802.11n: The Next Generation of Wireless Performance

The IEEE 802.11 Working Group has been working for the last few years to standardize an upgrade to the 802.11 radio that provides a new set of capabilities dramatically improving the reliability of 802.11 communications, the predictability of 802.11 coverage, and the overall throughput of 802.11 devices. This white paper describes these new capabilities, provides insight into how the benefits provided by 802.11n are achieved, and details how this new standard is compatible with existing 802.11 deployments. This white paper also describes the issues to confront when planning migration of an existing 802.11 deployment to 802.11n and the results that can be expected from such a migration. But first, it is important to understand just what 802.11n is, and what it is not.

The IEEE Version and Ratification of 802.11n

The official version of 802.11n is the one produced by the 802.11 Working Group and ratified by the IEEE. This version does not exist yet. The 802.11 Working Group is in the process of developing a draft of 802.11n. This draft has been approved by an internal Working Group ballot, but a significant number of technical issues still need to be resolved before this internal draft is complete and ready for its next phase of approval. The Working Group anticipates that this phase of its work will be done in July 2008.

Once the 802.11 Working Group completes its work on the 802.11n draft, the draft is sent for a "sponsor ballot." This sponsor ballot is run by the IEEE and consists of a variety of experts who review the draft once again. Based on comments received from the sponsor balloters, the 802.11 Working Group continues to update the draft until the sponsor balloters reach consensus that the draft is ready for final approval. The official 802.11 timeline for concluding this process and achieving IEEE Standards Board approval of 802.11n is March 2009.

Only after 802.11n achieves final approval by the IEEE Standards Board is there an actual standard.

The Wi-Fi Alliance Version of 802.11n

The Wi-Fi Alliance, an industry organization that provides interoperability certification for 802.11 devices, began certifying the interoperability of 802.11n devices in June 2007. The certification uses an early draft of the IEEE 802.11n standard, draft 2.0. 802.11n draft 2.0 is the first draft of 802.11n to achieve sufficient consensus in the 802.11 Working Group to be considered an "approved" internal draft. This internal Working Group approval requires 75 percent of the voters to vote in favor of the draft.

802.11n draft 2.0 establishes the basic requirements for 802.11n devices. Draft 2.0:

- Includes the radio requirements for channels and data rates
- Describes the use of multiple-input multiple-output (MIMO) technology.
- Modifies the frame formats used by 802.11n devices from those of existing 802.11 devices.
- Describes some mechanisms for backward compatibility with existing 802.11a/b/g deployments.

Since draft 2.0 was approved, many areas of the draft standard have been changed. Other areas of the draft are still under intense technical discussion.

However, 802.11n draft 2.0 is an early, unratified draft of the eventual 802.11n standard. There will be changes between this draft and the eventually ratified standard. A very important question that remains to be answered is how much will the ratified standard differ from 802.11n draft 2.0.

The Wi-Fi Alliance and its members are working to ensure that the difference between the final ratified 802.11n standard an 802.11n draft 2.0 is such that any device built to the 802.11n draft 2.0 specification can be upgraded to the final ratified standard with only software changes. How successful this strategy will be remains to be seen. The stated position of the Wi-Fi Alliance is that it will update its certification testing to include the ratified 802.11n standard, once it is available.

Pre-802.11n Products

Even before there was a draft of the 802.11n standard that had reached any degree of consensus in the 802.11 Working Group, there were many products available that claimed to be “pre-n.” These products were based on one or another of the many proposals that were made to the 802.11 Working Group. Most of these products were not compatible or interoperable between different vendors.

802.11n Technology

The goal of the work on 802.11n is to dramatically increase the effective throughput of 802.11 devices, not to simply build a radio capable of higher bit rates. The difference between these goals is like the difference between the mileage you achieve with your own, personal driving habits and the EPA-rated mileage for your model of car. To increase the effective throughput of an 802.11 device requires more than providing a higher bit rate. There are aspects of the 802.11 standard that are “overhead” for the 802.11 protocol. Many of these overhead aspects can’t be reduced or eliminated. The effect is that, without using other methods, there is an absolute upper bound on the effective throughput.

802.11n is much more than just a new radio for 802.11. In addition to providing higher bit rates (as was done in 802.11a, b, and g), 802.11n makes dramatic changes to the basic frame format that is used by 802.11 devices to communicate with each other. This section will describe the changes incorporated in 802.11n, including MIMO, radio enhancements, and MAC enhancements.

MIMO

Multiple-input multiple-output (MIMO) is the heart of 802.11n. This technical discussion of MIMO provides a basis for understanding how 802.11n can reach data rates of 600 Mbps.

Radio Operation Basics

To understand the improvement brought by MIMO technology, it is important to understand some of the basics that determine how well a traditional radio operates. In a traditional, single-input single-output radio, the amount of information that can be carried by a received radio signal depends on the amount by which the received signal strength exceeds the noise at the receiver, called the signal-to-noise ratio, or SNR. SNR is typically expressed in decibels (dB). The greater the SNR, the more information that can be carried on the signal and be recovered by the receiver.

To understand this situation, imagine the analogy of your eye as the receiver. Is your eye able to detect whether a table lamp is on or off in the house next door? In this analogy, ambient light is the noise. At night, detecting that the lamp is on or off is quite easy. However, in full daylight, it is much

more difficult to make the same determination, because the ambient light is much brighter, and the tiny amount of additional light from the lamp can be undetectable.

Light, like a radio wave, disperses uniformly from its source. The farther the receiver is from the source, the less power is received from the source. In fact, the amount of power received decreases more rapidly than the square of the distance from the source. Noise, unfortunately, is often constant in the environment, due to both natural and man-made causes.

So, returning to the table lamp example, when it is too bright to determine if the lamp next door is on or off, it might be possible to make that determination from just outside the neighbor's window. Alternatively, it might be possible to make the determination if the neighbor changed the 40 watt bulb for a 150 watt bulb. In both cases, the SNR increases—in the first case, because the distance to the source is reduced, and in the second case, because the power of the transmitter is increased.

Once the minimum SNR is achieved to allow information to be exchanged at the desired rate, any additional SNR is like money in the bank. That additional SNR can be spent on increasing the information rate, increasing the distance, or a little bit of both. However, you can't spend the same dB more than once, just as you can't spend the same dollar more than once (at least not without encountering some unpleasant consequences).

All this is background to understand the improvements that MIMO technology brings to 802.11.

MIMO Technology: Beamforming

MIMO technology takes advantage of other techniques to improve the SNR at the receiver. One technique is *transmit beamforming*. When there is more than one transmit antenna, it is possible to coordinate the signal sent from each antenna so that the signal at the receiver is dramatically improved. This technique is generally used when the receiver has only a single antenna and when there are few obstructions or radio-reflective surfaces—for example, open storage yards.

To understand transmit beamforming, consider a radio signal as a wave shape, with a wave length that is specific to the frequency of the signal. When two radio signals are sent from different antennae, these signals are added together at the receiver's antenna (see Figure 1). Depending on the distance that each radio signal travels, they are very likely to arrive at the receiver out of phase with each other. This difference in phase at the receiver affects the overall signal strength of the received signal. By carefully adjusting the phase of the radio signals at the transmitter, the received signal can be maximized at the receiver, increasing SNR. This is what transmit beamforming does—it effectively focuses the transmitters on a single receiver, as shown in Figure 2.

Figure 1. Destructive Interference

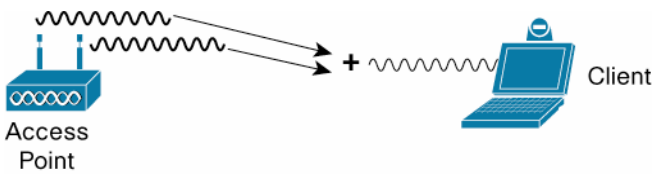
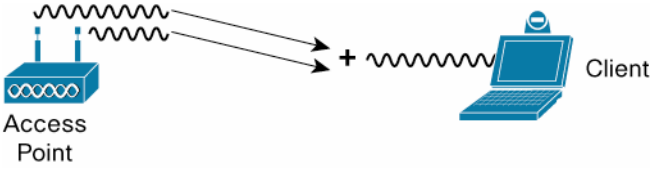


Figure 2. Transmit Beanforming (Constructive Interference)



Transmit beamforming is not something that can easily be done at the transmitter without information from the receiver about the received signal. This feedback is available only from 802.11n devices, not from 802.11a, b, or g devices. To maximize the signal at the receiver, feedback from the receiver must be sent to the transmitter so that the transmitter can tune each signal it sends. This feedback is not immediate and is only valid for a short time. Any physical movement by the transmitter, receiver, or elements in the environment will quickly invalidate the parameters used for beamforming. The wave length for a 2.4-GHz radio is only 120mm, and only 55mm for 5-GHz radio. So, a normal walking pace of 1 meter per second will rapidly move the receiver out of the spot where the transmitter's beamforming efforts are most effective.

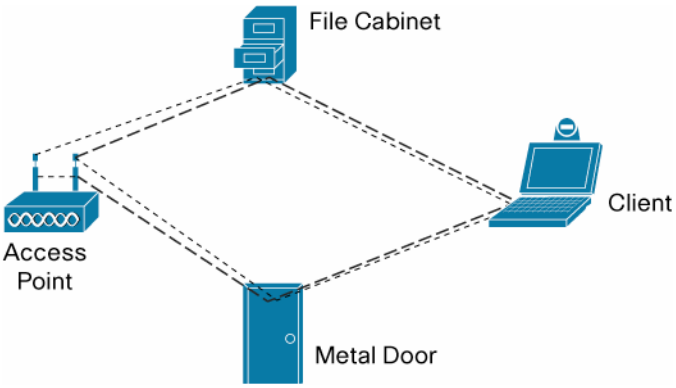
Transmit beamforming is useful only when transmitting to a single receiver. It is not possible to optimize the phase of the transmitted signals when sending broadcast or multicast transmissions. For this reason, in general networking applications, the utility of transmit beamforming is somewhat limited, providing improved SNR at the receiver for only those transmissions that are sent to that receiver alone. Transmit beamforming can increase the data rate available at greater distances from the AP. But, it does not increase the coverage area of an access point, since that is determined, in large part, by the ability to receive the beacons from the access point. Beacons are a broadcast transmission that does not benefit from transmit beamforming.

MIMO Technology: Multipath or Spatial Diversity

In typical indoor WLAN deployments—for example, offices, hospitals, and warehouses—the radio signal rarely takes the direct, shortest path from the transmitter to the receiver. This is because there is rarely "line of sight" between the transmitter and the receiver. Often there is a cube wall, door, or other structure that obscures the line of sight. All of these obstructions reduce the strength of the radio signal as it passes through them. Luckily, most of these environments are full of surfaces that reflect a radio signal as well as a mirror reflects light.

Imagine that all of the metallic surfaces, large and small, that are in an environment were actually mirrors. Nails and screws, door frames, ceiling suspension grids, and structural beams are all reflectors of radio signals. It would be possible to see the same WLAN access point in many of these mirrors simultaneously. Some of the images of the access point would be a direct reflection through a single mirror. Some images would be a reflection of a reflection. Still others would involve an even greater number of reflections. This phenomenon is called *multipath* (see Figure 3).

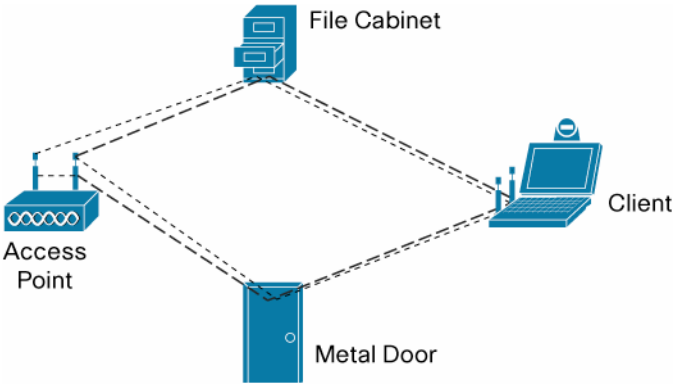
Figure 3. Multipath



When a signal travels over different paths to a single receiver, the time that the signal arrives at the receiver depends on the length of the path it traveled. The signal traveling the shortest path will arrive first, followed by copies or echoes of the signal slightly delayed by each of the longer paths that the copies traveled. When traveling at the speed of light, as radio signals do, the delays between the first signal to arrive and its copies is very small, only nanoseconds. (A rule of thumb for the distance covered at the speed of light is roughly one foot per one nanosecond.) This delay is enough to be able to cause significant degradation of the signal at a single antenna because all the copies interfere with the first signal to arrive.

A MIMO radio sends multiple radio signals at the same time and takes advantage of multipath. Each of these signals is called a *spatial stream*. Each spatial stream is sent from its own antenna, using its own transmitter. Because there is some space between each of these antennae, each signal follows a slightly different path to the receiver. This is called *spatial diversity*. Each radio can also send a different data stream from the other radios. The receiver has multiple antennas as well, each with its own radio. Each of the receive radios independently decode the arriving signals (see Figure 4.) Then, each radio's received signal is combined with the signals from the other receive radios. With a lot of complex math, the result is a much better receive signal than can be achieved with either a single antenna or even with transmit beamforming. One of the two significant benefits of MIMO is that it dramatically improves the SNR, providing more flexibility for the WLAN system designer.

Figure 4. Spatial Multiplexing



MIMO systems are described using the number of transmitters and receivers in the system—for example, 2x1 is “two by one,” meaning two transmitters and one receiver. 802.11n defines a number of different combinations for the number of transmitters and the number of receivers, from 2x1, equivalent to transmit beamforming, to 4x4. Each additional transmitter or receiver in the system increases the SNR. However, the incremental gains from each additional transmitter or receiver diminish rapidly. The gain in SNR is large for each step from 2x1 to 2x2 and to 3x2, but the improvement with 3x3 and beyond is relatively small. The use of multiple transmitters provides the second significant benefit of MIMO, the ability to use each spatial stream to carry its own information, providing dramatically increased data rates.

802.11n Radio Enhancements

In addition to MIMO technology, 802.11n makes a number of additional changes to the radio to increase the effective throughput of the WLAN. The most important of these changes are increased channel size, higher modulation rates, and reduced overhead. This section will describe each of these changes and the effect they have on WLAN throughput.

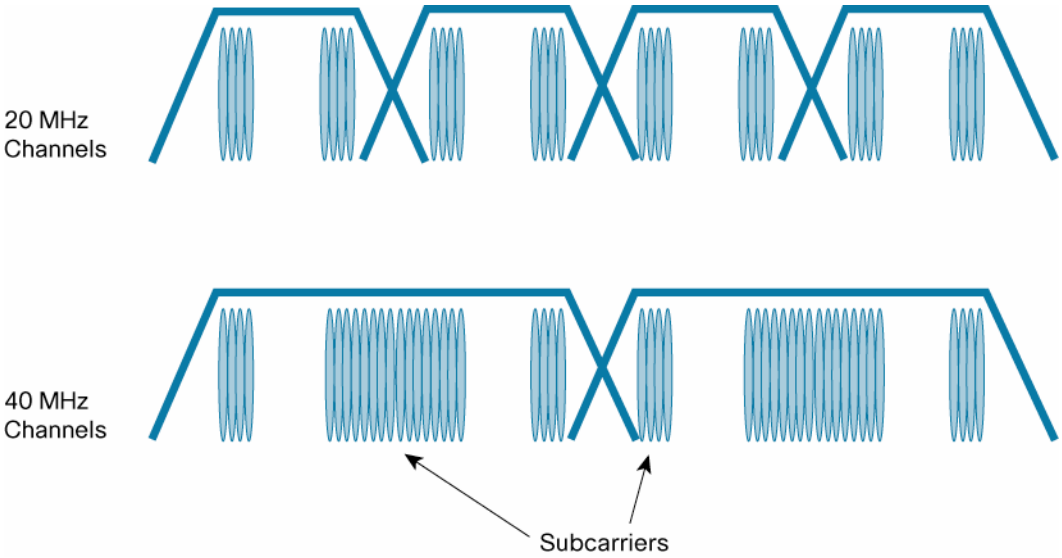
20- and 40-MHz Channels

The original 802.11 direct sequence radio and the 802.11b extension to the base standard use a radio channel spacing that is 22-MHz wide. 802.11a and 802.11g use radio channel spacings that are 20-MHz wide. Because 802.11g is an extension to 802.11b, 802.11g spaces its channels just as 802.11b does, every 22-MHz. The size, or bandwidth, of the radio channel is an important measure of the efficiency of the radio. This is called the *spectral efficiency* and is measured in bits per hertz. The spectral efficiency of 802.11b is one-half the bits per hertz (for example, 11 Mbps in 22 MHz). 802.11a and 802.11g have higher spectral efficiency, as much as 2.7 bits per hertz at 54 Mbps.

Using exactly the same technology as 802.11a and 802.11g, some proprietary WLAN systems are available that provide up to 108 Mbps. These proprietary systems use a simple technique to double the data rate of 802.11a and 802.11g. They use two channels at the same time. This is called *channel bonding*. With channel bonding, the spectral efficiency is the same as 802.11a and 802.11g, but the channel bandwidth is twice as great. This provides a simple way of doubling the data rate.

802.11n uses both 20-MHz and 40-MHz channels. Like the proprietary products, the 40-MHz channels in 802.11n are two adjacent 20-MHz channels, bonded together. When using the 40-MHz bonded channel, 802.11n takes advantage of the fact that each 20-MHz channel has a small amount of the channel that is reserved at the top and bottom, to reduce interference in those adjacent channels. When using 40-MHz channels, the top of the lower channel and the bottom of the upper channel don’t have to be reserved to avoid interference. These small parts of the channel can now be used to carry information. By using the two 20-MHz channels more efficiently in this way, 802.11n achieves slightly more than doubling the data rate when moving from 20-MHz to 40-MHz channels (see Figure 5).

Figure 5. 20 MHz and 40 MHz Channels



Higher Modulation Rates

The original 802.11 direct sequence radio transmitted a *symbol* to represent each bit (or set of bits) sent from a transmitter to a receiver. Each symbol lasted one microsecond. A symbol consisted of a fixed series of 11 *chips*. Each chip was modulated on the radio signal using a phase shift key (PSK) technique. For the 1-Mbps data rate, a single symbol was sent using binary PSK every microsecond. The 2-Mbps rate sent two symbols each microsecond, using quaternary (4-phase) PSK (QPSK). 802.11b extended the direct sequence radio by coding more bits into each symbol, while continuing the use of the QPSK modulation method. This allowed the extension of data rates to 11 Mbps.

802.11a and 802.11g changed the way information is transmitted on the radio signal. These standards adopted a method called *orthogonal frequency division multiplexing* (OFDM). OFDM divides a radio channel into a large number of smaller channels, each with its own *subcarrier* signal (see Figure 5 above). Each of these carrier signals is able to convey information independent of all the other carrier signals. It is roughly the same as having a group of independent radios bunched together.

For 802.11a and 802.11g, a symbol lasts 4 microseconds, including an 800 nanosecond *guard interval*. For the highest data rate, 54 Mbps, each symbol carries 216 data bits. These data bits are spread out over 48 subcarriers. In addition, there are 72 error- correction bits sent in each symbol at 54 Mbps, for a total of 288 bits in the symbol. To pack this many bits on each subcarrier, the subcarrier is modulated using 64 QAM (Quadrature Amplitude Modulation), 16 times the highest modulation rate of 802.11b. This means that each subcarrier is able to carry 6 bits (a combination of data and error correction bits).

802.11n continues to use OFDM and a 4-microsecond symbol, similar to 802.11a and 802.11g. However, 802.11n increases the number of subcarriers in each 20-MHz channel from 48 to 52. This marginally increases the data rate to a maximum of 65 Mbps, for a single-transmit radio. 802.11n provides a selection of eight data rates for a transmitter to use and also increases the number of transmitters allowable to four. For two transmitters, the maximum data rate is 130 Mbps. Three transmitters provide a maximum data rate of 195 Mbps. The maximum four transmitters can deliver 260 Mbps. In total, 802.11n provides up to 32 data rates for use in a 20-MHz channel.

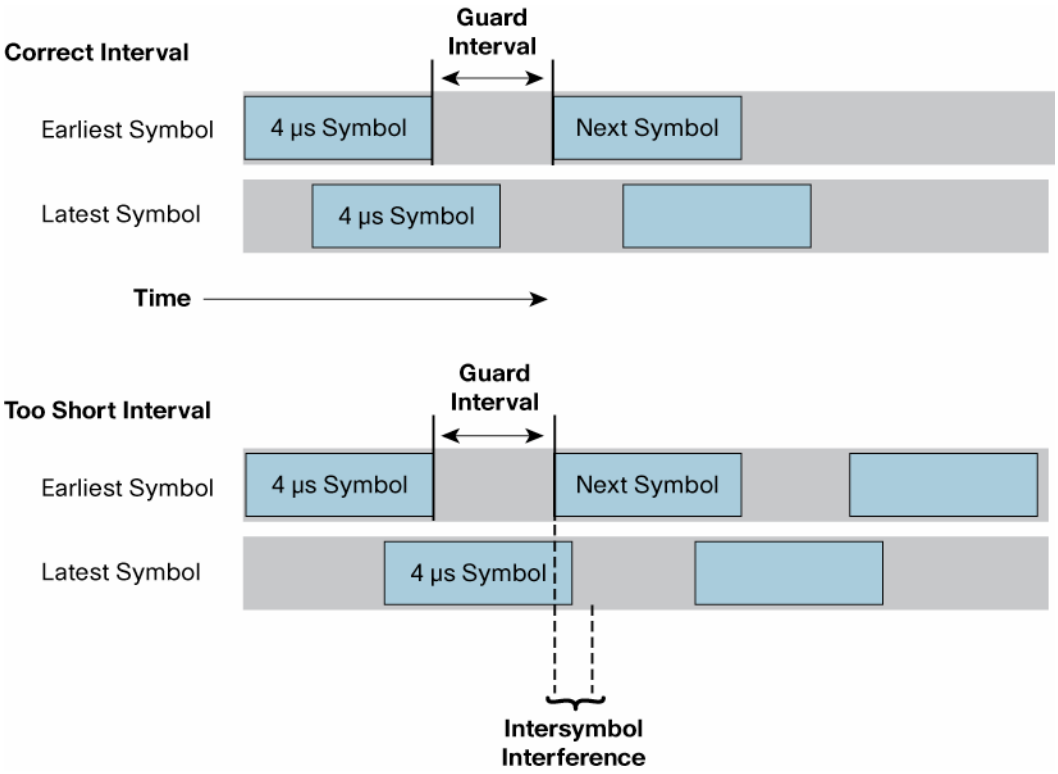
When using 40-MHz channels, 802.11n increases the number of subcarriers available to 108. This provides a maximum data rate of 135 Mbps, 270 Mbps, 405 Mbps, and 540 Mbps for one through four transmitters, respectively. Similarly, there are eight data rates provided for each transmitter, 32 in total, for the 40-MHz channel.

The rates described so far use the same modulation (*equal* modulation) on all of the subcarriers—for example, all the subcarriers use QPSK or 64 QAM.. This is the same as 802.11a and 802.11g. 802.11n adds the ability to modulate different spatial streams using different modulation methods—that is, some spatial streams use QPSK, some other spatial streams use 16 QAM, and yet other spatial streams use 64 QAM. This dramatically increases the number of data rates available to be used. In fact, there are dozens more possible data rates, using this *unequal modulation* method. However, it is unlikely that many practical implementations would be able to take advantage of this method, as it requires a significant amount of feedback from the receiver to the transmitter to identify the individual spatial streams that must use each of the different modulation methods.

Lowered Overhead: Guard Interval

The guard interval that is part of each OFDM symbol is a period of time that is used to minimize intersymbol interference. This type of interference is caused in multipath environments when the beginning of a new symbol arrives at the receiver before the end of the last symbol is done. These two symbols arrive over two different paths. The “late” symbol that has not yet been completely received when the new symbol arrives traveled a longer path than the new symbol (see Figure 6). When this situation occurs, the interference it causes reduces the effective SNR of the radio link. The guard interval is a quiet period between symbols that provides for the arrival of late symbols over long paths. The length of the guard interval is selected for the severity of the multipath environment. 802.11a and 802.11g use 800 nanoseconds as the guard interval, allowing for path differences of 800 feet.

Figure 6. Guard Interval



The default mode for 802.11n is also to use 800 nanoseconds as the guard interval. However, if the multipath environment is not as strict about a required allowance for 800 foot differences in paths between transmitter and receiver, 802.11n also provides a reduced guard interval of 400 nanoseconds. This reduces the symbol time from 4 microseconds to 3.6 microseconds. This reduced symbol time has a corresponding effect in increasing data rates. For 20-MHz channels, maximum data rates for one to four transmitters with the reduced guard interval are 72, 144, 216, and 288 Mbps for a 20- MHz channel and 150, 300, 450, and 600 Mbps for a 40-MHz channel.

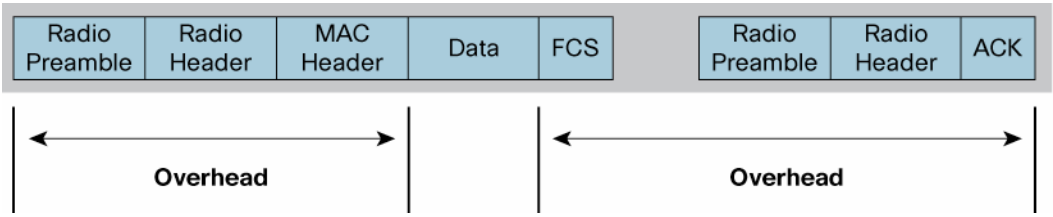
MAC Enhancements

There is only so much improvement that can be obtained in 802.11 by increasing the data rate of the radio. There is a significant amount of fixed overhead in the MAC layer protocol, and in the interframe spaces and acknowledgements of each frame transmitted, in particular. At the highest of data rates, this overhead alone can be longer than the entire data frame. In addition, contention for the air and collisions also reduce the maximum effective throughput of 802.11. 802.11n addresses these issues by making changes in the MAC layer to improve on the inefficiencies imposed by this fixed overhead and by contention losses.

Aggregation

Every frame transmitted by an 802.11 device has fixed overhead associated with the radio preamble and MAC frame fields that limit the effective throughput, even if the actual data rate was infinite (see Figure 7).

Figure 7. Overhead



To reduce this overhead, 802.11n introduces *frame aggregation*. Frame aggregation is essentially putting two or more frames together into a single transmission. 802.11n introduces two methods for frame aggregation: Mac Service Data Units (MSDU) aggregation and Message Protocol Data Unit (MPDU) aggregation. Both aggregation methods reduce the overhead to only a single radio preamble for each frame transmission (see Figure 8).

Figure 8. Aggregation

Because multiple frames are now sent in a single transmission, the number of potential collisions and the time lost to backoff is significantly reduced. The maximum frame size is also increased in 802.11n, to accommodate these large, aggregated frames. The maximum frame size is increased from 4 KB to 64 KB. One limitation of frame aggregation is that all the frames that are aggregated into a transmission must be sent to the same destination; that is, all the frames in the aggregated frame must be addressed to the same mobile client or access point. Another limitation is that all the frames to be aggregated have to be ready to transmit from the client or access point at the same time, potentially delaying some frames to wait for additional frames, in order to attempt to send a single aggregate frame. A third limitation of aggregation is that the maximum frame size that can be successfully sent is affected by a factor called *channel coherence time*. Channel coherence time depends on how quickly the transmitter, receiver, and other items in the environment are moving. The faster things are moving the smaller the maximum frame size can be as the data rate is reduced, i.e., the time for the transmission must be less than the channel coherence time.

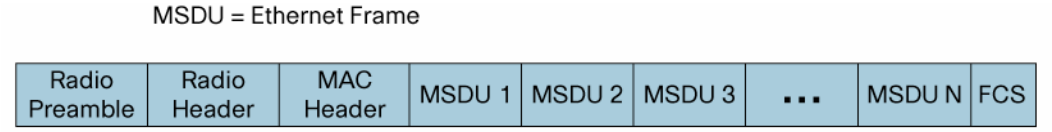
There are slight differences in the two aggregation methods that result in differences in the efficiency gained. These two methods are described here.

Mac Service Data Units Aggregation

MSDU aggregation is the more efficient of the two aggregation methods. It relies on the fact that an access point receives frames from its Ethernet interface, to be translated to 802.11 frames and then transmitted to a mobile client. Similarly, most mobile client protocol stacks create an Ethernet frame, which the 802.11 driver must translate to an 802.11 frame before transmission. In both these cases, the “native” format of the frame is Ethernet, and it is then translated to 802.11 format for transmission.

Theoretically, MSDU aggregation allows frames for many destinations to be collected into a single aggregated frame for transmission. Practically, however, MSDU aggregation collects Ethernet frames for a common destination, wraps the collection in a single 802.11 frame, and then transmits that 802.11-wrapped collection of Ethernet frames (see Figure 9). This method is more efficient than MPDU aggregation, because the Ethernet header is much shorter than the 802.11 header.

Figure 9. MSDU Aggregation



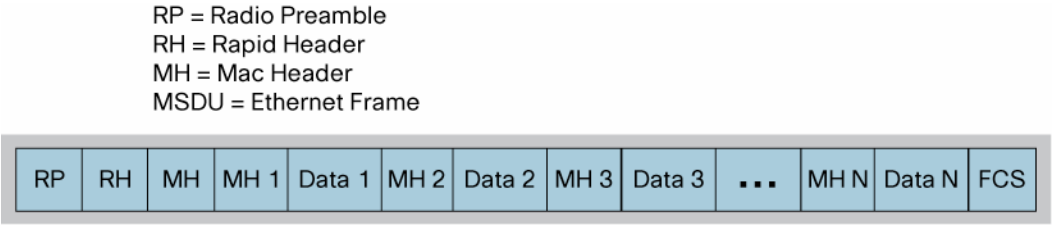
For a mobile device, the aggregated frame is sent to the access point, where the constituent Ethernet frames are forwarded to their ultimate destinations. For an access point, all of the constituent frames in the aggregated frame must be destined to a single mobile client, since there is only a single destination in each mobile client.

With MSDU aggregation, the entire, aggregated frame is encrypted once using the security association of the destination of the outer 802.11 frame wrapper. A restriction of MSDU aggregation is that all of the constituent frames must be of the same quality-of -service (QoS) level. It is not permitted to mix voice frames with best-effort frames, for example.

Mac Protocol Data Units Aggregation

MPDU aggregation is slightly different from MSDU aggregation. Instead of collecting Ethernet frames, MPDU aggregation translates each Ethernet frame to 802.11 format and then collects the 802.11 frames for a common destination. The collection doesn't require a wrapping of another 802.11 frame, since the collected frames already begin with an 802.11 MAC header (see Figure 10).

Figure 10. MPDU Aggregation



MPDU aggregation does require that all the 802.11 frames that constitute the aggregated frame have the same destination address. However, this results in the same behavior as MSDU aggregation, since the destination of all frames sent by a mobile client is that client's access point, where the 802.11 frames are translated to Ethernet and forwarded to the ultimate destination. Similarly, the destination of any frame sent by the access point is a single mobile client.

With MPDU aggregation, it is possible to encrypt each constituent frame independently, using the security association for each individual 802.11 destination address. This does not have any effective difference from the encryption done in MSDU aggregation, as all frames sent by a mobile client are encrypted using the security association for the access point, and all frames sent by the access point are encrypted using the security association for the single mobile client that is the intended recipient of the frame.

Similar to MSDU aggregation, MPDU aggregation requires that all of the constituent frames be of the same QoS level.

The efficiency of the MPDU aggregation method is lower than that of the MSDU aggregation method, because of the extra overhead of the individual 802.11 frame headers for each constituent frame. The efficiency is further reduced when the encryption is used. Encryption adds overhead to

each of the constituent frame in MPDU aggregation, where MSDU aggregation incurs overhead for a single encryption of the outer 802.11 wrapper.

Block Acknowledgement

For the 802.11 MAC protocol to operate reliably, each of the frames transmitted an individual address—that is, not multicast or broadcast frames—is immediately acknowledged by the recipient. MSDU aggregation requires no changes to this operation. The aggregated frame is acknowledged, just as any 802.11 frame is acknowledged. This is not the case for MPDU aggregation. For MPDU aggregation, each of the individual constituent 802.11 frames must be acknowledged. The mechanism to deal with this requirement that 802.11n introduces is called *block acknowledgement*.

Block acknowledgement compiles all the acknowledgements of the individual constituent frames produced by MPDU aggregation into a single frame returned by the recipient to the sender. This allows a compact and rapid mechanism to implement selective retransmission of only those constituent frames that are not acknowledged. In environments with high error rates, this selective retransmission mechanism can provide some improvement in the effective throughput of a WLAN using MPDU aggregation over that of one using MSDU aggregation, because much less is retransmitted when an error affects some of the constituent frames of an MPDU aggregated frame as compared to an MSDU aggregated frame.

Lower Overhead: Reduced Interframe Space

When aggregation of frames is not possible, 802.11n provides a mechanism to reduce the overhead involved with transmitting a stream of frames to different destinations. This mechanism reduces the interframe space between receiving a frame, typically an acknowledgement frame, and sending a subsequent frame. The 802.11e extension for quality of service added the ability for a single transmitter to send a burst of frames during a single, timed *transmit opportunity*. During the transmit opportunity, the sender does not need to perform any random backoff between transmissions, separating its frames by the smallest allowable interframe space, the short interframe space (SIFS).

802.11n improves on this mechanism, reducing the overhead between frames, by specifying an even smaller interframe space, called the *reduced interframe space* (RIFS). RIFS cuts down further on the dead time between frames, increasing the amount of time in the transmit opportunity that is occupied by sending frames. The one unfortunate aspect of using RIFS is that it is restricted to being used only in *greenfield deployments*—that is, only deployments where there are no legacy 802.11a, b, or g devices in the area.

Power Savings

Radios are power hungry. Operating several radios requires even more power. To address this situation, 802.11n has extended the power management capability of the 802.11 MAC. There are two extensions beyond the existing mechanisms established in the original standard and the automatic power save delivery added in 802.11e. The two new mechanisms provided by 802.11n are Spatial Multiplexing Power Save and Power Save Multi-Poll.

Spatial Multiplexing Power Save

The spatial multiplexing (SM) power save mode allows an 802.11n client to power down all but one of its radios. This power save mode has two submodes of operation: static operation and dynamic operation.

The static SM power save mode has the client turn off all but a single radio, becoming essentially equivalent to an 802.11a or 802.11g client. The client's access point is notified that the client is now operating in the static single-radio mode, requiring the access point to send only a single spatial stream to this client until the client notifies the access point that its additional radios are again enabled and operating. This notification of the access point is done using a new management frame, defined by 802.11n, telling the access point that the client is in static SM power save mode.

The dynamic SM power save mode also turns off all but one of the client's radios. But in this mode of operation, the client can rapidly enable its additional radios when it receives a frame that is addressed to it. The client can immediately return to the low power state by disabling its additional radios immediately after its frame reception is complete. In this mode of operation, the access point typically sends a request-to-send (RTS) frame to the client, to wake its radios, prior to sending the client a data or management frame. On receiving the RTS frame, the client enables its radios and responds with a *clear-to-send* (CTS) frame. All of its radios are now ready to receive the multiple spatial streams sent by all the radios in the access point. To use this power save mode, the 802.11n client sends a new management frame to its access point, informing the access point that it is in dynamic SM power save mode.

Power Save Multi-Poll

The Power Save Multi-Poll (PSMP) mode extends the Automatic Power Save Delivery (APSD) mechanism defined in 802.11e. Using APSD, a client informs an access point that frames of some specified QoS levels should be buffered (delivery enabled) until the client requests them, while frames sent by the client at another set of specified QoS levels are to be considered triggers that will cause the delivery of buffered frames. In 802.11e, this is typically used by WLAN handsets to reduce power expenditures during active calls, by specifying that voice packets be buffered at the access point until a voice packet is sent by the client. Once the access point receives the client's voice packet, it delivers the voice packet that is buffered and waiting to be sent.

PSMP provides the same delivery-enable and trigger concepts, extending the ability of the client to schedule the frames that it transmits as the trigger for delivering the downlink frames. This scheduling mechanism reduces the contention between clients and between the client and the access point. Reducing contention also reduces the time the client spends in backoff and reduces the number of times a frame must be transmitted before it is delivered successfully. This dramatically improves power conservation in the clients. PSMP is also a dynamic method that immediately adjusts to changes in traffic demand by the clients using it.

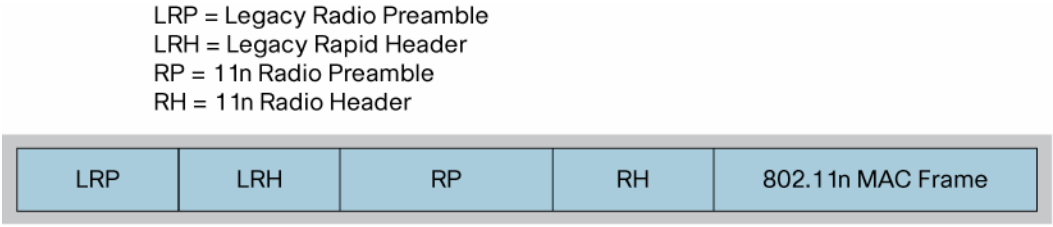
Backward Compatibility

Compatibility with existing 802.11a, b, and g devices is a critical issue addressed in 802.11n. Just as 802.11g provides a protection mode for operation with 802.11b devices, 802.11n has a number of mechanisms to provide backward compatibility with 802.11 a, b, and g devices, allowing these devices to understand the information necessary to allow 802.11n devices to operate in the same area.

For quite a long time, 802.11n will need to operate in the presence of legacy 802.11a, b, and g devices. This mixed-mode operation will continue until all the devices in an area have been upgraded or replaced with 802.11n devices. The mixed-mode protection mechanism for 802.11n is quite similar to the protection mechanism of 802.11g.

Like 802.11g, 802.11n transmits a signal that can't be decoded by devices built to an earlier standard. To avoid descending into absolute chaos in the presence of massive interference and collisions, 802.11n operating in mixed mode transmits a radio preamble and signal field that can be decoded by 802.11a and 802.11g radios (see Figure 11). This provides enough information to the legacy radios to allow them to indicate that there is another transmission on the air and how long that transmission will last. Following the legacy preamble and signal field, the 802.11n device sends the remaining information using 802.11n rates and its multiple spatial streams, including an 802.11n preamble and signal field.

Figure 11. Backward Compatibility



In addition to the legacy preamble and signal field, it can also be necessary to use additional protection mechanisms provided by 802.11g to allow the MAC in legacy devices to correctly determine when it is allowed to transmit and when it must perform backoff before transmission. The mechanism provided by 802.11g and utilized by 802.11n when either 802.11g or 802.11a devices are present is the CTS-to-self mechanism. CTS-to-self allows the 802.11n device to transmit a short CTS frame, addressed to itself, that includes the timing information necessary to be communicated to the neighboring legacy MACs that will protect the 802.11n transmission that will follow. The CTS frame must be transmitted using one of the legacy data rates that a legacy device will be able to receive and decode.

The cost of this additional legacy preamble and signal field, as well as the CTS-to-self, is more overhead on every 802.11n transmission. This reduces the benefits of all the 802.11n improvements, resulting in significantly lower effective throughput by 802.11n devices in mixed environments. Similar to 802.11g, legacy devices don't need to be associated to the same access point as an 802.11n device to require the use of protection mechanisms. If there are legacy devices on the same channel on any nearby access points, that will cause protection mechanisms to be invoked as well.

It can be expected that protection mechanisms will be in use in the 2.4-GHz band (802.11b and 802.11g)) until nearly every legacy device has disappeared. This is because there are too few channels available in that band to effectively overlay pure 802.11n WLANs in the same areas as legacy 2.4-GHz WLANs. Given the larger number of channels available in the 5-GHz band in many countries, it is possible that two completely separate WLANs could be operating in the same area, with 802.11a operating on one set of channels and 802.11n operating on a different, nonintersecting set of channels. This would result in 802.11n operating in pure high-throughput (greenfield) mode, achieving the highest effective throughput offered by this new standard.

Summary of 802.11n Technology

To summarize the benefits of 802.11n technology, it is simplest to say that there are two major areas of improvement over previous 802.11 devices. The first area of improvement is in the use of MIMO technology to achieve greater SNR on the radio link. The second area of improvement is in

the greater efficiencies in both radio transmissions and the MAC protocol. These improvements translate into benefits in three areas: reliability, predictable coverage, and throughput.

Reliability

Greater SNR on the radio link translates directly to more reliable communication, often at higher data rates. Higher SNR means that more interference is needed to corrupt a transmission. This means greater client densities can be supported.

Predictable Coverage

The use of multiple spatial streams provided by MIMO technology means that there will be fewer dead spots in a coverage area. Areas that previously suffered from destructive multipath interference now make use of that same multipath effect to provide robust communication.

Throughput

The efficiency improvement in 802.11n provides a greater transfer of the high bit rates of the 802.11n radio to effective throughput seen by actual applications, at least in greenfield deployments. Even in mixed-mode deployments with legacy 802.11 devices, 802.11n will provide greater effective throughput, although significantly less than the greenfield mode.

Migration to 802.11n

The migration to 802.11n, at least 802.11n draft 2.0, has already begun. 802.11n client devices, beginning with laptops, are already available. New client devices will begin to appear over the next several quarters, until 802.11n draft 2.0 will be the default WLAN adapter in any portable or mobile device. These client devices are completely compatible with existing 802.11a, b, and g access points and will operate just as existing devices do today. Planning to migrate the infrastructure portion of a network to support 802.11n on these new client devices is straightforward.

Planning

There are several areas to consider when planning the migration of a network to support 802.11n. Because of the higher speeds and greater power requirements of 802.11n access points, planning the migration needs to take into account more than just the access point.

Radio Bands

802.11n operates in both the 2.4-GHz (802.11b and g) and 5-GHz (802.11a) radio bands. Planning for each of the radio bands should be done independently, because of the constraints that are sometimes very different for each band.

The 2.4-GHz Band

The 2.4-GHz band is no more than 100 MHz wide, and often much less than that in many countries. The same channelization that is used for 802.11b and 802.11g can be used for 802.11n operating in this band. However, the use of the 40-MHz mode of operation of 802.11n is not recommended in this band, because a significant portion of the band will suffer from interference from a single 40-MHz transmitter. In addition, it is required that the it is required that the second 20 MHz channel, concatenated with the original 20 MHz channel to form the 40 MHz channel, must be free of any legacy transmissions. This drastically reduces the chance that any 40 MHz operations will be feasible in this band.

In much of the world where it is typical to utilize three nonoverlapping channels in this band, a single 40-MHz access point will present a significant challenge to developing a channel plan that will provide adequate capacity in most enterprises. Even when all legacy 802.11b and g devices are removed from the band, it will be difficult to deploy access points utilizing the 40-MHz channels

in this band. There is just not enough bandwidth available to even begin to duplicate the three nonoverlapping channels of the legacy layout.

The 5-GHz Band

The 5-GHz band has been opened up significantly in much of the world, due to recent changes by many regulatory agencies. There are significantly more channels available in the 5-GHz band than in the 2.4-GHz band. The larger number of channels in this band makes planning the deployment of an 802.11n network much simpler, even while allowing for 40-MHz operation.

There are at least two possible ways to migrate to 802.11n in the 5-GHz band. The first way is to replace individual legacy access points with 802.11n access points as budget allows and user demand for additional capacity dictates. This gradual migration can be accomplished over a planned period of time or as the need arises. This migration method would have the new 802.11n access point operating on the channel of the legacy access point it replaced. The new 802.11n access point would support 802.11n clients, as well as legacy 802.11a clients. It would operate in mixed mode, providing protection for the legacy 802.11a clients. Eventually, as the last legacy access point is replaced and the last legacy client is retired, the entire set of new 802.11n access points could be switched to operate in greenfield mode.

The second way to migrate to 802.11n would be to reassign the channels on some of the legacy access points to free a set of channels that could be used for 802.11n exclusively. Then as budget allows and demand dictates, new 802.11n access points would be *added* to the existing WLAN deployment, operating in parallel in overlapping areas with the legacy access points. The new 802.11n access points, however, would support only 802.11n devices and be able to operate in the greenfield mode, providing the greatest benefits of the new standard. Eventually, as 802.11n access points cover the entire area covered by the legacy access points—802.11n clients would have the ability to operate in greenfield mode everywhere, while the legacy access points still provide service to the legacy clients. Once the last legacy client is retired, the legacy access points can also be retired.

Wired Infrastructure Stresses

Today's dual-band access points can theoretically put a load on their Ethernet connections of as much as 108 Mbps. Practically, however, due to the inefficiencies of the 802.11 protocol, they top out at a peak load of 50 to 60 Mbps.

802.11n access points can demand much more of their Ethernet connection. With the higher bit rates on the air and the improved efficiency of the protocol, it is possible that a single dual-band 802.11n access point supporting a 20-MHz channel in the 2.4-GHz band and a 40-MHz channel in the 5-GHz band can place a peak demand on its Ethernet connection of as much as 300 to 400 Mbps. Obviously, this is greater than a single or double 100-Mbps Ethernet connection can support.

For this reason, planning for a migration to support 802.11n should also include planning to upgrade the edge Ethernet switching capabilities to support a 1-Gbps connection to each 802.11n access point. This will eliminate any bottlenecks that might occur in areas of high-capacity demand by the 802.11n clients.

Power Requirements

Most current 802.11 access points can be operated using Power over Ethernet (PoE) or 802.3af. This provides up to 15 watts of DC power over the standard category 5 Ethernet cable. 802.11n, with its multiple radios, requires more power than 802.3af can deliver. Fortunately, the IEEE 802.3

(Ethernet) Working Group has a solution for this. The 802.3at standard provides double the power of the original 802.3af standard. At 30 watts, 802.3at provides sufficient power to operate an 802.11n access point.

Migration to 802.11n should involve considering how to provide the source for this new PoE standard. There are at least two alternatives. The first alternative is to provide the power directly from the Ethernet switch to which the access point is connected. This might involve upgrading the Ethernet switch to one that supplies the new PoE standard, or, since a new 1-Gbps switch might be necessary to support the access point, to add a small 1-Gbps Ethernet switch at the edge of the network that also supplies 802.3at PoE.

The second alternative to supplying 802.3at PoE is to use an inline power injector. The power injector is installed in the wiring closet, along with the Ethernet switches, and is inserted into the Ethernet cable between the switch and the access point. If upgrading or adding an Ethernet switch is not part of the migration plan, this is a sound alternative solution.

Access Point Deployment

Planning the positioning of the 802.11n access points should also be considered during migration planning. If the migration plan is to gradually replace the existing legacy access points, there is no further deployment planning necessary. However, if the new 802.11n access points are to be deployed in a new installation or along side an existing deployment, it is possible to use the increased SNR provided by 802.11n to cover greater areas per access point, although at a cost of reducing the overall capacity of the resulting 802.11 WLAN. Remember that SNR is like money in the bank. It can be used for either increased data rate, increased range, or a little of both. But it can’t be used for the maximum of both at the same time.

Investment Protection

One of the most significant decisions in a migration plan is which manufacturer’s equipment to use. This decision should include consideration of the fact that the final 802.11n standard is not yet ratified. The ability of any draft 802.11n equipment to be upgraded must be considered a paramount requirement of the migration. While there have been many claims that draft 802.11n devices will be upgradeable to the final, ratified standard with only a software download, there is no telling how the final standard may differ from the very first draft approved by the 802.11 Working Group. The ability to upgrade the hardware of a draft 802.11n access point, without replacing the entire access point, is an important feature to provide protection of the investment made in the new draft 802.11n access points.

Promises and Expectations

There is great promise of tremendous throughput in the 802.11n standard. A migration plan should take into consideration that much of this promise might not be realized for several years. In some radio bands, the promise of high throughput might never be realizable in an enterprise environment. Only when operating in greenfield mode, can an 802.11n deployment reach its full potential.

Importance of Certification

Until a final, ratified 802.11n standard is available, and even after that, certification of the equipment used in the network is crucial to having any chance for a successful migration to 802.11n. The Wi-Fi Alliance is currently certifying equipment to 802.11n draft 2.0, providing a measure of certainty that equipment from different vendors will interoperate using this new technology. However, most equipment that is certified by the Wi-Fi Alliance is tested against a

small number of devices that are present in the Wi-Fi Alliance's testbed. For 802.11n draft 2.0, an access point is tested against only five 802.11n clients, only one of which is commercially available for sale to the general public. Only the access points in the Wi-Fi Alliance's testbed itself are tested against every 802.11n draft 2.0 client. For the greatest assurance of interoperability with any 802.11n client that might arrive in the WLAN, only the access points in the testbed can claim to have been tested against them all.

Conclusions

There is no reason not to begin installing a new 802.11n WLAN or migrating an existing WLAN to support 802.11n. Some care, though, should be taken when selecting the 802.11n equipment to install.

In particular, to reap the greatest benefits from 802.11n while protecting your investment in the newly installed access points from becoming obsolete should the final ratified 802.11n standard not be compatible with the draft 2.0 radio, the access points selected for the 802.11n network should provide for hardware upgradeability, as well as software upgradeability. Those same access points should be tested against the widest variety of client devices possible, to help ensure that whatever equipment a user might bring into the 802.11n network is compatible with the access points deployed.

802.11n has the ability to dramatically increase the capacity of a WLAN and the effective throughput of every client. The time to begin moving to this new standard is as soon as it is necessary to add a new access point, in order to address the demand for additional capacity in the WLAN. There is no need to wait and see how the standard development wends its way to completion, if careful decisions are made when planning to move to 802.11n.

Americas Headquarters
Cisco Systems, Inc.
170 West Tasman Drive
San Jose, CA 95134-1706
USA
www.cisco.com
Tel: 408 526-4000
800 553-NETS (6387)
Fax: 408 527-0883

Asia Pacific Headquarters
Cisco Systems, Inc.
168 Robinson Road
#28-01 Capital Tower
Singapore 068912
www.cisco.com
Tel: +65 6317 7777
Fax: +65 6317 7799

Europe Headquarters
Cisco Systems International BV
Haarlerbergpark
Haarlerbergweg 13-19
1101 CH Amsterdam
The Netherlands
www-europe.cisco.com
Tel: +31 0 800 020 0791
Fax: +31 0 20 357 1100

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at www.cisco.com/go/offices.

Printed in USA C11-427843-00 8/07