CHAPTER 5

Cisco Unified Wireless QoS

Introduction

This chapter describes quality of service (QoS) in the context of WLAN implementations. This chapter describes WLAN QoS in general, but does not provide in-depth coverage on topics such as security, segmentation, and voice over WLAN (VoWLAN), although these topics have a QoS component. This chapter also provides information on the features of the Cisco Centralized WLAN Architecture.

This chapter is intended for those who are tasked with designing and implementing enterprise WLAN deployments using the Cisco Unified Wireless technology.

QoS Overview

QoS refers to the capability of a network to provide better service to selected network traffic over various network technologies. QoS technologies provide the following benefits:

- Provides building blocks for business multimedia and voice applications used in campus, WAN, and service provider networks
- Allows network managers to establish service level agreements (SLAs) with network users
- Enables network resources to be shared more efficiently and expedites the handling of mission-critical applications
- Manages time-sensitive multimedia and voice application traffic to ensure that this traffic receives higher priority, greater bandwidth, and less delay than best-effort data traffic

With QoS, bandwidth can be managed more efficiently across LANs, including WLANs and WANs. QoS provides enhanced and reliable network service by:

- Supporting dedicated bandwidth for critical users and applications
- Controlling jitter and latency (required by real-time traffic)
- Managing and minimizing network congestion
- Shaping network traffic to smooth the traffic flow
- Setting network traffic priorities

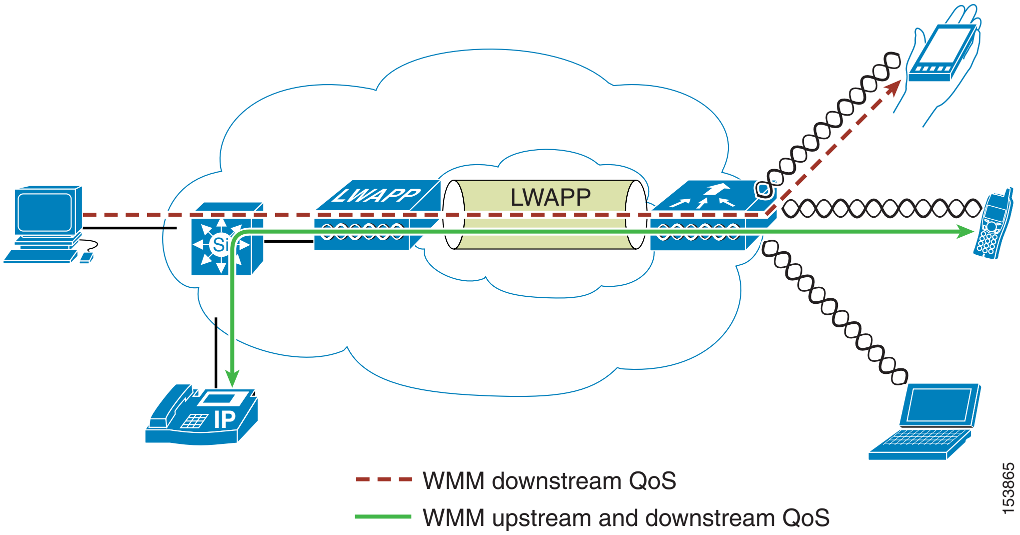
Wireless QoS Deployment Schemes

In the past, WLANs were mainly used to transport low-bandwidth, data-application traffic. Currently, with the expansion of WLANs into vertical (such as retail, finance, and education) and enterprise environments, WLANs are used to transport high-bandwidth data applications, in conjunction with time-sensitive, multimedia applications. This requirement led to the necessity for wireless QoS.

Several vendors, including Cisco, support proprietary wireless QoS schemes for voice applications. To speed up the rate of QoS adoption and to support multi-vendor time-sensitive applications, a unified approach to wireless QoS is necessary. The IEEE 802.11e working group within the IEEE 802.11 standards committee has completed the standard definition, but adoption of the 802.11e standard is in its early stages, and as with many standards there are many optional components. Just as occurred with 802.11 security in 802.11i, industry groups such as the WiFi Alliance and industry leaders such as Cisco are defining the key requirements in WLAN QoS through their WMM and CCX programs, ensuring the delivery of key features and interoperation through their certification programs.

Cisco Unified Wireless Products support Wi-Fi MultiMedia (WMM), a QoS system based on the IEEE 802.11e draft that has been published by the Wi-Fi Alliance. An example deployment of wireless QoS based on Cisco Unified Wireless technology features is shown in Figure 5-1.

Figure 5-1 Wireless QoS Deployment Example



QoS Parameters

QoS is defined as the measure of performance for a transmission system that reflects its transmission quality and service availability. Service availability is a crucial element of QoS. Before QoS can be successfully implemented, the network infrastructure must be highly available. The network transmission quality is determined by latency, jitter, and loss, as shown in Table 5-1.

Table 5-1 QoS Parameters

Transmission Quality	Description
Latency	Latency (or delay) is the amount of time it takes for a packet to reach the receiving endpoint after being transmitted from the sending endpoint. This time period is called the *end-to-end delay* and can be divided into two areas: fixed network delay and variable network delay. *Fixed network delay* includes encoding and decoding time (for voice and video), and the finite amount of time required for the electrical or optical pulses to traverse the media en route to their destination. *Variable network delay* generally refers to network conditions, such as congestion, that can affect the overall time required for transit.
Jitter	Jitter (or delay-variance) is the difference in the end-to-end latency between packets. For example, if one packet requires 100 mSec to traverse the network from the source endpoint to the destination endpoint and the next packet requires 125 mSec to make the same trip, then the jitter is calculated as 25 mSec.
Loss	Loss (or packet loss) is a comparative measure of packets successfully transmitted and received to the total number that were transmitted. Loss is expressed as the percentage of packets that were dropped.

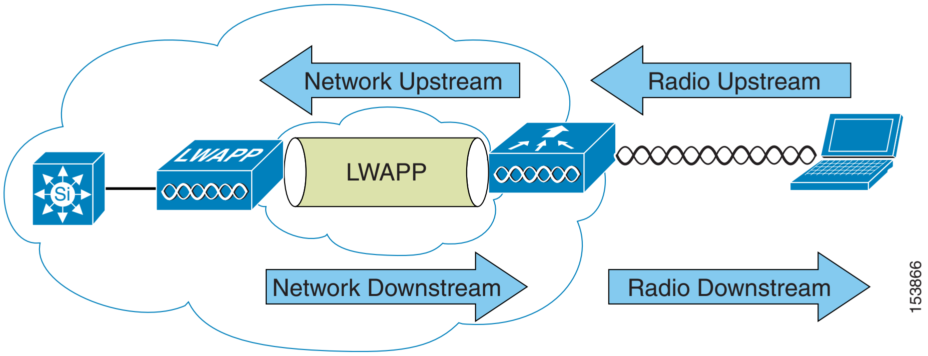
Upstream and Downstream QoS

Figure 5-2 illustrates the definition of QoS radio *upstream* and *downstream*.

The notations in Figure 5-2 refer to the following:

- *Radio downstream* QoS refers to the traffic leaving the AP and traveling to the WLAN clients. Radio downstream QoS is the primary focus of this chapter, because this is still the most common deployment. The client upstream QoS depends on the client implementation.
- *Radio upstream* QoS refers to traffic leaving the WLAN clients and traveling to the AP. WMM provides upstream QoS for WLAN clients supporting WMM.
- *Network downstream* refers to traffic leaving the WLC traveling to the AP. QoS can be applied at this point to prioritize and rate limit traffic to the AP. Configuration of Ethernet downstream QoS is not covered in this chapter.
- *Network upstream* refers to traffic leaving the AP, traveling to the WLC. The AP classifies traffic from the AP to the upstream network according to the traffic classification rules of the AP.

Figure 5-2 Upstream and Downstream QoS



QoS and Network Performance

The application of QoS features might not be easily detected on a lightly loaded network. If latency, jitter, and loss are noticeable when the media is lightly loaded, it indicates a system fault, poor network design, or that the latency, jitter, and loss requirements of the application are not a good match for the network.

QoS features start to impact application performance as the load on the network increases. QoS works to keep latency, jitter, and loss for selected traffic types within acceptable boundaries.

When providing only radio downstream QoS from the AP, radio upstream client traffic is treated as best-effort. A client must compete with other clients for upstream transmission as well as competing with best-effort transmission from the AP. Under certain load conditions, a client can experience upstream congestion and the performance of QoS-sensitive applications might be unacceptable despite the QoS features on the AP.

Ideally upstream and downstream QoS can be operated either by using WMM on both the AP and WLAN client, or by using WMM and a client's proprietary implementation.

Note Even without WMM support on the WLAN client, the Cisco Unified Wireless solution is able to provide network prioritization in both network upstream and network downstream situations.

Note WLAN client support for WMM does not mean that the client traffic automatically benefits from WMM. The applications looking for the benefits of WMM assign an appropriate priority classification to their traffic, and the operating system needs to pass that classification to the WLAN interface. In purpose-built devices, such as VoWLAN handsets, this is done as part of the design, but if implementing on a general purpose platform, such as a PC, application traffic classification and OS support must be implemented before the WMM features can be used to good effect.

802.11 DCF

Data frames in 802.11 are sent using the Distributed Coordination Function (DCF), which is composed of two main components:

- Interframe spaces (SIFS, PIFS, and DIFS)
- Random backoff (contention window)

DCF is used in 802.11 networks to manage access to the RF medium. A baseline understanding of DCF is necessary to deploy 802.11e-based enhanced distributed channel access (EDCA). See the IEEE 802.11 specification for more information on DCF at the following URL:
http://ieeexplore.ieee.org/xpl/standardstoc.jsp?isnumber=14251&isYear=1997

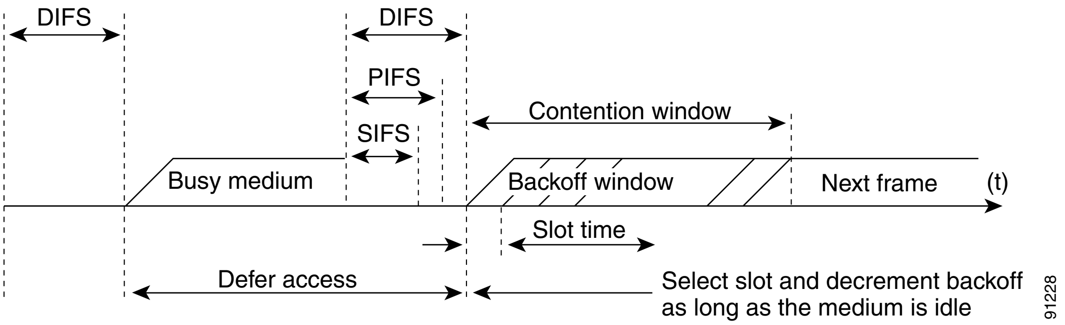
Interframe Spaces

802.11 currently defines three interframe spaces, as shown in Figure 5-3:

- Short interframe space (SIFS) 10 µs
- Point interframe space (PIFS) SIFS + 1 x slot time = 30 µs
- Distributed interframe space (DIFS) 50 µs SIFS + 2 x slot time = 50 µs

The interframe spaces, SIFS, PIFS, and DIFS allow 802.11 to control which traffic gets first access to the channel after carrier sense declares the channel to be free.

Figure 5-3 Interframe Spaces (IFS)



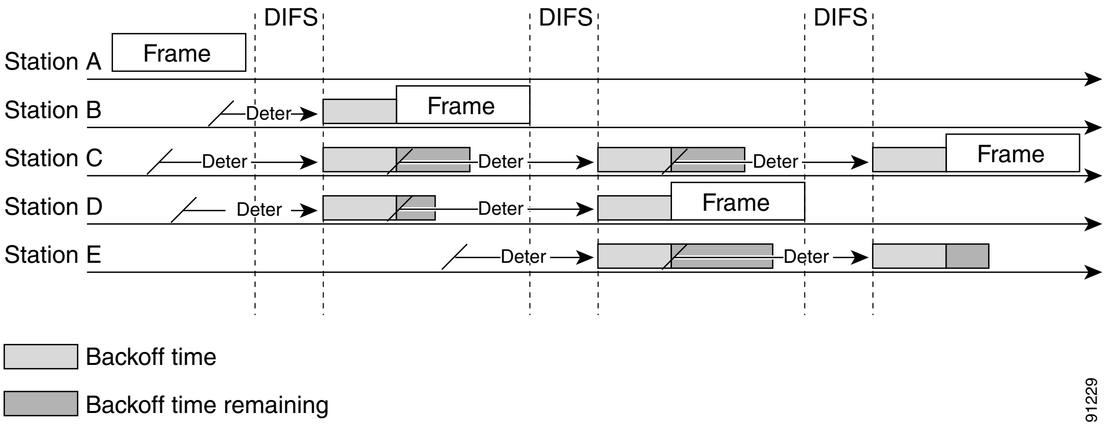
Random Backoff

When a data frame using Distributed Coordination Function (DCF), shown in Figure 5-4, is ready to be sent, it goes through the following steps:

1. Generates a random backoff number between 0 and a minimum Contention Window (CWmin).
2. Waits until the channel is free for a DIFS interval.
3. If the channel is still free, begins to decrement the random backoff number, for every slot time (20 µs) the channel remains free.
4. If the channel becomes busy, such as another station getting to 0 before your station, the decrement stops and steps 2 through 4 are repeated.
5. If the channel remains free until the random backoff number reaches 0, the frame can be sent.

Figure 5-4 shows a simplified example of how the DCF process works. In this simplified DCF process, no acknowledgements are shown and no fragmentation occurs.

Figure 5-4 Distributed Coordination Function Example



The DCF steps illustrated in Figure 5-4 are as follows:

1. Station A successfully sends a frame, and three other stations also want to send frames but must defer to Station A traffic.
2. After Station A completes the transmission, all the stations must still defer for the DIFS. When the DIFS is complete, stations waiting to send a frame can begin to decrement the backoff counter, once every slot time, and can send their frame.
3. The backoff counter of Station B reaches zero before Stations C and D, and therefore Station B begins transmitting its frame.
4. When Station C and D detect that Station B is transmitting, they must stop decrementing the backoff counters and defer until the frame is transmitted and a DIFS has passed.
5. During the time that Station B is transmitting a frame, Station E gets a frame to transmit, but because Station B is sending a frame, it must defer in the same manner as Stations C and D.
6. When Station B completes transmission and the DIFS has passed, stations with frames to send begin to decrement the backoff counters. In this case, Station D's backoff counter reaches zero first and it begins transmission of its frame.
7. The process continues as traffic arrives on different stations.

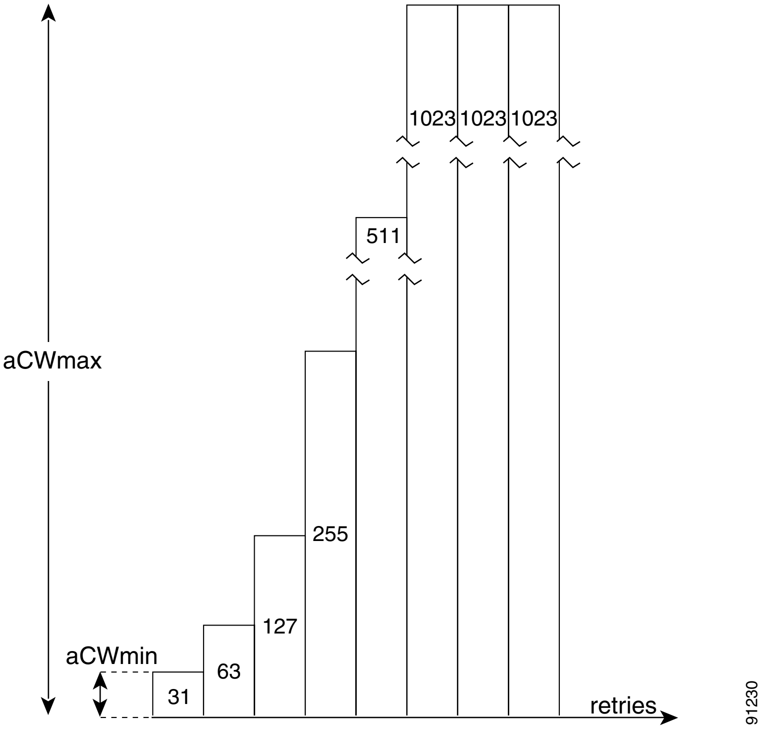
CWmin, CWmax, and Retries

DCF uses a contention window (CW) to control the size of the random backoff. The contention window is defined by two parameters:

- aCWmin
- aCWmax

The random number used in the random backoff is initially a number between 0 and aCWmin. If the initial random backoff expires without successfully sending the frame, the station or AP increments the retry counter, and doubles the value random backoff window size. This doubling in size continues until the size equals aCWmax. The retries continue until the maximum retries or time to live (TTL) is reached. This process of doubling the backoff window is often referred to as a *binary exponential backoff*, and is illustrated in Figure 5-5.

Figure 5-5 Growth in Random Backoff Range with Retries



Wi-Fi Multimedia

This section describes three Wi-Fi Multimedia (WMM) implementations:

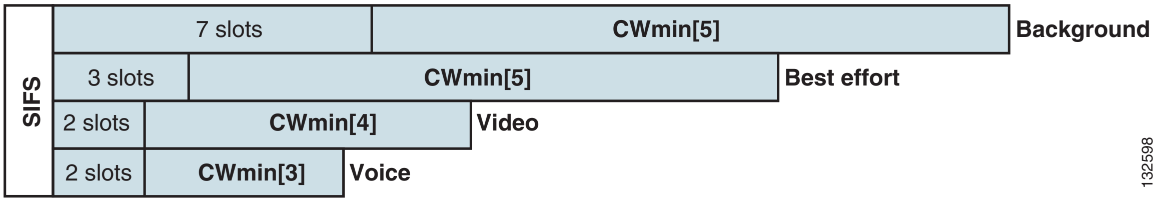
- WMM access
- WMM power save
- WMM access control

WMM Access

WMM is a Wi-Fi Alliance certification of support for a set of features from the 802.11e draft. This certification is for both clients and APs, and certifies the operation of WMM. WMM is primarily the implementation of the EDCA component of 802.11e. Additional Wi-Fi certifications are planned to address other components of the 802.11e.

Figure 5-6 shows the principle behind EDCF, where different interframe spacing and CWmin and CwMax values are applied per traffic classification. Different traffic types can wait different interface spaces before counting down their random backoff, and the CW value used to generate the random backoff number also depends on the traffic classification. High priority traffic has small interframe space and a small CWmin value, giving as short random backoff, whereas best-effort traffic has a longer interframe space and large CWmin value that on average gives a large random backoff number.

Figure 5-6 Access Category Timing



WMM Classification

WMM uses the 802.1p classification scheme developed by the IEEE (which is now a part of the 802.1D specification).

This classification scheme has eight priorities, which WMM maps to four access categories: AC_BK, AC_BE, AC_VI, and AC_VO. These access categories map to the four queues required by a WMM device, as shown in Table 5-2.

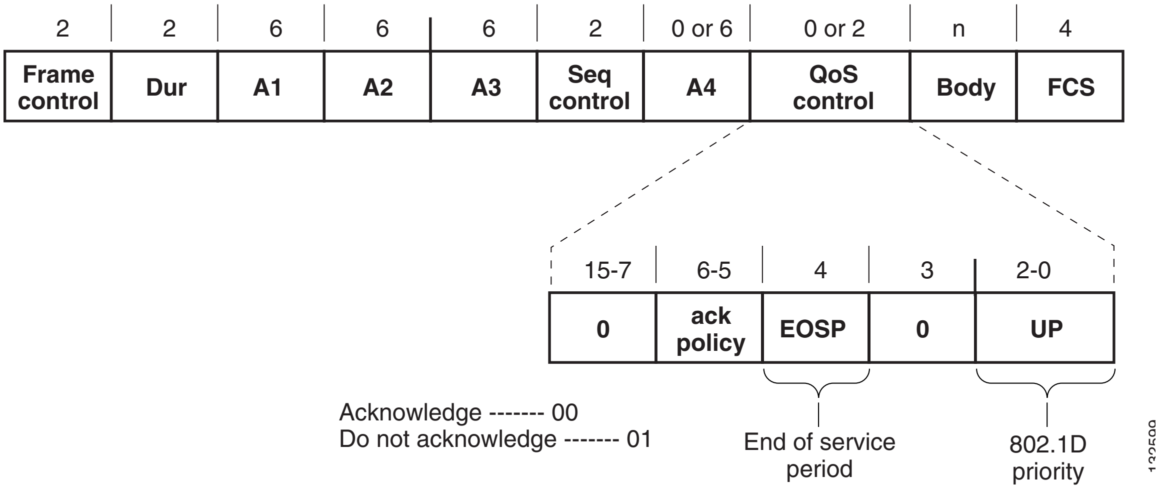
Table 5-2 802.1p and WMM Classification

Priority	802.1 Priority (=User Priority)	802.1p Designation	Access Category	WMM Designation
Lowest	1	BK Background	AC_BK	Background
	2	-Spare		
	0	BE Best-effort		
	3	EE Excellent Effort	AC_BE	Best-effort
	4	CL Control Load		
	5	VI Video <100ms	AC_VI	Video
	6	VO Voice <10ms	AC_VO	Voice
Highest	7	NC Network Control “must get there”		

Figure 5-7 shows the WMM data frame format. Note that even though WMM maps the eight 802.1p classifications to four access categories, the 802.11D classification is sent in the frame.

Note The WMM and IEEE 802.11e classifications are different from the classifications recommended and used in the Cisco network, which are based on IETF recommendations. The primary difference in classification is the demoting of voice and video traffic to 5 and 4, respectively. This allows the 6 classification to be used for Layer 3 network control. To be compliant with both standards, the Cisco Unified Wireless solution performs a conversion between the various classification standards when the traffic crosses the wireless-wired boundary.

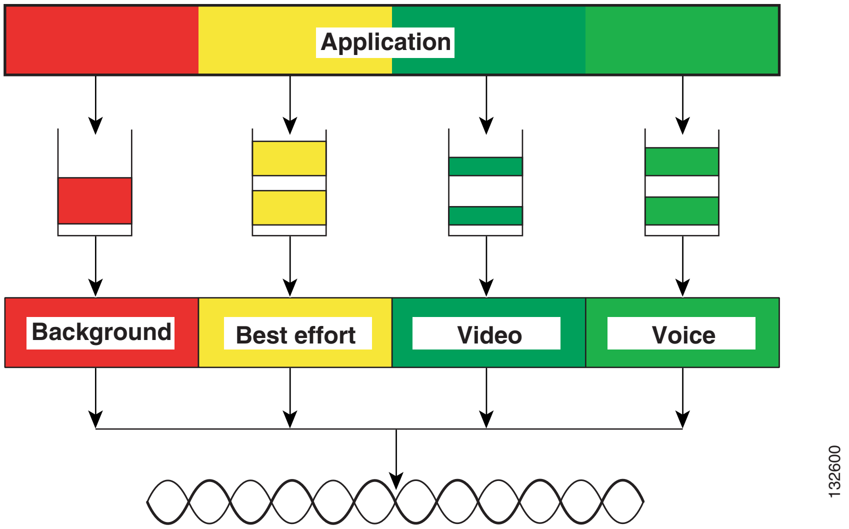
Figure 5-7 WMM Frame Format



WMM Queues

Figure 5-8 shows the queuing performed on a WMM client or AP. There are four separate queues, one for each of the access categories. Each of these queues contends for the wireless channel in a similar manner to the DCF mechanism described previously, with each of the queues using different interframe space, CWmin, and CWmax values. If more than one frame from different access categories collide internally, the frame with the higher priority is sent, and the lower priority frame adjusts its backoff parameters as though it had collided with a frame external to the queuing mechanism. This system is called enhanced distributed channel access (EDCA).

Figure 5-8 WMM Queues



EDCA

The EDCA process is illustrated in Figure 5-9, using data from Figure 5-10, and follows this sequence:

1. While Station X is transmitting its frame, three other stations determine that they must send a frame. Each station defers because a frame was already being transmitted, and each station generates a random backoff.
2. Because station Voice has a traffic classification of voice, it has an arbitrated interframe space (AIFS) of 2, and uses an initial CWmin of 3, and therefore must differ the countdown of its random backoff for 2 slot times, and has a short random backoff value.
3. Best-effort has an AIFS of 3 and a longer random backoff time, because its CWmin value is 5.
4. Voice has the shortest random backoff time, and therefore starts transmitting first. When Voice starts transmitting, all other stations defer.
5. After Voice Station finishes transmitting, all stations wait their AIFS, then begin to decrement the random backoff counters again.
6. Best-effort then completes decrementing its random backoff counter and begins transmission. All other stations defer. This can happen even though there might be a voice station waiting to transmit. This shows that best-effort traffic is not starved by voice traffic because the random backoff decrementing process eventually brings the best-effort backoff down to similar sizes as high priority traffic, and that the random process might, on occasion, generate a small random backoff number for best-effort traffic.
7. The process continues as other traffic enters the system. The access category settings shown in Table 5-3 and Table 5-4 are, by default, the same for an 802.11a radio, and are based on formulas defined in WMM.

Note Table 5-3 refers to the parameter settings on a client, which are slightly different from the settings for an AP. This is because an AP is expected to have multiple clients, and therefore needs to send frames more often.

Figure 5-9 EDCA Example

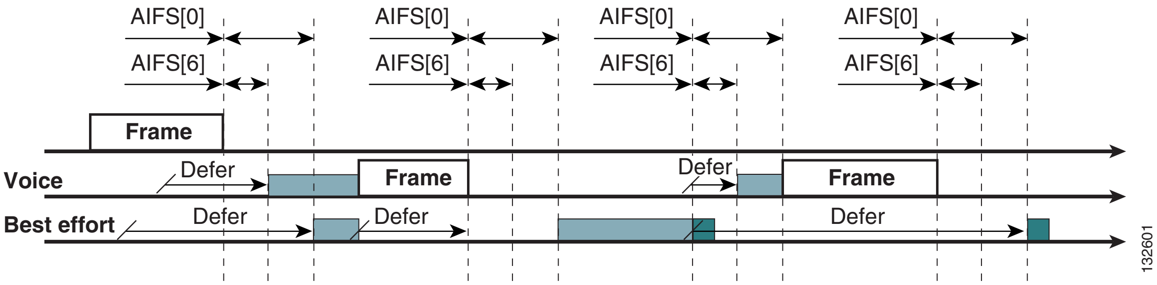


Table 5-3 WMM Client Parameters

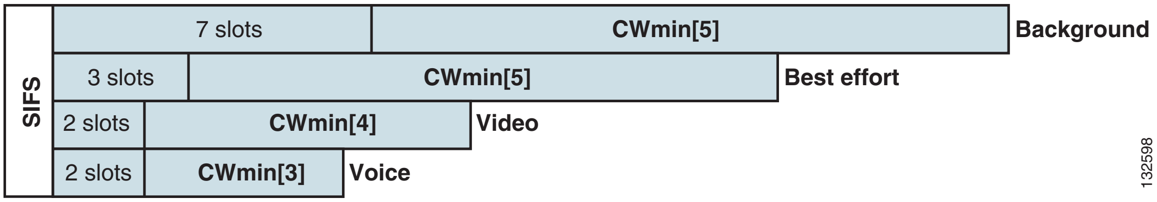
AC	CWmin	CWmax	AIFSN	TXOP Limit (802.11b)	TXOP Limit (802.11a/g)
AC_BK	aCWmin	aCWmax	7	0	0
AC_BE	aCWmin	4*(aCQmin+1)-1	3	0	0
AC_VI	(aCWmin+1)/2-1	aCWmin	1	6.016ms	3.008ms
AC_VO	(aCWmin+1)/4-1	(aCWmin+1)/2-1	1	3.264ms	1.504ms

Table 5-4 WMM AP Parameters

Access Category	CWmin	CWmax	AIFSN	TXOP Limit (802.11b)	TXOP Limit (802.11a/g)
AC_BK	aCWmin	aCWmax	7	0	0
AC_BE	aCWmin	4*(aCQmin+1)-1	3	0	0
AC_VI	(aCWmin+1)/2-1	aCWmin	2	6.016ms	3.008ms
AC_VO	(aCWmin+1)/4-1	(aCWmin+1)/2-1	2	3.264ms	1.504ms

The overall impact of the different AIFS, CWmin, and CWmax values is difficult to illustrate in timing diagrams because their impact is more statistical in nature. It is easier to compare the AIFS and the size of the random backoff windows, as shown in Figure 5-10.

Figure 5-10 AIFS and CWmin for Different Access Categories



When comparing voice and background frames as examples, these traffic categories have CWmin values of 3 (7) and 5 (31), and AIFS of 2 and 7, respectively. This an average delay of 5 slot times before sending a voice frame, and an average of 22 slot times for background frame. Therefore, voice frames are statistically much more likely to be sent before background frames.

U-APSD

Unscheduled automatic power-save delivery (U-APSD) is a feature that has two key benefits:

- The primary benefit is the saving of WLAN client power, by allowing the transmission of frames from the WLAN client to trigger the forwarding of data frames for a client that has been buffered at the AP for power saving purposes.

  The client remains listening to the AP until it receives a frame from the AP with an end of service period (EOSP) bit set. This tells the client that it can now go back into its power save mode. This triggering mechanism is considered a more efficient use of client power than the regular listening for beacons method, at a period controlled by the delivery traffic indication message (DTIM) interval, as the latency and jitter requirements of voice are such that a WVoIP client would either not save power during a call, resulting in reduced talk times, or use a short DTIM interval, resulting in reduced standby times. The use of U-APSD allows the use of long DTIM intervals to maximize standby time without sacrificing call quality. The U-APSD feature can be applied across access categories; U-APSD can be applied to the voice ACs in the AP, but the other ACs can still use the standard power save feature.

- The secondary benefit of this feature is increased call capacity. The coupling of transmission buffered data frames from the AP with the triggering data frame from the WLAN client allows the frames from the AP to be sent without the accompanying interframe spacing and random backoff, thereby reducing the contention experience by call.

Figure 5-11 shows an example frame exchange for the standard 802.11 power save delivery process. The client in power save mode first detects that there is data waiting for it at the AP via the presence of the TIM in the AP beacon. The client must power-save poll (PS-Poll) the AP to retrieve that data. If the data sent to the client requires more than one frame to be sent, the AP indicates this in the sent data frame. This process requires the client to continue sending power save polls to the AP until all the buffered data is retrieved by the client.

This presents two major problems. The first is that it is quite inefficient, requiring the PS-polls, as well as the normal data exchange, to go through the standard access delays associated with DCF. The second issue, being more critical to voice traffic, is that retrieving the buffered data is dependent on the DTIM, which is a multiple of the beacon interval. Standard beacon intervals are 100mS and the DTIM interval can be integer multiples of this. This introduces a level of jitter that is generally unacceptable for voice calls, and voice handsets switch from power save mode to full transmit and receive operation when a voice call is in progress. This gives acceptable voice quality but reduces battery life.

Figure 5-11 Standard Client Power Save

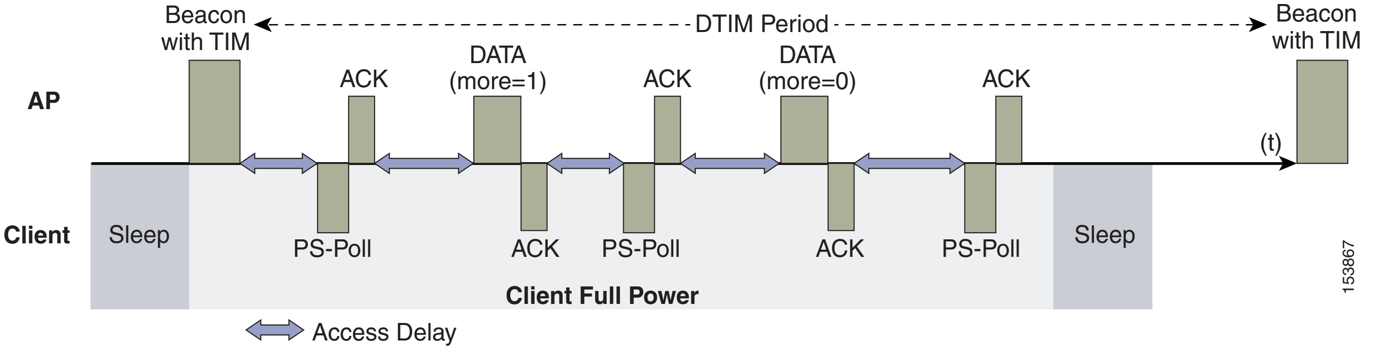
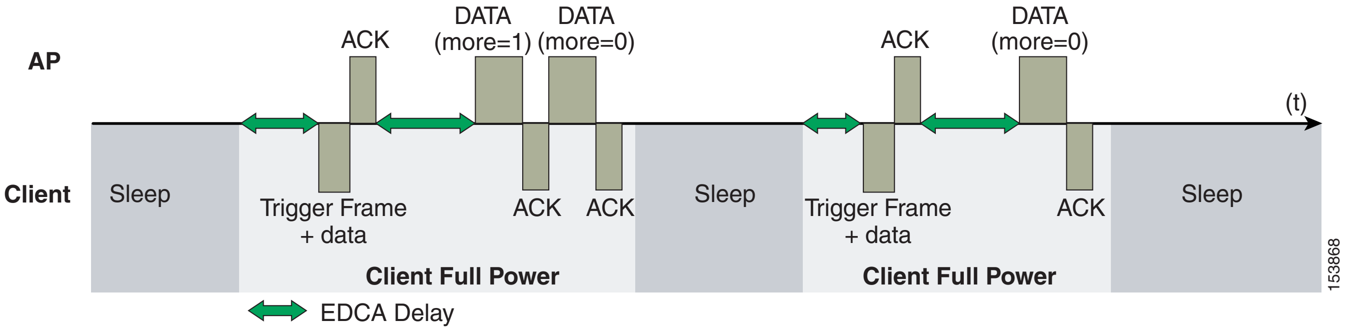


Figure 5-12 shows an example of traffic flows with U-APSD. In this case, the trigger for retrieving traffic is the client sending traffic to the AP. The AP, when acknowledging the frame, tells the client that data is queued for it, and that is should stay on. The AP then sends data to the client typically as a TXOP burst where only the first frame has the EDCF access delay. All subsequent frames are then sent directly after the acknowledgment frame.

This approach overcomes both of the disadvantages of the previous scheme in that it is much more efficient. The timing of the polling is controlled via the client traffic, which in the case of voice is symmetric, so if the client is sending a frame every 20mSec, it would be expecting to receive a frame every 20mSec as well. This would introduce a maximum jitter of 20mSec, rather than an n * 100mSec jitter.

Figure 5-12 U-APSD



TSpec Admission Control

Traffic Specification (TSpec) allows an 802.11e client to signal to the AP its traffic requirements. In the 802.11e MAC definition, there are two mechanisms to provide prioritized access. These are the contention-based EDCA option and the controlled access option provided by the transmit opportunity (TXOP).

When describing TSpec features where a client can specify its traffic characteristics, it is easy to assume that this would automatically result in the use of the controlled access mechanism, and have the client granted a specific TXOP to match the TSpec request.

This does not have to be the case; a TSpec request can be used to control the use of the various access categories (ACs) in EDCA. Before a client can send traffic of a certain priority type, it must have requested to do so via the TSpec mechanism. For example, a WLAN client device wanting to use the voice AC must first make a request for use of that AC. Whether AC use is controlled by TSpec requests or is openly configurable can be controlled by TSpec requests, but best-effort and background ACs can be open for use without a TSpec request.

The use of EDCA ACs, rather than the HCCA, to meet TSpec requests is possible in many cases because the traffic parameters are sufficiently simple to allow them to be met by allocating capacity, rather than creating a specific TXOP to meet the application requirements.

The TSpec admission control to an AC acts in a manner similar to Cisco CallManager in that it knows how much capacity is available at a branch (AP), and does not allow additional calls when that capacity is consumed. This is done without the CallManager having to specifically allocate resources in the path of the VoIP call.

Note The Cisco 7920 WVoIP handset does not support TSpec admission control.

Add Traffic Stream

The Add Traffic Stream (ADDTS) function is how a WLAN client performs an admission request to an AP. Signalling its TSpec to the AP.

An admission request is in one of two forms:

- ADDTS action frame—This happens when a phone call is originated or terminated by a client associated to the AP. The ADDTS contains TSpec and might contain a traffic stream rate set (TSRS) IE (CCXv4 clients).
- Association and re-association message—The association message might contain one or more TSpecs and one TSRS IE if the STA wants to establish traffic stream as part of the association. The re-association message might contain one or more TSpecs and one TSRS IE if a STA roams to another AP.

The ADDTS contains the TSpec element that describes the traffic request (see Table 5-5). Apart from key data describing the traffic requirements, such as data rates and frame sizes, the TSpec element also tells the AP the minimum physical rate that the client device will use. This allows the calculation of how much time that station can potentially consume in sending and receiving in this TSpec, and therefore allowing the AP to calculate whether it has the resources to meet the TSpec.

TSpec admission control is used by the WLAN client (target clients are VoIP handsets) when a call is initiated and during a roam request. During a roam, the TSpec request is appended to the re-association request.

Table 5-5 WMM TSpec Element Field

Field	Value
Element ID	221
Length	6+55=61
OUI	00:50:f2(hex)
OUI	Type2
OUI	Subtype2
Version	1

Table 5-5 WMM TSpec Element Field (continued)

TS Info	• Traffic stream ID, which combined with the addressing of the frame containing the TSpec element, uniquely identifies the Traffic for which a request is being made. • 802.1D priority information, and is the same value used in QoS data frames associated with this traffic stream. • Traffic is for upstream, downstream, or bi-directional traffic. • Power save is traditional or U-APSD.
Nominal MSDU Size	Size of MSDU, if fixed Nominal Size if variable
Maximum MSDU Size	-
Minimum Service Interval	-
Maximum Service Interval	-
Inactivity Interval	-
Suspension Interval	-
Service Start	Time -
Minimum Data	Rate -
Mean Data Rate	Average data rate, in units of bits per second; does not include overhead.
Peak Data Rate	-
Maximum Burst Size	-
Delay Bound	-
Minimum PHY Rate	The minimum 802.11 PHY rate that is used.
Surplus Bandwidth Allowance	-
Medium Time	-

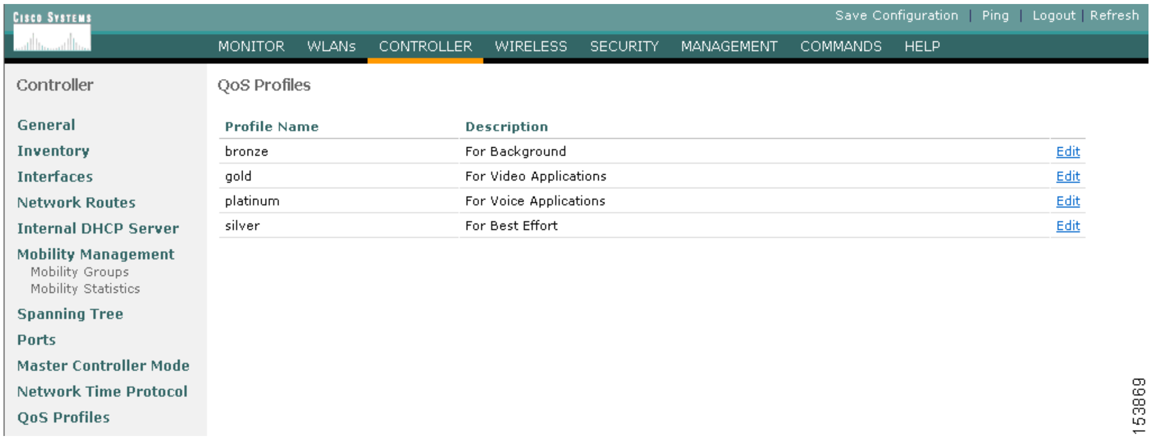
Sample TSpec Decode

```
Vendor Specific: WME            Tag Number: 221 (Vendor Specific)
Tag length: 61
Tag interpretation: WME TSPEC: type 2, subtype 2, version 1
Tag interpretation: WME TS Info: Priority 4 (Controlled Load) (Video), Contention-based
access set, Bi-directional
Tag interpretation: WME TSPEC: Fixed MSDU Size 5632
Tag interpretation: WME TSPEC: Maximum MSDU Size 56448
Tag interpretation: WME TSPEC: Minimum Service Interval 5
Tag interpretation: WME TSPEC: Maximum Service Interval 0
Tag interpretation: WME TSPEC: Inactivity Interval 0
Tag interpretation: WME TSPEC: Service Start Time 4294967040
Tag interpretation: WME TSPEC: Minimum Data Rate 255
Tag interpretation: WME TSPEC: Mean Data Rate 40960
Tag interpretation: WME TSPEC: Maximum Burst Size 40960
Tag interpretation: WME TSPEC: Minimum PHY Rate 40960
Tag interpretation: WME TSPEC: Peak Data Rate 0
Tag interpretation: WME TSPEC: Delay Bound 0
Tag interpretation: WME TSPEC: Medium Time 23437
```

QoS Advanced Features for WLAN Infrastructure

Cisco Centralized WLAN Architecture has multiple QoS features, in addition to WMM support. Primary among these is the QoS profiles in the WLC. Four QoS profiles can be configured: platinum, gold, silver, and bronze, as shown in Figure 5-13.

Figure 5-13 WLC QoS Profiles

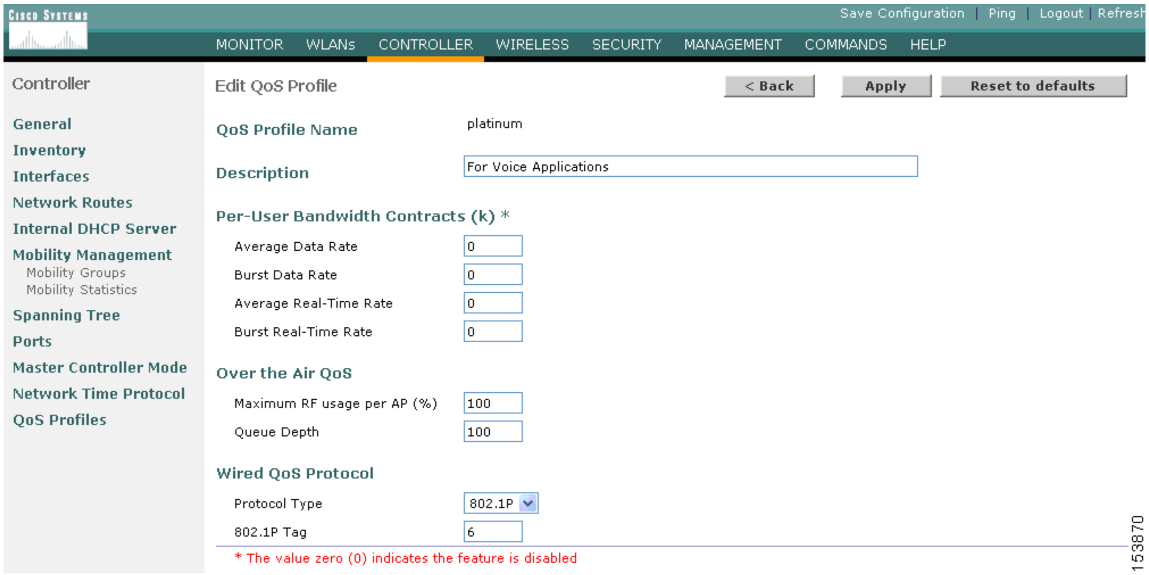


QoS Profiles

Profile Name	Description	
bronze	For Background	Edit
gold	For Video Applications	Edit
platinum	For Voice Applications	Edit
silver	For Best Effort	Edit

Each of these profiles (see Figure 5-14) allows the configuration of bandwidth contracts, RF usage control, and the maximum 802.1p classification allowed. It is generally recommended that these settings be left at their default values, and that the 802.11 WMM features be used to provide differentiated services.

Figure 5-14 QoS Profile Options



Edit QoS Profile

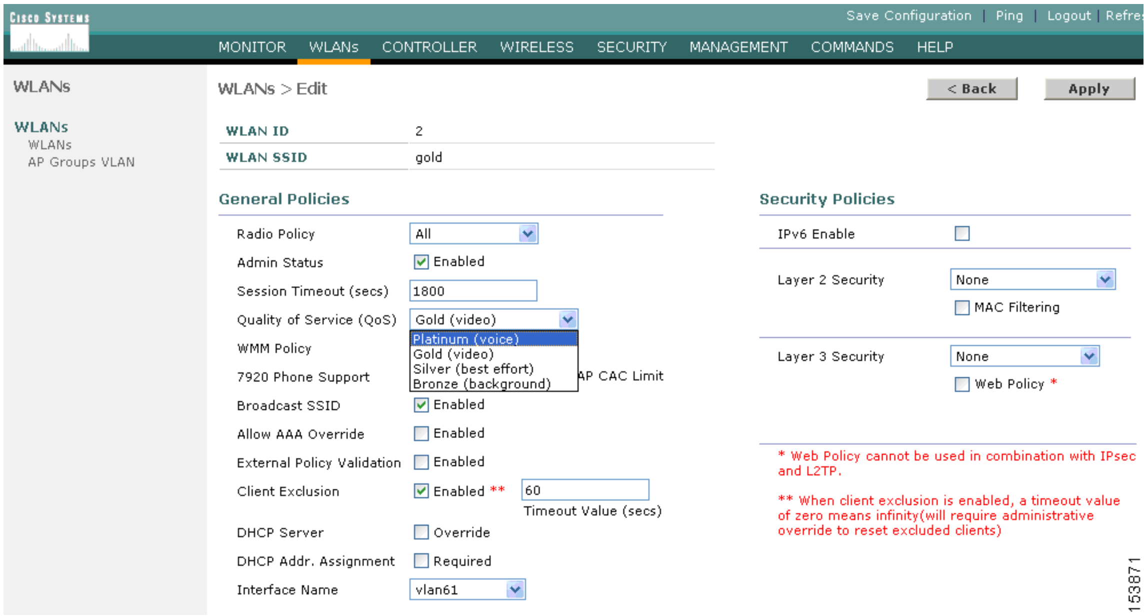
QoS Profile Name	platinum
Description	For Voice Applications
Per-User Bandwidth Contracts (k) *	
Average Data Rate	0
Burst Data Rate	0
Average Real-Time Rate	0
Burst Real-Time Rate	0
Over the Air QoS	
Maximum RF usage per AP (%)	100
Queue Depth	100
Wired QoS Protocol	
Protocol Type	802.1P
802.1P Tag	6

* The value zero (0) indicates the feature is disabled

The WLAN can be configured with different default QoS profiles, as shown in Figure 5-15. Each of the profiles (platinum, gold, silver, and bronze) are annotated with their typical use. In addition, clients can be assigned a QoS profile based on their identity, through AAA.

For a typical enterprise, WLAN deployment parameters, such as per-user bandwidth contracts and over the air QoS should be left at their default values, and standard QoS tools, such as WMM and wired QoS, should be used to provide optimum QoS to clients.

Figure 5-15 WLAN QoS Profile Settings



WLANs > Edit

WLAN ID	2
WLAN SSID	gold

General Policies

Radio Policy	All
Admin Status	☑ Enabled
Session Timeout (secs)	1800
Quality of Service (QoS)	Gold (video) Platinum (voice) Gold (video) Silver (best effort) Bronze (background)
WMM Policy	
7920 Phone Support	AP CAC Limit
Broadcast SSID	☑ Enabled
Allow AAA Override	□ Enabled
External Policy Validation	□ Enabled
Client Exclusion	☑ Enabled ** 60 Timeout Value (secs)
DHCP Server	□ Override
DHCP Addr. Assignment	□ Required
Interface Name	vlan61

Security Policies

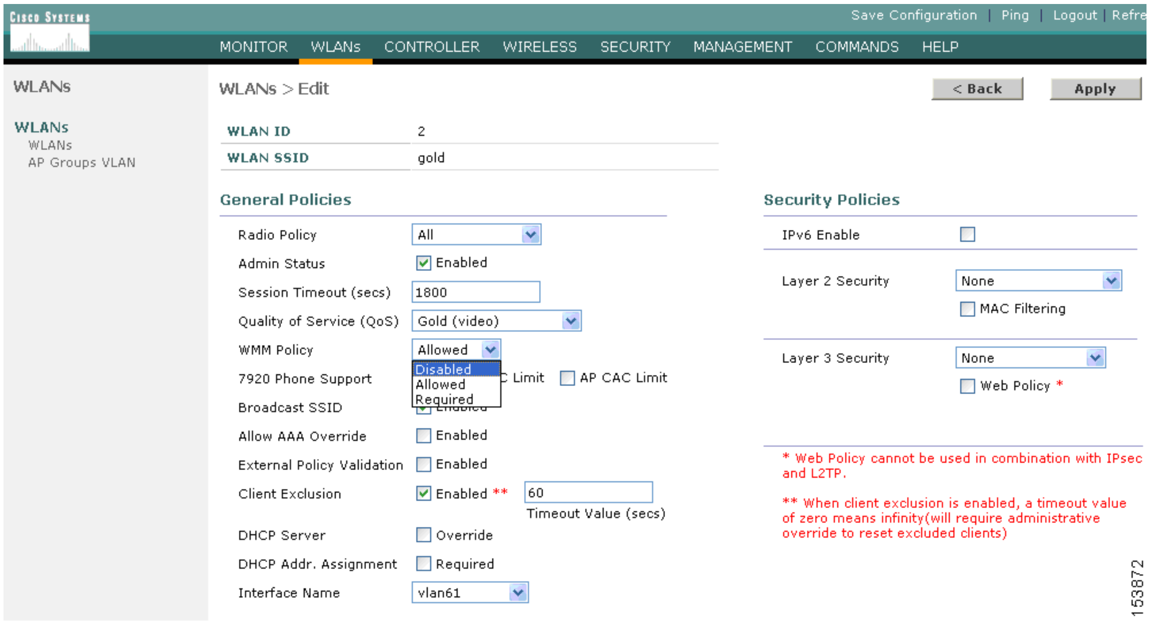
IPv6 Enable	□
Layer 2 Security	None □ MAC Filtering
Layer 3 Security	None □ Web Policy *

* Web Policy cannot be used in combination with IPsec and L2TP.

** When client exclusion is enabled, a timeout value of zero means infinity(will require administrative override to reset excluded clients)

In addition to the QoS profiles, the WMM policy per WLAN can also be controlled, as shown in Figure 5-16. The three WMM options are disabled, allowed, or required. Disabled means that the WLAN does not advertise WMM capabilities, or allow WMM negotiations, Allowed means that the WLAN does allow WMM and non-WMM clients, and required means that only WMM-enabled clients can be associated with this WLAN.

Figure 5-16 WLAN WMM Policy


Cisco Systems
Save Configuration | Ping | Logout | Refre
MONITOR WLANs CONTROLLER WIRELESS SECURITY MANAGEMENT COMMANDS HELP
WLANs
WLANs
WLANs
AP Groups VLAN
< Back
Apply
153872

WLANs > Edit

WLAN ID	2
WLAN SSID	gold

General Policies

Radio Policy	All
Admin Status	☑ Enabled
Session Timeout (secs)	1800
Quality of Service (QoS)	Gold (video)
WMM Policy	Allowed Disabled Allowed Required
7920 Phone Support	… C Limit ☐ AP CAC Limit
Broadcast SSID	☑ Enabled
Allow AAA Override	☐ Enabled
External Policy Validation	☐ Enabled
Client Exclusion	☑ Enabled ** 60 Timeout Value (secs)
DHCP Server	☐ Override
DHCP Addr. Assignment	☐ Required
Interface Name	vlan61

Security Policies

IPv6 Enable	☐
Layer 2 Security	None ☐ MAC Filtering
Layer 3 Security	None ☐ Web Policy *

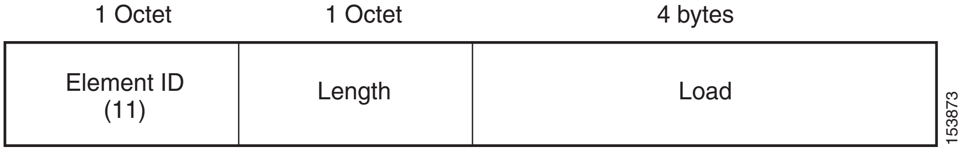
* Web Policy cannot be used in combination with IPsec and L2TP.

** When client exclusion is enabled, a timeout value of zero means infinity(will require administrative override to reset excluded clients)

IP Phones

Figure 5-17 shows the basic QBSS information element (IE) advertised by a Cisco AP. The Load field indicates the portion of available bandwidth currently used to transport data on that AP.

Figure 5-17 QBSS Information Element



1 Octet	1 Octet	4 bytes
Element ID (11)	Length	Load

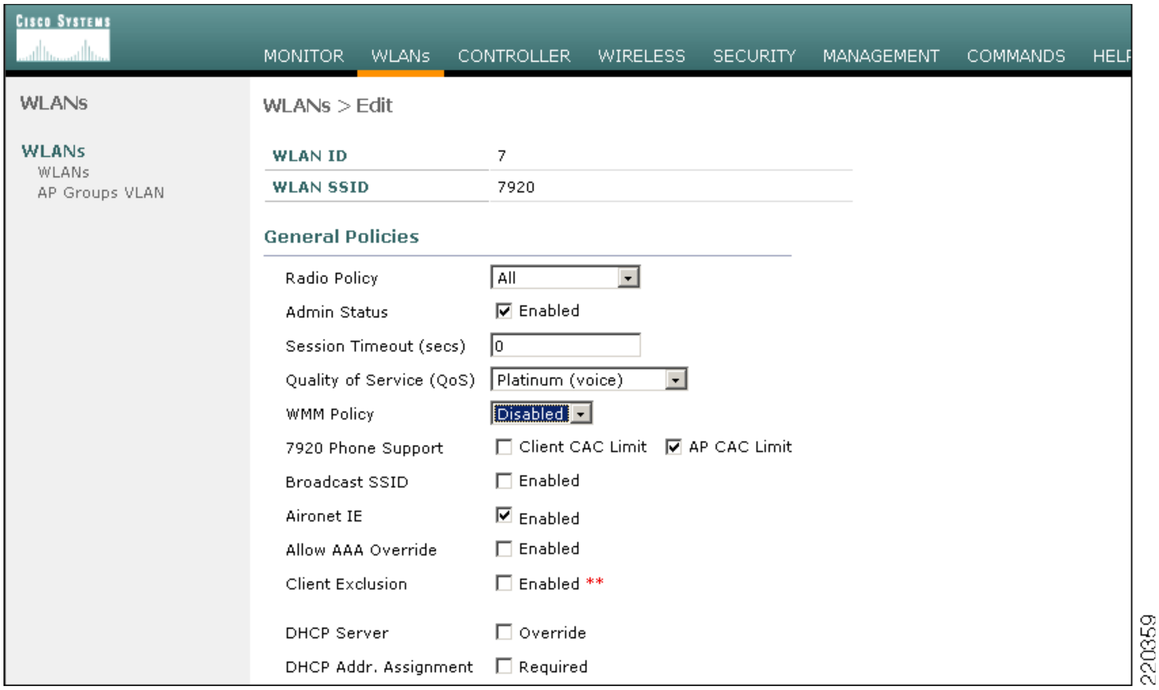
There are actually three QBSS IEs that need to be supported in certain situations:

- Old QBSS (Draft 6 (pre-standard))
- New QBSS (Draft 13 802.11e (standard))
- New distributed CAC load IE (a Cisco IE)

This QBSS depends on the WMM and 7920 settings on the WLAN.

7920 phone support, shown in Figure 5-18, is a component of the WLC WLAN configuration that enables the AP to include the appropriate QBSS element in its beacons. WLAN clients with QoS requirements use these advertised QoS parameters to determine the best AP with which to associate.

Figure 5-18 7920 Phone Support



WLANs > Edit

WLAN ID	7
WLAN SSID	7920

General Policies

Radio Policy	All
Admin Status	☑ Enabled
Session Timeout (secs)	0
Quality of Service (QoS)	Platinum (voice)
WMM Policy	Disabled
7920 Phone Support	□ Client CAC Limit ☑ AP CAC Limit
Broadcast SSID	□ Enabled
Aironet IE	☑ Enabled
Allow AAA Override	□ Enabled
Client Exclusion	□ Enabled **
DHCP Server	□ Override
DHCP Addr. Assignment	□ Required

The WLC provides 7920 support through the client CAC limit, or AP CAC limit. These features provide the following:

- Client CAC limit—The 7920 uses a call admission control setting that is set on the client. This supports legacy 7920 code -pre 2.01.
- AP CAC limit—The 7920 uses call admission control settings learned from WLAN advertisement.

The various combinations of WMM, client CAC limit, and AP CAC limit result in different QBSS IEs being sent:

- If WMM only is enabled, IE number 2 (802.11e standard) QBSS Load IE is sent out in the beacons and probe responses.
- If 7920 client CAC limit is to be supported, IE number 1 (the pre-standard QBSS IE) is sent out in the beacons and probe responses on the border gateway (bg) radios.
- If 7920 AP CAC limit is to be supported, the number 3 QBSS IE is sent in the beacons and probe responses for border gateway (bg) radios.

Note The various QBSS IEs use the same ID, and therefore the three QBSSs are mutually exclusive. For example, the beacons and probe responses can contain only one QBSS IE.

Setting the Admission Control Parameters

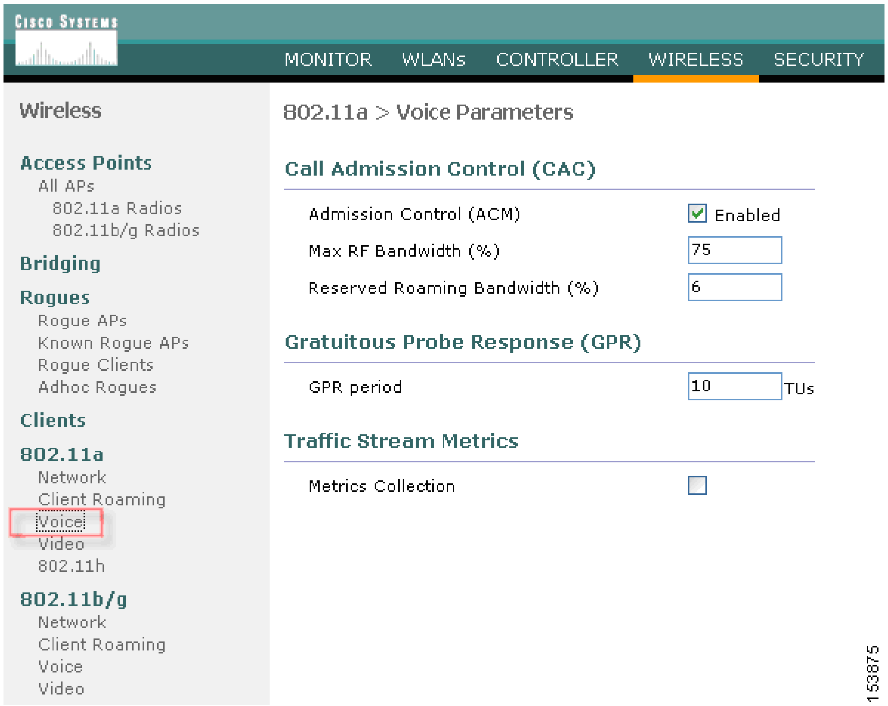
Figure 5-19 shows an example configuration screen for setting the voice parameters on the controller. The admission control parameters consist of the maximum RF capacity that a radio can have and still accept the initiation of a VoIP call through a normal ADDTS request. The reserved roaming bandwidth is how much capacity has been set aside to be able to respond to ADDTS requests during association or re-association, which are WVoIP clients with calls in progress that are trying to roam to that AP.

When checked, the Gratuitous Probe Response checkbox causes APs to send a probe response at a regular interval (default 10mSec) to assist certain WVoIP phones in making roaming decisions. The use of the probe response is considered more efficient and less disruptive than increasing the beacon rate of APs.

The Traffic Stream Metrics Collection option determines if data is collected on voice or video calls for use by the WCS.

Note Call admission control is performed only for voice and video QoS profiles.

Figure 5-19 Configuring Voice Parameters



802.11a > Voice Parameters

Call Admission Control (CAC)

Admission Control (ACM) ☑ Enabled

Max RF Bandwidth (%) 75

Reserved Roaming Bandwidth (%) 6

Gratuitous Probe Response (GPR)

GPR period 10 TUs

Traffic Stream Metrics

Metrics Collection □

Impact of TSpec Admission Control

The purpose of TSpec admission control is not to deny clients access to the WLAN; it is to protect the high priority resources. Therefore, a client that has not used TSpec admission control does not have its traffic blocked; it simply has its traffic re-classified if it tries to send (which it should not do if the client is transmitting WMM compliant-traffic in a protected AC).

Table 5-6 and Table 5-7 describe the impact on classification if Access Control is enabled and dependent on whether a traffic stream has been established.

Table 5-6 Upstream Traffic

	Traffic Stream established	No Traffic Stream
No admission control	No change in behavior; the packets go into the network as they do today – UP is limited to max = WLAN QoS setting.	No change in behavior; the packets go into the network as they do today – UP is limited to max = WLAN QoS setting.
Admission control	No change in behavior; the packets go into the network as they do today – UP is limited to max = WLAN QoS setting.	Packets are remarked to BE (both CoS and DSCP) before they enter the network for WMM clients. For non-WMM clients, packets are sent with WLAN QoS.

Table 5-7 Downstream Traffic

	Traffic Stream established	No Traffic Stream
No admission control	No change	No change
Admission control	No change	Remark UP to BE for WMM client. For non-WMM clients, use WLAN QoS.

802.11e, 802.1p, and DSCP Mapping

WLAN data in a centralized WLAN deployment is tunneled via LWAPP. To maintain the QoS classification that has been applied to data traffic, a process transferring or applying classification is required.

For example, when WMM classified traffic is sent by a WLAN client, it has an 802.1p classification in its frame. The AP needs to translate this classification into a DSCP value for the LWAPP packet carrying the frame to ensure that the packet is treated with the appropriate priority on its way to the WLC. A similar process needs to occur on the WLC for LWAPP packets going to the AP.

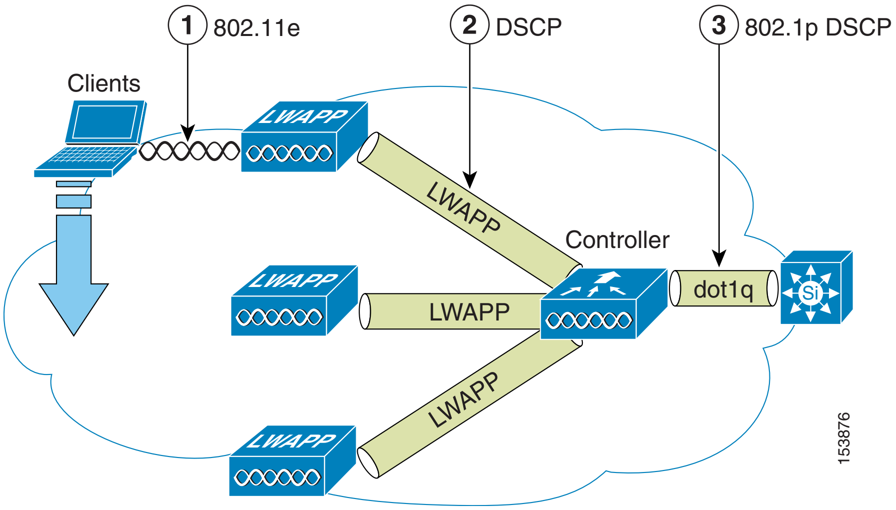
A mechanism to classify traffic from non-WMM clients is also required, so that their LWAPP packets can also be given an appropriate DSCP classification by the AP and the WLC.

Figure 5-20 shows the various classification mechanisms in the LWAPP WLAN network. The multiple classification mechanisms and client capabilities require multiple strategies:

- LWAPP control frames require prioritization, and LWAPP control frames are marked with a DSCP classification of CS6.
- WMM-enabled clients have the classification of their frames mapped to a corresponding DSCP classification for LWAPP packets to the WLC. This mapping follows the standard IEEE CoS to DSCP mapping, with the exception of the changes necessary for AVVID compliance. This DSCP value is translated at the WLC to a CoS value on the wired interfaces.

- Non-WMM clients have the DSCP of their LWAPP tunnel set to match the default QoS profile for that WLAN. For example, the QoS profile for a WLAN supporting 7920 phones would be set to platinum, resulting in a DSCP classification of EF for data frames packets from that AP WLAN.
- LWAPP data packets from the WLC have a DSCP classification that is determined by the DSCP of the wired data packets sent to the WLC. The WMM classification used when sending frames from the AP to a WMM client is determined by the AP table converting DSCP to WMM classifications.

Figure 5-20 WMM, DSCP, and 802.1p Relationship



AVVID Priority Mapping

The LWAPP AP and WLC perform AVVID conversion, so that WMM values as shown in Table 5-8 are mapped to the appropriate AVVID DSCP values, rather than the IEEE values.

Table 5-8 Access Point QoS Translation Values

AVVID 802.1p UP-Based Traffic Type	AVVID IP DSCP	AVVID 802.1p UP	IEEE 802.11e UP
Network control	-	7	-
Inter-network control (LWAPP control, 802.11 management)	48	6	7
Voice	46 (EF)	5	6
Video	34 (AF41)	4	5
Voice control	26 (AF31)	3	4
Background (Gold)	18 (AF21)	2	2
Background (Gold)	20 (AF22)	2	2
Background (Gold)	22 (AF23)	2	2
Background (Silver)	10 (AF11)	1	1
Background (Silver)	12 (AF12)	1	1
Background (Silver)	14 (AF13)	1	1

Table 5-8 Access Point QoS Translation Values (continued)

Best Effort	0 (BE)	0	0, 3
Background	2	0	1
Background	4	0	1
Background	6	0	1

Deploying QoS Features Cisco on LWAPP-based APs

When deploying WLAN QoS on the APs, consider the following:

- In the absence of Layer 2 classification (802.1p) information, the WLC and the APs depend on Layer 3 classification (DSCP) information. This DSCP value is subject to modification by intermediate routers and therefore the Layer 2 classification received by the destination might not reflect the Layer 2 classification marked by the source of the LWAPP traffic.
- The APs no longer use NULL VLAN ID. As a consequence, L2 LWAPP does not effectively support QoS because the AP does not send the 802.1p/Q tags, and in L2 LWAPP there is no outer DSCP to fall back on.
- APs do not classify packets; they prioritize packets based on CoS value, or WLAN.
- APs carry out EDCF like queuing on the radio egress port only.
- APs only do FIFO queueing on the Ethernet egress port.

QoS and the H-REAP

For WLANs that have data traffic forwarded to the WLC, behavior is same as non H-REAP APs.

For locally switched WLANs with WMM traffic, the AP marks the dot1p value in the dot1q VLAN tag for upstream traffic. This occurs only on tagged VLANs; that is, not native VLANs. For downstream traffic for locally switched WLANs, the H-REAP uses the incoming dot1q tag from the Ethernet side and marks the WMM values on the wireless side. The WLAN QoS profile is applied both for upstream and downstream packets; that is, for downstream if an 802.1p value that is higher than the default WLAN value is received, the default WLAN value is used. For upstream, if the client sends an WMM value that is higher than the default WLAN value, the default WLAN value is used.

For non-WMM traffic, there is no CoS marking on the client frames from the AP.

Guidelines for Deploying Wireless QoS

The same rules for deploying QoS in a wired network apply to deploying QoS in a wireless network. The first and most important guideline in QoS deployment is to know your traffic. Know your protocols, your application's sensitivity to delay, and traffic bandwidth. QoS does not create additional bandwidth, it simply gives more control of where the bandwidth is allocated.

Throughput

An important consideration when deploying 802.11 QoS is to understand the offered traffic, not only in terms of bit rate, but also in terms of frame size, because 802.11 throughput is sensitive to the frame size of the offered traffic.

Table 5-9 shows the impact that frame size has on throughput: as packet size decreases, so does throughput. For example, if an application offering traffic at a rate of 3Mbps is deployed on an 11Mbps 802.11b network, but uses an average frame size of 300 bytes, no QoS setting on the AP allows the application to achieve its throughput requirements. This is because 802.11b cannot support the required throughput for that throughput and frame size combination. The same amount of offered traffic, having a frame size of 1500 bytes, does not have this issue.

Table 5-9 Throughput Compared to Frame Size

	300	600	900	1200	1500	Frame Size (Bytes)
11g - 54Mbps	11.4	19.2	24.6	28.4	31.4	Throughput bps
11b - 11Mbps	2.2	3.6	4.7	5.4	6	Throughput bps

Traffic Shaping, Over the Air QoS and WMM Clients

Traffic shaping and over the air QoS are useful tools in the absence of WLAN WMM features, but they do not address the prioritization of 802.11 traffic directly. For WLANs that support WMM clients or 7920 handsets, the WLAN QoS mechanisms of these clients should be relied on; no traffic shaping or over the air QoS should be applied to these WLANs.

WLAN Voice and the Cisco 7920

The Cisco 7920 is a Cisco 802.11b VoIP handset, and its use is one of the most common reasons for deploying QoS on a WLAN.

Deploying voice over WLAN infrastructure involves more than simply providing QoS on WLAN. A voice WLAN needs to consider site survey coverage requirements, user behavior, roaming requirements and admission control. This is covered in the *Cisco Wireless IP Phone 7920 Design and Deployment Guide*, at the following URL: http://www.cisco.com/en/US/docs/voice_ip_comm/cuipph/7920/5_0/english/design/guide/7920ddg.html.