blue. Each oval symbol on the top or bottom of the timeline represents a data set gathered.

2. **Feature selection** – This panel allows developers to select the features to use to compose the fingerprints. It supports single or multiple selections according to the context of use.

3. **Main map** – The main map panel is the display area for the geospatial visualization. A preloaded map is displayed in the background to provide the geospatial context for the visualization. By selecting different viewing perspectives, datasets and feature sets, this panel shows a grid-based heat map for the selected parameters. The heat map representation is very useful in showing the relative query results between different locations on the map. This visualization technique is particularly effective for examining a fingerprinting-based localization system because we are most interested in the spatial differentiability of the location fingerprints. At the bottom of this perspective is the status bar. It shows the current selected feature set, the mouse interaction mode and information about the heat map being presented.

4. **Perspective control** – This panel is used to control the viewing perspectives. The system provides three different viewing perspectives, each showing a different type of information of the datasets, features and complementing each other when the developer intends to drill down to a specific problem.

- **Data Variance Perspective** (Figure 48) shows the raw data of all the datasets with their corresponding feature sets.

- **Spatial Variance Perspective** (Figure 53) shows the spatial variance of between fingerprints in the high dimensional feature space using the selected features.

- **Test Classification Perspective** (Figure 49) provides a geospatial representation to show the results of the location classification using the generated fingerprints.

We use a green-gray-red color scheme for the heat maps displayed in the main map panel. Green indicates better results and red indicates worse. As for other colors used in the system, we avoid using green or red to avoid any semantic confusion.

Table 18: Feature transformation for ranking version of PLP

Feature	Value	Rank
antenna_amp_447	-56.7354796	2
antenna_amp_448	-56.9711881	1
antenna_amp_500	-55.5012413	5
antenna_amp_600	-56.3032417	4
antenna_amp_601	-56.3032471	3

B.5 Scenario

B.5.1 PLP Ranking Dataset

To illustrate use of this visual analytics system, we present an actual analysis scenario we conducted using the PLP data. From our previous research, we knew the original feature values (the raw signal data) from the power line is useful for localization. However, since the original data was real valued, it is sometimes more clustered in the high dimensional feature space. As a result, when the location fingerprints contains certain amount of noise in the signal, the classification would be incorrect. Therefore, one of the researchers proposed to transform the features of the datasets from raw amplitude values into the relative ranks of raw amplitude values as illustrated in Table 1. Using the ranking of the original feature values will create a unified spacing in between the them for each block. In theory, this approach can be more robust to noise because the real values are dynamically ranged and rounded up into a ranking form. Our task is to see if the PLP ranking version is better than the original PLP system.

One major evaluation criteria for PLP ranking version is to compare an optimal set of fingerprints built for an area of interest of it to the original PLP. The following scenario will show how to use the system to rapidly build a good location fingerprint database that is capable of maximizing the classification accuracy of an area of interest for the PLP ranking version. The same procedure is conducted on the original PLP for comparison. For the scenario, we assume that the kitchen area in the residential

lab (lower left area) is our area of interest as shown in Figure 48.

B.5.2 Scenario: Building an optimized fingerprint for an area of interest

B.5.2.1 Temporal stability feature selection

After importing all the datasets, ranked feature sets and the residential lab map into the system, the system will begin with the Data Variance Perspective (Figure 48). It shows the raw feature values as a heat map on the main map view. The greener blocks represents higher raw values (stronger signal). The first thing we would like to determine for feature selection is whether the datasets we gathered at different times are consistent enough to build reliable fingerprints. Therefore, we check the "Calculate STD between datasets" checkbox and select all the datasets to calculate the standard deviation of each feature throughout all the datasets. Previous research found that the smaller this standard deviation is, the higher the overall system accuracy will be [48]. Because our focus is to compare the consistency of different features, we then check the "Global color over all features" checkbox to dynamically range the colors properly for inter-feature comparison. The main map now shows a mostly green heat map. This means that most of the locations on the map for the selected feature are roughly consistent. By cycling through the features, we find that 9 features are exhibiting less consistent values (red blocks) at our area of interest such as the one shown in Figure 48. Therefore, they are eliminated from our potential feature set for the target fingerprints.

B.5.2.2 Preview classification result

We now switch to the Test Classification Perspective to preview location classification results on the map (Figure 49). By default, it will use all the features to perform a leave-one-out cross validation on the datasets. As explained earlier in Section 3.2, we generally prefer a smaller fingerprint with fewer features. As a result, we first eliminate the 9 features identified in the last section from the full 44 features.

Figure 48: Standard deviation view of a selected feature in the Data Variance Perspective that shows several temporally unstable blocks. One is in the kitchen area and two are in the rooms at the back of the house.

Then, we click the Auto Selection button to use a correlation-based feature selection algorithm to automatically filter out some irrelevant features from the remaining 35 features. This results in an elimination of 14 more features. The classification results are shown in Figure 49. However, by cycling through different test datasets to use for cross validation, we notice that although some of them have all the kitchen's blocks correctly classified, some test datasets like **N2-1**, still have a couple of blocks misclassified.

B.5.2.3 Debugging problematic blocks by finding spatial variance problems

In order to find the problem of the two misclassified blocks, we switch into the Spatial Variance Perspective (Figure 50). In this perspective, we click on the problematic block **a1.0n** located at the lower left corner of the map in the kitchen area. From the heat map shown in Figure 50, we can see the reddest block on the map is **b2.0n**, the

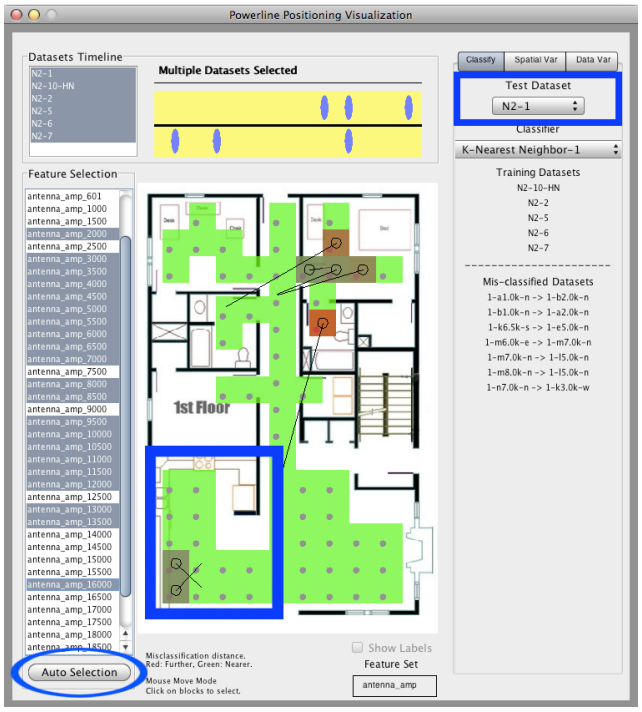

Figure 49: Features selected using automatic correlation-based feature selection in the Test Classification Perspective. 21 features are selected in this view. **N2-1** have misclassified two blocks in the lower left corner of the kitchen when used as the test dataset.

block where **a1.0n** was misclassified to. In this case, **a1.0n** was probably misclassified because of the closeness of the fingerprints in the high dimensional feature space of these two blocks. Therefore, if we can find a few features to change this closeness, the classification could potentially be corrected. So by the clicking on the block, we can pull up the parallel coordinates view that shows the differences of all the raw feature values from other blocks to further inspect the data.

In the parallel coordinates view, we can visualize all the raw feature values for a selected block to directly identify its degree of spatial differentiability and temporal stability. The default view shows the difference of raw feature values between **a1.0n**, the block selected, and the block on the y-axis. A higher value shown in the x-axis indicates there is more spatial variance between the blocks in the high dimensional feature space. Moreover, if we only select one feature to investigate and plot all the datasets' values together, we will be able to identify the degree of temporal stability too by visually observing the pattern overlapping amount in the plot. Therefore, the ideal form of parallel coordinates for a good feature should have a pattern like the one shown in Figure 51 (a). However, due to RF interference and multi-path propagation, in many cases we will see patterns like Figure 51 (b) or (c) which either do not have sufficient spatial differentiability or temporal stability. As a result, we could identify features like **antenna_amp_500** (b) and **antenna_amp_10000** (c) in Figure 51 to be potential removal candidates.

Now with these removal candidate features identified, we can go back to the Test Classification Perspective and conduct some trial and error with these features included or excluded to see the effect on the overall classification results. It turns out that removing **antenna_amp_9500** will give us the distinction needed for the lower left block (Figure 52(a)). We can continue this procedure several times to optimize all the classifications of the blocks in the area of interest.

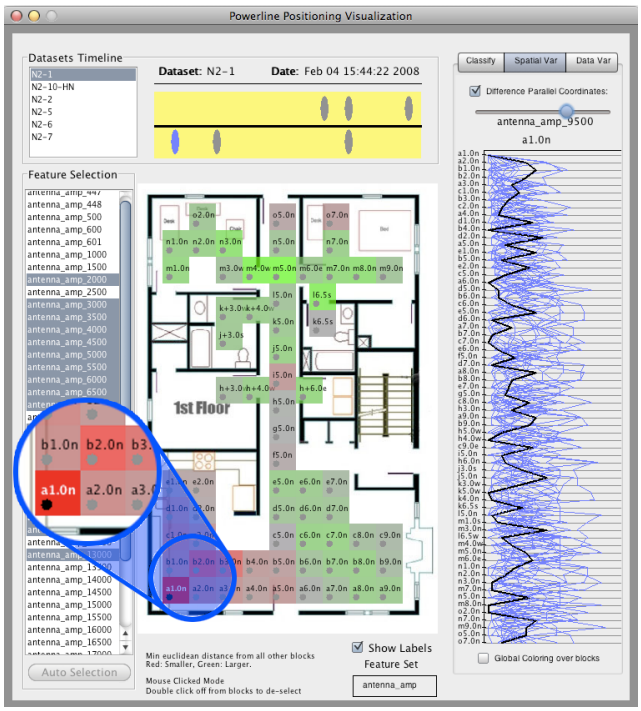

Figure 50: Selected block **a1.0n** is clearly closest to block **b1.0n** in the Spatial Variance Perspective.

B.5.2.4 Result comparison

Within a few minutes of experimenting with the feature selection, we managed to find 15 features for the fingerprints that can best classify locations in the kitchen (97.44 percent accuracy) as shown in Figure 52 (a). For comparison, we also used this method to find the best fingerprints in the original PLP datasets. Five features were initially filtered from the temporal stability test and 12 features were further eliminated through the automatic feature selection algorithm. After trial and error selection of features, the resulting fingerprints contains 11 features with good accuracy in the kitchen (94.87 percent accuracy) as illustrated in Figure 52 (b). To sum up, the PLP ranking version was not obviously better than the original version on the numbers. However, the misclassified blocks for the PLP ranking version were misclassified to closer blocks than the original PLP. These geospatial differences on the classification results are not easy to spot when using text-based machine learning

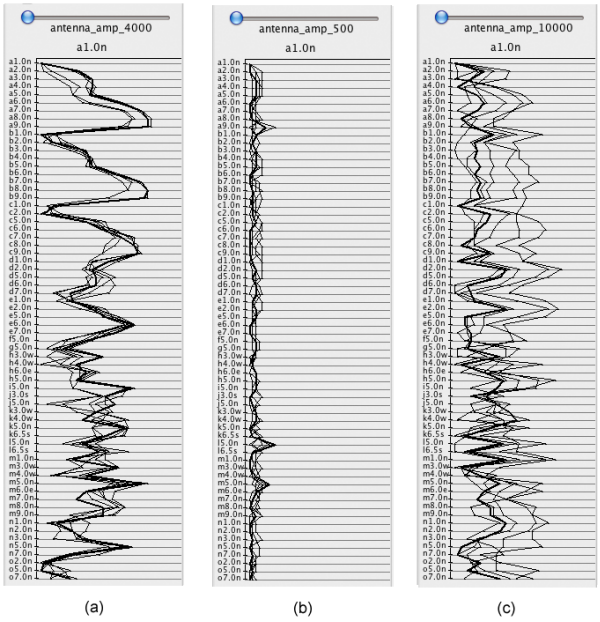

Figure 51: Parallel Coordinates of three different features plotted from block **a1.0n**. Each of the lines represents a different dataset. (a) An ideal feature with difference of feature values consistent across datasets and have sufficient spatial variance to most of the blocks (b) Problematic feature with high temporal stability but low spatial differentiability (c) Problematic feature with high spatial differentiability but low temporal stability.

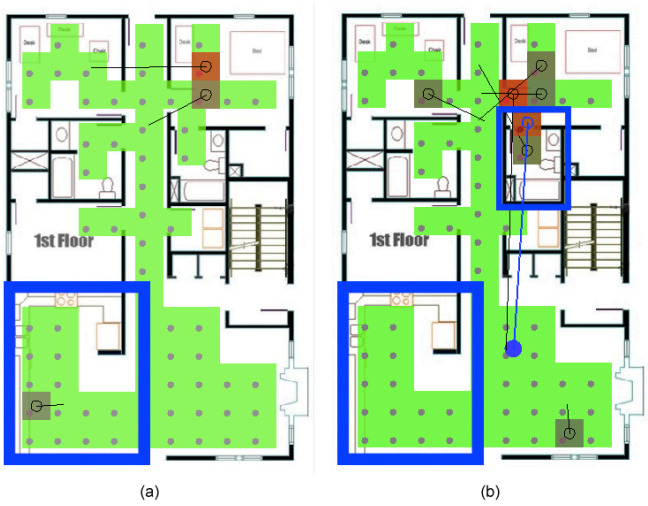

Figure 52: (a) PLP ranking version classification results with selected features using **N2-1** as test dataset. The overall classification accuracy for the kitchen area is 97.44 percent. (b) PLP original real valued data classification results with selected features using **N2-1** as test dataset. The overall classification accuracy for the kitchen area is 94.87 percent. However, the misclassified locations outside of the kitchen is far worse than the ranking version.

programs. In conclusion, the PLP ranking version does seem to do better overall for this scenario.

B.6 Visualization Design

This section will give more details on the design of the three main viewing perspectives and their specific use case in the PLP system.

B.6.1 Data Variance Perspective

B.6.1.1 Raw feature value view

This perspective provides a spatial view for the raw data collected for location fingerprinting. In this perspective, developers can choose which type of feature set to use for fingerprinting and be able to see the relative raw feature values on a heat map. In PLP, the coloring of this perspective is based on the raw signal strength values. The higher the value is, the greener the block is. The heat map can give us a

view of the data variance between the blocks for each feature selected. The coloring can be dynamically ranged either over one specific feature or over all the features. Developers can compare the colors of the blocks directly between different features when the colors are dynamically ranged over all the features.

In the PLP system, we discovered from the coloring patterns that the received signal strength (feature values) at the multiples of 3500 Hz are in general much stronger than the others. As stronger signals are easier to pickup and less susceptible to noise, they could potentially be better candidates for building the fingerprints. We also noticed that by placing two instances of the heat map visualization side by side, the color patterns of the blocks for several specific features of the first two datasets are closer to each other while the other four datasets are closer to each other. Since the first two datasets are gathered earlier than the others, we learned that these features are less temporally stable.

B.6.1.2 Standard Deviation

To see the data variance through time at a specific location of a certain feature, developers can chose to calculate the standard deviation of the data between a set of datasets selected in this perspective as shown in Figure 48. By the highlights of the datasets selected in the timeline, it is easy to tell the temporal stability of a certain feature at different locations on the map. One can compare the temporal stability between all the features when the colors are dynamically ranged over all the features. For example, in Figure 48, because smaller temporal variance is preferred, the blocks at the lower left corner and upper right corner showing redder colors in this view exhibited more temporal instability with the feature **antenna_amp_19000** over all the datasets. By simply selecting and deselecting different sets of datasets for this specific feature, we noticed that if we exclude the first two datasets gathered on the timeline, these two blocks will have much smaller variance (greener). Therefore, if we

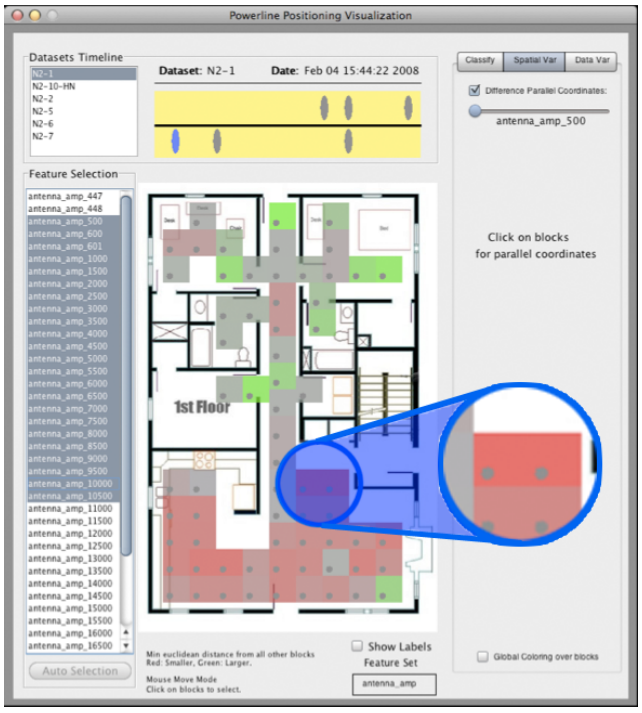

Figure 53: Minimal Euclidean distance view in the Spatial Variance Perspective that shows spatial variance. This set of features should not be selected for the location fingerprints if the two red blocks at the doorway is our area of interest.

wish to have a more temporally stable location fingerprint, we better not include this feature in our fingerprints when the two blocks at the bottom are important areas of interest.

B.6.2 Spatial Variance Perspective

B.6.2.1 Euclidean distance

This perspective shows the spatial variance between the locations inspected using the Euclidean distance of the selected features. The Euclidean distance is a very commonly used function to find the distance between two points in a high dimensional space. The function is also used by the K-nearest neighbor algorithm, the most widely used algorithm for RF location fingerprinting techniques [52]. Using the Euclidean distance of the features selected between the blocks in the high dimensional feature space provides the developers a view of how spatially differentiable a potential

fingerprint is.

B.6.2.2 Minimal Euclidean distance view

The default heat map shows the minimal Euclidean distance from all other blocks of the selected features as shown in Figure 53. As we prefer to avoid sets of features that generates little spatial variance, by showing the minimal Euclidean distance from all other blocks on the focused block can give us a general idea of how likely this block can be misclassified. The smaller this distance is, the redder the block is. Since the colors overlaid are by default dynamically ranged over the values shown in this view, it is fine for a red block to be present as long as it can be correctly classified. However, the redder blocks will have a relatively closer neighbor when represented in the high dimensional feature space so we certainly do not want them to be at our area of interest. For example, in Figure 53, if the area in front of the the doorway leading to the stairs is our area of interest, this set of selected features is probably not optimal for this dataset because it is more likely to cause misclassifications at those locations.

B.6.2.3 Euclidean distance from others view

When hovering the mouse over a block in the minimal Euclidean distance view, another heat map will show the Euclidean distance of the block selected from all the other blocks as shown in Figure 50. The closer the Euclidean distance of a block is from the block selected, the redder the block is. Developers can dynamically range the colors over all the Euclidean distances from of the blocks to directly compare the colors of the blocks of one selected block to other selected blocks. In this case, by hovering through different blocks with certain selected features, the more greener blocks are shown, the more spatially differentiable this hovered-over block is.

B.6.2.4 Parallel coordinates for all the datasets

For more details of the spatial differentiability for the multidimensional features, one can select a block and bring up a parallel coordinates plot as shown in Figure 50. Parallel coordinates can transform the analysis of the relations of multidimensional data into a two-dimensional pattern recognition problem [44]. Many works in geovisual analytics for multivariate visualization have used parallel coordinates for further exploration of the underlying data [24, 33, 47]. for The y-axis of the parallel coordinates lists all the blocks ordered by their physical distance from the selected block. By default, the value on the x-axis shows the individual feature value differences from the selected block's. The slider on top can highlight individual features in the parallel coordinates.

This view is particularly useful when used to identify temporal stability and spatial differentiability problems for a specified block when the x-axis is showing the feature value differences, the higher this value is for a specific feature generally means the more essential this feature is for creating spatial variance between the y-axis block and the block selected. If we choose to display multiple datasets' values for a specific feature, the overlapping pattern of these lines will directly indicate how temporally stable this feature is. An ideal visual pattern of a feature at a block should be like the one shown in Figure 51 (a). On the contrary, features of (b) and (c) in Figure 51 are probably not good candidates because they do not show both the preferred properties mentioned above.

By selecting an area of interest, developers can add in features one by one to generate the best minimalist set of features that make this location fingerprint more spatially differentiable when used together with the heat map. Developers can also select all the features at first and use the feature highlighting slider on the top of the parallel coordinates to find out which features are more problematic and eliminate them.

The parallel coordinates view also provides the developers a view of all the original feature values of all the blocks in the selected dataset. This view is very useful for finding extreme values in the raw feature data. In PLP, this view helped us find features with higher amplitudes that may be more distinguishable and less prone to noise. An extreme feature value might be caused by a temporary RF interference that occurred during the site survey process. Clearly, we normally do not want to include it in the location fingerprints. Therefore, avoiding using features that produce extreme peaks at our area of interest is one way to optimize the location fingerprints generated.

B.6.3 Test Classification Perspective

B.6.3.1 Leave-one-out cross validation

This perspective shows a geospatial view of the location classification results as shown in Figure 49. In this perspective, developers can select the training datasets, test dataset, and the machine learning algorithm to classify fingerprints with their selected features. All the datasets used for the classification are first selected in the dataset selection panel. Then, by selecting one of the selected datasets as the test dataset on the Perspective Control panel, the system will use the rest of the selected datasets as the training datasets to classify the instances in the test dataset, performing a step in the leave-one-out cross validation. By default, the K-nearest neighbor classifier using one nearest neighbor is used as it is the most frequently used classifier for RF location fingerprinting systems [52]. Several other classifiers provided by Weka machine learning toolkit [92] such as J48 decision trees, K-Star, Naive Bayes etc. are also provided in this perspective. The result of the classification will be shown as a heat map that is color-coded by how physically far a location is misclassified. The further a block is misclassified, the redder the block is. A line pointing to the misclassified location will also be drawn on the map. To avoid confusion introduced by overlapping lines when many locations are misclassified, the lines will jitter upon

mouse movement. Hovering the mouse over the block will place a highlighted circle on the misclassified block and a static highlighted line pointing to the misclassified location. For example in Figure 52(b), the block in the bathroom (the smaller highlighted area) was misclassified to the block far away, therefore, it is red.

B.6.3.2 Feature selection

Developers can select features that are used for location classification in the Test Classification Perspective through the feature selection panel as shown in Figure 49. They can use this panel to test the features they selected and use a trial and error approach to select the features that show the most promising results on the heat map. The visual analytics system provides an automatic feature selection function that uses Weka's correlation-based feature subset selection (CFS) algorithm with best-first search. The result of this feature selection function is not optimal but provides a fast and relatively good result as a starting point for the trial and error approach. The automatic feature selection button uses all the currently selected features as the full feature set to select from. Therefore, developers can first select a subset of features that they would like to automatically select from and execute the CFS algorithm on them to further narrow down the features selected. The major advantage of this approach of feature selection is that developers can see how each feature contributes to the accuracy of the classification for each block. We would like to select a subset of features for fingerprinting that can better classify our area of interest with a lower overall accuracy than a subset of features that cannot but with a higher overall accuracy. In PLP, from the scenario given in Section 5, with the trial and error feature selection method, we can quickly identify 11 out of 44 features that matters most for the fingerprinting system to achieve an over 90 percent classification accuracy in the kitchen within a few minutes as shown in Figure 52 (b).

B.7 Discussion and Future Directions

Two major issues of this visual analytics system are its scalability and generalizability. With regard to scalability, the system was developed with the PLP data which consists only 6 datasets, 44 features, and 66 blocks surveyed. The low number of instances allowed the system to produce almost instantaneous results when running K-nearest neighbor classifier on it. The fast calculation of all the Euclidean distances in the spatial variance perspective also benefited by the small amount of instances present. If there are hundreds of datasets, each dataset has hundreds of features, or a slower classifier is used, the system will very likely be running too slow to be usable. Moreover, developers will also have more trouble navigating through the current implementation when hundreds of datasets and features are shown. A more scalable design of the system should support distributed computing for the machine learning algorithms when dealing with large and high dimensional datasets.

The second issue is generalizability. The system design can be generalized for different kinds of features provided by different types of RF systems such as WiFi and GSM. However, its support for different types of machine learning algorithms is very limited, especially on the Spatial Variance Perspective. If using a different algorithm other than K-nearest neighbors for classification, the notion of "distance" will be different. Future work for generalizability should explore having more distance functions built in for showing spatial variances between different blocks and coupling them with the classification algorithm in the Test Classification Perspective. Systematic user evaluations on other fingerprinting-based location systems can be further conducted to better understand the generalizability and scalability issues of our system.

The visual analytics system presented in this paper can be further improved by delegating more currently required manual work to the computational system. For example, in the Data Variance Perspective, a color-coded feature list that can show

the "reddest" block color or the average color presented in its individual view can save developers the manual work of going through all the features in the list to find the temporally unstable features. In the Test Classification Perspective, instead of requiring developers to cycle through the more detailed leave-one-out cross validation, providing one meta 3-fold cross-validation view should be sufficient for them to learn whether the selected features are good enough.

B.8 Conclusion

The visual analytics system presented in this paper aids in the development of RF fingerprinting-based localization systems by helping developers select appropriate features for a good location fingerprint. The geospatial visualization in the Test Classification Perspective gives developers not only the accuracy results of the location classification, but also the exact location of misclassified blocks. The Data Variance Perspective with the standard deviation heat map can give the developers an idea of which features are better with respect to temporal stability. For more detailed spatial differentiability information, the Spatial Variance Perspective allows users to go down level by level into the raw feature value differences. This system supports many visual aids for location fingerprint feature selection and a clear geospatial output representation that text-based analytic machine learning algorithms can not deliver.

The system was able to effectively identify good location fingerprints on the PLP system as shown in this paper. By combining the power of interactive visualization and computational data analysis, we think this system is a great example of how visual analytics can support the development of technologies that will potentially shape our lives in the future.

REFERENCES

[1] "Agilent 33220a 20 MHz Function / Arbitrary Waveform Generator." http://cp.literature.agilent.com/litweb/pdf/5988-8544EN.pdf.

[2] "Crossbow MPR-MIB Users Manual, Revision A. (June 2007)." http://www.xbow.com/support/support_pdf_files/mpr-mib_series_users_manual.pdf.

[3] "Ettus Research LFRX Daughterboard." http://www.ettus.com/downloads/ettus_ds_USRP_TXRX_v5b.pdf.

[4] "Ettus Research Universal Software Radio Peripheral." http://www.ettus.com/downloads/ettus_ds_usrp_v7.pdf.

[5] "GNU Octave." http://www.gnu.org/software/octave/.

[6] "GNU Radio." http://gnuradio.org.

[7] "Homeplug Powerline Alliance: Frequently Asked Questions." http://www.homeplug.org/about/faqs/.

[8] "Insteon." http://www.insteon.net.

[9] "International Telecommunication Union: Frequently Asked Questions." http://www.itu.int/ITU-R/terrestrial/faq/index.html.

[10] "Sun Microsystems SunSPOT wireless sensors." http://www.sunspotworld.com/.

[11] "Texas Instruments CC2420 Datasheet." http://focus.ti.com/lit/ds/symlink/cc2420.pdf.

[12] "Texas Instruments MSP430 Ultra-Low-Power Microcontrollers Datasheet." http://focus.ti.com/lit/sg/slab034q/slab034q.pdf.

[13] "Ubisense." http://www.ubisense.net.

[14] "Vicon MX." http://www.vicon.com/products/systems.html.

[15] "Weka: Data Mining with Open Source Machine Learning Software in Java." http://www.cs.waikato.ac.nz/ml/weka/.

[16] "Wellbrook Communications ALA-1530+ Antenna." http://www.wellbrook.uk.com/products.html.

[17] "X10: Standard and Extended X10 Protocol." http://software.x10.com/pub/manuals/xtdcode.pdf.

[18] "Calculating Radiated Power and Field Strength for Conducted Power Measurements," Tech. Rep. TN1200.04, Semtech Corporation, 2007.

[19] "Indoor GPS." `http://www.indoorgps.com`, 2007.

[20] "United States Federal Communications Commission: Part 15 Regulations." `http://www.fcc.gov/oet/info/rules/part15/PART15_07-10-08.pdf`, July 2008.

[21] Abramson, N., "The ALOHA System: Another Alternative for Computer Communications," in *Proceedings of the Fall Joint Computer Conference 1970*, 1970.

[22] Akyildiz, I. F., Su, W., Sankarasubramaniam, Y., and Cayirci, E., "Wireless Sensor Networks: A Survey," *Computer Networks*, vol. 38, pp. 393–422, March 2002.

[23] Ananthanarayanan, G., Haridasan, M., Mohomed, I., Terry, D., and Thekkath, C. A., "StarTrack: A Framework for Enabling Track-Based Applications," in *Proceedings of the Seventh Annual International Conference on Mobile Systems, Applications and Services (MobiSys 2009)*, 2009.

[24] Andrienko, G. L. and Andrienko, N. V., "Interactive visual tools to support spatial multicriteria decision making," in *UIDIS '01: Proceedings of the Second International Workshop on User Interfaces to Data Intensive Systems (UIDIS'01)*, (Washington, DC, USA), p. 127, IEEE Computer Society, 2001.

[25] Andrienko, G. and Andrienko, N., "Data mining with c4.5 and interactive cartographic visualization," in *User Interfaces to Data Intensive Systems, 1999. Proceedings*, pp. 162–165, 1999.

[26] Anthony LaMarca and Jeff Hightower and Ian Smith and Sunny Consolvo, "Self-Mapping in 802.11 Location Systems," in *Proceedings of the Seventh International Conference on Ubiquitous Computing (UbiComp 2005)*, 2005.

[27] Bahl, P. and Padmanabhan, V., "RADAR: An In-Building RF-Based User Location and Tracking System," in *Proceedings of the Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies (Infocom 2000)*, pp. 775–784, 2000.

[28] Bahl, P., Padmanabhan, V. N., and Balachandran, A., "Enhancements to the RADAR User Location and Tracking System," Tech. Rep. MSR-TR-2000-12, Microsoft Research, 2000.

[29] Bian, X., Abowd, G. D., and Rehg, J. M., "Using Sound Source Localization in a Home Environment," in *Proceedings of the Third International Conference on Pervasive Computing (Pervasive 2005)*, 2005.

[30] Burchfield, T. R., Venkatesan, S., and Weiner, D., "Maximizing Throughput in ZigBee Wireless Networks through Analysis, Simulations and Implementations," in *Proceedings of the International Workshop on Localized Algorithms and Protocols for Wireless Sensor Networks*, 2007.

[31] Castro, P., Chiu, P., Kremenek, T., and Muntz, R., "A Probabilistic Room Location Service for Wireless Networked Environments," in *Proceedings of the Third International Conference on Ubiquitous Computing (UbiComp 2001)*, pp. 18–34, 2001.

[32] Collins, R., Lipton, A., Fujiyoshi, H., and Kanade, T., "Algorithms for Cooperative Multi-sensor Surveillance," *Proceedings of the IEEE*, vol. 10, no. 89, pp. 1456–1477, 2001.

[33] Edsall, R. M., "The parallel coordinate plot in action: design and use for geographic visualization," *Comput. Stat. Data Anal.*, vol. 43, no. 4, pp. 605–619, 2003.

[34] Ekahau. `http://www.ekahau.com`, 2007.

[35] Fleming, R., "Line Cord Antenna." United States Patent #3324473, 1965.

[36] Fritz, G., Seifert, C., and Paletta, L., "A Mobile Vision System for Urban Detection with Informative Local Descriptors," in *Proceedings of The Fourth IEEE International Conference on Computer Vision Systems (ICVS 2006)*, 2006.

[37] G. Cohn and E. Stuntebeck and J. Pandey and B. Otis and G. D. Abowd and S. N. Patel, "SNUPI: Sensor Nodes Utilizing Powerline Infrastructure," in *Proceedings of the 12th ACM International Conference on Ubiquitous Computing (UbiComp 2010)*, 2010.

[38] Global Search Communications. `http://www.globalsearch.ws`, 2007.

[39] Han, Y., Stuntebeck, E., Stasko, J., and Abowd, G., "A Visual Analytics System for Radio Frequency Fingerprinting-based Localization," in *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology (VAST 2009)*, 2009.

[40] Hashemi, H., "The Indoor Radio Propagation Channel," *Proceedings of the IEEE*, vol. 81, no. 7, pp. 943–968, 1993.

[41] Hazas, M. and Ward, A., "A novel broadband ultrasonic location system," in *Proceedings of The Fourth International Conference on Ubiquitous Computing (UbiComp 2002)*, 2002.

[42] Hazas, M. and Ward, A., "Broadband ultrasonic location systems for improved indoor positioning," in *IEEE Transactions on Mobile Computing*, 2006.

[43] Hightower, J. and Borriello, G., "A Survey and Taxonomy of Location Systems for Ubiquitous Computing," Tech. Rep. CSC-01-08-03, University of Washington, 2001.

[44] Inselberg, A., "Multidimensional detective," in *Proceedings of the 1997 IEEE Symposium on Information Visualization (InfoVis 1997)*, IEEE Computer Society, 1997.

[45] Jeffrey Hightower and Gaetano Borriello and Roy Want, "SpotON: An Indoor 3D Location Sensing Technology Based on RF Signal Strength," Tech. Rep. UW-CSE-2000-02-02, University of Washington, 2000.

[46] Ji, Y., Biaz, S., Pandey, S., and Agrawal, P., "ARIADNE: A Dynamic Indoor Signal Map Construction and Localization System," in *The Fourth International Conference on Mobile Systems, Applications, and Services (MobiSys 2006)*, 2006.

[47] Johansson, S. and Jern, M., "Geoanalytics visual inquiry and filtering tools in parallel coordinates plots," in *GIS '07: Proceedings of the 15th annual ACM international symposium on Advances in geographic information systems*, (New York, NY, USA), pp. 1–8, ACM, 2007.

[48] Kaemarungsi, K. and Krishnamurthy, P., "Modeling of Indoor Positioning Systems Based on Location Fingerprinting," in *Proceedings of IEEE Infocom (INFOCOM 2004)*, pp. 1012–1022, 2004.

[49] Kaemarungsi, K. and Krishnamurthy, P., "Properties of Indoor Received Signal Strength for WLAN Location Fingerprinting," in *Proceedings of the Annual International Conference on Mobile and Ubiquitous Systems (MOBIQUITOUS 2004)*, pp. 14–23, 2004.

[50] Kidd, C. D., Orr, R., Abowd, G. D., Atkeson, C. G., Essa, I. A., MacIntyre, B., Mynatt, E., Starner, T. E., and Newstetter, W., "The Aware Home: A Living Laboratory for Ubiquitous Computing Research," in *Proceedings of the Second International Workshop on Cooperative Buildings (CoBuild 1999)*, 1999.

[51] King, T., Haenselmann, T., and Effelsberg, W., "Deployment, calibration, and measurement factors for position errors in 802.11-based indoor positioning systems," in *LoCA*, 2007.

[52] Kjærgaard, M., "A Taxonomy for Radio Location Fingerprinting," in *In the Proceedings of the Third International Symposium on Location and Context Awareness (LoCA 2007)*, 2007.

[53] Krumm, J., Cermak, G., , and Horvitz, E., "Rightspot: A novel sense of location for a smart personal object," in *Ubicomp*, 2003.

[54] Krumm, J., ed., *Ubiquitous Computing Fundamentals*. CRC Press, 2010.

[55] Krumm, J., ed., *Ubiquitous Computing Fundamentals*, pp. 285–319. CRC Press, 2010.

[56] Krumm, J. and Horvitz, E., "LOCADIO: Inferring Motion and Location From Wi-Fi Signal Strengths," in *Proceedings of the First International Conference on Mobile and Ubiquitous Systems, Networking, and Services (MOBIQUITOUS 2004)*, 2004.

[57] Krumm, J. and Platt, J., "Minimizing Calibration Effort for an Indoor 802.11 Device Location Measurement System," Tech. Rep. MSR-TR-2003-82, Microsoft Research, 2003.

[58] Laasonen, K., "Clustering and prediction of mobile user routes from cellular data," in *PKDD*, 2005.

[59] LaMarca, A., Chawathe, Y., Consolvo, S., Hightower, J., Smith, I., Scott, I., Sohn, T., Howard, J., Hughes, J., Potter, F., Tabert, J., Powledge, R., Borriello, G., , and Schilit, B., "Place lab: Device positioning using radio beacons in the wild," in *Pervasive 2005*, 2005.

[60] LaMarca, A. and de Lara, E., *Location Systems: An Introduction to the Technology Behind Location Systems*, vol. 4 of *Synthesis Lectures on Mobile and Pervasive Computing*. Morgan & Claypool Publishers, 2008.

[61] Lee, Y.-W., Stuntebeck, E. P., and Miller, S., "MERIT: Mesh of RF Sensors for Indoor Tracking," in *Proceedings of the Third Annual IEEE Communications Society Conference on Sensor, Mesh, and Ad Hoc Communications and Networks (SECON 2006)*, 2006.

[62] Lemelson, H., King, T., and Effelsberg, W., "Pre-processing of Fingerprints to Improve the Positioning Accuracy of 802.11-based Positioning Systems," in *Proceedings of the First ACM International Workshop on Mobile Entity Localization and Tracking in GPS-less Environments (MELT 2008)*, pp. 73–78, 2008.

[63] Li, B., Salter, J., Dempster, A. G., and Rizos, C., "Modeling of Indoor Positioning Systems Based on Location Fingerprinting," in *Proceedings of the First IEEE International Conference on Wireless Broadband and Ultra Wideband Communications*, 2006.

[64] Lionel M. Ni and Yunhao Liu and Yiu Cho Lau and Abishek P. Patil, "LANDMARC: Indoor Location Sensing Using Active RFID," *Wireless Networks*, no. 10, pp. 701–710, 2004.

[65] Madhavapeddy, A. and Tse, T., "Study of bluetooth propagation using accurate indoor location mapping," in *Ubicomp 2005*, 2005.

[66] Milligan, T. A., *Modern Antenna Design.* John Wiley & Sons, second ed., 2005.

[67] Mitchell, T., *Machine Learning.* McGraw Hill, 1997.

[68] Moustafa A. Youssef and Ashok Agrawala and A. Udaya Shankar, "WLAN Location Determination via Clustering and Probability Distributions," in *In the Proceedings of the First IEEE International Conference on Pervasive Computing and Communications (PerCom 2003)*, 2003.

[69] Otsason, V., Varshavsky, A., LaMarca, A., and de Lara, E., "Accurate GSM Indoor Localization," in *The Proceedings of the Seventh International Conference on Ubiquitous Computing (UbiComp 2005)*, 2005.

[70] Papliatseyeu, A., Kotilainen, N., Ibarra, O. M., and Osmani, V., "FINDR: Low-cost Indoor Positioning Using FM Radio," in *The Proceedings of MobilWare 2009 (MobilWare 2009)*, pp. 15–26, 2009.

[71] Patel, S. N., Reynolds, M. S., and Abowd, G. D., "Detecting Human Movement by Differential Air Pressure Sensing in HVAC System Ductwork: An Exploration in Infrastructure Mediated Sensing," in *Proceedings of the Sixth International Conference on Pervasive Computing (Pervasive 2008)*, 2008.

[72] Patel, S. N., *Infrastructure Mediated Sensing.* PhD thesis, Georgia Institute of Technology, 2008.

[73] Patel, S. N., Truong, K. N., and Abowd, G. D., "PowerLine Positioning: A Practical Sub-Room-Level Indoor Location System for Domestic Use," in *Proceedings of the Eighth International Conference on Ubiquitous Computing (UbiComp 2006)*, pp. 441–458, 2006.

[74] Priyantha, N. B., Chakraborty, A., and Balakrishnan, H., "The Cricket Location-Support System," in *The Proceedings of Mobile Computing and Networking (MobiCom 2000)*, 2000.

[75] Rekimoto, J. and Ayatsuka, Y., "Cybercode: Designing augmented reality environments with visual tags," in *Proceedings of Designing Augmented Reality Environments (DARE 2000)*, 2000.

[76] Rekimoto, J. and Katashi, N., "The world through the computer: Computer augmented inter-action with real world environments," in *UIST 1995*, 1995.

[77] Rice, S. O., "Mathematical Analysis of Random Noise," *Bell System Technical Journal*, vol. 23, 1944.

[78] Seidel, S. Y. and Rappaport, T. S., "914 MHz Path Loss Prediction Model for Indoor Wireless Communications in Multi-floored Buildings," *IEEE Transactions on Antennas and Propagation*, February 1992.

[79] Stüber, G. L., *Principles of Mobile Communication*. Kluwer Academic Publishers, second ed., 2001.

[80] Stuntebeck, E. P., Patel, S. N., Robertson, T., Reynolds, M. S., and Abowd, G. D., "Wideband PowerLine Positioning for Indoor Localization," in *The Proceedings of the Tenth International Conference on Ubiquitous Computing (UbiComp 2008)*, 2008.

[81] Teemu Roos and Petri Myllymakiäki and Henry Tirri and Pauli Misikangas and Juha Sievänen, "A Probabilistic Approach to WLAN User Location Estimation," *International Journal of Wireless Information Networks*, vol. 9, no. 3, pp. 155–164, 2002.

[82] Tektronix, "http://www.tek.com/."

[83] Thompson, M., "Line-Cord Antenna." United States Patent #2581983, 1947.

[84] Tulli, T., "Television Line Antenna Apparatus." United States Patent #4401989, 1982.

[85] ur Rehman, W., de Lara, E., and Saroiu, S., "CILoS: A CDMA Indoor Localization System," in *The Proceedings of the Tenth International Conference on Ubiquitous Computing (UbiComp 2008)*, pp. 104–113, 2008.

[86] Varshavsky, A., Chen, M. Y., de Lara, E., Froehlich, J., Haehnel, D., Hightower, J., LaMarca, A., Potter, F., Sohn, T., Tang, K., and Smith, I., "Are GSM Phones THE Solution for Localization," in *Proceedings of the Seventh IEEE Workshop on Mobile Computing Systems and Applications (WMCSA 2006)*, 2006.

[87] Varshavsky, A., de Lara, E., Hightower, J., LaMarca, A., and Otsason, V., "GSM Indoor Localization," *Pervasive and Mobile Computing*, vol. 3, pp. 698–720, December 2007.

[88] Varshavsky, A., LaMarca, A., Hightower, J., and de Lara, E., "The SkyLoc Floor Localization System," in *Proceedings of the Fifth IEEE International Conference on Pervasive Computing and Communications (PERCOM 2007)*, 2007.

[89] Want, R., Hopper, A., Falcao, V., , and Gibbons, J., "The Active Badge Location System," *ACM Transactions on Information Systems*, vol. 10, pp. 91–102, January 1992.

[90] Ward, A., Jones, A., and Hopper, A., "A New Location Technique for the Active Office," *IEEE Personal Communications*, vol. 4, no. 5, pp. 42–47, 1997.

[91] Weiser, M., "The Computer for the 21st Century," *Scientific American*, September 1991.

[92] Witten, I. H. and Frank, E., *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, 2 ed., 2005.

[93] Woodman, O. and Harle, R., "Pedestrian Localisation for Indoor Environments," in *The Proceedings of the Tenth International Conference on Ubiquitous Computing (UbiComp 2008)*, 2008.

[94] Woodman, O. J., "An Introduction to Inertial Navigation," Tech. Rep. UCAM-CL-TR-696, University of Cambridge, 2007.

[95] Yin, J., Yang, Q., and Ni, L., "Adaptive Temporal Radio Maps for Indoor Location Estimation," in *In the Proceedings of the Third Annual IEEE International Conference on Pervasive Computing and Communications (IEEE PerCom 2005)*, 2005.

[96] Youssef, M. and Agrawala, A., "The Horus Location Determination System," *Wireless Networks*, vol. 14, no. 3, pp. 357–374, 2008.