Chapter 4

Large Area Coverage Testing

.

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 [122, 217, 95]. Many such deployments have been fraught with controversy around deployment motivations, performance expectations, and business models [228]. One possible explanation for these issues, offered by this thesis, is a failure to understand and effectively communicate performance expectations of the networks. 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.

As a first step towards developing a statistically robust method for coverage mapping, this section approaches the sub-problem of "coverage testing", that is, making a strong statement about the *percentage* of area within a region that is covered by a given network. Coverage testing has its own important applications in contractual verification. Because rigorously solving the coverage mapping problem involves solving the coverage testing problem first, this chapter begins here. The method developed in this chapter relies on basic and well-established statistical methods, including a selection of a Simple Random Sampling (SRS) of points, and testing via the Binomial Theorem. Combining these robust statistical methods with a data

[0] Work in this chapter has appeared in [171, 168]. Data collected for the experiments in this chapter has been made publicly available at [170].

collection methodology that is appropriate for sampling the radio environment is able to produce a simple method for coverage testing that requires a *minimum amount of work* required to make an accurate statement about coverage.

Most coverage and performance testing of large networks is carried out by contractors (e.g., [239, 4]) who use proprietary and sometimes nonrigorous techniques to perform their tests. By comparison, the methodology presented here not only comports with Occam's razor, it is also 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.

4.1 Method

The 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 [183] and of infrastructure wireless networks [120]. This section outlines domain appropriate guidelines for coverage testing. The following section will apply these guidelines to develop a practical coverage testing methodology for a municipal wireless network in Portland, Oregon.

Signal Strength Alone Is Not Enough

Bidirectional communication in wireless networks requires a symmetric concept of a link: *just because a client device can hear an AP does not guarantee that the AP can hear the client device* [120]. In practice, wireless APs are often much more powerful than wireless clients. A typical outdoor AP may include a 400 mW radio connected to a high gain antenna, resulting in an equivalent isotropically radiated power (EIRP) as high as 4 W[1] . In comparison, a common client device might have a 30mW radio attached

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

to a meager antenna (2-5 dBi is common in our experience) providing an EIRP of closer to 17.8 dBm (60.26 mW). 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[2] . Therefore, Neither RSS, nor SNR are appropriate measures of link quality [20] alone. By themselves, they 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.

Environmental Diversity

As discussed in [183], 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 that 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. It might not be 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 nonrepresentative hardware is used, then a normalization procedure should be adopted and independently confirmed. In all likelihood the easiest approach here is to use representative hardware and avoid the onerous task of normalization.

Other Operators, Other Networks

Measurements of a live network must consider effects of other users on that network, and of interference from neighboring, but unrelated, networks. The former can be addressed by testing the network during

[2] This is especially a concern in the case when a user is indoors and the AP is outdoors; in such cases it may simply be impossible to achieve high quality of service without using a more powerful antenna on the client side.

a time when it is not in use. The latter is a concern that a network designer must address when they deploy their network (choosing channels to minimize cochannel interference, etc.).

Application Layer Testing

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.

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. Section 5.2.1 provides a discussion of classic sampling schemes and the tradeoffs involved in design selection. Because it is least likely to be aligned with sources of error and is easy to implement and put to use, SRS is the sampling strategy advocated here.

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 freeway, 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 of 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 that 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.

Temporal Variability

Because the behavior of wireless networks has been shown to be variant in time [20], long term tests are required to determine temporal variability in network performance. Ideally, these tests would run for as long as possible and the testing points would be distributed using the same random sampling technique used for coverage testing. However, long term testing introduces some logistical complexities that may require some compromises. For instance, the test device is likely to need AC power and a good vantage from which to test. It is unlikely an SRS will choose positions that are appropriate for long-term tests. As such, it may be necessary to deploy long term test devices in locations where the testers can acquire permission and access at the sacrifice of proper sampling design. In any case, given choice of locations, the more convincingly representative the subset, the more useful the results will be.

4.2 Case Study: Portland, Oregon

In September of 2005, the city of Portland, Oregon issued a Request for Proposals (RFP) to build and operate a "citywide broadband wireless system". In April of 2006, the city chose MetroFi (Mountain View, California) as the winning bidder, and in the following summer the city and MetroFi signed a nonexclusive 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. Specifically, the contract stated that the network should be able to provide a connection to at least 90% of the outdoor POC area (defined as all locations within 500 feet of the 72 APs). Further, for a given stationary connection, the network should support a 1 Mbps downstream/256 Kbps upstream throughput, have 99% availability, and a total within-network latency of 100 milliseconds. During this POC testing phase, an independent analysis of the network coverage was conducted.

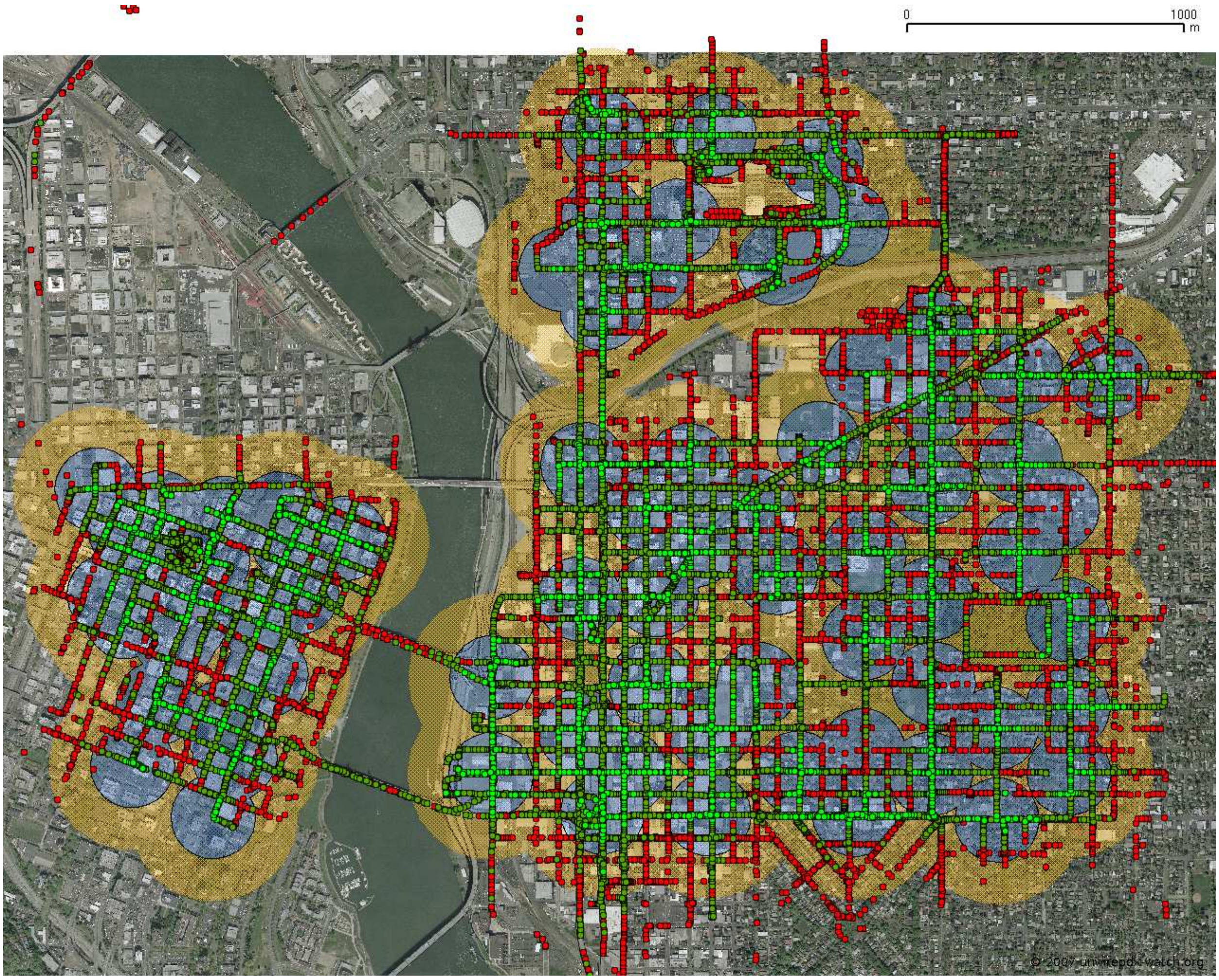

Figure 4.1: Signal strength from APs in the POC area. Lighter dots (green) indicate stronger signal.

4.2.1 Method

Because the tests were carried out without any access to the network infrastructure, the first task was to locate the APs in the POC area and obtain signal strength measurements over the entire area. To this end, every publicly accessible street was driven, collecting signal strength measurements using a battery powered embedded computer with an external 7 dBi omnidirectional antenna and a GPS device. Figure 4.1 plots the measured signal strengths. This data was used to triangulate the position of the APs. Not surprisingly, as other researchers have shown that signal strength is poorly correlated with distance [20], a satisfactory level of precision was unable to be obtained. To obtain the desired precision, triangulation was used to locate each AP, and then a reading was taken with a handheld GPS device directly under the AP. To compress this data set slightly, precision of GPS coordinates was truncated to five significant digits, which has the effect of grouping data points within a 0.74 m circle.

4.2.1.1 Sampling Metholodogy

From the list of 72 MetroFi APs that were considered to be in the POC network, a bounding box in latitude and longitude[3] was constructed extending 1,000 feet beyond the extremities of the AP locations. Because it was expected that many locations in the bounding box would fall outside of the POC areas, and because it was not certain how many locations we would be able to measure, an excessive sample of 1,001 locations was computed using a random number generator such that each location in the bounding box had an equal probability of being chosen. Locations not within 1,000 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 the sample of 1001 were either excluded on the basis of the criteria above or were visited and measured (see Figure 4.2). It was decided to stop after surveying 250 points because the results had sufficient statistical power at that point.

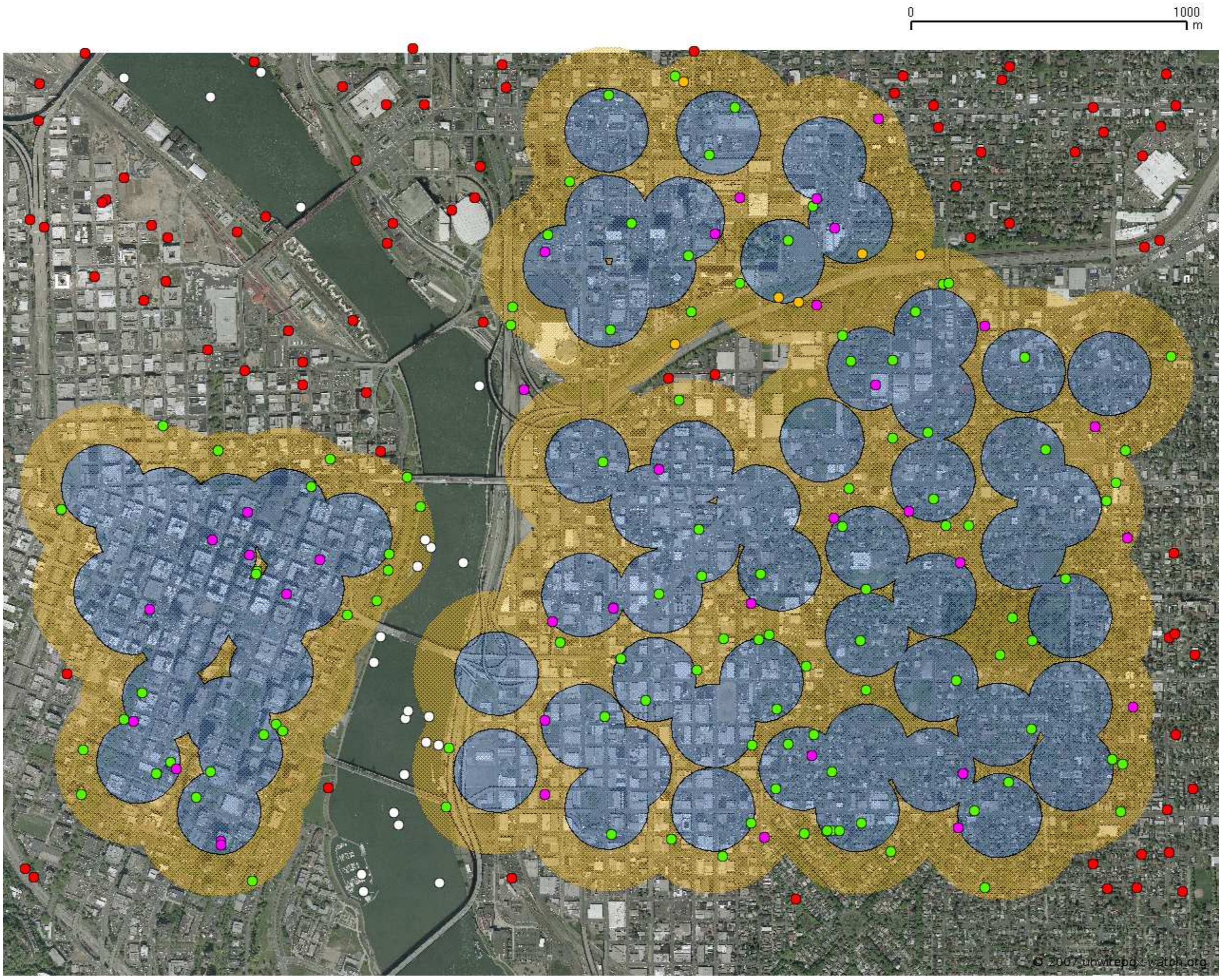

Figure 4.2: Random locations and their categorization. Green (light grey) dots were tested, purple and orange (grey) were points within the POC that were excluded because they were inaccessible, and red (dark grey) were excluded because they were not within the POC.

Figure 4.3: Testing apparatus. A battery powered Netgear WGT634u wireless router outfitted with a GPS device, USB storage, speakers, and an enable key.

4.2.1.2 Measurement Apparatus and Procedure

To act as a coverage point tester, a low cost single board computer (a Netgear WGT634u router) was combined with a reliable Linux-based firmware (OpenWRT GNU/Linux), a lithium-ion battery, USB GPS receiver, and USB compact-flash storage. In addition to the mandatory components, a USB sound card and a pair of small speakers were used to "speak" status updates along with a small Bluetooth USB dongle that was used as an "enable key"[4] . All together, this testing apparatus cost less than $200 USD to build. Additionally, the Atheros 5213 802.11b/g radio and attached 2 dBi omnidirectional antenna fulfilled the requirement that the testing apparatus be representative of a typical client device. The test device was rigged to be freestanding at six feet off the ground so that the operators would not interfere with the measurements. When enabled, the test device was programmed to carry out a series of tests. The outline of the testing

[3] All latitude/longitude coordinates are with respect to the WGS84 ellipsoid, unless otherwise noted.

[4] A small test was conducted using a WiSpy spectrum analyzer to test whether the Bluetooth device was radiating (and thus causing interference with the test device) when used this way. It was concluded that the bluetooth dongle does not emit noticeable radiation when it is not in use.

procedure is given in algorithm 1.

Algorithm 1 Point testing procedure

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

Standard Unix tools were used: ttcp, to test upstream throughput; Internet Control Message Protocol (ICMP) ping, to test latency and loss; and wget, to test downstream throughput. A small script was used to bypass advertisement traps[5]. It was also found to be necessary to use several watchdog scripts to check for a lost association, GPS issues, and stalled tests (for example, ttcp 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 that was waited for an association) to around 7 minutes. In addition to these steps, GPS position and time-stamp were also recorded throughout the test.

The results of each test were stored on the USB storage device. At the conclusion of the tests the results where retrieved and analyzed. In the analysis each visited location was categorized according to the states in table 4.1. By categorizing points by their success state in table 4.1, the set of test points can be treated as a binomially distributed Bernoulli trial—states 1 to 5 indicating failure and state 6 indicating success. Hence, classic binomial hypothesis testing can be used to analyze the results.

In addition to coverage testing, three more test devices were deployed for long term tests. As noted earlier, finding appropriate locations for long term testing poses some logistical challenges. Thus, the devices were positioned at the best locations permitted for use, and not at positions dictated by a simple random sample. At each location, continuous tests were run for a minimum of a week, collecting throughput, la-

[5] The "free" public MetroFi network was configured to inject banner advertisements into Hypertext Transport Protocol (HTTP) results and periodically redirect HTTP requests to a full-page advertisement. These advertisement traps, if not otherwise bypassed, would have interfered with the downstream throughput test, which involved downloading a fixed-size file over the HTTP protocol on port 80.

State	Description
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

Table 4.1: Point test state categorization

tency, and link-quality information. The hardware and test methods here are identical to those used for coverage testing.

4.2.2 Results

The first task in analyzing the results from the coverage tests is to infer a coverage percentage and a confidence interval for this inference. Figure 4.4 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 any area where the p-value is less than $\alpha = 0.05$ is rejected, which is essentially all of the combinations outside the prominent "ridgeline". In effect, the width of the ridgeline at any radius provides the 95% confidence bounds for the coverage percentage. For instance, at 150 meters, there are acceptable p-values only between about 50% and 70%. The 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 that can achieve a coverage criterion of 90% are 50 meters or less (where the p-value is near 1). It is worth noting that some of the results here differ substantially from those of the contracted company, [4]. A discussion of these differences is outside the scope of this thesis, but can be found at [169].

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 poorly performing locations i.e., locations providing a connection with slower throughput or higher latency than required by the contract, are included, it can be said that at a 95% confidence level, the percentage of locations acheiving *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, it is not clear that this coverage criterion was formulated in the best possible way. Instead of defining an arbitrary POC area as a certain radius from each AP, a more useful metric would be to define a (more conservative) percentage goal for the entire region to be covered. Additionally, the contract should be straightforward 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 AP density[6]. Moreover, since this testing was conducted exclusively outdoors, it can be at best looked at as an extremely optimistic estimate of indoor coverage.

Interestingly, signal strength is normally distributed among points where it was possible to associate with the AP. A Shapiro-Wilkes test gives a p-value of 0.297, i.e., unwilling to reject the null hypothesis that the samples are normal. Overall, signal is highly variable among those points that had successful connections, providing a mean value of -63.06 dBm and standard deviation of 9.63 dBm. Among those points where association was successful, but the test failed somewhere upstream, the mean signal strength is -77.13 dBm with a standard deviation of 5.80.

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 well linearly correlated well with state or signal (correlation coefficient is -0.36). Information about the performance of the network was collected at points that were successful. Averaging across the random sample provides an "expected

[6] The hardware vendor (SkyPilot [225]) claimed that this particular network was underdeployed relative to their recommendation[3].

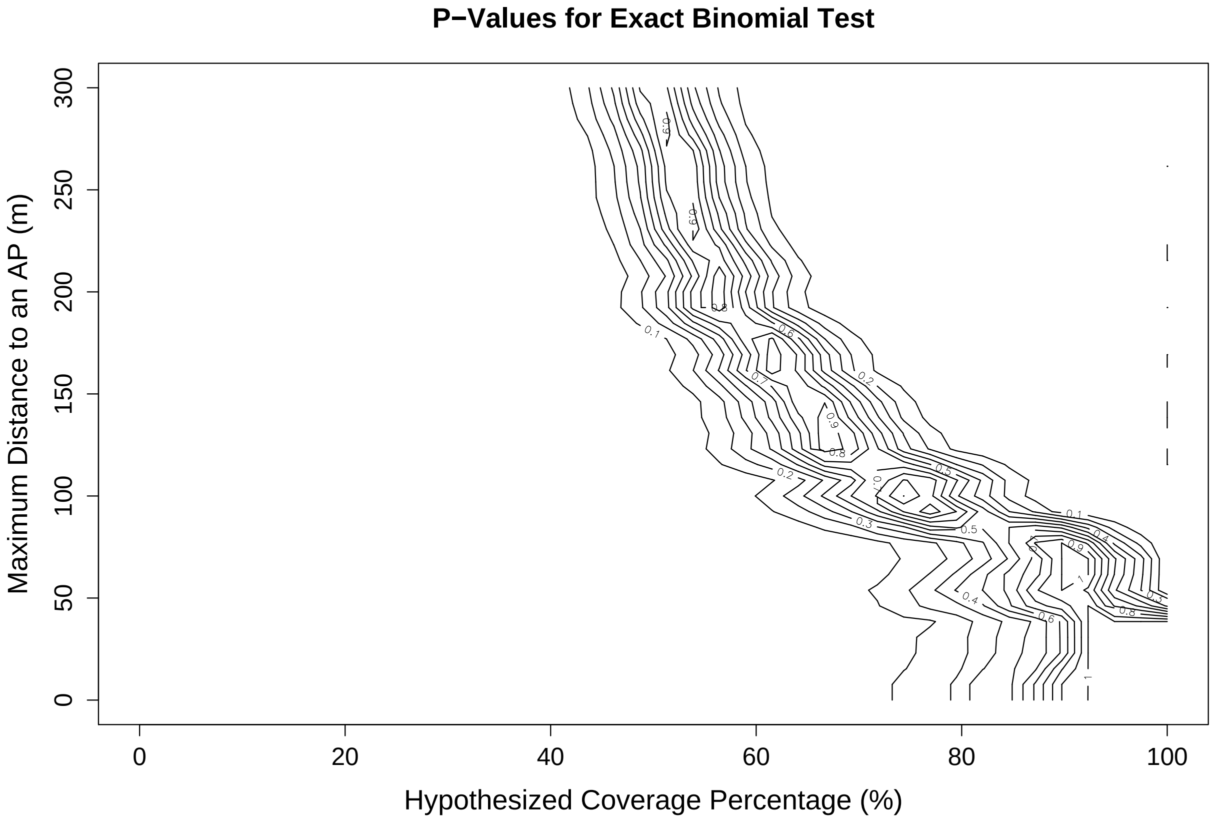

Figure 4.4: Contour map of p-values for an exact binomial test as a function of maximal distance to an AP (i.e., only concerning samples within some radius) and hypothesized coverage percent. p-values below $\alpha = 0.05$ reject the null hypothesis that the hypothesized coverage percent is possible given the observations.

Area (ft)	N	Down/Up Throughput (Kbps)		Latency (ms)		Loss (%)		Signal (dBm)	
		Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
<250	7	1672.4/373.28	1281.3/83.235	95.700	72.319	5.7143	15.119	-50.857	3.7607
<500	16	1508.7/373.42	1002.8/79.181	105.15	69.808	3.125	10.145	-57.938	8.4417
>1000	27	1437.2/370.52	875.72/74.682	97.459	59.344	3.33	8.77	-59.333	8.6425
>500	11	1333.1/366.30	657.23/71.159	86.273	40.182	3.6364	6.7420	-61.364	8.9249

Table 4.2: Random sample performance summary

view" of performance for those locations with a usable connection. These statistics are summarized in table 4.2.

Although the long term tests are not clearly representative because of logistical limitations, a large amount of continuous data at three locations within the POC area was still able to be collected. Site A was collected on the first floor of a house in a residential area, very close to an AP; site B was collected on the second floor of an office building on the edge of downtown Portland; site C was collected in the window of a fifth-floor office in the heart of downtown Portland. A summary of some of the statistics from these tests is in table 4.3. In terms of the performance requirements of the Portland network, all three locations passed.

One interesting observation, however, is that the performance can vary highly as a function of time of day. Figure 4.5 plots packet loss for site C as a function of time of day. Notice that site C, which was in a densely populated area (both in terms of people and wireless networks), exhibits large packet loss during the bulk of the typical business day. It is hypothesized that this is a result of internetwork interference. If nothing else, this plot should be yet another warning to network operators that interference from neighboring and third-party 802.11x wireless networks must not be neglected in the design and performance expectations of future networks.

Site	Duration (h)	Disassoc. Probability	Mean Percent Packet Loss	Percent Packet Loss Std. Dev
A	456.44	0.00149	1.562%	4.789%
B	173.83	0.00106	2.549%	7.418%
C	197.53	0.00449	33.031%	28.983%

Table 4.3: Summary of a selection of long-term test statistics.

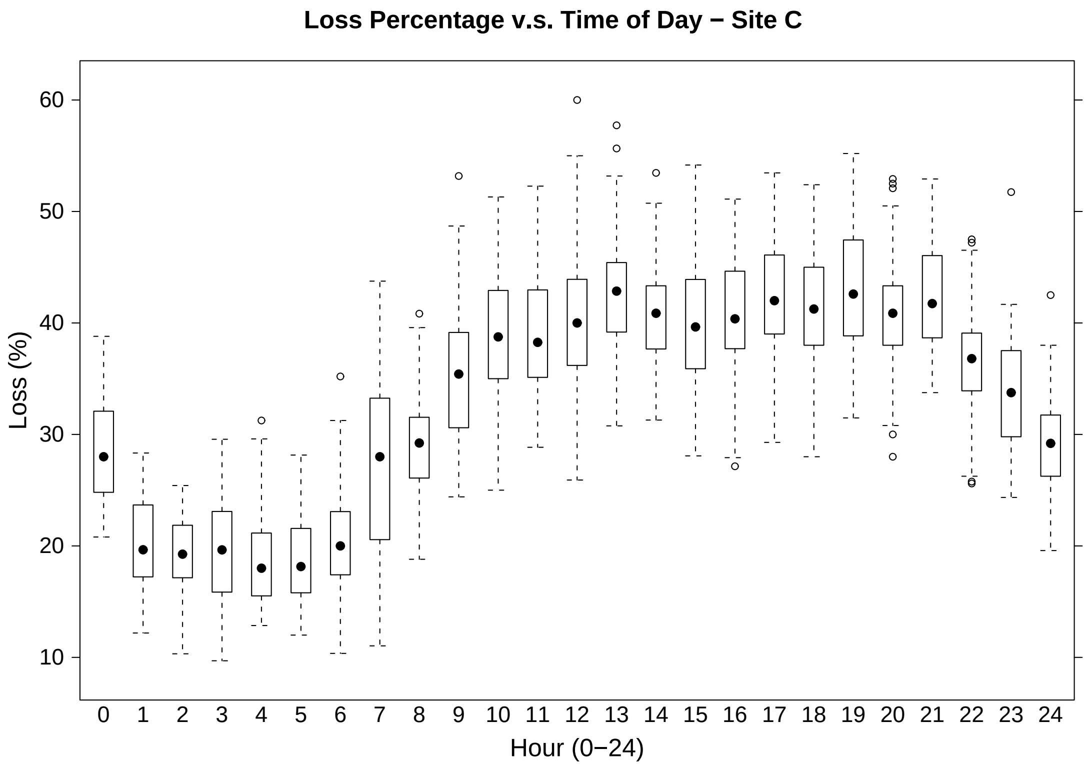

Figure 4.5: Packet loss for long term test at site C as a function of time of day. Measurements are averaged across days and bucketed per hour.

4.3 Discussion

This chapter has outlined a simple but powerful method for coverage and performance testing of large-scale wireless networks. The proposed method utilizes a random sample of points within the coverage area to make inferences about *usable coverage* and expected performance. For test results to be meaningful, it is crucial that an appropriate spatial sampling design be paired with a testing approach that both considers the perspective of the user and the complexities of the wireless medium.

This testing method was applied to a large municipal wireless mesh network in Portland, Oregon and results from that study have been presented. 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. The lessons learned in this study can be applied directly to developing appropriate methods for the larger problem of coverage mapping, the topic of chapter 5.