Addressing Repeatability in Wireless Experiments using ORBIT Testbed

Sachin Ganu, Haris Kremo, Richard Howard[1] and Ivan Seskar
WINLAB, Rutgers University, 73 Brett Road, Piscataway, NJ 08854
{sachin, harisk, reh, seskar}@winlab.rutgers.edu

Abstract

With the rapid growth in research activity on future wireless networking applications and protocols, experimental study and validation is becoming an increasingly important tool to obtain realistic results that may not be possible under the constrained environment of network simulators. However, experimental results must be reproducible and repeatable for them to be used to compare proposed systems and to build prototypes. In this paper, we address the issue of repeatability in wireless experiments in the Open Access Research Testbed for Next-Generation Wireless Networks (ORBIT) testbed[2] and propose a mechanism to promote reproducible experiments using periodic calibration of the equipment. Several experimental results that capture repeatability in time and space using our initial testbed setup are also provided.

1. Introduction

Widespread application of wireless networking is currently hampered by many issues including radio propagation, link reliability, and complexity of use and maintenance. Many of the underlying causes of these user problems are unique to wireless (e.g. the "hidden node problem", rapidly changing link quality, power control, and high bit-error rate) and can only be addressed by experiments using systems which incorporate realistic emulation of the wireless physical layer behavior.

Unfortunately, most of the work done so far is based on simulations that make simplifying assumptions and have limited real-world physical layer modeling capabilities. This often affects the quality of the results and also their reproducibility

Simulations often may provide repeatable results in wireless experiments; however they sometimes lack credibility [3,4] and may not truly represent the underlying phenomenon due to inaccurate real-world modeling. Such results may be inadequate to build working prototypes and test end-user applications under real-life conditions. On the other hand, experimental results, based on actual devices, do provide realistic results, but they are unusable unless they can be faithfully reproduced.

In the recent NSF-sponsored Network Testbeds Workshop Report [1], it was concluded that "open wireless multi-user experimental facility (MXF) testbeds" for wireless networking would be increasingly important to the research community in view of the limitations of available simulation methodologies and the growing importance of "cross-layer" protocol research. These considerations motivated the ORBIT testbed project [2] which aims to provide a flexible, open-access multi-user experimental facility to support research on next-generation wireless networks.

The key to success in the experiments on the ORBIT testbed is the ability to control and measure important network properties, such as transmit power, throughput, or error rate, accurately, reproducibly, and quickly enough to characterize complex systems.

In this paper, we intend to address the important issue of repeatability in experimental results obtained from the ORBIT testbed and present a few approaches to ensure that experimental results can be made reproducible. This is an important factor to consider for any testbed that uses commercially available hardware and does not have *anechoic* environments to

---

[1] Richard Howard is also Senior VP of Technology, PnP Networks.

[2] Research supported by NSF ORBIT Testbed Project, NSF NRT Grant #ANI0335244 and DARPA Contract NBCHC300016

guarantee RF isolation, which are expensive to provide for 400 nodes. We describe initial experiments on the ORBIT testbed designed to exercise its measurement and control capabilities at a basic level. Comparisons over time and space (using different hardware across the grid) are used to guide a calibration strategy that will assure the needed accuracy.

The organization of the paper is as follows: Section 2 describes the factors that influence repeatability in experimentation and some earlier proposed approaches to tackle this issue. Also, the results of card calibration are presented in this section. Section 3 discusses the experimental results obtained to capture repeatability of experimental results over a period of time and distributed in space. Section 4 concludes the paper and describes ongoing and future work to ensure repeatable experiments on the ORBIT testbed.

2. Parameters affecting repeatability

There are several factors in a wireless experiment that may affect the repeatability of experiments and reproducibility of results. First, the differences could be attributed to commercial hardware. This may be due to low-cost design constraints in commercial products intended for the Industrial, Scientific and Medical (ISM) band. Additional issues include broad tolerances and ageing for low cost components. Also, during the lifetime of the testbed, there may be a need to replace wireless cards that malfunction or are superseded by new models. This may lead to differences in experimental results over time. There is also a possibility of subtle software or firmware bugs that may manifest as inconsistent experimental results. Finally, for a wireless testbed, the environment poses the biggest challenge to repeatability due to uncontrolled interference over time and space. This could be due to interference from co-located infrastructure access points, movement of people, opening and closing of doors etc. In [5], the authors propose methods to reduce the effects of the environment by using cables instead of wireless links while, in [6], this approach is extended by using an emulator that can emulate different channel behavior. For the ORBIT testbed, however, we intend to retain the wireless link since this helps to capture some of the realistic wireless channel effects that may be lost by using RF cables or emulators. In this discussion, we address the issues that may arise due to hardware differences even across multiple devices from the same vendor.

2.1. Differences in reported RSSI across different cards supplied by the same vendor

RSSI (Received Signal Strength Indicator) is the primary measurement of the radio environment in which a wireless network card is operating. Unfortunately, it is a poorly described and understood parameter and, as such, is of limited utility in network testing. This is because the IEEE 802.11b standard [8] (Section 14.2.3.2) does not impose any restriction on how the RSSI should be determined and hence different vendors use different algorithms and scales to calculate RSSI. Also, reporting of RSSI is not mandatory; hence some card manufacturers do not support such measurements at all. In our initial experimental study, we conducted a simple test in order to measure the reported RSSI reading for five different cards that support RSSI measurement and are supplied by the same vendor.

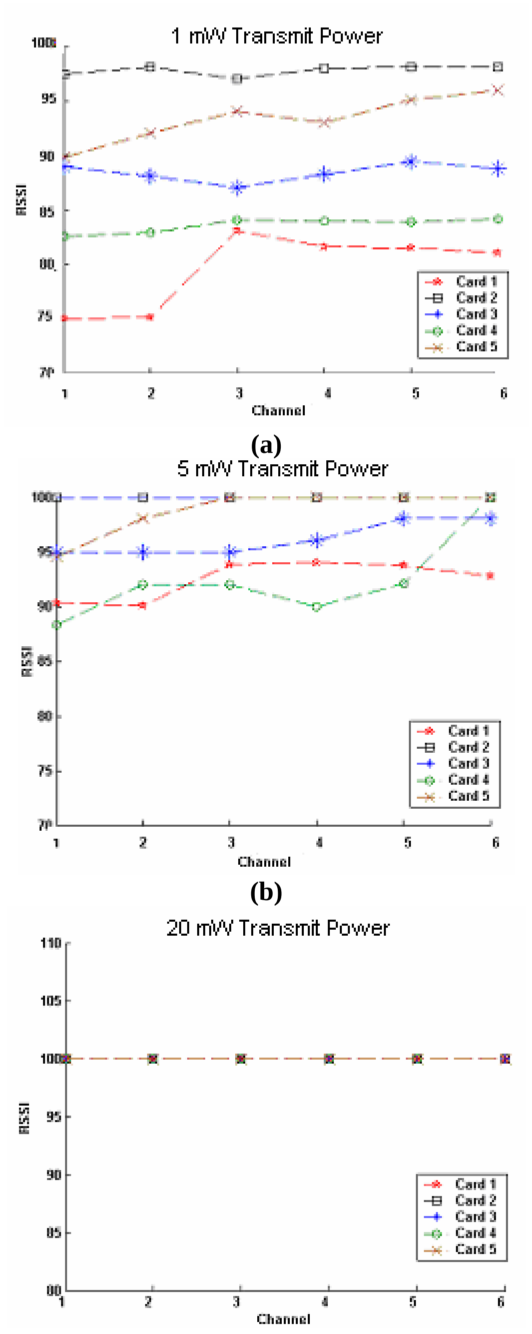
Our experimental setup consisted of specialized 2.4 GHz IEEE 802.11b nodes in a rectangular grid with a spacing of about a meter between nearest neighbors. The nodes were monitored and controlled through a wired backbone. In these initial experiments, we had a dedicated sender node with the same wireless card held constant throughout the course of the experiment. At the receiving side, separated by a distance of about 3 meters, we had a dedicated receiving node using one of the five different cards under test. For each card at the receiver, the sender (at 1 mW and a constant offered load) and receiver were set on channels 1 through 6. This was repeated for a different sender power levels (5 mW and 20 mW).

Throughout the course of the experiment that lasted for about 30 minutes, the sender and receiver nodes were separated by the same distance and for each iteration, only the wireless cards at the receiver were changed.

Fig. 1 shows the reported RSSI measurements for five cards with the sender set to transmit at 1 mW, 5 mW and 20 mW[3].

---

[3] Note that the transmit power settings for the cards used have to be verified. We have observed that some cards and drivers do not return an error message when set to a power level not supported by the hardware



(a)

(b)

(c)

Figure 1 (a,b,c) RSSI variation across different cards at 1 mW, 5 mW and 20 mW

As it can be seen from the figures, the reported RSSI value is significantly different across cards at a lower transmit power and the differences start diminishing as the transmit power increases. This is because of the saturation at the receiving end at higher transmitter power levels.

Many proposed cross-layer adaptive algorithms such as [7] and the proposed cognitive network management systems being studied under the DARPA contract mentioned in the section use the RSSI (or signal strength) reported by the card as a basis for finding stable routes or other adaptive approaches to improve wireless system performance. One important assumption in the above study is the availability of a reliable reading reported by the card. As seen from our simple experiment, even with a small sample set of five cards over a relatively short interval of time, there is an inherent discrepancy of nearly 20dB in the readings reported by different cards. In addition, this is not a simple scaling factor, but varies widely between channels. Hence, the experimental results obtained would be highly dependent on the choice of cards, thereby seriously hampering repeatability.

In order to address this issue, we propose calibration of the cards to be used, in terms of transmit power settings and RSSI values reported. Since we have little information on drift in these values, initially, frequent calibration should be used until there is statistical confidence in the probable rate of drift. For our initial experiments, we chose the readily available Cisco Aironet 350 series 802.11b wireless adapters with the following specifications [9] as shown in Table 1.

Table 1 Cisco 350 series client adapter specifications

Data Rates	1, 2, 5.5 and 11 Mbps
Receiver Sensitivity	1 Mbps: -94 dBm 2 Mbps: -91 dBm 5.5 Mbps: -89 dBm 11 Mbps: -85 dBm
Available transmit power settings	100 mW (20 dBm); 50 mW (17 dBm); 30 mW (15 dBm); 20 mW (13 dBm); 5 mW (7 dBm); 1 mW (0 dBm)
Frequency bands	2.4 to 2.4897 GHz

The Cisco 350 series card reports the RSSI as a number between 0 and 100. To the best of our knowledge, no prior work apart from [10] has been done to map the reported RSSI from the card to appropriate dBm values, for meaningful interpretation by adaptive algorithms such as in [7]. In the next section, we explain the card calibration procedure and results obtained for a sample set of four cards chosen from the above group.

2.2. Card calibration procedure

In this section, we explain the details of the calibration procedure applied to each wireless card in order to test the operating range for each card and to record any discrepancies. We form a database of the corrections to be applied for each card (if applicable) during the analysis of experimental results. The card calibration is carried out for both the transmitter as well as the receiver using the setup shown in Fig. 2.

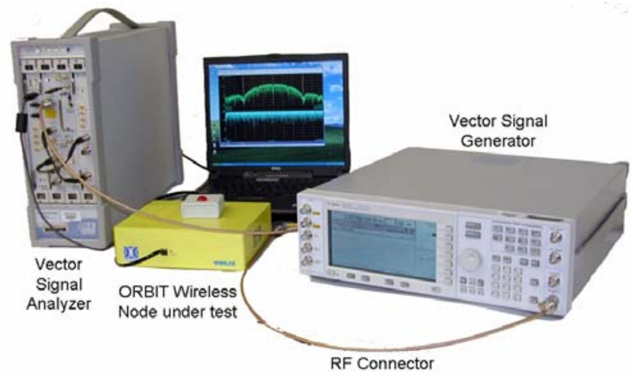


Figure 2 ORBIT Calibration Setup

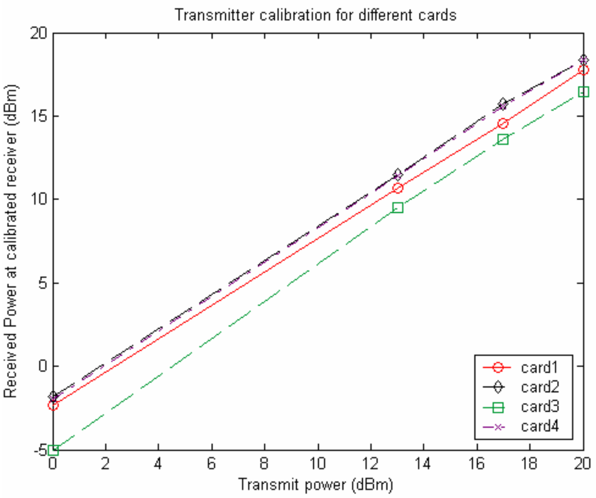
2.2.1. Transmitter calibration

In order to calibrate the transmitting side of each card, we use Agilent 89600S Vector Signal Analyzer (VSA) as the calibrated receiver with the following specifications [12] as shown in Table 2.

Table 2 Vector Signal Analyzer Specifications

Frequency Range	DC to 2.7 GHz
Amplitude Accuracy	±2 dB
Spurious response	<-65dBm
Sensitivity	-158 dBm/Hz
Frequency Accuracy	Drift: 100 ppb/year Temperature: 50 ppb

The output of the cards was connected through an RF-cable and a pair of connectors (with 2 dB attenuation loss) into the front end of the VSA. The transmitting card was fixed on channel one at four different power levels and was configured to send a continuous stream of packets through the wireless interface. The VSA measured the corresponding received band energy for each of the transmitter power settings. This was repeated for four different cards under test.



Note that received power is not corrected for the 2dB cable loss

Figure 3 Transmitter calibrations for different cards (without 2dB cable loss correction)

As seen in Fig. 3, the received power from cards 1, 2 and 4 matches their corresponding transmit power settings (after taking into account the 2 dB RF-cable attenuation loss). However, there is a slight deviation from this trend for card 3. The received power for this card is about 3 dB lower than the other three cards. It is precisely this information that we intend to capture for each card and store in the form of a correction factor to be applied during the experiments. The reason for this deviation could be attributed to the ageing of the components as well as differences in their tolerance levels. However, it needs to be accounted for in order to support repeatability in experimental results.

2.2.2. Receiver calibration

The receiver side is calibrated by using Agilent E4438C Vector Signal Generator (VSG) as the calibrated transmitter with the following specifications [11] as shown in Table 3.

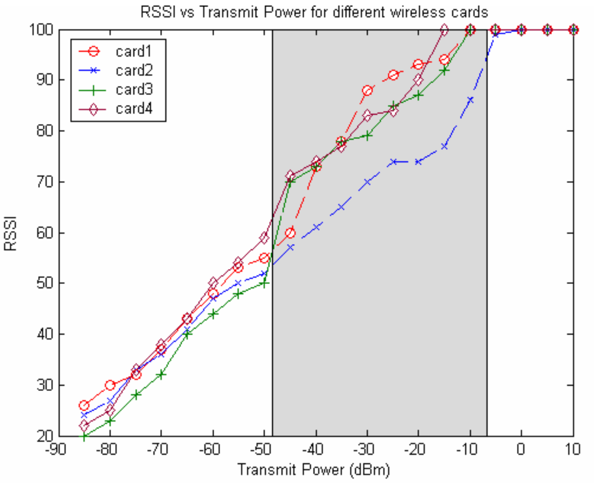
Table 3 Level Accuracy for Vector Signal Generator (dB)

	7 to -50dBm	-50 to -110dBm	-110 to -27dBm	<-127 dBm
250 kHz-2 Ghz	±0.6	±0.8	±0.8	(±1.5)
2-3 GHz	±0.6	±0.8	±1.0	(±2.5)
3-4 GHz	±0.8	±0.9	±1.5	(±2.5)
4-6 GHz	±0.8	±0.9	(±1.5)	

The internal reference oscillator for this product has an ageing of < ±1 ppm/year and a temperature variation of ±1 ppm over the range of frequencies being measured [11]. The VSG supports the capability to injected modulated 802.11b packets (with custom payload) at a desired frequency and power level. We used this feature to generate and transmit test beacons at precise frequencies and powers to exercise the entire range of RSSI measurements at the receiver. We chose a basic data rate of 1Mbps and beacon size of 59 bytes. As before, the card under test was connected to the VSG using an RF-cable with a 2 dB attenuation loss. The same procedure was repeated for the same four cards used in our earlier transmitter calibration.

Fig. 4 shows the reported RSSI values by each card for each of the transmit power settings. All the cards are unable to receive packets below a VSG transmit power of -88 dBm (which is equal to -90 dBm at the front end of the card taking into account the 2 dB RF cable and connectors' loss). This roughly corresponds to the receiver sensitivity of each card which is slightly worse than the specification value of -94 dBm as reported in Table 1.

Note that while cards 1, 3 and 4 report similar RSSI values for different transmit powers at the VSG, the RSSI readings reported by card 2 are as much as 10dB lower than the rest for some ranges of the transmit power. It is also interesting to note that card 3 had the largest deviations for the transmit calibration, but card 2 was the outlier for the receive calibration.



Note that the RSSI readings includes the 2 dB attenuation loss in the RF connector

Figure 4 Receiver calibrations for different cards

As shown in Fig. 4, all the cards exhibit similar behavior at very low powers levels, but reach saturation at power levels ranging from -10 dBm to 10dBm. Thus, for all the power levels above -10 dBm, all the cards report the same RSSI value. The deviation from the mean RSSI as shown in Fig. 5 is the greatest in the shaded portion corresponding to Fig. 4.

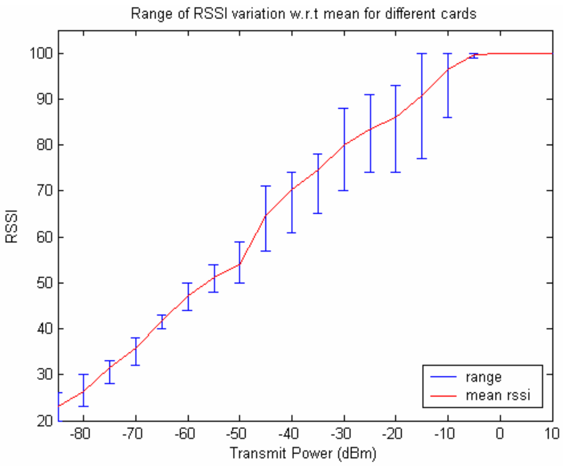


Figure 5 Dynamic range of reported RSSI across different cards

These power levels (-50 dBm to -10 dBm) are typically to be expected when using the cards in an indoor wireless testbed and it is in this range that the behavior of different cards differs significantly.

Our goal, as explained before, is to document these patterns during card calibration, in order to account for them later during actual experimentation to ensure repeatable results.

3. Tests to Characterize Repeatability in Experiments Results

In this section, we discuss the experiments conducted to measure repeatability of results in our initial testbed setup in an environment that is not optimized for RF stability. This includes identical experiments conducted over the span of a month (in order to capture time variations) and also on different sets of nodes, while maintaining the same topology (in order to capture the spatial effects and other hardware issues).

3.1. Temporal Repeatability

To investigate repeatability of results, we conducted the same experiment at random times over an extended period of about a month. In this section, we report the results for five sample runs chosen out of this duration ensuring that they span across a time period of a month. To reduce the scope of experimental error, we used the same set of nodes, same wireless cards and the same settings for each of these experiments for the entire duration. Over that period, there were some changes in the physical environment and positioning of the nodes that contribute to any changes noted. When the testbed is fully operational in its final location, these variables will be eliminated.

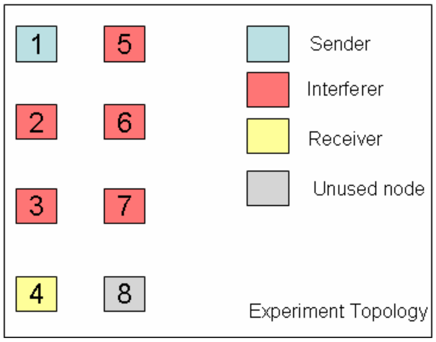


Figure 6 Experiment to study temporal repeatability

The experimental setup, as shown in Fig.6, consisted of 7 nodes, with a sender sending UDP packets of 1024 bytes to a receiver that formed the Link Under Test (LUT). Five other interfering nodes broadcasted UDP packets (1024 bytes) on the same channel as the sender-receiver pair. Both the sender and all interferers transmit at 1 mW. All the nodes are configured to be on Channel 1 initially.

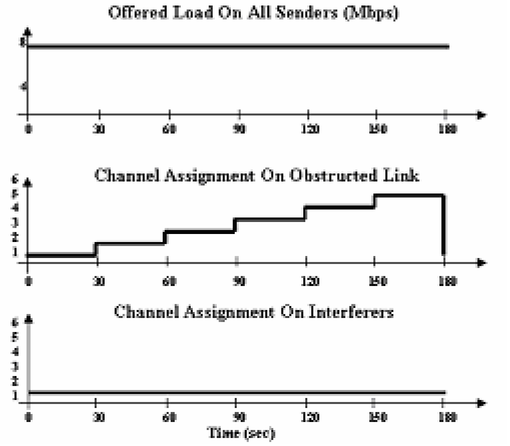


Figure 7 Experiment dynamics (to study temporal repeatability)

In order to combat interference, the channel used by the LUT is incremented one channel at a time until it operates on a completely orthogonal channel (Channel 6) as shown in Fig. 7. We observe the effect on the throughput of the LUT as it is moved to an orthogonal channel away from the interferers. The LUT dwells on each channel for 30 seconds. Hence, the entire experiment duration is 180 seconds.

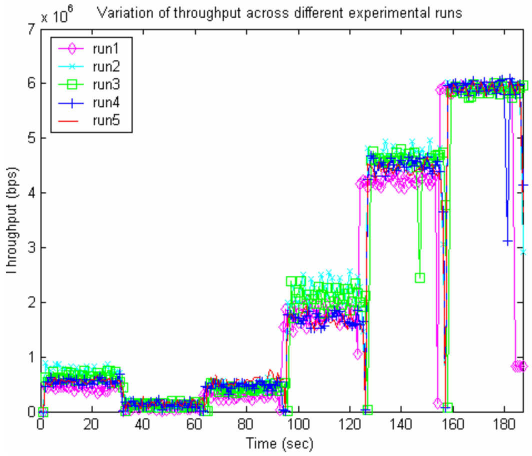


Figure 8 Throughput variations across different experimental runs in time

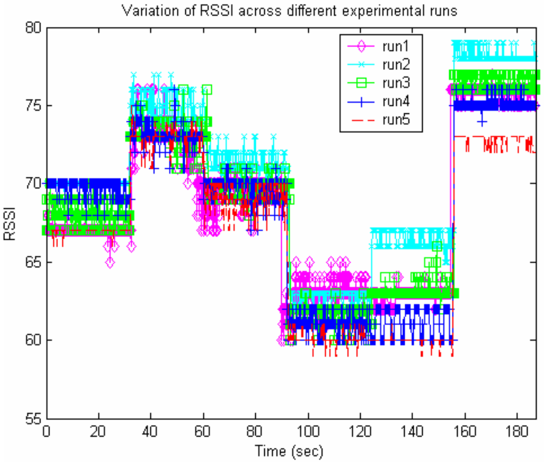


Figure 9 RSSI variations across different experimental runs

Figures 8 and 9 shows the throughput and measured RSSI of the LUT for each repeated experiment. Figure 10 shows the maximum deviation of throughput amongst different experimental runs with respect to the mean throughput. It is seen that the differences are slightly greater when channel separation is 3 (partial channel overlap) corresponding to the time interval of 90-120 seconds. It is much smaller for the cases when channel separation is 0 (0 to 30 seconds) or greater than 4 (150-180). These cases correspond to LUT being on the same channel as the interferers or on an orthogonal channel respectively.

Note that the concept of orthogonality is only valid for perfectly linear transmitters and receivers. Given the variability observed in these cards, it is unlikely that strict linearity will be achieved in these low-cost devices and thus a power dependence of these results is expected and needs to be included in any calibration strategy.

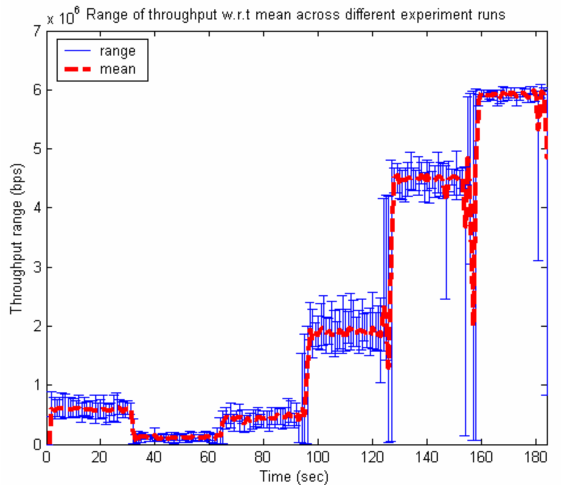


Figure 10 Throughput variation w.r.t mean across different experimental runs in time

3.2. Spatial Repeatability

Another concern regarding testbed operation is whether different (symmetric) assignment of nodes for different experiment runs produces similar results. In order to study the effects of node positions on the outcome of the experiment, we performed a simple test across twelve different node topologies in the three basic arrangements as shown in Fig 11.

In each run, we used four nodes: two senders and two receivers. The senders operated at 1 mW transmit power, on channel 1, using 1280 bytes UDP packets (40 packets/sec) for an offered load of 409.6 kbps per flow. Each experiment was conducted for 60 seconds.

For each of the basic arrangements, we rotated the topology four times giving us a total of twelve experimental runs. Since, the two flows were also symmetric in terms of offered load, the total number of sample runs for the experiment was 12 topologies × 2 flows = 24 samples runs.

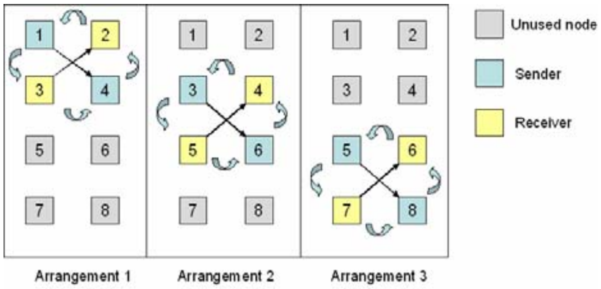


Figure 11 Experiment to test spatial repeatability

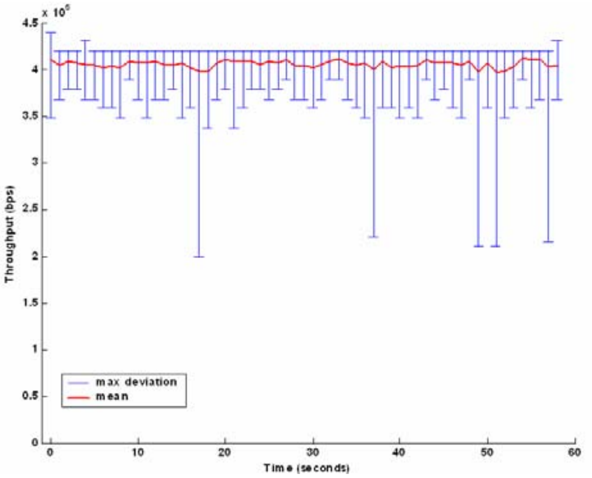


Figure 12 Spatial throughput variations w.r.t mean for experiment duration averaged over different experimental runs

In Figure 12, we show the variations of throughput with respect to the mean taken over the entire duration of the experiment. For each second on the X-axis, we found the average throughput and the maximum deviation from the average throughput using the 24 sample runs. Figure 13 shows the results from a per experiment perspective. Here, we show the throughput averaged over a single experiments' duration for each of the 24 different sample runs, and the maximum deviation from this mean.

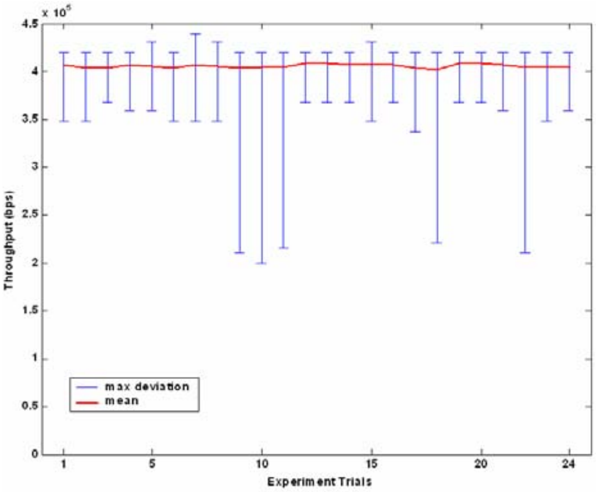


Figure 13 Spatial throughput variation w.r.t mean for different experimental runs averaged over experiment duration

Table 4 summarizes the mean and the standard deviation over all the samples (24 sample runs × 60 seconds = 1440 samples).

Table 4 Mean and standard deviation across all samples

Offered Load	Mean Throughput	Standard deviation
409.6 Kbps	406 Kbps	18.44 Kbps (~4%)

The primary observation we can make from these experiments is that over time periods of weeks, measurement variations associated with environmental changes or drift in the equipments is that they are demonstrably non-Gaussian. We have included the standard deviation for simplicity and further work may be needed to understand the distribution better. However, these observed variations are still less than the initial differences between the cards, even operating in a ramp-up mode in a temporary laboratory that is not optimized for RF environmental stability. This gives us confidence that a calibration procedure of the type we describe can be used to improve substantially the repeatability of the measurements, and thus their utility in wireless networking research. In addition, even this initial configuration is stable enough to allow definitive measurements on many experimental configurations and, thus, begin to increase our understanding of the complex behavior of wireless networks.

4. Conclusions

In this paper, we have addressed the important issue of repeatability in wireless experiments using the ORBIT testbed. A careful calibration procedure to resolve the dependency of experimental results on the hardware is also proposed. In order to make use of the card corrections obtained during the calibration process, it is important to identify the relationships between channel settings, observed RSSI values and the corresponding measured throughputs (or packet losses). We also plan to calibrate the wireless antennas that will be used in the actual experiments on the testbed using the above procedure. These tests were conducted on the preliminary ORBIT testbed setup with a 4-by-4 grid in a partially controlled environment. As future work, we intend to extend these tests to a larger and final version of the testbed consisting of 400 nodes in a 20-by-20 grid in a completely automated manner.

5. References

[1] NSF Workshop on Network Research Testbeds, Chicago, Il, Oct 2002. http://www net.cs.umass.edu/testbed_workshop/

[2] D. Raychaudhuri, I. Seskar, M. Ott, S. Ganu, K. Ramachandran, H. Kremo, R. Siracusa, H. Liu, and M. Singh, "Overview of the ORBIT Radio Grid Testbed for Evaluation of Next-Generation Wireless Network Protocols," *submission under review at IEEE WCNC 2005,* New Orleans, USA.

[3] K. Pawlikowski, H.-D.J Jeong, and J.-S.R. Lee., "On credibility of simulation studies of telecommunication networks", *IEEE Communications Magazine*, 40(1):132–139, January 2002

[4] David Kotz, Calvin Newport, Robert S. Gray, Jason Liu, Yongu Yuan and Chip Elliott, "Experimental Evaluation of Wireless Simulation Assumptions, *Proceedings of the 7th ACM/IEEE International Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM'04),* October 4-6, 2004. Venice, Italy.

[5] Judd, Glenn and Steenkiste, Peter, "Repeatable and Realistic Wireless Experimentation through Physical Emulation", *2nd Workshop on Hot Topics in Networks (HotNets-II),* November 2003, Cambridge, MA, USA.

[6] J. T. Kaba and D. R. Raichle, "Testbed on a desktop: strategies to support multi-hop MANET routing protocol development," *ACM MobiHoc,* 2001.

[7] Rohit Dube, Cynthia D. Rais, Kuang-Yeh Wang, Satish K. Tripathi, "Signal Stability Based Adaptive Routing (SSA) for Mobile Ad-hoc Networks", *IEEE Personal Communications,* February 1997.

[8] IEEE 802 LAN/MAN Standards Committee, "Wireless LAN medium access control (MAC) and physical layer (PHY) specifications", IEEE Standard 802.11, 1999.

[9] Cisco Aironet 350 Series Client Adapter Specifications, http://www.cisco.com/univercd/cc/td/doc/pcat/ao350ca.htm.

[10] "Converting Signal Strengths to dBm values", White paper, http://www.wildpackets.com/elements/whitepapers/Converting_Signal_Strength.pdf

[11] Agilent E4438C Vector Signal Generator Data Sheet,http://cp.literature.agilent.com/litweb/pdf/5988-4039EN.pdf

[12] Agilent 89600 Vector Signal Analyzer Data Sheet, http://cp.literature.agilent.com/litweb/pdf/5988-7811EN.pdf