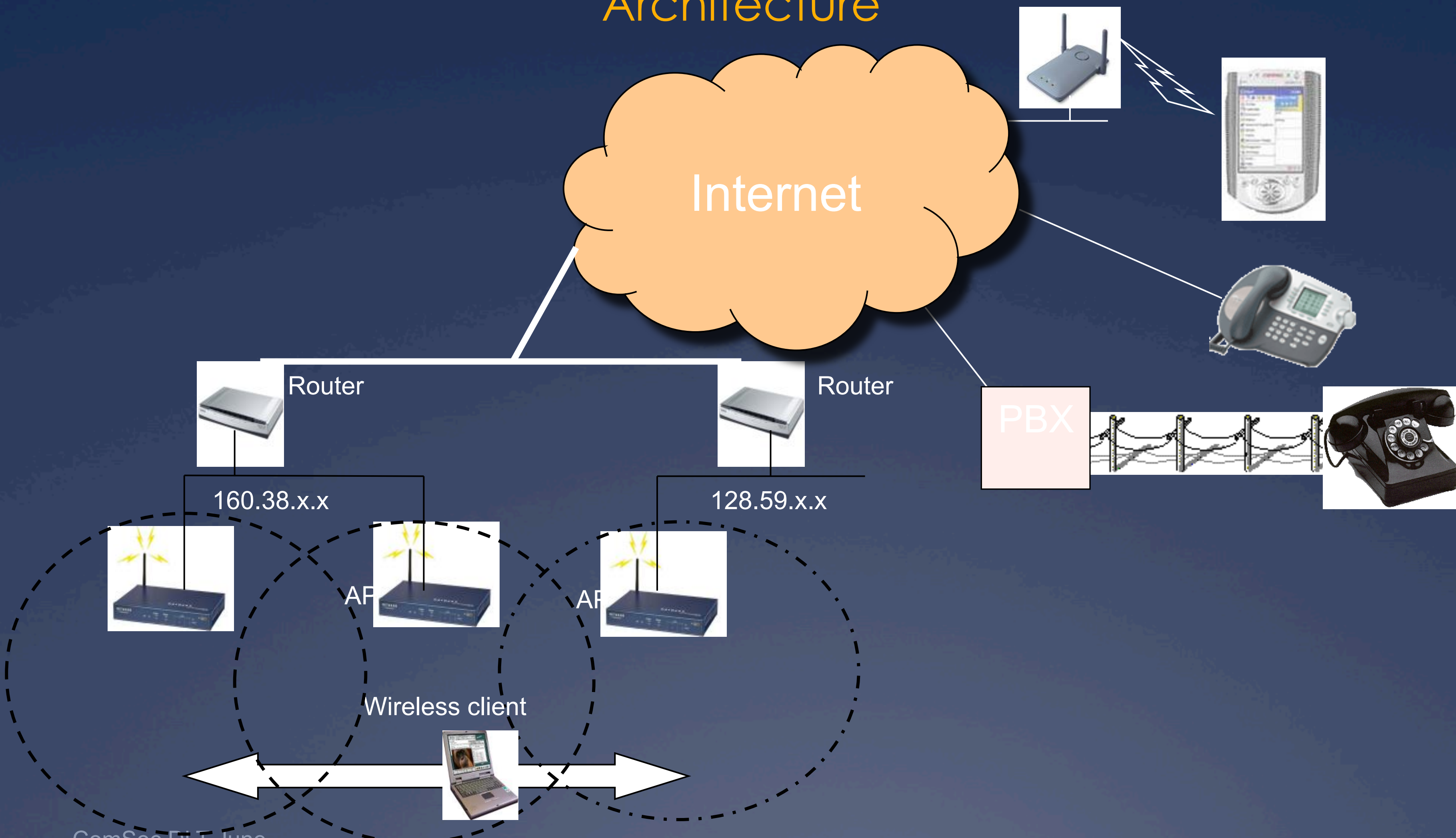
VoIP in 802.11 Wireless Networks

Henning Schulzrinne
with Andrea G. Forte, Sangho Shin
Department of Computer Science
Columbia University

VoIP and IEEE 802.11

Architecture



VoIP and IEEE 802.11

Problems

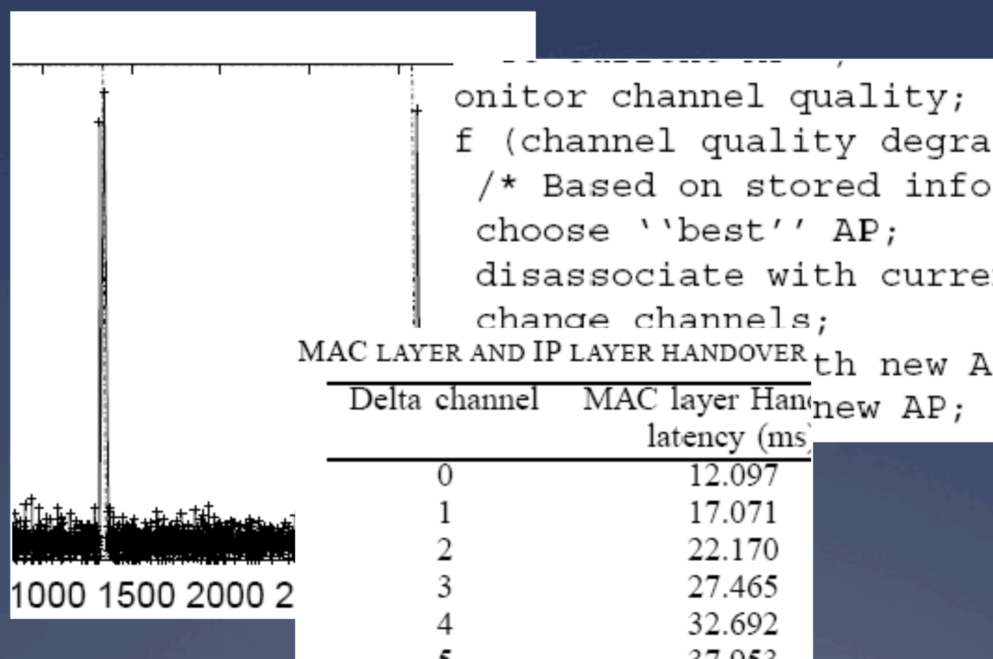
- Support for real-time multimedia
  - Handoff
    - L2 handoff
      - Scanning delay
    - Authentication
      - 802.11i, WPA, WEP
    - L3 handoff
      - Subnet change detection
      - IP address acquisition time
    - SIP session update
      - SIP re-INVITE
  - Low capacity
    - Large overhead
    - Limited bandwidth
    - Unfair resource distribution between uplink and downlink
  - Call Admission control
    - Difficult to predict the impact of new calls
- Wireless coverage
  - both 802.11 and cellular coverage has holes

VoIP and IEEE 802.11

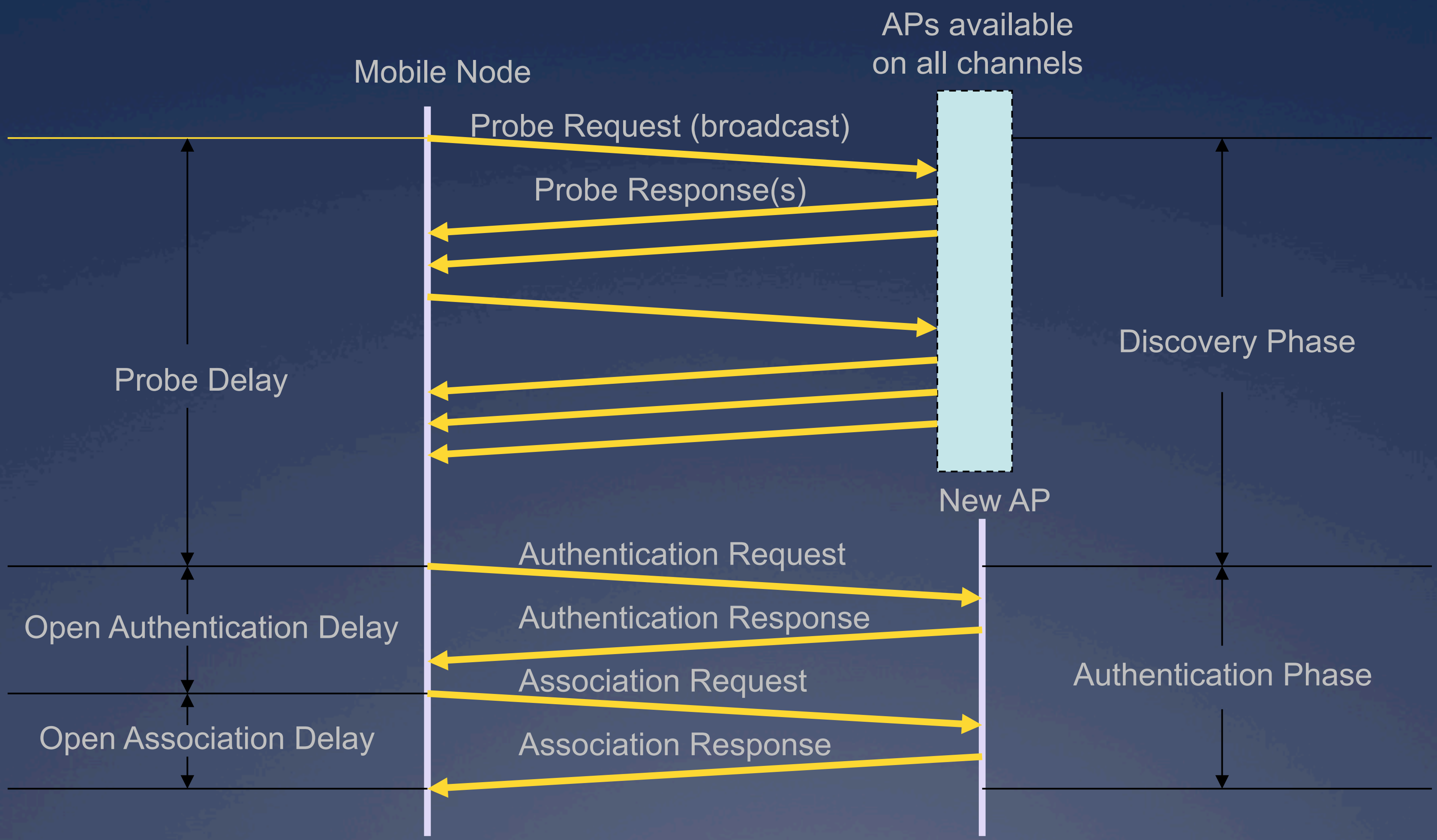
Solutions

- Support for real-time multimedia
  - Handoff
    - Fast L2 handoff
    - Fast L3 handoff
    - Passive DAD (pDAD)
    - Cooperative Roaming (CR)
  - Low capacity
    - Dynamic PCF (DPCF)
    - Adaptive Priority Control (APC)
  - Call admission control
    - Queue size Prediction using Computation of Additional Transmissions (QP-CAT)
  - Signaling hand-off
    - GSM + SIP

Reducing MAC Layer Handoff in IEEE 802.11 Networks



Layer 2 Handoff



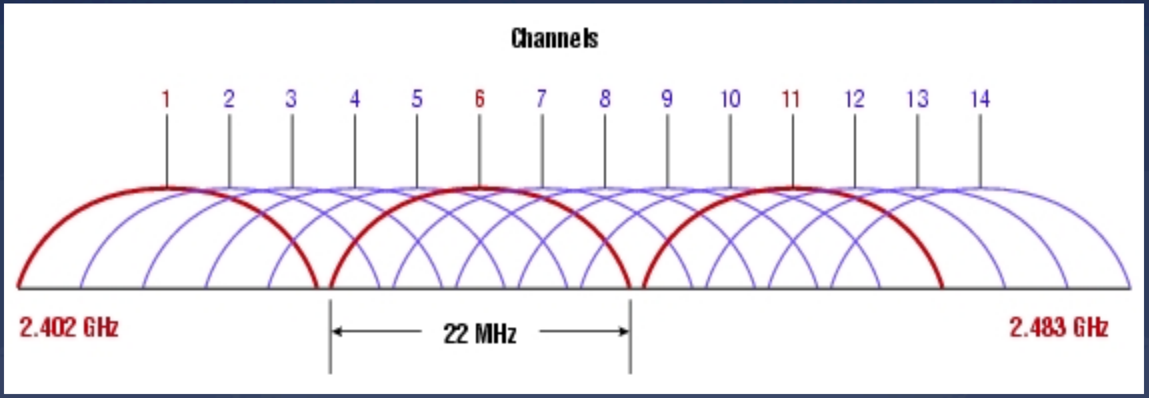
Fast Layer 2 Handoff
Overview

- Problems
  - Handoff latency is too big for VoIP
    - Seamless VoIP requires less than 90ms latency
    - Handoff delay is from 200ms to 400ms
  - The biggest component of handoff latency is probing (over 90%)

- Solutions
  - Selective scanning
  - Caching

Fast Layer 2 Handoff

Selective Scanning



♪

* In most of the environments (802.11b & 802.11g), only channel 1, 6, 11 are used for APs
* Two APs that have the same channel are not adjacent (Co-Channel interference)

Scan 1, 6, 11 first and give lower priority to other channels that are currently used

Fast Layer 2 Handoff

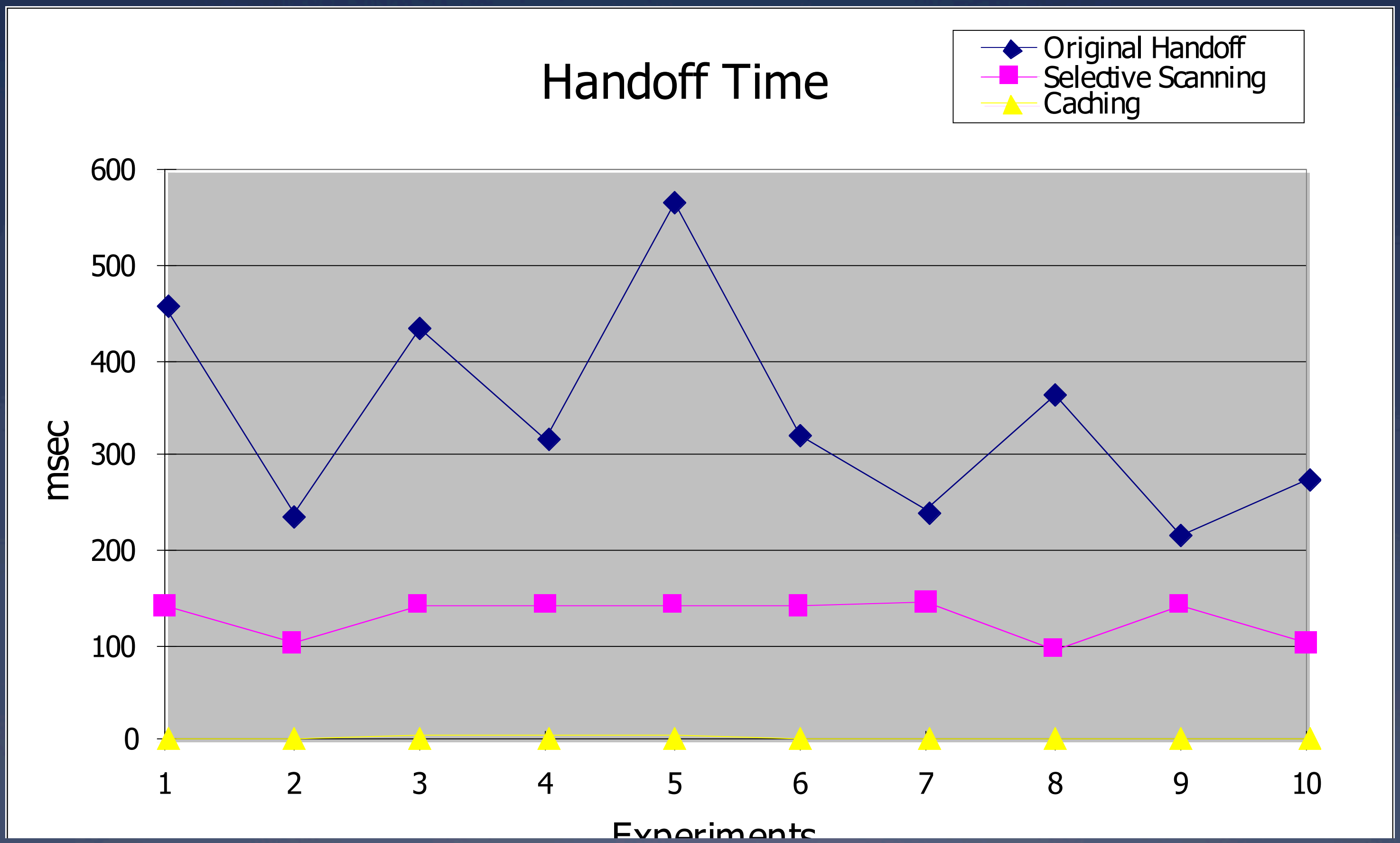
Caching

- Background
  - Spatial locality (Office, school, campus…)
- Algorithm
  - After scanning, store the candidate AP info into cache (key=current AP).
  - Use the AP info in cache for association without scanning when handoff happens.

	Key	AP1	AP2
1	Current AP	Next best AP	Second best AP
	….	….	….
N			

Fast Layer 2 Handoff

Measurement Results – Handoff time



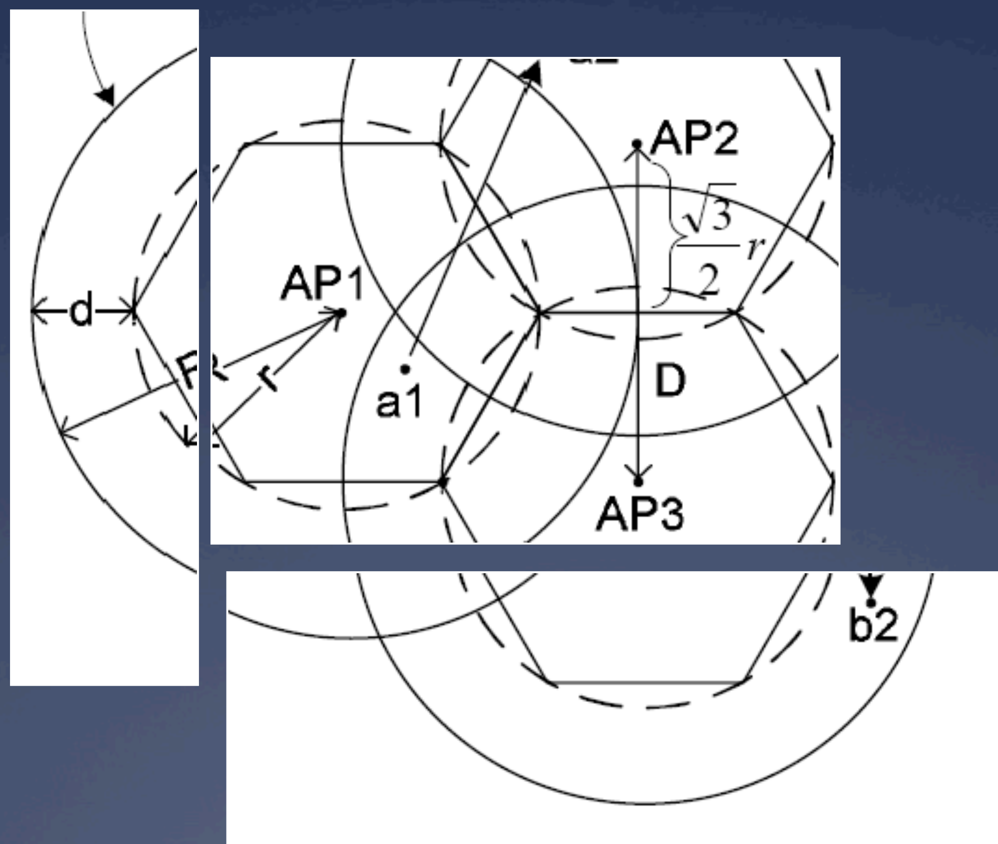
Fast Layer 2 Handoff

Conclusions

* Fast MAC layer handoff using selective scanning and caching
* Selective scanning : 100-130 ms
* Caching : 3-5 ms
* Low power consumption (PDAs)

- Don't need to modify AP, infrastructure, or standard. Just need to modify the wireless card driver!

Layer 3 Handoff



L3 Handoff

Motivation

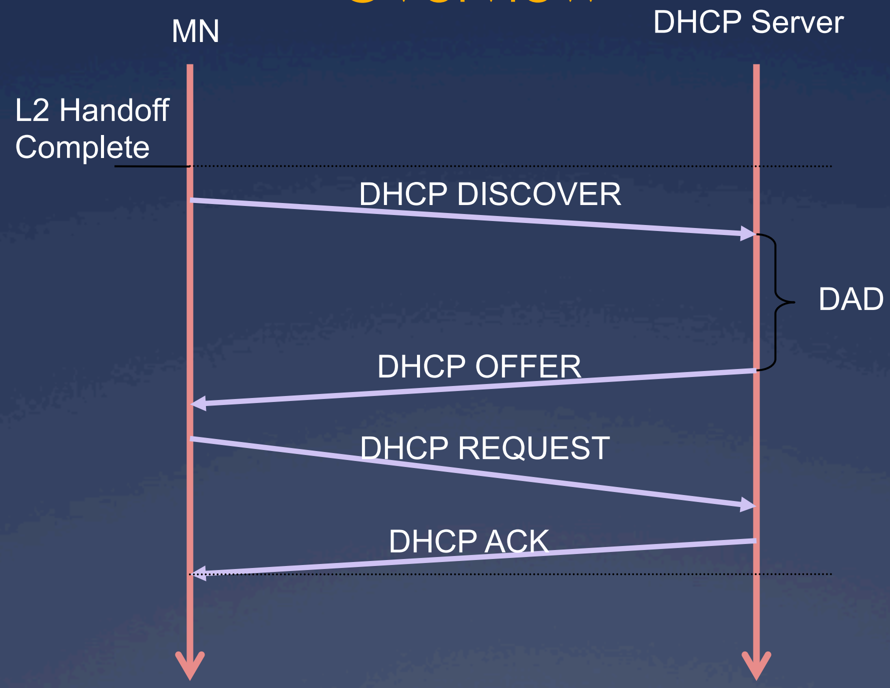
- Problem
  - When performing a L3 handoff, acquiring a new IP address using DHCP takes on the order of one second

➡ The L3 handoff delay too big for real-time multimedia sessions

- Solution
  - Fast L3 handoff
  - Passive Duplicate Address Detection (pDAD)

Fast L3 Handoff

Overview



- We optimize the layer 3 handoff time as follows:
  - Subnet discover
  - IP address acquisition

Fast Layer 3 Handoff
Subnet Discovery (1/2)

- Current solutions
  - Router advertisements
    - Usually with a frequency on the order of several minutes
  - DNA working group (IETF)
    - Detecting network attachments in IPv6 networks only

No solution in IPv4 networks for detecting a subnet change in a timely manner

Fast Layer 3 Handoff
Subnet Discovery (2/2)

- Our approach
  - After performing a L2 handoff, send a bogus DHCP_REQUEST (using loopback address)
  - DHCP server responds with a DHCP_NAK which is relayed by the relay agent
  - From the NAK we can extract subnet information such as default router IP address (IP address of the relay agent)
  - The client saves the default router IP address in cache
  - If old AP and new AP have different default router, the subnet has changed

Fast Layer 3 Handoff

Fast Address Acquisition

- IP address acquisition
  This is the most time consuming part of the L3 handoff process → DAD takes most of the time
  We optimize the IP address acquisition time as follows:
  - Checking DHCP client lease file for a valid IP
  - Temporary IP ("Lease miss") → The client "picks" a candidate IP using particular heuristics
  - SIP re-INVITE → The CN will update its session with the TEMP_IP
  - Normal DHCP procedure to acquire the final IP
  - SIP re-INVITE → The CN will update its session with the final IP

While acquiring a new IP address via DHCP, we do not have any disruption regardless of how long the DHCP procedure will be.
We can use the TEMP_IP as a valid IP for that subnet until the DHCP procedure ends.

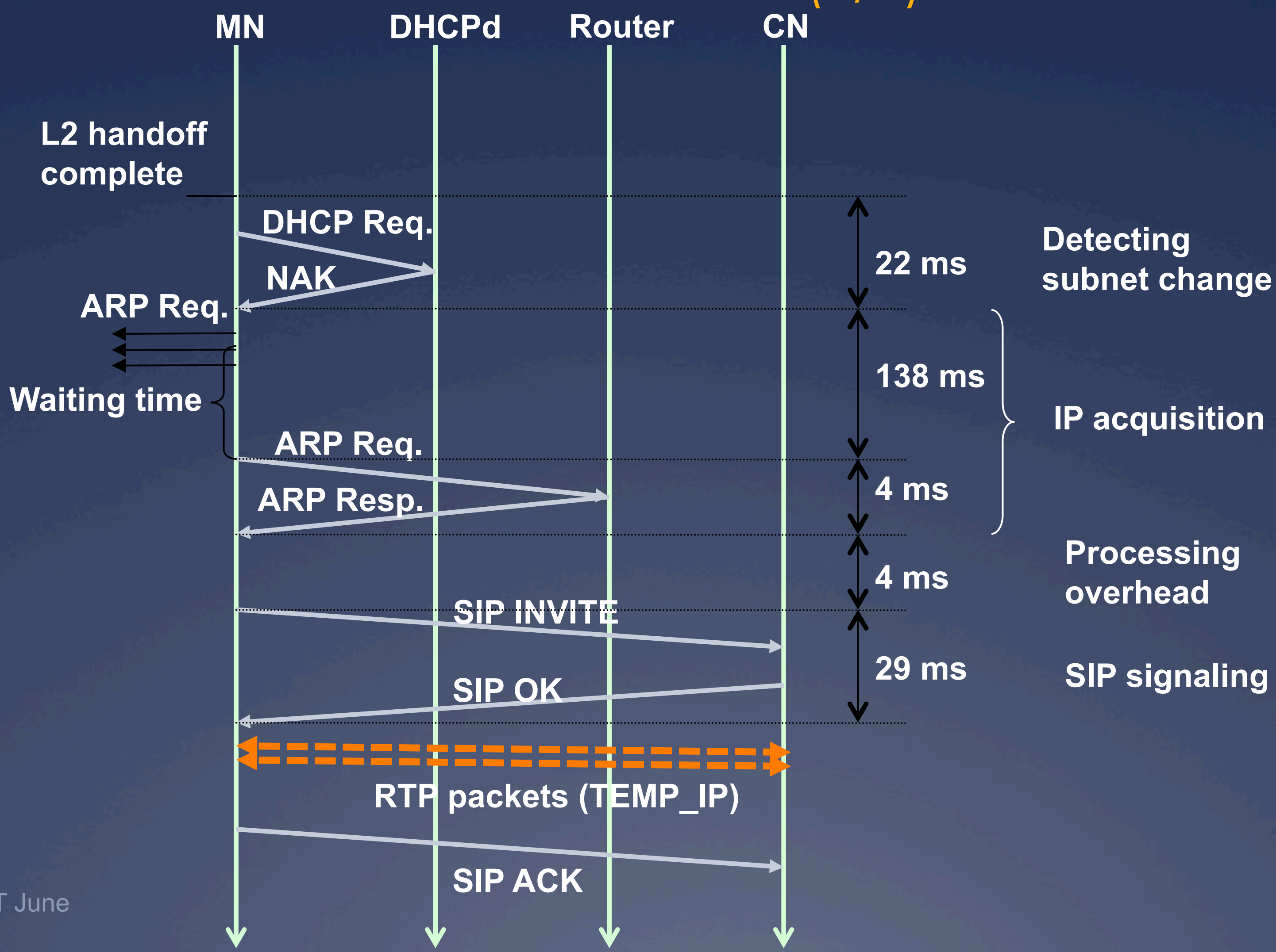
Fast Layer 3 Handoff

TEMP_IP Selection

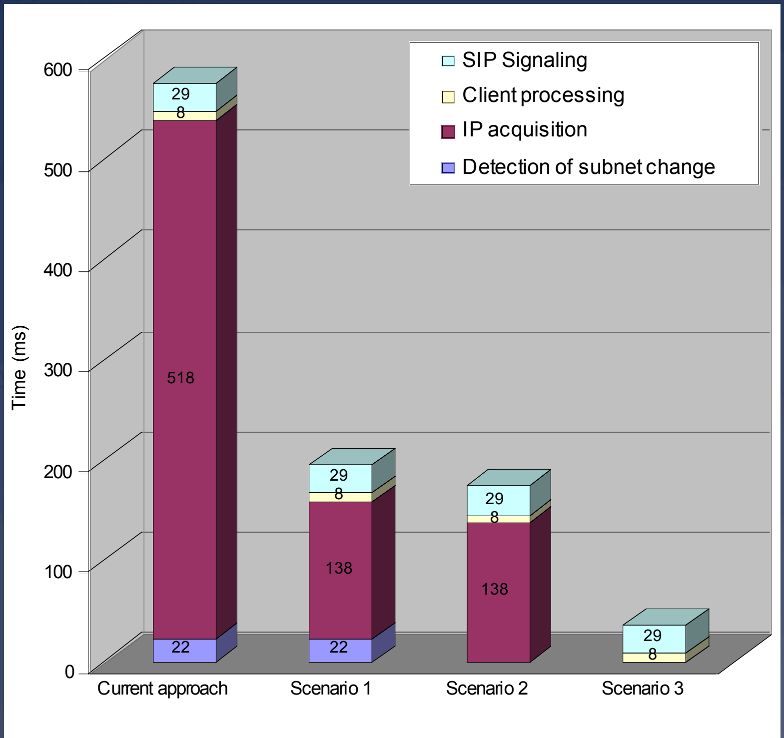
- Roaming to a new subnet
  - Select random IP address starting from the router's IP address (first in the pool). MN sends 10 ARP requests in parallel starting from the random IP selected before.
- Roaming to a known subnet (expired lease)
  - MN starts to send ARP requests to 10 IP addresses in parallel, starting from the IP it last used in that subnet.
- Critical factor: time to wait for an ARP response.
  - Too small → higher probability for a duplicate IP
  - Too big → increases total handoff time
- TEMP_IP: for ongoing sessions only
- Only MN and CN are aware of the TEMP_IP

Fast Layer 3 Handoff

Measurement Results (1/2)



Fast Layer 3 Handoff
Measurement Results (2/2)



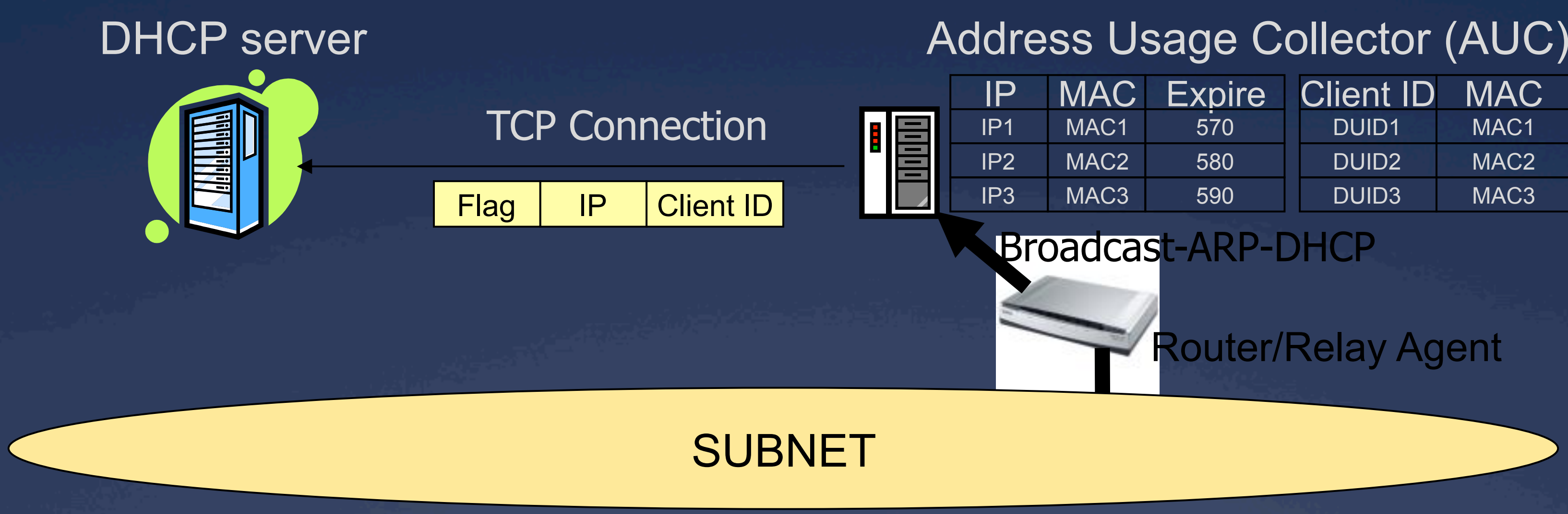
- Scenario 1
  - The MN enters in a new subnet for the first time ever
- Scenario 2
  - The MN enters in a new subnet it has been before and it has an expired lease for that subnet
- Scenario 3
  - The MN enters in a new subnet it has been before and still has a valid lease for that subnet

Fast Layer 3 Handoff

Conclusions

- Modifications in client side only (requirement)
  - Forced us to introduce some limitations in our approach
    Works today, in any network
- Much faster than DHCP although not always fast enough for real-time media (scenarios 1 and 2)
- Scenario 3 obvious but … Windows XP
- ARP timeout → critical factor → SIP presence
- SIP presence approach (Network support)
  - Other stations in the new subnet can send ARP requests on behalf of the MN and see if an IP address is used or not. The MN can wait for an ARP response as long as needed since it is still in the old subnet.

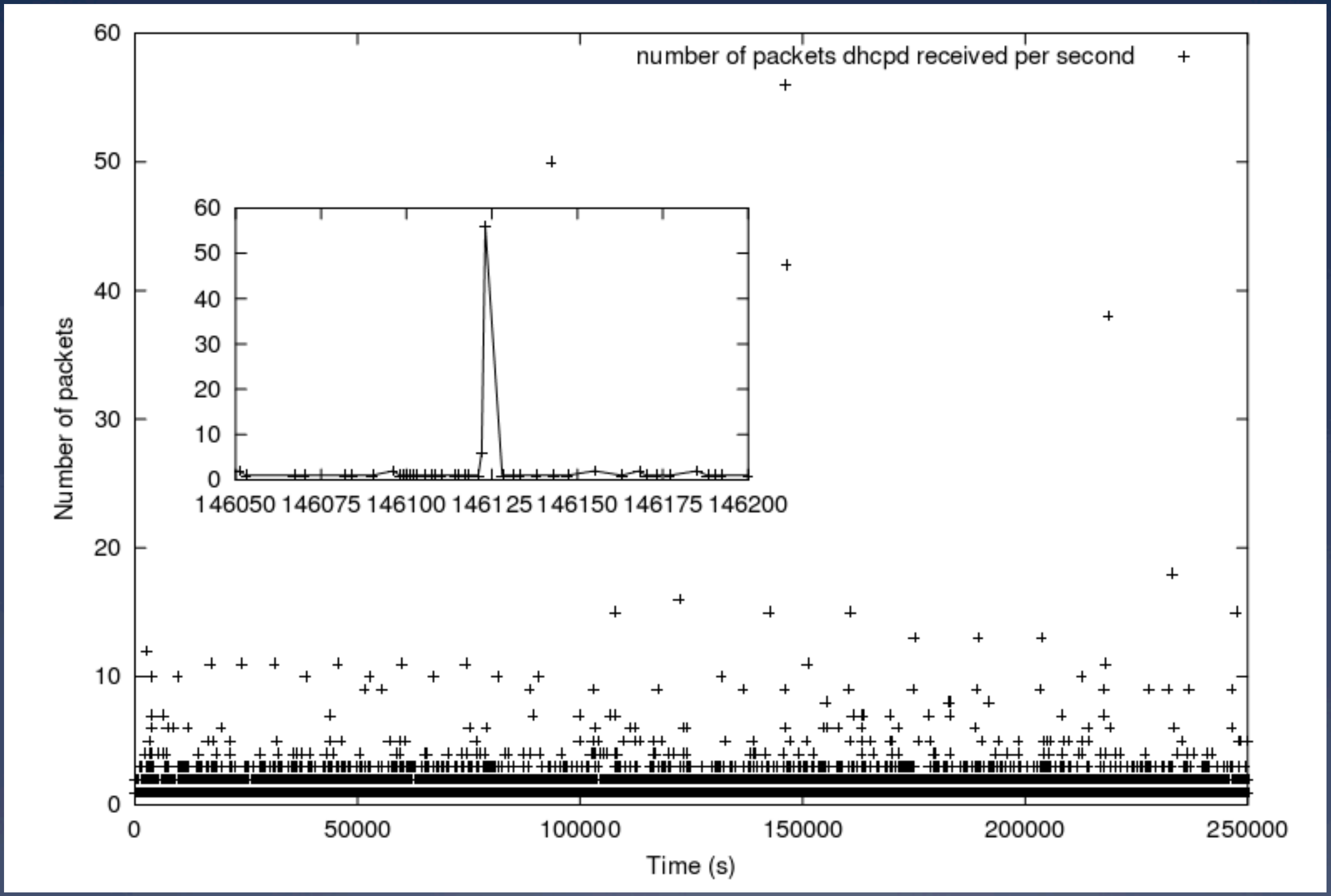
Passive DAD
Overview



- AUC builds DUID:MAC pair table (DHCP traffic only)
- AUC builds IP:MAC pair table (broadcast and ARP traffic)
- The AUC sends a packet to the DHCP server when:
  - a new pair IP:MAC is added to the table
  - a potential duplicate address has been detected
  - a potential unauthorized IP has been detected
- DHCP server checks if the pair is correct or not and it records the IP address as in use. (DHCP has the final decision!)

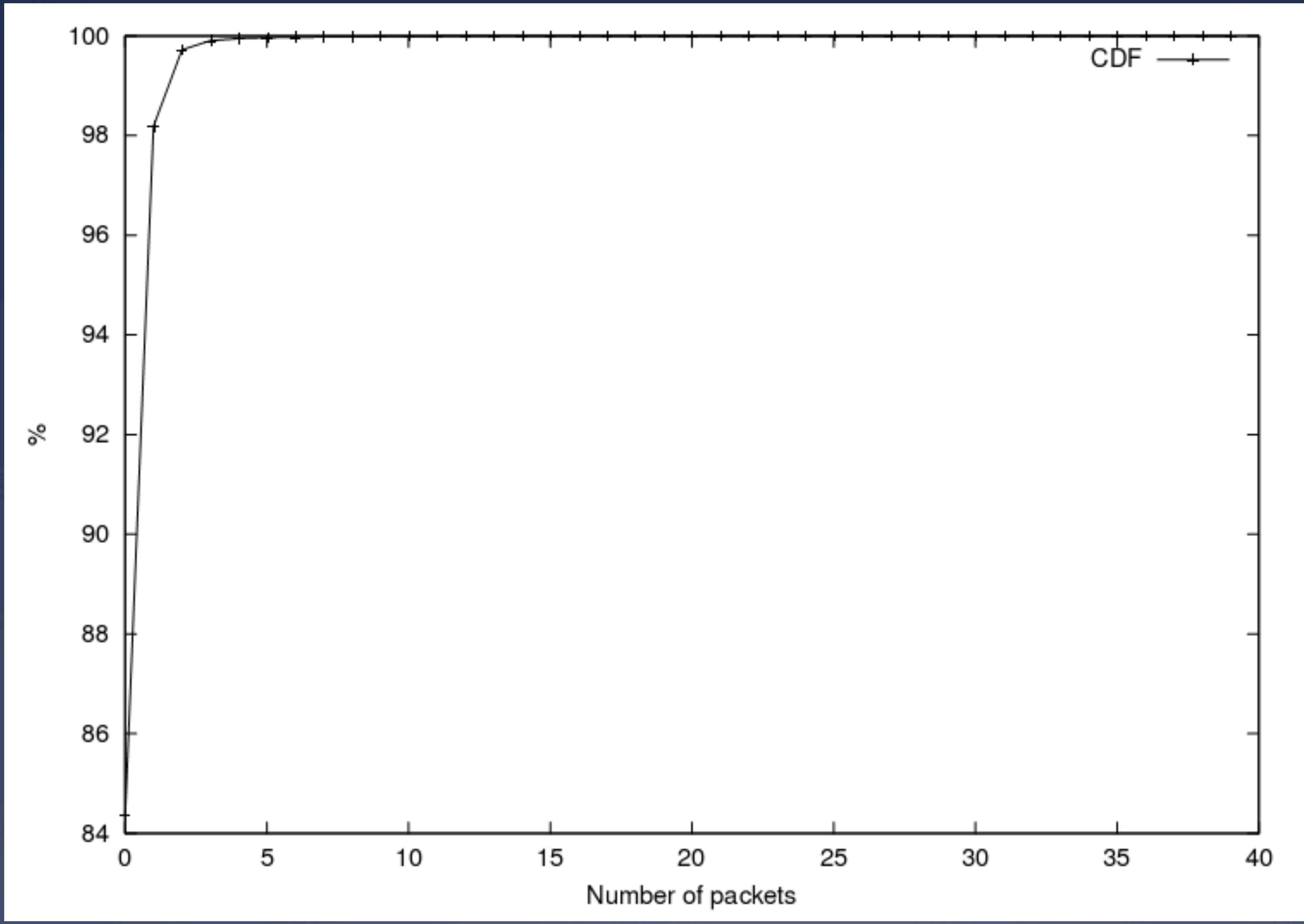
Passive DAD

Traffic load – AUC and DHCP



Passive DAD

Packets/sec received by DHCP

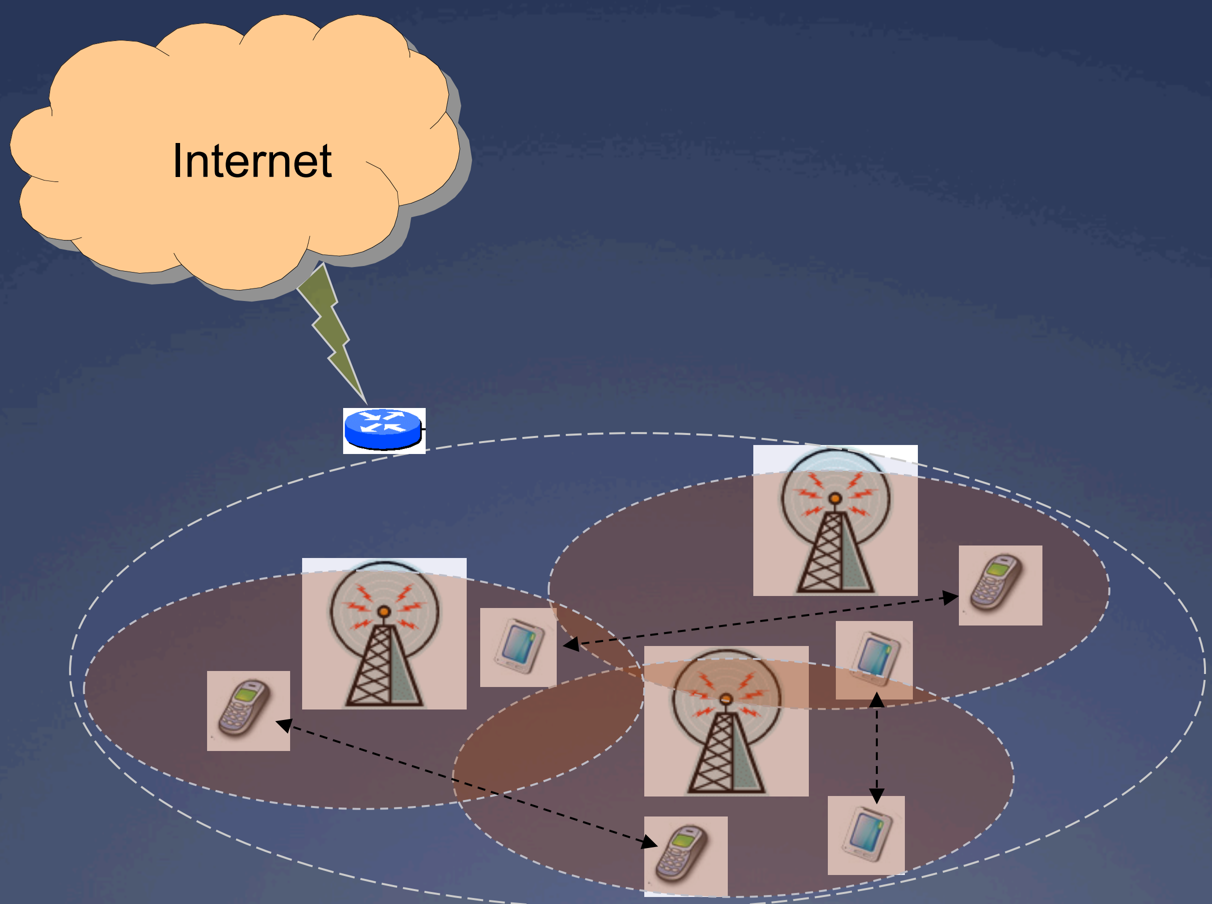


Passive DAD

Conclusions

- pDAD is not performed during IP address acquisition
  - Low delay for mobile devices
- Much more reliable than current DAD
  - Current DAD is based on ICMP echo request/response
    - not adequate for real-time traffic (seconds - too slow!)
    - most firewalls today block incoming echo requests by default
  - A duplicate address can be discovered in real-time and not only if a station requests that particular IP address
  - A duplicate address can be resolved (i.e. FORCE_RENEW)
- Intrusion detection …
  - Unauthorized IPs are easily detected

Cooperation Between Stations in Wireless Networks



Cooperative Roaming

Goals and Solution

* Fast handoff for real-time multimedia in any network
  * Different administrative domains
  * Various authentication mechanisms
  * No changes to protocol and infrastructure
  * Fast handoff at all the layers relevant to mobility
    * Link layer
    * Network layer
    * Application layer

* New protocol → Cooperative Roaming
  * Complete solution to mobility for real-time traffic in wireless networks
  * Working implementation available

Cooperative Roaming

Why Cooperation ?

- Same tasks
  - Lay
  - Lay
  - Aut
  - Mul sess

- Same information
  - To
  - D
  - G
  - Se

- Same goals
  - Low latency
  - QoS
  - Load balancing
  - Admission and congestion control
  - Service discovery

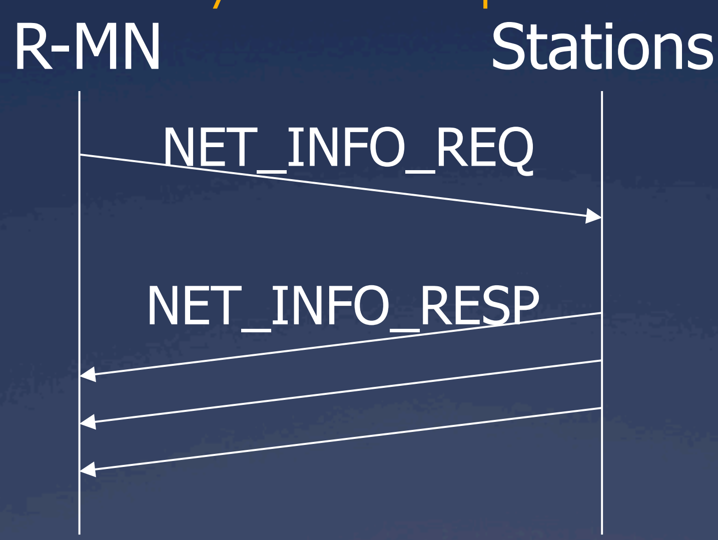
Cooperative Roaming

Overview

- ✓ Stations can cooperate and share information about the network (topology, services)
- ✓ Stations can cooperate and help each other in common tasks such as IP address acquisition
- ✓ Stations can help each other during the authentication process without sharing sensitive information, maintaining privacy and security
- ✓ Stations can also cooperate for application-layer mobility and load balancing

Cooperative Roaming

Layer 2 Cooperation



- Random waiting time
  - Stations will not send the same information and will not send all at the same time
- The information exchanged in the NET_INFO multicast frames is:

APs {BSSID, Channel}
SUBNET IDs

Cooperative Roaming

Layer 3 Cooperation

- Subnet detection
  - Information exchanged in NET_INFO frames (Subnet ID)
- IP address acquisition time
  - Other stations (STAs) can cooperate with us and acquire a new IP address for the new subnet on our behalf while we are still in the OLD subnet

    → Not delay sensitive!

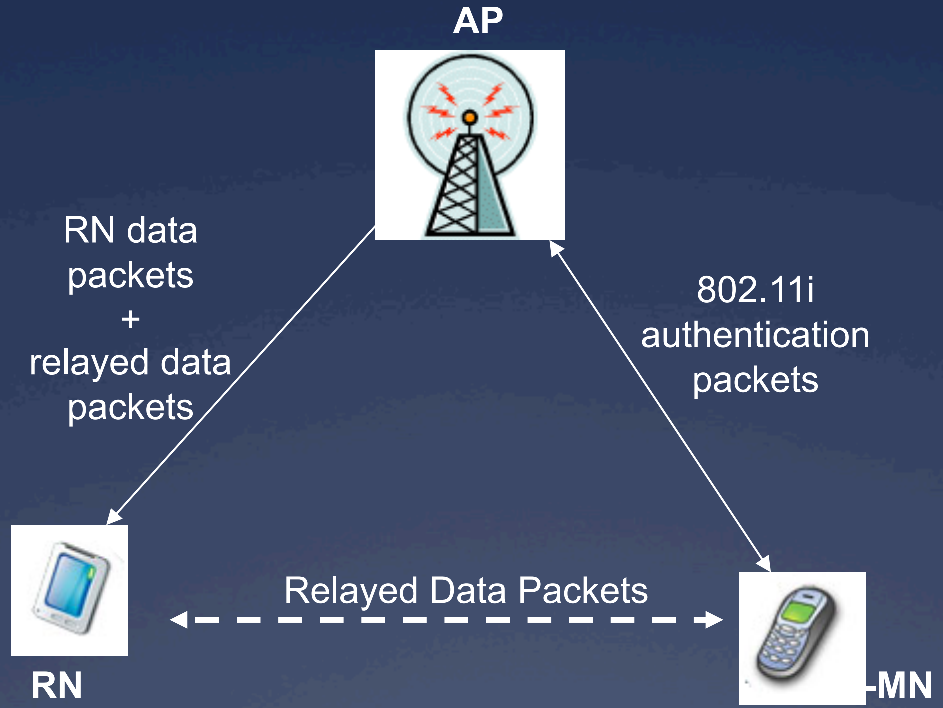
Cooperative Roaming

Cooperative Authentication (1/2)

- Cooperation in the authentication process itself is not possible as sensitive information such as certificates and keys are exchanged.
- STAs can still cooperate in a mobile scenario to achieve a seamless L2 and L3 handoff regardless of the particular authentication mechanism used.
  - In IEEE 802.11 networks the medium is "shared".
    - Each STA can hear the traffic of other STAs if on the same channel.
  - Packets sent by the non-authenticated STA will be dropped by the infrastructure but will be heard by the other STAs on the same channel/AP.

Cooperative Roaming

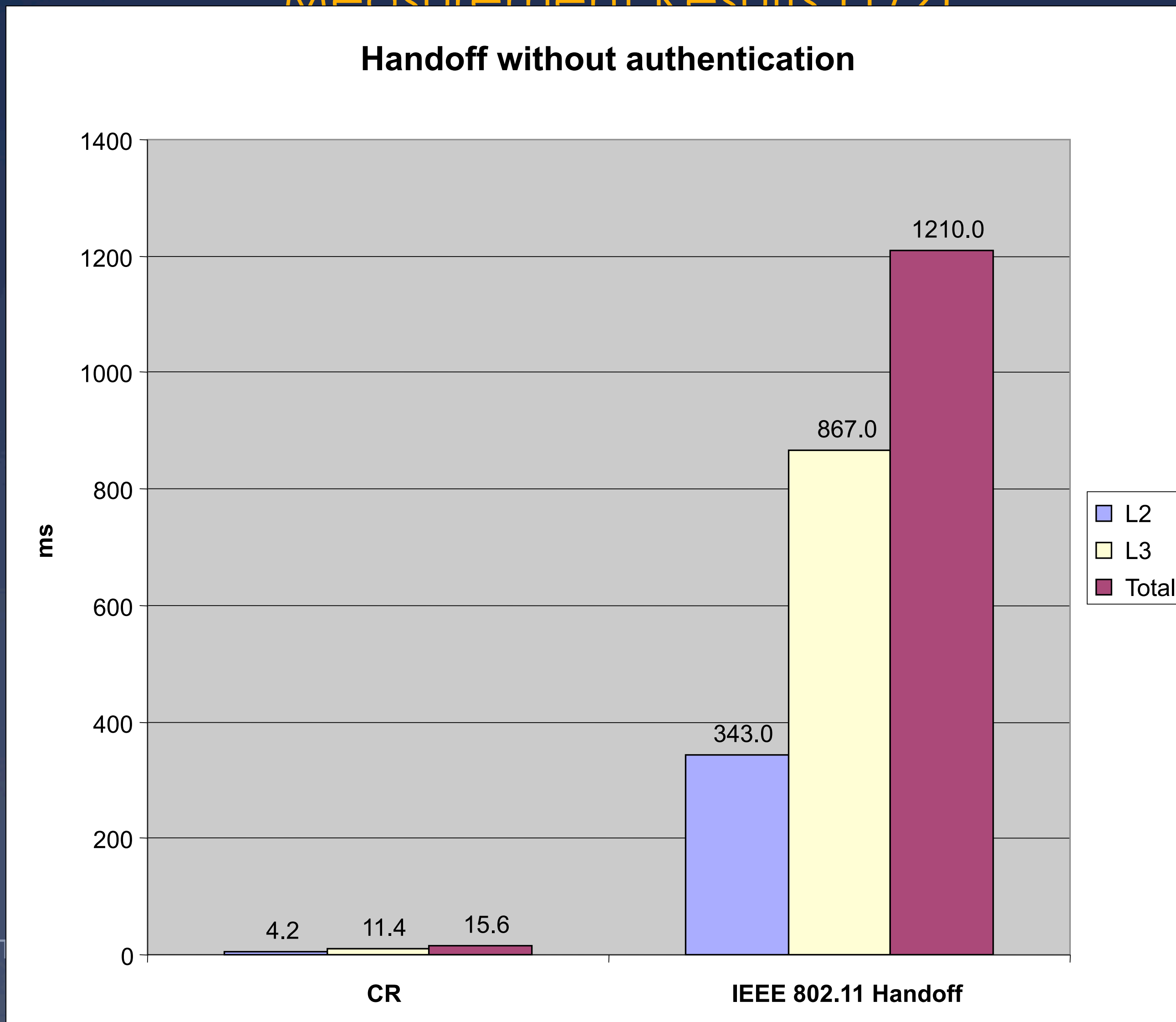
Cooperative Authentication (2/2)



* One selected STA (RN) can relay packets to and from the R-MN for the amount of time required by the R-MN to complete the authentication process.

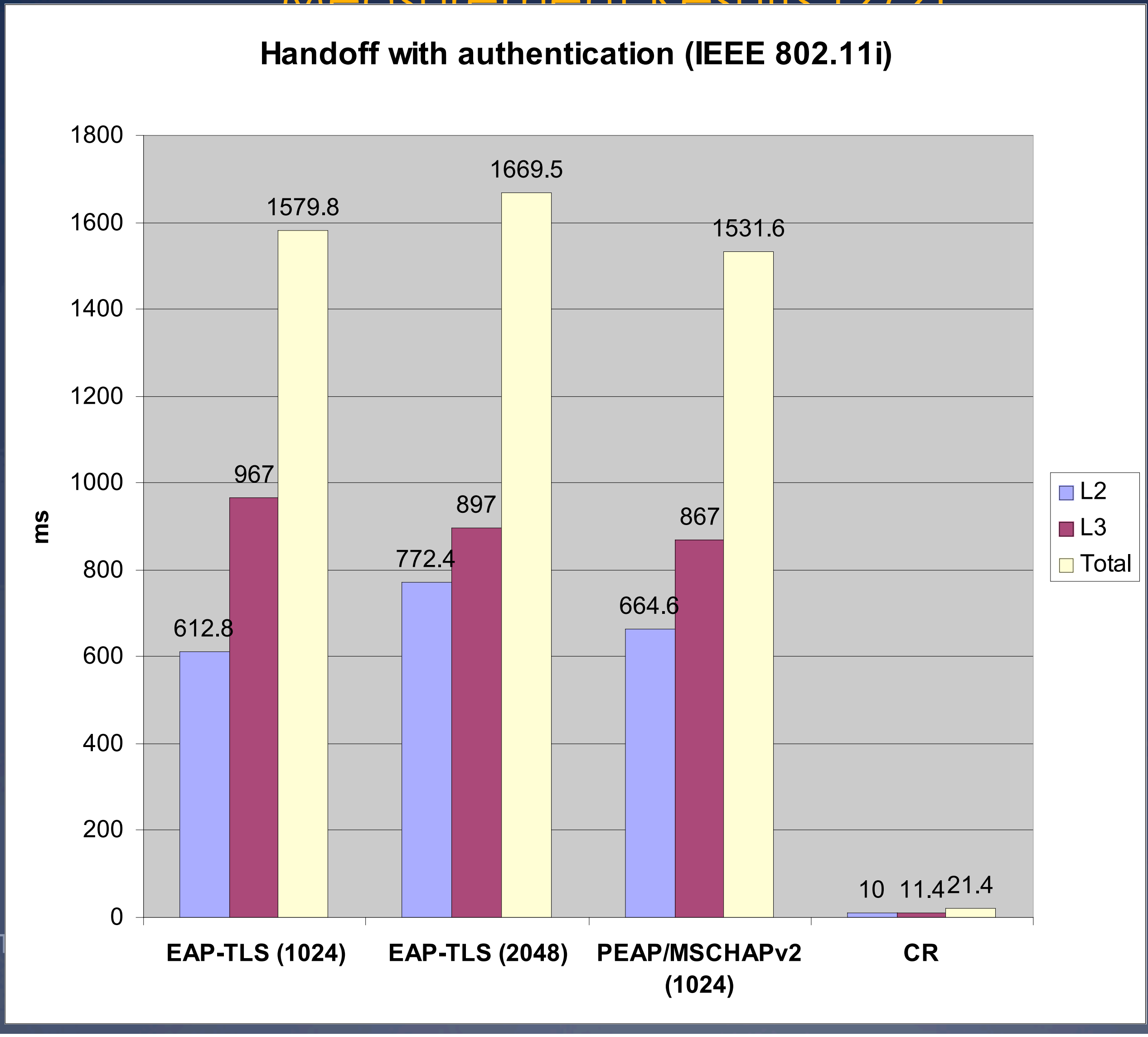
Cooperative Roaming

Measurement Results (1/2)



Cooperative Roaming

Measurement Results (2/2)



Cooperative Roaming

Application Layer Handoff - Problems

- SIP handshake
  - INVITE → 200 OK → ACK
    (Few hundred milliseconds)
- User's direction (next AP/subnet)
  - Not known before a L2 handoff
  - MN not moving after all

Cooperative Roaming

Application Layer Handoff - Solution

* MN builds a list of {RNs, IP addresses}, one per each possible next subnet/AP
* RFC 3388
  * Send same media stream to multiple clients
  * All clients have to support the same codec
* Update multimedia session
  * Before L2 handoff
    * Media stream is sent to all RNs in the list and to MN (at the same time) using a re-INVITE with SDP as in RFC 3388
    * RNs do not play such streams (virtually support any codec)
  * After L2 handoff
    * Tell CN which RN to use, if any (re-INVITE)
    * After successful L2 authentication tell CN to send directly without any RN (re-INVITE)
* No buffering necessary
  * Handoff time: 15ms (open), 21ms (802.11i)
  * Packet loss negligible

Cooperative Roaming
Other Applications

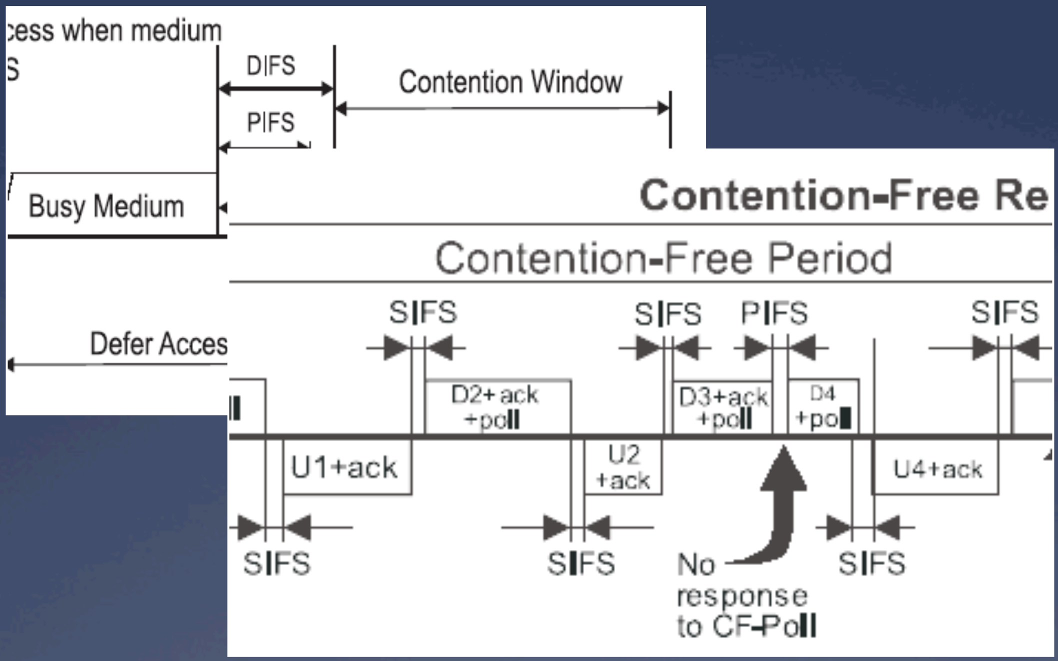
- In a multi-domain environment Cooperative Roaming (CR) can help with choosing AP/domain according to roaming agreements, billing, etc.
- CR can help for admission control and load balancing, by redirecting MNs to different APs and/or different networks. (Based on real throughput)
- CR can help in discovering services (encryption, authentication, bit-rate, Bluetooth, UWB, 3G)
- CR can provide adaptation to changes in the network topology (common with IEEE 802.11h equipment)
- CR can help in the interaction between nodes in infrastructure and ad-hoc/mesh networks

Cooperative Roaming

Conclusions

- Cooperation among stations allows seamless L2 and L3 handoffs for real-time applications (10-15 ms HO)
- Completely independent from the authentication mechanism used
- It does not require any changes in either the infrastructure or the protocol
- It does require many STAs supporting the protocol and a sufficient degree of mobility
- Suitable for indoor and outdoor environments
- Sharing information → Power efficient

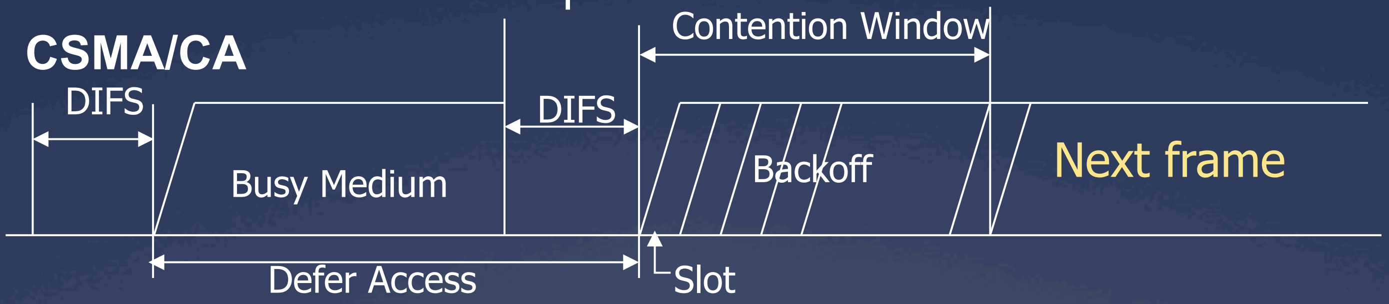
Improving Capacity of VoIP in IEEE 802.11 Networks using Dynamic PCF (DPCF)



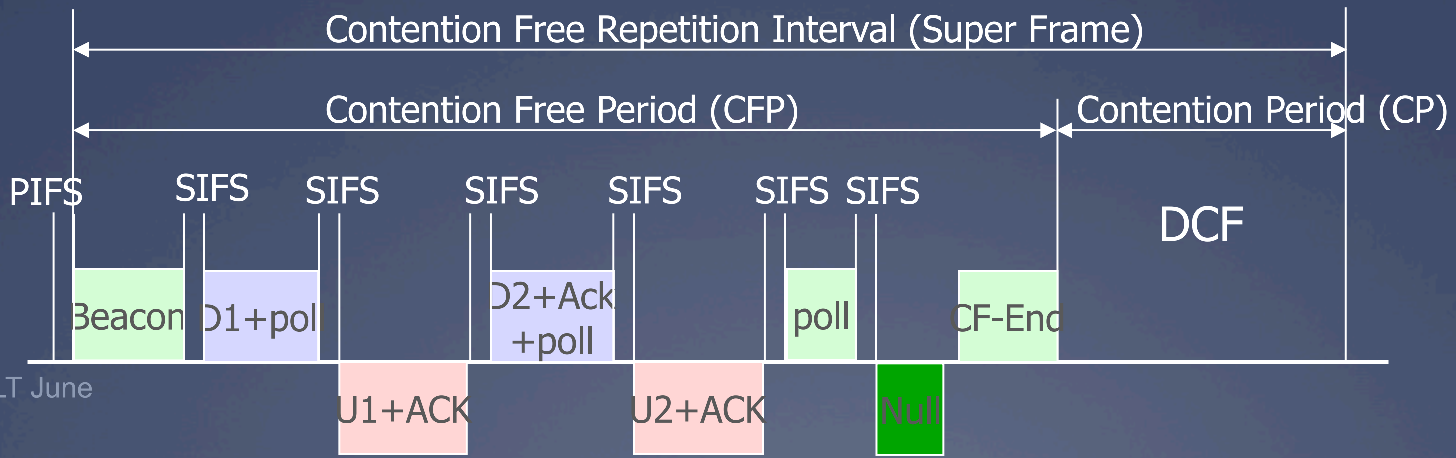
Dynamic PCF (DPCF)

MAC Protocol in IEEE 802.11

- Distributed Coordination Function (DCF)
  - Default MAC protocol



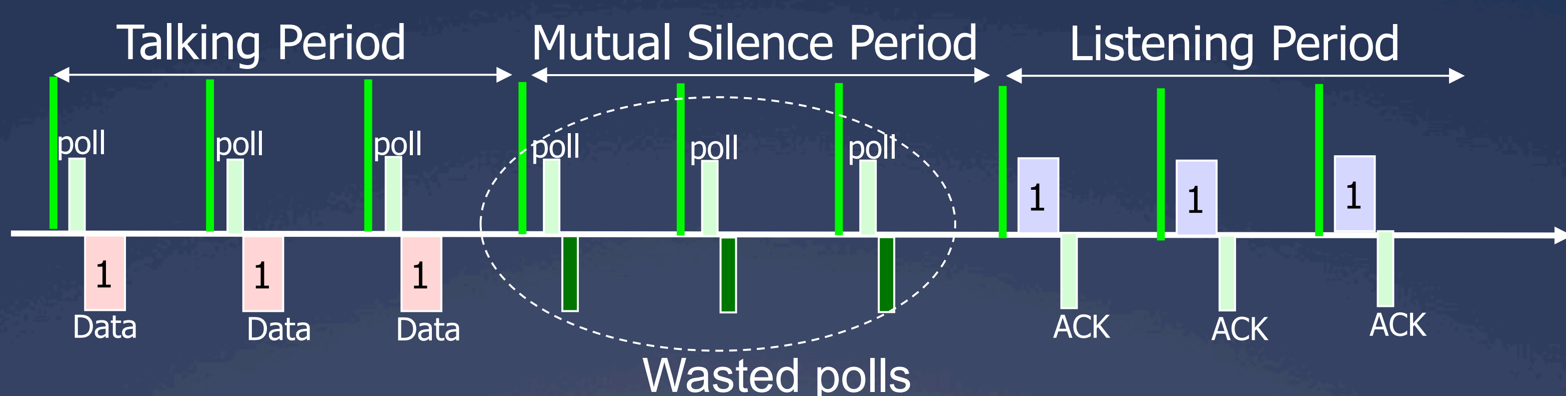
- Point Coordination Function (PCF)
  - Supports rudimentary QoS, not implemented



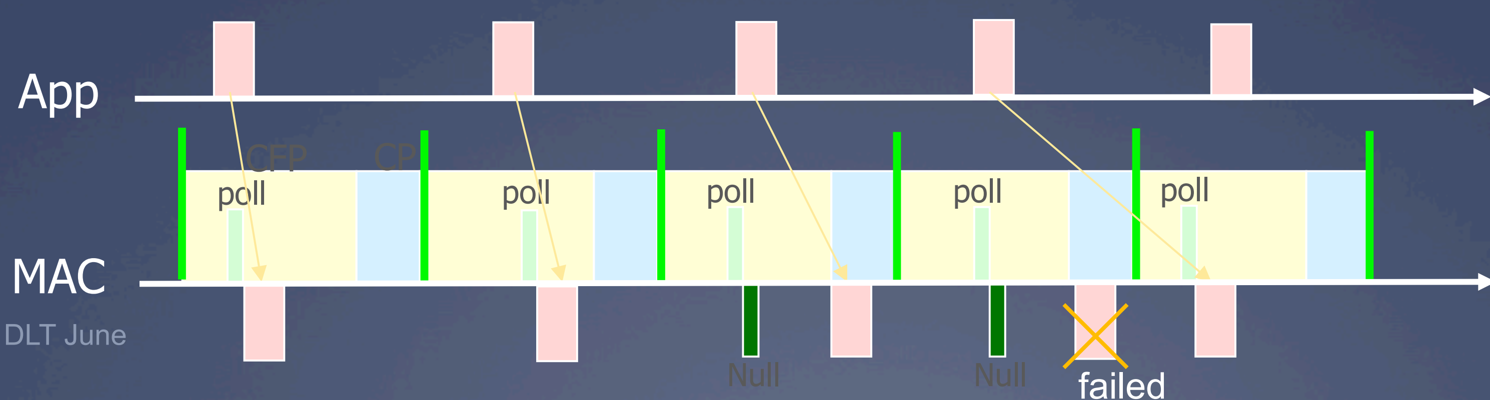
Dynamic PCF (DPCF)

Problems of PCF

- Waste of polls
  - VoIP traffic with silence suppression



  - Synchronization between polls and VoIP packets



Dynamic PCF (DPCF)

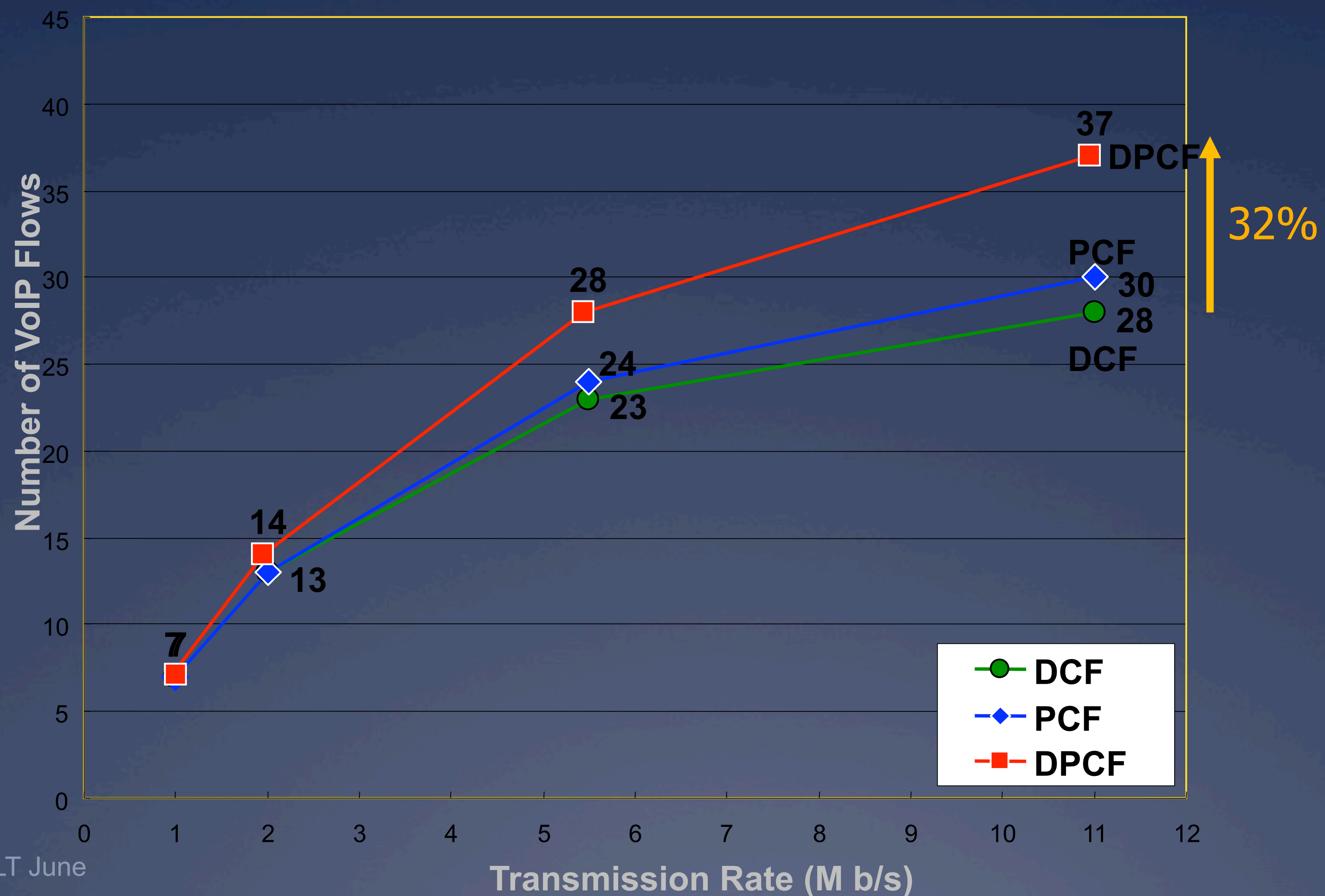
Overview

* Classification of traffic
  * Real-time traffic (VoIP) uses CFP, also CP
  * Best effort traffic uses only CP
  * Give higher priority to real-time traffic
* Dynamic polling list
  * Store only "active" nodes
* Dynamic CFP interval and More data field
  * Use the biggest packetization interval as a CFP interval
  * STAs set "more data field" (a control field in MAC header) of uplink VoIP packets when there are more than two packets to send → AP polls the STA again
  * Solution to the various packetization intervals problem
* Solution to the synchronization problem
  * Allow VoIP packets to be sent in CP only when there are more than two VoIP packets in queue

Dynamic PCF (DPCF)

Simulation Results (1/2)

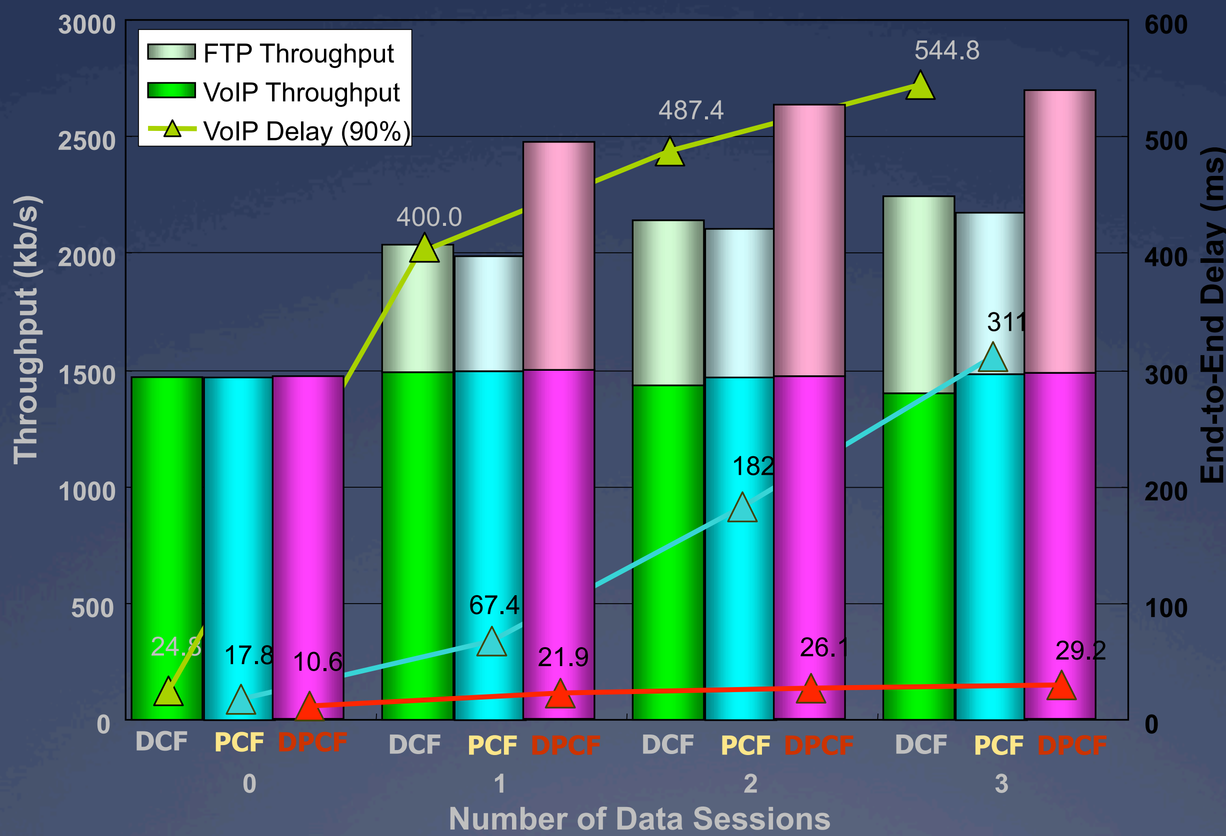
Capacity for VoIP in IEEE 802.11b



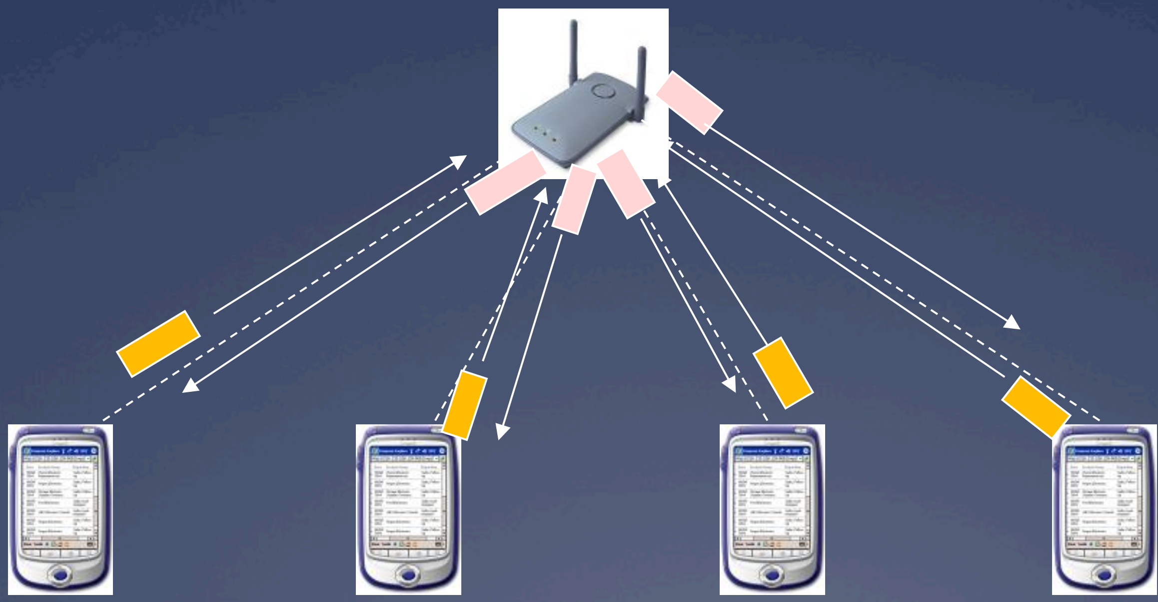
Dynamic PCF (DPCF)

Simulation Results (2/2)

Delay and throughput of 28 VoIP traffic and data traffic

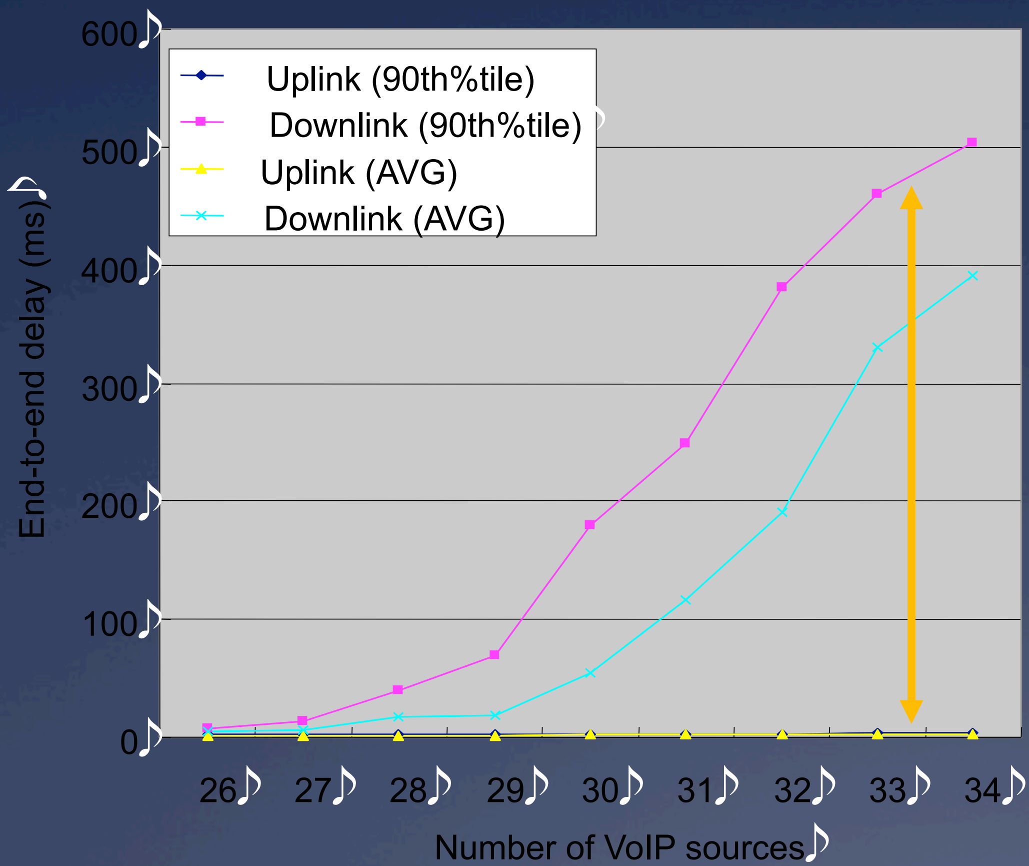


Balancing Uplink and Downlink Delay of VoIP Traffic in 802.11 WLANs using Adaptive Priority Control (APC)



Adaptive Priority Control (APC)

Motivation



20 ms packetization interval (64kb/s)

- Big difference between uplink and downlink delay when channel is congested
- AP has more data, but the same chance to transmit them than nodes

Solution?

- AP needs have higher priority than nodes
- What is the optimal priority and how the priority is applied to the packet scheduling?

Adaptive Priority Control (APC) Overview

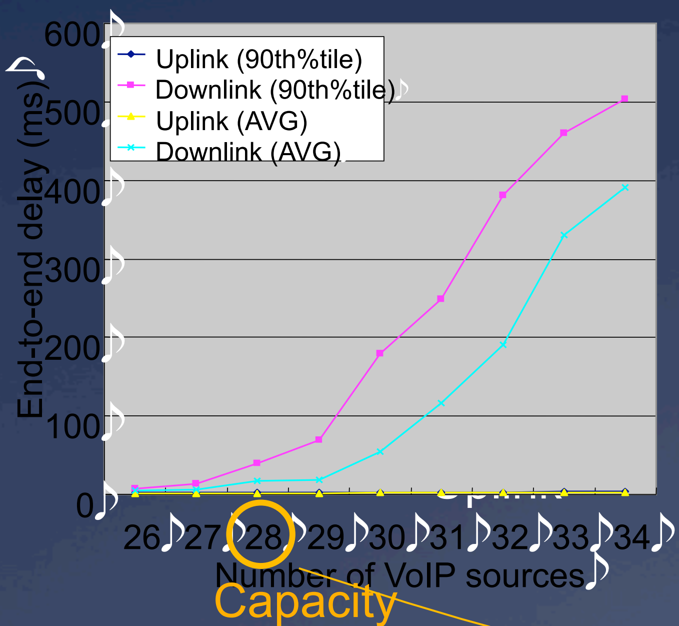
Number of packets in queue of AP

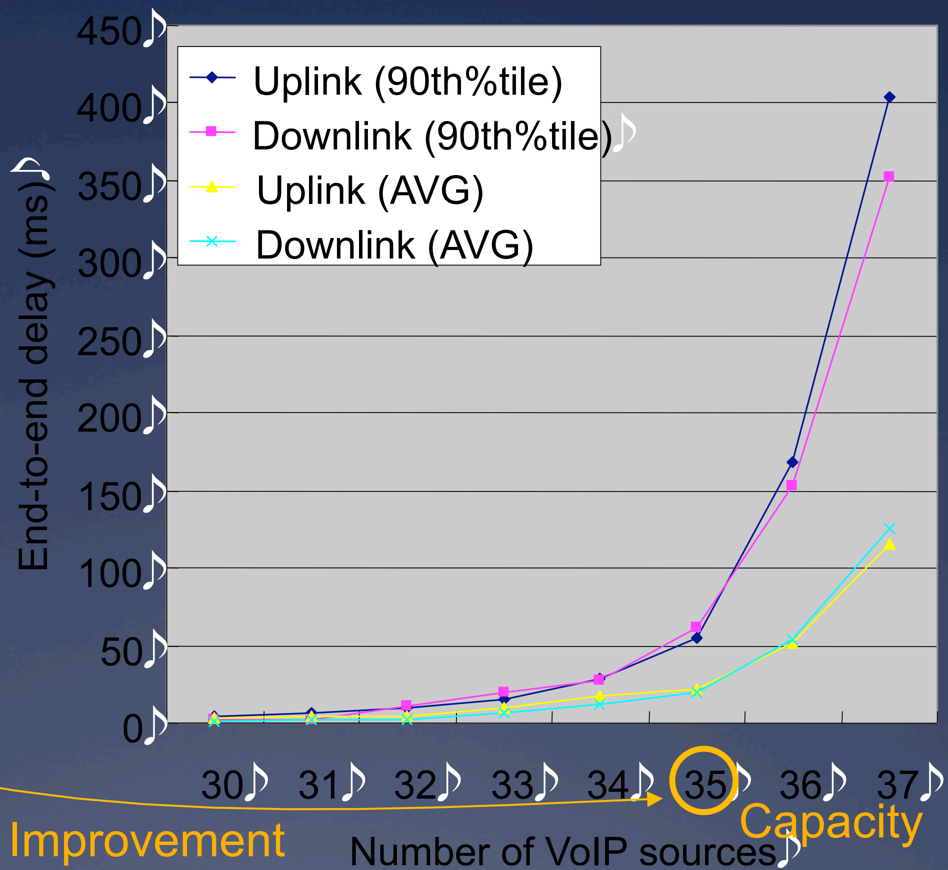
- Optimal priority (*P*) = Q_{AP}/Q_{STA}
  - Simple
  - Adaptive to change of number of active STAs
  - Adaptive to change of uplink/downlink traffic volume

Average number of packets in queue of STAs

- Contention free transmission
  - Transmit *P* packets contention free
  - Precise priority control
    - *P* → Priority
    - Transmitting three frames contention free → three times higher priority than other STAs.
  - No overhead
  - Can be implemented with 802.11e CFB feature

Adaptive Priority Control (APC) Simulation Results





25% Improvement

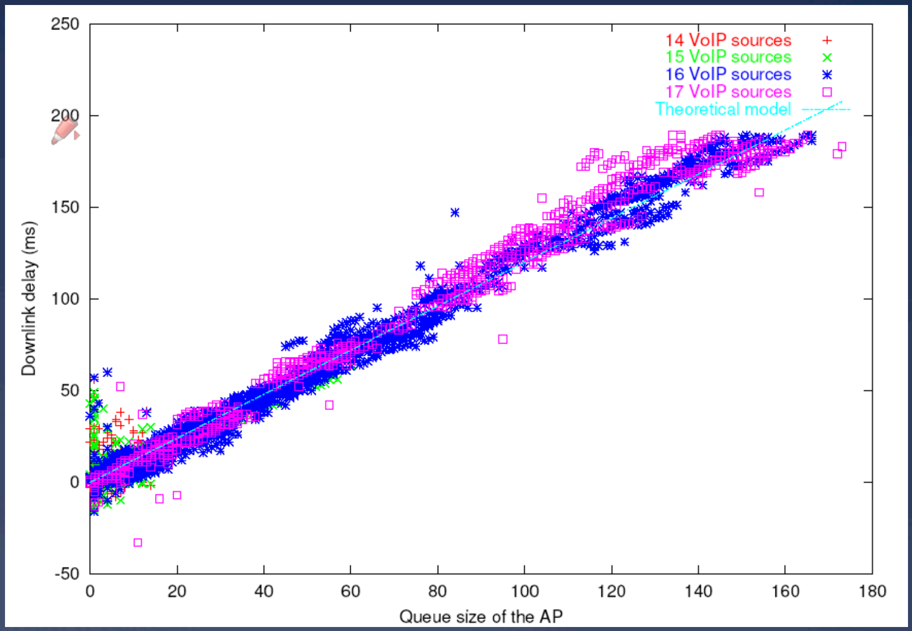
20 ms packetization interval (64kb/s)

Call Admission Control using QP-CAT

Admission Control using QP-CAT

Introduction

- QP-CAT
  - Metric: Queue size of the AP
    - Strong correlation between the queue size of the AP and delay

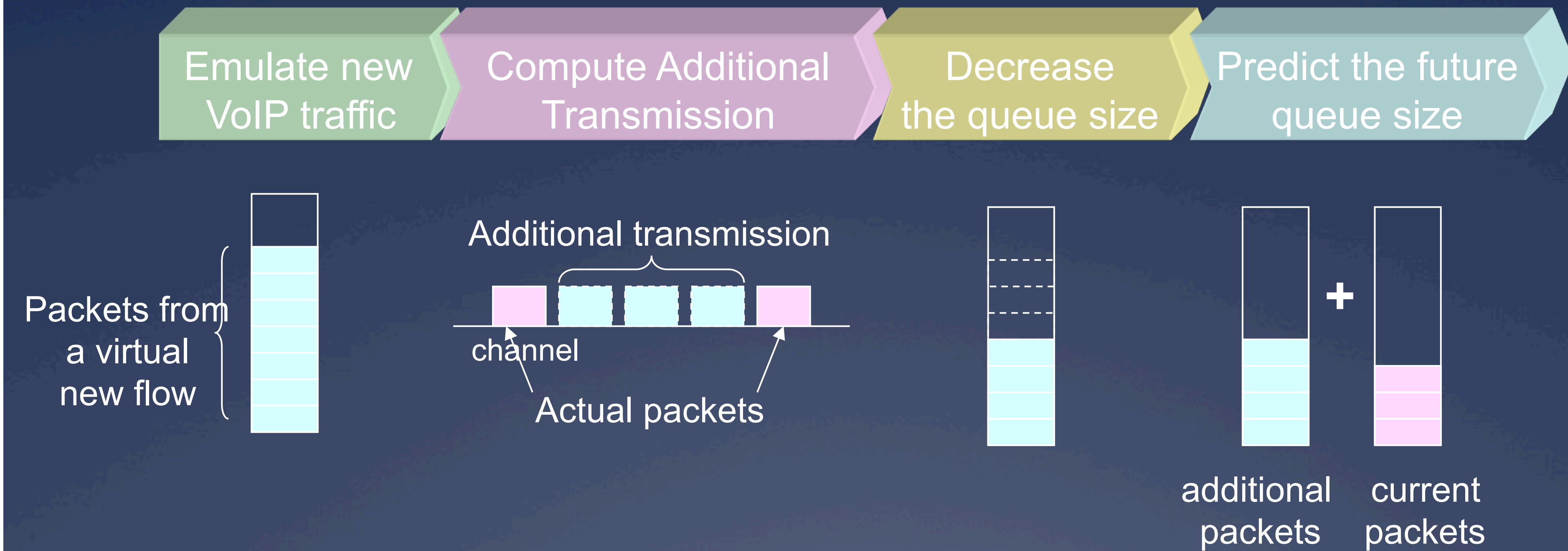


Correlation between queue size of the AP and delay
(Experimental results with 64kb/s VoIP calls)

- Key idea: predict the queue size increase of the AP due to new VoIP flows, by monitoring the current packet transmissions

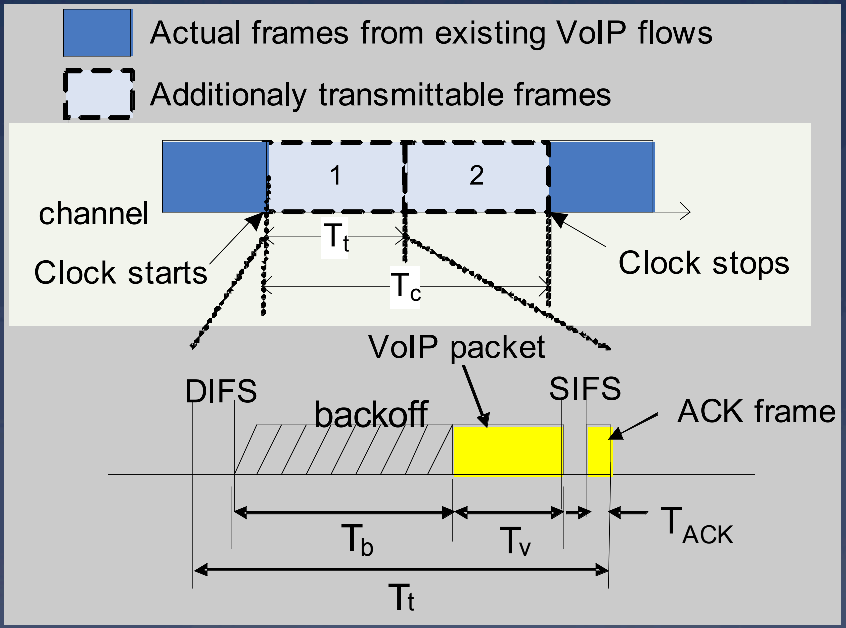
QP-CAT

Basic flow of QP-CAT



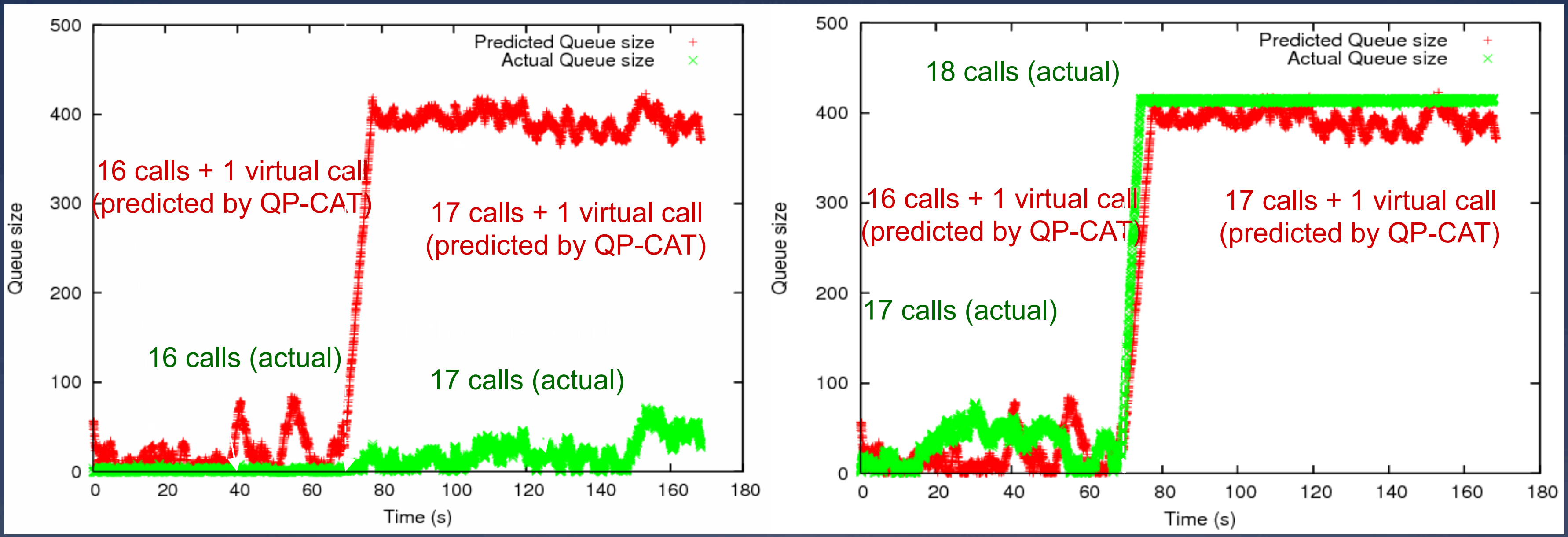
QP-CAT

Computation of Additional Transmission



- Virtual Collision
- Deferrals of virtual packets

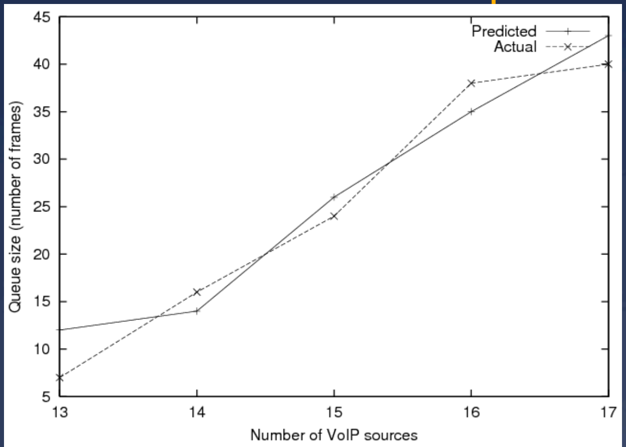
QP-CAT
Simulation results

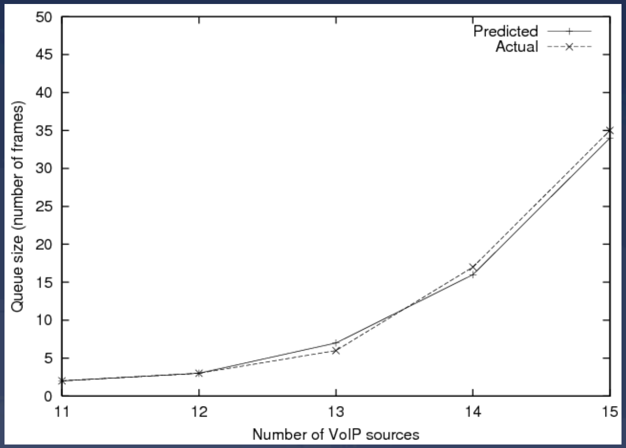


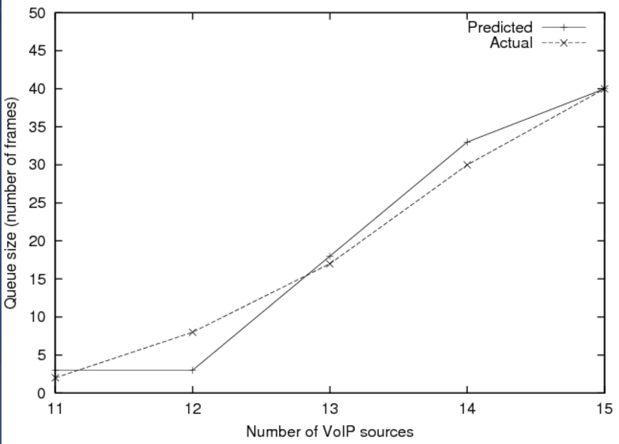
VoIP traffic with 34kb/s 20ms Packetization Interval

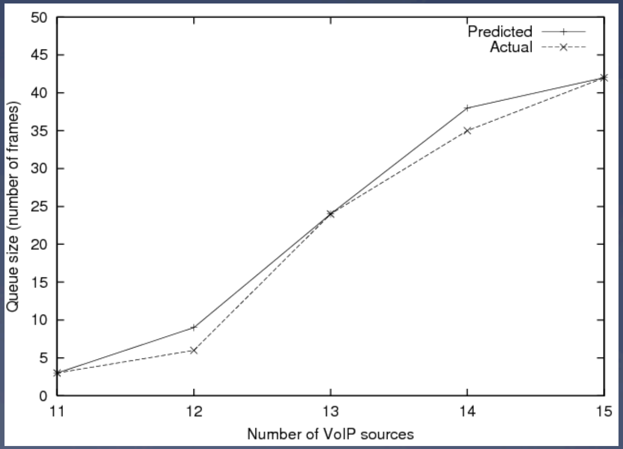
QP-CAT

Experimental results







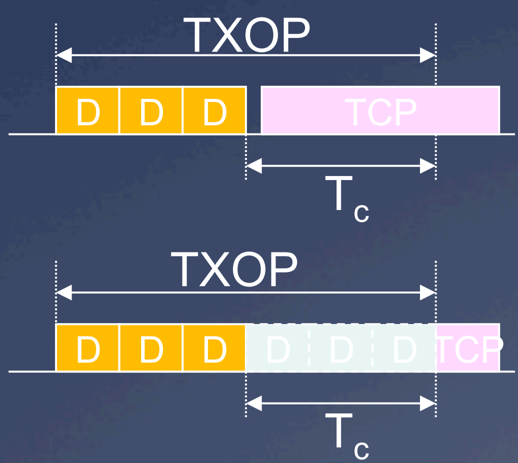


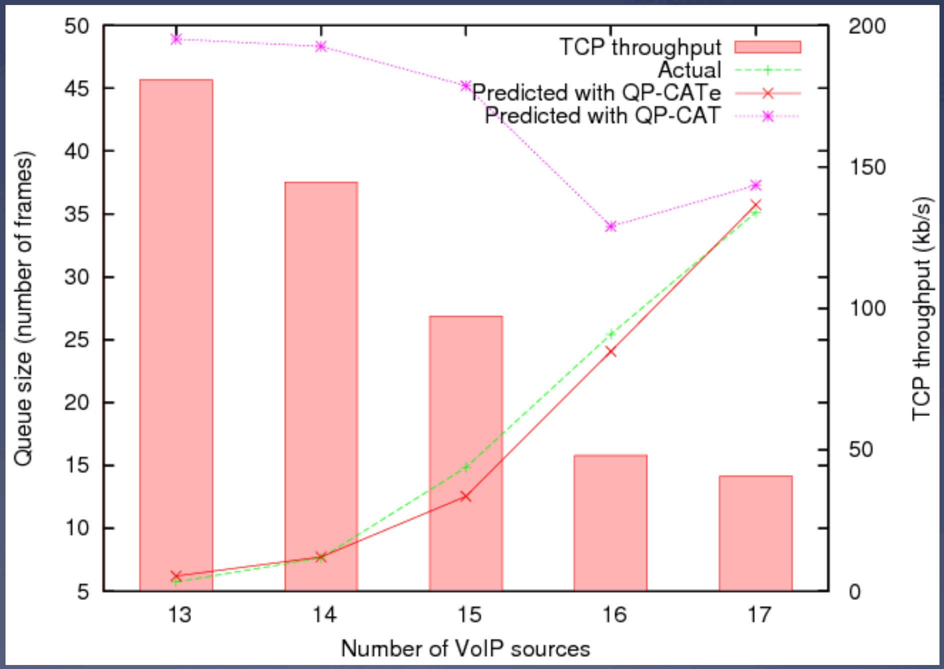
VoIP traffic with 64kb/s 20ms Packetization Interval

QP-CAT

Modification for IEEE 802.11e

- QP-CATe
  - QP-CAT with 802.11e
  - Emulate the transmission during TXOP



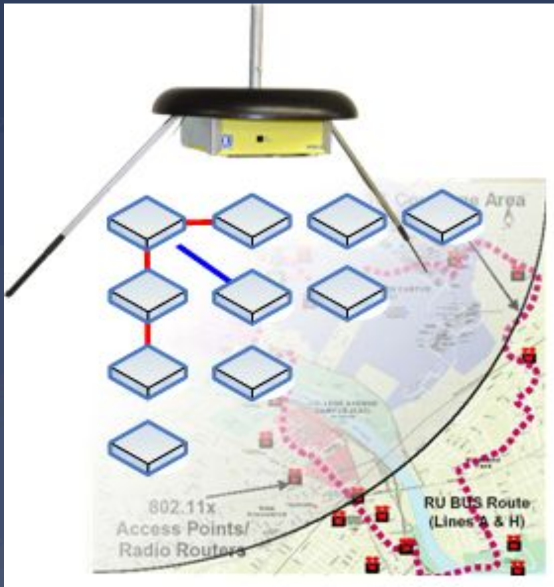


QP-CAT
Conclusions

- What we have addressed
  - Fast handoff
    - Handoffs transparent to real-time traffic
  - Fairness between AP and STAs
    - Fully balanced uplink and downlink delay
  - Capacity improvement for VoIP traffic
    - A 32% improvement of the overall capacity
  - 802.11 networks in congested environments
    - Inefficient algorithms in wireless card drivers
  - Call Admission Control
    - Accurate prediction of impacts of new VoIP calls

- Other problems
  - Handoff between heterogeneous networks

Experimental Capacity Measurement in the ORBIT Testbed



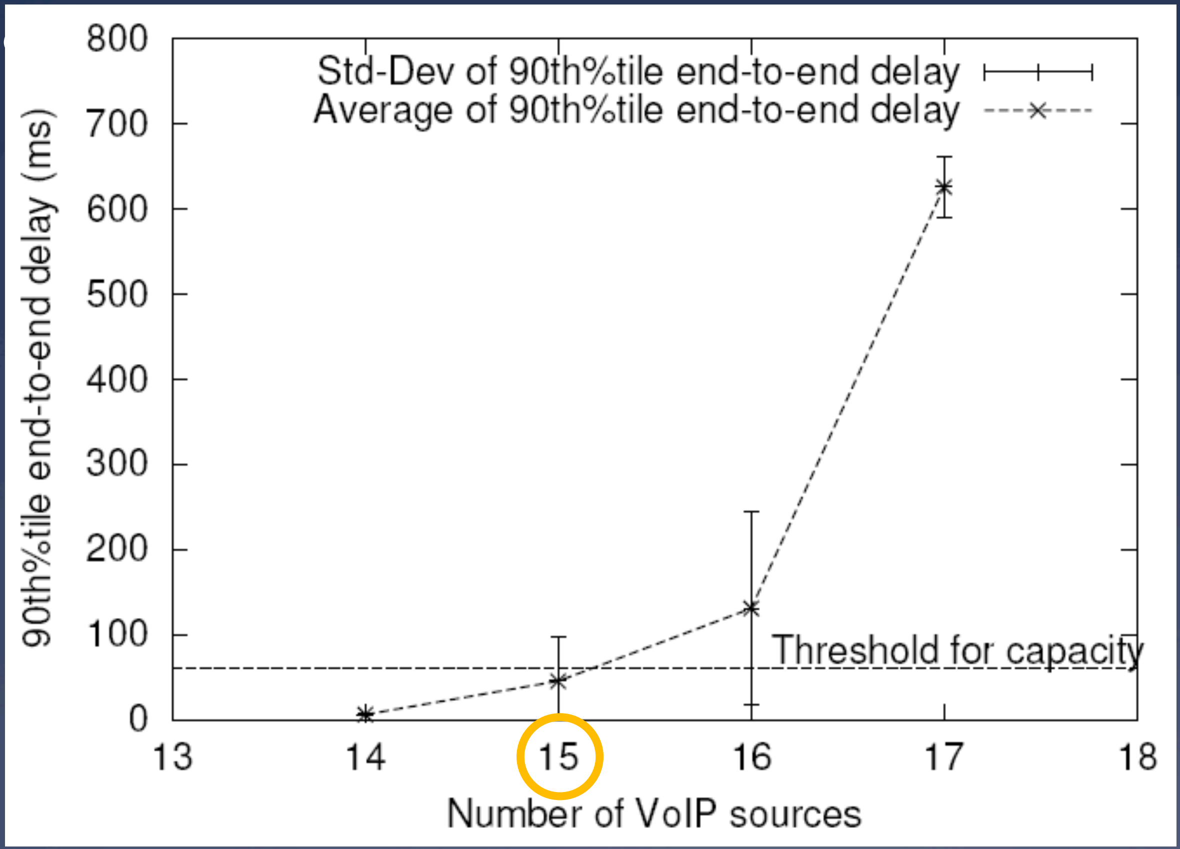
Capacity Measurement

ORBIT test-bed

- Open access research test-bed for next generation wireless networks
- WINLab in Rutgers University in NJ

Capacity Measurement

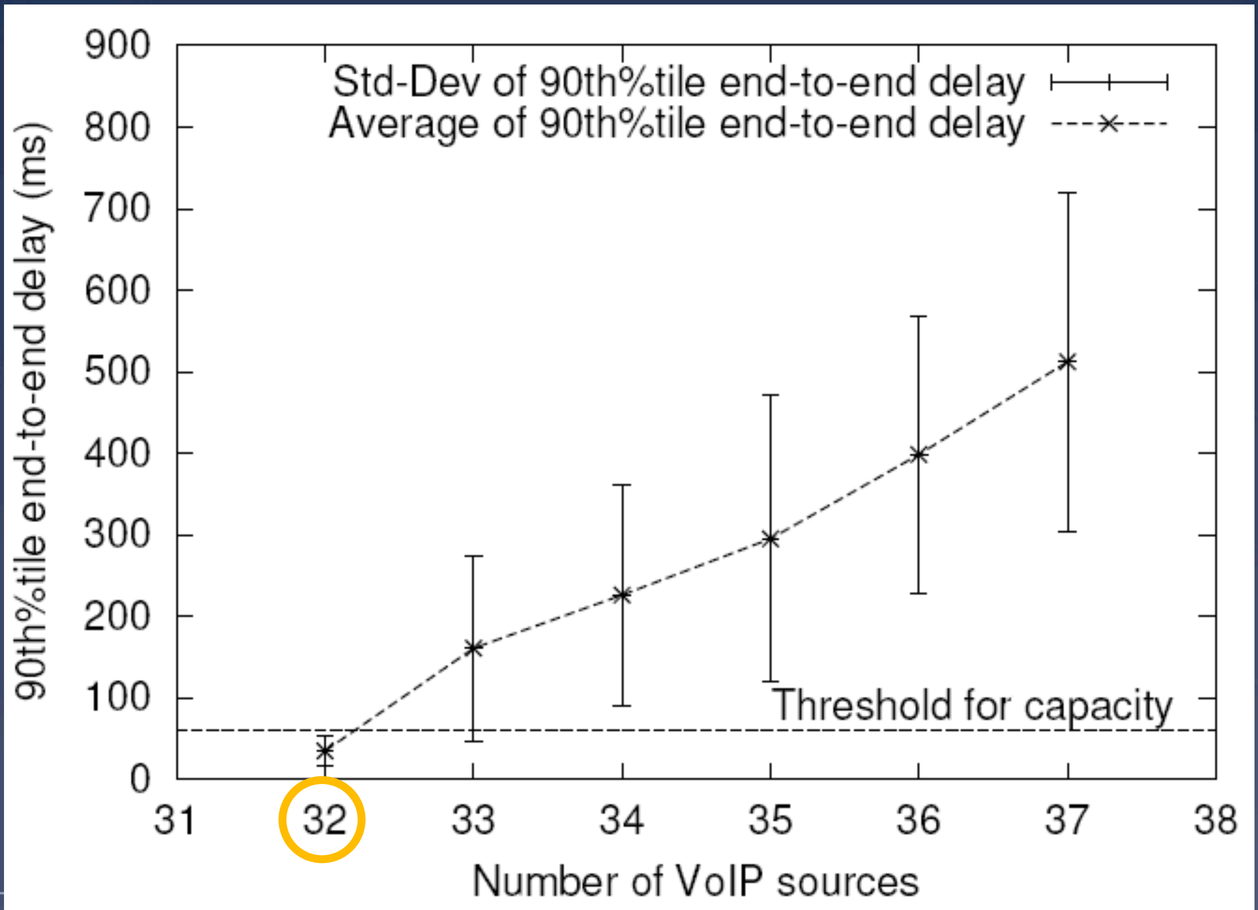
Experimental Results - Capacity of CBR VoIP traffic



Capacity Measurement

Experimental Results - Capacity of VBR VoIP traffic

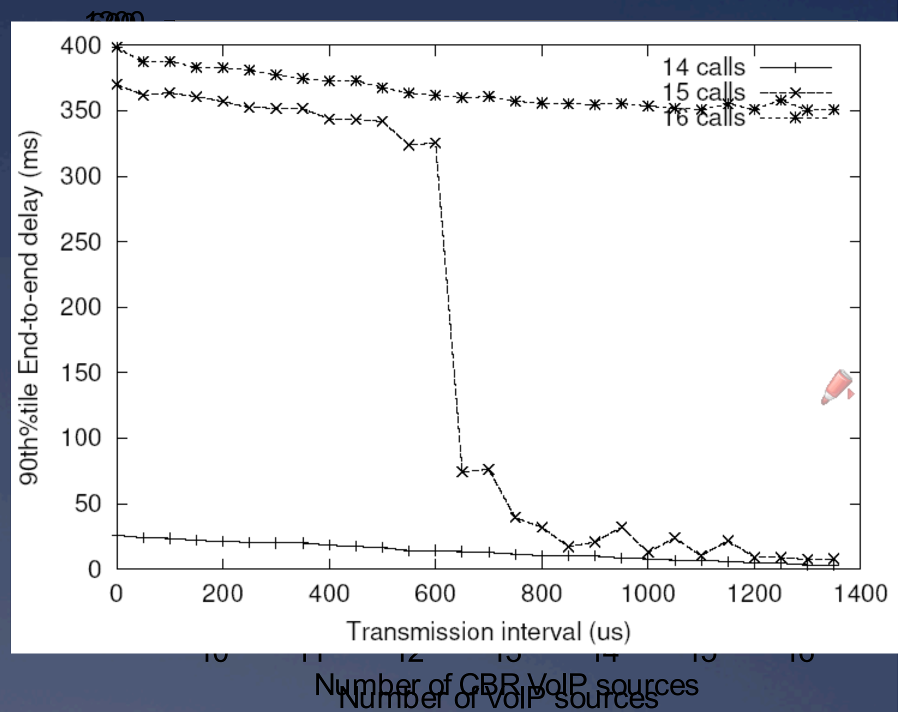
*



Capacity Measurement

Factors that affects the capacity

- Auto Rate Fallback (ARF) algorithms
  - 13 calls (ARF)→ 15 calls (No ARF)
  - Because reducing Tx rate does not help in alleviating congestion
- Preamble size
  - 12 calls (long) → 15 calls (short)
  - Short one is used in wireless cards
- Packet generation intervals among VoIP sources
  - 14 calls → 15 calls
  - In simulation, random intervals needs to be used

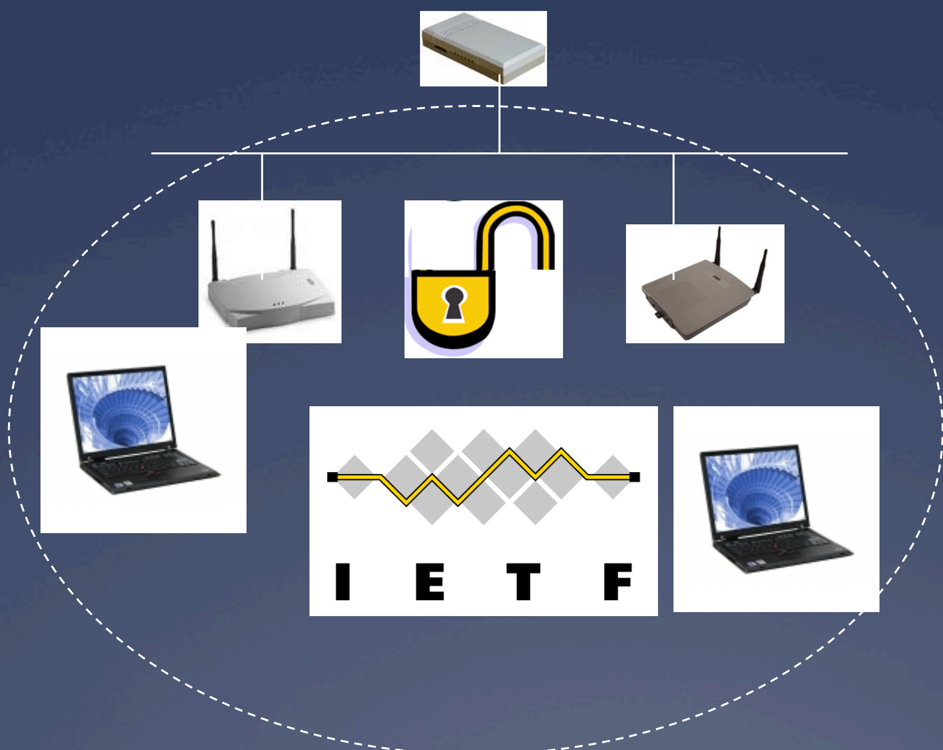


Capacity Measurement

Other factors

* Scanning APs
  * Nodes start to scan APs after experiencing many frame losses
  * Probe request and response frames could congest channels
* Retry limit
  * Retry limit is not standardized and vendors and simulation tools use different values
  * Can affect retry rate and delay
* Network buffer size in the AP
  * Bigger buffer → lower packet loss, but long delay

IEEE 802.11 in the Large: Observations at an IETF Meeting



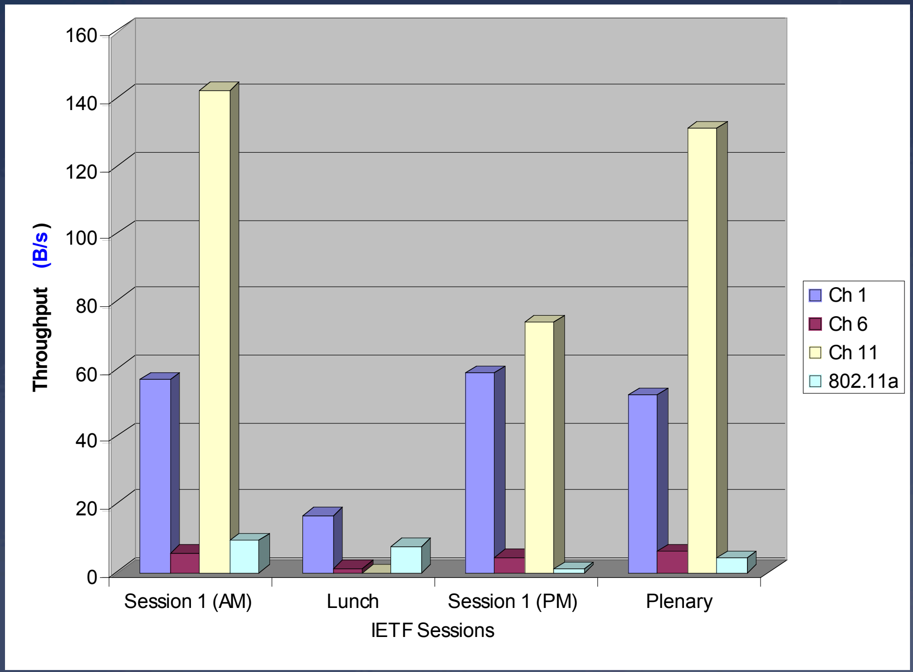
Observations at the IETF Meeting

Introduction

- 65^{th} IETF meeting
  - Dallas, TX (March 2006)
  - Hilton Anatole hotel
  - 1,200 attendees
- Data collection
  - 21^{st} - 23^{rd} for three days
  - 25GB data, 80 millions frames
- Wireless network environment
  - Many hotel 802.11b APs, 91 additional APs in 802.11a/b by IETF
  - The largest indoor wireless network measured so far
- We observed:
  - Bad load balancing
  - Too many useless handoffs
  - Overhead of having too many APs

Observations at the IETF Meeting
Load balancing

- Throughput per client



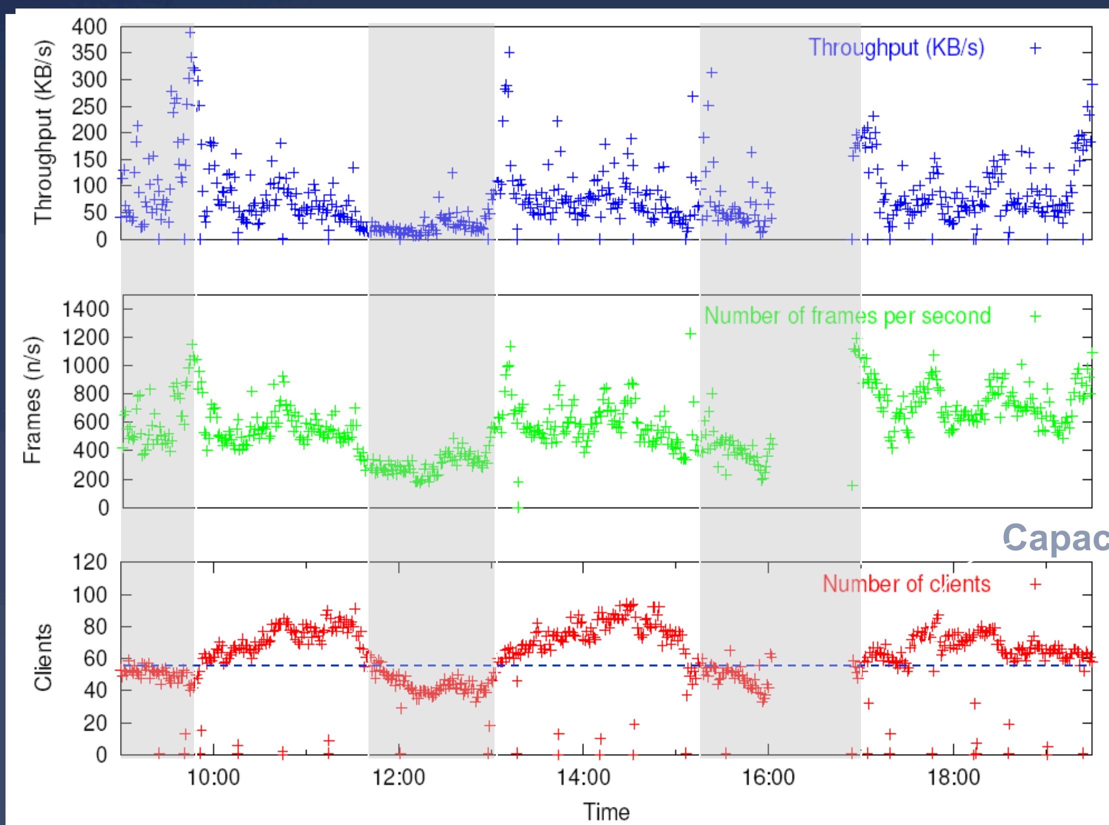
Average throughput per client in 802.11a/b

- No load balancing feature was used
- Client distribution is decided by the relative proximity from the APs
- Big difference in throughput among channels

Observations at the IETF Meeting

Load balancing

- Number of clients vs. Throughput

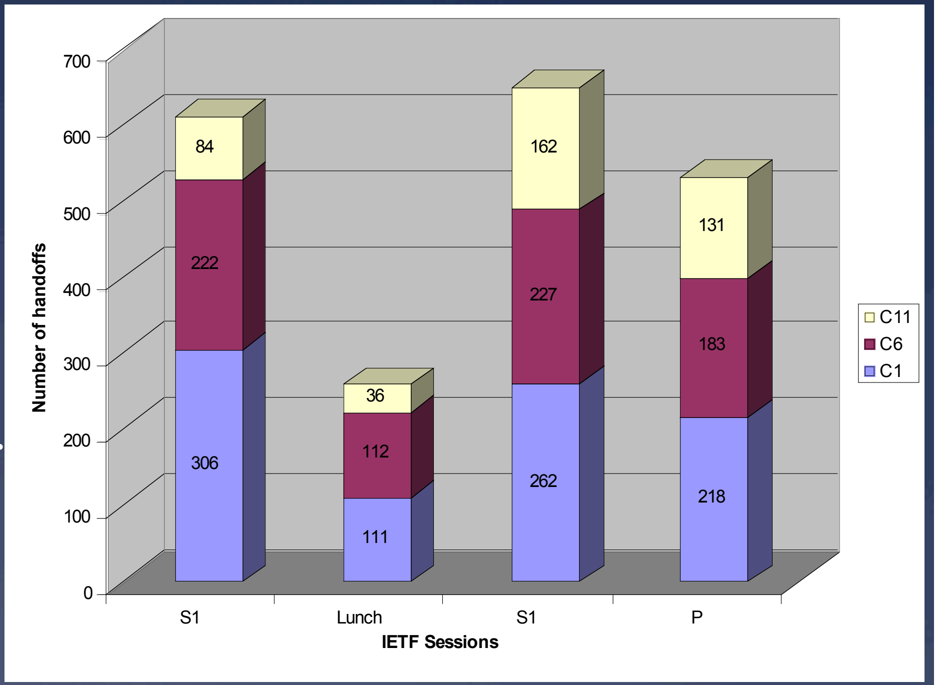


- ∗ Clear correlation between the number of clients and throughput
- ∗ The number of clients can be used for load balancing with low complexity of implementation, in large scale wireless networks

Capacity in the channel

Observations at the IETF Meeting Handoff behavior

- Too many handoffs are performed due to congestion
  - Distribution of session time : time (x) between handoffs
    - 0< x < 1 min : 23%
    - 1< x < 5 min : 33%
  - Handoff related frames took 10% of total frames.
- Too many inefficient handoffs
  - Handoff to the same channel : 72%
  - Handoff to the same AP : 55%

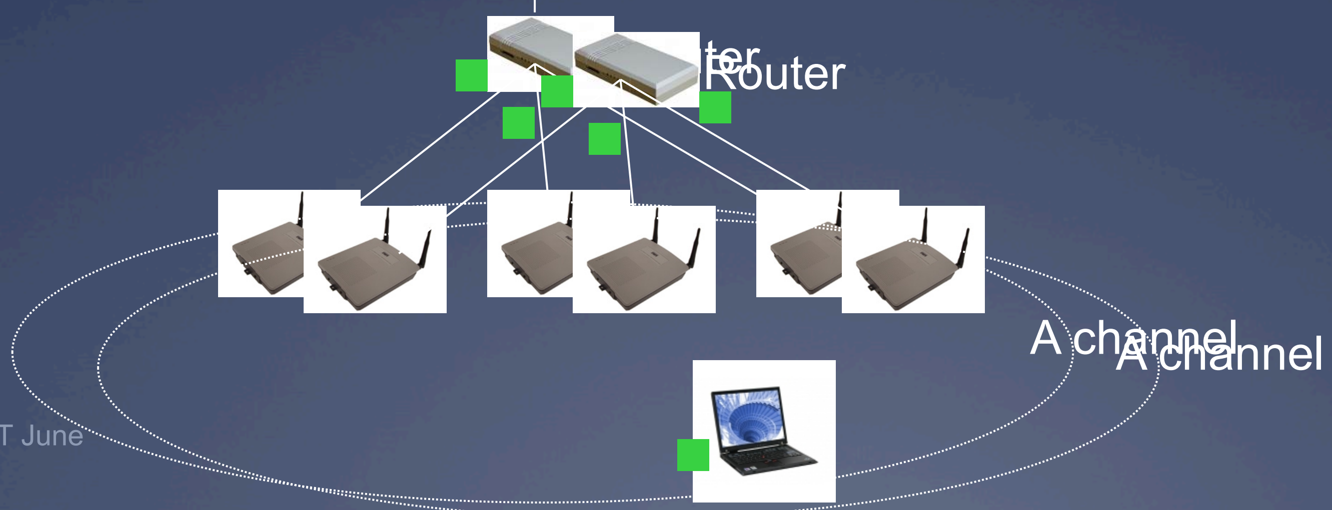


The number of handoff per hour in each IETF session

Observations at the IETF Meeting

Overhead of having multiple APs

- Overhead from replicated multicast and broadcast frames
  - All broadcast and multicast frames are replicated by all APs → increase traffic
  - DHCP request (broadcast) frames are replicated and sent back to each channel
  - Multicast and broadcast frames: 10%



Deploying dense 802.11 networks – conventional wisdom meets measurements

Andrea G. Forte and Henning Schulzrinne

Site Survey – Columbia University

[Google Map!]

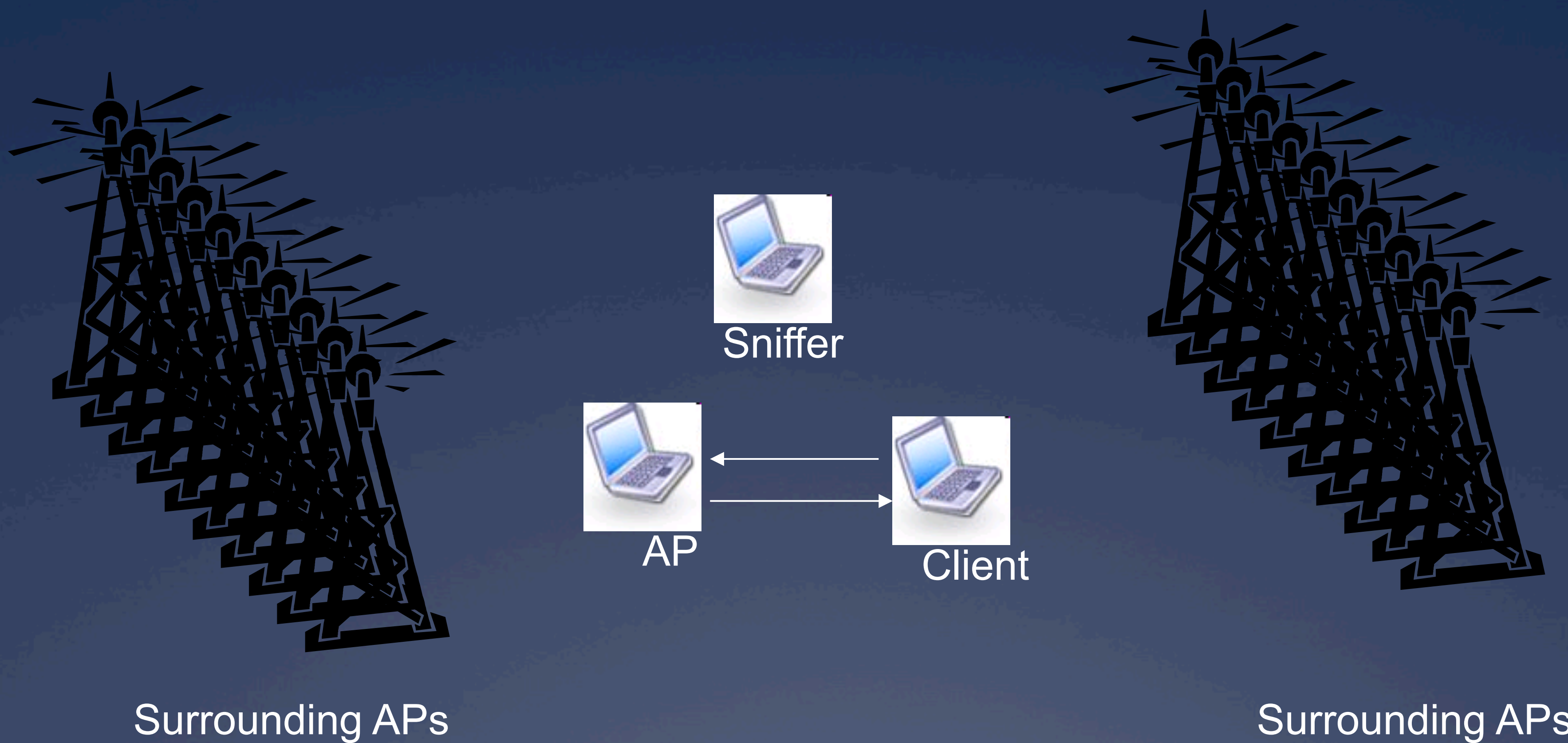
Site Survey – Columbia University

- Found a total of 668 APs
  - 338 open APs (49%)
  - 350 secure APs (51%)
  - Best signal: -54 dBm
  - Worst signal: -98 dBm

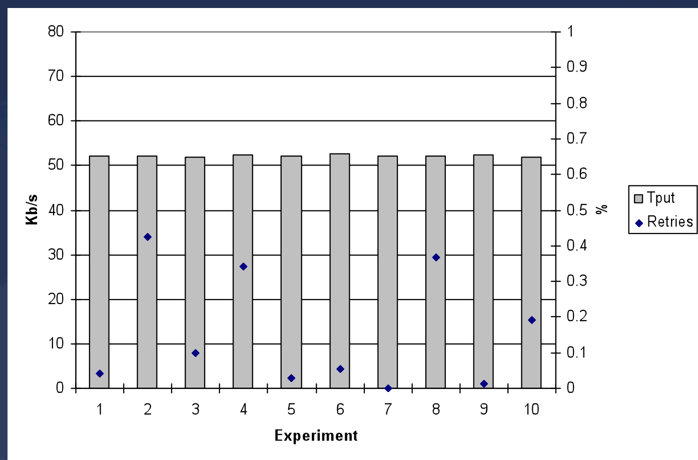
- Found 365 unique wireless networks
  - "private" wireless networks (single AP): 340
  - "public" networks (not necessarily open): 25
    - Columbia University: 143 APs
    - PubWiFi (Teachers College): 33 APs
    - COWSECURE: 12 APs
    - Columbia University – Law: 11 APs
    - Barnard College: 10 APs

Experiment 1

Experimental setup

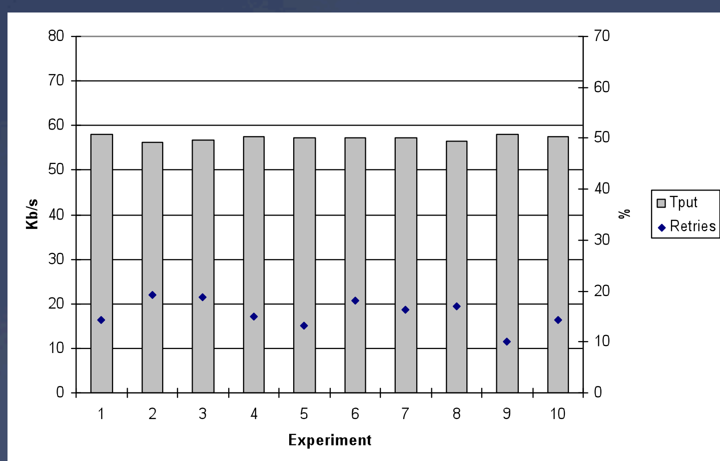


Using non-overlapping channels



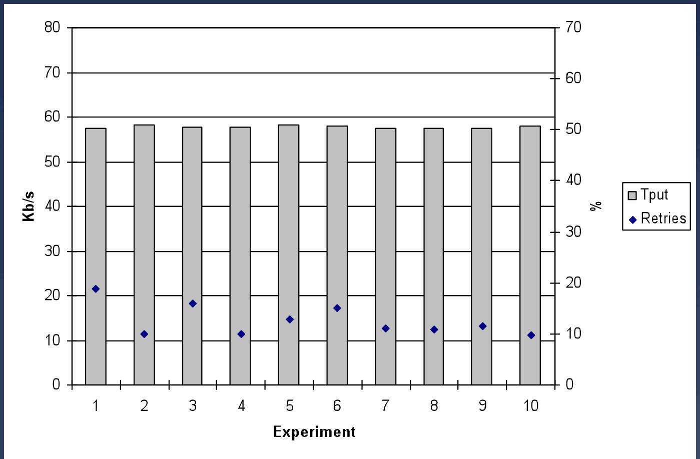
- Throughput and retry rate with no interference

→ Same for any channel



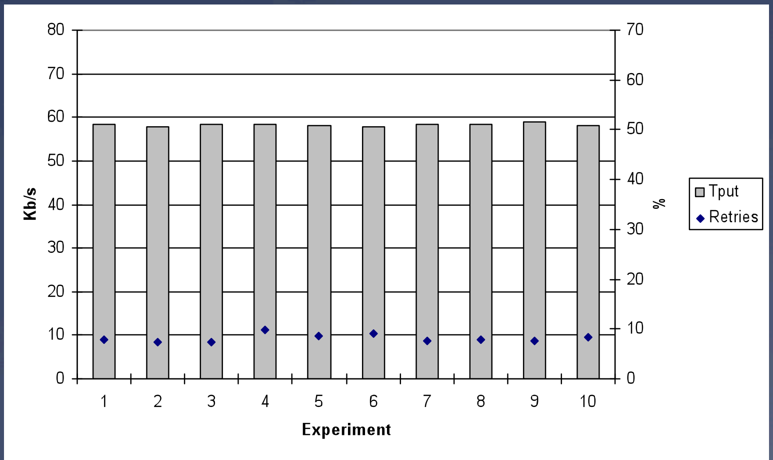
- Throughput and retry rate with interference on channel 1

Overlapping channels



- Throughput and retry rate with interference on channel 4

→ Better than channel 6



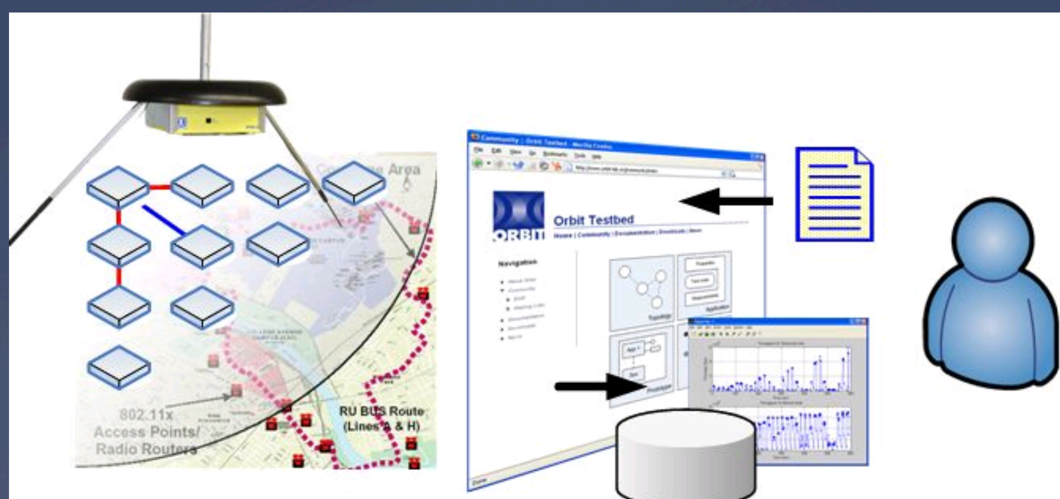
- Throughput and retry rate with interference on channel 8

→ Better than channel 6

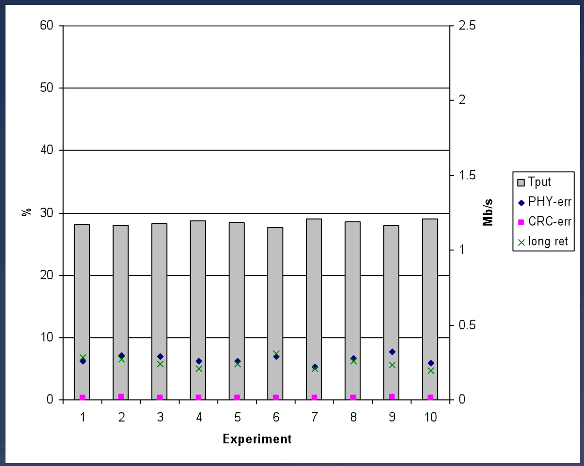
Experiment 2

Experimental setup

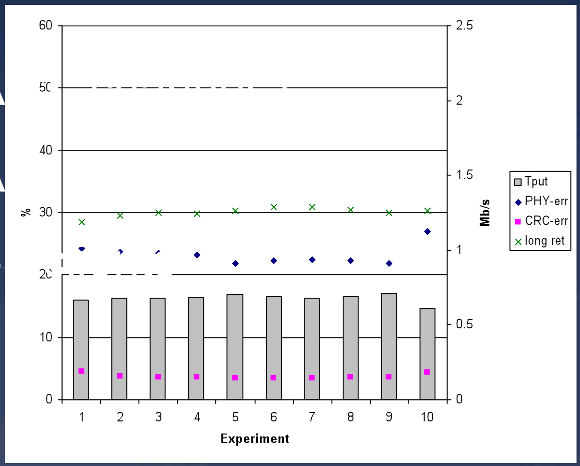
- ORBIT wireless test-bed
  - Grid of 20x20 wireless nodes
  - Used only maximum bit-rate of 11 Mb/s (no ARF)
  - G.711 CBR
  - Number of clients always exceeding the network capacity (CBR @ 11Mb/s → 10 concurrent calls)



Non-overlapping channels

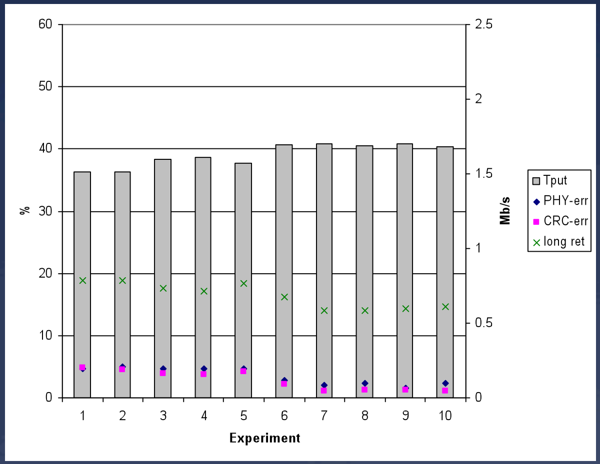


- A
- A
- 4

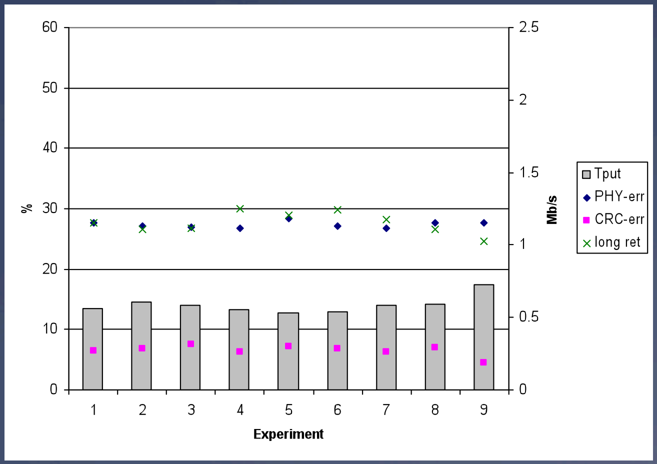


- AP1 and AP2 use channel 1
- 43 clients

Overlapping channels



- AP1: channel 1
- AP2: channel 4
- 67 clients



- AP1 and AP2 use ch. 4
- 67 clients

Results

* When using two APs on the same channel
  * Throughput decreases drastically
  * Physical-error rate and retry rate increase

* Using two APs on two overlapping channels performs much better than using the same non-overlapping channel

  → Do not deploy multiple APs on the same non-overlapping channels

  → USE OVERLAPPING CHANNELS!

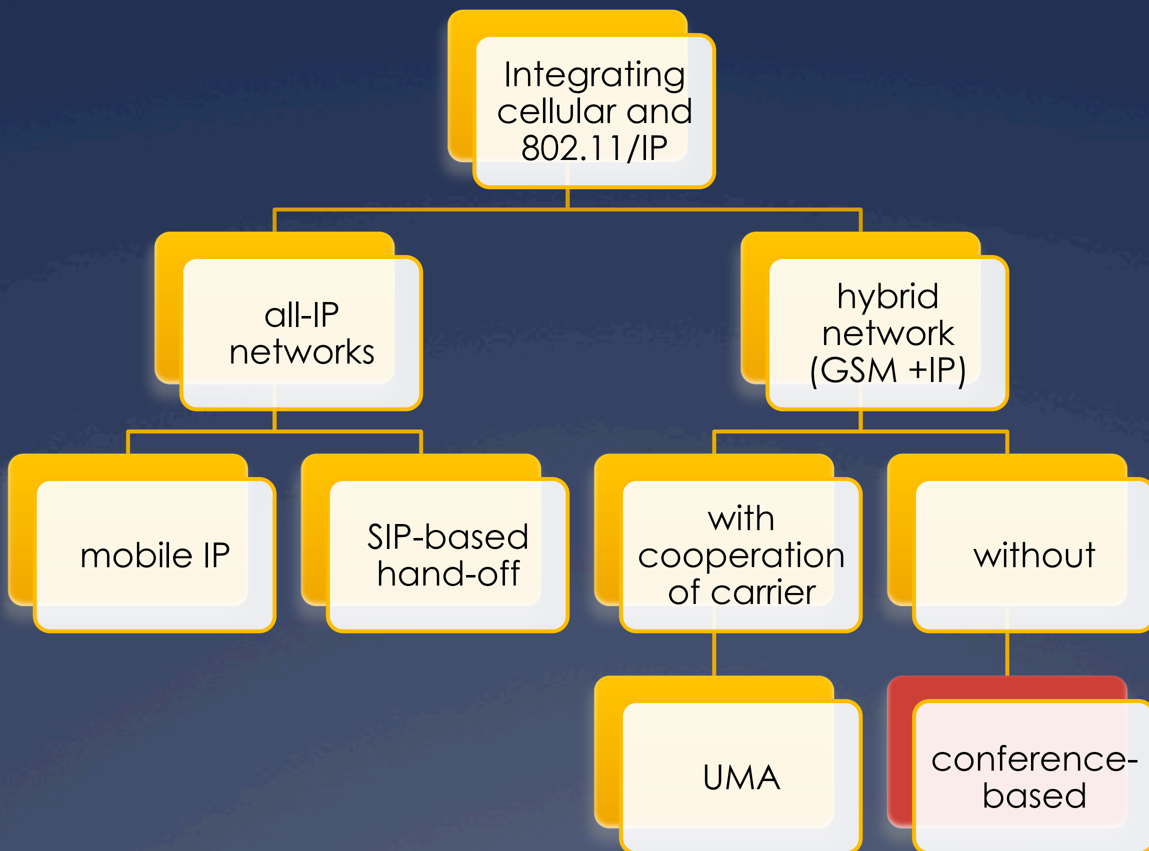
Channel selection algorithm

- Using overlapping channels does not reduce performance
  - Use at least channels 1, 4, 8 and 11
- Do not deploy multiple APs on the same non-overlapping channels
- Using two APs on the same channel worse than using a single AP!
  - Just increasing the number of APs does not help
- Impact on automated channel assignment mechanisms

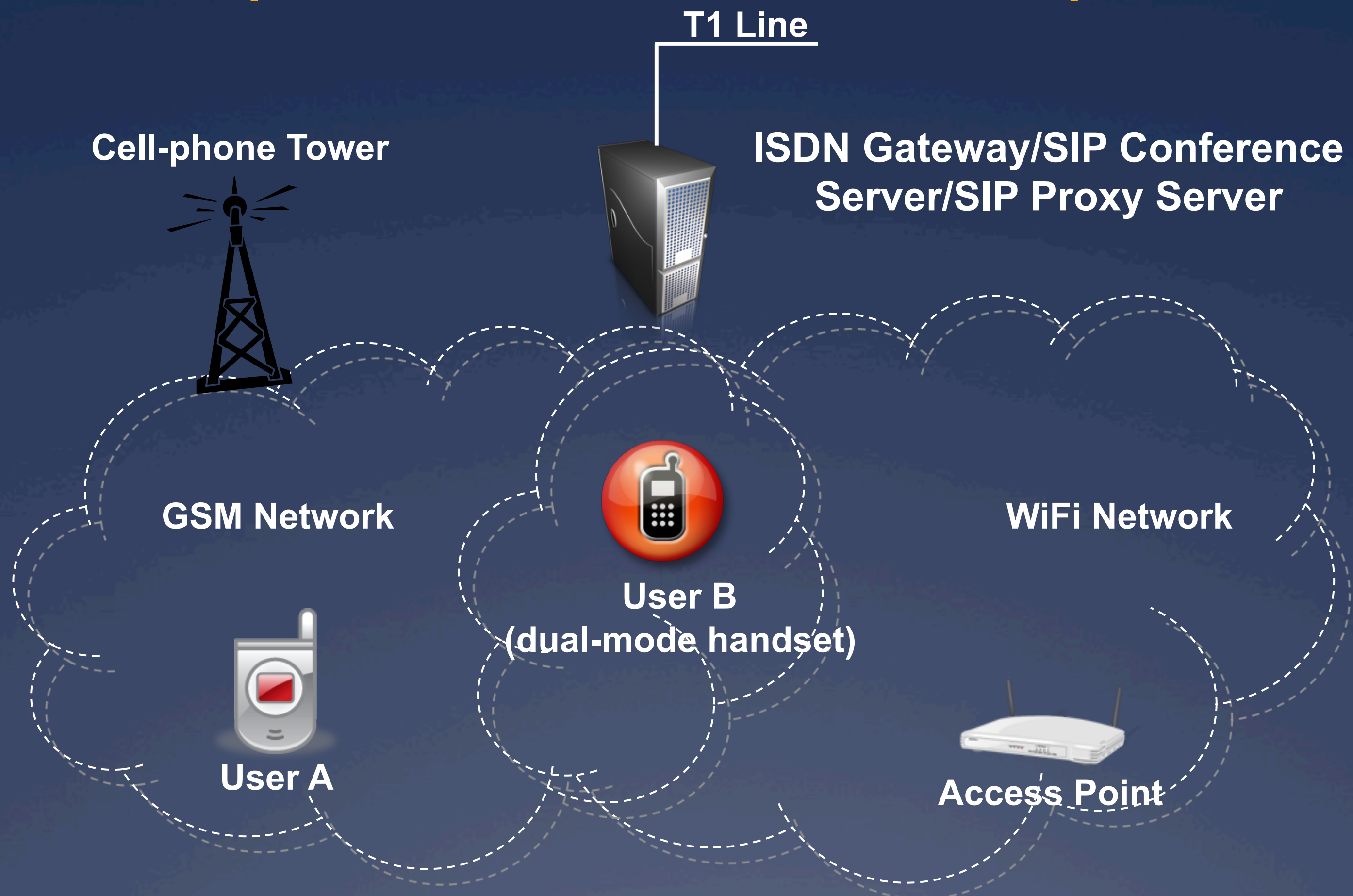
Integrating cellular and 802.11

ComSoc DLT June 2009

Options

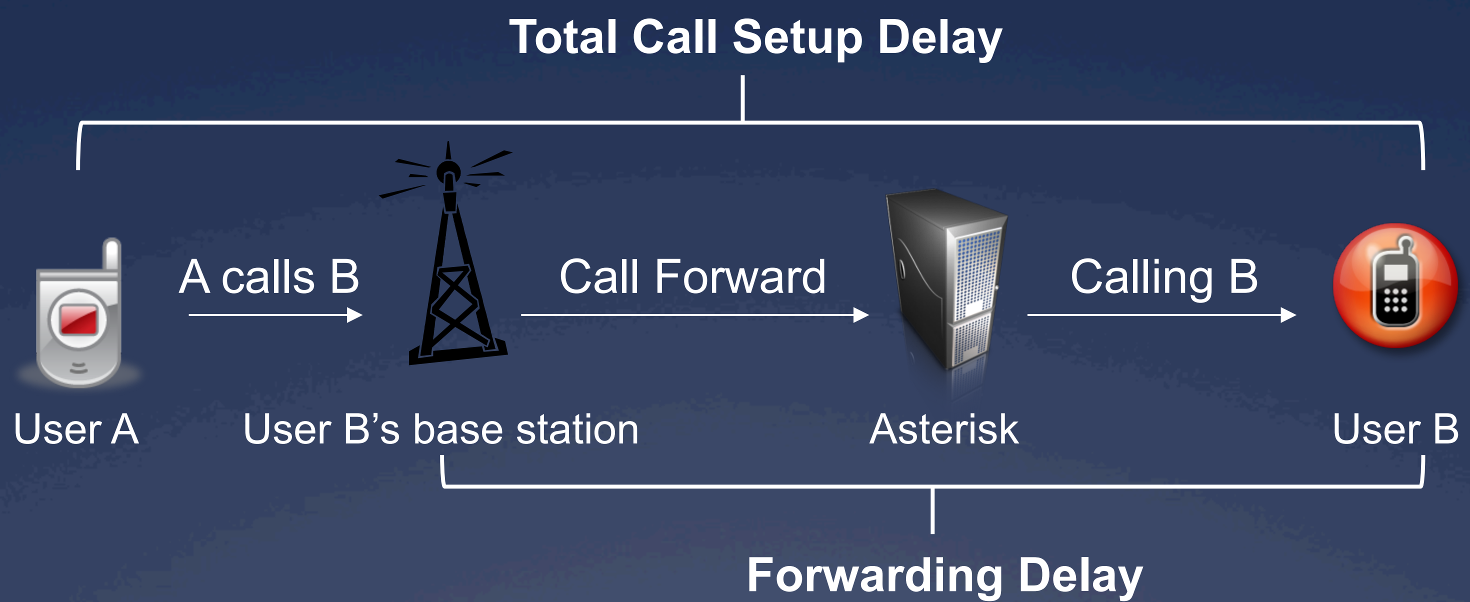

Integrating cellular and 802.11/IP
all-IP networks
hybrid network (GSM +IP)
mobile IP
SIP-based hand-off
with cooperation of carrier
without
UMA
conference-based

Experimental Setup



- Dual-mode handset
- IP interface: X-Lite client
- GSM interface: Nokia cellphone

Experiments



Type of call (A → B)	Forwarding delay	Call-setup delay
Cell-to-cell *	6.7 s	9.6 s
Cell-to-IP **	3.1 s	6.2 s

* Call set-up delay for B →A is higher because of DTMF: ~15 s

** Call set-up delay for B → A: ~6.9 s

Conclusions

- VoIP requires multi-faceted re-engineering of 802.11
  - Hand-off
    - focused on local, client-based approaches
    - need systematic comparison with infrastructure approaches
      - pro-active probably most promising
      - needs discovery, L3 remoting of AA operations
  - QoS
    - About 20% utilization - but most WLANs will carry mixed traffic
    - Admission control remains challenging - need NSIS or similar

More information & papers

- http://www.cs.columbia.edu/IRT/wireless
- http://www.cs.columbia.edu/~andreaf
- http://www.cs.columbia.edu/~ss202