Copyright Transfer Agreement:

"The authors represent that the work is original and they are the author or authors of the work, except for material quoted and referenced as text passages. Authors acknowledge that they are willing to transfer the copyright of the abstract and the completed paper to the SDR Forum for purposes of publication in the SDR Forum Conference Proceedings, on associated CD ROMS, on SDR Forum Web pages, and compilations and derivative works related to this conference, should the paper be accepted for the conference. Authors are permitted to reproduce their work, and to reuse material in whole or in part from their work; for derivative works, however, such authors may not grant third party requests for reprints or republishing."

COGNITIVE GEOLOCATION: LEARNING LOCATION BY LISTENING TO THE RADIO

Arash Farhang (Dept. of Electrical & Computer Engineering, Salt Lake City, Utah, USA, arash.farhang@utah.edu) and Neal Patwari (Dept. of Electrical & Computer Engineering, Salt Lake City, Utah, USA, npatwari@ece.utah.edu)

ABSTRACT

Global positioning system (GPS)-based localization is often unavailable in indoor environments, in urban areas, and is susceptible to jamming. We evaluate *cognitive geolocation*, the use of local ambient wireless signals for the estimation of a receiver's position. Cognitive geolocation can be seen as localization fingerprinting, in which a database of pre-measured spectrum vs. location is stored, and a receiver's current spectrum measurement is compared to the database to estimate location. The database includes spectral measurements at different locations and times of day. We present a feasibility study, in which extensive measurements of spectral activity are recorded across time and space in a campus building, in both a training and test set. In the study, receiver locations are correctly identified between 80 to 90% of the time. Our results show the capability of cognitive geolocation and point to a promising area of localization research.

1. INTRODUCTION

GPS is very useful for determining an exact location, however, within buildings, GPS signals may be quite low due to shadowing caused by walls and ceilings. In urban areas, position estimates can also be denied or degraded due to the lack of line-of-sight to GPS satellites. In general, GPS is limited due to the tight link budget of satellite links. The emergence of frequency agile software defined radios enables the reception of a broader range of radio signals. In general, it makes sense to rely on a range of ambient wireless signals for the purposes of localization. Many such signals have sufficient signal strength indoors and in urban locations; in fact, many types of signals are more prevalent in urban areas and in buildings, where GPS is least available. The use of complementary signals may thus improve the robustness of localization receivers.

One example of ambient signal localization is the use of analog TV signals for localization. A technology commercialized by Rosum, Inc. measures the time delays of TV synchronization signals (ghost canceling reference) sent as part of the US analog broadcast television standard, ATSC [4]. Multiple delay measurements from different TV broadcast towers can then be used to estimate receiver position. Similar to a differential GPS receiver, the time delays of the TV signals can be measured and provided to the receiver to aid in the location solution.

We also note the progress in indoor localization using ambient signals from WiFi access points. The received signal strength (RSS) fingerprinting method is credited to Bahl and Padmanabhan [5], and has been commercialized by real-time location services (RTLS) providers AeroScout Inc. and Ekahau Inc. In RSS fingerprinting, the RSS at a WiFi tag of multiple access points' signals is measured and used as a 'RSS fingerprint'. The fingerprint is compared to a pre-measured database of RSS fingerprints for many locations within the building. This database is either manually and densely measured in the environment of interest, or extrapolated from less dense measurements based on propagation models. Reported accuracies from RSS fingerprint methods (5m [5], 1.7 m [6]) demonstrate accuracies as high, or higher than, achieved by multi-lateration approaches, which use RSS to estimate range and then use range estimates to compute position.

In this paper, we study an approach to localization from ambient signals that does not rely on any particular radio standard and which extends fingerprinting beyond WLAN signals. We use the ability of a spectrum-agile radio to measure the spectral occupancy across a wide bandwidth. The cognitive localization idea is that if the electromagnetic (EM) spectrum of ambient radio signals remains relatively constant and unique for any given location, then it can be used as a 'fingerprint' to identify that location. A database of spectrum measurements may be created for many locations, and then whenever a cognitive radio is near one of those locations, it may measure the spectral occupancy and compare its current spectrum measurement to the database in order to determine location. Geolocation is commonly seen as a burden for cognitive radios – a radio must determine its location in order to look up in a database the location's typical spectrum usage. In contrast, in this work, we look at geolocation as an opportunity for cognitive radios, which may be able to self-localize based on their wideband measurement of the spectrum. Cognitive geolocation may have significant applications for localization of people and objects in environments where GPS is denied.

This paper presents the results of a measurement campaign, which confirms the ability of cognitive geolocation to estimate location within the third floor of the Merrill Engineering Building (MEB) at the University of Utah. Our research relied on extensive temporal and spatial measurements of the spectrum at nine different locations. First, measurements were made over a three-day period during the morning and afternoon hours to be used as training set for a database. Next, a set of measurements was taken two weeks later, one set during the day and one set at night. These served as a test set. We found a 91% probability of correct detection of location for the day set, and an 80% probability of correct detection for the night set, even though nighttime measurements were not part of the database. The results show the potential of using cognitive radio spectrum measurements of ambient radio sources in order to self-localize.

2. EQUIPMENT

We evaluate cognitive geolocation using the universal software radio peripheral (USRP) from Ettus Research [1]. It is capable both of measuring the spectrum across a wide bandwidth and saving the data on a PC. As a standard platform for software defined radio (SDR), it represents the capabilities which a typical SDR or cognitive radio might possess. The USRP can be programmed in software to switch frequency bands and to compute the power in each frequency band.

We use the GNU Radio software development toolkit [2]. GNU Radio is based on the object oriented programming languages of Python and C++. Python programs connect the inputs and outputs of modules which are implemented in C++ code. By using Python, users can put together signal processing blocks (modules) with data from the USRP device as the source (input to the signal processing blocks) and direct the output of the signal processing blocks to a sink (a file to which data is written or a display on the screen). GNU Radio provides many Python modules for sampling a received radio signal and performing signal processing on that signal. The user may use these modules as they are, they may alter them to suit their needs, or they may write new modules if needed.

The module used to sense the spectrum in this work was `usrp_spectrum_sense.py`, a python block available from GNU Radio [2]. This module measures the spectrum power across a frequency range and directs the output to be written to a file. In addition, the user can specify details such as the band over which measurements are made, the number of times to measure the band, and the gain. A couple modifications were made to this module. The first modification set the number of bands measured to 89 for a specified bandwidth. The second modification was made to systematically generate the filename for purpose of data storage and retrieval.

3. EMULAB

We use the Emulab network testbed to perform our experimental measurements [3]. Emulab is a public facility at the University of Utah, maintained by the Flux Lab, containing many different types of nodes which are freely available to researchers to be programmed to emulate a wide variety of wireless and wired networks. The nodes range from software defined radio nodes (equipped with USRP devices) to mobile wireless nodes located on robots with full mobility. The nodes used for this project are USRP-equipped nodes. These nodes are deployed throughout the 3rd floor of the Merrill Engineering Building (MEB) at the University of Utah. To use these nodes an experiment is started by creating a new script (NS) file. Specifications are made in the file as to which nodes the user wishes to use and which programs they wish to run. By logging into many nodes at once, measurements can be made simultaneously at multiple locations. These measurements can then used to create a database. The USRP devices provided by the Emulab include the RFX900 transceiver daughterboard, capable of operation across the frequency range from 750 to 1050 MHz. However, the receiver is limited in frequency by its antenna, a "rubber duck" antenna with a 150 MHz bandwidth centered at 900 MHz. We made spectral measurements over the frequency band from 824 to 960 MHz. In addition, it was found by trial and error that a receiver gain of 50dB was best for receiving signals without saturating the receiver.

4. METHODS

In this section, we describe the experimental methodology for conducting a proof-of-concept test of cognitive geolocation. First, we describe how power vs. frequency spectrum data is collected using the USRP and the Emulab nodes. Then we describe the collection of the training set of data, which is used to fill the database. Next, we describe the collection of the test set. Finally, we describe our method for estimating receiver location using the test data.

A. Data Collection

The Emulab nodes' USRP receivers were programmed to take simultaneous measurements at various locations throughout the MEB. An NS file was created so that by running a command at the prompt, ten measurements would be taken at each node with a range of 824 – 960 MHz and a gain of 50dB. These measurements were organized in files distinguishable by the location and time of the measurement, which were recorded as part of the filename. Each measurement requires about 0.5 seconds, and the ten measurements are thus completed in about five seconds.

The `usrp_spectrum_sense.py` program operates by sequentially operating on narrowband received signals, and then increasing the center frequency. The i^{th} received signal is centered at frequency f_i. The receiver records a signal vector length 512, and then takes the FFT. The output contains the frequency representation of the signal, with samples separated by Δf = 3 kHz. We use the root mean squared (RMS) power across the first two samples from the frequency domain representation. This RMS power near the center frequency is thus used to represent the narrowband signal at band i. If $P_1(i)$ and $P_2(i)$ are the powers recorded at frequencies f_i and $f_i+\Delta f$, respectively, the RMS power $x(i)$ in frequency band i is given by

$$x(i) = \sqrt{(P_1^2 + P_2^2)/2} \qquad (1)$$

The choice of the statistic is arbitrary – any consistently applied measure of power vs. frequency band could be used to identify the position at which the measurement was taken. A total of 89 distinct frequency bands from 824 to 960 MHz, each with a bandwidth of 1.5 MHz, are measured in this manner.

B. Training Data

Training data is collected in order to represent the power spectrum at each location, at different times of day, and on different days of the week. Care is taken to ensure that an equal quantity of measurements is included in the training set for each location. For example, a database which included many more measurements from one location might influence a location estimation algorithm to be biased towards that one location. We also ensure that measurements at different locations are nearly simultaneous to ensure that bias due to temporal variations is not introduced. For example, if the spectral map for one location is created using measurements taken mainly before noon and the spectral map for another location is created using measurements taken after 3pm then there could be a bias in the location estimation algorithm which tends to select measurements from the same time of day, rather than the same location, when comparing a current measurement to those in the training set. By using simultaneous and an equal quantity of measurements from all measurement locations, we avoid these potential biases.

Using the Emulab nodes and the usrp_spectrum_sense, the training data is recorded at nine different locations throughout the MEB on a Sunday, Monday, and Tuesday the 18th - 20th of November of 2007 during the morning and afternoon hours. These three days were chosen for the following reasons. The days in the three groups Monday/Wednesday/Friday (MWF), Tuesday/Thursday (TT), and Saturday/Sunday (SS) will yield similar patterns of spectral usage when compared to other days in the same group because of MWF or TT class scheduling scheme at the University of Utah. For this reason, one day from each of these groups is sufficient to sample the day-of-week variation of spectrum measurements. Throughout these three days, measurements for the nine different locations were taken in 53 *sets*, and in each set ten measurements were taken for each location. Thus there are 530 total measurements for each location in the training set. A Matlab program then averages the 530 measurements for each location and thus created nine database files for nine locations.

C. Test Data

The database is then tested by taking measurements at night on December 4, 2007 and in the morning on December 5, 2007, from the hours of 9:30-11:15pm and 10:00am-12:30pm. Note that the nighttime test set is measured at a time which had not been incorporated into the database. Measurements are taken in ten different sets during each time period at the nine locations. Just as for the database, ten measurements are taken in each set for each location. A second Matlab program inputs the ten measurements for a given location in a given set and averages them. This average spectral measurement is then compared to the database to estimate the receiver location.

Location is estimated to be the location of the measurement in the database that matches most closely the current average spectral measurement. A match is quantified by a distance. The distance between two spectrum measurements is defined to be a distance between the decibel valued measurement vectors. Let $\mathbf{x}$ and $\mathbf{y}_k$ denote the current average spectral measurement and the measurement vector from location k stored in the database respectively. Specifically, $\mathbf{x} = [x(1), \ldots, x(N)]^T$ is the vector of averaged linear RMS received power in the spectrum in all N measured bands, and $\mathbf{y}_k = [y_k(1), \ldots, y_k(N)]^T$ is the RMS power measurement across bands in the database from location k. Then the distance $d(\mathbf{x},\mathbf{y})$ is given by,

$$d(\mathbf{x},\mathbf{y}) = \sum_i 10 \cdot \left|\log_{10} x(i) - \log_{10} y(i)\right| \qquad (2)$$

While many distance metrics are possible, the chosen metric accounts for the fact that changes in spectrum power are typically measured in decibels. In contract to the linear Euclidean (l_2) distance metric, the l_1 decibel distance takes into account that low amplitude signals can also play an important role in identifying the spectral characteristic of a location and should not be overwhelmed when higher power signals also exist. In radio propagation, losses are multiplicative and thus are additive in the dB scale. The chosen distance metric deals with additive losses appropriately by increasing distance linearly for each additive change in the dB power.

Using this distance metric, the database algorithm is as follows: for a new measurement $\mathbf{x}$, the location k associated with measurement $\mathbf{y}_k$ with the lowest distance $d(\mathbf{x},\mathbf{y}_k)$ is estimated to be the radio's current location:

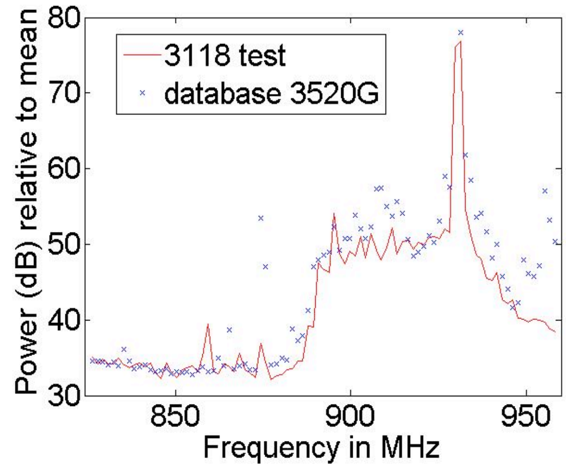
$$\hat{k} = \arg\min_k \; d(\mathbf{x},\mathbf{y}_k) \qquad (3)$$

This estimate $\hat{k}$ can then be compared with the actual location to determine the correctness of the estimate.

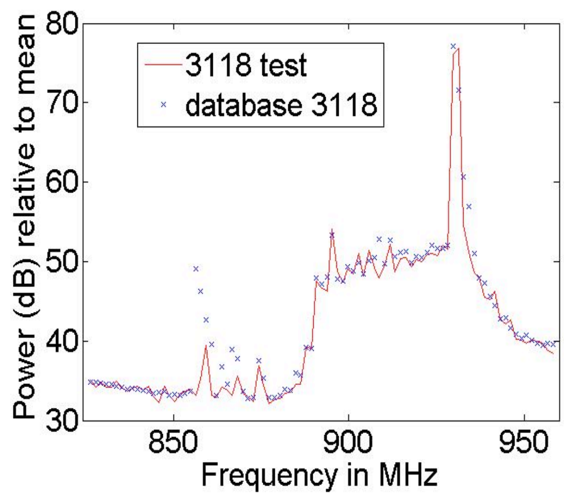
5. RESULTS

With this method, it was found that the correct location could be determined about 80% of the time at night and 91% of the time during the day. Note that the database was created using only daytime measurements. It was also observed that whenever location was determined incorrectly, the correct location had the second smallest distance $d(\mathbf{x},\mathbf{y}_k)$ and was not much larger then the distance of the incorrectly determined location. Thus the high correct detection rate indicates that the spatial differences in spectrum are very significant compared to the temporal changes (over the course of the two weeks between the recording of the training set and recording of the test set).

Additional observations of spectral measurements led to the following conclusions. It was noticed that measurements taken at night lacked significant signal activity recorded during the day. For example, the large spike at 930 MHz that is present in Figure 1 and almost always present during daytime measurements was seldom present in measurements taken at night. Big differences in the spectrum such as this lead to the conclusion that having separate spectral maps for day and night would result in greater localization accuracy. Furthermore, from this conclusion it is likely safe to say that further refinement of a database by creation of separate spectral maps for different hours of the day, different days of the week, and different times of the year would be beneficial for determining location.



(a)



(b)

Figure 1: Spectrum measurement comparisons where test measurement is compared to a database measurement recorded (a) at a different location, and (b) at the same location.

6. FUTURE RESEARCH

The experiment presented in this paper provides a proof of concept for the concept of using spectral occupancy as a means for location identification. There are opportunities for performance improvement for cognitive localization as well as opportunities for more extensive and realistic experimental tests.

In particular, solely using power in the spectrum is likely not the solution which achieves best localization performance. In particular, we do not distinguish between transmitters when measuring the power in spectrum, because many transmitters may be sharing individual channels. Future work should identify the transmitter, for example by MAC address or other identifying number, so that the power in a band can be further isolated by source. This would truly be the extension of WiFi RSS fingerprint-based localization systems to a wider range of sources.

Furthermore, the l_1 distance calculated on decibel power in (2) is likely not the optimal choice for distance. For best performance, we should weight the bands and the sources that are most reliable for purposes of localization. Certain sources may be only infrequently active, and thus may confuse the algorithm more often than they help. In addition, some bands are completely inactive and contribute only additive noise, which does not contribute any useful information in determining location. It would therefore be best to ignore such bands when determining location.

Future research could also involve speeding up database searches. When large-scale systems are deployed, the large quantity of measurements may require better search algorithms to speed up database lookup. Characteristics of the spectrum, unique to each smaller area, could be used to determine the general area in which the receiver is located, and then more quickly determine the position within that area. If such imprints can be made for a general area then the process of determining exact location could be sped up dramatically.

We also note that more extensive measurement campaigns could be conducted. The ability to make simultaneous measurements is a benefit of the Emulab network; but a rigorous test of cognitive geolocation could be done with a measurement campaign which measured spectral content with a very high density, for example, a few measurements per room in the building. Such a dense map would allow the study of the characteristics of spectral occupancy as a spatial random field. It could also be used in a real-time deployment system which tested cognitive geolocation over a long period of time, for example, months of operation. We note that the same USRP device in one location which made the training measurement also was the device which made the test measurement. Thus it is impossible to separate the device characteristic from the location characteristic in our test. A rigorous test experiment would use different devices for training and test.

7. ACKNOWLEDGEMENTS

Neal Patwari's research is supported in part by NSF CAREER Award ECCS-0748206. Arash Farhang's research is supported by an assistantship from the University of Utah Undergraduate Research Opportunities Program (UROP). Thanks to Jim Gaines for assistance with the setup and advice on operation of the USRP and GNU Radio software. Major thanks to David Johnson of the Flux Group at the University of Utah for his help setting up and debugging the experiments run on the Emulab network, without which the experiments could not have been run.

8. REFERENECES

[1] Ettus Research LLC, [Online]: http://www.ettus.com/

[2] E. Blossom, The GNU software radio, [Online]: http://gnuradio.org/trac

[3] B. White, J. Lepreau, L. Stoller, R. Ricci, S. Guruprasad, M. Newbold, M. Hibler, C. Barb, and A. Joglekar, "An Integrated Experimental Environment for Distributed Systems and Networks," in Proc. of OSDI 2002, Boston, MA, December 2002, pp 255-270.

[4] M. Rabinoqitz, J.J. Spilker Jr., "A new positioning system using television synchronization signals", *IEEE Trans. Broadcasting*, vol. 51, no. 1, March 2005, pp. 51-61.

[5] P. Bahl and V.N. Padmanabhan, "RADAR: an in-building RF-based user location and tracking system", in Proc. INFOCOM 2000, vol. 2, pp. 775-784, March 2000.

[6] T. King, S. Kopf, T. Haenselmann, C. Lubberger, W. Effelsberg, "COMPASS: A Probabilistic Indoor Positioning System Based on 802.11 and Digital Compasses," in Proc. WiNTECH'06, Sept. 2006.