CILoS: A CDMA Indoor Localization System

Waqas ur Rehman
waqas@cs.toronto.edu

Eyal de Lara
delara@cs.toronto.edu

Stefan Saroiu
stefan@cs.toronto.edu

Department of Computer Science
University of Toronto
Toronto, ON M5S 3G4 Canada

ABSTRACT

CILoS is an indoor localization system based on CDMA mobile phone signal fingerprinting. CDMA networks vary their transmission power to accommodate fluctuations in network load. This affects signal intensity and therefore limits the practicality of traditional fingerprinting approaches based on receiver signal strength (RSSI) measurements. Instead, CILoS uses fingerprints of signal delay that are robust to cell resizing. We demonstrate that CILoS achieves a median accuracy of 5 meters, and compares favourably to RSSI fingerprinting systems. We highlight the significance of wide fingerprints, constructed through scanning multiple channels, for achieving high localization accuracy. We also show that our system can accurately differentiate between floors of a multifloor building.

Author Keywords
Location, Localization system, Radio fingerprinting

ACM Classification Keywords
C.2.8 [Communication/Networking and Information Technology]: Mobile Computing—Support Service

INTRODUCTION

This paper presents CILoS, a novel indoor localization system based on the CDMA mobile phone system. CDMA is one of the two most popular mobile phone systems in use today with an estimated 431 million subscribers in 99 countries around the world [23]. The key advantage of our approach is that it leverages the phone's existing hardware and can provide location estimates anywhere CDMA cellular service is available. This system can provide a localization service in places where GPS does not work well, such as in indoor environments or in urban canyons. Accurate indoor localization is important in the context of emergency response [24], as well as other emerging applications, such as location aware advertising and gaming [25].

CILoS is based on signal fingerprinting, an empirical localization technique that involves a *training* or *mapping* phase in which a radio map of the environment is constructed by collecting a series of fingerprints in multiple locations. A radio fingerprint captures a certain property of a group of radio sources heard at a specific location. After performing a training phase, CILoS can help a client determine its location by searching for the closest matches of the current measurement to the set of measurements collected in the training phase.

CILoS is different from previous fingerprinting systems, such as the ones using 802.11 [2] and GSM [18], because it is based on signal delay rather than the receiver signal strength (RSSI). While 802.11 and GSM networks operate with fixed cell sizes, CDMA has a dynamic architecture that supports the frequent reconfiguration of cell-sizes to accommodate fluctuations in network load. Cell resizing affects the power at which beacons are transmitted; this alters the intensity at which signals are perceived at a given location and severely limits the practicality of RSSI fingerprinting. Instead, transmissions from CDMA towers are tightly synchronized with each other making it possible to construct fingerprints that capture the relative time difference at which signals emanating from different base stations are heard at a given location. We show that fingerprints of signal delay are stable even in the face of changes in beacon transmit power such as when cell sizes change.

Experiments conducted on two multi-floor buildings in the Toronto metropolitan area show that CILoS achieves a median accuracy of 5 meters and succeeds in detecting the current floor 90% of the time. This performance is comparable to systems based on 802.11 and GSM. We demonstrate that the key to high accuracy is the use of wide fingerprints. While this finding is consistent with previously reported results for GSM [15], our experience with CDMA indicates that obtaining these wide fingerprints requires scanning of multiple frequency bands from the same or different operators, as interference from nearby base stations limits the number of neighbouring nodes that can be heard.

The rest of this paper is organized as follows. The following section provides a brief introduction into the technological aspects of CDMA that are relevant to radio fingerprinting. The data collection section describes the process we followed to collect our experimental data and the special modem we used. This is followed by a description of our localization algorithms. The evaluation section presents the

UbiComp'08, September 21-24, 2008, Seoul, Korea.
Copyright 2008 ACM 978-1-60558-136-1/08/09...$5.00.

results of our experiments. Finally, the related work section compares our work with previous efforts in indoor localization in general and in signal fingerprinting in particular, and the conclusion summarizes our findings.

CDMA PRIMER

The Code Division Multiple Access (CDMA) mobile phone system was first introduced in 1995. Today 248 CDMA commercial operators provide third-generation (3G) services to 431 million customers in 99 countries [23]. CDMA is a spread spectrum technique in which all base stations owned by one operator share the same spectral bandwidth. To avoid interference, the transmissions from base stations and from mobile hosts are encoded with orthogonal pseudo-random codes.

To enable mobiles to meaningfully compare nearby base stations, all base stations participate in the transmission of a pilot signal. It is this pilot signal that we will use for radio fingerprinting. The pilot signal consists of a pseudo-random sequence of 32768 chips, or symbols as illustrated in Figure 1. Each base station is assigned a unique 64 chip range of the pilot sequence, known as the PN offset, for a total of 512 individual offsets assigned to as many base stations. The PN offset uniquely identifies a base station within a CDMA deployment. The pilot signal is transmitted continuously with the different base stations taking turns to transmit their portion of the sequence. This requires all the base stations to be highly synchronized to a common timing reference, also called system time. This timing reference is achieved using GPS.

A mobile that monitors the pilot signal can determine three key properties that are useful for signal fingerprinting: Ec, Ec/Io, and the PN delay. Ec measures the signal strength of an individual base station's pilot expressed in dBm, and Ec/Io is the power in an individual base station's pilot divided by the total power in the channel expressed in dB. The PN delay measures the difference between the expected and the actual arrival time of the pilot signal. To get around the requirement for tight synchronization between the mobile and the base stations, the PN delays are calculated relative to a reference base station. The mobile selects one base station (e.g., the one with strongest signal) as its timing reference setting its clock with the arrival of its pilot (i.e., set the PN delay for the reference base station to zero). Because all base station transmissions are tightly synchronized, the mobile can then determine when it expects to hear the pilot transmissions from other base stations based on their PN offset. For example, if the mobile uses as its time reference base station with PN offset equal to ten (PN 10), it can expect that the pilot from base station with PN 20 will arrive 640 chips later. The mobile determines the PN delay for base stations by comparing the actual and expected arrival time of their pilots.

Later in the paper we will show that due to the practice of reconfiguring cell-sizes in CDMA networks to accommodate fluctuation in network load, pilot Ec is not an appropriate property for signal fingerprinting. In contrast, we will show that PN delay is amenable to fingerprinting.

PN delay has been used in CDMA networks to determine localization by means of Time Difference of Arrival (TDOA) trilateration [6]. TDOA accuracy, however is low, ranging between 50 and 500 meters depending on interference, system geometry and multipath effects.

DATA COLLECTION

Our experimental setup consisted of a Dell laptop running Windows XP connected to a Condor CDMA scanner via a serial port (see Figure 2). Condor is a dual band PN scanner that can scan both the PCS and Cellular bands and supports CDMAOne and CDMA2000. Condor can measure all the 512 pilots in less than a second and reports Ec, Ec/Io and signal delay for each pilot. The laptop was running the Condor Data Logger software that communicates with Condor and logs the binary data provided by it. Offline processing of the binary data requires specialized software that understands the binary format of files produced by the Condor Data Logger. For this we used BVS Chameleon, a data conversion and filtering tool for CDMA receivers. The result of offline processing is a single file for each channel containing tuples of the form (x-coordinate,y-coordinate BS$_1$=[delay,Ec,Ec/Io] BS$_2$=[delay,Ec,Ec/Io] . . . , BS$_n$=[delay,Ec,Ec/Io]).

We collected measurements during normal business hours in two university buildings: the Bahen Centre for Information Technology at the St. George campus and the South Building at the Mississauga campus. These building are located in geographical regions that differ widely in their network coverage characteristics. The Bahen Centre for Information Technology is located in a busy downtown while the South Building is located in a suburb. In the rest of this paper we refer to these buildings as Downtown and Suburb, respectively.

Downtown, is a modern 8-storey building with dimensions of $88m \times 113m$ per floor and has good cellular coverage. The building is home to lecture rooms, labs and offices. We collected fingerprints on the 5^{th} and 7^{th} floors of Downtown. Limited access to 6^{th} floor forced us to skip this floor. Suburb is an old 5-storey building. While this is a very large building, we limited our data collection to a $66m \times 48m$ region. Suburb is home to labs and faculty offices. The cellular coverage in the building was poor and we noticed no reception at numerous locations on each floor. We collected fingerprints on all floors of Suburb except the basement which had no coverage. For practical considerations all the fingerprints were collected in the hallways of both buildings.

To find active CDMA networks we scanned both the PCS and Cellular bands. We found 6 frequency bands in Downtown and 4 frequency bands in Suburb used by the two cellular operators that provide CDMA service in the Toronto metropolitan area. We will be using OP1 and OP2 throughout the paper to distinguish between these two operators. Each frequency band or channel occupies 1.25 MHz of spectral bandwidth. OP1 uses 4 channels in Downtown and 3

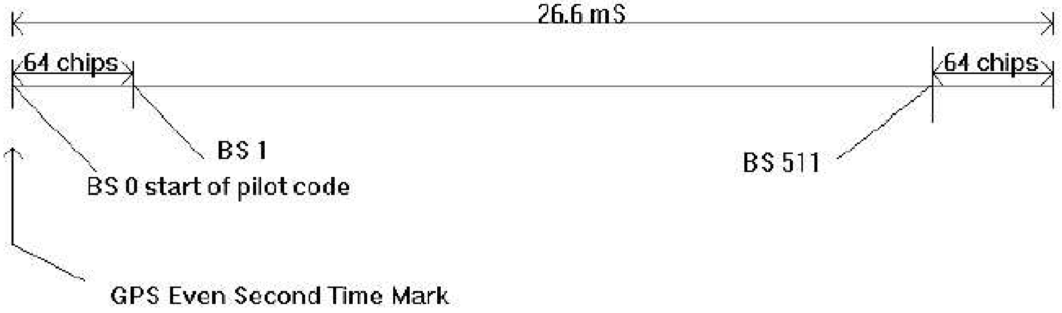


Figure 1. PN offsets in CDMA.

Figure 2. Experimental setup.

channels in Suburb. OP2, on the other hand, uses 2 channels in Downtown and only 1 channel in Suburb. The channels of OP1 operate on 1946.25, 1947.50, 1948.75, and 1981.25 MHz frequencies in Downtown and 1946.25, 1947.50, and 1948.75 MHz frequencies in Suburb. OP2's channels, on the other hand, use 1955.0, and 1957.50 MHz frequencies in Downtown and 1957.50 MHz frequency in Suburb. More channels are used in Downtown to increase the system capacity and provide better service to the denser user population found in this area.

Table 1 summarizes the number of fingerprints collected per floor for each building. In both buildings we collected fingerprints at locations chosen 2 meters apart. In each location we collected 120 measurements per available channel. The reason for collecting multiple measurements per location will be discussed later in the paper. Although we fingerprinted more floors in Suburb than in Downtown, we have collected fewer fingerprints in Suburb due to its smaller size and fewer available channels.

CDMA FINGERPRINTING

Radio fingerprinting requires that the radio signal characteristics that are been recorded vary from one location to another (i.e., have high spatial variation) while remaining constant over time at any single location (i.e., have low temporal variation). GSM localization schemes use signal strength to fingerprint an environment because the strength of GSM signals have high spatial and low temporal variations. However, as we presented in our background section, CDMA base stations vary the power of their signal dynamically to support cell resizing. This is likely to make signal strength have high temporal variation, making it unsuitable for use as a fingerprint.

We verified this assertion by conducting the following experiment. We recorded the signal strength of the same CDMA base station at one single location continuously for two hours. We repeated this experiment on four different days. Figure 3 shows the signal strength recorded for each of our four experiments. For each experiment, the signal strength varied over the course of one experiment by as much as 15dB. Even worse, each of the four experiments showed different signal strength characteristics. These experiments confirmed our intuition that signal strength was an unsuitable radio characteristic for fingerprinting using CDMA.

Instead, we focused on a different radio characteristic – the signal delay. As we described in the background section, a CDMA base station transmits at predefined time intervals. In fact, CDMA base stations use highly accurate clocks to synchronize their signal transmission. In turn, this leads to a CDMA signal whose signal delay does not vary over time.

We verified whether the signal delay is a suitable metric for fingerprinting by measuring its temporal and spatial variation. To verify this we used the data from above experiments conducted to show temporal variations of signal strength. In addition to recording the signal strength, we also measured the signal delay of multiple base stations in these experiments. Figure 4 plots the signal delay measurements for each of our four experiments. The signal delay metric appears very stable within one experiment and also across different experiments although occasional erroneous readings do occur.

We eliminate these bad readings with a simple two-step filtering technique. In the first step, we remove all readings with very low signal-to-noise ratio. As we described in the background section, Ec/Io is a common metric for CDMA to measure the signal-to-noise ratio; we filter out all readings whose Ec/Io is lower than -21dB. In the second step, we remove all spurious errors using a simple windowing technique. At each location, we take several consecutive measurements of signal delay over a short time interval. We then

	Downtown		Suburb			
	5^{th} Floor	7^{th} Floor	2^{th} Floor	3^{th} Floor	4^{th} Floor	5^{th} Floor
per floor	732	786	248	400	248	248
per building	1518		1144			

Table 1. The number of fingerprints collected for two buildings.

record the most common value out of these measurements (i.e., the *mode*) as the reading for a particular location. We will present an in-depth sensitivity analysis of our two-step filtering technique in the evaluation section.

Figure 5 shows the effects of our filtering technique on our four experiments. The signal delay readings remain very stable over time for the same location over the course of all the experiments.

In addition to low temporal variation, the signal delay metric must have high spatial variation to be a suitable metric for fingerprinting. To verify this, we measured the signal delay in 10 different locations chosen two meters apart. All these measurements were collected on one floor in Downtown. Figure 6 shows the signal delays of six base stations (i.e., the fingerprint) at each of these 10 locations. While one base station can have the same signal delay at different locations, when combined, all six base stations form unique fingerprints at each of the 10 locations. These experiments show that signal delay is a suitable radio characteristic for CDMA fingerprinting.

LOCALIZATION ALGORITHMS

CILoS estimates a client's location by comparing the client's current measurement with the fingerprint map collected in the training phase. At a high-level, CILoS' localization algorithm is simple – find the k closest fingerprints to a signal delay reading and use some form of arithmetic mean to estimate the measurement's location. We use Euclidean distance to measure the distance between a measurement and each of the fingerprints; for a given client measurement of signal delay $< PN_1^r, PN_2^r, \ldots, PN_n^r >$ and a given entry in the fingerprint map $< PN_1^{fp}, PN_2^{fp}, \ldots, PN_n^{fp} >$, we measure the distance d as:

$$d = \sqrt{\sum_{i=1}^{n} (PN_i^r - PN_i^{fp})^2} \qquad (1)$$

If any of the PNs are missing either in a fingerprint or in the client measurement (e.g., due to an error in the measurement or due to our filtering scheme), we assign maximum signal delay (e.g. 64 chips) to that particular base station. Once we compute the distances to each of the fingerprints, we select the k closest fingerprints and estimate the client's location by taking the weighted average of selected k closest fingerprints. Our weighted average assigns to each distance a weight equal to the distance's reciprocal; in this way, closer distances have higher weights. Finally, the choice of k is important to the estimate's accuracy. We experimented with different values of k and we found that setting $k = 3$ leads to the best accuracy of our estimates.

Our localization algorithms can be classified into two broader categories: simple algorithms and feature selection algorithms.

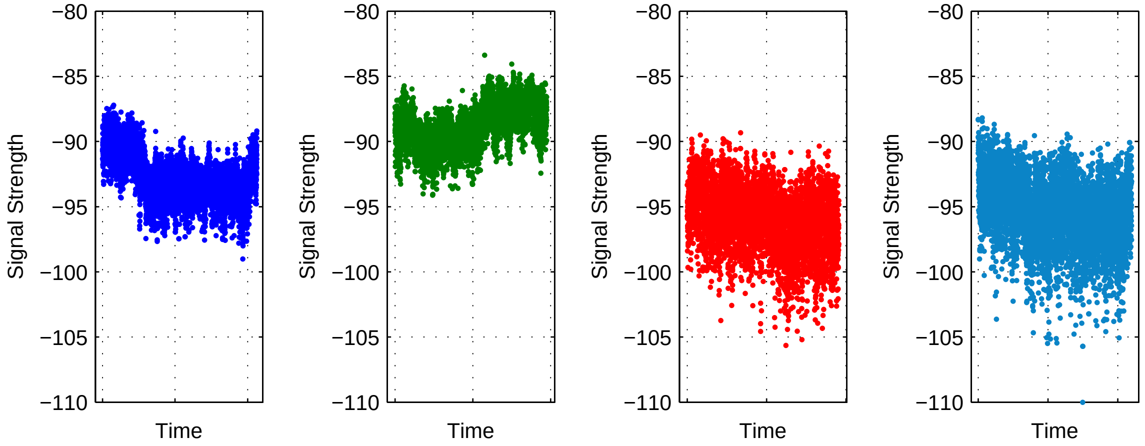
Simple Algorithms

The *simple* algorithms use signal delay readings of all the base stations in training and testing points to calculate the Euclidean distance. We implemented two variants of *simple* algorithms that differ in the number of channels used in measurements: (i) *allChannels* uses PN delay readings from all the available channels; (ii) *oneChannel* uses PN delay readings from a single channel.

Feature Selection Algorithms

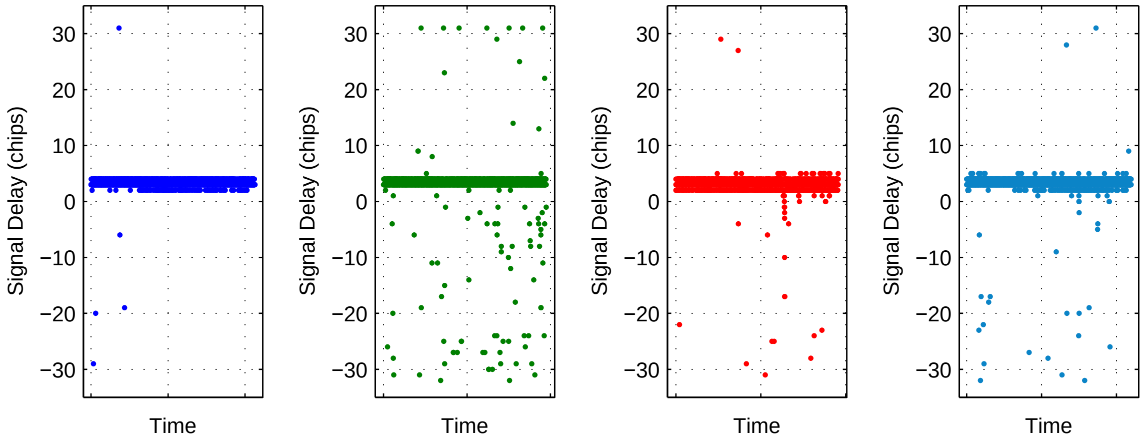
The *simple* algorithms assume that the accuracy of estimation increases as we add more PN readings to the fingerprints. In practice, some radio sources could be so noisy or so unstable that the localization algorithm should always ignore them. Identifying the set of all channels and PNs that leads to the best accuracy is intractable; to do so, we would have to verify all combinations of channels and PNs from our radio map that includes 4 to 6 channels and 60 to 90 PNs. Instead, we use a machine learning approach called *feature selection* to identify these sources of error. Feature selection uses two standard greedy techniques to remove the source of error – forward selection and backward elimination [3]. With forward selection, the algorithm starts with an empty set and adds one "feature" (i.e., a channel or a PN) at a time. At each step, the feature is selected greedily to be the one leading to the best increase in accuracy out of all possibilities. With backward elimination, the algorithm starts with all features and removes one feature at a time, again greedily selecting the ones that contribute the most to the error. Both techniques stop when adding or removing a feature does not lead to any accuracy improvements. We tried both forward selection and backward elimination to compute two sets of features. The results using these two feature sets were comparable so we only report the results using forward selection.

We implemented two variants of feature selection algorithms that differ on the basis of how they filter noisy PNs: (i) *fs*$_{ch}$ uses the set of all channels as a feature set. In this case we use PN readings from those channels that lead to best accuracy; (ii) *fs*$_{pn}$ uses all the channels and filters noisy PNs individually as opposed to filtering them in group based on channels.



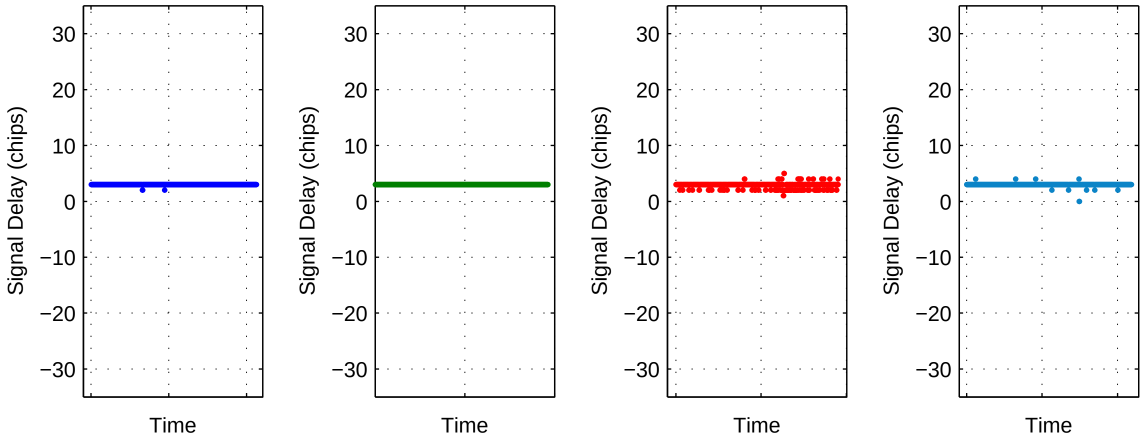
(a) February 26th, 2007 (b) February 27th, 2007 (c) March 1st, 2007 (d) March 2nd, 2007

Figure 3. CDMA's signal strength varies over time. The signal strength of a CDMA base station was measured over two hours.



(a) February 26th, 2007 (b) February 27th, 2007 (c) March 1st, 2007 (d) March 2nd, 2007

Figure 4. CDMA's signal delay remains stable over time. The signal delay of a base station was measured over two hours.



(a) February 26th, 2007 (b) February 27th, 2007 (c) March 1st, 2007 (d) March 2nd, 2007

Figure 5. CDMA's signal delay shows very low variation after filtering and windowing.

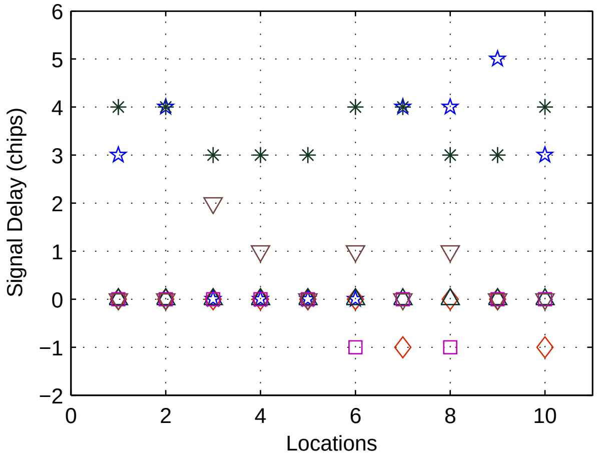


Figure 6. The variation of signal delay over locations. Each location has a unique signal fingerprint using 6 different base stations.

Changing Reference Base Station

As discussed earlier, PN delays are determined in relationship to a reference base station (usually the one with the highest signal strength). However we noticed that the reference base station can vary over time at a location depending on network traffic and signal power. This means the two fingerprints obtained using different reference base stations at a single location may appear completely different. This problem can be eliminated by converting the two fingerprints to a common reference base station. Since the signal delay measurement for each base station reflects the delay in chips from actual arrival time, we can use Figure 1 to change the reference base station and calculate new signal delay readings using Algorithm 1.

```
Algorithm 1 Change reference base station
1. Calculate pilot arrival in chips for each PN.
       $PA_{actual} = delay_{actual} + 64 \times PN$
2. Calculate the expected pilot arrival for the new reference PN$_x$.
       $PA_{expected} = 64 \times PN_x$
3. Calculate the difference in actual and expected pilot arrival of PN$_x$.
       $t = PA_{actual} - PA_{expected}$
4. Subtract $t$ from PA$_{actual}$ of each PN. This changes the reference to PN$_x$.
       $PA_{new} = PA_{actual} - t$
5. Calculate the new signal delay for each PN.
       $delay_{new} = PA_{new} - 64 \times PN$
```

EVALUATION

In this section we first analyze the data we collected and then evaluate the accuracy of our localization algorithms.

Figure 7 shows the average number of PN offsets (i.e., base stations) recorded per location for different channel combinations. We observe that while in principle the Condor scanner can listen simultaneously to 512 base stations per channel, in practice interference from nearby base stations limits the number of effective PN offsets (those with an Ec/Io value above the -21db threshold) to an average of 4 and 2 for Downtown and Suburb, respectively.

As expected, figure 7 also shows that it is possible to dramatically increase fingerprint width by scanning multiple frequency bands from the same or different operators. We show later in this section that increasing the width of the fingerprint (i.e., the number of distinct PN offsets) leads to substantial improvements in localization performance.

The higher Downtown numbers reflect the larger number of available channels (6 vs. 4) as well as an average of twice as many recorded PN offsets per channel. We hypothesize that this is the result of differences in base station density between downtown and suburban deployments as well differences in building materials.

During data analysis we noticed that it is common for base station to use the same PN offset for transmission on multiple channels. We exploit this observation to reduce the effect of PN aliasing. PN aliasing occurs if the pilot of a base station does not arrive in the search window allocated for it. We discover PN aliasing when we compared the recorded PNs to the actual layout of base stations in our area, and noticed the presence of some signal delay reading for base stations which were not physically present.

Localization Accuracy

We evaluate the accuracy of our algorithms by removing a training point from the radio map and then try to infer its location. We repeat this process for all the training points. The approach is somewhat pessimistic since no point in the radio map matches with the testing point. Similarly, for our machine learning algorithms we use leave-one-out cross validation.

Table 2 shows the 50^{th} and 90^{th} percentile within floor localization error for both buildings. Shown are results for algorithms that use different numbers of channels from the two operators in our area, as well as two algorithms that use feature selection. fs_{ch} uses the set of all channels as a feature set and fs_{pn} uses all PNs from all channels as a feature set. The error is calculated as the Euclidean distance between the actual and inferred location of the testing point. Results for the additional floors of Suburb are similar and are not shown.

Figure 8 provides an alternative view of the data with additional details for the 7^{th} floor of Downtown. The plot shows the cumulative distribution (CDF) of the localization error.

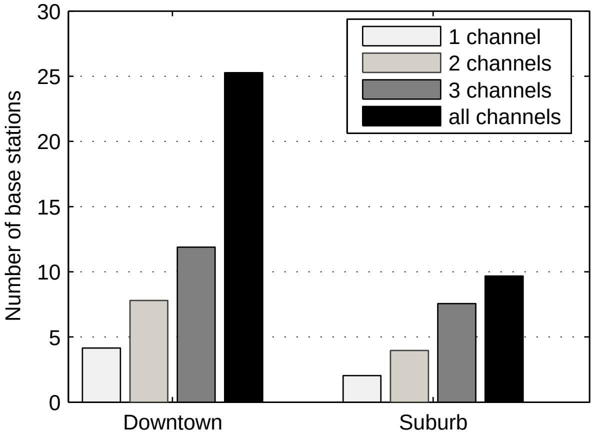


Figure 7. The average number of PNs recorded per location for different channel combinations.

As expected *oneChannel* performs the worst across all the floors. The poor performance of *oneChannel* is the result of limited number of PN readings recorded using a single channel. We notice a significant improvement in the localization accuracy as we widen the fingerprint by adding readings from additional channels. For example, *allChannels*$_{op1+op2}$ improves the median accuracy of *oneChannel*$_{op1}$ by up to 40%. The two feature selection algorithms show that choosing channels and PNs wisely further improves system performance by removing noisy radio sources from the fingerprint. Specifically, fs$_{pn}$ performs the best and achieves median accuracy between 4.5 and 6.7 meters. Overall fs$_{pn}$ achieves improvements in median accuracy of up to 50% over *allChannels*$_{op1+op2}$.

Comparison with 802.11 and GSM

Table 3 shows the within-floor median localization error of 802.11, GSM and CDMA for Downtown. GSM and 802.11 experiments use traces collected in our previous work [15]. The median width of 802.11 and GSM fingerprints is 5 access points and 25 base stations, respectively. The table shows that when all radio sources are used, 802.11 and GSM significantly outperform CDMA. Once feature selection is used, however, the performance of CDMA matches that of 802.11 and GSM. While feature selection also results in improvements for 802.11 and GSM, it is clear that it plays a critical role for good CDMA performance.

Table 4 reports the effectiveness of 802.11, GSM and CDMA to differentiate between floors for Downtown. Since the concrete floors significantly attenuate 802.11 signals, 802.11 achieves 100% classification accuracy. CDMA also shows high classification accuracy and slightly performs better then GSM.

Sensitivity Analysis

In this section we analyze the sensitivity of localization accuracy as a function of the Ec/Io threshold used for filtering

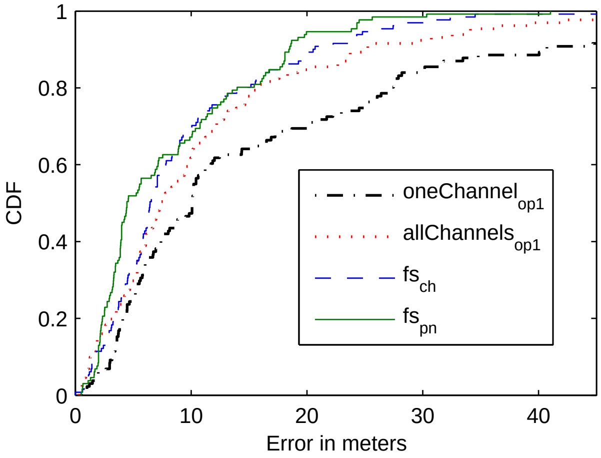


Figure 8. CDF of Localization error for 7th floor of Downtown.

base stations and the number of measurements collected per location.

Ec/Io Threshold

We observed in our initial experiments that signal delay measurements are stable as long as Ec/Io of a base station stays higher than -21dB. Figure 9 shows how we selected this threshold. The figure plots the localization error for the 5th floor of Downtown as a function of decreasing Ec/Io cutoff value. We observe that accuracy increases as we decrease the Ec/Io value. Lower Ec/Io values ensure that the more remote base stations having low signal strength are also included in the fingerprints making them wider, which in turn increases system performance. But there is limit to which we can decrease the Ec/Io. The threshold is around -21dB. By decreasing Ec/Io below threshold we start including those base stations in the fingerprints that have unstable signal delay measurements. These unstable base stations act as a noise and decrease the localization accuracy as illustrated in the figure. Experiments conducted on the 7th floor of Downtown and in the Suburb building show a similar trend and are therefore omitted.

Number of Measurements per Location

The results reported so far take 120 measurements per location for each channel. We use the *mode* of these measurements as a signal delay reading for a particular location. Each measurement takes about one second so recording of 120 measurements in a practical system imposes some timing constraints. The purpose of multiple measurements is to get a stable reading. This implies we only require the number of measurements that stabilize the signal delay reading. We conducted the following experiment to estimate this value. We measured the signal delay of several PNs at one location for two hours. This resulted in approximately 14400 signal delay measurements for each PN. We calculate the mode of these measurements and then chopped the entire set of measurements into smaller segments. We used the segment or window sizes of 1, 5, 10, 15, 20, and 30 measurements. We compute the mode of measurements in each of

	Downtown				Suburb			
	7^{th} Floor		5^{th} Floor		5^{th} Floor		4^{th} Floor	
	50%-ile	90%-ile	50%-ile	90%-ile	50%-ile	90%-ile	50%-ile	90%-ile
fs_{pn}	4.5	18.4	4.7	21.3	6.7	23.3	6.0	19.9
fs_{ch}	6.4	20.5	8.5	25.8	8.6	21.2	9.8	19.4
$\text{allChannels}_{op1+op2}$	7.6	22.2	9.8	23.9	8.8	21.8	13.2	22.7
allChannels_{op1}	7.4	25.1	10.2	24.6	14.4	33.2	12.1	38.1
oneChannel_{op1}	10.1	40.2	13.1	32.7	15.2	36.0	12.7	38.1

Table 2. Within-floor localization error in meters.

	7^{th} Floor			5^{th} Floor		
	802.11	GSM	CDMA	802.11	GSM	CDMA
AllRadioSources	4.6	5.2	7.6	4.5	7.0	9.8
FeatureSelection	3.6	3.8	4.5	3.2	6.4	4.7

Table 3. Within floor median localization error in meters for Downtown.

802.11	GSM	CDMA
100%	84%	87%

Table 4. Percentage of successful floor classification for Downtown.

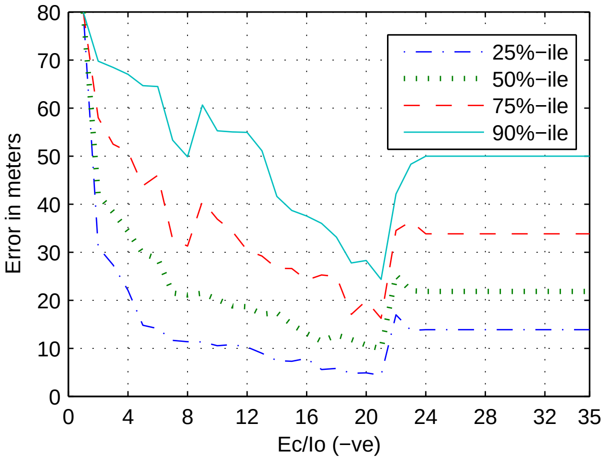


Figure 9. Localization error as a function of Ec/Io.

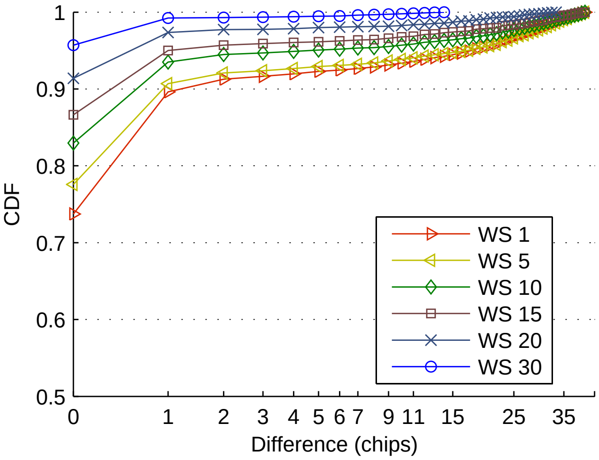


Figure 10. CDF of difference in signal delay for a window size to signal delay of entire experiment. Window sizes of 1, 10 15,20 and 25 are used.

these segments and compare with the mode of entire experiment. Figure 10 plots the CDF of the difference in the signal delay reading of various window sizes to the signal delay reading of the entire experiment. The graph includes signal delay measurements of all the visible base stations. We observe that $15 - 20$ measurements are sufficient to stabilize the signal delay reading.

CELL PHONE DEPLOYMENT ISSUES

The Condor PN scanner we used for our experiments is a bulky unit that commands a significant price well above 10,000 USD. We had to resort to using a special modem because current CDMA phones limit the availability of PN delay information, e.g., only monitor PNs which are in the *active list* and *neighbour list*, and do not give third party applications control over which channel is used. We observe, however, that these are software and not hardware limitations (the phone after all has the ability to monitor all 512 PNs and switch between channels), and that with the appropriate changes it should be possible to implement CILoS on a standard CDMA phone.

RELATED WORK

The growing interest in location-aware systems and service has resulted in a wealth of research on accurate localization technology. Although the Global Positioning System (GPS) provides accurate location information outdoors, it does not operate well in indoor environments and other areas with limited view of the sky. To address this limitation many systems have been pursued using a variety of techniques [12].

The original Active Badge system [7] and follow on commercial systems like Versus [21] use infrared emitters and detectors to achieve 5-10 m accuracy. Both the Cricket [17] and the Batt [20] systems use ultrasonic ranging to estimate location. Depending on the density of infrastructure and degree of calibration, ultrasonic systems have accuracies between a few meters and a few centimetres. Most recently, ultra-wideband emitters and receivers have been used to achieve accurate indoor localization [22]. The common drawback of all of these systems is that they require custom infrastructure for every area in which localization is to be performed. As a result, these systems have not seen significant deployment outside of high-value applications like hospital process management. In contrast CILoS leverages the existing CDMA cellular infrastructure for localization thus eliminating the cost associated with infrastructure deployment.

The earliest work in fingerprinting systems was done by Bahl et al. who observed that the signal strength of a radio source exhibits spatial variation but is consistent in time. They used this observation to build Radar [2]. Using four 802.11 access points Radar located a laptop of its true position with an accuracy of 2-3 meters. Since the first version there have been many improvements to Radar's fingerprint matching algorithm to improve its accuracy [1, 5, 9].

Localization based on fingerprinting of mobile phone signals has been the focus of several recent research efforts. Compared to 802.11, mobile phone networks provide better coverage and have a more stable infrastructure that guarantees a radio map that degrades at a slower rate. Prior to this work, most of the work on mobile phone fingerprinting has concentrated on GSM and has been based on receiver signal strength (RSSI) fingerprinting. Laitinen et al. [13] used GSM-based fingerprinting for outdoor localization. They have collected sparse fingerprints from the 6-strongest cells, achieving 67th percentile accuracy of 44 m. Similarly, Laasonen et al. [10] used the transition between GSM cell towers to determine the places a user goes. PlaceLab [11] is another system that uses sparse traces of GSM and 802.11 radios to estimate user location with $100 - 150$ meter accuracy in a metropolitan area. While these systems provide outdoor localization, Otsason et al. [15] used GSM for indoor localization. In addition to using 6-strongest cells for fingerprinting, they also included the cells that are strong enough to be detected. Using fine grained fingerprints with granularity of 1.5 meter they showed that their system can achieve the accuracy comparable to 802.11 based systems. Similar, SkyLoc [19] uses GSM fingerprints to identify the floor where a mobile user is located in large multifloor buildings.

In contrast, this paper focuses on CDMA-based fingerprinting. CDMA support for cell-size reconfiguration to accommodate dynamic fluctuations in network load prevents the use of RSSI fingerprinting – cell resizing affects the power at which beacons are transmitted. Instead, the approach introduced in this paper is based on the fingerprinting of signal delay, which we have shown is not affected by cell resizing.

The only other research that we are aware of that explores CDMA fingerprinting is by Li et al. [14]. This work, however, is based on RSSI fingerprinting and as a result is not likely to be robust due to frequent cell resizing typical of CDMA systems.

CDMA cellular network have also been used in past to provide mobile positioning based on time of arrival (TOA) and time difference of arrival (TDOA) measurements [4, 8]. These techniques suffer from line of sight and multipath errors, have low accuracy (50-500 meters) and are not applicable to indoor environments. Instead, CILoS overcomes these challenges by using signal delay fingerprints. Our technique works indoors and has higher accuracy.

Finally, powerline positioning (PLP) [16] is another system that uses fingerprinting for localization. PLP used tones transmitted along the residential powerline to fingerprint different locations in a home.

CONCLUSIONS

We presented CILoS – an accurate indoor localization system based on the fingerprinting of CDMA mobile phone signals. Traditional fingerprinting approaches based on the receiver signal strength (RSSI) do not work in CDMA systems because CDMA cell sizes frequently change. Instead, CILoS is based on the fingerprinting of signal delay, which we have shown is resilient to cell resizing.

Experiments conducted in two geographically dispersed locations show that our system achieves a median accuracy between 4.5 and 6.7 meters in large multifloor building. Moreover, CILoS also correctly differentiated between floors 90% of time. The high localization accuracy and floor classification of CILoS is the result of wide fingerprints obtained using multiple CDMA channels.

ACKNOWLEDGMENTS

We thank Derek McAvoy for helping us to understand the nuances of CDMA and the Bell network. We also extend our gratitude to our shepherd Shwetak Patel and the anonymous reviewers for their many useful suggestions.

This research was supported by the Bell University Labs (BUL), the National Science and Engineering Research Council (NSERC), the Canadian Foundation for Innovation (CFI), and the Ontario Innovation Trust (OIT). Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the BUL, NSERC, CFI, OIT, or the University of Toronto.

REFERENCES

1. Agrawala, A.K. and Shankar, A.U. WLAN Location Determination via Clustering and Probability Distributions. In *IEEE PerCom 2003*, 143–150.

2. Bahl, P. and Padmanabhan, V. RADAR: An In-Building RF-Based User Location and Tracking System. In *INFOCOM 2000*, 775–784.

3. Blum A. and Langley P. Selection of relevant features and examples in machine learning. In *Journal on Artificial Intelligence 1997*, 245–271.

4. Caffery, J. and Stuber, G.L. Vehicle Location and Tracking for IVHS in CDMA Microcells. In *Proc. IEEE PIMRC 1994*, 1227–1231.

5. Haeberlen, A., Flannery, E., Ladd, A., Rudys, A., Wallach, D., and Kavraki, L. Practical robust localization over large-scale 802.11 wireless networks. In *Proc.ACM MOBICOM 2004*.

6. Hepsaydir, E. Mobile positioning in CDMA cellular networks. In *Proc. IEEE VTC 1999*, 795–799.

7. Hopper, A., Harter, A. and Blackie, T. The Active Badge System. In *Proc. INTERCHI 1993*, 474–481.

8. Ko, J.L. Tracking Of Mobile Phone Using Imm in CDMA Environment. In *Proc. IEEE ICASSP 2001*, 2829–2832.

9. Ladd, A., Bekris, K., Marceau, G., Rudys, A., Kavraki, L., and Wallach, D. Robotics-based location sensing using wireless Ethernet. In *Proc.ACM MOBICOM 2002*.

10. Laasonen, K., Raento, M. and Toivonen, H. Adaptive On-Device Location Recognition. In *Proc. PERVASIVE 2004*, 287–304.

11. LaMarca, A., Chawathe, Y., Consolvo, S., Hightower, J., Smith, I.E., Scott, J., Sohn, T., Howard, J., Hughes, J., Potter, F., Tabert, J., Powledge, P., Borriello, G. and Schilit, B. N. Place Lab: Device Positioning Using Radio Beacons in the Wild. In *Proc. PERVASIVE 2005*, 116–133.

12. LaMarca, A., de Lara, E. Location Systems: An Introduction to the Technology Behing Location Awareness. *Morgan & Claypool Publishers*. 2008.

13. Laitinen, H., Lahteenmaki, J., Nordstrom, T., Database correlation method for GSM location. In *Proc. of the 53rd IEEE Vehicular Technology Conference*, 2001.

14. Li, B., Dempster, A.G., Barnes, J., Rizos, C., Li, D. Probabilistic algorithm to support the fingerprinting method for CDMA location. In *Proc. Int. Symp. on GPS/GNSS 2005*.

15. Otsason, V., Varshavsky, A., LaMarca, A. and de Lara, E. Accurate GSM Indoor Localization. In *Proc. UbiComp 2005*, 141–158.

16. Patel, S.N., Truong, K.N. and Abowd, G.D. PowerLine Positioning: A Practical Sub-Room-Level Indoor Location System for Domestic Use In *Proc. UbiComp 2006*, 441–458.

17. Priyantha, N.B., Chakraborty, A. and Balakrishnan, H. The Cricket location-support system. In *Proc. ACM MOBICOM 2000*, 32–43.

18. Varshavsky, A., de Lara, E., LaMarca, A., Hightower, J., and Otsason, V. GSM Indoor Localization. In *Pervasive and Mobile Computing Journal 2007*, 698–720.

19. Varshavsky, A., LaMarca, Hightower, J. and de Lara, E. The SkyLoc Floor Localization System. In *PerCom 2007*, 125–134.

20. Ward, A. and Jones, A. A New Location Technique for the Active Office. In *IEEE Personnel Communications 1997*, 42–47.

21. Versus Technologies. http://www.versustech.com

22. Ubisense. http://www.ubisense.com

23. CDMA Development Group. http://www.cdg.org/

24. 911 Services. http://www.fcc.gov/pshs/911

25. Dodgeball.com: Social Mobile Software. http://www.dodgeball.com