

# Robust Coverage and Performance Testing for Large-Area Wireless Networks

Caleb Phillips, Russell Senior, Douglas Sicker, and Dirk Grunwald

University of Colorado, Computer Science Department, Boulder, CO, 80303, USA  
caleb.phillips@colorado.edu, russell@unwirepdx-watch.org,  
{sicker,grunwald}@cs.colorado.edu

**Abstract.** As large-scale wireless networks continue to proliferate, a reliable way to test coverage and communicate requirements becomes increasingly important. In this paper we discuss concerns and provide guidelines to consider when developing a coverage testing methodology for large-scale wireless networks. We propose a method which complies with these guidelines and apply it to a large municipal mesh network in Portland, Oregon. This approach is at the same time simple, cost-effective, and rigorous. We use commodity hardware to perform both high and low-layer tests at a random sample of points. Our method provides insights into the “expected” performance and coverage of the network. We find that a greater density of nodes is required in Portland to provide the required level of coverage, but that at covered areas, the performance is within specification. We are able to make these extrapolations with high statistical confidence, on a large network, using only 53 measurement points and a single measurement device which cost less than \$200 USD to build.

## 1 Introduction

Over the past several years more and more cities, townships, and institutions have been deploying large-scale wireless networks. On the largest scale, combination infrastructure and mesh networks are being used in municipalities to cover very large areas [1,2,3]. Many such deployments have been fraught with controversy around deployment motivation, performance expectations, and business model [4]. Our position is that with a robust and rigorous coverage testing methodology, many of these controversies and unfulfilled expectations are mapped into a clear and quantifiable problem and solution space. Indeed, the best way for a municipality to ensure that expectations are met is to be clear about the coverage and performance criterion of the network, and to ensure that this is tested in a thorough way.

To date, most coverage and performance testing of large networks is carried out by contractors (e.g., [5,6]) who use proprietary and sometimes non-rigorous techniques to perform their tests. The authors of this paper devised a testing methodology while doing an independent analysis of the municipal wireless network in Portland, Oregon. This methodology not only comports with Occam’s

razor, it is based on low-cost and readily obtainable commodity hardware. Additionally, all techniques are passive, requiring no more access to the network than any casual observer would have. Because the methodology is simple and the hardware inexpensive, it may even be possible for some testing to be carried out by institutions and municipalities themselves. At the very least, simple and well defined approaches to coverage testing will serve to encourage transparency in the testing of contractors, which will go a long way to making results easier to interpret and validate.

Additionally, coverage and performance testing has numerous uses outside of contractual verification. Using the results from tests, network operators can create maps that accurately reflect the *usable coverage* of their network. Moreover, this testing can find gaps and problems in coverage which will help inform the deployment of large and complex networks.

In the next section we give an overview of related work. In section 3 we'll discuss the approach we take and the method we have developed. Section 4 will discuss the application of the method to a large municipal wireless mesh network in Portland and section 5 provides a summary and conclusions.

## 2 Related Work

To date, large-scale wireless networks have mostly been the domain of commercial vendors. As such, the technical literature contains scarce analysis. The work of Tang et al. in [2] serves as one of the few examples of an academic analysis of a large commercial wireless network. Due to limitations on the dataset, the authors are limited to offering some high-level conclusions about usage patterns.

Some researchers have looked at mapping of access points by using passive sniffing techniques while driving (colloquially called “stumbling” or “war-driving”). In [7], Byers and Kormann provide a good overview of AP mapping and in [8], geography researchers provide their mapping technique for the unplanned networks of Salt Lake City, Utah. Although such information about unplanned networks is compelling ([9] presents a scheme for architecting ubiquitous connectivity using such networks), seldom is completeness, or statistical significance considered.

The vast majority of related work has considered the planning phase of network deployment. Software such as [10,11] have been used successfully by cell network engineers and amateur radio operators. This software makes use of elevation data and RF propagation models (a Longly-Rice model is used in [10]) to make inferences about coverage. While this approach may be effective for long-range radios with high vantages, it simply does not scale down to shorter-range IEEE 802.11x mesh networks in urban environments exhibiting complex fading. It should be noted, however, that technologies such as IEEE 802.16 (WiMax), which are also being considered for municipal wireless applications, might be able to safely make use of this software for coarse-resolution planning.

More examples of work in wireless network planning include [12], where the authors propose an AP-placement algorithm which uses a measured ray-tracing

model as input. The derivation of the ray-tracing model is not described in detail, and in any case, this approach is unlikely to scale to large networks – developing a ray-tracing model on the scale of a city is simply infeasible. In [13], Rodrigues et al. plan for an indoor wireless network. They divide the “demand area” into small spots. The spots are then grouped, and measurements are made at each spot-group. The criterion for size, grouping, and alignment of spots is not explained. In [3], Hills discusses the wireless network at CMU and argues for an iterative deployment process where coverage testing feeds back into deployment decisions. However, he uses signal strength alone for coverage mapping, which we will argue is insufficient. Finally, in [14], Dartmouth researchers map one access point using a GPS-enabled PDA, but do not explain how they choose where to make measurements.

### 3 Method

In this section we will introduce our proposed method for coverage testing of large-scale wireless networks and address concerns about accuracy and appropriateness. Overall, our method is focused on the goal of accurately inferring the usability of the network from the perspective of a typical client.

#### 3.1 Considerations for Radio

The vagaries and complexities of the wireless medium require that measurement strategies are approached carefully. We want to make experimental assumptions that are enlightened with respect to both the properties of RF propagation [15] and of infrastructure wireless networks [14].

**Signal Strength Alone Is Not Enough.** We argue that neither received signal strength (RSS), nor signal to noise ratio (SNR) are appropriate measures of link quality [16], and hence, form a poor basis for inferring about usable coverage. If one wants to use distance, SNR, or any other variable alone as a single-value indication of link quality, a relationship should be experimentally derived based on the appropriate environment and the equipment. If this is done with acceptable thoroughness, it may produce coverage extrapolations that are acceptable using this value alone.

The rationale behind this reasoning is based on the premise that bidirectional communication in wireless networks requires a symmetric concept of a link: *just because a client device can hear an access point does not guarantee that the access point can hear the client device* [14]. In practice, wireless access points are often much more powerful than wireless clients. A typical outdoor access point may include a 400mW radio connected to a high-gain antenna, resulting in an equivalent isotropically radiated power (EIRP) as high as  $4W^1$ . In comparison,

---

<sup>1</sup> The Skypilot-brand radios used in Portland, Oregon, for instance, have a transmit power of 400mW and a 7.4dBi omnidirectional antenna, resulting in an EIRP of 2.2W (33.4 dBm).

a common client device might have a 30mW radio attached to a meager antenna (2-5 dBi is common in our experience) providing an EIRP of closer to 17.8 dBm (60.26mW). Although the AP's antenna will provide gain on receive as well as transmit, this cannot make up for the clear asymmetry in power and sensitivity of the two devices, which results in many situations where a client device can see a strong signal from an AP, but is unable to get its communications back to the AP<sup>2</sup>.

**Environmental Diversity.** As discussed in [15], the quality of a wireless signal can vary substantially due to the location and the characteristics of the environment in which it is measured. Due to this, any scheme which purports to quantify the performance or coverage of a wireless network must give careful consideration to where measurements are made so that they do not skew the results in one direction or another. Additionally, it is not safe to use information drawn from one wireless environment to make conclusions about another – any such extrapolations should be treated with extreme skepticism.

**Variation in Hardware.** Wireless networking hardware varies greatly. Principally, variations in receiver sensitivity, transmit power, and antenna gain are most troublesome. Any equipment used in testing should be convincingly representative and should be carefully calibrated. If non-representative hardware is used, then a normalization procedure should be adopted and independently confirmed. Because normalization is non-trivial, we believe the easiest approach here is to use representative hardware.

**Application Layer Testing.** Ultimately, we believe the best way to model the usability of the network is to approach problems with the perspective of real use cases. This means that when we do a point-test of network quality we gain the most by doing application layer tests, such as throughput and latency testing in addition to low level tests (such as signal strength and noise level). Ideally, the endpoint for such tests would be very near the endpoint of the network to remove effects from outside the network.

### 3.2 Sampling Design

For a small network, it may be feasible to measure the entire expected coverage area. However, this quickly becomes intractable for larger networks. Choosing an appropriate statistical sampling design is crucial to draw a useful conclusion from the results. Although there are many approaches to spatial (sometimes called regional) statistical sampling, not all are appropriate for the problem.

**Placing Samples.** *Simple Random Sampling* (SRS) (see figure 1a) is the classic approach to this type of problem. It is simple, straight-forward, and well understood. It also applies well to spatial sampling problems. In particular, SRS

---

<sup>2</sup> This is especially a concern in the case when a user is indoors and the access point is outdoors, in such cases it may simply be impossible to achieve high quality of service without using a more powerful radio on the client-side.

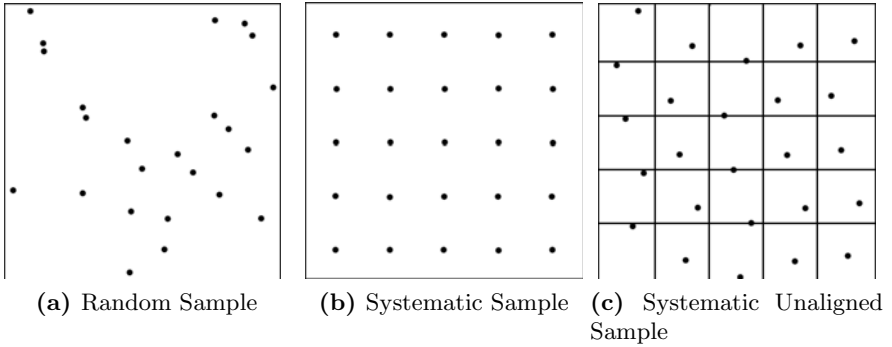


Fig. 1. Examples of sampling schemes

is convenient in that any consecutive subset of a simple random sample is also a simple random sample itself. This means that one can create a sample of points, and then test them consecutively up until a statistical significance criterion is met. For these reasons, we believe that SRS is an excellent choice for the problem at hand.

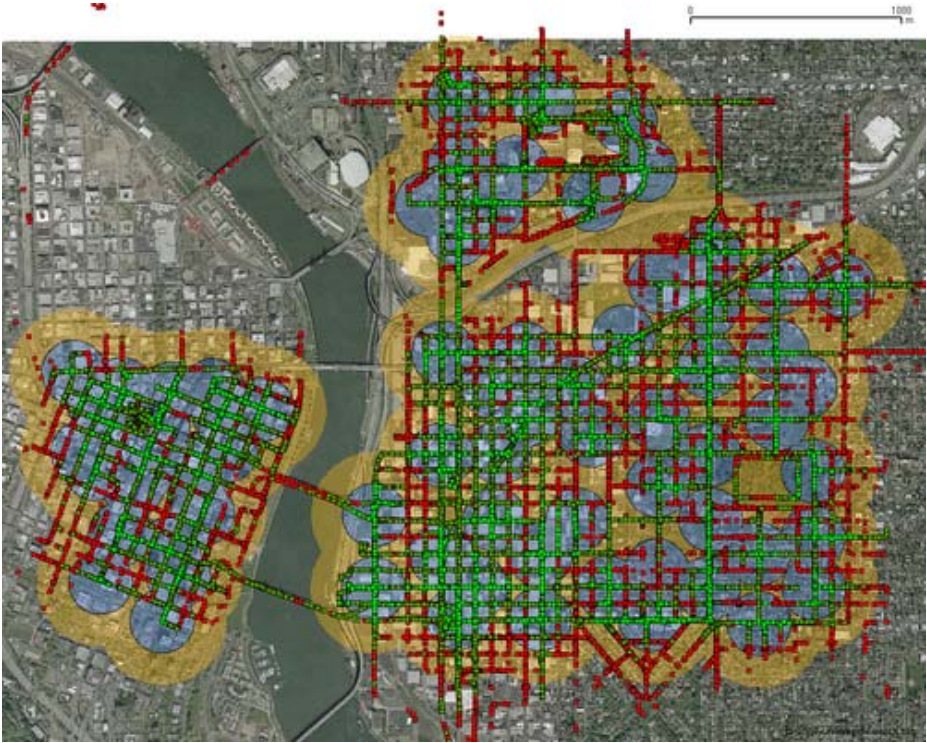
Competing sampling schemes might include *systematic*, *systematic unaligned*, or *stratified*. Systematic sampling (see figure 1b) can be dangerous because it risks alignment bias. It is admittedly tempting, when measuring coverage, to align samples along an obvious geographic feature of cities - streets. However, we argue that aligning tests along streets has the capacity to highly bias results. Aside from degenerating to a type of one-sided stratified sampling, streets also have the capacity to act as RF wave-guides (sometimes called “street canyons” in the literature). Systematic unaligned sampling (see figure 1c) can be a good compromise between SRS and systematic as it is more robust to alignment bias, but guarantees an even distribution of sample points among the test area. Stratified sampling is typically used when there are differences and/or differences in variability in different areas. For instance, a municipality may wish to prioritize or set different performance and coverage criterion for different areas of the city.

**Dealing with Unreachable Points.** It is inevitable that when testing sample points in any well designed spatial sampling scheme, some points will not be reachable. They might, for instance, be in the middle of a free-way, or a river, or on private property. These points should be measured on a best-effort basis as close to the original sample point as possible and the deviation should be carefully documented. Often, an assumption in spatial sampling is that values at geographically close points are similar. While the wireless medium is highly variable, with the exception of extreme shadowing scenarios, it is unlikely two close points will differ substantially in coverage. Hence, making a best-effort measurement in some small set of pathological cases is unlikely to significantly bias results. In the case that it does, careful documentation will be rewarded.

**Sample Size.** The required sample size for a certain confidence interval is dependent on the variability of the results. If an SRS is used, points can be tested up until the confidence interval narrows to the desired value.

#### 4 Case Study: Portland, Oregon

In September of 2005, the City of Portland, Oregon issued a Request for Proposals to build and operate a “citywide broadband wireless system”. In April of 2006, the City chose MetroFi (Mountain View, CA) as the winning bidder, and in the following summer the City and MetroFi signed a Non-exclusive License Agreement. Thereafter, MetroFi began to deploy their network in preparation for a December 2006 launch of a Proof-of-Concept (POC) network, as called for in the agreement. The deal was structured such that the POC network would first be built and afterward an independent third party would test it. When the City was satisfied that the POC network met its performance criteria, it would issue a Certificate of Acceptance.



**Fig. 2.** Signal strength from APs in the Proof-of-Concept area. Lighter dots (green) indicate stronger signal.

Two of the authors of this paper participated in a bid for this testing, but were not selected. Despite this, on January 11, 2007, they chose to do the testing anyway for their own edification and for the benefit of others<sup>3</sup>.

#### 4.1 Locating Access Points

Because our tests were carried out without any access to the network infrastructure, our first task was to locate the access points in the POC area. To this end, we drove to every publicly accessible street, collecting signal strength measurements using a battery-powered embedded computer with an external 7 dBi omnidirectional antenna and a GPS device. Figure 2 plots the measured signal strengths. We used this data to triangulate the position of the APs. Not surprisingly, as other researchers have shown that signal strength is poorly correlated with distance [16], we were unable to reach a satisfactory level of precision. To obtain the desired precision, we used triangulation information to locate each AP and then took a reading with a hand-held GPS device directly under the AP.

#### 4.2 Sampling Design

Because it is straight-forward, simple, and incurs little risk of alignment bias, we chose to use a SRS of points. From the list of 72 MetroFi access points that we considered to be in the POC network, we constructed a bounding box in latitude and longitude<sup>4</sup> extending 1000 feet beyond the extremities of the access point locations. Because we expected that many locations in the bounding box would fall outside of the POC areas, and because we were not certain how many locations we would be able to measure, we computed an excessive sample of 1001 locations using a random number generator such that each location in the bounding box had an equal probability of being chosen. Locations not within 1000 feet of an access point were immediately excluded. Each remaining location was plotted against orthoimagery using Google Maps. If the location fell in the Willamette River, was inside a building, or was not practically reachable, it was also excluded.

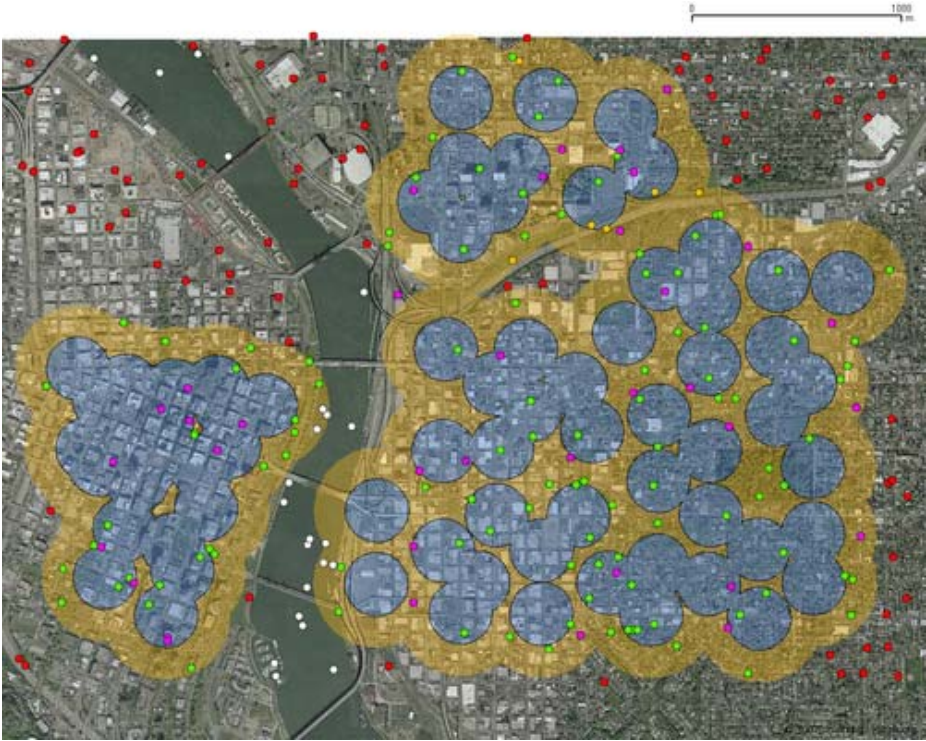
Ultimately, the first 250 locations in our sample of 1001 were either excluded on the basis of the criteria above or were visited and measured (see Figure 3). We chose to stop at 250 points after finding that this well bypassed our needs in terms of statistical power, both in the POC and overall.

#### 4.3 Test Device and Procedure

To act as a coverage point-tester, we combined a low-cost single-board computer (a Netgear WGT634u router) with a reliable Linux-based firmware (OpenWRT GNU/Linux), a lithium-ion battery, USB GPS receiver, and USB compact-flash

<sup>3</sup> All of our data, software, and configuration for the tests is available for public inspection and reuse, as well as our full report, at <http://unwirepdx-watch.org>

<sup>4</sup> All latitude/longitude coordinates are with respect to the WGS84 ellipsoid, unless otherwise noted.



**Fig. 3.** Random locations and their categorization. Green (light-grey) dots were tested, purple and orange (grey) were points within the POC that were excluded because they were inaccessible, and red (dark-grey) were excluded because they were not within the POC.

storage. Figure 4 shows the device in-situ. In addition to the mandatory components, we also used a USB sound-card and a pair of small speakers to “speak” status updates along with a small bluetooth USB dongle which was used as an “enable key”<sup>5</sup>. All together, this testing apparatus cost less than \$200 USD to build. Additionally, the Atheros 5213 802.11b/g radio and attached 2 dBi omni-directional antenna fulfill our requirement that the testing apparatus be representative of a typical client device. The test device was rigged to be free-standing at 6-feet off the ground so that the operators would not interfere with the measurements. When enabled, the test device was programed to carry out a series of tests. The outline of the testing algorithm is in table 1.

We use the standard Unix tools `ttcp` to test upstream throughput, ICMP ping to test latency and loss, and `wget` to test downstream throughput. A small script was used to bypass advertisement traps. We also found it necessary to use several watchdog scripts to check for a lost association, GPS issues, and stalled tests (for

<sup>5</sup> We conducted a small test to prove to ourselves that this bluetooth device was not radiating (and thus causing interference with the test device) when used this way.





**Fig. 4.** Testing apparatus. A battery-powered Netgear WGT634u wireless router outfitted with a GPS device, USB storage, speakers, and an enable key.

example, `tcp` has a tendency take a very long time on unstable connections). Depending on the results, a random location test might take anywhere from about 60 seconds (the length of time we would wait for an association) to around 7 minutes. In addition to these steps, we also recorded GPS position and time-stamp throughout the test.

The results of each test were stored on the USB storage device. At the conclusion of the tests we retrieved and analyzed the results. In our analysis we categorized each visited location according the states in table 2. Note that this is essentially a binomially distributed Bernoulli trial - states 1 to 5 indicating failure and state 6 indicating success. Hence, we can use classic binomial hypothesis testing to analyze the results.

#### 4.4 Results

Our first task in analyzing the results from the coverage tests is to infer a coverage percentage and a confidence interval for this inference. Figure 5 shows the P-value for an exact binomial test as the radius of points from the nearest AP changes and the hypothesized coverage percentage changes. Notice that the we reject any area where the P-value is less than  $\alpha = 0.05$ , which is essentially all of the combinations outside the prominent “ridgeline”. In effect, the width of

**Table 1.** Point-testing procedure

1	Disassociate
2	Try to associate with an access point for 60 seconds
3	Record information about the physical layer (BSSID, Signal, etc.)
4	Try to obtain a DHCP lease by sending up to 10 DHCP requests
5	Attempt to pass traffic to the Internet, if unable, bypass the captive-portal
6	Test latency and loss using ping
7	Test downstream throughput with a 1MB file, and a 5MB file
8	Test upstream throughput using ttcp
9	Store the contents of the ARP table
10	Store some statistics about our test device (memory and CPU utilization, etc.)
11	Perform a traceroute to an internet host to record routing topology

**Table 2.** Point-test state categorization

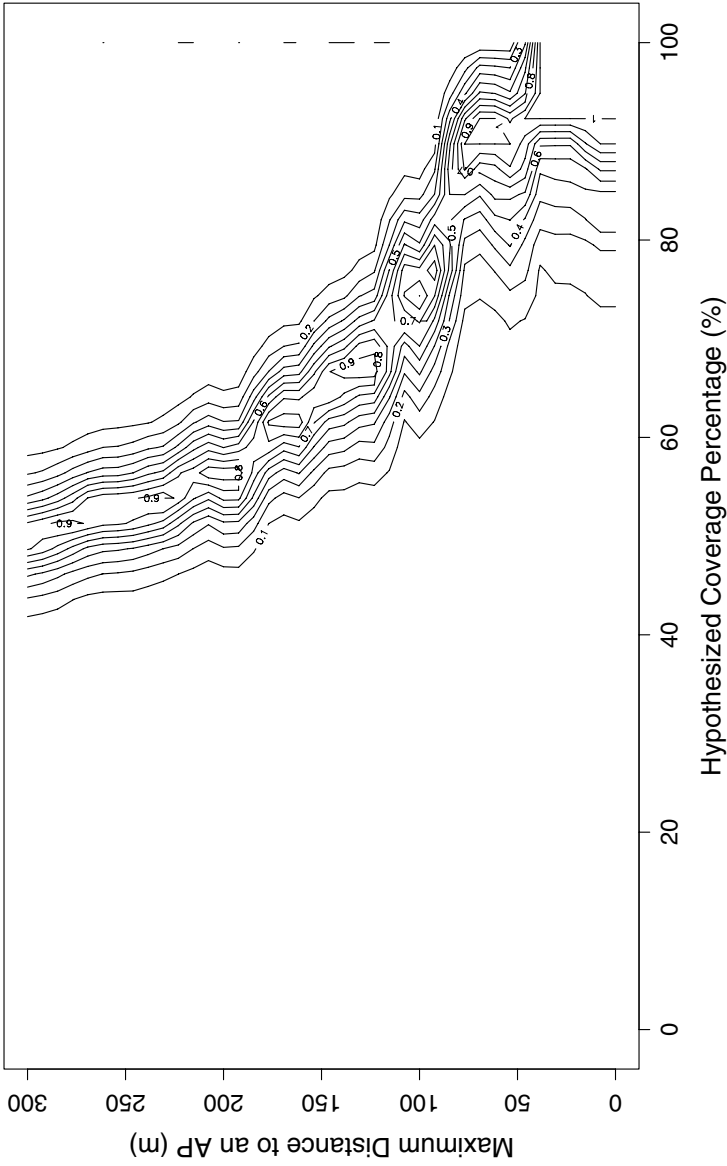
1	Could not associate
2	Lost association mid-test
3	Could not get a DHCP lease
4	Could not pass traffic
5	Performance below specified
6	Success

the ridgeline at any radius provides the 95% confidence bounds for the coverage percentage. For instance, we can see that at 150 meters, we have acceptable P-values only between about 50% and 70%. The city of Portland contract required 90% coverage within 500 feet (approximately 150 meters) of each AP. The measured percentage covered was 44.4% overall and 63.46% within the 500-foot radius. The probability of the coverage requirement being satisfied given the overwhelming evidence against it is one in 4,451,872. According to this map, the only radii which can achieve a coverage criterion of 90% are 50 meters or less (where the P-value is near 1).

It should be noted that this value, 44.4%, indicates that less than half of *locations* within the coverage area are expected to be able to achieve a connection at the performance required by the contract. Additionally, if we include poor-performing locations, we can say that at a 95% confidence level, the percentage of locations achieving *any connection* is between 36.08% and 54.77%. From the perspective of municipalities hoping to deploy a wireless network for the purpose of automated meter reading and other such applications, these numbers are fairly dismal and further serve to highlight the fact that it is essential that requirements are well specified and tested to ensure that both the needs of the network operator, and that of the institution or city are met.

Although the network in Portland does not meet the coverage criterion defined in the contract, we are not sure this coverage criterion was formulated in the best possible way. Instead of defining an arbitrary proof-of-concept area as a certain radius from each AP, we suggest that a more useful metric would be to define a (more conservative) percentage goal for the entire region to be covered.

### P-Values for Exact Binomial Test



**Fig. 5.** Contour map of P-values for an Exact Binomial test as a function of maximal distance to AP (i.e., only concerning samples within some radius) and hypothesized coverage percent. P-values below  $\alpha = 0.05$  reject the null hypothesis that the hypothesized coverage percent is possible given the observations.

Additionally, the contract should be straight-forward about the way this coverage will be tested in terms of sampling and performance goals. In the case of the network in Portland, at 44.4% it is still very low, indicating that the network operator should seriously consider increasing access-point density. Moreover, since this testing was conducted exclusively outdoors, it can be at best looked at as an extremely liberal estimate of indoor coverage.

Interestingly we find that signal strength is normally distributed among points where we could associate. A Shapiro-Wilkes test gives a p-value of 0.297 - unwilling to reject the null hypothesis that the samples are normal. Overall signal is highly variable among those points where we are successful, providing a mean value of -63.058 dBm and standard deviation of 9.625 dBm. Among those points where we could associate but failed somewhere upstream, the mean signal strength is -77.125 dBm with a standard deviation of 5.793.

State and signal are reasonably linearly correlated, showing a correlation coefficient of 0.47. This correlation is very strong if we assume signal strength -95 dBm (essentially, the noise floor) for those trials that failed to associate (the coefficient is 0.90 in this case). Distance, however, is not linearly correlated well with state or signal (correlation coefficient is -0.36).

For those tests that were successful, we collected information about the performance of the network. Averaging across the random sample allows us to provide an “expected view” of performance for those locations with a usable connection.

## 5 Conclusion

In this paper we have outlined a simple, but powerful, method for coverage and performance testing of large-scale wireless networks. Our proposed method utilizes a random sample of points within the coverage area to make inferences about *usable coverage* and expected performance. We argue that appropriate spatial sampling design, paired with a testing approach that both considers the perspective of the user and the complexities of the wireless medium is crucial for test results to be meaningful.

We applied this testing method to a large municipal wireless mesh network in Portland, Oregon and presented results from that study. As similar networks continue to proliferate, having a practical and effective method to test them is vital to their success and to achieving a rational way of communicating expectations.

## References

1. Kramer, R.D.J., Lopez, A., Koonen, A.M.J.: Municipal broadband access networks in the netherlands - three successful cases, and how new europe may benefit. In: AccessNets 2006: Proceedings of the 1st international conference on Access networks, Athens, Greece, p. 12. ACM Press, New York (2006)
2. Tang, D., Baker, M.: Analysis of a metropolitan-area wireless network. *Wireless Networks* 8(2-3), 107–120 (2002)

3. Hills, A.: Large-scale wireless lan design. *Communications Magazine*, IEEE 39(11), 98–107 (2001)
4. Urbina, I.: Hopes for wireless cities fade as internet providers pull out. *The New York Times* (22) (March 2008)
5. Edx wireless (March 2008), <http://www.edx.com/>
6. Uptown services (March 2008), <http://www.uptownservices.com/>
7. Byers, S., Kormann, D.: 802.11b access point mapping. *Commun. ACM* 46(5), 41–46 (2003)
8. Hsu, J.: Wi-fi cloud hovers over salt lake city. *IEEE Spectrum Online* (February 2008), <http://spectrum.ieee.org/feb08/6025>
9. Sastry, N., Crowcroft, J., Sollins, K.: Architecting citywide ubiquitous wi-fi access. In: *Hotnets IV* (2007)
10. Magliacane, J.A.: Splat! a terrestrial rf path analysis application for linux/unix (March 2008), <http://www.qsl.net/kd2bd/splat.html>
11. Ericsson, “Planet ev” (March 2008), [http://www.ericsson.com/solutions/tems/network\\_plan/planetev.shtml](http://www.ericsson.com/solutions/tems/network_plan/planetev.shtml)
12. Kamenetsky, M., Unbehaun, M.: Coverage planning for outdoor wireless lan systems. *Broadband Communications, 2002. Access, Transmission, Networking. 2002 International Zurich Seminar on*, pp. 49–1–49–6 (2002)
13. Rodrigues, R., Mateus, G., Loureiro, A.: On the design and capacity planning of a wireless local area network. In: *Network Operations and Management Symposium, 2000. NOMS 2000. 2000 IEEE/IFIP*, pp. 335–348 (2000)
14. Kotz, D., Newport, C., Elliot, C.: The mistaken axioms of wireless-network research. *Dartmouth College of Computer Science, Tech. Rep.* (2003)
15. Anderson, J.B., Rappaport, T.S., Yoshida, S.: Propagation measurements and models for wireless communications channels. *IEEE Communications Magazine* 33, 42–49 (1995)
16. Aguayo, D., Bicket, J., Biswas, S., Judd, G., Morris, R.: Link-level measurements from an 802.11b mesh network. *SIGCOMM Comput. Commun. Rev.* 34(4), 121–132 (2004)