

Evaluating Enterprise-Class Wireless LANs: Real and Virtual Benchmarking

A Farpoint Group Technical Note

Document FPG 2006-351.1
October 2006



Farpoint Group has spent a significant amount of time developing benchmarks for testing and comparing the throughput performance of wireless LANs (WLANs). As we discussed in our Technical Notes 2006-109.1, *Benchmarking MIMO – Examining the Performance of “Draft N” Products*, and 2006-314.1, *Benchmarking Wireless LANs – Recommended Practice*, there are a number of techniques which can be used to temper the highly-variable and highly-statistical nature of radio propagation in freespace. These have allowed us to perform meaningful evaluations and comparisons of a wide variety of products and architectures – but with one key drawback: we are only able to test small configurations of APs (one or perhaps two) with small numbers of clients. This is fine for comparing residential-class products, but it became very clear some time ago that we would need an entirely different set of techniques and procedures to test and compare enterprise-class systems. It would be great to be able to test a large number of clients operating against a large number of APs, but the sheer enormity of attempting this task in the real world serves as a major disincentive to even trying. One would require not just hardware, but also a suitable facility, spectrum assurance monitoring tools (to check for interference; see our White Paper 2006-321.1, *The Invisible Threat: Interference and Wireless LANs* for more information), significant effort to perform an installation, a reasonable staff, a suitable benchmarking tool, and lots of time. Moreover, results could vary based on a wide variety of externalities – the geometric relationship of all radio-equipped devices involved, the prevailing radio conditions (which indeed can also vary from moment to moment), and many more. Reproducibility and control are key to creating the level playing field from which meaningful comparisons result.

This problem is more than academic, however. As more enterprises begin to roll out production-scale WLANs, there’s a clear need to be able to specify, configure, and test large-scale WLAN installations – but the challenge we outlined above is a serious roadblock to these ends. It being completely impractical to test even a single large-scale configuration in freespace, there is no hope whatsoever for being able to conduct a meaningful real-world comparison of two different systems. This state of affairs provided the incentive for the exercise discussed in this Tech Note. Specifically, we set out to define a methodology for conducting large-scale WLAN benchmarking exercises, but without having to deal with the artifacts of real-world testing as noted above. We call this approach *virtual benchmarking*, as it is conducted in an environment defined by test equipment instead of real-world settings.

For this project, we decided to use a test-equipment system most often employed by engineers developing wireless LANs, but which also has (as we learned) significant applicability to solving the problem outlined above. We also decided to use the techniques we developed in this work to explore an architectural debate with significant real-world value. Specifically, we set out to compare two different enterprise-class Wi-Fi systems, produced by Aruba Networks and Meru Networks, respectively, to find out which performed best under virtual, simulated real-world conditions. And while it was not (as we discussed above) possible to test exactly the simulated configuration in the real world, we did perform a set of exercises that provided us with a high degree of confidence that the results of our virtual benchmark would in fact be seen in real-world installations.

Test Scenarios

Our core emphasis in these tests was to examine a number of performance-related elements. We primarily wanted to compare the performance of the Aruba and Meru systems based on both raw throughput and user density, but also regarding parameters relating to voice over Wi-Fi (VoFi) connections. The latter includes latency, jitter, call capacity, and roaming speed. We felt such an exercise would be interesting because, while Aruba and Meru produce systems with similar architectures and market focus, they use very different implementations. Most notably, Aruba uses a traditional multi-radio-channel access-point allocation strategy, while Meru puts all APs on the same radio channel. We really wanted to discover if Meru's approach offered any advantages.

We decided to use for our virtual benchmarking environment the WaveTest traffic generator/performance analyzer from VeriWave, Inc. [<http://www.VeriWave.com/products/wavetest.asp>]. VeriWave is a leading manufacturer of test equipment aimed primarily at product engineers building wireless LAN systems. It's important to be able to precisely test new and finished designs for performance, compliance to the IEEE 802.11 standard and Wi-Fi specifications, as well as other compatibility parameters. VeriWave has mentioned to us in the past that their products have found significant, well, bugs in a number of production wireless-LAN products that likely would have been missed in smaller installations but became critical in larger-scale deployments – very much in line with the intent of our testing here. And they have also developed significant performance-evaluation capabilities in their products.

The WaveTest system consists of a backplane that can hold a variety of feature cards, called *WaveBlades*, which do the actual testing and performance evaluation. Most of these are used to connect to enclosed (i.e., isolated from outside radio interference) chambers used hold an individual AP under test. APs are placed in these chambers and their antennas jacks are connected directly to cables leading to a WaveBlade. The WaveBlades also generate synthetic client traffic, so no real clients are required.

But, being fundamentally electronic test equipment, the WaveTest does not come immediately to mind as a tool for the comparative benchmarking of enterprise-class WLAN systems by end-user organizations. While clearly a highly-technical product, though, we were impressed with the WaveTest's ease-of-use and believe it could be valuable to technically-savvy staff within IT operations groups. VeriWave provides a large number of testing scripts (written in Python) that make most testing exercises no more complex than running any other benchmarks. The key differences that this approach brings, however, are high performance, and, more importantly, complete control over the RF environment.

We used a number of scripts from VeriWave for our tests, and left all unmodified. We used an Aruba 6000 Mobility Controller with AP-70 APs with Version 2.5.4 software, and a Meru MC 3025 Mobility Controller with AP-200 APs and Version 3.1.3.7 software. We were limited to three APs from each vendor for our testing; while we would have liked to have used more, it was only important for our purposes here to be able to compare the synthetic (VeriWave) and equivalent real-world configurations to see if they would indeed correlate. Our real-world tests

were run in a “Faraday cage”, which is a room built from copper mesh fabric. Radio waves in the 2.4 GHz. band cannot pass through this mesh, so we have in the cage a very high degree of isolation from potentially-interfering signals and a well-controlled environment amenable to repeatability – not as good as the VeriWave environment, to be sure, but quite good regardless. See Figures 1 and 2 for diagrams of the test configurations used.

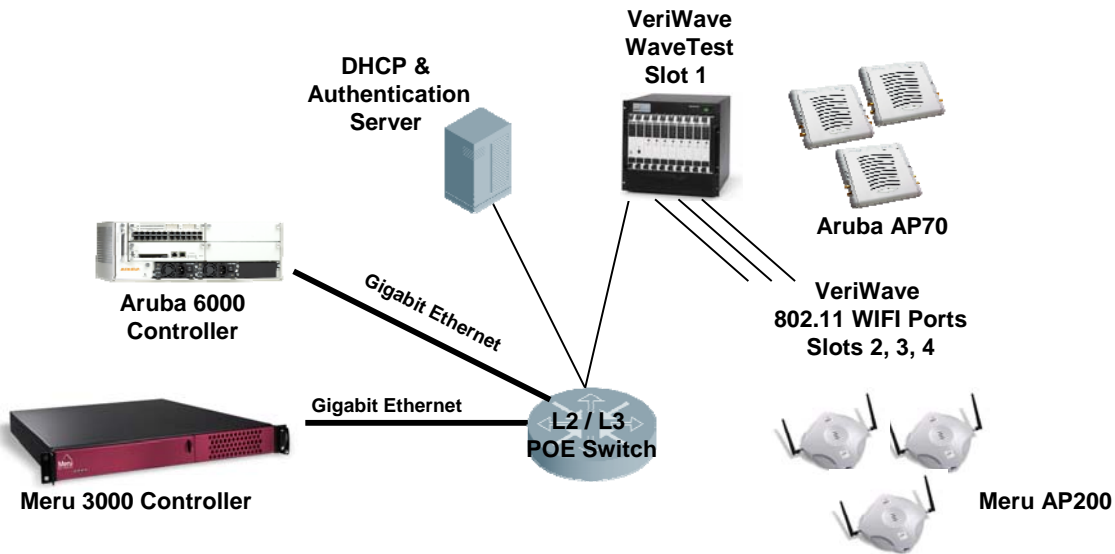


Figure 1 - VeriWave test configuration. Source: Farpoint Group

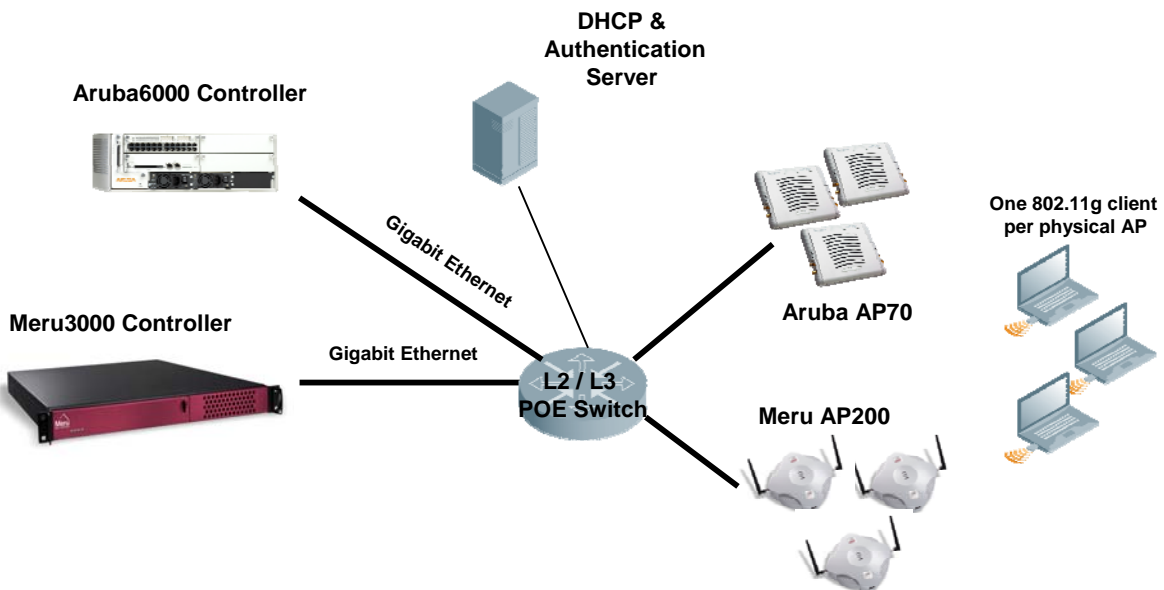


Figure 2 - Faraday cage test configuration. Source: Farpoint Group

Test Results

VeriWave test results are summarized in Table 1. By way of explanation, each of the throughput tests represent the average of seven test runs, each with a different packet size (88, 128, 256, 512, 1024, 1280, and 1518 bytes). The “Maximum Client Throughput” test is designed to find the absolute maximum number of clients that a given AP will support at a given throughput rate. Note that we could not get the Meru AP to run with more than 50 (simulated) clients. Latency and jitter tests were run with near-minimal and near-maximal loads, and (telephone) call capacity tests were run for both 802.11b (because there are so many .11b-only telephony devices on the market, although we no longer recommend the use of 802.11b at all) and 802.11g. These tests assumed as a level of quality a packet loss rate of no more than 5%. Finally, we measured the Roaming Duration of the Aruba system, which is the amount of time required to hand off a connection from one AP to another. This number is assumed to be zero for the Meru system, but both were well below the 50 ms. usually assumed to be an acceptable upper bound for this activity. The Aruba system easily outperformed the Meru system in all cases, with the exception of Roaming Duration.

Test	Aruba Result	Meru Result
One AP, One Client Throughput (Mbps)	22.36	9.7
Three APs, Three Clients Throughput (Mbps)	66.96	29.52
One AP, 40 Clients Throughput (Mbps)	22.79	10.33
One AP Maximum Client Throughput (Mbps)	22.66 Mbps (500 Clients)	10.41 (50 Clients)
One AP, 40 Clients Latency (ms., 90% throughput load)	2.62	15.96
One AP, 40 Clients Jitter (ms., 90% throughput load)	2.65	11.45
One AP, 40 Clients Latency (ms., 10% throughput load)	.44	1
One AP, 40 clients Jitter (ms., 10% throughput load)	.16	.45
One AP Call Capacity with 802.11b VoFi handsets	22	16
One AP Call Capacity with 802.11g VoFi handsets	74	14
Roaming Duration (ms.)	6.88	(assumed 0)

Table 1 - VeriWave Results. *Source:* Farpoint Group

We then used the Iperf benchmark [<http://dast.nlanr.net/Projects/lperf/>] on similar configurations to the above inside the Faraday cage. The command lines we used were as follows (using 1472-byte packets in each case):

- *TCP* – This script was used to generate stream suitable for testing throughput:

Clients: iperf -c <ip> -p 77 -i 1 -t 300 -w 128K
 Server: iperf -s -p 77 -i 1 -w 128K

- *UDP* – This test was used to simulate real-time voice traffic:

Clients: iperf -c <ip> -p 77 -u -b 40M
 Server: iperf -s -p 77 -u

Test results for these runs are summarized in Table 2. All results are in Mbps.

Traffic Type	Meru 3AP	Meru 1AP
TCP Upload	17.9	10.3
TCP Download	16.3	14.2
UDP Download	16.9	10.6
TCP/UDP Average	17.0	11.7
	Aruba 3AP	Aruba 1AP
TCP Upload	68	22
TCP Download	69	23
UDP Download	105	36
TCP/UDP Average	80.6	26.9

Table 2 - Faraday cage results. *Source:* Farpoint Group

Analysis and Conclusions

First of all, we did see a very good correlation between the synthetic and real-world results with both systems, and we can conclude at this point that synthetic benchmarking will likely be quite valuable as a predictive tool going forward - and we certainly intend to do more of it. While more can certainly be done to increase the scale of projects like this in terms of number of access points and clients tested (limited here, as we noted earlier, by logistics and cost), we are looking forward to further exploring the virtual benchmarking technique regardless.

We also found that Meru’s single-channel allocation technique does not appear to be effective in boosting either performance or capacity or provide any additional benefits in supporting voice over WLAN traffic. We have always found the single-channel concept a bit counterintuitive, as the use of more radio channels clearly provides greater capacity, and we cannot imagine that any installation would strategically restrict the number of channels to which it has access. Of course, one could deploy multiple single-channel “clouds” in a given location, but this is really not that different from deploying multiple multi-radio APs and then interleaving channels as is traditionally done (via RF Spectrum Management functionality provided by WLAN system vendors, of course). While it can be argued that using multiple APs as synchronized receivers on a single channel at any given moment in time should boost

reliability (acting, in effect, as a single large distributed antenna), such must be balanced against the overall loss in capacity that necessarily results. We are also concerned about the effect of interference on a single-channel allocation – while we did not empirically evaluate this impact during our testing, interference will, as we have noted in other reports, likely play an increasing role in the configuration and management of future wireless LAN installations.

One potential advantage that Meru's approach brings is essentially zero delay in roaming handoffs between APs, as such handoffs occur only in the controller. We found, however, that the roaming delays introduced by the Aruba system were sufficiently low that no artifacts would be noticeable in practice, including in latency-sensitive applications like voice.

In conclusion, we found that a three-channel approach to channel allocation, as tested using the Aruba 6000, provides much higher performance than is available using the single-channel technique as embodied in the Meru system. But perhaps even more importantly, we are encouraged that virtual benchmarking using test equipment will become a common technique in the comparative evaluation of enterprise-class WLAN systems.



Ashland MA 01721
508-881-6467
www.farpointgroup.com
info@farpointgroup.com

The information and analysis contained in this document are based upon publicly-available information sources and are believed to be correct as of the date of publication. Farpoint Group assumes no liability for any inaccuracies which may be present herein. Revisions to this document may be issued, without notice, from time to time.

Copyright 2006 — All rights reserved

Permission to reproduce and distribute this document is granted provided this copyright notice is included and no modifications are made to the original.