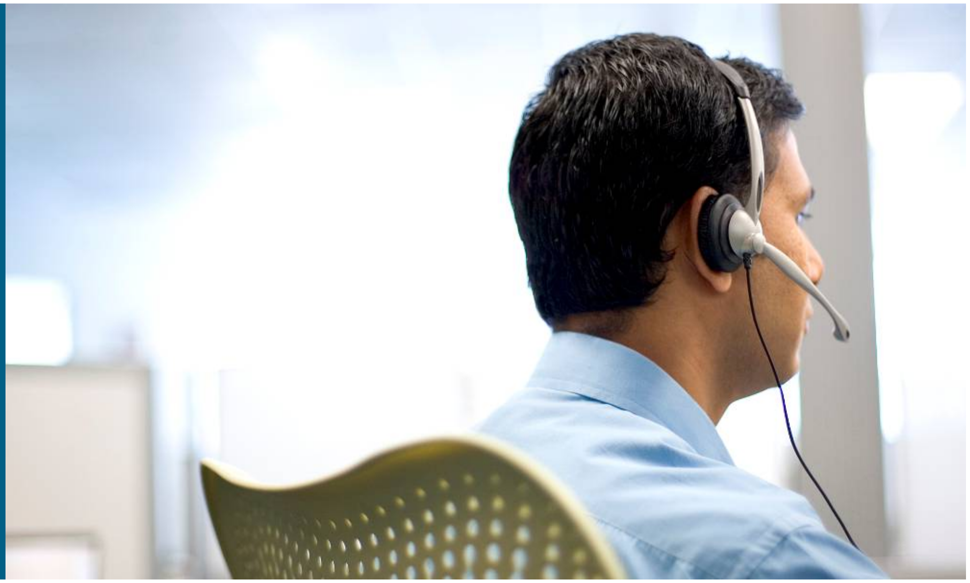
Advanced Qos and Voice over Wireless

BRKCAM-3012



Jesse Lindeman

Cisco Networkers 2007

HOUSEKEEPING

- We value your feedback, don't forget to complete your online session evaluations after each session and complete the Overall Conference Evaluation which will be available online from Friday.
- Visit the World of Solutions on Level -01!
- Please remember this is a 'No Smoking' venue!
- Please switch off your mobile phones!
- Please remember to wear your badge at all times including the Party!
- Do you have a question? Feel free to ask them during the Q&A section or write your question on the Question form given to you and hand it to the Room Monitor when you see them holding up the Q&A sign.

Session Objectives

This session WILL cover:

- Wireless Quality of Service Principles
- WLAN System-wide QoS Requirements
- Where QoS Fits into the Cisco Unified Wireless Network

Anti-Session Objectives

This session WON'T cover:

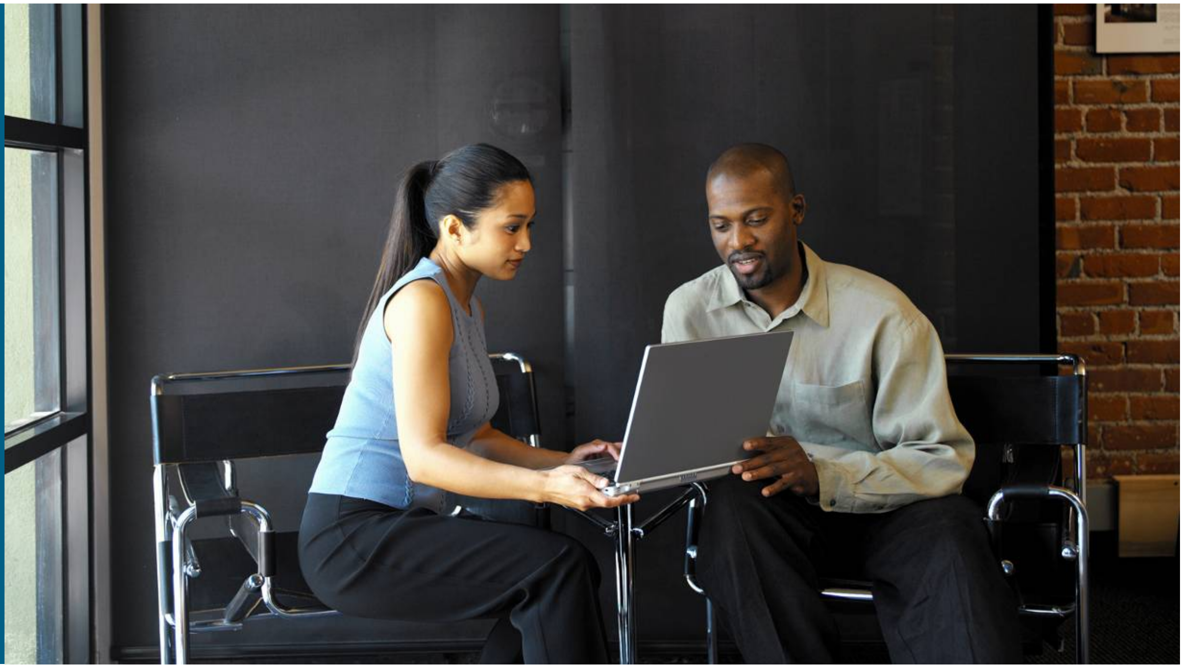
- Basic 802.11/WLAN topics and design
- RF cell planning and site surveying
- In-depth switching and routing QoS principles

Key Takeaways

The Key Takeaways of this presentation are:

- **Understand when, where, and why QoS is necessary**
- **Appreciate the end-to-end requirements of QoS**
- **Understand where VoWLAN fits into the QoS discussion**
- **Apply QoS/VoWLAN knowledge to the CUWN**

Introduction to QoS Principles



First: Why QoS for WLAN?

- **Wireless is fundamentally different from wired**
  - **Far more stringent bandwidth limitations**
    - **Limited spectrum (few non-overlapping channels)**
    - **Half-duplex medium**
    - **Every directed data and management frame is ACK'd**
    - **'Listen Before Talk' contention model**
- **This all makes WLAN highly susceptible to latency and jitter**
- **Can't really 'throw bandwidth' at the problem, either**

QoS Concepts

- **Latency**

  Fixed Delay

  Variable Delay

- **Jitter**

  Delay Variance

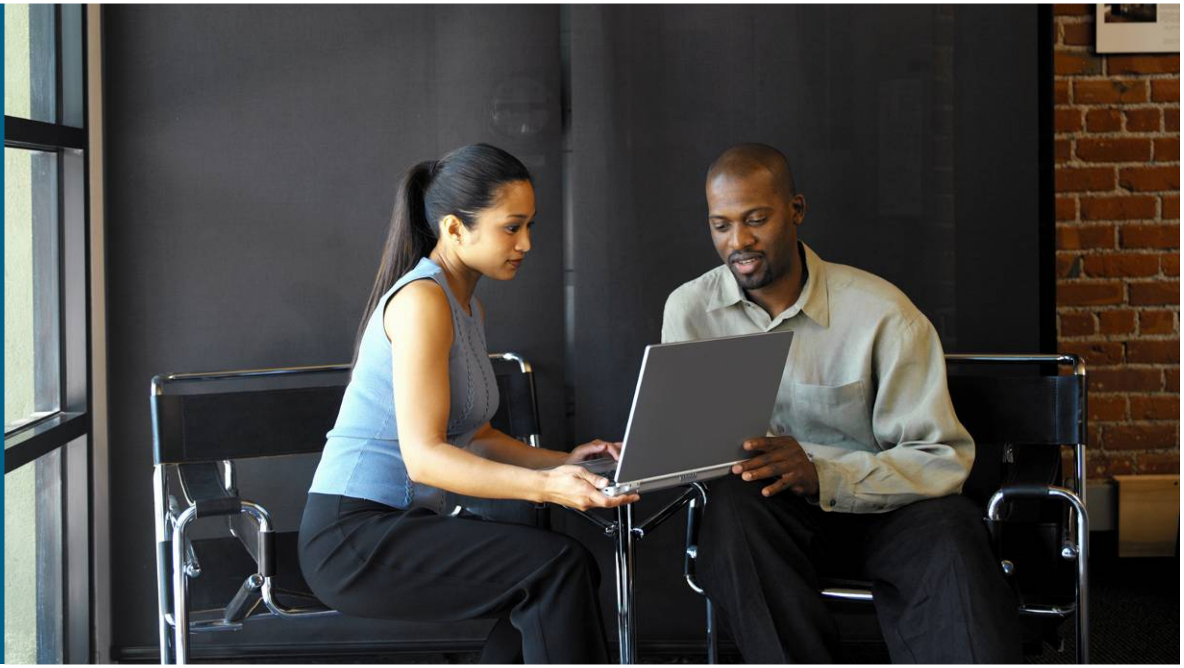
- **Loss**

  Packet/Frame Loss

So, Where’s the Need for QoS in WLAN?

- In the past, there hasn’t been much QoS need
  - WLANs designed for coverage, basic data access
- Now, WLANs designed for mission-critical applications
  - Shift from coverage to capacity to allow for
    - More client devices
    - More requiring data apps (higher bandwidth needs)
    - Emergence of voice- and video-over-WLAN needs
      - Proliferation of voice handsets (such as Cisco’s 7920)
      - Latency-sensitive applications (softphone applications, IPTV, etc)

How Does QoS Work Today in 802.11?



… It Doesn't

- **802.11 networks are completely egalitarian**
  - **Every device, AP included, has equal access to transmit**
    - **No device has precedence over any other**

Example: voice handsets abide by the same access rules as laptops

- **All transmissions for each individual device have the same access, transmitting in 'FIFO' fashion**
  - **No application has more transmit 'weight' than any other**

Example: on a single laptop, a voice frame has the same right to transmit as any other frame, such as a web frame

802.11's Access Rules

- **Distributed Coordinated Function (DCF)**

  Transmission rules followed by all clients

  DCF is 802.11's rules of the road

- **Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA)**

  Responsible for mediating access to the air

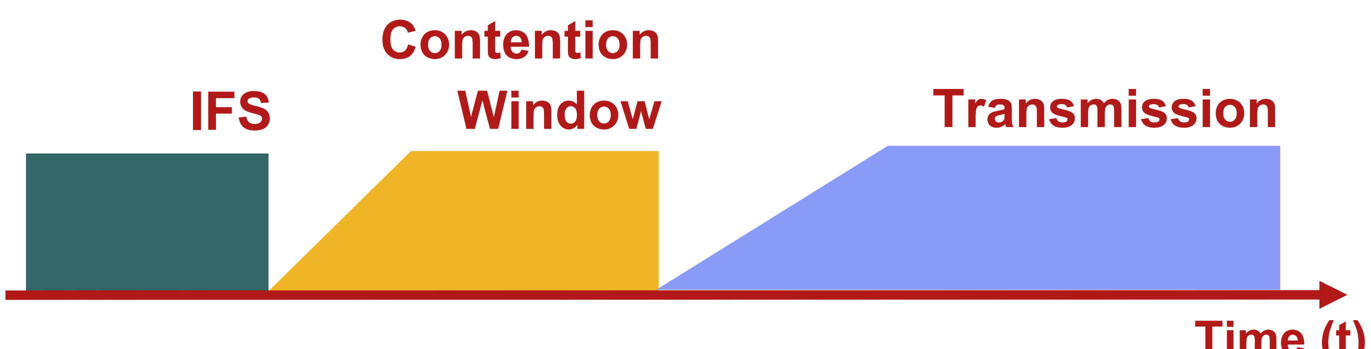
  CSMA/CA is 802.11's traffic lights

802.11’s Access Rules

- **Distributed Coordinated Function (DCF)**
  - **Transmission rules followed by all clients**
  - **Interframe spaces (IFS) are used to ‘prioritize’ traffic**
    - **IFS are very short delays before transmissions are allowed**
      - **The Short Interframe Space (SIFS) is used for transmission of management and control frames**
      - **The DFS Interframe Space (DIFS) is used before the transmission of data frames**
- **CSMA/CA allows ‘peaceful’ coexistence of many devices trying to transmit simultaneously**

802.11's Access Mediation

- Carrier Sense Multiple Access with Collision Avoidance
  - CSMA/CA responsible for mediating access to the air
    - Reduces the likelihood of a transmission collision
  - Provides probabilistically fair access to every device
- CSMA/CA provides a framework clients follow before being allowed to transmit: 'Listen before talk'
  - Wait the appropriate interframe space (SIFS or DIFS)
  - If medium is free, wait to make sure no one else is beginning to transmit (this is called the 'backoff')
  - If medium is still free after 'backoff', then transmit



Dissecting CSMA/CA

- **Two sensing methods**
    - **Physical Carrier Sense**
    - **Virtual Carrier Sense**
        - **Directed management and data frames include a Duration ID**
            - **Used to indicate to non-sending/receiving devices how long the medium will be occupied**
            - **Clients then set a timer: Network Allocation Vector (NAV)**
            - **This timer is decremented, and once at 0, physical carrier sense is invoked and the process starts over**
- **Only when both sensing mechanisms indicate the air is free do clients then begin to prepare to wait before transmission**
- **'Carrier Sense' allows more than one device to 'talk' on the network, granting 'Multiple Access'**

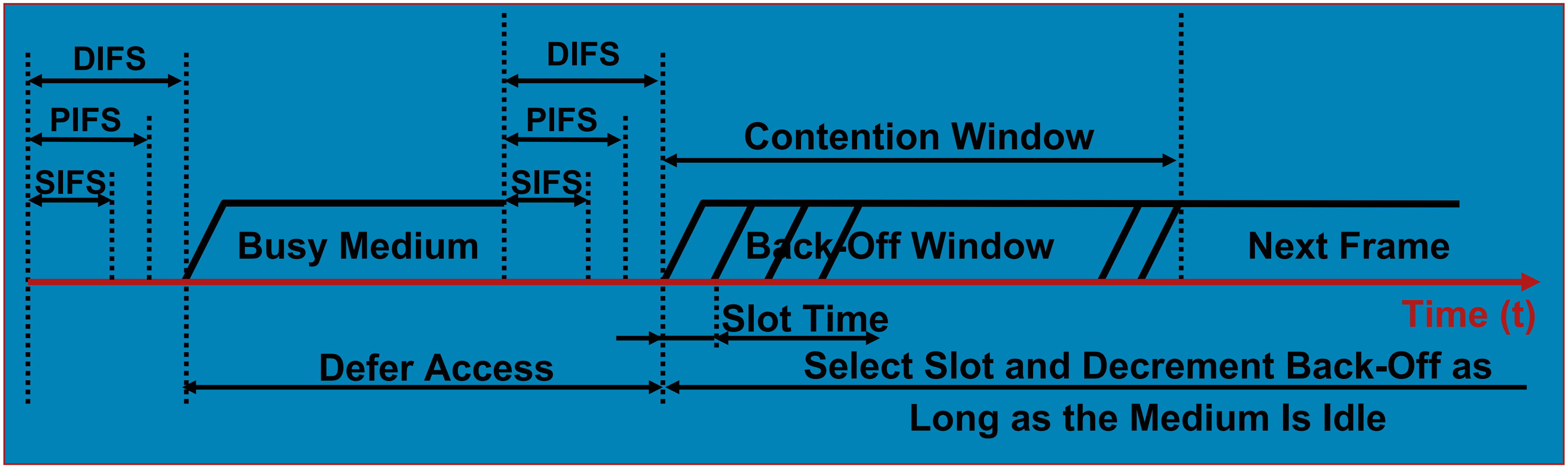
Dissecting CSMA/CA (continued)

- **After the medium is determined free, the transmitting station waits the appropriate IFS**
- **The 'Contention Window' then starts when the station selects a random backoff duration**

  Backoff is derived by multiplying slot time (20µs for 11b, 9µs for 11a/g) by a 'randomly' selected number between 0 and aCWmin (31 for 11b/g, 15 for 11a)

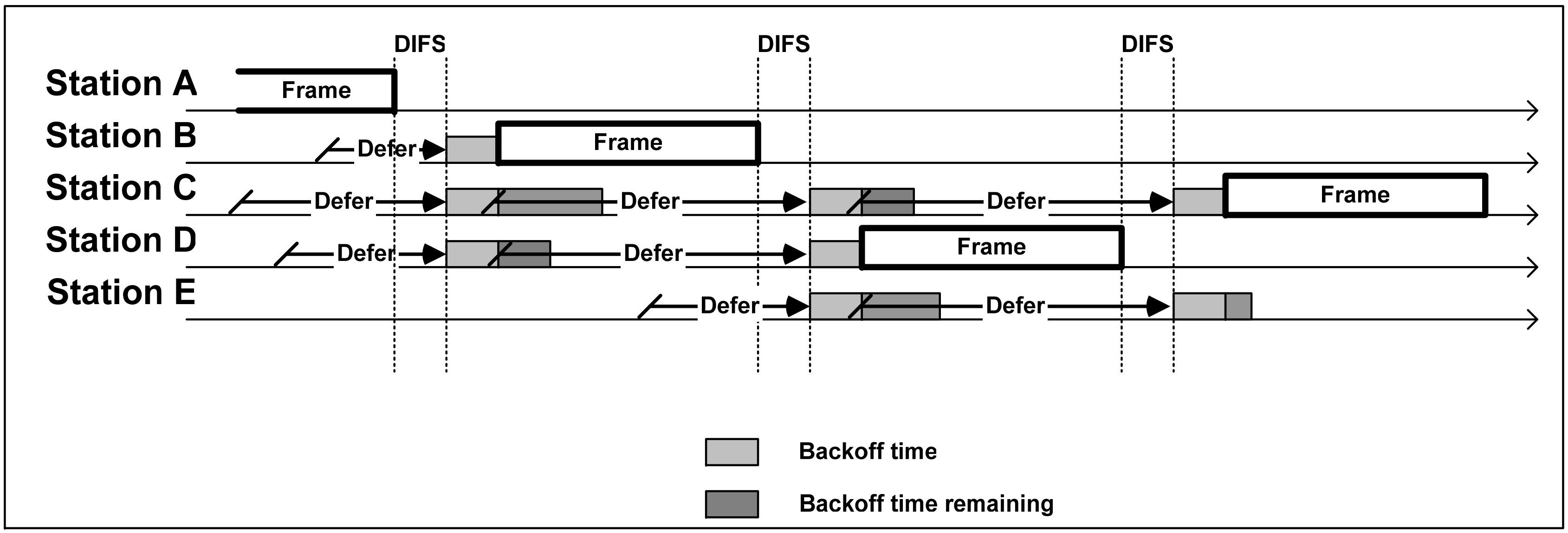
Simple Transmission Example

- **Transmission access only when the medium is free for the appropriate IFS, plus the contention window**

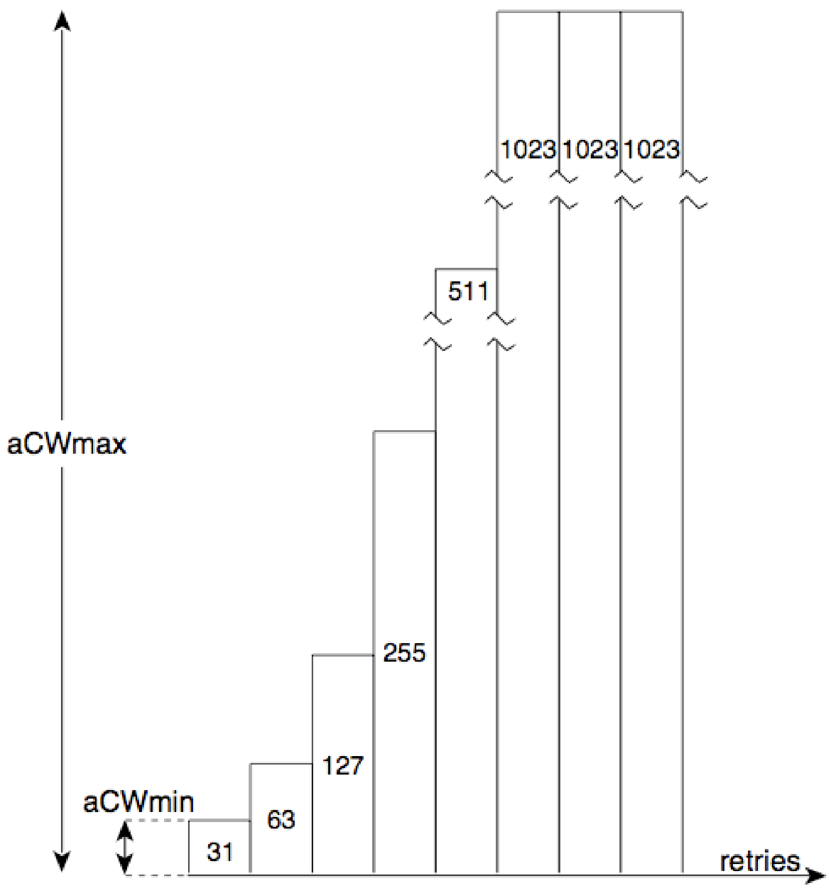


Extended Transmission Example

- If another station begins to transmit before the backoff counter fully decrements, the countdown is simply deferred until the present transmission ends



Retransmissions with CSMA/CA



For 11b/g, aCWmin is 31 and aCWmax is 1023

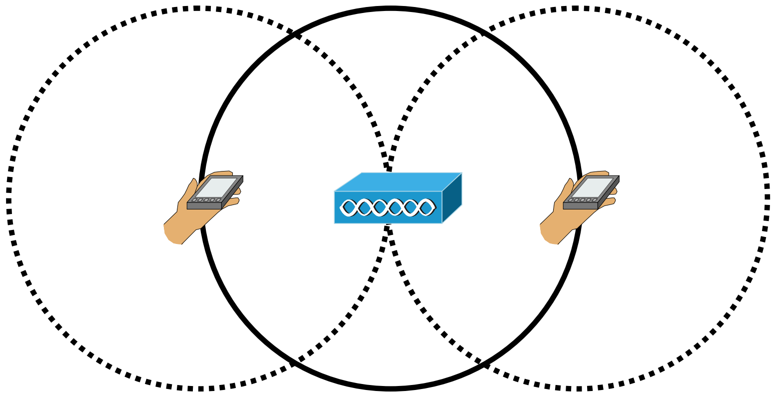
For 11a, aCWmin is 15 and aCWmax is 1023

- 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 successful transmission, or the size of the window equals aCWmax. If necessary, retries continue until the maximum retries or time to live (TTL) is reached.

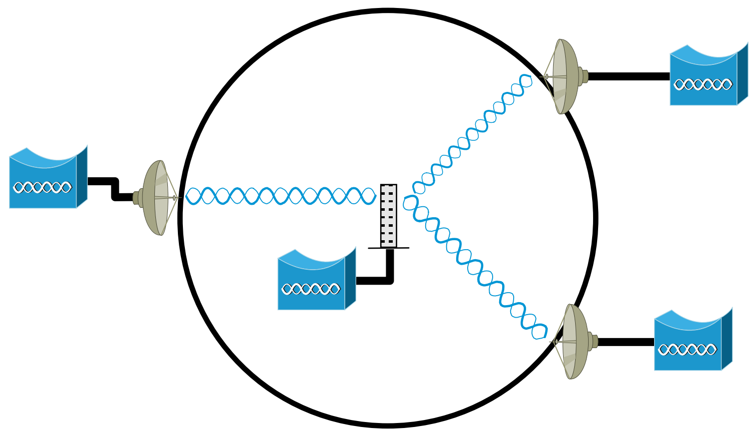
Hidden Nodes Are Also a QoS Issue

- If you can't hear a frame, you can't avoid colliding with it
- Only the AP can see and be seen by all nodes
- The virtual carrier sense mechanism can be aided by RTS/CTS and CTS-to-self
- 11b and 11g coexistence creates a hidden node potential

  CTS-to-self is typically used



Hidden Nodes Due to Range



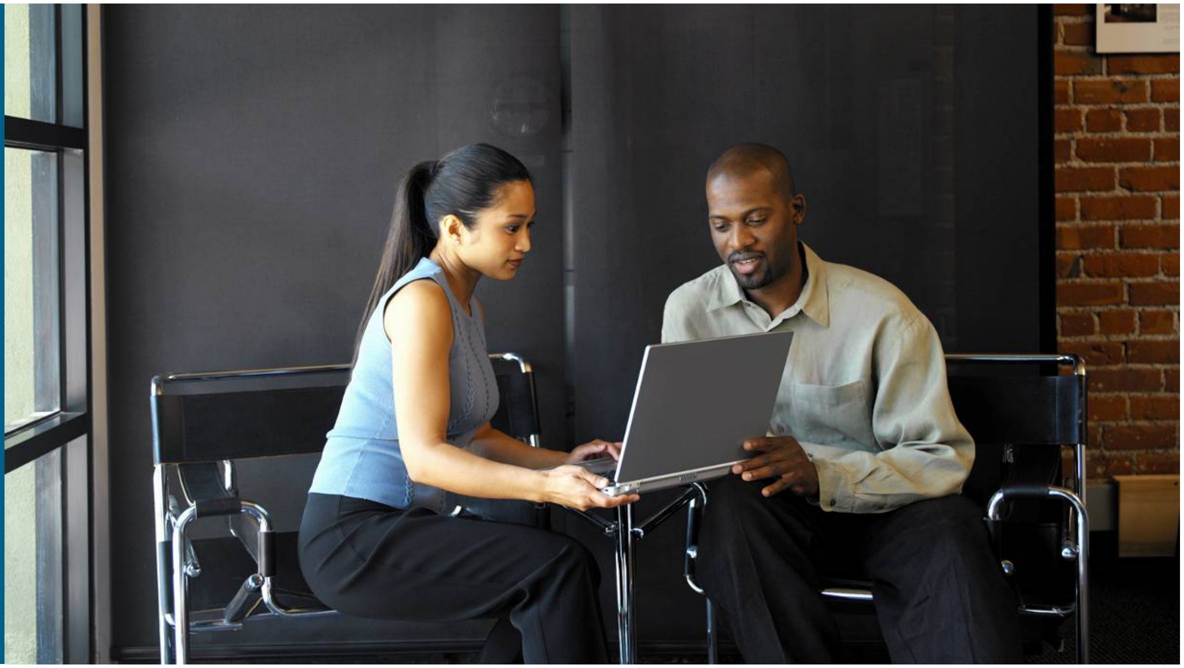
Hidden Nodes Due to Directional Antennas

Retrofitting 802.11 with QoS

- **Intelligent queuing at the AP allows the WLAN to realize downstream, over-the-air QoS**
  - **Upstream, from AP to controller QoS applied, as well**
- **Prioritization done per-WLAN**
  - **Additionally/alternatively, QoS can be assigned per user via Identity-based Networking Services (IBNS)**
- **DiffServ and 802.1p priority preserved upstream and downstream between AP and controller**

--- To complete WLAN QoS, add WMM/802.11e ---

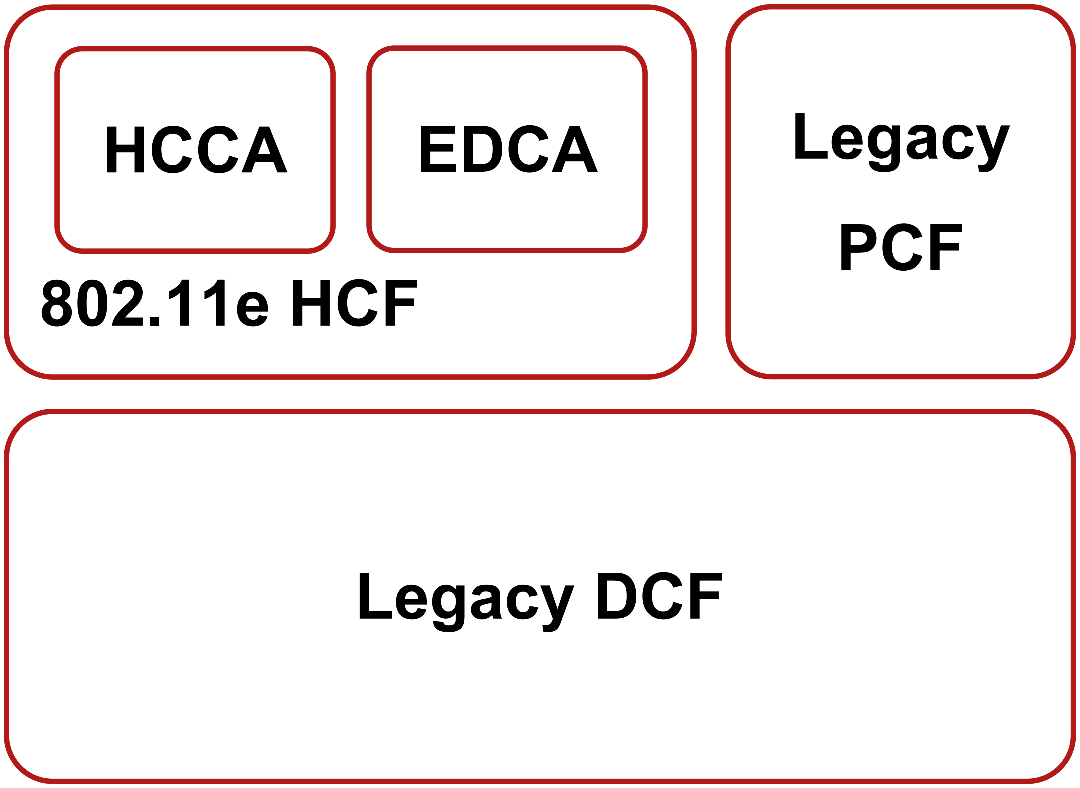
Introduction to 802.11e and WMM



QoS with WMM/802.11e

- IEEE TGe ratified 802.11e in late 2005
- The WFA moved early with a subset of 11e called Wireless Multimedia (WMM), similar to the way they did with WPA/11i
- WMM specifies a subset of 11e functionality, called EDCA
- 11e outlines two modes of operation, EDCA and HCCA

802.11e Architectural Relationship



EDCA and HCCA

- **EDCA is similar to DCF's contention-based access model in that it is up to each individual device to determine when it is allowed to access the medium**

  Prioritization is done by allowing differing traffic types varied access levels based on how long they wait to transmit

- **HCCA is conceptually similar to 802.11's PCF model by prioritizing traffic via polling from the AP**

  More advanced than PCF because traffic classes are supported and bandwidth can be more intelligently allocated to devices that require it

HCCA Scheduled Access

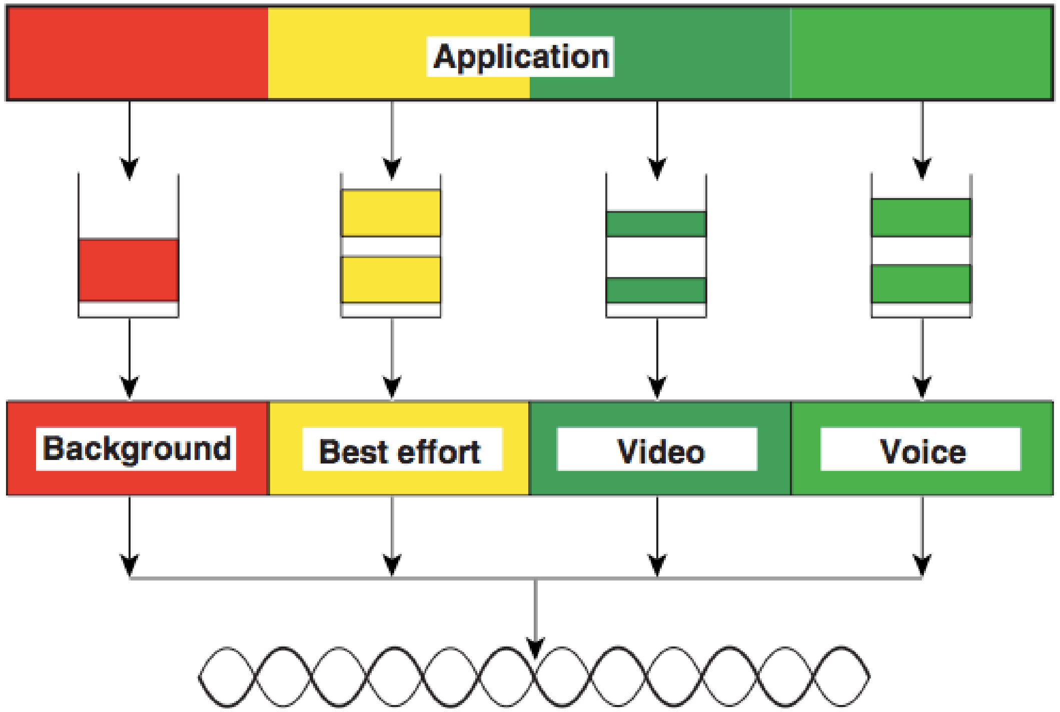
- **HCCA allows the AP to ‘poll’ supporting devices for transmissions**
  - Granting of parameterized ‘TXOPs’ allow for intelligent allocation of bandwidth
  - Reduces the overhead of contention
- **Presently no client support**
  - WFA continues to discuss plans of additions to WMM to include the HCCA portion of 11e
- **EDCA is the present go-to-market QoS strategy**

Prioritizing with EDCA Access Categories

- Prioritizing only on a per-station basis overlooks the need to prioritize different types of traffic within a single device
    - Classifying traffic allows different levels of priority within a single device
- WMM/11e specifies four different classifications of traffic called access categories, or ACs
    - Background
    - Best-effort
    - Video
    - Voice

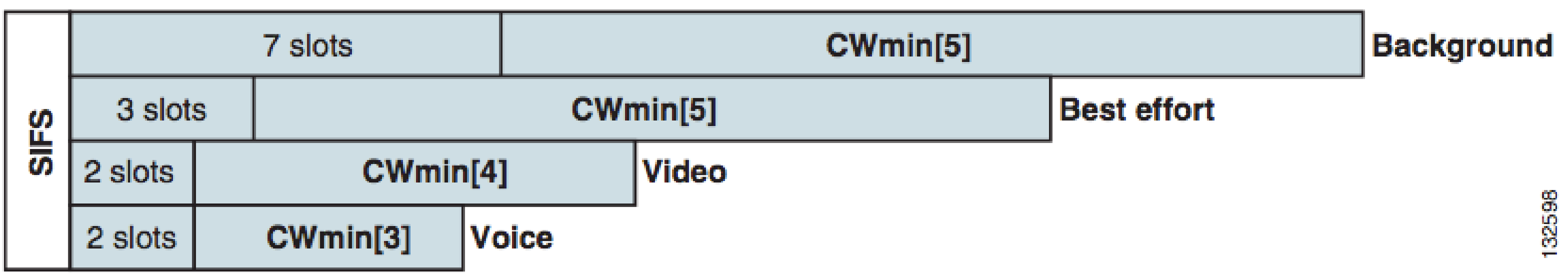
From Application to Transmission

- **Based on application tagging, frames are passed to the appropriate access categories, and then queued for transmission**



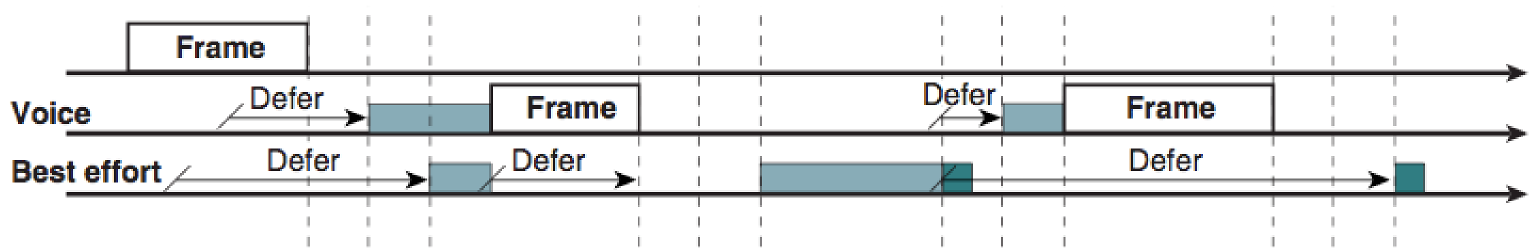
EDCA Backoff

- **Frames in each access category queue have differing backoff delays, referred to as the Arbitration Interframe Space Number (AIFSN), followed by varied contention windows**
  - **This is advertised in the AP's beacon frames and probe responses**



EDCA Contention

- **Within a station, frames in different access categories contend with each other according to their respective access properties**
- **Frames from different stations then contend with each other in the same fashion over the air**



QoS Frame Tagging

- **Each category corresponds to an 802.1p/D classification and is tagged accordingly for upstream QoS preservation**
- **In the downstream, WMM/11e category and associated backoff interval is determined by frame/packet tagging**

Priority	802.1 Priority (=User Priority)	802.1p Designation	Access Category	WMM Designation
Highest	1	BK Background	AC_BK	Background
	2	-Spare		
Lowest	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
	7	NC Network Control "must get there"		

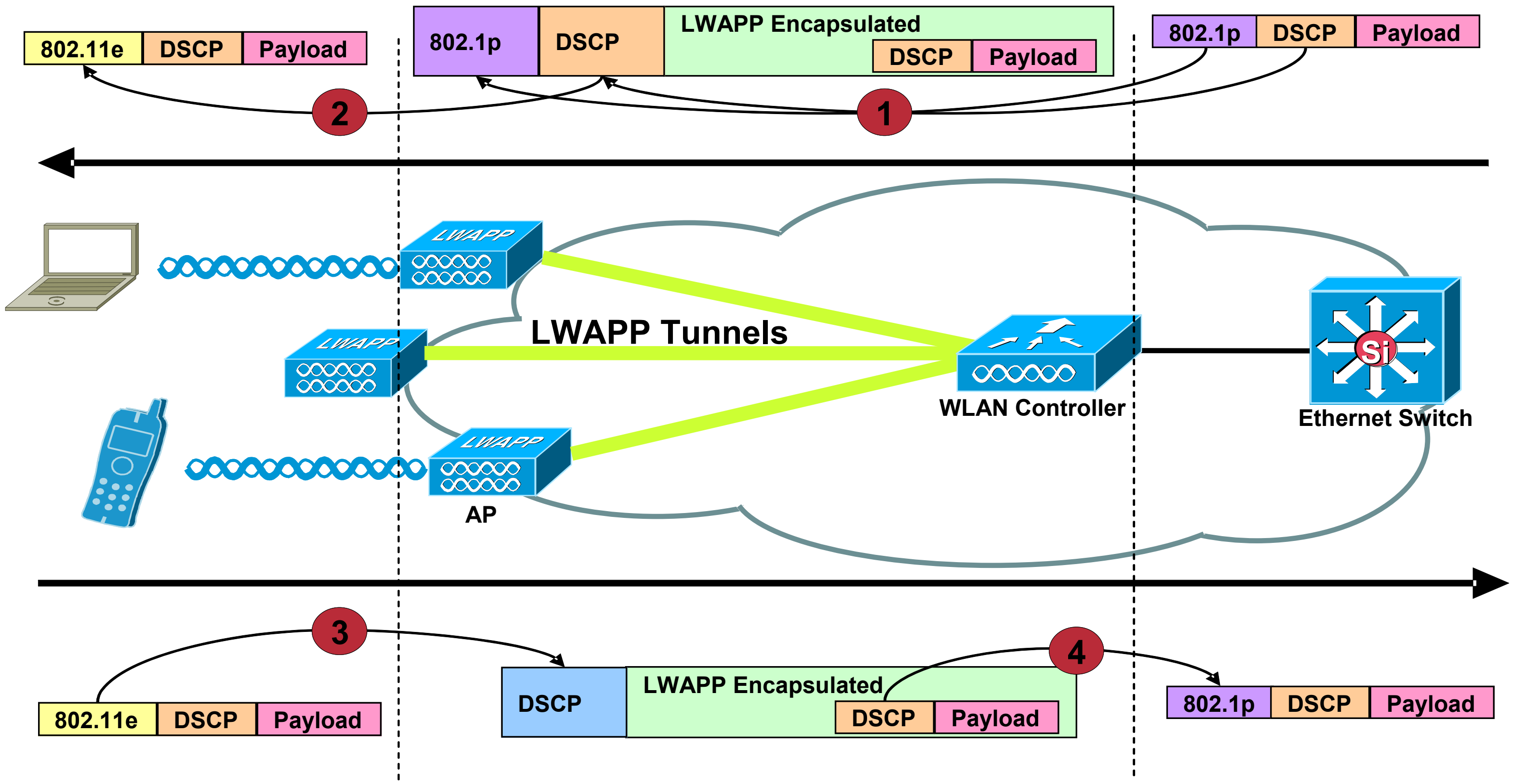
Simplified 802.1p/D to EDCA AC Mapping

UP - Same as 802.1p/D)	802.1p/D Designation	AC	802.11e Designation
1	BK	1	Background
2	-	1	Background
0	BE	0	Best Effort
3	EE	0	Best Effort
4	CL	2	Video
5	VI	2	Video
6	VO	3	Voice
7	NC	3	Voice

11e and 802.1p/D to DSCP Tagging

802.1p/D UP-based Traffic Type	DSCP UP	11e UP
Inter-network control (LWAPP, 802.11 mgmt)	48	7
Voice	46 (EF)	6
Video	34 (AF41)	5
Voice Control	26 (AF31)	4
Background (Gold)	18 (AF21)	2
Background (Gold)	20 (AF22)	2
Background (Gold)	22 (AF23)	2
Background (Silver)	10 (AF11)	1
Background (Silver)	12 (AF12)	1
Background (Silver)	14 (AF13)	1
Best Effort	0 (BE)	0, 3
Background	2	1

WMM/11e, DSCP, Dot1p Relationship



Note: Over-the-air packets to the AP are policed allowing edge switches to trust the QoS marking of the packets from the AP

Client-Side Trusted QoS

- Whether using EDCA today or HCCA tomorrow, upstream QoS from the client isn't a possibility without the necessary client piece
- Not only does the client need to support WMM/11e QoS, but the client needs to know how to mark traffic to the appropriate access category
- Even with the necessary AC marking, this upstream marking needs to be trusted
    - Without trusted QoS, DoS attacks may be sanctioned

Trusted Client-Side QoS Configuration

- **Cisco's Security Agent (CSA) allows centralized control of client-side QoS marking on a per-application basis**

    The CSA Management Center (CSAMC) allows QoS rules to be securely passed to CSA client-side software

    This not only allows applications' traffic to be properly applied to correct access categories, but it restricts users, and more likely, self-appraising applications, from overriding such classification

- **Such a client-side software piece isn't necessary for purpose-built devices such as Cisco's 7920**

Admission Control

- **Admission control allows metered client access based on available resources at the AP**
- **Client can intelligently select access points based on advertised load information**
- **Admission control is typically performed to optimize for voice, called call admission control (CAC)**
  - **By indicating to a phone when too many AP resources are occupied, it may alternatively seek out a less-loaded access point**
- **Admission control is performed in a couple of ways**

Approaches to Admission Control

- Admission control grants clients access on a per-WLAN (SSID) basis
- This is typically done in one of two ways
  - Load – number of calls based on channel load
  - TSpec – based on a host of additional parameters
- WMM/11e's Traffic Specification (TSpec) takes much more into account
  - Clients request admission based on: traffic priority (access category), power save, mean data rate, frame sizes, minimum PHY data rate, etcetera

Optional WMM/11e Components

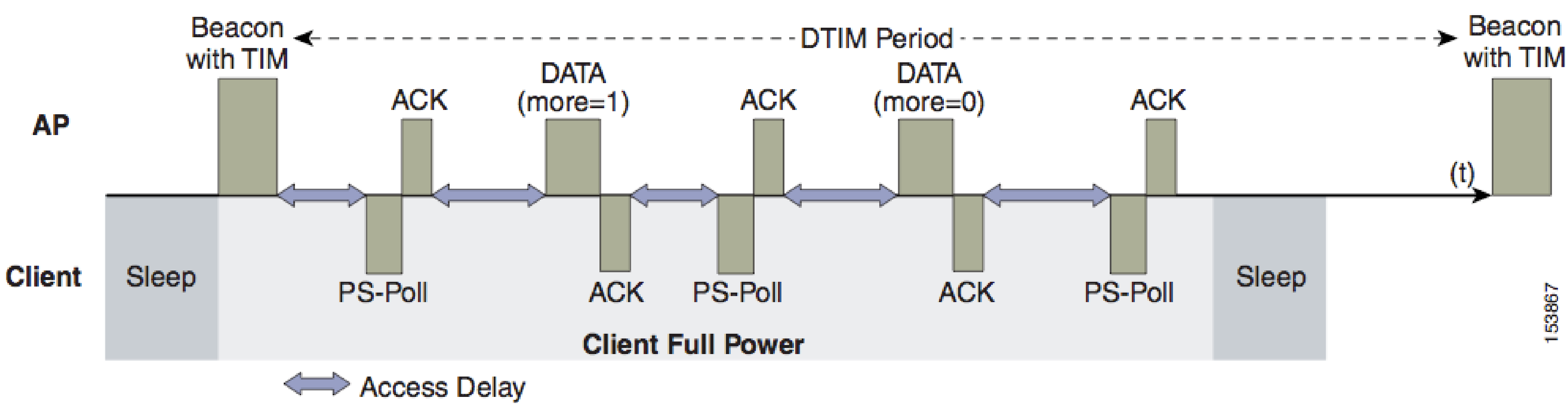
- Support for Block Acknowledgements
  - With the option of not requiring any ACKs
- Direct Link Setup (DLS)
  - Allows Station-to-Station direct access without first transmitting data through the access point
- Contention-free Bursts (CFB)
  - Allows stations to make further use of transmission opportunities to send frames without contending for the medium
    - May be optionally used to grant more time to faster 11g stations in coexistence scenarios
- Automatic Power-Save Deliver (APSD)

Automatic Power-Save Delivery

- Power saving mechanisms are very important for many battery powered devices, specifically phone handsets
  - Allows devices to ‘sleep’ when neither transmitting nor receiving
- EDCA uses the unscheduled portion of APSD (U-APSD)
  - Instead of being required to wake up periodically to see if data is waiting for the sleeping stations (à la PSP), clients trigger the AP to forward queued traffic
  - This drastically lowers latency and still allows devices to save battery power

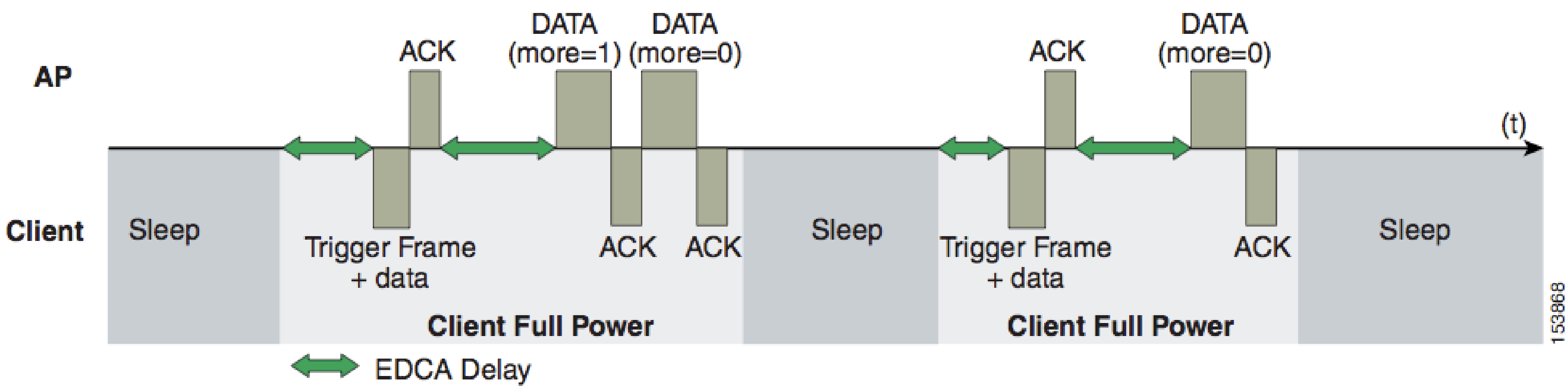
802.11's Power-Save Polling

- **Disadvantageous balance of power savings and increased latency**

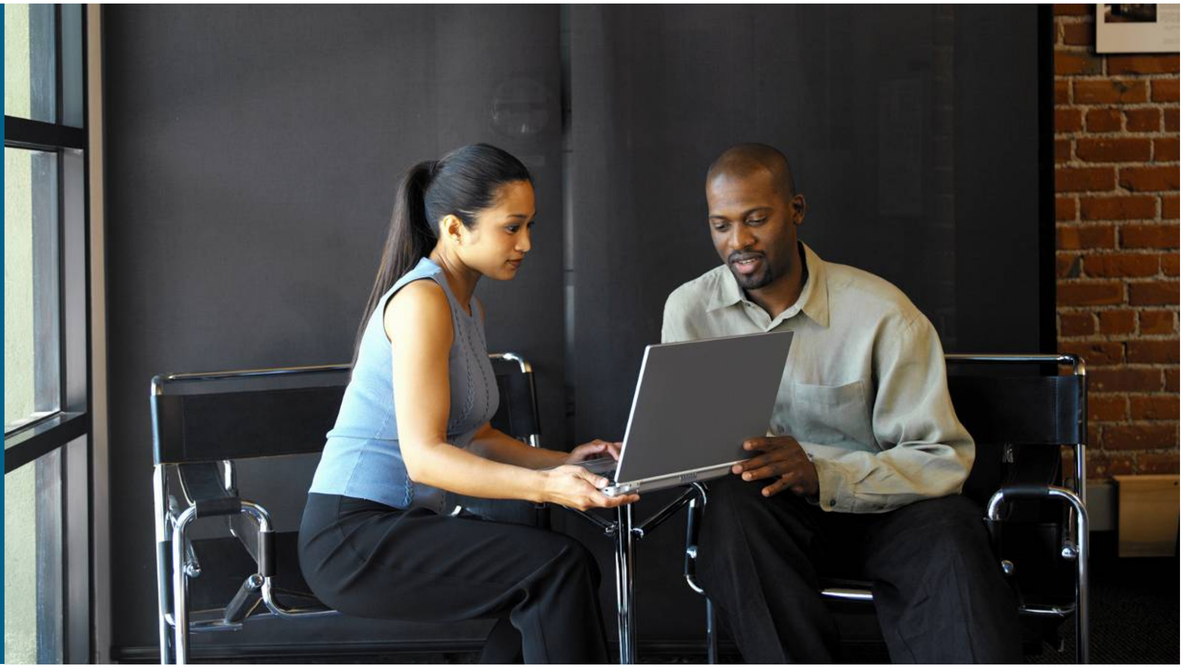


WMM/11e's U-APSD

- **Blends battery savings with increased efficiency to yield lower latencies**



Designing for Voice Beyond QoS



Designing for Voice Beyond QoS

- Don’t underestimate the importance of fast, secure roaming (with CCKM)
- Continue to rely on RRM for dynamic power and channel assignment and automated RF fault tolerance
- Disable Aggressive Load Balancing (a global parameter)
- Look to CCX for assurance that not only much-needed WMM/802.11e features are present, but other important voice features, as well

How CCX Benefits Voice

Feature		Benefit
CCKM Support for EAP-Types	⟺	Locally Cached Credentials Means Faster Roams
Unscheduled Automatic Power Save Delivery (U-APSD)	⟺	More Channel Capacity and Better Battery Life
TSPEC-Based Call Admission Control (CAC)	⟺	Managed Call Capacity for Roaming and Emergency Calls
Voice Metrics	⟺	Better and More Informed Troubleshooting
Neighbor List	⟺	Reduced Client Channel Scanning
Load Balancing	⟺	Calls Balanced Between APs
Dynamic Transmit Power Control (DTPC)	⟺	Clients Learn a Power to Transmit At
Assisted Roaming	⟺	Faster Layer 2 Roams

References and Further Reading

- The IEEE 802.11 Handbook: A Designer's Companion by Bob O'Hara and Al Petrick (Second Edition)
- Cisco's Enterprise Mobility Design Guide (Version 3.0)

  http://www.cisco.com/univercd/cc/td/doc/solution/emblty30.pdf
- Cisco 7920 Phone Design and Deployment Guide

  http://www.cisco.com/en/US/products/ps6366/prod_technical_reference09186a00805e75a1.html

Meet the Experts
Campus and Wireless Evolution

- Mark Montanez
  Corporate Dev Consulting Engineer

- Tim Szigeti
  Technical Leader

- Sujit Ghosh
  Technical Mktg Eng

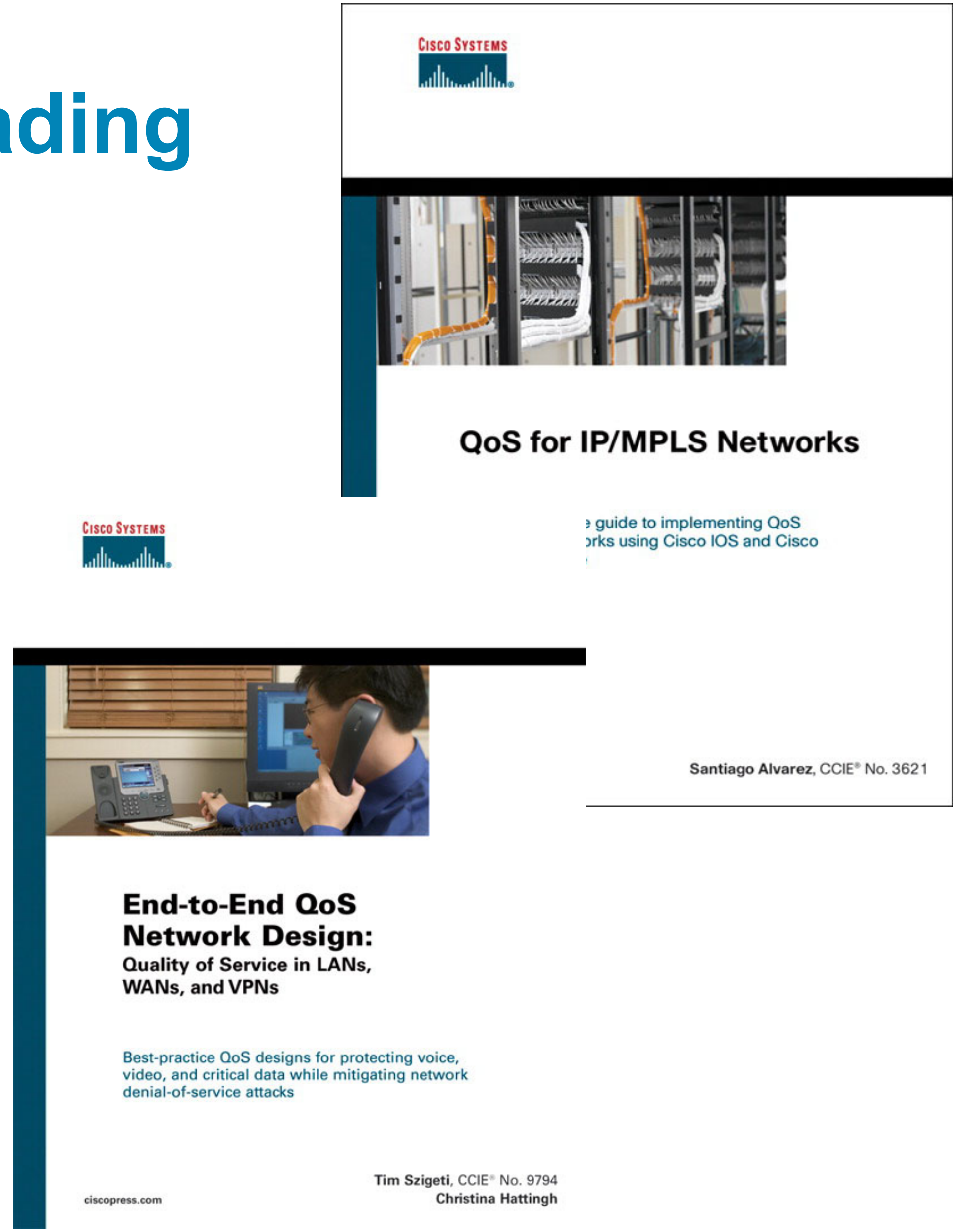
- Victor Moreno
  Technical Leader

- Mike Herbert
  Technical Leader

Recommended Reading

BRKCAM -3012

- QoS in Multiservice Networks
- End-to-End QoS Network Design: Quality of Service in LANs, WANs, and VPNs



Available in the Cisco Company Store

Q & A