Enterprise Wireless LAN Scale Testing

November 1, 2007

1.0 Overview

We had a unique opportunity to test wireless LANs in a vacant, but fully built out office building. Most comparisons of wireless LAN systems have been done with one access point and 10 to 20 clients in an RF chamber. While that is an excellent controlled environment for repeatable tests with wireless clients, it doesn't reveal the subtlety of the complete systems or demonstrate their ability to handle large scale deployments. We wanted to examine the behavior of wireless LAN systems in a more realistic environment.

We tested wireless LAN systems from the leading enterprise vendors Cisco, Aruba Networks and Meru Networks. All of them are enterprise class wireless LAN systems with integrated security and management tools that are designed to handle very large deployments. All of the systems are 802.11 a/b/g and are Wi-Fi certified. All of them employ a wireless LAN controller that addresses the complexity of managing, securing and deploying these systems.

There are two different system architectures represented. The Aruba and Cisco systems are examples of the micro-cellular architecture which has been the primary approach for deploying large scale enterprise wireless LANs. The Meru system uses a novel system architecture based on single channel spans and AP coordination.

This paper examines what happens when we push these systems to their limits. We explore how much data capacity these systems deliver, how many voice calls are possible, and how the systems react with a mix of voice and data. The results are surprising, and illustrate some challenges for the 802.11 MAC protocol and highlight the differences between these two architectural approaches to large wireless LANs.

2.0 Test Environment

2.1 The Facility

We conducted the testing at an office building in Sunnyvale, California that was being configured for a new company to move in. The space is a typical high tech office. The entire second floor, illustrated below, was dedicated to our testing. It is approximately 20,000 square feet with a mix of hard walled offices around the perimeter and cubicles in the interior.

Figure 1 - Second Floor Laptop Locations

A few employees had already moved in during our testing, but no one in the building was using the wireless LAN or generating any traffic in the 2.4 GHz band. RF scans showed that we could “hear” other access points and wireless LAN traffic from neighboring businesses, but at a very low power level. The testing was conducted between 9 AM and 9 PM on normal work days.

2.2 Infrastructure Equipment

We used Chariot Console 4.2 to generate the traffic for these tests and gather performance statistics. Wildpackets AiroPeekNX 2.0.3 was our 802.11 packet sniffer and we used Cognio Spectrum Expert 3.1.67 to scan the airwaves and ensure that we did not have any unusual interference.

A flat Gigabit Ethernet network was used to interconnect all wired devices in the test. We use an HP Procurve 48 port 10/100/1000 switch. All of the APs were connected at 100 Mbps. The Chariot console was running on a laptop PC connected to the Ethernet network at 1 gigabit. The same laptop acted as an endpoint for the Chariot tests, essentially acting as a server on the wired network. The Berkeke OnDO SIP server was used to enable access for the VoIP handsets.

2.3 Wireless LAN Systems

	Controller	Software	Access Points
Aruba	Alcatel-Lucent OmniAccess 4308, same as Aruba 800	3.1.1.4	15 Aruba AP 70
Cisco	Cisco 4402 Wireless LAN Controller	4.0.217 and 4.1.185	15 Cisco AP1242
Meru	Meru MC3000 controller	3.4	15 Meru AP208

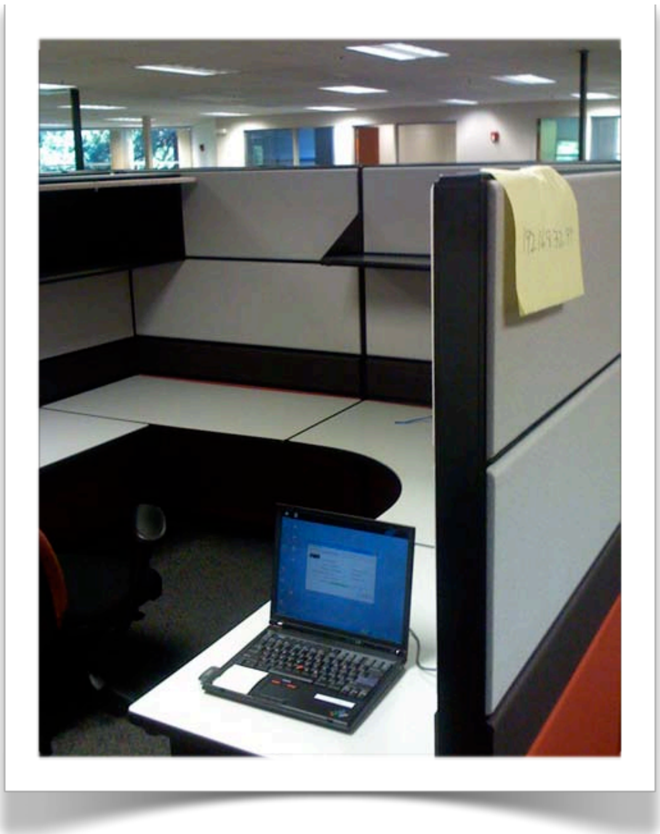
The APs from all of the vendors are dual radio. However, we only enabled the 2.4 GHz radio for our testing.

2.4 Wireless LAN Client Equipment

PC Laptops

We used 72 laptops a mix of Lenovo Thinkpad T43, HP 510 and HP Compaq nc6200 notebook PCs. All notebook PCs were running Windows XP. To keep the Wireless LAN behavior as consistent as possible, we used the same PC Card on all of the laptops - the Cisco Aironet 802.11a/b/g Cardbus Adapter[1]. Along with the Cisco AIR CB21 cards, each laptop was running Aironet Desktop Utility version 3.6.0.122 which is compatible with the latest Cisco Compatible Extensions, CCX version 5. The laptops were spread around the second floor in cubicles and conference rooms. The laptop locations are shown in the floor plan in figure 1. They remained in the same locations for all of the testing and were placed where they would likely be used - on desktops or conference tables.

1 http://www.cisco.com/en/US/products/hw/wireless/ps4555/products_data_sheet09186a00801ebc29.html



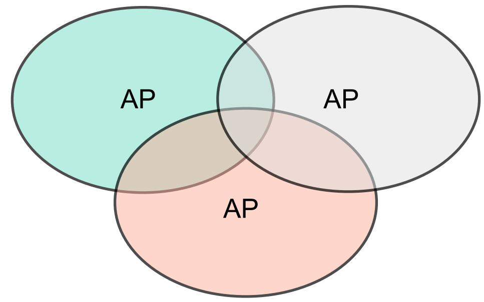
Laptop 79 shown in its testing location.

Phones

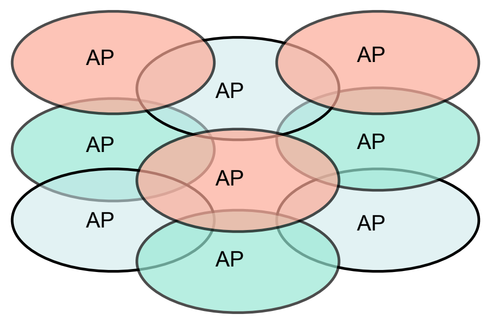
We used 50 Ascom i75 VoIP phones running 1.2.19 and 2 Cisco 7921g VoIP phones running WLAN firmware ID 4.1.0.85 w/CCX v4. The Cisco phones have a nice management tool that shows a MOS estimate.

3.0 Wireless LAN Architecture

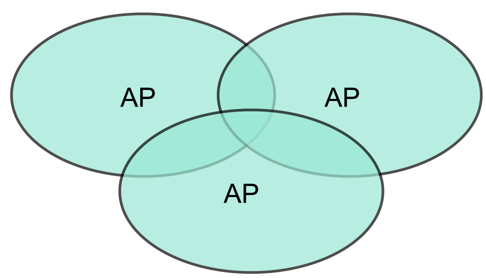
The Cisco and Aruba systems use a micro-cellular architecture for AP deployment which is the typical way of deploying enterprise wireless LANs with many access points like this. The APs act independently, with no awareness of their neighboring APs in terms of Wireless LAN access other than the low level 802.11 collision avoidance protocol. The WLAN controllers in these systems are connected to all of the APs and do have a system wide view, but they are only used to set the channel and the output power of the APs. The strategy for micro-cellular deployments is to lay out the APs such that adjacent APs are operating on different channels. In the 2.4 GHz band there are three non-overlapping channels. A simple network could be deployed as shown at left where each color represents coverage on a different channel.



To increase capacity in the same area, more APs are added and spread evenly throughout the desired coverage area. The power for each of the APs is reduced and adjacent APs are placed on alternate channels. This is an attempt to provide continuous coverage and increase capacity, while minimizing the overlap of adjacent cells to reduce the co-channel interference that the APs might cause for each other. It is difficult to create a coverage map with no overlap of adjacent cells on the same channel because there are only three non-overlapping channels in the 2.4 GHz band used by 802.11b/g networks. (Figure 3 is an example of this.) A major benefit of the controller based enterprise WLAN systems is that they can automate the process of channel selection and power level for all the APs in the system. In fact, the Aruba and Cisco systems can be configured to do this dynamically as the environment changes.

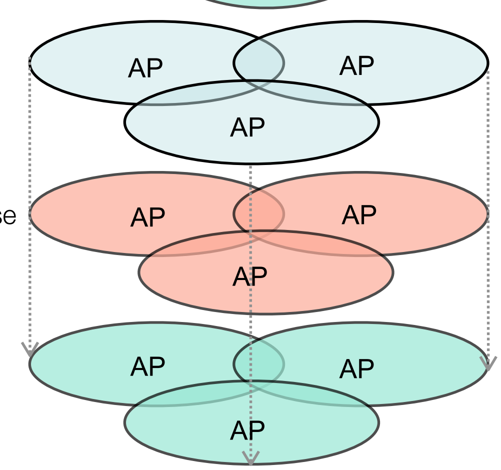


The Meru system uses a unique architecture and very different deployment strategy. For smaller systems, all APs are deployed on the same channel to form a channel span that covers the entire area. In a Meru system, the controller does much more than management and security - it also coordinates access to the airwaves by the APs and their clients. With this system-wide view and coordination of the airwaves, the Meru system can prevent collisions and reduce interference from neighboring APs on the same system.



For high density deployments, when more capacity is needed in the Meru system, additional channel spans on different channels are layered in the same area. The AP power remains at high power, since each of the APs are still covering the same size area.

The testing described in this paper illustrates the differences between these two architectures - the micro-cellular systems from Aruba and Cisco, and the coordinated AP channel spans of Meru.



3.1 Test Network Configuration

The entire office area can be covered by 5 APs. All of the systems we tested can achieve this coverage with APs mounted in the same locations as shown in figure 2 below. When configured in this fashion with the APs operating at a high power level, coverage is sufficient for 54 Mbps 802.11g operation throughout the area.

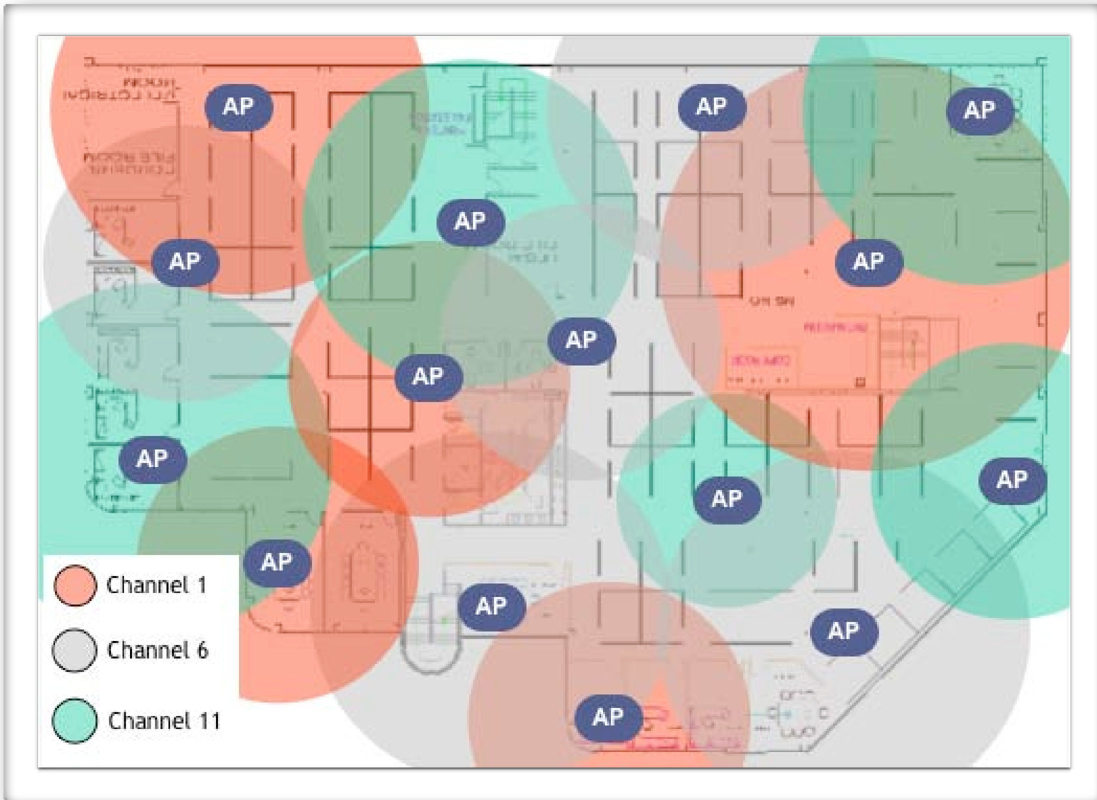
Figure 2 - 5 AP locations with Micro-cellular Coverage

All of the testing is done in the 2.4 GHz band using the non-overlapping 802.11 channels 1, 6, or 11. The micro-cellular systems are configured with adjacent access points operating on different channels as shown in the example above. The coverage shown in figure 2 is just an example. We used the Cisco and Aruba tools to automatically set channels and power levels of the APs as recommended by the manufacturers.

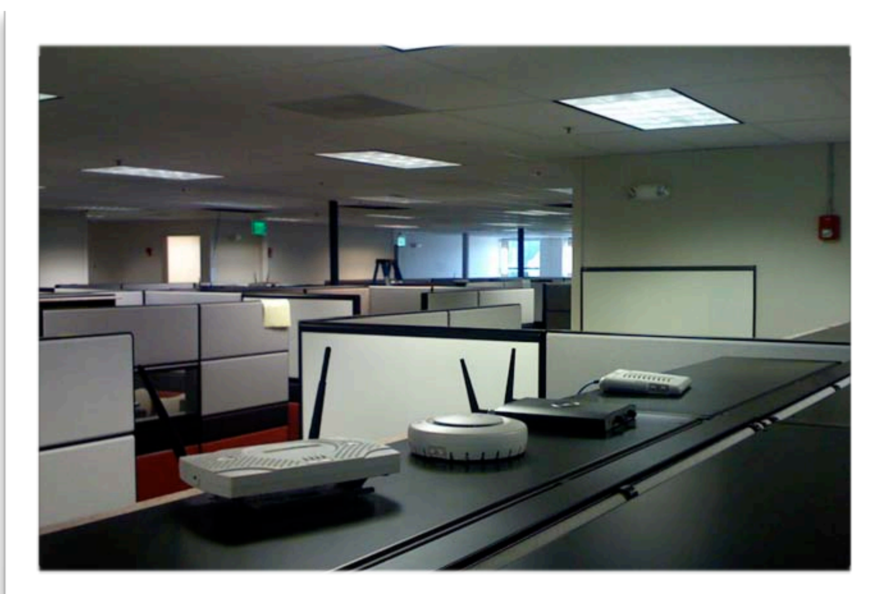
The Meru system organizes all the access points on the same channel to form a single channel span across the entire coverage area. A five AP Meru network would use the same locations as figure 2 with all the APs on the same channel. To increase capacity, the Meru system layers additional channel spans in the same area. Since there are three non overlapping channels in the 2.4 GHz band, we could layer up to 3 channel spans using 5 APs each for a total of 15 APs. We decided to use 15 APs for the high capacity testing in this office.

The 15 AP locations for Aruba and Cisco are shown in figure 3 below, along with an estimated coverage and channel plan. (This is just an example. Actual AP power levels and channel plan was done using the vendors' automated tools.) We picked the additional AP locations for Cisco and Aruba according to their recommended practice by trying to fill in the space between the original 5 AP locations and spreading the additional AP as evenly as possible. The Cisco or Aruba RF tools were used to set AP power level and channel. We made sure that the Auto RF (Cisco) or ARM (Aruba) process was given time to complete before each test. The result is that the AP power is reduced significantly and the pattern of alternating channels in the micro-cellular architecture continues as much as possible, with each AP covering a smaller area.

Figure 3 - 15 AP Micro-cellular Locations



Most of the APs are mounted on top of cubicles as shown in the picture below. Although this is not optimal and not recommended by any of the three vendors, we followed the same practice for each of the vendors as ceiling mounting was not practical for this test. We expect that the impact of AP placement on system operation was negligible for these tests.



The Meru high capacity configuration is shown in figure 4 below. Three APs are deployed at each of the original 5 locations - one on channel 1, one on channel 6, and one on channel 11. The APs at each location are separated by at least 7 feet in a "Y" configuration as Meru recommends for layered deployments.

Figure 4 - Meru 15 AP Locations

4.0 Scale Testing

The focus of this testing is enterprise wireless LAN systems operating at scale in a real environment. We concentrated on the high capacity network configurations and tested all three systems with a variety of different data and voice traffic. All of the tests were run in the 15 AP configurations described in section 3 above. However, after problems occurred with the original configurations, we repeated some of the tests with only 10 APs in the infrastructure.

We had 72 Wi-Fi equipped laptops and 52 voice over Wi-Fi handsets that we could use to generate traffic for these tests. The tests we ran are:

1. Data only with 72 clients. TCP flows from a server on the wired network to the wireless clients. 15 and 10 APs.
2. Voice only. 72 simulated VoIP conversations with up and downstream flows.
3. Voice only. 48 simulated VoIP conversations with up and downstream flows.
4. Voice only. 24 simulated VoIP conversations with up and downstream flows. 15 and 10 APs.
5. Mixed Voice and Data. 24 simulated VoIP conversations and 48 data clients. 15 and 10 APs.
6. Maximum handset tests. Using real VoiP handsets, how many calls does the system support? Up to 52.
7. Mixed Voice and Data. 12 VoIP handsets and 72 data clients.

We ran each of the tests three times for each vendor and averaged the results of the three runs. The detailed results are listed in Appendix B. The Chariot tests are designed to run for a duration of 90 seconds when they complete successfully. For some of the high density voice tests, the systems were not able to complete the tests. The results in that case are the average of the successful flows.

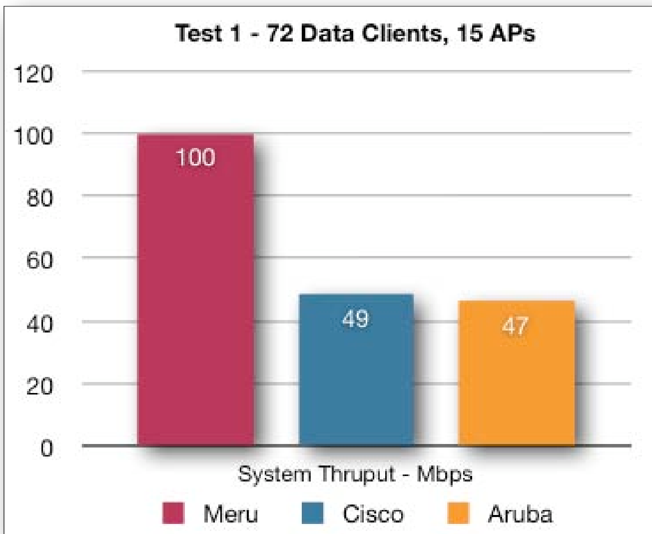
Our goal was to use current hardware and the latest shipping software deployed in the manner recommended by the manufacturer. We used the standard integrated antennas with all of the APs. We were not always able to achieve our goal of using the latest and greatest hardware and software as we tried to optimize the performance for each vendor. The configuration exceptions are described in Appendix C.

4.1 Test 1 - Data Only, 72 data clients TCP down

Setup

15 APs and 72 laptop clients connected using 802.11g. The clients and access points are configured to support WPA2 security with AES encryption. Each laptop is a Chariot endpoint. The Chariot console generates a stream of TCP traffic to each of the laptops, using a script[2] that continuously sends a 100,000 byte file as fast as TCP will allow, on each pair for 90 seconds. The only modification to the standard Chariot throughput script is setting the transmit buffer to 1400 bytes. (We did this to accommodate the reduced packet size in the tunnels between the APs and controller for each vendor, as the version of Chariot we have tends to employ TCP buffer flushing assuming 1500-byte MTUs, a behavior not seen with ordinary traffic and which can be worked around by reducing the packet size.) The same script is used for all runs of this test with all vendors.

Results



In this high density test, the wireless network is flooded with traffic from the wired network. The Meru system delivered 100 Mbps, more than twice the system throughput of Cisco or Aruba. This large difference is surprising. It appears that the co-channel interference from adjacent APs has a significant impact on the performance of the micro-cellular systems when they are loaded. The performance reduction from this is more than we expected.

Micro-Cellular Challenges

The access points and clients from adjacent cells are interfering with each other. Figure 3 illustrates the planning part of this problem - there are not enough channels in the 2.4 GHz band to truly separate APs operating on the same channel. Also, the interference range of 802.11 APs and clients is greater than their useful communication range. So even when coverage areas from APs on the same channel do not overlap, (meaning a wireless client is only able to communicate successfully with one of the adjacent APs at a reasonable data rate), there is still interference.

The micro-cellular systems from Cisco and Aruba do not deal with this co-channel interference directly. They rely on the underlying 802.11 protocol running on individual APs to resolve the issue. The 802.11 protocol is very resilient in the face of interference and poor signal quality. This is one of the great virtues of 802.11 and Wi-Fi. However, it comes with a price - protocol efficiency. The 802.11 MAC protocol uses collision avoidance, and has a built in positive acknowledgment and retry protocol. Most data messages are ACKed (acknowledged) when they are received correctly. If the sender does not receive an ACK, it will send the message again. Typically, up to 10

[2] The throughput.scr script is one of the default scrips included with the Chariot software. It is often used for network product reviews and is part of the Wi-Fi Alliance certification testing.

retransmissions are allowed per message. In these high density tests, every AP and client station has a message waiting to transmit. The probability of a collision increases due to traffic from nearby cells that is received as noise. When collisions occur that generates a retransmission, which causes more traffic, which increases the collision probability, which causes more retries, etc. The constant load can drive these systems to unstable behavior. In this test both of micro-cellular systems are filling the airwaves with packets, however, a high percentage of them are retransmissions. Therefore the throughput (the useful data delivered by the system) is lower than expected.

AP Coordination

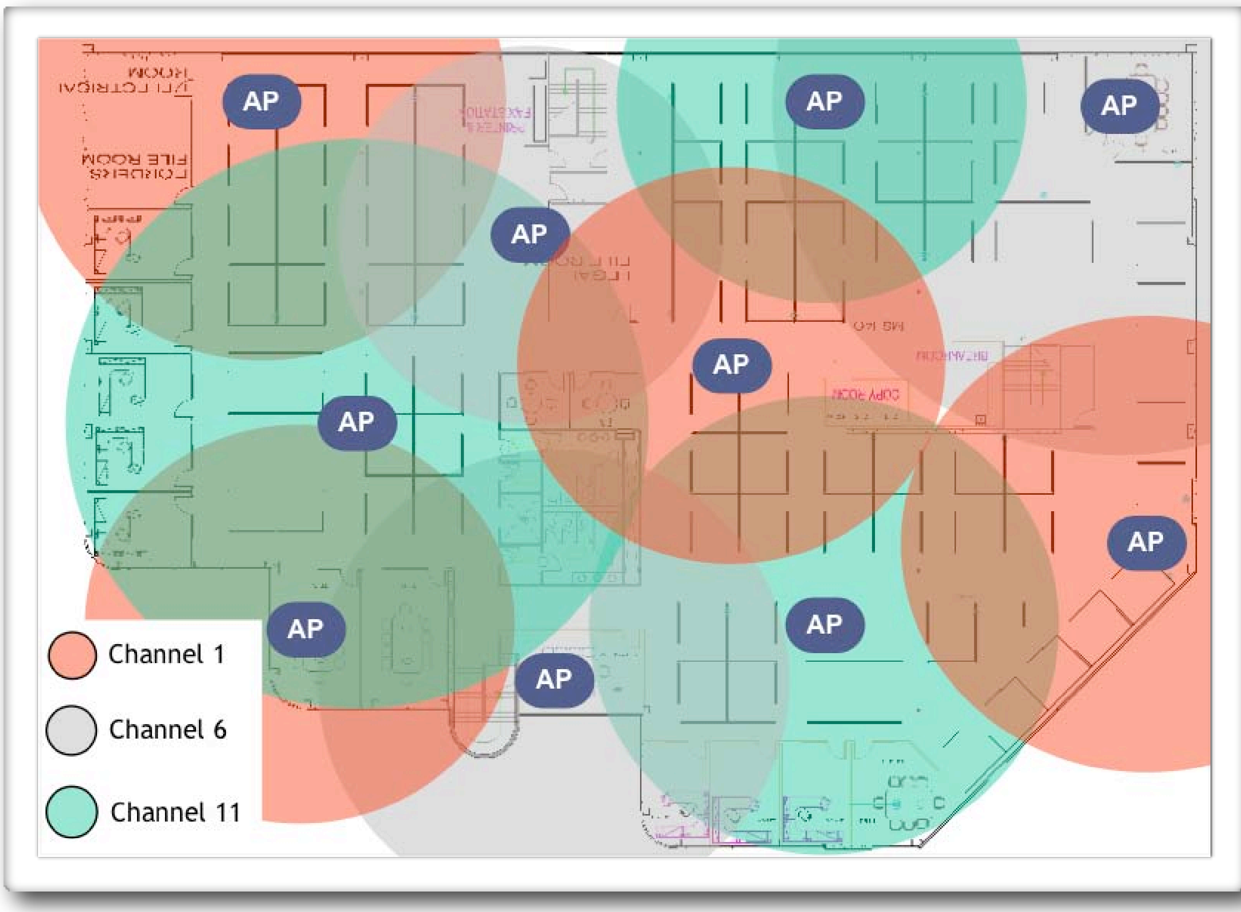
The Meru system coordinates the airtime access across all of the APs. It assumes that there will be interference from neighboring cells on the same channel and coordinates the airtime access across the collection of APs (the "channel span"). The Meru controller prevents neighboring APs from beginning transmissions that could cause collisions. The Meru controller is also aware of the higher layer protocols flowing through the controller. It uses this protocol awareness to control the behavior of the clients as well. At a very high level, the Meru system is able to reduce the collision probability through AP coordination and intelligent queueing. Essentially, they are making the network less aggressive. The Meru system prevents the 802.11 MAC protocol from getting into an overload situation. Like the micro-cellular systems, the Meru system also fills the airwaves with packets during this test. However, a higher percentage of the packets carry useful traffic rather than retransmissions, and therefore their system throughput is much higher.

Testing Fewer APs

The whole point of this exercise is to test the systems at scale - maximum APs, maximum data clients, maximum voice clients. We thought that 15 APs in this office space was reasonable in order to support 124 independent wireless clients. We followed the guidelines for configuring all of the systems in a high density deployment. However, Cisco and Aruba may not recommend using this many APs in a 20,000 square foot space. After completing the core tests with the 15 AP configuration we repeated a few of the key tests with 10 APs to see how the systems faired.

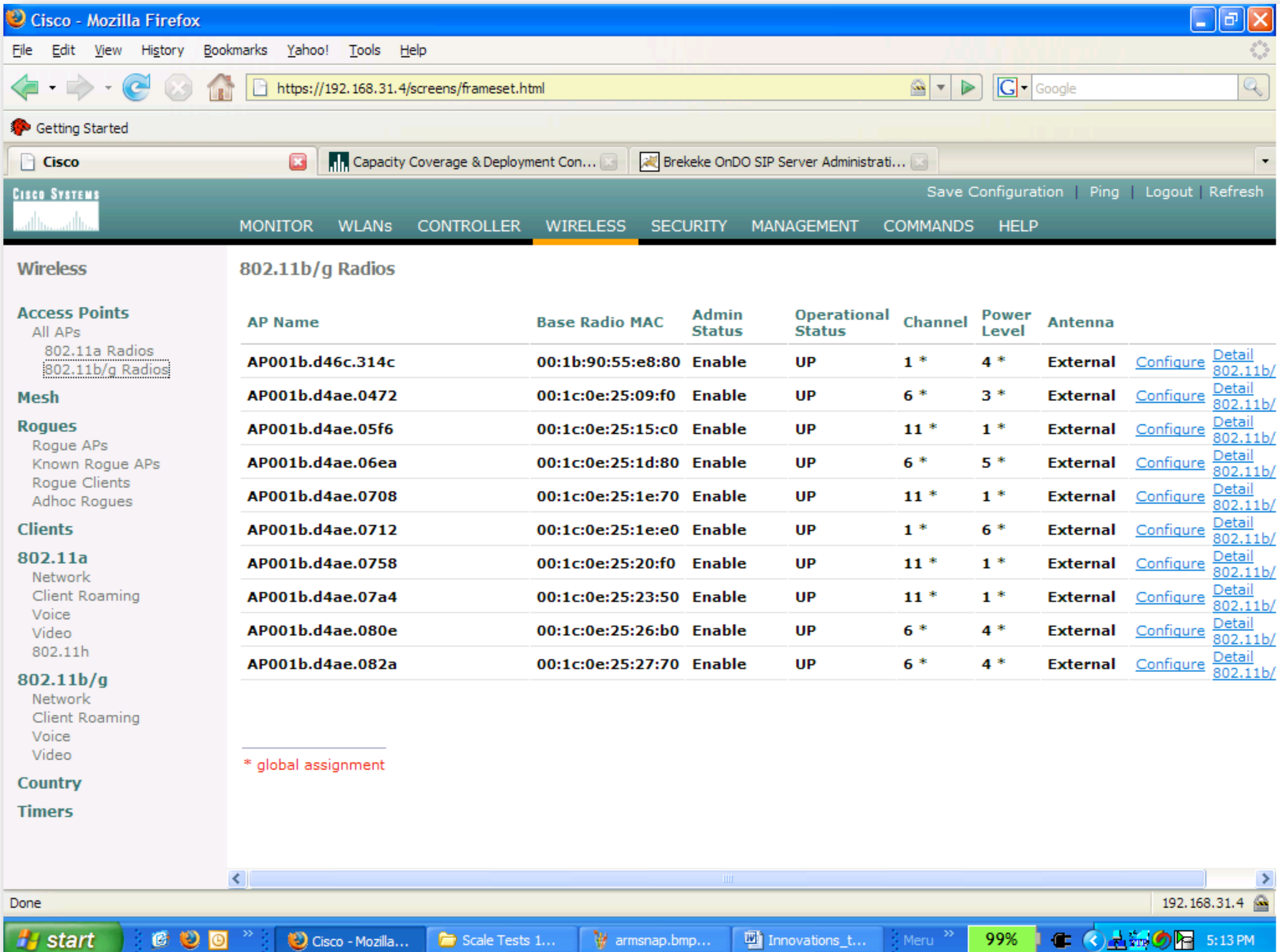
We ran the exact same Chariot script with 72 TCP clients on all three vendors systems with 10 APs to see what would happen. We configured Meru as two channel spans on channels 1 and 11 for this test. The Meru AP locations remained the same as in figure 4 with two APs in each location. For Cisco and Aruba, we placed the APs as shown in figure 5.

Figure 5 - 10 AP Micro-cellular locations



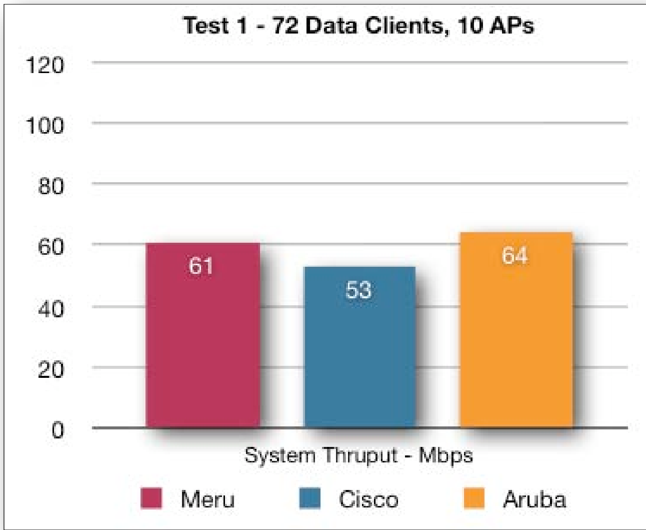
We ran the Aruba ARM and Cisco Auto RF after changing the AP configuration. Below is the Cisco screen showing the new 10 AP configuration. All three channels are used and the AP power is adjusted accordingly.

Figure 6 - Cisco AP Radio Status after Auto RF



AP Name	Base Radio MAC	Admin Status	Operational Status	Channel	Power Level	Antenna		
AP001b.d46c.314c	00:1b:90:55:e8:80	Enable	UP	1 *	4 *	External	Configure	Detail 802.11b/
AP001b.d4ae.0472	00:1c:0e:25:09:f0	Enable	UP	6 *	3 *	External	Configure	Detail 802.11b/
AP001b.d4ae.05f6	00:1c:0e:25:15:c0	Enable	UP	11 *	1 *	External	Configure	Detail 802.11b/
AP001b.d4ae.06ea	00:1c:0e:25:1d:80	Enable	UP	6 *	5 *	External	Configure	Detail 802.11b/
AP001b.d4ae.0708	00:1c:0e:25:1e:70	Enable	UP	11 *	1 *	External	Configure	Detail 802.11b/
AP001b.d4ae.0712	00:1c:0e:25:1e:e0	Enable	UP	1 *	6 *	External	Configure	Detail 802.11b/
AP001b.d4ae.0758	00:1c:0e:25:20:f0	Enable	UP	11 *	1 *	External	Configure	Detail 802.11b/
AP001b.d4ae.07a4	00:1c:0e:25:23:50	Enable	UP	11 *	1 *	External	Configure	Detail 802.11b/
AP001b.d4ae.080e	00:1c:0e:25:26:b0	Enable	UP	6 *	4 *	External	Configure	Detail 802.11b/
AP001b.d4ae.082a	00:1c:0e:25:27:70	Enable	UP	6 *	4 *	External	Configure	Detail 802.11b/

10 AP Results



With 10 APs, the spread between the highest and lowest system throughput dropped dramatically. Aruba delivered the best system throughput, 64 Mbps. Both Cisco and Aruba had higher throughput in the 10 AP test than they did in the 15 AP test. This supports the co-channel interference explanation above. Adding more APs reduced system throughput for the micro-cellular systems even though we ran their automated tools to optimize the network and turn down the AP power. It appears that increased interference is an issue for the micro-cellular systems in the 15 AP tests.

The Meru system had lower throughput with 10 APs than 15 APs. Interestingly, the Meru system achieved 61 Mbps of system throughput using just two channels. thus dropping throughput by a third when one third of the APs were removed. The performance of the channel layering approach seems to be linearly scalable with the number of channels.

4.2 Test 2 - Voice Only 72 VoIP Pairs

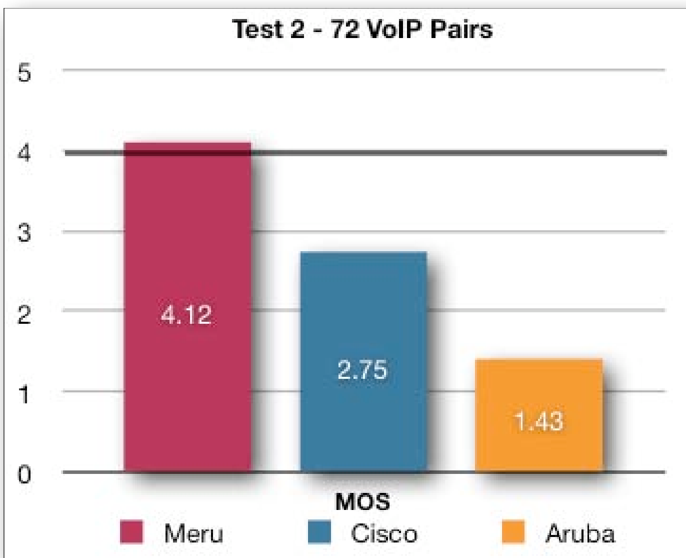
Setup

In this 15 AP test, we used Chariot to simulate 72 VoIP phone calls to the laptops running the Chariot endpoint software. We turned every laptop into a voice client. Each pair of endpoints has an upstream flow and a downstream flow to simulate a complete conversation. Chariot simulates a G.711 codec with a 30 millisecond frame size for this voice test, the same as the Ascom phones that we test later. All Chariot simulated voice traffic was marked with the WMM access class for voice in both directions using Chariot's Service Quality configuration feature.

We defined an SSID for voice traffic and activated the relevant QoS mechanisms in each of the systems within that SSID. For Cisco we used WMM allowed and Platinum. For Aruba we also used WMM for the voice SSID. For Meru we defined a new QoS rule in the controller that recognized these Chariot flows as voice traffic.

The Mean Opinion Score (MOS), R-value and other statistics that Chariot reports are averaged for up and down flows and across all of the clients. R-Value scores above 80 are considered good voice quality. The detailed results with all of these statistics are in Appendix B.

Results



A MOS of 4.0 or higher is considered toll quality voice. The MOS scale is not linear. A 2.75 MOS Score is very poor and 1.43 is completely unusable. The results show that the Cisco and Aruba systems were completely overloaded by this test. The Meru system maintained toll-quality voice averaged across all of the flows.

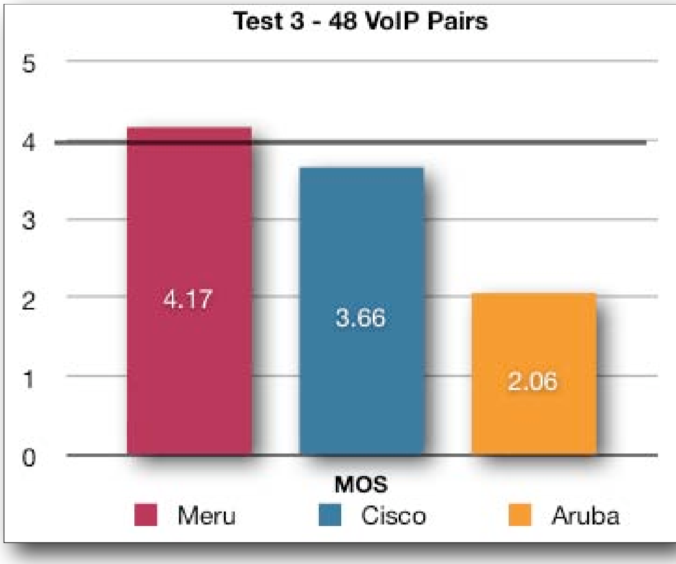
The Chariot tests run for 90 seconds and then all the results are reported to the Chariot console. "Abandoned flows" means that a test was not completed between one of the pairs. The abandoned flows have to be stopped manually after the test completes. Usually that pair "hung" in the middle of the test and stopped transmitting data. For Aruba, more than half the flows were abandoned during this test. Abandoned flows, in Chariot, do not lower the MOS score reported, but rather remove them from the average entirely.

4.3 Test 3 - Voice Only 48 VoIP Pairs

Setup

Since some of the systems were overloaded with 72 VoIP calls, we reduced the number of pairs tested. Test 3 is the same as test 2, except we tested with 48 simulated VoIP calls.

Results



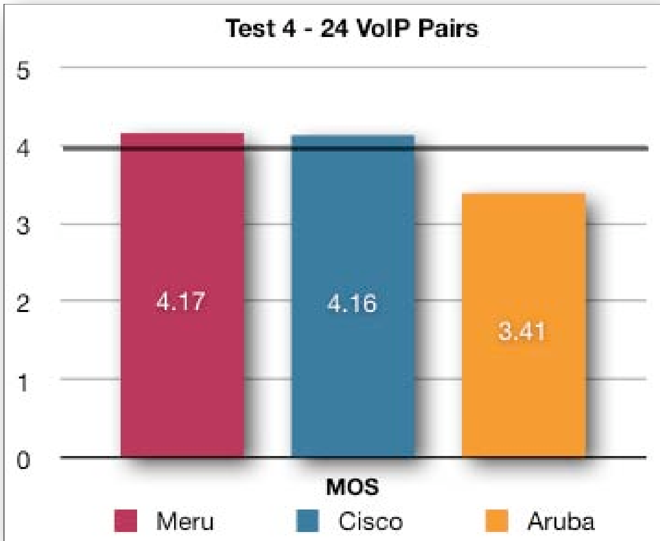
Meru delivered toll quality voice. With fewer VoIP pairs Cisco improved to the point where it was usable. Aruba, however, was still overloaded.

4.4 Test 4 - Voice Only 24 VoIP Pairs

Setup

Continuing in this reduction, trying to avoid overload, we reduced the number of pairs tested. Test 4 is the same as test 2 with 24 simulated VoIP calls.

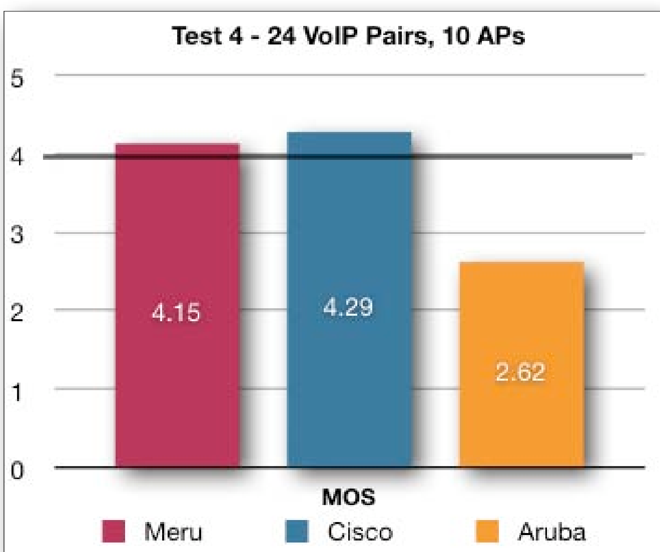
Results



Meru again delivered toll quality. Cisco was toll quality and was slightly better than Meru on some statistics, with lower loss percentage and higher throughput. Aruba was usable, but not toll quality and still had abandoned flows.

All three vendors can handle 24 VoIP calls with the 15 AP configuration. On these voice tests, Cisco and Aruba seem to be suffering from a combination of the co-channel interference described in test 1 and some sort of issues with the QoS mechanisms for voice. We tested with 10 APs to see what happens in a lower interference environment.

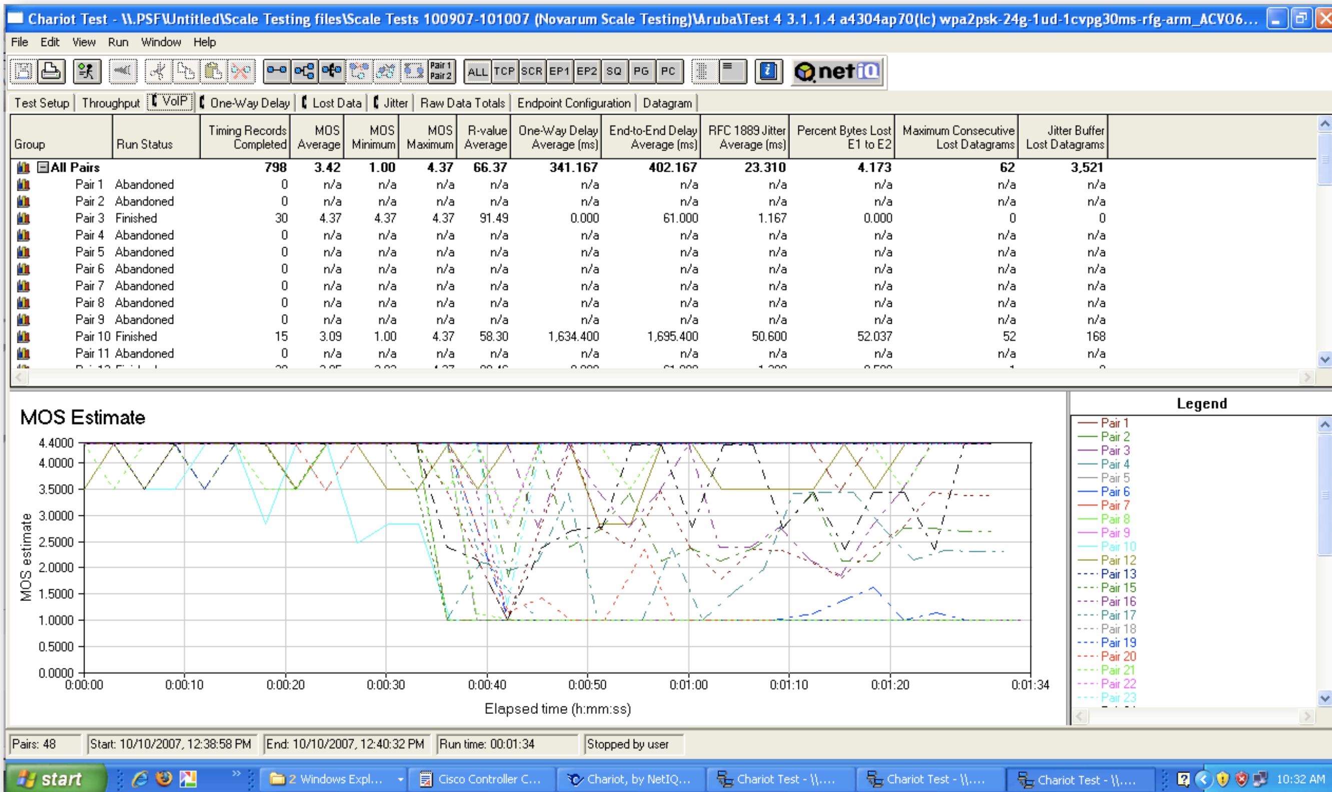
10 AP Results



Cisco worked better with 10 APs instead of 15 and achieved good toll quality with a 4.29 MOS although their delay went up compared to the 15 AP test. Aruba's MOS score was worse in the 10 AP voice test. Meru still delivered toll quality voice with results very similar to the 15 AP test.

The Aruba system had widely varying results in this test. One of the runs in the 15 AP test scored 4.0 on the MOS, but many others were much lower. The Aruba WMM firmware appears to be unstable with voice traffic. We saw some very strange behavior with the Aruba system as shown on the Chariot screen below.

Figure 7 - Chariot VoIP Screen for Aruba Test 4, 15 APs

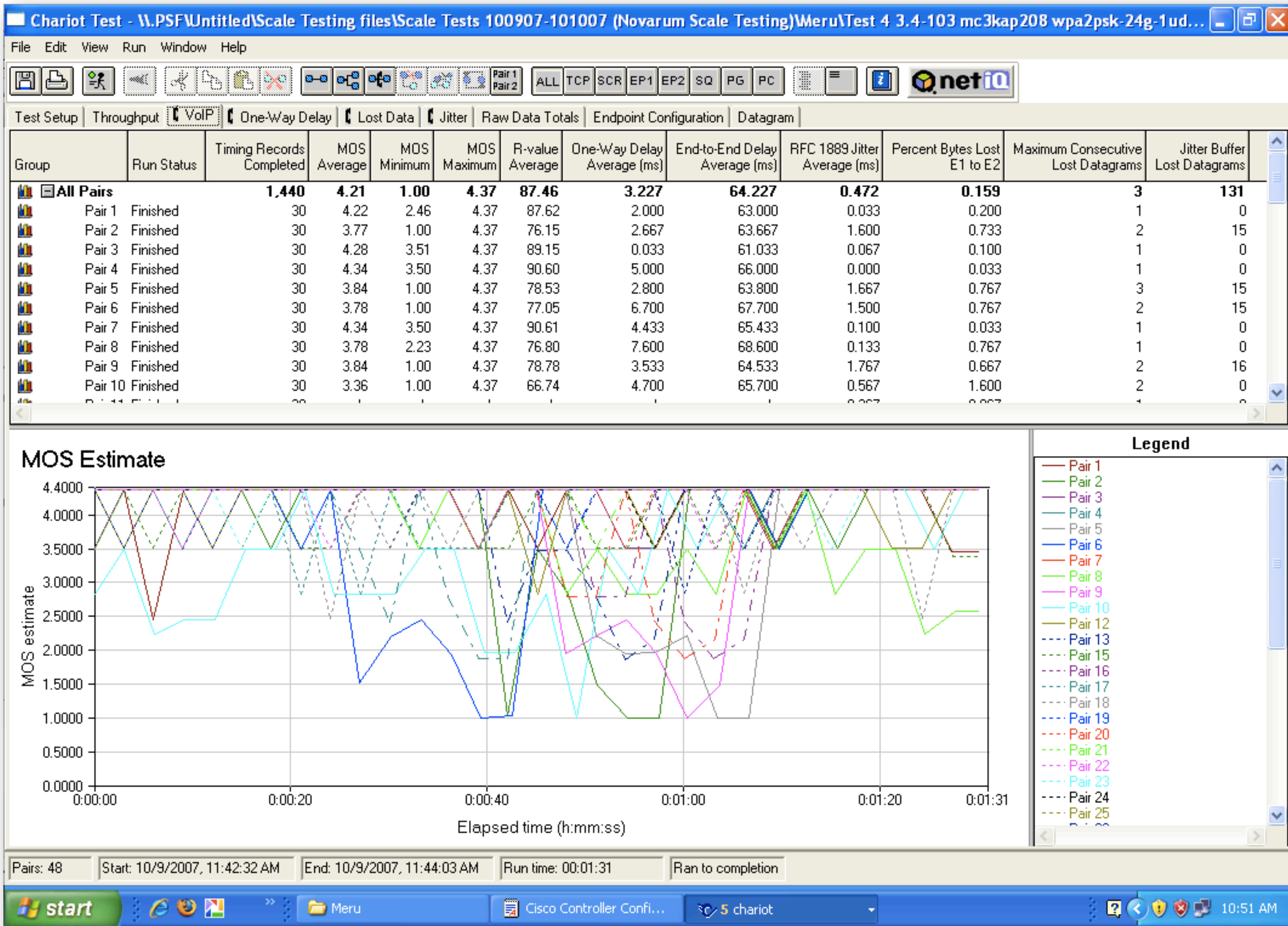


Group	Run Status	Timing Records Completed	MOS Average	MOS Minimum	MOS Maximum	R-value Average	One-Way Delay Average (ms)	End-to-End Delay Average (ms)	RFC 1889 Jitter Average (ms)	Percent Bytes Lost E1 to E2	Maximum Consecutive Lost Datagrams	Jitter Buffer Lost Datagrams
All Pairs		**798**	**3.42**	**1.00**	**4.37**	**66.37**	**341.167**	**402.167**	**23.310**	**4.173**	**62**	**3,521**
Pair 1	Abandoned	0	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Pair 2	Abandoned	0	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Pair 3	Finished	30	4.37	4.37	4.37	91.49	0.000	61.000	1.167	0.000	0	0
Pair 4	Abandoned	0	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Pair 5	Abandoned	0	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Pair 6	Abandoned	0	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Pair 7	Abandoned	0	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Pair 8	Abandoned	0	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Pair 9	Abandoned	0	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Pair 10	Finished	15	3.09	1.00	4.37	58.30	1,634.400	1,695.400	50.600	52.037	52	168
Pair 11	Abandoned	0	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a

The graph from Chariot above shows the real time MOS estimate from Chariot. For the first 30 seconds of the test, the results are good, then many of the flows degrade to very poor MOS and do not recover. Only a few of the flows finish with good results. Just above the graph is the status by pair. Only two of the first eleven flows completed the test. The rest are abandoned flows. These are all downlink flows.

This behavior is very different than the other vendors in the voice tests. The detailed Chariot results for the others are not perfect, but show much better behavior as illustrated in the Meru results from the same test in figure 8 below. In this case, there are no abandoned flows in the first 10 shown. The MOS estimate for most of the flows is very good. There are some poor estimates on a few of the flows, but they all recover by the end of the test.

Figure 8 - Chariot VoIP Screen for Meru Test 4, 15 APs

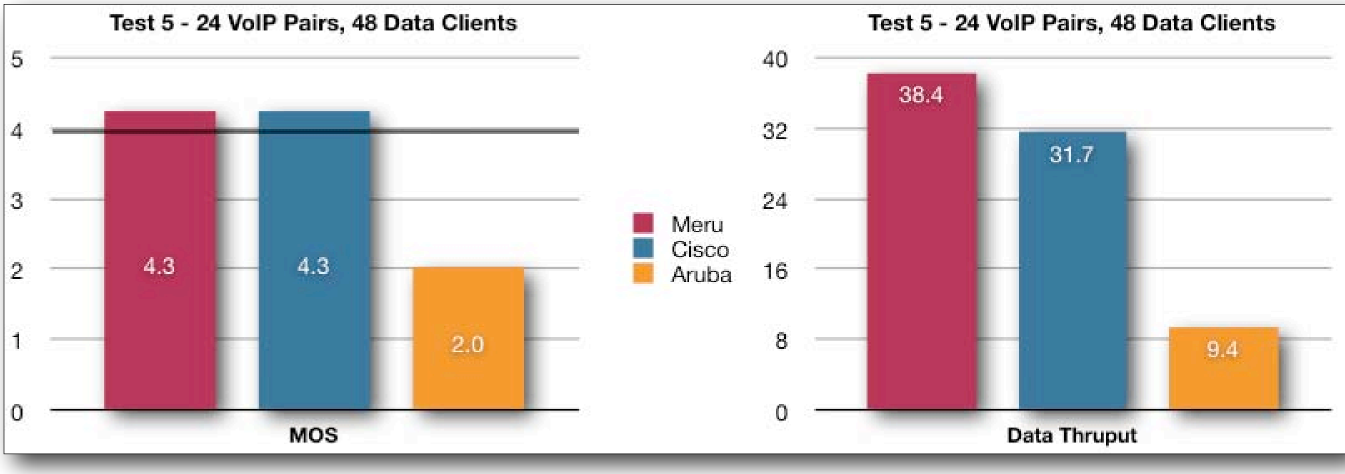


4.5 Test 5 - Mixed Voice and Data, 24 VoIP Pairs and 48 Data Clients

Setup

Having established 24 VoIP pairs as a reasonable number that all systems can handle, we use the remaining Chariot endpoints as data clients. Will the systems be able to maintain appropriate QoS for the voice traffic with a moderate load of data traffic in the background? We set up two SSIDs - one for voice and one for data and applied the appropriate QoS rules to each. The data traffic is WPA2, AES encrypted TCP flows from the Chariot console to the laptops. The voice traffic is an up flow and a down flow to each of 24 Chariot endpoints.

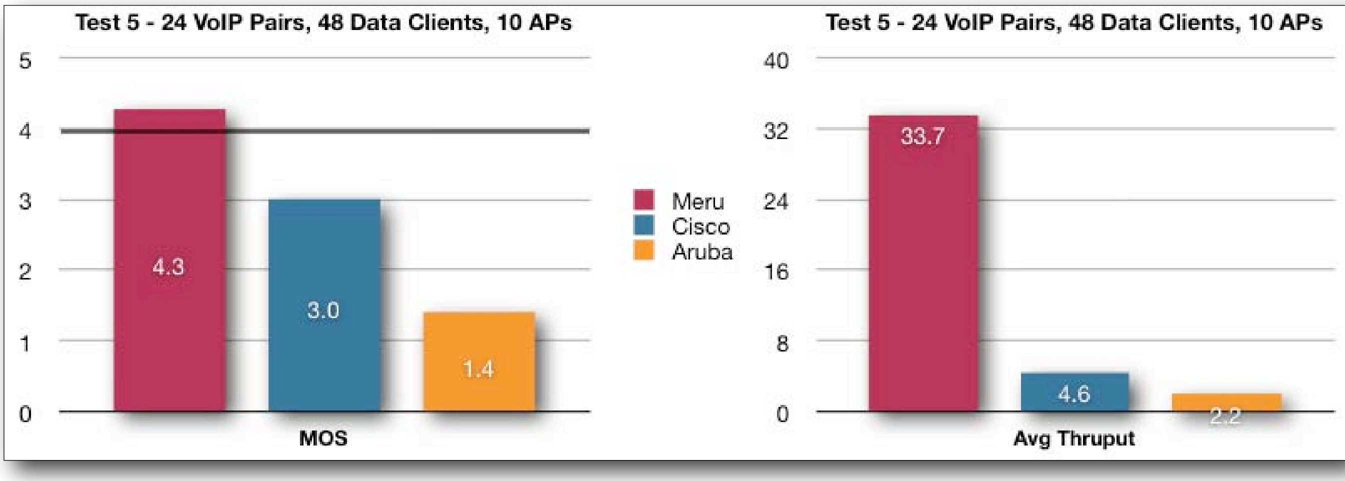
Results



Both Meru and Cisco did very well with a MOS score of 4.27 and reasonable data throughput. Meru delivered the highest total throughput - 38.4 Mbps. Once again, the Aruba results were disappointing.

We repeated test 5 with 10 APs instead of 15.

10 AP Results



Meru delivered toll quality voice and good data throughput. The voice results for Meru were the same with fewer APs, but the data throughput was slightly lower. Cisco results were worse when we reduced the number of APs in this test. This is different than test 1 and test 4 where the micro-cellular systems improved with fewer APs. With 15 APs

Cisco performed very well in this test, but they did not deliver toll quality with 10 APs. We may have uncovered a per AP limit on the number of toll quality calls mixed with data.

With the unusual Aruba behavior, we decided to turn off WMM and see what would happen. There appears to be something wrong with the new Aruba firmware and voice operation. (Or perhaps our configuration of Aruba...)

Aruba Test 3	WMM Off
MOS	4.25
rValue	88.0
1 way delay	3.2
Jitter	0.6
% loss	0.1
Avg Through	6.1
Abandoned flows	0.00

In the voice only test 3, with 48 simulated VoIP calls, Aruba performed quite well with WMM off. This is the best MOS reported for Aruba on any voice test, and there were no abandoned flows.

Aruba Test 5	WMM Off
MOS	2.49
rValue	41.2
1 way delay	691.2
Jitter	59.7
% loss	1.2
Avg Through	27.2
Abandoned flows	14.00

However, in the mixed voice and data test 5, the voice results for Aruba were poor with WMM turned off. The MOS was 2.49 compared with 2.0 for the same test with WMM on.

4.6 Test 6 - Max Voice Test with Handsets

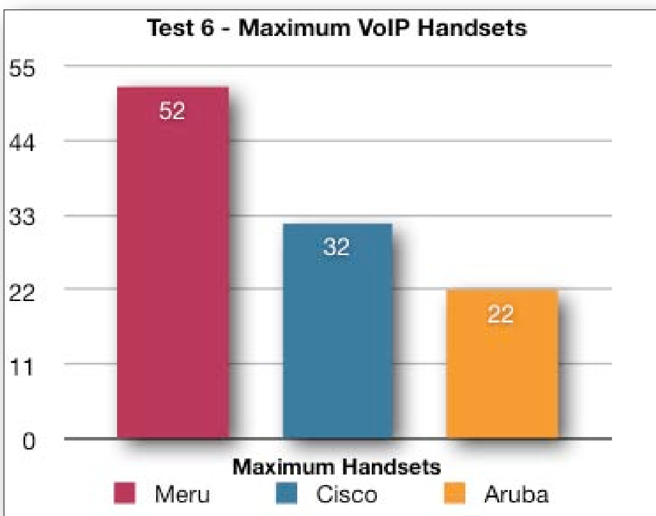
Setup

In this 15 AP test we used real VoIP handsets. The WLAN infrastructure is in the same configuration as tests 2, 3 and 4. We tested with 50 Ascom i75 phones running 1.2.19 firmware and 2 Cisco 7921g phones running firmware 4.1.0.85 and Cisco CCX v4. The newer Cisco phones have a management tool that estimates the MOS value. All phones used the standard SIP protocol for registration to the OnDo SIP server and call setup and tear down.

The previous voice tests have used Chariot to simulate calls. This is excellent because we run an automated script to set up and simulate the calls, it is easily repeatable, and we get detailed statistics on each flow. Using actual handsets makes the tests a bit more realistic and brings in the human factor. We can get an idea of how the calls sound, but it makes it more difficult to gather detailed performance statistics. We always included the two Cisco phones in the mix since they report MOS.

The phones were located in 5 different conference rooms in the four corners and the center of the floor. We established calls between the handsets and tested the voice quality subjectively. After switching from Chariot to real phones, the metric we were looking for was: the maximum number of phones that we could place into calls, by hand, and still be able to carry on a reasonable conversation across the phones. We tested each pair, using one person on each end, and recorded the number of phones that produced usable results.

Results



For Meru, we were able to initiate and conduct voice calls with all of the available handsets. Once the calls were established, we walked around the floor with the two Cisco phones during a continuous conversation. We tried to observe any clicks or gaps as we transitioned from one AP to another. We are not sure if we ever transitioned to a different AP or if the transitions were flawless. Either way, we heard no problems. The MOS reported by the Cisco handset was 4.10 before we finished establishing all of the calls. It dropped to 4.0 as we were walking around. The voice quality was good on all of the calls that we checked.

With Cisco, we could no longer initiate calls after 34 handsets were communicating. The voice quality of all of the calls suffered and was not very good at that point. We turned off two of the handsets and the voice quality improved. With 32 handsets, the voice quality was OK.

Aruba made it to 26 before we were unable to make further calls. The voice quality on some of the calls was OK, but others were not working at all. After dropping to 22 calls and waiting a bit, the call quality improved and was usable.

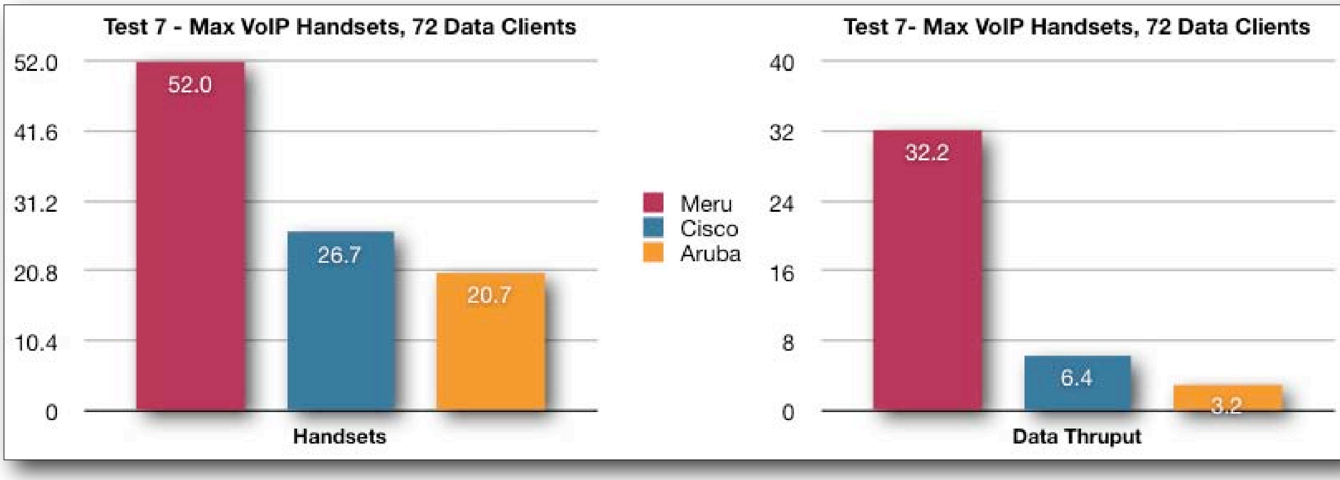
4.7 Test 7 - Max Mixed Voice and Data Test with Handsets

Having determined the maximum number of calls that the different systems could handle in our environment, we added data traffic to see how they respond.

Setup

15 APs, 2 Cisco 7921g phones and the maximum number of usable Ascom phones and 72 data clients with TCP flows from the Chariot console to the laptops.

Results



In this test, Meru handled all of the voice calls and delivered data throughput of 32.23 Mbps, about one third of the data only throughput for the same configuration in test 1. Cisco and Aruba delivered 13% and 7% of the test 1 data throughput, and it appeared these systems were struggling to support the number of voice calls.

We initiated the voice calls first and then started the data test. For Cisco and Aruba when the data started, the sound quality degraded and some phones dropped out altogether. We had to reduce the number of handsets in the test for each vendor slightly.

The MOS for Cisco started out at 3 (as reported on one of the Cisco handsets) and then dropped to 2.5 on average. Most of the calls were usable, but not great. The calls with the Cisco handsets sounded a bit better than most of the call on the Ascom phones. For Aruba, we had to reset the controller after test 7 because we were unable to start new calls for the next test.

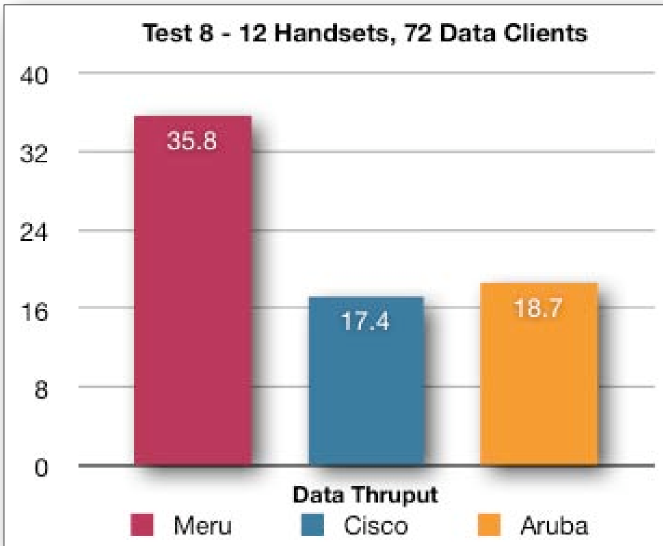
4.8 Test 8 - Mixed Voice and Data Test with 12 Handsets

We reduced the number of voice calls further to focus on data capacity the systems could deliver in a mixed voice and data test.

Setup

15 APs 2 Cisco 7921g phones, Ascom phones, and 72 data clients TCP flows from the Chariot console to the laptops.

Results



It is clear from the results that even when only 12 handsets were used, that data throughput was impacted by the voice traffic. Meru and Cisco delivered 36% of the data throughput they delivered in the data only test 1 and Aruba delivered 40% of its test 1 throughput. The cost of delivering voice calls with QoS is about the same for all the systems, but Meru starts out with a much higher system throughput, so again they deliver twice the data throughput in this test.

The subjective voice quality on the handsets was better with Meru on all of the voice tests regardless of the number of handsets.

5.0 Conclusions

The story that emerges from this enterprise wireless LAN scale testing is broader than the difference between products from 802.11 infrastructure vendors. It is really about the 802.11 protocol and how it responds when pushed to the limits in an enterprise environment.

Co-channel interference is a real factor in enterprise wireless LAN deployments whether they are hand tuned or configured with the vendors' automated RF tools. The interference range of Wi-Fi devices is greater than their useful communication range. There are not enough independent channels in the 2.4 GHz band to allow for deployments with continuous coverage that also have APs on the same channel spaced far enough apart to avoid self interference. (See figure 3 above for an example of the challenge.) This co-channel interference causes packet errors and retransmissions, limiting the overall performance of 802.11 systems under load.

Cisco and Aruba are classic micro-cellular architecture systems. In the data only test 1, with 15 APs and 72 wireless clients they delivered less than 50 Mbps of system throughput. However, in a 10 AP configuration, both systems delivered more throughput in the same test. Aruba throughput increased from 46 Mbps to 64 Mbps - almost 40%. If the APs in the system were perfectly isolated from each other, we would expect the 15 AP configuration to deliver 50% more throughput than the 10 AP configuration with the constant load in this test. But more APs allow more simultaneous transmissions which caused more interference and lowered performance for these systems.

The Meru system has a very different architecture and deployment strategy that is designed to deal with these enterprise deployment issues. The recommended Meru deployment starts out with all APs on the same channel. In our test environment, we were able to cover the entire floor with 5 Meru APs operating at high power. On the surface, the Meru approach seems like it would be low capacity since it essentially groups APs together on the same channel and the same collision domain. However this deployment approach allows the Meru WLAN controller to coordinate the airtime access of the APs and (indirectly) wireless clients in the system. To increase capacity in the Meru system, a new set of APs is added in the same area, all tuned to a different channel and still operating at high power. The 15 AP configuration we tested is three independent channel spans with 5 APs each on channels 1, 6 and 11. This approach runs contrary to the micro-cellular deployment exemplified by Cisco and Aruba, which adds more APs at lower power distributed throughout the coverage area in order to increase capacity.

The Meru system delivered twice the system throughput of the micro cellular systems in the 15 AP high density data test. This dramatic difference was surprising. Co-channel interference is worse that we expected at this scale, and **AP coordination is a significant benefit for enterprise WLAN systems**. The micro-cellular systems had a very high retry rate during the testing. (We saw retry rates greater than 40% during some of the tests.) The Meru system with AP coordination had a much lower retry rate. and the result is better system throughput.

The micro-cellular systems did not scale well. We thought 15 APs would be reasonable for high capacity testing in our environment, but Cisco and Aruba did not perform well with that many APs in this space. They both delivered higher system throughput with 10 APs. The best result for the Aruba system was on Test 1 with 10 APs where it delivered the highest system throughput of 64 Mbps. Cisco delivered the best voice performance on the test 4 with moderate load, 24 VoIP pairs, and 10 APs. Clearly, adding more APs did not increase the capacity of the micro-cellular systems and there is a limit to the number of APs that can be effectively deployed in these systems. The Meru system delivered better performance than the best micro-cellular system in all of the 15 AP tests. Coordinating access with neighboring APs is a promising area that should be pursued by 802.11 for both increased performance within the same system and better co-existence with other systems in the unlicensed bands.

All of these systems work well with just a few APs or with moderate load on larger systems. We were able to overload the micro-cellular systems with 72 data clients and 52 voice clients operating simultaneously in a 20,000 square foot office. We were surprised that we were able to cause the negative micro-cellular/802.11 behavior at this scale. The only unusual thing about our testing is the constant load during the data tests. That type of load is similar to 72 people on the same network downloading a movie from iTunes store at the same time. Though that is not typical behavior for many enterprises, we would expect any of these systems to be able handle it gracefully.

Voice is a different story. The Cisco systems did a good job with 24 simulated calls, with or without data traffic mixed in. However, Cisco was unable to deliver toll quality for 48 or higher simulated calls. With real handsets, the Cisco system seemed to max out at 26 or 28 simultaneous calls. It is very likely that many enterprises deploying voice on their wireless LAN will expect to support 30 or more simultaneous calls in a 20,000 foot office space.

The Meru system delivered better voice, and mixed voice and data performance. Meru maintained toll quality on all of the voice tests. We did not find the limit of the Meru system in terms of number of calls supported by a system of this scale. Test 7, the mixed voice and data test, showed the most dramatic difference between these systems. The Meru system delivered 52 simultaneous toll quality calls and 32 Mbps of data throughput, far exceeding the performance of the Cisco or Aruba systems. It appears that there are two factors that lead to this voice performance advantage under load. The AP coordination already discussed keeps the system stable, and the Meru system includes pre-WMM QoS mechanisms and additional enhancements for voice that go beyond the standard IEEE 802.11e.

Appendix A - Full Disclosure

These tests were conducted at a Meru facility with the help of Meru engineers.

For Novarum, we took our role as independent analysts very seriously. We adjusted the test deployment of the micro-cellular systems so that they would deliver the best possible performance. We followed the manufacturers' guidelines for deployment and configuration. We made sure that all of the tests were conducted under the same conditions. We defined new test configurations that showed the micro-cellular systems in a better light - the 10 AP tests.

We did not contact Cisco or Aruba directly and ask for their help with the testing.

For the Meru system we had the benefit of working with Meru engineers to stage and configure their system. They did so with loving care. There is no doubt that the system was configured properly for the best performance. It is possible that the Cisco and Aruba systems were not optimally configured. Maybe there is a different version of code in the lab that will perform better for these tests. Maybe a Cisco or Aruba engineer would find a way to tweak their configuration to get better performance. In our judgement, the Cisco and Aruba systems were configured and deployed as well as they would be if one of their authorized VARs did the installation.

It is possible that these systems could be optimized further, but that might be a 5 or 10% improvement. That would not change the fundamental story here. In the high density data test, Cisco and Aruba delivered half of the throughput of Meru. In the voice only and mixed voice and data tests Meru supported many more toll quality calls and much higher data throughput across the board. We had to reduce the number of handsets tested in order to prevent the Cisco and Aruba systems from going into unstable behavior.

Personal note:
If you are looking for personal bias behind this, I am in the micro-cellular, Cisco and Aruba camp. I was a major proponent of the micro-cellular approach during the development of the 802.11 standard. My early papers for 802.11 may be where the term "micro-cellular" was first applied to wireless LANs. At Aironet, I led the software and systems team that developed the first micro-cellular system software that worked for both frequency hopping and direct sequence systems. Much of that work on the client roaming, scanning, re-association and probing is still embodied in the Cisco CCX software today.

I am directly responsible for RTC/CTS being in the MAC protocol and was a major proponent of the distributed control MAC with positive ACK and retry. Over the years I have taken a lot of flak for the protocol inefficiencies of the 802.11 MAC. Looking back, I wouldn't change a thing. Those MAC protocol mechanisms - "protocol inefficiencies" - are what has enabled 802.11 to flourish in the unlicensed bands.

Conceptually, I knew that the 802.11 MAC protocol overload was possible. I also know that the interference range is greater than the communication range of 802.11 products, and co-channel interference is occurring in multiple AP networks more than people realize. Most of the time the resiliency of the MAC protocol masks these issues and they are seen as a throughput reduction. This scale testing was an eye opener for me. I was surprised to see how easily we could drive these systems to unstable behavior. There is nothing wrong with the Aruba and Cisco systems. They simply chose not to address this. Meru addresses these issues head on, and the test results speak for themselves.

Phil Belanger, November 2007

Appendix B - Detailed Test Results

Test 1	Meru	Cisco	Aruba
System Thruput - Mbps	**100.4**	48.7	46.6

10 APs Test 1	Meru	Cisco	Aruba
System Thruput - Mbps	60.9	53.2	**64.4**

Test 2	Meru	Cisco	Aruba
MOS	**4.12**	2.75	1.43
rValue	**85.3**	48.8	11.9
1 way delay	**5.2**	394.6	2262.9
Jitter	**0.5**	48.6	197.6
% loss	**0.4**	17.6	17.3
Avg Thruput	**9.1**	5.2	1.4
Abandoned flows	**0.00**	31.00	73.00

Test 3	Meru	Cisco	Aruba
MOS	**4.17**	3.66	2.06
rValue	**86.4**	72.9	29.4
1 way delay	**3.3**	107.4	1180.5
Jitter	**0.5**	13.4	99.4
% loss	**0.3**	7.0	9.9
Avg Thruput	**6.1**	5.3	2.0
Abandoned flows	**0.33**	4.00	40.33

Test 4	Meru	Cisco	Aruba
MOS	**4.17**	4.16	3.41
rValue	**86.3**	86.1	65.9
1 way delay	**2.8**	11.5	530.4
Jitter	**0.5**	3.2	28.8
% loss	0.5	**0.2**	5.3
Data Thruput	3.0	**3.1**	1.8
Abandoned flows	0.33	**0.00**	13.67

10 APs Test 4	Meru	Cisco	Aruba
MOS	4.15	**4.29**	2.62
rValue	85.9	**89.6**	45.0
1 way delay	**4.4**	59.7	937.2
Jitter	**0.4**	0.5	57.6
% loss	0.6	**0.0**	9.9
Avg Thruput	3.0	**3.1**	1.5
Abandoned flows	1.00	**0.00**	16.00

Test 5	*Meru*	*Cisco*	*Aruba*
MOS	**4.27**	**4.27**	2.05
rValue	88.6	**88.7**	30.0
1 way delay	8.9	**7.6**	1747.0
Jitter	**1.4**	4.1	69.1
% loss	**0.1**	**0.1**	9.5
Data Thruput	**38.4**	31.7	9.4
Abandoned flows	0.33	**0.00**	10.33

10 APs Test 5	*Meru*	*Cisco*	*Aruba*
MOS	**4.29**	3.03	1.41
rValue	**89.0**	56.9	12.9
1 way delay	**10.8**	448.3	1364.3
Jitter	**1.8**	29.1	71.7
% loss	**0.1**	2.3	13.0
Data Thruput	**33.7**	4.6	2.2
Abandoned flows	**0.00**	5.67	16.00

Test 6	*Meru*	*Cisco*	*Aruba*
Maximum Handsets	**52.0**	32.0	22.0

Test 7	*Meru*	*Cisco*	*Aruba*
Handsets	**52.0**	26.7	20.7
Average Throughput	**32.23**	6.43	3.18

Test 8	*Meru*	*Cisco*	*Aruba*
Handsets	**12.0**	**12.0**	12.0
Average Throughput	**35.78**	17.38	18.75

Appendix C - Configuration Exceptions

Aruba - We did have the current Aruba firmware and were able to deploy the system as recommended with one exception. The Aruba AP70 includes an integrated antenna that is attached to the AP with a hinge. The AP can operate with the antenna down (flush to one side of the case) or up (perpendicular to the top of the case). The recommended configuration for our deployment was antennas up. When we ran the tests in this fashion, the results were very disappointing. We tried putting the antennas down and the tests ran much better. This dramatic difference surprised us. We thought the automated RF tool for Aruba would adjust for the coverage difference between up and down by changing the transmit power level. The Aruba system was tested with the antennas down.

Cisco- The current controller firmware, version 4.1.185, did not work well in most of our tests. After much experimenting and re-configuring controller parameters, we switched to an earlier version of the Cisco firmware for most of the tests. Version 4.0.217 of the firmware worked much better in all of our scale tests. We refer to that version as "older firmware" here. We used the older Cisco firmware in these tests in order to optimize performance.

Meru - We tested with Meru's latest shipping firmware available for the United States. As of the time of writing, this code did not have WMM support, but included Meru's pre-WMM voice enhancements instead. Meru claims that the pre-WMM voice enhancements use the same mechanisms as WMM, but that this version of code does not have WMM certification and thus is not able to turn on WMM in the clients. Meru says that a firmware version that supports WMM and includes some additional Meru enhancements for voice will be available in the US shortly. We were unable to try that code for these tests.

