UNIVERSITY OF TARTU
FACULTY OF MATHEMATICS AND COMPUTER SCIENCE
Institute of Computer Science
Chair of Software Systems

Veljo Otsason

ACCURATE INDOOR LOCALIZATION USING WIDE GSM FINGERPRINTING

Master's Thesis

Supervisors: Prof. Eyal de Lara
Prof. Jüri Kiho

Tartu 2005

ACCURATE INDOOR LOCALIZATION USING WIDE GSM FINGERPRINTING

Master's Thesis

Veljo Otsason

Abstract

Accurate indoor localization has long been an objective of the ubiquitous computing research community, and numerous indoor localization solutions based on 802.11, Bluetooth, ultrasound and infrared technologies have been proposed. This Thesis presents the first accurate GSM-based indoor localization system that achieves median accuracy of 5 meters in large multi-floor buildings. The key idea that makes accurate GSM-based indoor localization possible is the use of wide signal-strength fingerprints. In addition to the 6 strongest cells traditionally used in the GSM standard, the wide fingerprint includes readings from up to 32 additional cells whose signals are strong enough to be detected, but too weak to be used for efficient communication. We evaluate our GSM-based indoor localization system in three multi-floor buildings located in two metropolitan areas. Experimental results show that our system achieves accuracy comparable to an 802.11-based implementation, and can accurately differentiate between floors in both wooden and steel-reinforced concrete structures.

To the memory of my father,

for his love, encouragement and great example

To my mother,

for her love, understanding and support

Acknowledgements

This research was done in Canada during my exchange year at the University of Toronto. It would have not been possible without the extremely valuable support, guidance and contribution from my co-supervisor Eyal de Lara, and his student Alex Varshavsky from the University of Toronto, as well as Anthony LaMarca from Intel Research, Seattle. Thank you very much for working closely with me on this project.

I would like to thank my co-supervisor Prof. Jüri Kiho from the University of Tartu, for his support, and for encouraging me to research the things I liked the most. I'd like to thank all the great people from the Systems Lab at the University of Toronto for making me feel like home from the day one. Your company, help and inspiration was invaluable. Also, I'm grateful to Neil Ernst for working with me on the initial experiments that finally led to this project.

I would like to thank people and organizations who have supported me with my academic pursuit. Great thanks to Estonian Scholarship Fund, Elmar Tampõld, Jüri Nurmberg and Tartu College in Toronto, Merli Tamtik and the Educational Advising Center in Tartu, Sirje Üprus at the International Relations Office of the University of Tartu, Miranda Cheng from ISXO at the University of Toronto, Tiit Roosmaa and the Institute of Computer Science, colleagues from Mobi Solutions, and many others.

Last but not least, I'm especially grateful to all the friends and family on both sides of the ocean for your love and support through difficult times, and for making my days in Toronto so unforgettably nice.

Table of Contents

Introduction

Developments in the wireless technology have enabled creating applications that are aware of the user's location. These applications use location to provide relevant information or use it otherwise for the benefit of the user. Different location aware applications are meant for different environments and require different accuracy. While many outdoors applications, such as friend-finder, can successively work with accuracy of hundreds of meters, indoor applications, like printing to the nearest printer or guiding people indoors, usually require granularity of a few meters.

The accurate localization of objects and people in indoor environments has long been considered an important building block for ubiquitous computing applications [37, 18]. Most research on indoor localization systems has been based on the use of short-range signals, such as Wi-Fi [4, 8, 21], Bluetooth [1], ultrasound [29, 39], infrared [37, 38], or RFID [14, 27]. This Thesis shows that contrary to popular belief an indoor localization system based on wide-area GSM signal fingerprints can achieve high accuracy, and is in fact comparable to an 802.11-based implementation.

GSM-based indoor localization has several benefits:

- GSM coverage is almost pervasive, far outreaching the coverage of 802.11 networks.
- The wide acceptance of cellular phones makes them ideal conduits for the delivery of ubiquitous computing applications.
- A localization system based on cellular signals, such as GSM, leverages the phone's existing hardware and removes the need for additional radio interfaces.
- Because cellular towers are dispersed across the covered area, a cellular-based localization system would still work in situations where a building's electrical infrastructure has failed. Moreover, cellular systems are designed to tolerate power failures. For example, the cellular network kept working during the massive power outage that left most of the North-Eastern United States and Canada in the dark in the summer of 2003.

- GSM, unlike 802.11 networks, is operating in a licensed band, and therefore does not suffer from interference from nearby devices operating on the same frequency (*e.g.*, microwave ovens, cordless phones, wireless keyboards, garage door openers, all Bluetooth devices etc).

- The significant expense[1] and complexity of cellular base stations results in a network that evolves slowly and is only reconfigured infrequently. While this lack of flexibility (and high configuration cost) is certainly a drawback for the cellular system operator, it results in a stable environment that allows the localization system to operate for a long period before having to be recalibrated.

This Thesis presents the first fine-grained GSM-based indoor localization system. We present results for experiments conducted on datasets collected from three multi-floor buildings in two large North American cities spanning a wide spectrum of urban densities, ranging from a busy downtown core to a quiet residential neighborhood. The results show that this fine-grained GSM-based indoor localization system can effectively differentiate between floors and achieves median within-floor accuracy as low as 2.5 meters.

The key idea that makes accurate GSM-based indoor localization possible is the use of *wide* signal-strength fingerprints. The wide fingerprint includes the 6 strongest GSM cells and readings of up to 32 additional GSM channels, most of which are strong enough to be detected, but too weak to be used for efficient communication. The higher dimensionality introduced by the additional channels dramatically increases localization accuracy.

We make the following contributions:

- We present the first accurate GSM-based indoor localization system and show that it achieves accuracy comparable to an 802.11-based implementation.

- We show that a GSM-based localization system can effectively differentiate between floors for both wooden and steal-reinforced concrete structures.

---

[1] A macro-cell costs $500,000 to $1-million (U.S.). Micro-cells cost about a third as much, but a larger number is needed to cover the same area [24].

- We show that there is significant signal diversity across metropolitan environments and that this diversity enables the GSM-based system to achieve high localization accuracy.
- We show that the availability of signal strength readings from cells other than the 6 strongest cells traditionally used in GSM increases localization accuracy by up to 50%.

The rest of this Thesis is organized as follows. First we give a general overview and motivation of the location-aware applications, after which common approaches to location sensing are described. Description of the wireless technologies that this Thesis is based on ends the first chapter. In the next chapter, related work is discussed. Several indoor and outdoor localization systems are described. Third chapter explains our methodology – our localization algorithms and methods, as well as data collection approach. Chapter 4 describes results of our experimental evaluation. Finally, Chapter 5 concludes the Thesis and discusses directions for future work.

1 Background

The rapid advancement of computing technology is constantly opening new possibilities. The shift from mainframes to personal computers as well as the shift from disconnected computers to networked ones both dramatically enhanced the way we use computer technology in our every-day lives. Technology advancements have made it possible to shrink the size of the computer, and connect it wirelessly, so that it can be easily used anywhere anytime. The paradigm shift from desktop to mobile computing opens new possibilities. Context awareness, the ability of applications to adjust their behavior based on the environmental information, is considered one of the essential aspects of this new paradigm. Context-awareness and context-based adaptation is particularly useful for mobile computing, enabling many valuable applications, such as locating people in case of emergencies, tracking patients and equipment in hospitals, finding friends or colleagues, or sending location and context-based advertisements to people inside shopping malls.

In this chapter we first give general overview of context awareness and location awareness in particular. We then discuss three common approaches for determining geographical location. Finally, we describe wireless technologies central to this Thesis, including GSM cellular system and 802.11 wireless LAN.

1.1 Context awareness

Context is any information that can be used to describe the general environment the application is used in. It can be information about people, their locations, activities and intentions, or anything else that is relevant to the application's functionality [6]. Abowd *et al.* point out five important parts of context ("five W's") [2]:

- *Who is the user and/or the other people around (identity)?*
- *What the user is doing (activity)?*
- *Where the user is (location)?*
- *When is the usage taking place, including relative time (time)?*
- *Why the user is doing what she is doing?*

The *Where* and *Who* of context (location and identity) have been widely researched. Olivetti Research Lab's Active Badge [37] and the Xerox PARCTab [38] were two of the first applications that used indoor location to provide context-aware services, such as automatic call forwarding and automatically updated maps of users' locations. Identity of the user is often used, but identities of other people have not gained that much attention. Time component is also widely used, but not with its full capacity. For example, relative changes in time could be used to interpret user's activity or intentions. Short visits at an exhibit could be used as an indication of lack of interest. Also, actions that diverge from the typical timeline can reveal useful information. For example, when an elderly person deviates from a typically active morning routine, a notification can be sent. The parts of context that deal with activity and user motivation (What, Why), are still widely unexplored, because of the complexity of extracting and representing this information. [2]

Additionally, basic contextual elements can be used (alone or in combination) to extract more sophisticated contextual information. For example, the identity of the user can be used to get user's phone number from the phone book. Location and identity can be leveraged to determine a list of friends near-by. Time and location can be leveraged together to get information about the weather conditions. [6]

1.1.1 Location awareness

Location is the most widely used contextual element. Location-aware applications take location into account to do their work or to show information to the user that is a function of their and/or other users' location. Different applications require different granularity of location information. For example, to show weather conditions where the user currently is, city-scale accuracy could be sufficient, but finding near-by friends requires accuracy of at least a few kilometers. Some applications are specifically meant for indoors use and therefore require higher accuracy. Even within different indoor applications, however, there is significant variation in accuracy requirements For example, locating the nearest printer requires different accuracy than locating a book in a library [4].

Location-awareness enables many useful commercial [1], educational [11], military and healthcare [37] applications. Efficient location and coordination of staff in large organizations is a recurring problem that can be relieved by location aware applications. Hospitals, for example, may utilize up-to date information about the location of staff and patients [37].

Location information is particularly useful, or sometimes even the matter of life and death, in case of emergencies. People needing help often do not know their exact location or are unable to communicate it, for example while having a heart attack and calling for help over a cell phone. U.S. Federal Communications Commission has approved the Enhanced 911 (E911) mandate [50], which requires wireless carriers to be able to locate, within 50 to 100 meters, any cell phone calling 911, the U.S. nation-wide emergency service number. E112 [49], the European equivalent of the American E911, does not require any particular accuracy – carriers only need to provide location capabilities that are compatible with their networks [36]. However, although 50 meters accuracy would significantly ease providing help in many situations, it does not help finding the person in high density areas, inside hospitals, office buildings, hotels or condominiums, with potentially tens of floors and hundreds of rooms. Therefore, satisfying the current E911 requirements is only the minimum that has to be done to accurately localize people and provide quick and effective help in these areas.

Cell phones are increasingly becoming the most ubiquitous mobile devices. In Europe more than 80% of the people carry cell phones, in North America the penetration is about 60% and growing rapidly. At the same time the capabilities of phones improve – screens become bigger, processing power and networking bandwidth increase etc. This makes mobile phones an ideal platform for ubiquitous computing and location-aware applications.

1.2 Location Sensing

An important building block of location-aware applications is the location sensing system, which provides the application with the actual geographical location of the user or other important entities. The location can be represented as absolute coordinates (longitude, latitude, elevation), relative coordinates (x,y relative to the corner of a building), or in symbolic form (such as "Room 5180", or "5th floor of the Bahen building"). This section describes three main approaches of extracting location using radio (infrared or ultrasound) signals. These approaches can be used separately or in combination to do actual location sensing.

1.2.1 Cell Identification

Most of the wireless radio networks make use of cellular architecture. This means that instead of one wide-range radio transmitter, many stations with smaller ranges are used. This allows more effective bandwidth and energy use. The general idea behind cell identification method is that such small stations (or cells) transmit unique location information that is only heard by mobile stations in the radio range of the particular cell. Reading this information, the stations can extract their own location. The transmitted information can be in a form of explicit location, or as a unique identifier, which requires further matching of identifiers and explicit locations to make the information useful.

Obviously, the accuracy of this approach depends on the size of cells. Unfortunately the optimal cell size of many technologies is quite large. In case of cellular telephony systems such as GSM, the cell size can be up to tens of kilometers [26]. Wireless local area networks such as 802.11 also use cellular organization, but cells are much smaller, usually up to hundreds of meters [31]. However, the usefulness of this approach also depends on the area that is covered by the base technology, *i.e.* the geographical area where any of the cells can be heard and thus location determined.

In the case of GSM, the Cell Identification (CI) method relies on the fact that a cell phone is constantly aware of the cell ID it is currently using. The size of cells is usually smaller (up to tens of meters) in urban areas and much larger (up to tens of kilometers) in rural environments [26]. CI's accuracy can be improved by TA (Timing Advance) [36]. TA is a delay time used to adjust the transmission timing to the propagation delay between cell phone and cell station that are farther away. In practice, TA is a discrete parameter; each unit of which represents about 500 meters.

1.2.2 Lateration

Lateration-based techniques extend basic Cell Identification by taking advantage of the fact that in cellular systems, coverage areas of cells usually overlap and mobile station can hear many cells simultaneously. Knowing that the station is located in the intersection of the areas of multiple cells increases the accuracy of localization. In addition to that, lateration methods try to estimate the angle or distance [35] between the mobile station and cells, increasing the accuracy even more.

The angle can be measured by cells with directional antennas that detect the direction of the signal transmitted by the mobile station. If at least two cells detect the angle, the intersection of the lines formed by the angles identifies the two-dimensional location of the mobile station. This method is also referred to as Angle of Arrival (AOA) [35].

The distance between the cell and mobile station can be estimated by measuring the time it takes the signal to travel between them or the amount of signal attenuation along the way.. Radio signals travel at the speed of light, so by knowing the time, the distance from the cell could be easily calculated. Knowing distances between the mobile station and at least three cell stations, the actual location can be calculated. Each distance forms a circle around a cell. The intersection of three circles is the position of the mobile station. Popular methods based on this approach are Time of Arrival (TOA) [35], Time Difference of Arrival (TDOA) [35], Enhanced Observed Time Difference (E-OTD) [36]. In case of TOA, the distance is derived from the absolute time for a radio signal to travel. TOA, however requires that the receiver knows the exact time of transmission. To overcome this requirement, round-trip time can be measured instead. TDOA measures time difference of the same signal at different cells. In E-OTD, cells broadcast messages to mobile stations, which then compare the relative times of arrival to estimate its distance from stations. [36]

Signal strength of the radio waves in vacuum decreases as the inverse of the squared distance (d^{-2}) [26]. Using this relation, the received signal strength and the transmission power of the cell, the distance from it can be estimated. Similarly to TOA, the location of the mobile station can be calculated if the distances from at least three cells are known.

Unfortunately, these methods are not very accurate in real life. In reality, radio signals are corrupted by unwanted random effects such as noise, interference from other sources, and interference between different radio channels. Signal propagation indoors is even more complex. Indoor environments cause harsh multi-path effects, interference and dead-spots [21]. Thus, these methods work ideally only in line-of-sight conditions, where no obstructions are on the way of the signal, which is rarely the case indoors. They do not take into account complicated radio signal propagation and therefore lack the accuracy required for indoor positioning.

1.2.3 Fingerprinting

One approach to overcome the problem of signal propagation peculiarities is to teach them to the system. The varying signal strengths, propagation times or angles can be measured at different known locations and recorded. When a new point needs to be localized, these quantities can be compared to the ones encountered before. The new location can then be assumed to be close to the previously collected points that have similar signal characteristics. This technique is called *fingerprinting* and the collected signal characteristics are called *fingerprints*, due to similarity to the fingerprint comparison in forensics. Therefore, to localize a mobile device using fingerprinting, the current signal fingerprint has to be compared to the fingerprints collected during a training period whose locations are known.

Two factors account for the good performance of radio fingerprinting. The first is that the signal characteristics observed by mobile devices exhibit considerable spatial variability. For example, a given radio source may be heard stronger or not at all a few meters away. The second factor is that these characteristics are consistent in time; for example a medium-weak signal from a given source at a given location is likely to be similar tomorrow and next week. In combination, this means that there is a radio profile that is feature-rich in space and reasonably consistent in time. Fingerprinting-based location techniques take advantage of this by capturing this radio profile for later reference.

The advantage of a fingerprinting based localization system is that it allows determining the location very accurately as all the signal propagation oddities can be taken into account. However, the more details are learned, the more vulnerable is this radio map to changes in the environment, such as moving furniture, construction of new buildings, weather conditions or even people and cars moving inside or outside the buildings. Therefore, this approach requires recalibration time after time to adapt to the changes in the environment. However, the parts of the environment that affect the signal propagation the most (buildings, walls) are usually stable, so recalibration is not needed often.

Fingerprinting approach can be used with different technologies (*e.g.*, GSM, 802.11), and with different types of input data. Most common is to use signal strength measurements, times or angles of arrival, or combinations of these. Another important part of fingerprinting based localization method is the *predictive algorithm*. The role of this algorithm is to calculate the locations of new points by building a generalizing model that matches the training samples, but more importantly, is able to predict the location of the yet unseen samples with high accuracy. In other words, determining the location if the fingerprint is identical to one of the training points is trivial; the algorithm has to be able to estimate the location in all the other cases as well, for example if the user is in between the training locations. Possible predictive algorithms include *k*-Nearest Neighbors, Support Vector Machines, Neural Networks, or other machine learning algorithms for supervised learning [25, 13, 5].

1.3 Wireless Technologies

This Thesis considers signal strength fingerprinting-based indoor localization systems that use two wireless technologies: GSM cellular phone system [26] and 802.11 (Wi-Fi) wireless LAN [31]. This section provides an overview of GSM and 802.11 emphasizing those aspects that are most relevant for building indoor localization systems.

1.3.1 GSM Cellular System

In the 1980s several analog systems for mobile communications were in use, such as AMPS in the United States, TACS in Britain and NMT in Northern Europe [41]. However, the need for a system that allowed roaming between countries was quickly recognized. Soon a standardization organization was created to develop a common standard – GSM (Global System for Mobile communications). GSM is fully digital system; it supports both speech and data services, and allows smooth roaming across wireless carriers and countries. [41]

GSM is the most widespread cellular telephony standard in the world, with deployments in 210 countries by 676 network operators[2] in the end of 2004. Last year, the number of subscribers was growing most rapidly in Latin America (more than 150%), Russia, North America[3] and India (all more than 70%). Asia Pacific region together with China is steadily becoming the largest GSM market. There were 1.6-billion GSM subscribers worldwide by the end of 2004, accounting for close to 80% of all the cellular subscribers. [43]

[2] Excluding China and Chinese operators, which are not members of the GSM Association.

[3] Here, North America includes United States and Canada, but not Mexico, which belongs to Latin American subdivision.

GSM is the only cell phone standard in Europe and many other regions. North American market is dominated by CDMA, the next popular technology worldwide. Only about 30% of the North American subscribers were using GSM in the end of 2004, but the annual growth of GSM subscribers was bigger than the one of CDMA subscribers [43], which suggests growing importance of GSM in North America as well.

The GSM network architecture can be divided into three general parts. The Mobile Station (MS) is the device (cell phone) carried by the user. The Base Station (BS) hosts *cells* and handles the radio link with MS's. The Network Subsystem (NS) switches calls between mobile users, and between mobile and fixed network. [33, 26]

1.3.1.1 Radio Resource Use

The radio interface of GSM uses a combination of Frequency Division Multiple Access (FDMA), Time Division Multiple Access (TDMA), and frequency hopping. The FDMA part divides the bandwidth by frequency into *Radio Frequency Channels* (RFCH) spaced 200 kHz apart. Each of these *carrier frequencies* is then divided into eight *timeslots* or *bursts* using TDMA. A timeslot lasts 0.577 ms and occupies a 200 kHz slice of bandwidth. The slots numbered from timeslot 0 to 7 form a TDMA frame with length 4.615 ms. The recurrence of one of the eight timeslots in each frame makes up one *physical channel*. [26, 33, 30]

Different radio frequencies are used for GSM networks around the globe. In Europe and most of the world, 900 MHz + 1800 MHz bands are used. In North America and some countries in Latin America and Caribbean, GSM is using 850 MHz + 1900 MHz bands. In several popular tourist destinations in Caribbean, all four bands are supported to make it easier for international travelers to use their cell phones. Different ranges are allocated for uplink (MS to BS) and downlink (BS to MS) communication (Table 1). In North America there are 124 bi-directional RFCHs in the 850 MHz band and 299 in the 1900 MHz band, totaling 423 channels. In Europe the total number of channels is 548. [46] In this Thesis, all the experiments are done in North American bands. However, as the frequencies are similar, results should be analogous in other bands as well.

Band	Frequencies used for RFCHs		Numbers
900 MHz	up	880.2, 880.4, …, 914.8 MHz	975..1023, 0..124
	down	925.2, 925.4, …, 959.8 MHz	
1800 MHz	up	1710.2, 1710.4, …, 1784.8 MHz	512..885
	down	1805.2, 1805.4, …, 1879.8 MHz	
850 MHz	up	824.2, 824.4, …, 848.8 MHz	128..251
	down	869.2, 869.4, …, 893.8 MHz	
1900 MHz	up	1850.2, 1850.4, …, 1909.8 MHz	512..810
	down	1930.2, 1930.4, …, 1989.8 MHz	

Table 1. Four frequency bands used for GSM, up- and downlink frequencies, and channel numbering

The GSM radio interface uses *slow frequency hopping*, changing the transmission frequency at regular intervals. The frequency is changed between bursts so that the whole burst is transmitted using the same frequency. Frequency hopping sequence is broadcast to all the MS's through *control channels*. [26]

The transmission power can be reduced to minimize the energy use and decrease interference, whilst maintaining the quality of the radio links. According to specifications, power control must be implemented in MS side, but is optional in the BS. BS can reduce its RF output power at most 30 dB from its maximum level. [47]

In each physical channel defined above, many *logical channels* can be transmitted, dividing physical channels further in time. GSM defines a variety of traffic and signaling/control logical channels of different bit rates. There are speech and data traffic channels, different control and synchronization channels, etc. [30]

1.3.1.2 BCCH Carrier

Particularly important in the context of this Thesis is the *broadcast control channel* (BCCH), which is used in the BS to MS direction to broadcast system information such as the synchronization parameters, available services, and cell ID. Each cell is allocated a subset of RFCH channels, defined as the *cell allocation* (CA). One radio frequency channel of the CA is used to carry BCCH (among other channels). This is called *BCCH carrier* (a.k.a. *beacon frequency* or C0, as it is the first frequency channel in a cell allocation). In this channel, no frequency hopping is permitted on the first timeslot carrying BCCH. Although the BCCH information is only transmitted on the first timeslot, all other timeslots of the BCCH carrier should also be continuously transmitted without variation of RF level. If there is no information to send on a timeslot, BS must transmit a dummy burst. This enables MS to measure the received signal level and estimate the potential for handover to surrounding cells by simply tuning to their BCCH carriers. As BCCH carriers are constantly transmitted, MS can listen to them whenever it can, without waiting for the particular timeslot. [45, 47, 26] In this Thesis, we also measure signal strengths on multiple BCCH channels and use this information to infer user's location.

When the MS is switched on and doesn't know which channels are BCCH carriers, it goes through all the channels within its bands of operation and searches for BCCH carriers. Once it has found a BCCH carrier, it can read all the channel numbers of other BCCH carriers near by. To achieve smooth handover and operation, MS measures signal strengths of 16 BCCH channels, but synchronizes to and reads the BCCH information from the 6 strongest ones. [47, 44]

The RFCHs allocated to a cell (including C0) may change dynamically in time, although this happens very rarely. Frequencies can be changed to install new hardware or remove some for maintenance, or due to unplanned interference. [26] A change in one cell must usually be coupled with changes to other cells in order to retain non-interference. The change is broadcast to all MSs in range together with the exact time (timeslot number) the change will occur.

1.3.2 802.11 Wireless Networks

In 1997, the IEEE approved 802.11 standard [48] that uses 2.4 GHz band to provide wireless networking at a maximum rate of 1-2 Mbps. In 1999, the 802.11b High Rate amendment was approved, which increased the maximum rate to 11 Mbps. The 802.11b (a.k.a. Wi-Fi or Wireless Ethernet) is now the most popular wireless LAN standard in the world. [31] However, many new improvements have been developed, for example 802.11a and 802.11g, which increase the rate even further. In the following, term "802.11" is used to refer to the IEEE 802.11b networks.

The 802.11 network consist of several mobile nodes and an access point (AP), which usually provides the connection to the Internet. The nodes communicate wirelessly with the AP and to each other.

1.3.2.1 Radio Resource Use

Wi-Fi networks operate in 2.4 GHz ISM (Industrial, Scientific and Medical) band, which is reserved for unlicensed use in most of the countries in the world. It means that anybody using equipment that complies with the technical requirements can send and receive radio signals on these frequencies without a license. One of the allowed uses of this band is spread-spectrum wireless data networks, like 802.11. The exact frequency allocations are slightly different from one part of the world to another. [31]

There are 14 possible carrier frequencies, different subsets of these in use in different countries. Table 2 shows frequencies allowed in United States, France, rest of the Europe and Japan. Most of the other countries use one of these four subsets (for example Canada uses the same channels as the U.S.).

Frequency	U.S.	Europe	Japan	France
2412 MHz	1	1	1	
2417 MHz	2	2	2	
2422 MHz	3	3	3	
2427 MHz	4	4	4	
2432 MHz	5	5	5	
2437 MHz	6	6	6	
2442 MHz	7	7	7	
2447 MHz	8	8	8	
2452 MHz	9	9	9	
2457 MHz	10	10	10	10
2462 MHz	11	11	11	11
2467 MHz		12	12	12
2472 MHz		13	13	13
2484 MHz			14	

Table 2. Radio frequency channels and channel numbers used for 802.11 networks [31]

All of these frequencies, most of them only 5 MHz apart from each other, are actually center frequencies of a 22 MHz channel. Therefore, each channel overlaps several others above and below it. The whole 2.4 GHz band provides only three completely separate channels – 1, 6 and 11. Different countries also put different limits on the allowed transmission power.

Each AP is assigned a single channel, which is used both for uplink and downlink communication with the nodes. Collisions and conflicts are avoided by using CSMA/CA (Carrier Sense Multiple Access with Collision Avoidance). [31]

Spread spectrum technology uses wide channels, which makes it in theory less sensitive to interference from other radio signals and electrical noise. However, as the ISM band is unlicensed, many different devices, like microwave ovens, Bluetooth equipment and cordless phones, are using the band without synchronization, and interference with them has turned out to be a difficult problem [17, 10].

2 Related Work

This Thesis examines the effectiveness of GSM signal strength fingerprinting as an indoor localization technique. While this combination is new, indoor localization, radio fingerprinting and use of GSM for localization have all been explored before. We describe these efforts and key distinctions between these efforts and our.

2.1 Indoor Localization

Indoor location systems have been successfully built using a variety of technologies. The Active Badge [37], PARCTab [38] and follow on commercial systems like Versus [55] used infrared emitters and detectors to achieve 5-10 meter accuracy. Both the Cricket [29] and Active Bat [39] used ultrasonic signals to estimate location. Depending on the density of infrastructure and degree of calibration, ultrasonic systems have accuracies from a few meters to a few centimeters. Radio Frequency ID (RFID) technology has been used in research systems, such as SpotON [14] and Landmark [27], and commercial solutions like PinPoint [40, 51] to perform three-dimensional localization using signal strength measurements. Most recently, ultra-wideband emitters and receivers have been used to achieve highly accurate indoor localization [54]. The common drawback of all of these systems is that they require custom infrastructure for every area in which localization is to be performed. Bluetooth based systems like [1] could use existing Bluetooth network, but the specific uses of this technology do not cause large indoor areas to be covered with signals from fixed Bluetooth devices. Thus, additional equipment still needs to be installed to make Bluetooth localization work. As a result, all these systems have not seen significant deployment outside of high-value applications like hospital process management. In contrast, GSM fingerprinting makes use of the existing GSM infrastructure, obviating the need for infrastructure investment and greatly increasing the possible area in which the system will work. This increases the likelihood of GSM fingerprinting achieving wider adoption.

2.1.1 Active Badge

Want *et al.* proposed the Active Badge [37] localization system in 1992. Their solution to the problem of automatically determining the location of an individual was to design a wearable tag that emits a unique code every 15 seconds. These signals were then picked up by sensors around the building. A master station polled the sensors for badge “sightings” and processed the data, making it available for client applications.

Infrared (IR) signals were used for signaling between the badge and sensor. The emitted signals operated in approximately 6 meters range, and didn’t travel through the walls. Thus, sensors needed to be installed at least in every room, more than one to bigger or more complicated rooms. People had to wear special badges. Because the signals were transmitted through an optical path, the badges had to be worn outside of the clothing, preferably clipped to a shirt or a blouse. Sensors needed to be placed high up on walls or ceiling tiles of offices and on the entrances and exits of corridors and other public areas. The total cost of sensors, badges, cabling and installation was high, especially when large buildings had to be covered. It was not expected that sensors had overlapping coverage areas. Even if multiple sensors received signals from the same badge, this information was not used to increase accuracy. No other characteristics of the signal (like signal strength or time-of-arrival) was used to pinpoint the exact location inside the room.

2.1.2 Cricket

Cricket [29], developed in MIT in 2000, is a location-support system for indoor location aware applications. It allows mobile and static nodes to learn their physical location by using listeners that hear and analyze information from beacons located throughout the building. On the contrary to the Active Badge, devices carried by users infer the location, not the central server. Thus the device controls the location information and can determine to whom it actually publishes it and to what extent. This alleviates privacy concerns, but on the other hand, makes it impossible to create applications that require guaranteed location (*e.g.*, location based billing). In that sense, Cricket is similar to our approach.

Cricket uses beacons that send location information to listeners. A beacon is a small device attached to some location within the geographic space it advertises. It is typically placed at an unobtrusive location like a ceiling or wall. The message sent out by beacon is a plain text string that identifies the location.

To obtain location information, every device has a listener attached to it. Devices measure the one-way propagation time of the ultrasonic signals emitted by a beacon. A beacon sends information over radio frequency, together with an ultrasonic pulse. When the listener hears the RF signal, it turns on its ultrasonic receiver and listens for the ultrasonic pulse, which will arrive a bit later, because the speed of sound is lower than the speed of light. The listener uses the time difference between the receipt of the RF signal and the ultrasonic signal to calculate the distance to the beacon. This is done to determine the nearest beacon, whose location information is then taken as an estimate to user's own location.

2.2 Indoor Localization Using 802.11 Fingerprinting

Bahl *et al.* observed that the strength of the signal from an 802.11 access point does not vary significantly in a given location. They used this observation to build RADAR [4], a system that performed localization based on which access points would be heard where, and how strongly. This was the first 802.11 *fingerprinting* system, and in the hallways of a small office building, fingerprints from three access points could localize a laptop within 2-3 meters of its true location. There have been improvements to RADAR's fingerprint matching algorithm that have advanced accuracy [3, 21, 42, 5, 19, 34, 32, 28] and differentiated floors of a building with a high degree of precision [12]. In addition, commercial localization products have been built using 802.11 fingerprinting [51].

The differences between our work and 802.11 fingerprinting systems are primarily due to the differences between 802.11 and GSM:

- Due to higher coverage, GSM fingerprinting works in more places than 802.11 fingerprinting.
- Due to more stable infrastructure, 802.11 radio maps will degrade more quickly than GSM radio maps.
- Due to the larger range of GSM cells, 802.11 fingerprinting will be more accurate than GSM fingerprinting given the same number of radio sources.

2.2.1 RADAR

RADAR [4] was the first attempt to use fingerprinting and an existing 802.11 infrastructure for localization. Instead of utilizing special equipment like infrared or ultrasound sensors and badges, RADAR used wireless networks already deployed in the building to localize hosts. The hosts periodically broadcast packets to the network and access points measured the signal strengths of these packets. The collected signal strength data was used to train the system and to determine later the location of a mobile host. RADAR describes two solutions: empirical and signal propagation modeling.

Empirical approach was based on *fingerprinting*, similarly to the solution described in this Thesis. The training phase consisted of measuring signal strengths in multiple locations a few meters apart across the floor of a building. To determine the actual location, the measured signal strengths were compared to the ones measured during testing phase. The closest one or a number of closest ones were used to estimate the location of the predicted point. Nearest neighbors in signal space (NNSS) algorithm was used to find closest matches. No weighting was done to give higher weights to nearest neighbors.

The main limitation of the empirical method is that significant effort is needed to construct the data set for each physical environment. Furthermore, the data collection process may need to be redone if the network changes, *e.g.*, when a base station is relocated. The purpose of the second approach, radio propagation model, was to decrease the amount of time required to take the measurements in the building. A simple Wall Attenuation Factor method was used to estimate the signal strengths in the building as a function of distance and the number of walls in the path from the access point to the mobile host. However, the exact map of walls was required, as well as a few measurements to determine the actual attenuation caused by each wall. The reported results were worse than the ones of the empirical method. In our case, modeling radio propagation would be much more complicated, because we would also need to take account other near-by buildings as GSM radio transmitters are located outside.

2.2.2 Improvements to 802.11 Fingerprinting

In their subsequent report [3] Bahl *et al.* proposed a number of enhancements to the RADAR system. Specifically, they describe a Viterbi-like algorithm for continuous user tracking. This algorithm takes into account the mobility pattern of the user to disambiguate between candidate user locations guessed by the system. They also describe access point-based environmental profiling scheme, where they automatically switch between two sets of fingerprints, taken in different environmental conditions (busy hour, non-busy hour).

Ladd *et al.* [21] have increased the accuracy of 802.11 fingerprinting by applying standard approaches from robotics-based localization, notably the explicit manipulation of noise distributions and the modeling of position as a probability distribution.

Battiti *et al.* [5] have used other statistical learning methods besides simple *k*-Nearest Neighbors. The experiments with Weighted *k*-Nearest Neighbors, Support Vector Machines, Neural Networks, Bayesian Nets and others resulted in slightly better accuracy.

Elnahrawy *et al.* [8] proposed interpolation technique to decrease the amount of time required to take measurements without losing too much accuracy. They used Interpolated Map Grid (IMG) to create additional training points between the existing ones. In addition to that, they described three new area-based localization methods, where the predicted location is an area not a single point.

Most of these ideas can also be useful and applicable to GSM fingerprinting. This is however left for future work.

2.3 Localization Using GSM Fingerprinting

The Place Lab system employed a map built using war-driving software and a simple radio model to estimate cell phone's location with 100-150 meter accuracy in a city environment [23]. Laitinen *et al.* [22] used GSM-based fingerprinting for outdoor localization. They collected sparse fingerprints from the 6 strongest cells, achieving 67th percentile accuracy of 44 meters. Laasonen *et al.* used the transition between GSM cells to build a graph representing the places user goes [20]. Like Place Lab, Laasonen's system used cell phones that only exported the single cell-tower the phone was associated with. In contrast to the other systems we have mentioned, Laasonen's system did not attempt to estimate absolute location, but rather assigned locations symbolic names like *Home* and *Grocery Store*.

These previous efforts to use GSM for localization differ from the work reported in this Thesis in that they are based on sparse fingerprints collected tens to hundreds of meters apart from each other. Moreover, these efforts used *narrow* fingerprints obtained from commercial GSM phones that report the signal strength for the current cell [23, 20] or the 6 strongest cells [22]. In contrast, we collected GSM fingerprints in a dense grid with 1.5 meters granularity. Moreover, in addition to the 6 strongest GSM cells, we collected *wide* fingerprints that include up to 32 different GSM channels. This additional information has helped to significantly increase the accuracy of our system, as we show in the following chapters.

2.3.1 Place Lab

Place Lab [23] provides wide scale localization by listening for the transmissions of wireless networking sources like 802.11 access points, fixed Bluetooth devices, and GSM cell towers. However, instead of relying on extensive training phase, they use a public database of measurements collected by people in volunteering basis. Many of these beacon databases can come from institutions that own a large number of wireless networking beacons. Companies, universities and departments often know the locations of their 802.11 access points since this information is commonly recorded as part of a deployment and maintenance strategy. Other sources of Place Lab mapping data are the large databases produced by the *war-driving* community. War-driving is the act of driving around with a mobile computer equipped with a GPS device and a radio (typically an 802.11 card but sometimes a GSM phone or Bluetooth device) in order to collect a trace of network availability.

They have used three methods for location calculation. A simple Centroid calculates the average coordinates of the beacons in range and uses this as estimation. Fingerprinting method takes also signal strength information into account. More complicated Bayesian Particle Filter method uses the information about previous locations of the user to pinpoint the location.

The goal of Place Lab was to provide coarse-grained accuracy with minimal mapping effort. This is different, and complementary to our goal of doing accurate indoor localization given a detailed radio survey. Another distinction is that Place Lab used a cell phone platform that only programmatically exported the single associated cell tower.

2.3.2 Database Correlation Method

Laitinen *et al.* have proposed Database Correlation Method (DCM) [22], which uses sparse GSM fingerprints to do localization outdoors. DCM is based on adjusted 1-Nearest Neighbor algorithm to find the best matching fingerprint from the collected fingerprint database. The location of that fingerprint is then used as a resulting prediction. They report 44m accuracy in 67% of times in urban environments and 90m accuracy 90% of the times. The location calculation method they use is a very simple version of ours. We use more fingerprints to average location, weight them based on the distance in the signal space and use clustering to eliminate outliers; techniques which all have improved indoors accuracy considerably. In addition to that, we use *wide* fingerprints that are taken with higher granularity.

2.4 Indoor Localization and Global Positioning System

Additionally, many cell phone manufacturers have integrated GPS [9] units into the phones. A technique called Assisted GPS (A-GPS) [7] is used to shorten the time it takes for MS to localize themselves. Although accurate outdoors, these solutions are not very useful indoors or in "urban canyons," because of the lack of line of sight (LoS) between phone and multiple satellites. Indoor GPS [52] installs expensive GPS repeaters inside buildings to make the GPS devices work. However, the technique is still based on trilateration, which does not consider complicated signal propagation inside buildings, and thus requires large empty rooms or huge number of repeaters to provide high accuracy.

3 Methodology

This chapter first gives an overview of signal strength fingerprinting, and the predictive algorithm we use in this Thesis. Then we describe the data collection process and the localization methods that we compare in our evaluation.

3.1 Signal Strength Fingerprinting

In our research, we use signal strength data measured in different GSM radio channels. As a comparison, we also use signal strength fingerprints from 802.11 wireless networks. Our initial assumption was that the signal strengths in GSM channels are relatively stable in time, but vary location by location, so that localizing mobile station with high accuracy is possible. To evaluate this assumption, we compared the stability of GSM and 802.11 signals. We recorded signal strengths of several 802.11 access points (AP) and GSM cells at a few fixed locations in a University building in downtown Toronto over a period of several days. In the reminder of this Thesis, we will refer to this building as University.

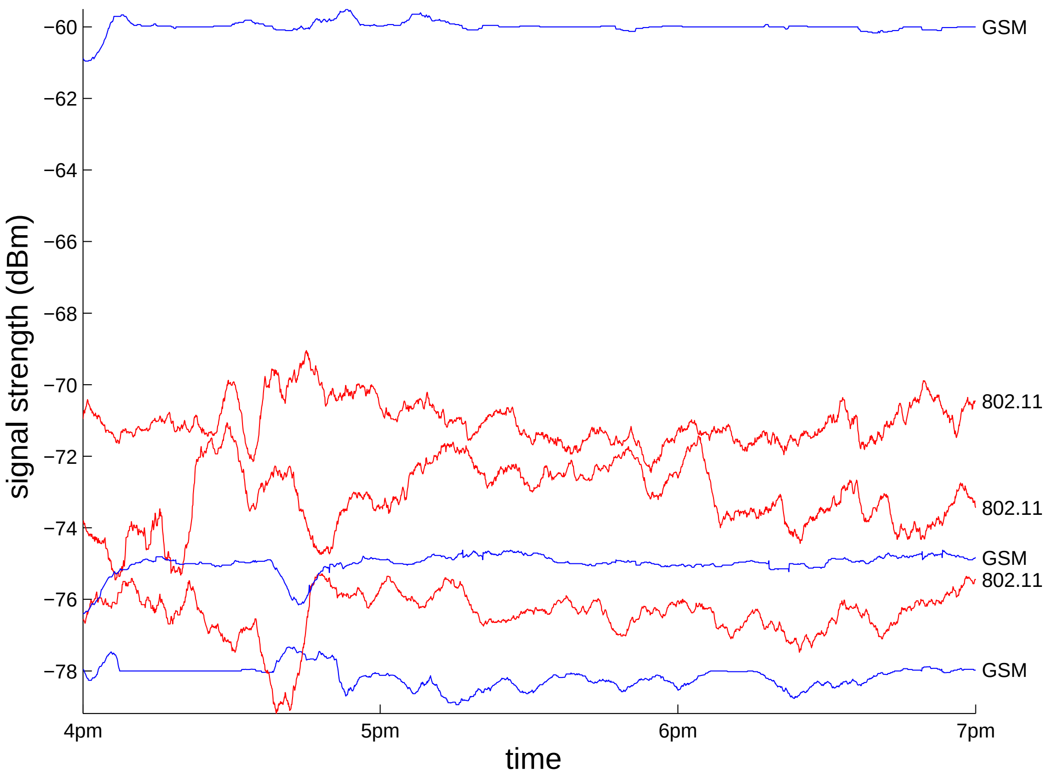

Figure 1. Temporal 802.11 and GSM signal stability

Figure 1 shows a 3-hour segment of the signal strength measurements at a fixed location on the 5th floor of the building during a workday afternoon. The plot shows signals from three strongest GSM cells and the three strongest 802.11 APs. GSM signals appear to be more stable than 802.11 signals. We believe that one reason for this is that 802.11 uses unlicensed overcrowded 2.4 GHz band, and therefore its signal strength suffers from interference from nearby appliances such as microwaves and cordless phones. An analysis of GSM signal stability under different weather conditions (*e.g.*, rain, snow, fog) is left for future work.

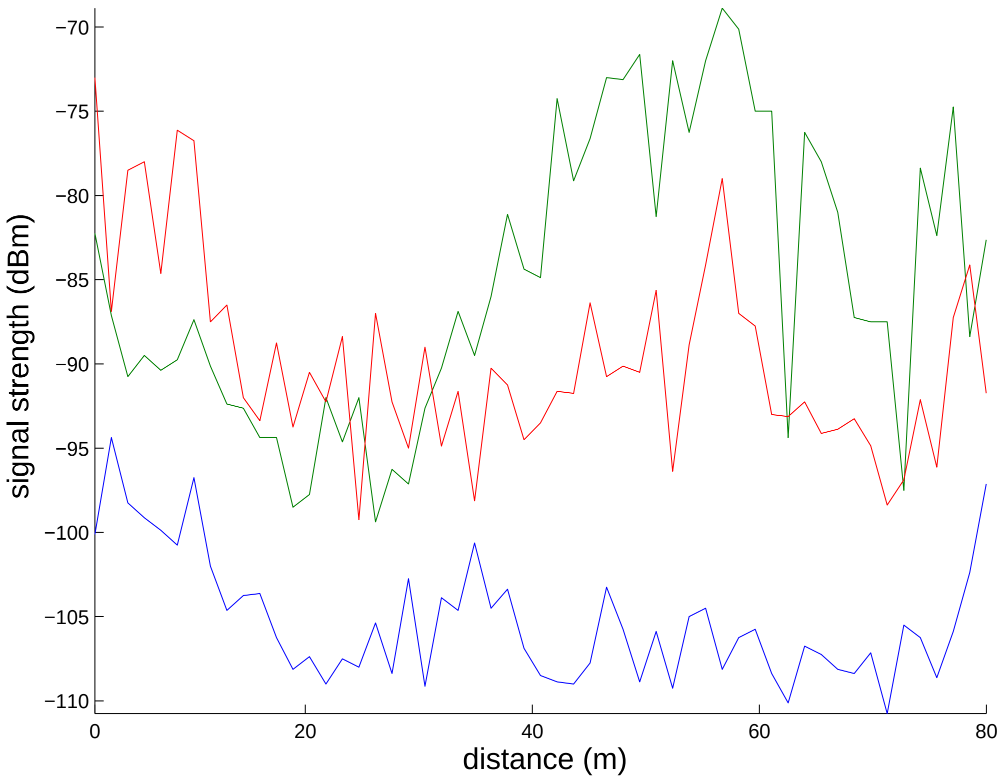

Figure 2. Signal strength of three GSM cells while walking through the University building

Figure 2 shows the changing signal strengths while taking a walk from one end to the other on the 5th floor of the University building. Measurements are taken about every 1.5 meters. It can be seen that the signal strengths change considerably and different locations have different patterns, which suggests that it may be possible to deduct the location of the mobile device from signal strength data.

Signal strength fingerprinting relies on a "training phase" in which a mobile device moves through the environment recording the strength of signals emanating from a group of radio sources (*e.g.*, 802.11 access points, GSM base stations, FM radio or TV stations). We refer to the physical position where the measurement is performed as a *location*, to the whole radio scan as a *measurement* and to the recording of the signal strength of a single source as a *reading*. That is, to build a radio map of the building, a mobile device takes a series of measurements in multiple locations of the building. Each measurement is composed of several readings; one for each radio source in range. The set of data recorded in a single location is also referred to as a *training point*. Since signal strengths have considerable spatial variability, a fairly dense collection of locations need to be collected to achieve good accuracy. The original RADAR experiments, for example, measured every square meter on average [4]. To achieve their advertised accuracy, the commercial 802.11 fingerprinting product from Ekahau [51] recommends similar density. Once the training phase is completed, the locations of new fingerprints (also referred to as *testing points*) can be calculated using the *predictive algorithm*.

3.1.1 Predictive Algorithm

A simple technique for estimating location is to choose the location of the training point with the closest Euclidean distance in a signal strength space. The Euclidean distance d can be calculated according to Equation 1, where $s_1 \ldots s_n$ are the signal strengths of n radio sources of the testing point and $s'_1 \ldots s'_n$ are the corresponding signal strengths of the training point.

$$d = \sqrt{\sum_{i=1}^{n} (s_i - s'_i)^2}$$

Equation 1. Euclidean distance in signal strength space

Better accuracy can be achieved by averaging the location of the k closest neighbors (or training points) in the radio map, where k is some small constant. It is also beneficial to use weighted averaging, so that neighbors closer in signal space are given higher weights. This method is further referred to as Weighted k-Nearest Neighbors (WKNN).

We calculated weights according to Equation 2, where w_i is the weight of i-th neighbor and d_i is the distance of that neighbor in signal space. Weighting factor b determined the amount of weighting being done. If $b = 0$, then there is no weighting and all the neighbors are given equal weights. The ideal value for b depends on the average distances d_i, which depends on the dimensionality n.

$$w_i = \frac{e^{-b \cdot d_i}}{\sum_{j=1}^{k} e^{-b \cdot d_j}}$$

Equation 2. Weight calculation using distances and weighting factor *b*

We use WKNN both for estimating the floor *(z)* and the location *(x, y)* on that floor. The first calculation is called *classification*, as the estimated value has final number of possible values (number of floors in the building). The latter calculation is called *regression*, as the coordinates on the floor have more continuous nature.

In case of regression, the continuous value is estimated according to Equation 3, where x_i are values of the training points and w_i the corresponding weights.

$$x = \sum_{i=1}^{k} w_i x_i$$

Equation 3. Regression using WKNN

In case of floor classification, the estimated floor is found according to Equation 4, where z_i are values of the training points, w_i the corresponding weights, and δ is a function, that returns 1 if the arguments are equal, and 0 if not.

$$z = \arg\max_{c} \sum_{i=1}^{k} w_i \delta(z_i, c)$$

Equation 4. Classification using WKNN

Our initial evaluation uncovered cases in which the algorithm selected points that are neighbors in the signal space, but are actually located far away from the true location of the testing point in the physical space. Often just a few of them lied away from the others. To ameliorate the effect of these false positives, we used simple *K*-mean clustering [16] in physical space. The *K*-mean clustering algorithm works in the following way:

1. Set randomly *K* initial locations, which are called *means*
2. Assign each of the *k* neighbors to the mean that is closest to it
3. Recalculate means to be the average values of the assigned locations
4. Repeat (go back to step 2) until the assignments don't change

We used *K*-mean clustering to split the set of nearest neighbors into two geographical clusters (*i.e.*, setting *K* equal to two[4]). We then compared the sizes of the clusters, and if one of the clusters was considerably larger than the other, we removed the points that belonged to the smaller cluster from the final location calculation.

In this Thesis, WKNN is used as the *predictive algorithm*. Similar approach has been compared with some others by Battiti *et al.* in [5] and reported as the most effective technique for spatial localization using 802.11 signal strength fingerprinting. Investigating the applicability of other predictive algorithms to GSM fingerprinting is a topic for future work.

3.2 Data Collection

We collected multi-floor measurements in two office buildings and one single-family detached house. The three buildings are located in two major North-American cities located on opposite coasts. The office buildings house part of the Computer Science Department at the University of Toronto and the Intel Research Lab in Seattle. The private house is located in Seattle as well. In the rest of this Thesis, we refer to these buildings as: University, Research Lab, and House. University is located in a busy

[4] We experimented with different vales for *K*, but 2 produced the best results.

downtown core, Research Lab is located in a commercial midtown area, and House is located in a quiet residential neighborhood.

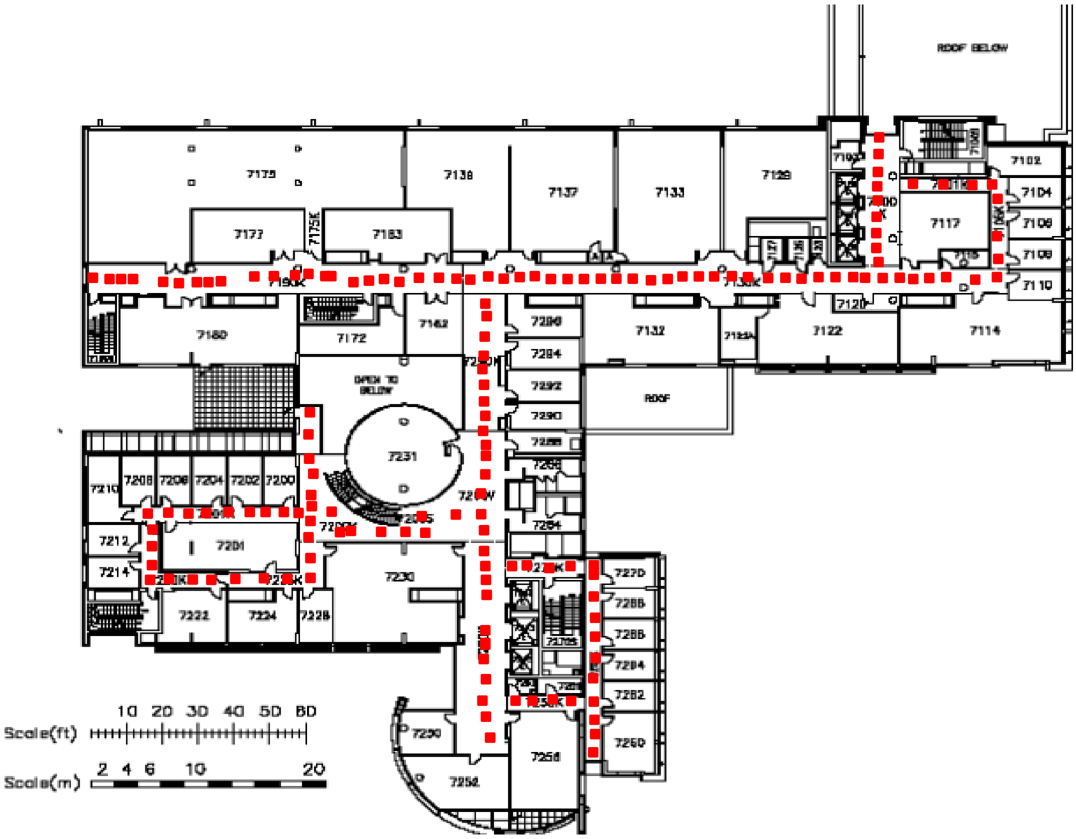

Figure 3. Map of the 7th floor of the University building with red squares as training points

University is a large *88m x 113m* 8-storey building with lecture rooms, offices and research labs. Since we had no access to the offices, we collected training points in the hallways[5] of the 5th and 7th floors of the building (Figure 3). Research Lab is a medium size (*30m x 30m*) 6-storey building. Space inside the building is partitioned with semi-permanent cubicles. Due to access restrictions, we collected readings from the whole 6th floor, but only a half of the 5th floor. House is a 3-storey wooden structure (*18m x 6m*) that includes a basement and two floors above ground. We collected measurements on all 3 floors. The distance between floors is about 6 meters for University and Research Lab, and about 3 meters for House.

[5] A localization system that should also work inside offices will in all likelihood not function properly if it is limited to relying on training points taken exclusively from hallways.

We collected 802.11 and GSM fingerprints using a laptop running Windows XP. To collect 802.11 fingerprints, we used an Orinoco Gold wireless card configured in active scanning mode, where the laptop periodically transmits probe requests and listens to probe responses from nearby 802.11 APs.

Figure 4. Sony Ericsson GM28 GSM modem

We collected GSM fingerprints using the Sony Ericsson GM28[6] GSM modem (Figure 4), which operates as an ordinary GSM cell phone, but exports a richer programming interface. The GSM modem provides two interfaces for accessing signal strength information: `cellsAPI` and `channelsAPI`. The `cellsAPI` interface reports the cell ID, signal strength, and associate channel for the *n* strongest cells. While the modem's specifications does not set a hard bound on the value of *n*, in practice in the 3 environments we measured *n* was equal to 6. The `channelsAPI` interface simultaneously provides the signal strength for up to 32 channels, 16 of which can be specified by the programmer, with up to 16 additional channels picked by the modem itself. In practice, 6 of the 32 channels typically correspond to the 6 strongest cells. Unfortunately, `channelsAPI` reports signal strength but does not report cell IDs. We speculate that the cell ID information for other than the 6 strongest cells cannot be determined because the signals of those cells are strong enough to be detected, but too weak to be used for efficient communication. In addition to that, many cells can use the same channel and signal strength measured in a single channel may in fact be a sum of the signals from all these cells, and detecting single cell ID is impossible (we refer to this as *aliasing*).

[6] Sony Ericsson GM28 works on North American 850+1900 MHz frequency bands. The exact same product for European 900+1800 MHz bands has a model number GM29.

	University (downtown)	Research Lab (midtown)	House (residential)
`cellsAPI`	-87.69	-76.74	-88.35
`channelsAPI`	-96.14	-102.19	-105.27

Table 3. The average signal strength (dBm) of the signals received from cells and channels

Table 3 shows the average signal strength returned by the `cellsAPI` and `channelsAPI` interfaces. As expected, the average signal strength reported by `cellsAPI` is significantly higher than the average reported by `channelsAPI`. Note that the average signal strength reported by the `channelsAPI` interface is close to modem's stated receiver sensitivity[7] of -102 dBm. Efficient GSM communication requires an SNR (signal to noise ratio) higher than -90 dB.

The lack of cell ID information for some channels raises the possibility of *aliasing*, *i.e.*, a situation when two or more cells transmitting simultaneously on the same channel appear to be a single radio source and therefore cannot be differentiated. In the extreme case, a fingerprinting system that relies exclusively on channel-based data may suffer from *world-wide aliasing*. Because channels are reused throughout the world, fingerprints taken in two far-away locations may produce similar fingerprints. To alleviate the aliasing problem, we combine the information returned by the `cellsAPI` and `channelsAPI` interfaces into a single fingerprint. We then restrict the set of fingerprints to which we compare a *testing point* to fingerprints that have at least one cell ID in common with the *testing point*. This practice effectively differentiates between fingerprints from our three indoor environments. However, for environments closer to each other (for example, several buildings in a campus), more precise method that considers more cells to differentiate between the buildings could be necessary.

As we show in Section 4, even in the presence of aliasing, our localization system based on wide GSM fingerprinting significantly outperforms GSM fingerprinting based on the 6 strongest cells, and is comparable to 802.11 based fingerprinting. This is be-

[7] In practice, the modem reports signal strength as low as -115 dBm.

cause our fingerprints are wide (have many readings), and therefore, in order for the aliasing to reduce accuracy, many readings in the fingerprints of distant locations need to match, which is highly unlikely in practice.

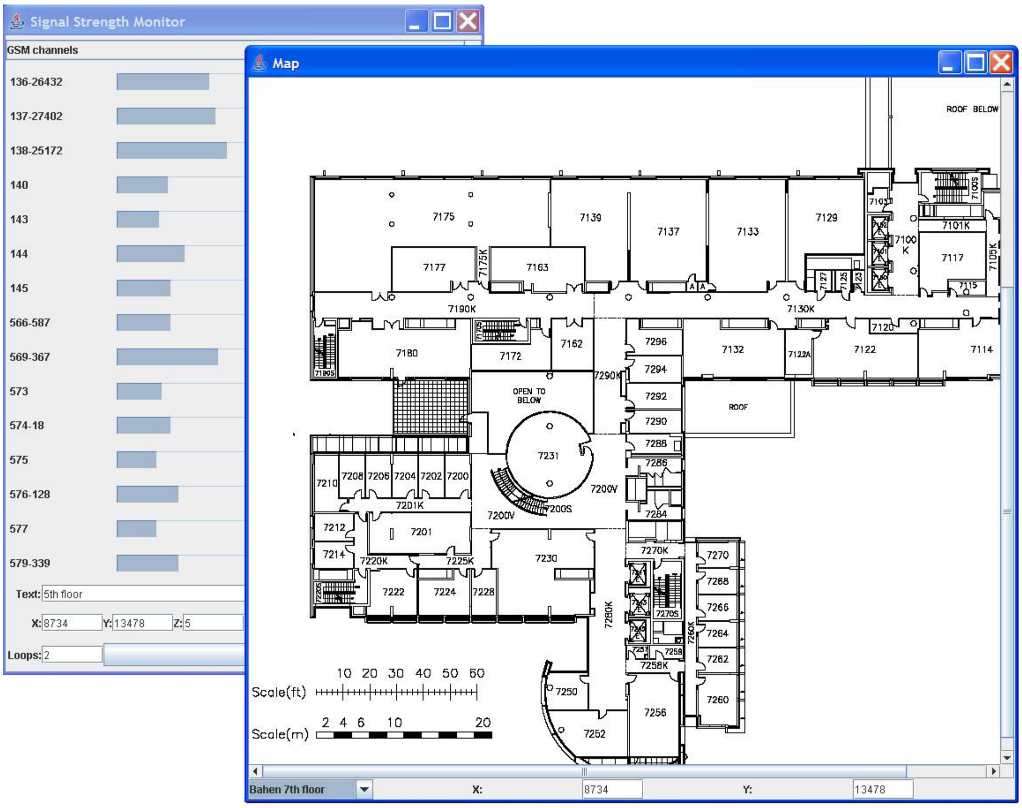

Figure 5. Application for measuring signal strengths and identifying location by clicking on the map

We developed a simple Java-based application to assist us in the process of gathering fingerprints. To record a fingerprint, we first identify the current position by clicking on a map of the building. The application then records the signal strengths reported by the 802.11 card and the `cellsAPI` and `channelsAPI` interfaces of the GSM modem. Figure 5 shows a screen shot of the Java-based application.

Figure 6. Experimental setup with laptop computer, GSM modem and antenna

To collect the measurements, we placed the laptop on an office chair and moved the chair around the building. While primitive, this setup assures measurements collected at a constant height. Figure 6 shows our experimental setup. Table 4 summarizes the number of training points collected on each of the floors of the three buildings. In all three indoor environments, we collected 802.11 and GSM fingerprints for points located 1.5 meters apart.

	University (downtown)		Research Lab (midtown)		House (residential)		
	5^{th}	7^{th}	5^{th}	6^{th}	B	1^{st}	2^{nd}
Per floor	130	154	53	181	17	44	50
Total	284		234		111		

Table 4. Training points collected on each floor for the three buildings

3.3 Localization Methods

All our localization algorithms use the weighted *k*-nearest neighbors algorithm described in Section 3.1.1. For each method, we varied the number of nearest neighbors to average over, and selected the value of *k* that gave the best results. In most cases, the best *k* was a small constant (2 or 4). The weighting factor *b* was chosen the same way and best values were in range 0 to 5.

We implemented four localization methods which differ in the structure of their fingerprints:

- `radar`, uses only readings from 802.11 access points;
- `onecell`, uses the reading of the single strongest GSM cell;
- `cell`, uses readings of the 6 strongest GSM cells;
- `ch`, uses readings from up to 32 GSM channels in addition to readings of the 6 strongest GSM cells.

To the best of our knowledge, `onecell` and `cell` are the methods that could be currently implemented using commercial cell phones [23, 20, 22]. `cell` method also provides a comparison with `ch` to show the advantage of wide fingerprinting using additional GSM channels.

An initial evaluation of `ch` revealed cases where the algorithm selected neighbors close in signal space, but far away in the physical space. To eliminate these neighbors, geographical *K*-mean clustering was used, as described in Section 3.1.1. In the rest of this Thesis, the version of `ch` that uses geographical clustering is referred as `clch`.

We also present results of a `random` algorithm that determines location by picking an arbitrary position in the particular floor or building by simply choosing one of the training samples. Therefore, `random` provides a lower bound on the performance of localization systems for a given floor and building.

3.4 Practical Considerations

We collected our wide fingerprints using a programmable Sony Ericsson GSM modem, which operates as an ordinary GSM cell phone, but exports a richer programming interface that provides access to readings from up to 32 GSM channels. In contrast, commercial phones limit access to signal strength information to the 6 strongest cells or even just the current cell. However, we speculate that the software on commercial phones could be easily enhanced to provide signal strength measurements for a richer set of channels. Once extended, those phones could take advantage of the wide-fingerprinting technique introduced in this Thesis. We base this speculation on the observation that the Sony Ericsson GSM modem is implemented using standard GSM electronics, and that the GSM specifications require phones to be able to scan all channels in the GSM band.

This Thesis presents a new localization method supported by an initial evaluation. However, many issues need to be addressed while creating a real life localization system based on this method. The following describes some of them.

While creating applications that run on mobile devices, the optimal use of scarce resources is important. Cell phones have limited computing power, memory and battery life. The network bandwidth is also limited and expensive, usually charged per kilobyte of data transferred. If some computations are done in the network side and used by many mobile users simultaneously, the network bandwidth and computing power of the server might also become bottle-necks. In case of emergency applications, such as E112/E911, the resource use is not so important, because situations when localization is needed are relatively rare and the high value created by it compensates most of the costs. In these cases, accuracy and precision are much more important. However, there are many other applications that provide less value, but use localization information more often. Especially resource consuming are applications that need to know the location of many people all the time, such as the ones that notify people when their friends or potential dates are near-by, or the ones that constantly track people indoors to help managing the workforce. These applications need the information to be constantly known by a central entity, so the network bandwidth and energy have to be used by the phone to transfer it. The localization using the predictive algorithm can be done by the mobile device and the extracted location used locally or transferred over the wireless network to a server that provides services. It can also be done by a server, where the device has to send the raw data. If the calculations are done by the mobile device, then the predictive model has to be sent to it. In case of WKNN, this means sending all the training points. For that matter, other methods such as SVM or Neural Networks can give an advantage, because their smaller models take less to transfer and store.

Another important issue is privacy [34, 15]. For some applications, it is possible to keep user's location undisclosed if the calculation is carried out in the mobile device's side. However, the problem is that in case of multiple areas and buildings, the whole model (in case of WKNN, a radio map) does not fit into the device's memory and parts of it need to be downloaded on demand. The sequence of downloading the parts reveals the location of the user to the server hosting the models [15].

As noted in [12, 23], different chipsets for different Wi-Fi devices can report different signal strength values in the same circumstances. However, as was discovered, there was a linear correlation between them. The same might be true in case of cell phones produced by different manufacturers and with different antenna sensitivities. To compensate against that, signal strengths relative to each other might be used instead of the absolute ones we have used in this Thesis. This is another interesting topic for future work.

4 Evaluation

In this section, we first analyze the collected data and then evaluate localization accuracy obtained by 802.11 and GSM fingerprinting.

4.1 Data analysis

The total number of 802.11 APs, GSM cells and channels recorded during the data collection phase is summarized in Table 5. The University building has a much denser 802.11 deployment than the Research Lab building both because the University building is much larger and because while the APs at the Research Lab building are centrally managed by IT personnel, numerous APs at the University building are owned and maintained by independent research groups.

The total number of GSM cells seen at the University building is larger than in other buildings because of the better coverage, larger building size and smaller cell size in higher density downtown area. The lower number of cells seen at the Research Lab is the consequence of both the much smaller building size and the much stronger signal received from nearby cells. Because of the proximity of a few base stations, the strongest cells reported by the modem in the Research Lab benefit from less variations than in other buildings (*i.e.*, the same group of cells appears in most of the cell measurements). The total number of channels seen in the residential area is slightly lower than in other areas due to lower coverage.

	University (downtown)	Research Lab (midtown)	House (residential)
802.11 APs	44	10	5
Cells	58	14	18
Channels	34	33	24

Table 5. The total number of different 802.11 APs, GSM cells and channels spotted in each of the areas

Figure 7 plots the cumulative distribution function (CDF) of the number of readings[8] per location of 802.11 APs, GSM cells and GSM channels at the University building. The figures showing the data for the Research Lab and the House show a very similar pattern and are therefore not presented. The number of readings for 802.11 and GSM cells is roughly the same but much lower than the number of GSM channel readings. As we will show in the next section, this has a dramatic effect on the localization performance.

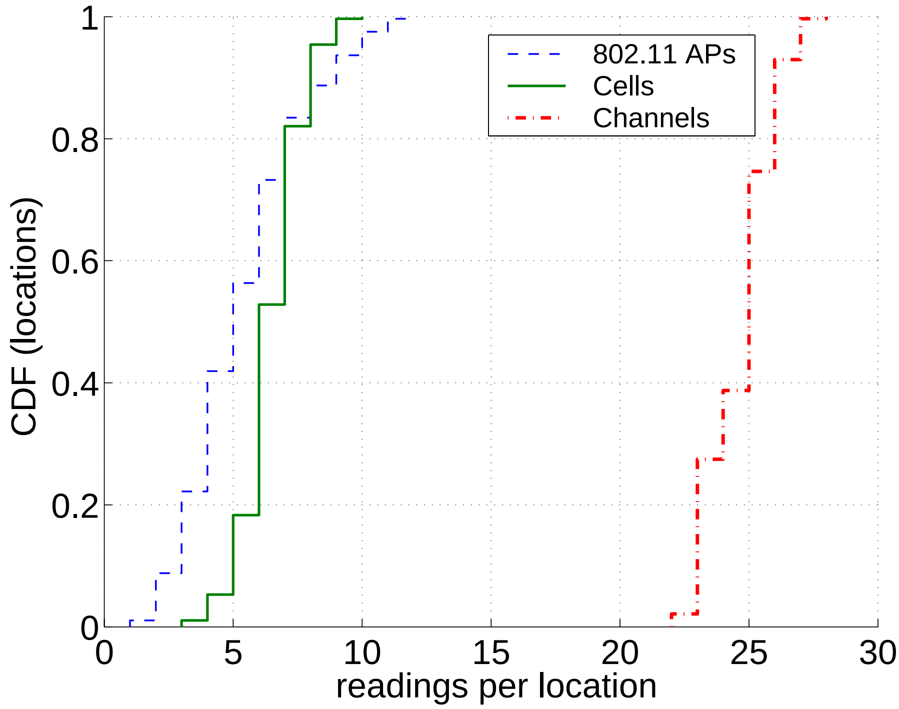

Figure 7. Readings per location of 802.11 APs, GSM cells and GSM channels in the University building

4.2 Channel Aliasing

Figure 8 shows the amount of channel aliasing seen in our three different experimental environments. There is significantly more aliasing in the University building, than in the other two buildings. Small percentage of channels was even used by five different cells in different parts of the building. This is because University building is much bigger than the others, and that it is located in a high density downtown area, where smaller cells are used. However, as shown in the following sections, aliasing does not decrease the localization accuracy significantly.

---

[8] recordings of the signal strength of a single source, e.g., 802.11 AP, GSM cell or GSM channel

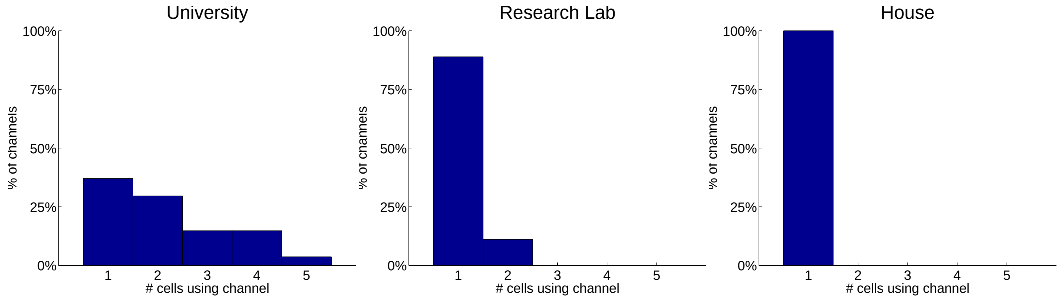

Figure 8. Channel reuse (aliasing) in different buildings

4.3 Relative performance

The results reported in this section were obtained by taking one point at a time out of the training set and using it as the testing point. This technique is similar to the one used by Bahl *et al.* [4], and is somewhat pessimistic approach since it takes the point with the best match out of the training set, and creates a hole in the set exactly at the location we are trying to estimate.

4.3.1 Floor Classification

Table 6 summarizes the effectiveness with which the localization methods introduced in Section 3.3 differentiate between floors in the three indoor environments. `clch` achieves similar performance to `ch` and is therefore not shown – grouping neighbors into floors already constitutes a form of clustering.

	University (downtown)	Research Lab (midtown)	House (residential)
`radar`	100	100	62.16
`ch`	89.08	97.01	93.69
`cell`	89.08	81.2	51.53
`onecell`	74.65	77.35	57.66
`random`	50.18	64.81	37.79

Table 6. Percentage of successful floor classifications

First, it is important to notice the different accuracy `random` method provides in our environments. Although randomly guessing the floor from 2 or 3 possible options would result in 50% or 33% accuracy, the actual random accuracy is higher due to different sizes of the floors in the building, which resulted in different number of testing points taken on each floor. The bigger the difference between floors (*i.e.*, lower entropy), the easier it is to randomly pick the right floor.

As expected, `radar` does an excellent job differentiating between floors in the University and Research Lab buildings. The reinforced concrete floors in these structures effectively block the propagation of 802.11 signals between floors, significantly simplifying the task of floor prediction. These results are consistent with previous findings [12].

In the House environment, however, `radar` achieves low classification accuracy as the house's wood structure presents little obstacle to radio propagation, making it harder to differentiate between signal fingerprints on different floors. Not surprisingly, all but 3 of the 42 misclassifications happen at locations on the first and second floors of the house. In the house scenario, 4 out of 5 of the available 802.11 signals emanate from neighboring residences. These signals propagate easily through the wooden frame of the first and second floors, but suffer significant attenuation propagating through dirt and the house's foundations to reach the basement. The low power at which neighboring access points are heard (if at all) in the basement helps to identify basement locations. On the other hand, the 802.11 signals from neighboring households contribute little to improving the accuracy of predictions for the above-ground floors.

In contrast, the GSM-based `ch` algorithm shows strong performance across all three buildings, and significantly outperforms `radar` for the House environment. It is interesting that in both the Research Lab and the House environments, `ch` achieves up to 42% better accuracy than `cell` and `onecell`. This is strong evidence that extending fingerprints to include signal strength information from channels other than the 6 strongest cells, even when the identity of the transmitter cannot be determined, can dramatically improve localization accuracy.

4.3.2 Within-Floor Localization Error

Table 7 summarizes the localization errors within specific floors for the 5 algorithms introduced in Section 3.3 for the three indoor environments. For each floor, the table shows the 50-percentile localization error, calculated as the Euclidean distance between the *actual* and *predicted* location of the point within the specific floor. All calculations assume a training set restricted to include only points that are on the same building and floor as the point whose position is being determined.

The localization error in `random` depends on the size of the covered area on each floor, which accounts for difference in its localization error across floor and building.

	University (downtown)		Research Lab (midtown)		House (residential)		
	5th	7th	5th	6th	B	1st	2nd
`radar`	4.22	4.78	2.20	2.59	3.49	3.43	3.87
`clch`	5.44	3.98	2.48	4.77	3.28	2.95	3.96
`ch`	6.47	4.07	3.40	4.82	3.28	3.36	4.55
`cell`	11.06	8.02	4.82	6.99	3.41	3.40	5.27
`onecell`	15.05	14.64	8.39	7.93	3.42	4.85	6.13
`random`	33.87	30.43	10.40	13.35	4.68	6.21	7.07

Table 7. Single-floor median localization error

Across the three buildings, `radar` achieves median accuracy between 2.2 and 4.8 meters. These results are consistent with results previously reported in the literature. Differences in accuracy between buildings reflect discrepancies in the areas of the floors, granularity of the measurement grid which varied between 1 and 1.5 meters, different number of radio sources in range etc.

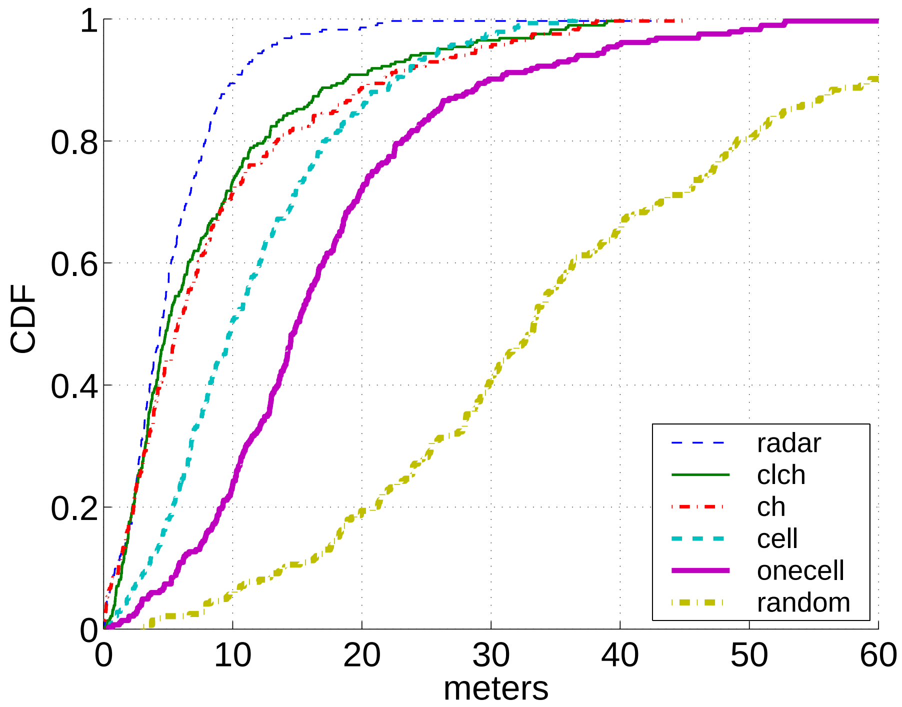

Figure 9. CDF of the localization error for 5th floor of the University building

There are large differences in the performance of the various GSM-based algorithms. `ch` and `clch` outperform `cell` and `onecell` in all cases. Moreover, `clch` achieves between 25% to 50% better performance than `cell` for at least one floor in each of the three buildings. Across the three buildings, `clch` achieves median accuracy between 2.5 and 5.4 meters, and in 3 out of the 7 floors, `clch` even achieves better accuracy than `radar` (*e.g.*, 7th floor of University building).

The strong performance of `clch` demonstrates the advantage of wide fingerprints including measurements from a large number of channels rather than just the 6 strongest cells. Moreover the significant accuracy improvement of `clch` over `ch` shows that geographical clustering manages to reduce the effect of false-positives introduced by channel aliasing. Geographical clustering, on the other hand, did not have a significant effect on the performance of `radar` as channel aliasing does not occur in this case.

Figure 9 shows the cumulative distribution function (CDF) of the localization error of all algorithms for the 5th floor of the University building. Most remarkable is the closeness with which `clch` approximates `radar`, and the large difference in performance between `clch` and `cell`.

4.3.3 Effects of Multi-Floor Fingerprints

In the previous section, we evaluated within-floor localization accuracy assuming that the training set was limited to fingerprints in the same floor, *i.e.*, we predicted the floor first, and then predicted position within that floor. In contrast, in this section, we evaluate the effects on within-floor localization accuracy of including in the training set fingerprints taken on different floors. For this purpose, we project the training points collected on different floors of a building onto a single XY-plane, therefore removing all floor information. We then ran the K-nearest neighbors on the extended training set. Table 8 shows the results of this experiment.

	University (downtown)		Research Lab (midtown)		House (residential)	
	50^{th} %	90^{th} %	50^{th} %	90^{th} %	50^{th} %	90^{th} %
`radar`	4.40	10.27	2.49	4.94	3.11	5.80
`clch`	4.98	18.74	4.41	9.43	3.66	7.02
`ch`	5.76	21.75	4.72	9.44	4.10	7.18
`cell`	9.86	22.31	6.41	11.64	4.35	8.05
`onecell`	14.92	29.80	8.55	14.31	4.67	8.95
`random`	35.61	59.36	13.85	21.33	6.46	15.18

Table 8. Median and 90^{th} percentile localization error with multi-floor fingerprints

Projecting the points collected on different floors onto a single plane has several effects. On one hand, this practice may reduce the localization accuracy as the training points of other floors add "noise" (*e.g.*, potential aliasing), which may result in larger localization errors. On the other hand, if the training points at a specific `<X,Y>` location on all floors have similar signal strength signatures, combining the training data from multiple floors will increase the density of the measurement's grid, which may result in higher accuracy. The ability to detect floor well suggests the latter is not the case, and points close to each other on X,Y dimensions buy on different floors have different signatures.

The multi-floor performance of `radar` in the House is better than in any of the single-floor experiments. We found that the signal strength from the APs outside the building varies more with distance within a floor than within similar position on different floor. As a result, the training data from multiple floors overlaps tightening the grid and increasing localization accuracy. The performance of `radar` in a multi-floor setting in the University and Research Lab buildings is close to the average of the single-floor experiments, which is further indication that `radar` can effectively differentiate between floors in office buildings with heavy concrete and steel frames.

The multi-floor localization error for GSM-based algorithms is also close to the average of the single-floor experiments. This suggests that GSM-based algorithms can differentiate between the floors with good accuracy.

Therefore, for most cases, first identifying the floor and then performing localization using single-floor training data results in higher accuracy than performing the localization using multi-floor data. However, when the number of readings per location is low or differences in signal strength across floors are small, combining the training sets of multiple floors may produce higher localization accuracy.

4.4 Sensitivity Analysis

In this section, we analyze the best GSM performer, `clch`, in more detail. Specifically, we test the localization accuracy of `clch` as a function of the number of channels used, the number of measurements collected per location and the training grid size.

4.4.1 Number of Channels

Figure 10 plots the median localization error for the multi-floor experiment as a function of the number of channels used. Increasing the number of channels results in a larger fingerprint, which allows for more accurate comparison between neighboring points and therefore for increased localization accuracy. The channels picked are sorted by popularity (*i.e.*, the number of readings of the specific channel in all measurements). For example, the median localization error for 6 channels, corresponds to an algorithm where 6 fixed most popular channels are picked from the training set. Notice that the accuracy of the algorithm that picks 6 most popular channels is lower than of the `cell` algorithm. This is because the `cell` algorithm picks 6 strongest cells for each measurement, which may result in much larger fingerprint vector (*i.e.*, completely different 6 cells may be picked in two far locations, increasing the fingerprint vector to at least 12 entries).

Figure 11 plots the percentage of incorrect floor classifications as a function of the number of channels. As expected, picking more channels decreases classification error. Interestingly, in all cases, picking about 20 channels is sufficient for achieving good localization accuracy.

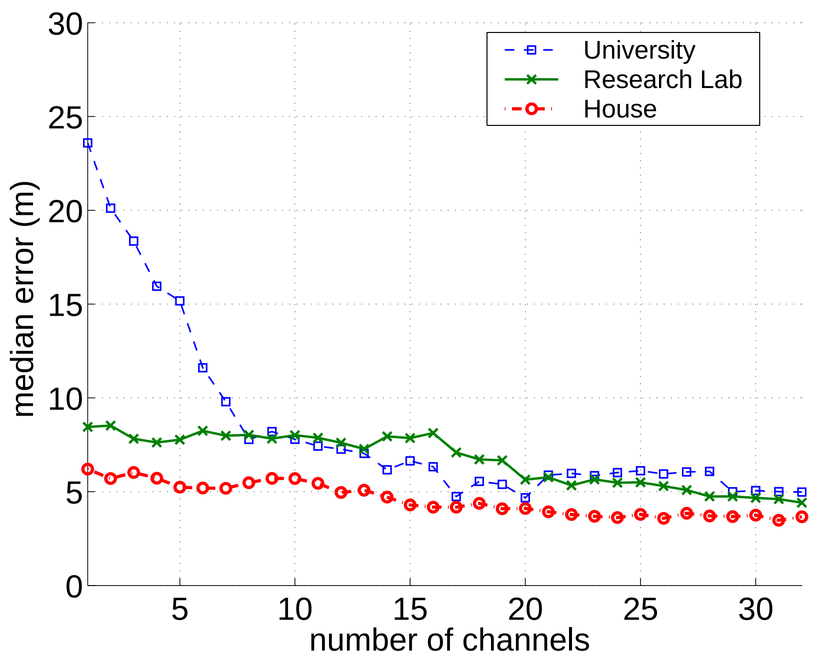

Figure 10. Number of channels and localization error with multi-floor fingerprints

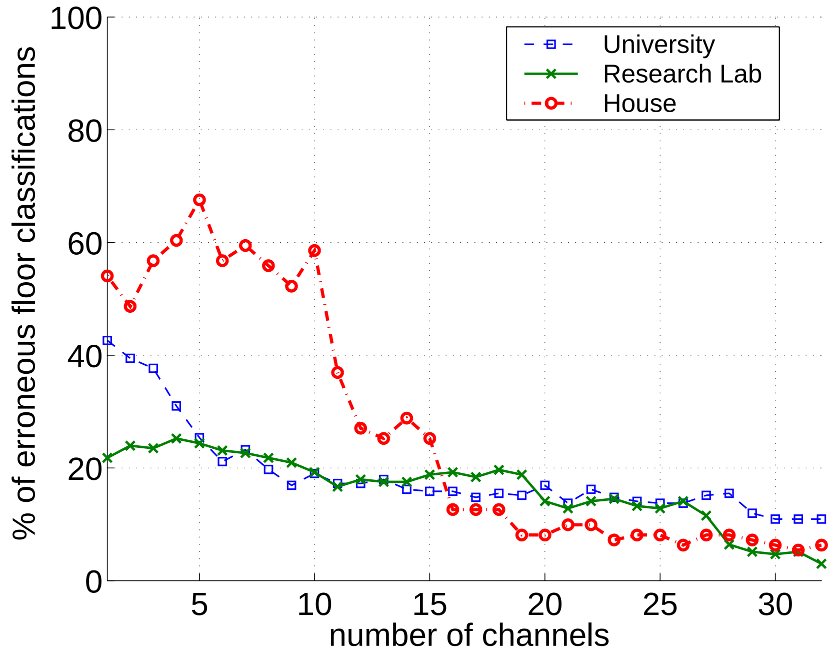

Figure 11. Number of channels and percentage of erroneous floor classifications

4.4.2 Number of Measurements per Location

Although all the results reported so far were based on the average of 2 measurements per location, we actually obtained 10 measurements per location for the University building dataset. However, experiments varying the number of measurements per location between 2 and 10 scan showed virtually no difference in the accuracy of the algorithms. This is because the readings are stable and therefore adding more measurements per location does not improve localization accuracy.

4.4.3 Data Collection Grid Size

We reduce the number of locations used for training by including only every *n*-th point we measured. As the points were collected sequentially while walking through the corridors, this approach still results in points that cover the whole area evenly; only the distance between measured points increases. Figure 12 and Figure 13 show the effects of reducing grid size on the median multi-floor localization error and the floor classification error, respectively. In most cases, reducing the grid size results in lower localization accuracy, but occasionally we do see anomalies. As it turns out, decreasing the size of the grid may eliminate (in some cases) “problematic” or “aliased” points, which in turn increases localization accuracy.

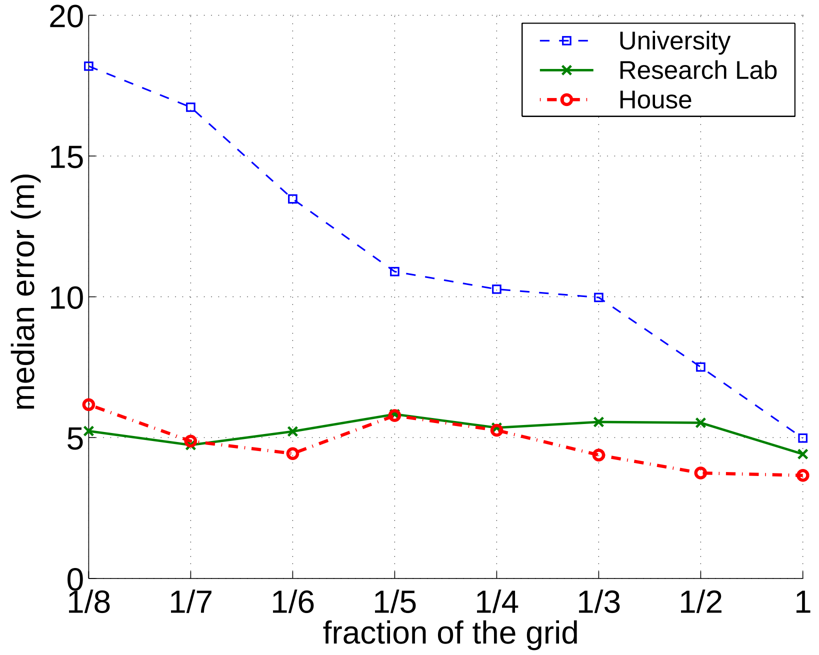

Figure 12. Grid size and localization error with multi-floor fingerprints

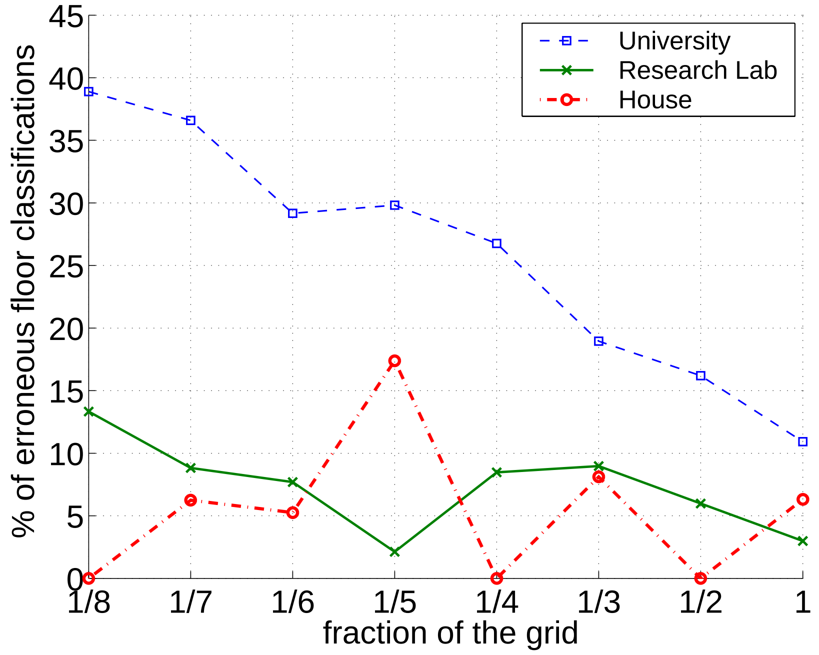

Figure 13. Grid size and percentage of erroneous floor classifications

4.5 Combined 802.11 and GSM localization

In this section, we present an initial attempt to combine 802.11 and GSM fingerprinting. Since we collected both 802.11 and GSM channels information simultaneously, we have been able to combine the readings of both into one large fingerprint. We refer to the method where this large fingerprint was used together with geographical clustering as `clchradar`. The results are summarized in Table 9. The combined algorithm achieves slightly better accuracy in the University building, underperforms `radar` in the Research Lab, and achieves similar performance in the House. An explanation for the lackluster performance of the combined algorithm may be found in the way in which we combine the fingerprint data. By simply concatenating fingerprint vectors we implicitly give more weight to the more numerous and less accurate GSM readings. Therefore, the additional accuracy that could be gained from 802.11 signal readings is overlooked and does not help localizing the user.

	University (downtown)		Research Lab (midtown)		House (residential)	
	50^{th} %	90^{th} %	50^{th} %	90^{th} %	50^{th} %	90^{th} %
`clchradar`	4.03	8.65	3.35	6.39	3.24	4.29
`radar`	4.40	10.27	2.49	4.94	3.11	5.80
`clch`	4.98	18.74	4.41	9.43	3.66	7.02
`random`	35.61	59.36	13.85	21.33	6.46	15.18

Table 9. Multi-floor localization error

5 Conclusions

We presented the first fine-grained GSM-based indoor localization system that achieves median accuracy comparable to an 802.11-based implementation. We show that accurate indoor GSM-based localization is possible thanks to the use of *wide* signal-strength fingerprints that include readings of up to 32 GSM channels in addition to the 6 strongest cells.

While the lack of cell ID information for majority of channels raises the possibility of *world wide aliasing*, we show that filtering fingerprints based on the subsets of the cell IDs of the 6 strongest cells is sufficient for differentiating between locations in our three indoor environments.

We presented evaluation results of our system in three multi-floor buildings located in two North American metropolitan areas, covering a wide range of urban densities. Our GSM-based indoor localization system achieves a median accuracy ranging from 2.48m to 5.44m in large multi-floor buildings. Moreover, our GSM-based system effectively differentiates between floors in both wooden and steel-reinforced concrete structures, achieving correct floor classifications between 89% and 97% of the time. In contrast, in the wooden building, the 802.11-based fingerprinting system achieved correct classifications only 62% of the time due to a limited fingerprint size.

5.1 Future Work

The applicability of additional predictive algorithms to GSM-based fingerprinting is an interesting question. Although this Thesis focuses on WKNN algorithm, other algorithms might improve accuracy or have other advantages. Probabilistic location tracking by taking account the previous locations can be applied to increase the accuracy even further. Also, it would be interesting to apply our method outdoors, where signal propagation is simpler, area is much larger and taking measurements tightly together is not feasible. Collecting measurements in large multi-floor buildings takes a lot of time. The applicability of extending the measurement grid using interpolation can be investigated.

Our initial experiments and analysis suggest that signal strength fingerprints are relatively stable in time. However, it would be interesting to see how much the changing environment changes the accuracy of this method in the long run. Also, GSM signal stability under different weather conditions (*e.g.*, rain, snow, fog) needs to be evaluated.

References

1. L. Aalto, N. Gothlin, J. Korhonen and T. Ojala, "Bluetooth and WAP push based location-aware mobile advertising system," in *MobiSYS '04: Proceedings of the 2nd international conference on Mobile systems, applications, and services*, pp. 49-58. ACM Press, 2004.
2. G. Abowd and E. Mynatt, "Charting past, present, and future research in ubiquitous computing," *ACM Transactions on Computer-Human Interaction (TOCHI)*, v.7 n.1, pp. 29-58, March 2000.
3. P. Bahl and V. Padmanabhan, "Enhancements to the RADAR User Location and Tracking System," Microsoft Research, Technical Report, Feb. 2000.
4. P. Bahl and V. Padmanabhan, "RADAR: An In-Building RF-Based User Location and Tracking System," in *INFOCOM*, March 2000.
5. R. Battiti, M. Brunato, and A. Villani, "Statistical Learning Theory for Location Fingerprinting in Wireless LANs," University of Trento, Informatica e Telecomunicazioni, Technical Report DIT-02-086, Oct. 2002.
6. A. Dey and G. Abowd, "Towards a Better Understanding of Context and Context-Awareness," *Proceedings of the CH12000 Workshop on The What, Who, Where, When, Why and How of Context-Awareness*, April 2000.
7. G. Djuknic and R. Richton, "Geo-Location and Assisted GPS," *IEEE Computer*, vol. 34. pp. 123-125, Feb. 2001.
8. E. Elnahrawy, X. Li and R. Martin, "The limits of localization using signal strength: a comparative study," In *Proceedings of the First IEEE International Conference on Sensor and Ad Hoc Communications and Networks, SECON 2004*, pp. 406-414, Oct. 2004.
9. P. Enge and P. Misra, "Special issue on GPS: The Global Positioning System," *Proceedings of the IEEE*, pp. 3-172, January 1999.
10. N. Golmie , R. E. Van Dyck , A. Soltanian , A. Tonnerre and O. Rébala, "Interference evaluation of Bluetooth and IEEE 802.11b systems," *Wireless Networks*, v.9 n.3, pp. 201-211, May 2003.

11. W. Griswold, P. Shanahan, S. Brown and R. Boyer, “ActiveCampus: Experiments in Community-Oriented Ubiquitous Computing,” *IEEE Computer*, Vol. 37, No. 10, pp. 73-81, October 2004.

12. A. Haeberlen , E. Flannery , A. Ladd , A. Rudys , D. Wallach and L. Kavraki, “Practical robust localization over large-scale 802.11 wireless networks,” *Proceedings of the 10th annual international conference on Mobile computing and networking (MOBICOM)*, September 26-October 01, 2004, Philadelphia, PA, USA

13. M. Hearst, “Support Vector Machines,” *IEEE Intelligent Systems*, v.13 n.4, pp.18-28, July 1998.

14. J. Hightower, R. Want and G. Borriello, “SpotON: An indoor 3D location sensing technology based on RF signal strength,” Technical Report, University of Washington, February 2000.

15. J. Hong, G. Borriello, J. Landay, D. McDonald, B. Schilit and D. Tygar, “Privacy and Security in the Location-enhanced World Wide Web,” In *Proceedings of Ubicomp 2003*, Seattle, October 2003.

16. H. Jiawei and M. Kamber, “Data Mining: Concepts and Techniques,” Academic Press, 2001.

17. A. Kamerman and N. Erkocevic, “Microwave oven interference on wireless LANs operating in the 2.4 GHz ISM band,” *Proceedings of the 8th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, Vol. 3 (1997), pp. 1221-1227.

18. T. Kindberg and A. Fox, "System Software for Ubiquitous Computing," *IEEE Pervasive Computing*, vol. 1, no. 1, pp. 70-81, Jan.-Mar. 2002.

19. P. Krishnan, A. Krishnakumar, W. Ju, C. Mallows and S. Ganu, “A system for LEASE: System for location estimation assisted by stationary emitters for indoor RF wireless networks,” in *IEEE Infocom*, Hong Kong, March 2004.

20. K. Laasonen, M. Raento, and H. Toivonen, “Adaptive On-Device Location Recognition,” In *Proceedings of the 2nd International Conference on Pervasive Computing (Pervasive 2004)*, Vienna, Austria, April 2004.

21. A. Ladd , K. Bekris , A. Rudys , L. Kavraki , D. Wallach and G. Marceau, "Robotics-based location sensing using wireless Ethernet," *Proceedings of the 8th annual international conference on Mobile computing and networking (MOBICOM)*, September 23-28, 2002, Atlanta, Georgia, USA.

22. H. Laitinen, J. Lähteenmäki and T. Nordström, "Database correlation method for GSM location," *IEEE VTC 2001 Spring Conference*, Rhodes, Greece, May 2001.

23. A. LaMarca, Y. Chawathe, S. Consolvo, J. Hightower, I. Smith, J. Scott, T. Sohn, J. Howard, J. Hughes, F. Potter, J. Tabert, P. Powledge, G. Borriello and B. Schilit, "Place Lab: Device Positioning Using Radio Beacons in the Wild," to appear, *Pervasive 2005*, Munich, Germany.

24. L. Luxner, "The Manhattan Project: AT&T Wireless invades the Big Apple with microcells," *Telephony*, 8(20), Feb. 1997.

25. T. Mitchell, "Machine Learning," McGraw-Hill Higher Education, 1997.

26. M. Mouly and M. Pautet, "The GSM System for Mobile Communications," Palaiseau, France, 1992.

27. L. Ni, Y. Liu, Y. Lau and A. Patil, "LANDMARC: indoor location sensing using active RFID," *Wireless Networks*, Volume 10, Issue 6, November 2004.

28. P. Prasithsangaree, P. Krishnamurthy, P. Chrysanthis, "On indoor position location with wireless LANs," *The 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, 2002*, Volume 2, pp. 720-724, Sept. 2002.

29. N. Priyantha, A. Chakraborty, and H. Balakrishnan, "The Cricket Location-Support system," in *ACM International Conference on Mobile Computing and Networking (MobiCom)*, Boston, MA, Aug. 2000.

30. M. Rahnema, "Overview of the GSM System and Protocol Architecture," *IEEE Communications Magazine*, pp. 92-100, April 1993.

31. J. Ross, "The book of Wi-Fi: Install, Configure, and Use 802.11b Wireless Networking," No Starch Press, 2003.

32. S. Saha, K. Chaudhuri, D. Sanghi and P. Bhagwat, “Location determination of a mobile device using IEEE 802.11b access point signals,” *IEEE Wireless Communications and Networking (WCNC) 2003*, Volume 3, pp. 1987-1992, March 2003.

33. J. Scourias, “Overview of the Global System for Mobile Communications,” University of Waterloo, 1995.

34. A. Smailagic, D. Siewiorek, J. Anhalt, D. Kogan and Y. Wang, “Location sensing and privacy in a context aware computing environment,” *Pervasive Computing*, 2001.

35. S. Tekinay, “Wireless Geolocation Systems and Services,” Special Issue of the *IEEE Communications Magazine*, April 1998.

36. E. Trevisani and A. Vitaletti, “Cell-ID location technique, limits and benefits: an experimental study,” *Proceedings of the Sixth IEEE Workshop on Mobile Computing Systems and Applications (WMCSA 2004)*, 2004.

37. R. Want, A. Hopper, V. Falcao, and J. Gibbons, “The Active Badge Location System,” *ACM Transactions on Information Systems*, vol. 10, no. 1, pp. 91–102, Jan. 1992.

38. R. Want, B. Schilit, D. A. Norman, R. Gold, D. Goldberg, K. Petersen, J. Ellis and M. Weiser, “An overview of the ParcTab ubiquitous computing experiment,” *IEEE Personal Communications Magazine*, Vol. 2, Issue 6, pp. 28-43, Dec. 1995.

39. A. Ward, A. Jones, A. Hopper, “A New Location Technique for the Active Office,” *IEEE Personal Communications Magazine*, Vol. 4, No. 5, October 1997. pp. 42-47.

40. J. Webr and C. Lanzl, “Designing a positioning systems for finding things and people indoors,” *IEEE Spectrum*, v.35 n.9, pp.71-78, Sept. 1998.

41. S. Willassen, “A Method for Implementing Mobile Station Location in GSM,” Diploma Thesis, Norwegian University of Science and Technology, Trondheim, Dec. 1998.

42. M. Youssef, A. Agrawala and A. Shankar, “WLAN Location Determination via Clustering and Probability Distributions,” *Proceedings of the First IEEE International Conference on Pervasive Computing and Communications*, p.143, March 23-26, 2003.

43. “Quarterly statistics for Q4 2004,” GSM Association.

44. “3GPP TS 03.22: Functions related to Mobile Station (MS) in idle mode and group receive mode (version 8.7.0 Release 1999); Digital cellular telecommunications system (Phase 2+),” ETSI Technical Specification, 2002.

45. “3GPP TS 05.02: Multiplexing and Multiple Access on the Radio Path (version 8.11.0 Release 1999); Digital cellular telecommunications system (Phase 2+),” ETSI Technical Specification, 2003.

46. “3GPP TS 05.05: Radio Transmission and Reception (version 8.17.0 Release 1999); Digital cellular telecommunications system (Phase 2+),” ETSI Technical Specification, 2004.

47. “3GPP TS 05.08: Radio subsystem link control (version 8.22.0 Release 1999); Digital cellular telecommunications system (Phase 2+),” ETSI Technical Specification, 2004.

48. “IEEE Standard for Wireless LAN Medium Access Control (MAC) and Physical Layer (PSY) Specification,” IEEE Standard 802-11.1997, Approved: June 1997.

49. Single European Emergency Call Number 112, `http://www.europa.eu.int/comm/environment/civil/prote/112/112_en.htm`, 16/05/2005

50. Enhanced 911 – Wireless Services, `http://www.fcc.gov/911/enhanced/`, 16/05/2005

51. Ekahau Inc, `http://www.ekahau.com`, 16/05/2005

52. Global Locate Inc – Indoor GPS, `http://www.globallocate.com/A-GPS/A-GPS_f3.htm`, 16/05/2005

53. RF Technologies, `http://www.rft.com`, 16/05/2005

54. Ubisense Limited, `http://www.ubisense.net`, 16/05/2005

55. Versus Technology Inc, `http://www.versustech.com`, 16/05/2005

GSM MOBIILTELEFONI TÄPNE POSITSIONEERIMINE SISERUUMIDES

Magistritöö

Veljo Otsason

Resümee

Traadita andmeside areng on teinud võimalikuks asukohatundlike rakenduste loomise. Need rakendused võtavad arvesse kasutaja geograafilist asukohta, et esitada talle parasjagu relevantset informatsiooni või pakkuda muid kasulikke teenuseid. Erinevad rakendused on mõeldud erinevates tingimustes kasutamiseks ning nõuavad erinevat täpsust. Kui vabas õhus kasutamiseks mõeldud teenused ei eelda enamasti väga suurt täpsust, siis mitmed spetsiifiliselt siseruumide tarvis loodud rakendused nõuavad mõnemeetrist täpsust, et tuua kasutajale maksimaalselt kasu.

Mobiilsete seadmete täpset positsioneerimist siseruumides on ammu peetud traadita andmeside võimaluste ärakasutamisel oluliseks, ning see on olnud paljude uurimistööde eesmärgiks. Kui vabas õhus on enamasti kasutatud globaalset positsioneerimissüsteemi (GPS), siis siseruumide jaoks on välja pakutud erinevaid lahendusi, mis baseeruvad Wi-Fi (802.11) või Bluetooth traadita andmesidel, ultrahelil või infrapuna lainetel. Käesolev töö esitleb esimest GSM mobiilside signaalidel baseeruvat siseruumide positsioneerimissüsteemi, mis on võimeline määrama mobiiltelefoni asukoha 5-meetrise mediaan-täpsusega suurtes mitmekorruselistes hoonetes. Kandev idee, mis teeb täpse GSM positsioneerimise võimalikuks, on laiendatud signaalitugevuste informatsiooni kasutamine. Lisaks kuue tugevaima raadiomasti signaalide mõõtmisele, mida tavaliselt GSM standardis kasutatakse, jälgime meie ka kuni 32 lisamasti, mille signaalid on piisavalt tugevad, et neid mõõta, kuid enamasti liiga nõrgad efektiivseks andmevahetuseks. Me tegime eksperimente ja mõõtsime oma süsteemi kolmes hoones, mis asuvad kahes suurlinnas. Eksperimentaalsed tulemused näitavad, et meie süsteemi täpsus on võrreldav 802.11-baseeruvate süsteemidega, ning et see suudab ka täpselt vahet teha erinevate korruste vahel nii puit- kui betoonehitistes.